UNIVERSITÀ CATTOLICA DEL SACRO CUORE

Dottorato di ricerca in Scienze della Persona e della Formazione – indirizzo "Persona, Sviluppo, Apprendimento. Prospettive Epistemologiche, Teoriche e Applicative"

Ciclo XXXVI

S.S.D. M/PSI-04

Developing Social Cognition within Human-Robot Interaction

Coordinatore:

Ch.ma Prof.ssa Antonella Marchetti

Tesi di Dottorato di:

Laura Miraglia

N. Matricola: 5013987

Anno Accademico 2022/2023

Table of Contents

CHAPTER 1
GENERAL INTRODUCTION

General abstract

The field of human-robot interaction (HRI) has evolved significantly since the late 1990s, witnessing the development of social robots capable of engaging with humans on a social level. These robots, ranging from Roomba vacuum cleaner to Ishiguro's Geminoid, have found applications in domestic work, caregiving, education, therapy, entertainment, and customer service. Despite these advances, challenges remain in designing robots capable of understanding and responding to dynamic and ambiguous human social behaviors. Designing robots that understand and respond appropriately to dynamic, context-dependent, and often ambiguous social behaviors is still a significant hurdle. Achieving truly human-like social interaction remains a technical challenge with potential psychological implications.

In everyday interactions, the success of social exchanges depends critically on a uniquely human capacity: social cognition. Social cognition encompasses a range of psychological and neuropsychological processes that underlie individuals' ability to make sense of their own and others' behavior, including embodied cognition and social metacognition. Embodied cognition emphasizes the role of bodily experiences, perceptions, and actions in shaping cognitive processes. This perspective suggests that our understanding of concepts and the world is grounded in the way our bodies physically engage with and experience the environment. On the other hand, social metacognition refers to the ability to reflect on and understand one's own and others' mental states, allowing individuals to interpret social cues, understand the perspectives of others, and adjust their behavior accordingly.

The present work proposes a comprehensive investigation of HRI through an integrated approach to social cognition mediated by both embodied dimension and social metacognition, focusing on two macro objectives: determining the behaviors for humanoid robots to be considered social agents, and understanding the psychological mechanisms essential for successful human-robot interactions. The conceptual framework follows the chronological progression of a child's developmental stages – from motor resonance to higher-order cognitive skills – and provides a basis for an in-depth examination of social cognition in the context of HRI. In conclusion, the three studies discussed here explore key dimensions of HRI and social cognition. Chapter 2 examines action understanding in preschool children, focusing on the activation of mirroring motor chains. Chapter 3 explores ostensive communication, critical to early human development, and its potential application to humanoid robots, contributing to human-robot communication and trust. Chapter 4 applies the Theory of Mind to robot interactions by introducing the Attribution of Mental States Questionnaire (AMS-Q) for several non-human entities. This thesis represents a critical exploration of the intricate dimensions within HRI by delving into the realm of social cognition. As the chapters unfold, they offer insights into action understanding, ostensive communication, and Theory of Mind, each contributing to a holistic understanding of how robots can evolve into genuine social agents.

Introduction

The inherent sociability of humans is a cornerstone of existence, driving the intricate web of connections that define our lives. Since the dawn of civilization, humans have formed bonds and relationships, creating a tapestry of interactions that shape our understanding of ourselves and others. This fundamental need for sociability stems from our evolutionary heritage, as social interactions provide survival benefits, cooperative strategies, and shared resources. Beyond mere survival, our innate drive for companionship fulfills emotional, psychological, and cognitive needs, contributing to our well-being and personal growth (Tomasello, 2014). Social relationships have an impact on shaping our identity, influencing our behavior, and ultimately forming the narrative of our shared human experience.

Throughout the history of mankind, there has been an enduring fascination with human simulation. This intrinsic curiosity has taken various forms, from the statues of ancient Egypt over 2,000 years ago to the pioneering creation of the first androids in the 19th century. This journey culminates in the development of cognitive robots with the ability to plan, reason, manipulate, and the subsequent addition of 'social skills'. The need for these robots to have social skills depends on the specific requirements of their application domains. For example, robots working in a factory may not require social skills. Conversely, a robot that delivers mail in an office and regularly interacts with customers will require the integration of social skills to facilitate smoother human-robot interactions. Similarly, a robot that serves as a home companion for the elderly or assists people with disabilities will require an extensive repertoire of social skills to ensure its acceptance by humans. The study of social skills in robots can be a rewarding exercise in the study of mechanisms of social cognition. Broadly defined, social cognition encompasses a set of neurocognitive processes that underlie the individual's ability to make sense of their own and others' behavior (Kunda, 1999). Delving deeper into this field reveals the mechanisms enabling us to decipher the thoughts, emotions, and intentions of others while shaping our own social responses. This multifaceted skill relies on the development of a variety of abilities, ranging from decoding perceptual social cues, such as faces and emotional expressions, to making inferences about the mental or emotional states of others, to making decisions that are consistent with social norms, the well-being of others, and specific social contexts. Ultimately, social cognition refers to individuals' efforts to interpret human actions. Consider the following passage from the novel *"I Baffi"* by the French author Emmanuel Carrère.

> *"But this was precisely what surprised him: she didn't seem surprised at all, not even for a second, just enough time to compose herself and assume a natural air. He had looked at her closely the moment he had seen her, as she placed the record back in its case: she hadn't blinked, hadn't changed her expression, nothing, as if she had all the time in the world to prepare for the show that awaited her. Of course, one could argue that he had warned her; Agnès had even said, laughing, that it would be a good idea. It was impossible to imagine that she had taken it seriously, that she had gone grocery shopping thinking to herself, « When I see him, I'll have to act as if nothing's wrong. » On the other hand, if she hadn't expected it, her composure was even less credible."*[1] (E. Carrère, *I Baffi, 1986,* page 17).

[1] *"Ma era proprio questo a stupirlo: non era affatto sembrata sorpresa, neanche per un secondo, il tempo di ricomporsi, di assumere un'aria naturale. L'aveva guardata bene nell'istante in cui l'aveva visto, mentre rimetteva il disco nella custodia: non aveva battuto ciglio, non aveva cambiato espressione, niente, come se avesse avuto tutto il tempo di prepararsi allo spettacolo che l'aspettava. Certo, si sarebbe potuto obiettare che l'aveva avvisata, Agnès aveva perfino detto, ridendo, che sarebbe stata una buona idea. Impossibile immaginare che l'avesse preso sul serio, che avesse fatto la spesa dicendosi: si sta radendo i baffi, quando lo vedo dovrò fare come*

> [Personal translation of the text from Italian into English; original text in footnote].

The novel centers on a recurring theme: the husband's decision to shave off his mustache and the wife's adamant denial that he ever had one. This peculiar motif weaves its way through the narrative, leaving us to ponder whether the mustache-less protagonist is descending into madness or is the victim of a plot concocted by his wife to drive him mad. In this excerpt, the character discusses the understanding of his wife's behavior based on her actions, emotional expressions, words, and gestures. We witness the husband's desperate efforts to decipher his wife's behavior by observing and interpreting these cues to deduce underlying thoughts, feelings, and intentions in a relentless pursuit to catch her in the act and reveal her elaborate prank. The protagonist constructs his interpretations of his wife's actions by meticulously decoding the social perceptual cues he scrutinizes in her, such as her facial expressions, as well as inferences about her mental and emotional states and the prior knowledge he possesses about her. Both husband and wife rely on their own observations and interpretations of each other's actions to make judgments and delve into each other's thoughts, feelings, and intentions. These intricate psychological processes, woven throughout the novel, serve as a reflection of the protagonist's social cognition, which is essential for successful relationships. This excerpt highlights the intricate mechanisms at play in human social interactions, including the ability to perceive social cues, share experiences, infer the thoughts and emotions of others, and manage emotional responses. These skills are central to understanding the emotions, perspectives, and mental states of individuals, and to interpreting, explaining, and predicting behaviors. Notably, these cognitive abilities not only facilitate interaction between humans but also extend to non-living agents such as social robots (Nass & Moon, 2000). Social robots are specifically designed to serve as human assistants and companions, playing roles in healthcare and everyday life. Their functions include fostering collaboration in the workplace or providing assistance in various settings such as homes, airports, and supermarkets. In these dynamic environments, social robots will be expected to engage in reciprocal social interactions. While these robots may be able to perform useful actions in complex and social contexts, the critical challenge is whether people will accept them and willingly engage with them. This remains a key issue for the successful integration of robots into a human-centered perspective.

se niente fosse. D'altra parte, se non se l'aspettava, il sangue freddo di cui aveva dato prova, era ancora meno credibile". (E. Carrère, *I Baffi, 1986,* page 17).

The Human-Robot Interaction

Human-robot interaction (HRI) is a young field currently undergoing a phase of unrest. Since the development of KISMET, one of the first social robots, at the MIT Media Lab in the late 1990s, significant progress has been made toward engineering robots capable of engaging humans on a social level. The new millennium has witnessed the emergence of increasingly autonomous humanoids, such as ASIMO in 2000, one of the pioneering humanoids capable of walking on legs and feet (Hirai et al., 1998; Sakagami et al., 2002). The concept of social robots encompasses fully or partially automated technologies that can understand, interpret, and respond to human social cues, including gestures, facial expressions, speech, and even emotional states. Social robots often take the form of humanoids resembling the human body and connect to online platforms to enhance their functionalities, incorporating voice and emotion recognition, human face recognition, and other artificial intelligence-related capabilities. These robots are designed to interact with humans in a more natural and human-like manner, enabling assistance in various tasks such as caregiving for the elderly, education, therapy, entertainment, and customer service (Breazeal et al., 2016; Broadbent et al., 2009).

Over the years, research in HRI has led to the development of different robots with varying degrees of social capabilities. Some of these robots can maintain eye contact, comprehend and generate speech, display emotional expressions, and adjust their behavior based on user responses. Examples include SoftBank’s NAO and Pepper, as well as Luxai’s QT Robot, each with its unique set of capabilities and applications. Social robots have transitioned from industrial settings (e.g., factories) to public domains (e.g., hospitality, healthcare, and educational) and even domestic environments (Fan et al., 2017; Robinson et al., 2014; Severinson-Eklundh et al., 2003; Young et al., 2009). The application of educational robotics holds immense potential, especially considering the challenges posed by aging populations in most countries, which have implications for healthcare, family structures, and financial markets. Social robots have the potential to address some of the challenges in service settings, particularly if they can exhibit human-like behavior (Čaić et al., 2019; Di Dio et al., 2020a; Manzi, Di Dio, et al., 2021). For these reasons, the understanding of human-robot interaction is a challenge that is more relevant than ever. Social interaction has been recognized as one of the ten grand challenges facing the field of robotics (Yang et al., 2018). This challenge stems from the complexity of human social behaviors, characterized by nuances, context, and cultural differences. Designing robots capable of seamlessly engaging in social interactions requires a multidisciplinary approach that combines expertise in robotics, psychology, cognitive science, neuroscience, and more. Despite the progress achieved, the field of HRI continues to face challenges and uncertainties. Firstly, human social interaction is highly dynamic, context-dependent, and often ambiguous. Designing robots that can understand and appropriately respond to these complexities remains a challenge. Moreover, ethical issues related to privacy, autonomy, and the potential displacement of human jobs need to be addressed. Despite advances in AI and robotics, achieving truly human-like social interaction remains a technical challenge. Robots may struggle to comprehend subtle cues or adapt to unexpected situations. Additionally, there are psychological concerns: understanding how humans perceive and react to robots is crucial for successful interactions. An example is the *uncanny valley effect*, where robots that closely resemble humans but not entirely, can evoke discomfort (MacDorman & Ishiguro, 2006; Mori, 1970; Mori et al., 2012). Just think of Geminoid or Erika from ATR Hiroshi Ishiguro Laboratories. Therefore, it has been suggested that the design of a social robot should balance *humanness*, which facilitates social interaction, with a degree of *robot-ness*, to prevent users from developing false expectations about the robot's emotional capabilities (DiSalvo et al., 2002). This claim is closely related to the concept of anthropomorphism, namely the tendency to attribute human characteristics, both physical and psychological, to non-human agents (Duffy, 2003; Epley et al., 2007). Anthropomorphism profoundly shapes the dynamics of human-robot interactions, influencing how individuals perceive and engage with robotic entities. The tendency to ascribe human-like characteristics to robots can have a significant impact on these interactions, both positive and challenging. Positively, anthropomorphic design elements such as humanoid appearance or facial expressions increase user engagement and acceptance (van Pinxteren et al., 2019). The familiarity of anthropomorphic robots often makes them more approachable and encourages more natural interactions (Złotowski et al., 2015). In addition, these design features contribute to the effective communication of social signals, allowing users to interpret the robot's intentions through familiar cues such as gestures and expressions (Fiore et al., 2013). On the other hand, mismatches between anthropomorphic expectations and a robot's actual capabilities can lead to communication breakdowns and misunderstandings during interactions, leading to potential disappointment (de Sá Siqueira et al., 2023; Ye et al., 2019). Design choices need to be carefully considered, taking into account how anthropomorphic features can contribute to relatability without creating unrealistic expectations. The improvement of robot-human interaction would benefit significantly from the inclusion of human-like motor resonance in the robot's behaviors (Chaminade & Cheng, 2009). Moreover, if robots are intended to serve as social companions, they should activate brain mechanisms of social cognition, akin to those triggered when humans interact with each other (Wiese et al., 2017). These mechanisms include joint attention (Charman et al., 1997; Moore et al., 2014), perspective-taking (Samson et al., 2010; Tversky & Hard, 2009; Zwickel, 2009), action understanding (Brass et al., 2007; Gallese et al., 1996; Rizzolatti et al., 1996; Rizzolatti & Craighero, 2004), turn-taking (Knapp et al., 2013), and mentalizing (Baron-Cohen, 1997; Frith & Frith, 2006; Marchetti et al., 2018). Before brain areas involved in social cognitive processing are activated to infer what others think, feel, and intend, it is necessary to perceive others as intentional beings. Thus, alongside human-like appearance and movement, the fact that robots may be perceived as intentional beings is a key factor in treating them as social entities.

In human interactions, people typically attribute minds to other people by detecting social signals that indicate another person's ability to perceive, feel, and intend (Meltzoff, 2007); this tendency extends to non-human agents when they induce perceptions of intentionality, leading people to attribute mental states to these agents (Marchetti et al., 2018; Waytz, Cacioppo, et al., 2010). The growing sense that humanoid robots possess "minds of their own" has important implications for robotic design, highlighting the importance of robots displaying human-like minds alongside their human-like appearances (Di Dio et al., 2020b; Manzi et al., 2020; Manzi, Massaro, et al., 2021; Waytz et al., 2014). The research in HRI initially emphasized the development of competence traits by enhancing robots' cognitive resources and improving their functionalities (Pineau et al., 2003). More recently, however, the field of robotics studies has been increasingly attuned to recognizing the importance of warmth traits arising from enriched affective resources, such as the incorporation of meaningful gestures like eye contact (Johnson et al., 2014). This nuanced shift underscores the recognition that effective human-robot interaction is as much about emotional resonance as it is about technical competence. The mechanisms of social cognition may activate when humans interact with cognitively and affectively endowed social robots, leading them to evaluate these robots despite their non-human status.

The field of neuroscience, meanwhile, offers fascinating insights. Through techniques like functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and electromyography (EMG), researchers delve into human information processing during robot interactions. Neuroscientific findings shed light on the cognitive processes at play when interacting with humanoid robots, with the potential to positively shape these interactions (Chaminade & Cheng, 2009; Henschel et al., 2020; Krach et al., 2008; Manzi, Di Dio, et al., 2021). For instance, activation of the social brain network involved in action understanding and mentalizing enhances feelings of social connection, empathy, prosociality, and performance during joint action tasks. This offers an opportunity to identify parallels and differences in the processing of social cues between robots and humans. By investigating the mechanisms underlying the formation of impressions, trust, and emotional bonds with robots, this knowledge can guide the creation of robots that display behaviors and gestures aligning with human social processing, thereby eliciting more authentic responses.

In summary, the state-of-the-art research involving artificial agents has illuminated various aspects of human social cognition: (i) automatic processing of social visual information, including motor resonance, remains intact during observation of artificial agents in comparison to humans; (ii) in contrast, higher-order social cognitive processes are influenced by whether an agent is categorized as 'natural' or 'artificial'; (iii) people are highly sensitive, albeit often at an implicit level, to subtle cues that indicate a robot's humanity, encompassing both appearance and behavior (Wykowska et al., 2016). Consequently, human-like behavior in humanoid robots may trigger social cognitive mechanisms to the same extent as other human interaction partners. Thus, it becomes necessary to deepen two key aspects: first, the behaviors that humanoid robots should exhibit in order to be considered social agents, and second, the underlying psychological mechanisms that are essential, and possibly even sufficient, to facilitate human-robot interactions. The present work is an attempt to investigate these two macro-objectives.

Social Cognition

Social cognition refers to the ability to make sense of the behavior of others and encompasses social abilities that emerge at least as early as 14 months (Scott & Baillargeon, 2017) and continue to play a pivotal role for a lifetime (Slaughter & Perez-Zapata, 2014). The development of awareness and understanding of others' thoughts, feelings, and actions is essential for effective functioning in social environments. Its centrality in everyday life is evident in conditions where impaired social cognition results in adverse outcomes, such as neurodegenerative, neuropsychiatric, and neurodevelopmental conditions, as well as acute brain damage (Kennedy & Adolphs, 2012). However, where does this understanding originate, and how does it evolve? How does social cognition influence our understanding of other people and ourselves, and what are its implications for social interactions and relationships? The following chapters address these questions and more, offering an overview of the development and consequences of social cognition in various aspects of social understanding.

The perspective proposed here is that social cognition is influenced by two factors: *embodied cognition*, facilitated by a mirroring system that offers experiential insight into others' sensations, actions, emotions, and intentions, and *social metacognition*, involving deliberate reflection on another person's mental content using abstract concepts. Exploring these questions requires considering the historical context that has shaped our present understanding of social cognition. This contextual exploration serves as a bridge between the theoretical framework and the broader narrative encompassing the development and implications of social cognition in diverse social understanding domains. A brief historical overview offers insights into the foundations upon which contemporary perspectives are built. The study of social cognition includes various theoretical and research traditions. Although Vygotsky assumed that the child is naturally social so that any nonsocial, egocentric use of language comes later in development (Vygotsky, 1962), it was the Piagetian theory, focusing on the egocentric child, that dominated cognitive development studies in the 1960s and 1970s. The late 1970s marked a shift influenced by Flavell and colleagues (1968) and Kohlberg (1969), breaking from Piagetian egocentrism and emphasizing role-taking abilities, that is the child's increasing ability to recognize and make allowances for differences between self and other. Simultaneously, infant research revealed capacities for sharing and reciprocity in early life, such as Meltzoff's work (Meltzoff & Moore, 1977) on imitative sensitivity to the parent's facial and hand gestures, and Trevarthen's theory of early intersubjectivity (Trevarthen, 1998). Infants show a remarkable ability to synchronize their vocalizations and gestures with those of others (Gopnik & Meltzoff, 1997). This interactive coordination establishes a rhythmic connection between the infant and caregiver that contributes directly to intersubjective understanding. The early skills involved in primary intersubjectivity represent an immediate, non-mentalizing mode of interaction. These skills imply that even before considering another person's beliefs or desires, there is a perceptual understanding of what the other person feels, whether they are attentive to us or not, whether their intentions are friendly or not, and so on. Primary intersubjectivity is characterized by a shared bodily intentionality, a mutual awareness between the perceiving subject and the observed other. Around the age of one year, infants move from the person-to-person immediacy of primary intersubjectivity to secondary intersubjectivity (Trevarthen & Hubley, 1978). At this stage they engage in shared attention, gaining insights into the meanings and purposes of things. Notably, this shift does not involve adopting an intentional stance, instead, intentionality is perceived through the embodied actions of others.

Wimmer and Perner found that 4- and 5-year-old children made predictions about the actions of a doll character based on their beliefs and correctly anticipated the doll's actions (Perner & Wimmer, 1985; Wimmer, 1983). This finding surprised supporters of Piaget's egocentrism, demonstrating that even 4-year-olds could complete Wimmer and Perner's false-belief task successfully. Research on metacognition (Flavell, Green, & Flavell, 1995) and emotion understanding (Harris, 1989; Harris et al., 1981) invigorated social cognition research. Attachment theory highlighted infants' sensitivity to caregiver differences, further illustrating early social cognition. These studies reveal that young children are attuned to others and more competent than classical Piagetian theory suggests. This heightened cognitive receptivity forms the basis for meaningful social interactions and relationships, which involve psychological mechanisms. These mechanisms include processes that allow individuals to empathize with the emotions of others and to recognize their actions as reflections of inner mental states and intentions. These processes ultimately contribute to a deeper understanding of others and guide adjustments in one's own social behavior.

The structure of this conceptual framework follows the chronological progression of a child's developmental stages and provides a basis for the in-depth examination of social cognition in the chapters that follow. Subsequent sections explore embodied cognition, social learning, and Theory of Mind. This deliberate sequence, reflecting the natural progression from embodied cognition to advanced Theory of Mind, allows for a coherent and comprehensive exploration of social cognitive development.

Embodied Cognition

During the early stages of cognitive development, our comprehension of others is primarily rooted in non-verbal communication – gestures, glances, movements, paralanguage – that is, a form of language that seeks synchronization with the other person's body. These embodied mechanisms find their origin in intersubjectivity, particularly within the mother-child dyad. Intentionality is perceived in the embodied actions of others. The capacities involved in intersubjectivity suggest that we have a specific perceptual understanding of what others feel. We understand the actions of others as we understand our own actions, at the highest and most relevant pragmatic level possible, ignoring possible subordinate descriptions, ignoring interpretations in terms of beliefs, desires, or hidden mental states (Gallagher, 2007).

Consider attachment, for instance. A secure bond is based on the timely contingency and appropriateness of responses from the mother. This is where the first social exchanges, such as gaze sharing, attention, pointing, and then, the emergence of agency, take place. These interactions gradually evolve into mental representations of one's own states, initially rooted in the physical and emotional realm and fused with the states of others. As children explore the world, they build their own understanding through sensory experience, physical movement, and emotions. Information from the environment is absorbed in an undifferentiated manner, leading to the construction of sensory-motor-affective patterns where the child's own states and those of others tend to overlap. This process relies on the well-known mirroring mechanism, which allows us to internally simulate and directly experience the states of others (Rizzolatti et al., 1996). It is only as cognitive development unfolds, and plausibly in conjunction with the maturation of control and inhibition systems (Di Dio et al., 2023), that the child gradually acquires the ability to decentralize their own thoughts from those of others.

The MNS has been proposed to play an important role in shaping social cognition, as these embodied mechanisms are intricately intertwined with the experiential aspects of interpersonal relationships. This experiential dimension of social cognition bridges the gap between the multifaceted experiential knowledge we possess about our own bodily experiences and the insight we gain into the experiences of others (Gallese, 2008). The notion of embodiment has therefore been proposed to underlie social cognition phenomena. According to the framework of *embodied cognition* (Wood et al., 2016), all cognitive representations and operations are fundamentally grounded in their physical sensorimotor context. Even emotional processing involves the reactivation of motor programs and feelings associated with their direct sensorimotor experiences (Winkielman et al., 2008).

The success of social interactions hinges on the ability to understand others' behavior and decode their intentions. This crucial ability is partially mediated by a specific mechanism known as the Mirror Neuron System (MNS; Gallese et al., 2006; Rizzolatti et al., 1996), a group of specialized neurons that fire both when an organism acts and when it observes the same action being performed by another organism (Rizzolatti et al., 1996). This network is considered fundamental to various action-related social functions, including action recognition, imitation, intention understanding, and empathy (Gallese et al., 2004; Rizzolatti, 2005; Rizzolatti & Craighero, 2004). The intentions of others can be inferred from the context in which the action is performed and even from the manner in which the object is manipulated. However, it has been suggested that, in complex situations, inferential processing may complement the mirror mechanism in intention understanding (Gallese et al., 2002; Rizzolatti et al., 2001). For example, when John sees Mary grasping a cup of coffee, he recognizes the goal of Mary's actions by observing her hand moving toward the cup. This recognition of the goal (grasping the cup) stems from the mirror system's ability to identify the action (grasp). More complex, the MNS also plays a role in coding the broader intention underlying a motor act (Iacoboni et al., 2005). Questions like "Why is Mary grasping a cup of coffee? Is she going to drink it or give it to someone else? Or perhaps throw it away?" reflect the complex nature of intention understanding and its central role in social interactions. When an individual initiates a movement, such as a grasp, they have a clear intention in mind, for example, drinking. This intention is present at the beginning of the action and is mirrored in each step of the sequence. A single action can arise from very different intentions – Mary could grasp the cup in order to drink from it or wash it. The direct matching between the observed action and its motor representation in the observer's brain not only reveals the action itself (grasping) but also the underlying intention. As a result, premotor mirror areas are implicated in action recognition, whereas the parietal lobe, originally associated solely with action organization, is actually involved in understanding intentions (Fogassi et al., 2005). While not the primary focus of this thesis, it is valuable to note that the mirror system extends beyond comprehending goals and intentions, and allows the observer to understand, on the basis of how an action is performed, the psychological state of the agent (Di Cesare et al., 2014, 2015, 2016). These aspects of action comprehension have been termed by Stern "vitality forms" (Stern, 1985, 2010). Furthermore, the MNS – in concert with activity in the anterior insula and amygdala – plays a role in deciphering the emotional states of others (Carr et al., 2003; Dapretto et al., 2006).

Our *social brain* (Adolphs, 1999) contains areas that are specifically activated during interactions with other social entities. Action understanding, a critical aspect of social interactions, is based on *motor resonance* (Gallese et al., 1996), that is shared representations that are activated both when action is executed and when a similar action is observed in others. Given the importance of action understanding in human-robot interaction, it is essential to examine whether motor resonance, a key element of this process, is exclusive to human agents or extends to robotic agents. A set of studies consistently showed that motor resonance can indeed be induced by robot agents; however, the extent of this resonance appears to depend on features such as physical appearance (Chaminade et al., 2007) and motion kinematics (Gazzola et al., 2007). These studies suggest that while robots have the potential to activate motor resonance, the degree of activation is contingent upon specific conditions. Taken together, these promising findings provide a solid empirical basis for investigating the activation of brain areas when interacting with robotic agents. For instance, an interesting area of research involves investigating motor anticipation, a mechanism enabling the observer to effectively predict the agent's intention. Over the past decade, fMRI studies (Hamilton & Grafton, 2008; Iacoboni et al., 2005) and EMG studies (Cattaneo et al., 2007; Fogassi et al., 2005) have shown that motor acts are organized in chains that encode specific intentional actions (e.g., grasping to drink or grasping to place). Most interestingly, these chains may be activated by merely observing the first motor act of a sequence, providing the observer with an internal copy of the agent's entire future action before execution, effectively prefiguring the agent's intention. This aspect of embodied cognition is poorly investigated in children and, consequently, whether and how the actions and intentions of a robot agent are anticipated is, to the author's knowledge, unknown.

Social learning

Crucial to social cognition is the social signals, such as facial expressions and eye gaze, which play an important role in gathering insights about the world. As outlined in the session above, from early infancy, humans are finely attuned to these social cues, guiding our interactions and helping us decipher the intentions and emotions of others. Infants can learn a great deal simply by observing others and the social signals they convey. Social signals serve as valuable sources of information about the world surrounding them. For instance, expressions of disgust or fear communicate the presence of something in the immediate environment that should be approached with caution or avoided. Through a mother's facial expression, an infant can discern whether an object is nice or nasty. Furthermore, social interactions often lead to the unconscious phenomenon of mirroring, in which individuals involuntarily synchronize their movements (Chartrand & Bargh, 1999) and engage in a chain of stimuli and responses triggered by social cues (Frith, 2008). For example, a fearful facial expression elicits a social response of fear without either the sender or the addressee being aware that they are exchanging signals. This phenomenon is rooted in the Mirror Neuron System (MNS), which effectively replicates observed actions in the observer's motor system, making it an ideal mechanism for learning through imitation (Buccino et al., 2004; Nishitani & Hari, 2000). Thus, it is reasonable to assume that systems connecting an individual's own actions and experiences with those of others play an important role in social cognition.

As opposed to learning through observation, social signals serve also distinct communicative functions, such as signaling danger or pointing out interesting things (Bandura, 1986), and form the foundation of social learning (Frith & Frith, 2007). A notable example of this lies in ostensive cues, which serve as indicators of intentional communication, setting the stage for instruction. A typical scenario unfolds as follows: a mother establishes eye contact with her infant, directing the child's attention toward an object through gaze and gestures, and subsequently labeling it. However, the initial signal - the mother's eye contact - serves a dual purpose. Beyond capturing the infant's attention, it acts as an ostensive cue, a signal to which the child is particularly sensitive (Fonagy & Allison, 2014). Additional ostensive signals encompass the calling of one's name (Mandel et al., 1995) and the use of 'motherese' during maternal communication with infants (Fernald, 1985). The ability to interpret social signals is vital not only for survival in a complex and ever-changing environment, but also for the development of empathy, emotional intelligence, and effective communication skills (Brothers, 1989) and for acquiring new and relevant knowledge about a referent object (Egyed et al., 2013). By observing and internalizing these cues, individuals learn the unwritten rules of their culture and adapt their behavior accordingly. The transmission and acquisition of shared knowledge are intricately tied to our developmental and evolutionary perspectives. According to Frith (2008), this connection hinges on our ability to interpret social signals as deliberate and informative. This concept is rooted in the Natural Pedagogy theory (Csibra & Gergely, 2009; Csibra & György, 2006), and according to the theory, infants are naturally inclined to learn generic knowledge when adults address them through ostensive communication, as it makes children feel recognized as subjects (Fonagy & Allison, 2014). Developmentally, ostensive communication encourages the establishment of epistemic trust in caregivers. It paints caregivers as benevolent, cooperative, and trustworthy sources of cultural information that facilitate the rapid learning of shared knowledge without the need to critically examine its validity or relevance (Fonagy & Allison, 2014). Infants' reliance on caregiver claims is likely to have far-reaching implications for social cognition. Thus, this perspective implies a complex level of understanding that underlies the acquisition of knowledge through intentional communication, referred to as metacognition. This learning process requires introspection into one's own cognitive framework, particularly as it relates to the exchange of signals, and involves a metacognitive mechanism in which both the communicator and the addressee recognize the communicative nature of social signals.

In the present thesis, the question is whether ostensive cues acted upon by a robotic agent can elicit effects that go beyond mere attentional arousal, potentially leading to the sharing of relevant information that facilitates the acquisition of new knowledge. Previous research has explored fundamental mechanisms of social cognition, such as gaze and joint attention using humanoid robots. For instance, studies have provided insights into the effect of a robot's gaze accompanied by verbalizations on infants' object learning. In particular, findings from studies by Okumura et al. (2013, 2020) suggests that the combination of robot gaze and verbalization is important in the design of robot agents from which infants can learn. However, little is actually known about the effect of ostensive cues on the transmission of relevant information. This thesis aims to address this gap by investigating how these cues contribute to the effective communication of new knowledge, even when conveyed by a robot.

Theory of Mind

Since others' behavior is not entirely predictable, the success of social interactions depends on the ability to decode mental states, such as beliefs, desires, intentions, goals, experiences, intentions, and emotions of other people. Interpreting others' behavior in terms of mental states becomes a pivotal step in predicting future actions. Taking into account the mental states of others (monitoring) and using this information to predict behavior (control) refers to the ability of mentalizing, which is an intrinsic disposition that leads to the development of a Theory of Mind (ToM). ToM is based on the awareness that individuals possess mental states, intentions, and motivations that guide their behaviors and may diverge from one's own (Perner & Wimmer, 1985; Premack & Woodruff, 1978; Wellman et al., 2001). A fundamental distinction, although with partial overlap, pertains to cognitive ToM, the capacity to attribute cognitive and epistemic states, *versus* affective ToM, the capacity to ascribe affective states. Furthermore, the act of mentalizing, encompassing the representation of others' thoughts, desires, feelings, and intentions, differs from empathetically comprehending and automatically sharing affective states. Mentalization holds the potential to influence social perception through top-down mechanisms, allowing us to perceive and appropriately respond to others' emotions, intentions, and behaviors based on context (Arioli et al., 2018).

The ontogeny of ToM faculties closely mirrors the maturation of other brain functions. The acquisition of ToM is profoundly social and inextricably intertwined with broader knowledge about individuals and their lives (Garfield et al., 2001). During the initial months of life, infants predominantly engage in dyadic, face-to-face interactions. As early as the second month, they fixate on the eyes and mouths of others and become distressed when interactions cease due to a still face (Tronick et al., 1978). By the end of the first year, infants achieve joint attention and participate in triadic person-object-person interactions (Striano & Reid, 2006). Around 24 months of age, children participate in "pretend play", demonstrating their understanding that others are engaged in pretense. At 3 years old, children grasp that mere physical contact with a box does not reveal its contents. By age 4, children succeed in the false belief task, recognizing when someone holds a false belief about the world. At this age, children can recognize that someone's behavior is determined by their beliefs and knowledge, even if they are clearly false. By age 9, they identify faux pas and understand what might hurt others' feelings, and remain unspoken (Baron-Cohen, 2008). Evidence suggests that the social brain continues to mature during adolescence, characterized by social change, heightened self-awareness, more intricate peer relationships, and an improved understanding of others (Blakemore, 2008).

Reflecting on the relations between our own or other's behaviors and knowledge is an example of explicit metacognition. Metacognition is the cognitive process involved in thinking about thinking. Thinking about our actions is an important feature of human mental life. We think about what actions to take and when to take them. Such introspection suggests that explicit metacognition determines human behavior and enables fruitful group interactions (Frith, 2012). The uniquely human capacity for collective intentionality captures the idea that humans do not simply act together but adopt a group-oriented stance when working together, creating a collective that shares intentions and knowledge. This attitude underpins collaborative behavior and the sharing of resources and information (Liszkowski et al., 2008; Rekers et al., 2011; Warneken et al., 2011). This extension of metacognition to social interactions is called *social metacognition*. Social metacognition refers to the ability to reflect on and understand one's own and others' mental states in social situations. It includes the ability to reflect on and regulate one's own social cognitive processes, such as beliefs, intentions, and emotions, as well as the ability to attribute similar mental states to others. Social metacognition is crucial for navigating complex social interactions, as it allows individuals to interpret social cues, understand the perspectives of others, and adjust their behavior accordingly. It plays a key role in various aspects of social cognition, contributing to effective communication, empathy, and social problem-solving. In human-human interactions, interpreting others' mental states empowers people as social agents (Frith & Frith, 2007). In the realm of adult social cognition, others are recognized not only as behaving "like me" and experiencing perceptions "like me," but also possessing a spectrum of mental states like beliefs, emotions, and intentions similar to my own (Garfield et al., 2001). In our social exchanges, explicit acts of interpretation are infrequent.

Instead, our comprehension of social situations is immediate, automatic, and nearly reflexive. This is feasible because social cognition hinges on the notion that you are "like me," distinct yet sufficiently similar to serve as a role model. Consequently, I become the interpreter of your intentions and behaviors.

Luckily for human-robot interaction, the perception of a mind is not confined to entities that actually possess a mind; it extends to agents like robots and avatars (Marchetti et al., 2018). The attribution of a mind is contingent upon various factors, including cognitive or motivational features of the perceiver, as well as the physical and behavioral traits of the perceived agent. This perceptual flexibility is evident in anthropomorphism, where the need for social connection increases the likelihood of ascribing human-like characteristics to non-human agents (Waytz et al., 2010a; 2010b). Conversely, instances of social rejection acts diminish the perceived humanity in others, a phenomenon known as dehumanization (Waytz et al., 2010b). Consequently, the study of the attribution of mind to robotic agents has attracted interest and revealed a gap in tools for the quantitative measurement of such attributions. This thesis aimed to address this gap through the validation and creation of a questionnaire specifically designed to assess the attribution of mental qualities to non-human agents, including robots. The development of this questionnaire was a central aim, recognizing the need for a systematic and validated questionnaire to capture aspects of how humans perceive and attribute mental qualities to non-human entities.

This Thesis

In the present thesis, I propose an integrated approach to social cognition mediated by both *embodied dimension* and *social metacognition*. Social metacognition involves explicitly thinking about the mental contents of others through abstract representations, while the embodied dimension employs a mirroring system that provides access to the sense of others' actions, emotions, and intentions experientially. The interplay between mentalizing ability and the mirror system suggests their complementary roles in comprehending others' behaviors, influenced by the presence of abstract information versus biological actions (Arioli et al., 2018). Within this framework, I attempted to explore the embodied and cognitive components that underlie the comprehension of robots' behavior and, therefore, may support and facilitate the interaction between humans and robots. More detailed, the primary focus of this thesis revolves around investigating the implications of social cognition within human-robot interaction across different stages of life. Human social requirements undergo significant shifts as individuals progress through different developmental phases. Furthermore, cognitive processes, that enable adaptation to novel social environments, evolve and deteriorate with age. Within this context, the thesis contends that an examination of social cognition in the context of perceiving and interacting with robots can yield profound insights into the human aspects of human-robot interaction. Analyzing how humans respond to robots in social contexts, interpreting robotic social cues, and identifying factors contributing to successful interactions can unveil underlying cognitive and emotional mechanisms. By emphasizing the universal dimensions of warmth and competence in social cognition (Fiske et al., 2007), this thesis aims to provide a social psychological perspective on human-robot interactions. The inclusion of human social cognitive processes in comparing human-human and human-robot interactions seeks to explore the potential complementary roles that social humanoid robots in the future. Finally, the investigation of the psychological mechanisms involved in HRI provides, in turn, interesting insight into the essentially social cognition.

The subsequent chapters will delve into the research conducted throughout my Ph.D. journey, focusing on the concepts introduced in this general introduction, particularly embodied cognition and action understanding in infancy (Chapter 2), Theory of Mind (Chapter 3), and social learning and epistemic trust (Chapter 4). The overarching theme binding these chapters is the concept of social cognition examined within human-robot interaction. The aim is to identify and comprehend the psychological and psychophysiological mechanisms supporting this form of interaction. Furthermore, the exploration of human-robot interactions offers a window into studying human social cognition, granting deeper insights into the cognitive processes that underlie everyday interactions between individuals.
The forthcoming chapters are briefly introduced below.

Chapter 2 delves into action understanding by exploring the activation of motor chains with mirror properties in preschool-aged children. The motor system’s role in comprehending others’ actions and intentions is vital, conveying essential communicative signals (Iacoboni & Dapretto, 2006). Mirror neurons, initially identified in monkeys’ premotor cortex, play a crucial role in action understanding. These neurons fire both when an individual performs an action and when they observe the same action executed by others. Beyond understanding actions, mirror neurons facilitate imitation, emotion recognition, and the discernment of intentions behind actions. Specifically, the chapter investigates whether a similar chain organization is present in typically developing preschool children aged 3 to 5 years, utilizing

electromyographic recordings (EMG). This project aims to determine, empirically and for the first time, whether the capacity to predict and comprehend the intentions underlying observed actions experientially is already developed at a young age.
This exploration is part of a larger project to investigate the embodied components of human-robot relationships. The study represents a first step that sheds light on one of the processes by which young children may engage with robotic entities.

Chapter 3 explores ostensive communication, a form of intentional communication where the sender provides explicit signals to direct the addressee's attention to a specific object or action. This type of communication plays a crucial role in early human development, establishing epistemic trust and facilitating secure attachments. The chapter examines whether this mechanism persists into adulthood, offering insights into shared knowledge through ostensive communication. This research project also investigates whether humanoid robots can activate the process of shared knowledge through ostensive communication, rendering the robot a trustworthy communication partner capable of conveying pertinent information.

Chapter 4 explores Theory of Mind (ToM), the ability to comprehend both one's own and others' mental states, allowing the prediction and interpretation of behaviors based on this mental representation. This ability is not limited to human relationships but extends to interactions with robots (Marchetti et al., 2018), facilitated by the human tendency to anthropomorphize. This chapter introduces the Attribution of Mental States questionnaire (AMS-Q), designed to outline the attribution of mental states across various entities, including robotic agents. The research aims to validate the AMS-Q through three studies: refining questionnaire items in a preliminary study, analyzing the psychometric properties of the questionnaire through exploratory factor analysis in Study 1, and examining the multidimensionality of AMS-Q through confirmatory factor analysis and its reliability and validity in Study 2.

In conclusion, this thesis aims to unravel the intricate dimensions of social cognition within human-robot interaction across the lifespan. By probing the mechanisms underlying human-robot interactions, it seeks to enhance the understanding of human social cognition and the cognitive processes shaping everyday interpersonal interactions.

References

Adolphs, R. (1999). The Human Amygdala and Emotion. *The Neuroscientist*, *5*(2), 125–137. https://doi.org/10.1177/107385849900500216

Arioli, M., Crespi, C., & Canessa, N. (2018). Social Cognition through the Lens of Cognitive and Clinical Neuroscience. *BioMed Research International*, *2018*, 4283427. https://doi.org/10.1155/2018/4283427

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory* (pp. xiii, 617). Prentice-Hall, Inc.

Baron-Cohen, S. (1997). *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press.

Baron-Cohen, Simon. (2008). The evolution of brain mechanisms for social behavior. *Foundations of Evolutionary Psychology*, 331–352.

Blakemore, S.-J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, *9*(4), 267–277. https://doi.org/10.1038/nrn2353

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating Action Understanding: Inferential Processes versus Action Simulation. *Current Biology*, *17*(24), 2117–2121. https://doi.org/10.1016/j.cub.2007.11.057

Breazeal, C., Dautenhahn, K., & Kanda, T. (2016). Social Robotics. In B. Siciliano & O. Khatib (Eds.), *Springer Handbook of Robotics* (pp. 1935–1972). Springer International Publishing. https://doi.org/10.1007/978-3-319-32552-1_72

Broadbent, E., Stafford, R., & MacDonald, B. (2009). Acceptance of Healthcare Robots for the Older Population: Review and Future Directions. *International Journal of Social Robotics*, *1*(4), 319–330. https://doi.org/10.1007/s12369-009-0030-6

Brothers, L. (1989). A biological perspective on empathy. *American Journal of Psychiatry*, *146*(1), 10–19.

Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H.-J., & Rizzolatti, G. (2004). Neural Circuits Underlying Imitation Learning of Hand Actions. *Neuron*, *42*(2), 323–334. https://doi.org/10.1016/S0896-6273(04)00181-3

Čaić, M., Mahr, D., & Oderkerken-Schröder, G. (2019). Value of social robots in services: Social cognition perspective. *Journal of Services Marketing*, *33*(4), 463–478. https://doi.org/10.1108/JSM-02-2018-0080

Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences*, *100*(9), 5497–5502. https://doi.org/10.1073/pnas.0935845100

Cattaneo, L., Fabbri-Destro, M., Boria, S., Pieraccini, C., Monti, A., Cossu, G., & Rizzolatti, G. (2007). Impairment of actions chains in autism and its possible role in intention understanding. *Proceedings of the National Academy of Sciences*, *104*(45), 17825–17830. https://doi.org/10.1073/pnas.0706273104

Chaminade, T., & Cheng, G. (2009). Social cognitive neuroscience and humanoid robotics. *Journal of Physiology-Paris*, *103*(3), 286–295. https://doi.org/10.1016/j.jphysparis.2009.08.011

Chaminade, T., Hodgins, J., & Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience*, *2*, 206–216. https://doi.org/10.1093/scan/nsm017

Charman, T., Swettenham, J., Baron-Cohen, S., Cox, A., Baird, G., & Drew, A. (n.d.). *Infants With Autism: An Investigation of Empathy, Pretend Play, Joint Attention, and Imitation*.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910. https://doi.org/10.1037/0022-3514.76.6.893

Cinzia Di Dio, Davide Massaro, & Antonella Marchetti. (2023). Inibizione e Teoria della Mente. *Giornale italiano di psicologia*, *1*, 157–164. https://doi.org/10.1421/106929

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

Csibra, G., & György, G. (2006). Social learning and social cognition: The case for pedagogy.

Attention and Performance, *21*, 249–274.

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., & Iacoboni, M. (2006). Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, *9*(1), Article 1. https://doi.org/10.1038/nn1611

De C. Hamilton, A. F., & Grafton, S. T. (2008). Action Outcomes Are Represented in Human Inferior Frontoparietal Cortex. *Cerebral Cortex*, *18*(5), 1160–1168. https://doi.org/10.1093/cercor/bhm150

de Sá Siqueira, M. A., Müller, B. C. N., & Bosse, T. (2023). When Do We Accept Mistakes from Chatbots? The Impact of Human-Like Communication on User Experience in Chatbots That Make Mistakes. *International Journal of Human–Computer Interaction*, *0*(0), 1–11. https://doi.org/10.1080/10447318.2023.2175158

Di Cesare, G., Di Dio, C., Marchi, M., & Rizzolatti, G. (2015). Expressing our internal states and understanding those of others. *Proceedings of the National Academy of Sciences*, *112*(33), 10331–10335. https://doi.org/10.1073/pnas.1512133112

Di Cesare, G., Di Dio, C., Rochat, M. J., Sinigaglia, C., Bruschweiler-Stern, N., Stern, D. N., & Rizzolatti, G. (2014). The neural correlates of 'vitality form' recognition: An fMRI study: This work is dedicated to Daniel Stern, whose immeasurable contribution to science has inspired our research. *Social Cognitive and Affective Neuroscience*, *9*(7), 951–960. https://doi.org/10.1093/scan/nst068

Di Cesare, G., Valente, G., Di Dio, C., Ruffaldi, E., Bergamasco, M., Goebel, R., & Rizzolatti, G. (2016). Vitality Forms Processing in the Insula during Action Observation: A Multivoxel Pattern Analysis. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00267

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

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P., Massaro, D., & Marchetti, A. (2020b). Come I bambini pensano alla mente di un robot. Il ruolo dell'attaccamento e della Teoria della Mente nell'attribuzione di stati mentali a un agente robotico [How children think about the robot's mind. The role of attachment and Theory of Mind in the attribution of mental states to a robotic agent]. *Sistemi Intelligenti*, *1*, 41–56.

DiSalvo, C. F., Gemperle, F., Forlizzi, J., & Kiesler, S. (2002). All robots are not created equal: The design and perception of humanoid robot heads. *Proceedings of the Conference on Designing Interactive Systems Processes, Practices, Methods, and Techniques - DIS '02*, 321. https://doi.org/10.1145/778712.778756

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3–4), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

Egyed, K., Király, I., & Gergely, G. (2013). Communicating Shared Knowledge in Infancy. *Psychological Science*, *24*(7), 1348–1353. https://doi.org/10.1177/0956797612471952

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Fan, J., Bian, D., Zheng, Z., Beuscher, L., Newhouse, P. A., Mion, L. C., & Sarkar, N. (2017). A Robotic Coach Architecture for Elder Care (ROCARE) Based on Multi-User Engagement Models. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(8), 1153–1163. https://doi.org/10.1109/TNSRE.2016.2608791

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*(2), 181–195. https://doi.org/10.1016/S0163-6383(85)80005-9

Fiore, S., Wiltshire, T., Lobato, E., Jentsch, F., Huang, W., & Axelrod, B. (2013). Toward understanding social cues and signals in human–robot interaction: Effects of robot gaze and proxemic behavior. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00859

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Flavell, John H. (1968). *The Development of Role-Taking and Communication Skills in Children.*

Flavell, John H, Green Frances L., & Flavell Eleanor R. (1995). The development of children's knowledge about attentional focus. *Developmental Psychology*, *31*(4), 706.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal Lobe: From Action Organization to Intention Understanding. *Science*, *308*(5722), 662–667. https://doi.org/10.1126/science.1106138

Fonagy, P., & Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy*, *51*(3), 372–380. https://doi.org/10.1037/a0036505

Frith, C. D. (2008). Social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1499), 2033–2039. https://doi.org/10.1098/rstb.2008.0005

Frith, C. D. (2012). The role of metacognition in human social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1599), 2213–2223. https://doi.org/10.1098/rstb.2012.0123

Frith, C. D., & Frith, U. (2006). The Neural Basis of Mentalizing. *Neuron*, *50*(4), 531–534. https://doi.org/10.1016/j.neuron.2006.05.001

Frith, C. D., & Frith, U. (2007). Social Cognition in Humans. *Current Biology*, *17*(16), R724–R732. https://doi.org/10.1016/j.cub.2007.05.068

Gallagher, S. (2007). Social cognition and social robots. *Pragmatics & Cognition*, *15*(3), 435–453. https://doi.org/10.1075/pc.15.3.05gal

Gallese, V. (2008). Embodied Simulation: From Mirror Neuron Systems to Interpersonal Relations. In G. Bock & J. Goode (Eds.), *Novartis Foundation Symposia* (pp. 3–19). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470030585.ch2

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593–609. https://doi.org/10.1093/brain/119.2.593

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

https://doi.org/10.1016/j.tics.2004.07.002

Gallese, Vittorio, Fadiga, Luciano, Fogassi, Leonardo, & Rizzolatti, Giacomo. (2002). 17 Action representation and the inferior parietal lobule. *The Cogn. Anim.*, 451-461.

Gallese, Vittorio, Morris N. Eagle, & Paolo Migone. (2006). La simulazione incarnata: I neuroni specchio, le basi neurofisiologiche dell'intersoggettività e alcune implicazioni per la psicoanalisi. *La Simulazione Incarnata*, 1000–1038.

Garfield, J. L., Peterson, C. C., & Perry, T. (2001). Social Cognition, Language Acquisition and The Development of the Theory of Mind. *Mind and Language*, *16*(5), 494–541. https://doi.org/10.1111/1468-0017.00180

Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *NeuroImage*, *35*(4), 1674–1684. https://doi.org/10.1016/j.neuroimage.2007.02.003

Gopnik, A., & Meltzoff, A. N. (1997). *Words, Thoughts, and Theories*. MIT Press.

Harris, P. L. (1989). *Children and emotion: The development of psychological understanding* (pp. viii, 243). Basil Blackwell.

Harris, P. L., Olthof, T., & Terwogt, M. M. (1981). Children's Knowledge of Emotion. *Journal of Child Psychology and Psychiatry*, *22*(3), 247–261. https://doi.org/10.1111/j.1469-7610.1981.tb00550.x

Henschel, A., Hortensius, R., & Cross, E. S. (2020). Social Cognition in the Age of Human–Robot Interaction. *Trends in Neurosciences*, *43*(6), 373–384. https://doi.org/10.1016/j.tins.2020.03.013

Hirai, K., Hirose, M., Haikawa, Y., & Takenaka, T. (1998). The development of Honda humanoid robot. *Proceedings. 1998 IEEE International Conference on Robotics and Automation (Cat. No.98CH36146)*, *2*, 1321–1326. https://doi.org/10.1109/ROBOT.1998.677288

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews. Neuroscience*, *7*(12), 942–951. https://doi.org/10.1038/nrn2024

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the Intentions of Others with One's Own Mirror Neuron System. *PLoS Biology*, *3*(3), e79. https://doi.org/10.1371/journal.pbio.0030079

Johnson, D. O., Cuijpers, R. H., Juola, J. F., Torta, E., Simonov, M., Frisiello, A., Bazzani, M., Yan, W., Weber, C., Wermter, S., Meins, N., Oberzaucher, J., Panek, P., Edelmayer, G., Mayer, P., & Beck, C. (2014). Socially Assistive Robots: A Comprehensive Approach to Extending Independent Living. *International Journal of Social Robotics*, *6*(2), 195–211. https://doi.org/10.1007/s12369-013-0217-8

Kennedy, D. P., & Adolphs, R. (2012). The social brain in psychiatric and neurological disorders. *Trends in Cognitive Sciences*, *16*(11), 559–572. https://doi.org/10.1016/j.tics.2012.09.006

Knapp, M. L., Hall, J. A., & Horgan, T. G. (2013). *Nonverbal Communication in Human Interaction*. Cengage Learning.

Kohlberg, L., & Kramer, R. (1969). Continuities and Discontinuities in Childhood and Adult Moral Development. *Human Development*, *12*(2), 93–120. https://doi.org/10.1159/000270857

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can Machines Think? Interaction and Perspective Taking with Robots Investigated via fMRI. *PLOS ONE*, *3*(7), e2597. https://doi.org/10.1371/journal.pone.0002597

Kunda, Z. (1999). *Social Cognition: Making Sense of People*. MIT Press.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, *108*(3), 732–739. https://doi.org/10.1016/j.cognition.2008.06.013

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, *7*(3), 297–337. https://doi.org/10.1075/is.7.3.03mac

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' Recognition of the Sound

Patterns of Their Own Names. *Psychological Science*, *6*(5), 314–317. https://doi.org/10.1111/j.1467-9280.1995.tb00517.x

Manzi, F., Di Dio, C., Di Lernia, D., Rossignoli, D., Maggioni, M. A., Massaro, D., Marchetti, A., & Riva, G. (2021). Can You Activate Me? From Robots to Human Brain. *Frontiers in Robotics and AI*, *8*, 633514. https://doi.org/10.3389/frobt.2021.633514

Manzi, F., Massaro, D., Di Lernia, D., Maggioni, M. A., Riva, G., & Marchetti, A. (2021). Robots Are Not All the Same: Young Adults' Expectations, Attitudes, and Mental Attribution to Two Humanoid Social Robots. *Cyberpsychology, Behavior, and Social Networking*, *24*(5), 307–314. https://doi.org/10.1089/cyber.2020.0162

Manzi, F., Peretti, G., Di Dio, C., Cangelosi, A., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020). 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., Manzi, F., Itakura, S., & Massaro, D. (2018). Theory of Mind and Humanoid Robots From a Lifespan Perspective. *Zeitschrift Für Psychologie*, *226*(2), 98–109. https://doi.org/10.1027/2151-2604/a000326

Meltzoff, A. N. (2007). The 'like me' framework for recognizing and becoming an intentional agent. *Acta Psychologica*, *124*(1), 26–43. https://doi.org/10.1016/j.actpsy.2006.09.005

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of Facial and Manual Gestures by Human Neonates. *Science*, *198*(4312), 75–78. https://doi.org/10.1126/science.198.4312.75

Miraglia, L., Di Dio, C., Manzi, F., Kanda, T., Cangelosi, A., Itakura, S., Ishiguro, H., Massaro, D., Fonagy, P., & Marchetti, A. (2023). Shared Knowledge in Human-Robot Interaction (HRI). *International Journal of Social Robotics*. https://doi.org/10.1007/s12369-023-01034-9

Moore, C., Dunham, P. J., & Dunham, P. (2014). *Joint Attention: Its Origins and Role in Development*. Psychology Press.

Mori, M. (1970). The Uncanny Valley. *IEEE Spectrum*.

Mori, M., MacDorman, K., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nishitani, N., & Hari, R. (2000). Temporal dynamics of cortical representation for action. *Proceedings of the National Academy of Sciences*, *97*(2), 913–918. https://doi.org/10.1073/pnas.97.2.913

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., & Itakura, S. (2013). Can infants use robot gaze for object learning?: The effect of verbalization. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, *14*(3), 351–365. https://doi.org/10.1075/is.14.3.03oku

Okumura, Y., Kanakogi, Y., Kobayashi, T., & Itakura, S. (2020). Ostension affects infant learning more than attention. *Cognition*, *195*, 104082. https://doi.org/10.1016/j.cognition.2019.104082

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

Pineau, J., Montemerlo, M., Pollack, M., Roy, N., & Thrun, S. (2003). Towards robotic assistants in nursing homes: Challenges and results. *Robotics and Autonomous Systems*, *42*(3–4), 271–281. https://doi.org/10.1016/S0921-8890(02)00381-0

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

Rekers, Y., Haun, D. B. M., & Tomasello, M. (2011). Children, but Not Chimpanzees, Prefer to Collaborate. *Current Biology*, *21*(20), 1756–1758. https://doi.org/10.1016/j.cub.2011.08.066

Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomy and Embryology*, *210*(5–6), 419–421. https://doi.org/10.1007/s00429-005-0039-z

Rizzolatti, G., & Craighero, L. (2004). THE MIRROR-NEURON SYSTEM. *Annual Review of Neuroscience*, *27*(1), 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131–141. https://doi.org/10.1016/0926-6410(95)00038-0

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*(9), 661–670. https://doi.org/10.1038/35090060

Robinson, H., MacDonald, B., & Broadbent, E. (2014). The Role of Healthcare Robots for Older People at Home: A Review. *International Journal of Social Robotics*, *6*(4), 575–591. https://doi.org/10.1007/s12369-014-0242-2

Sakagami, Y., Watanabe, R., Aoyama, C., Matsunaga, S., Higaki, N., & Fujimura, K. (2002). The intelligent ASIMO: System overview and integration. *IEEE/RSJ International Conference on Intelligent Robots and System*, *3*, 2478–2483. https://doi.org/10.1109/IRDS.2002.1041641

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255–1266. https://doi.org/10.1037/a0018729

Scott, R. M., & Baillargeon, R. (2017). Early False-Belief Understanding. *Trends in Cognitive Sciences*, *21*(4), 237–249. https://doi.org/10.1016/j.tics.2017.01.012

Severinson-Eklundh, K., Green, A., & Hüttenrauch, H. (2003). Social and collaborative aspects of interaction with a service robot. *Robotics and Autonomous Systems*, *42*(3–4), 223–234. https://doi.org/10.1016/S0921-8890(02)00377-9

Slaughter, V., & Perez-Zapata, D. (2014). Cultural Variations in the Development of Mind Reading. *Child Development Perspectives*, *8*(4), 237–241. https://doi.org/10.1111/cdep.12091

Stern, D. N. (1985). *The Interpersonal World of the Infant: A View from Psychoanalysis and Developmental Psychology*. (New York: Basic Books).

Stern, D. N. (2010). *Forms of Vitality: Exploring Dynamic Experience in Psychology, the Arts, Psychotherapy, and Development*. OUP Oxford.

Striano, T., & Reid, V. M. (2006). Social cognition in the first year. *Trends in Cognitive Sciences*, *10*(10), 471–476. https://doi.org/10.1016/j.tics.2006.08.006

Tomasello, M. (2014). The ultra-social animal. *European Journal of Social Psychology*, *44*(3), 187–194. https://doi.org/10.1002/ejsp.2015

Trevarthen, C., & Hubley, P. (2023). *Secondary intersubjectivity: Confidence, confiding, and acts of meaning in the first year. In A. Lock (Ed.), Action, gesture and symbol: The emergence of language*.

Trevarthen, C. (1998). *The concept and foundations of infant intersubjectivity*.

Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). The Infant's Response to Entrapment between Contradictory Messages in Face-to-Face Interaction. *Journal of the American Academy of Child Psychiatry*, *17*(1), 1–13. https://doi.org/10.1016/S0002-7138(09)62273-1

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition*, *110*(1), 124–129. https://doi.org/10.1016/j.cognition.2008.10.008

van Pinxteren, M. M. E., Wetzels, R. W. H., Rüger, J., Pluymaekers, M., & Wetzels, M. (2019). Trust in humanoid robots: Implications for services marketing. *Journal of Services Marketing*, *33*(4), 507–518. https://doi.org/10.1108/JSM-01-2018-0045

Vygotsky, L. S. (1962). *Thought and language.* (MIT Press).

Warneken, F., Lohse, K., Melis, A. P., & Tomasello, M. (2011). Young Children Share the Spoils After Collaboration. *Psychological Science*, *22*(2), 267–273. https://doi.org/10.1177/0956797610395392

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who Sees Human?: The Stability and

Importance of Individual Differences in Anthropomorphism. *Perspectives on Psychological Science*, *5*(3), 219–232. https://doi.org/10.1177/1745691610369336

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383–388. https://doi.org/10.1016/j.tics.2010.05.006

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005

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

Wiese, E., Metta, G., & Wykowska, A. (2017). Robots As Intentional Agents: Using Neuroscientific Methods to Make Robots Appear More Social. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01663

Wimmer, H. (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

Winkielman, P., Niedenthal, P. M., & Oberman, L. (2008). *The Embodied Emotional Mind.*

Wood, A., Rychlowska, M., Korb, S., & Niedenthal, P. (2016). Fashioning the Face: Sensorimotor Simulation Contributes to Facial Expression Recognition. *Trends in Cognitive Sciences*, *20*(3), 227–240. https://doi.org/10.1016/j.tics.2015.12.010

Wykowska, A., Chaminade, T., & Cheng, G. (2016). Embodied artificial agents for understanding human social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693), 20150375. https://doi.org/10.1098/rstb.2015.0375

Yang, G.-Z., Bellingham, J., Dupont, P. E., Fischer, P., Floridi, L., Full, R., Jacobstein, N., Kumar, V., McNutt, M., Merrifield, R., Nelson, B. J., Scassellati, B., Taddeo, M., Taylor, R., Veloso, M., Wang, Z. L., & Wood, R. (2018). The grand challenges of

Science Robotics. *Science Robotics*, *3*(14), eaar7650. https://doi.org/10.1126/scirobotics.aar7650

Ye, S., Neville, G., Schrum, M., Gombolay, M., Chernova, S., & Howard, A. (2019). Human Trust After Robot Mistakes: Study of the Effects of Different Forms of Robot Communication. *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1–7. https://doi.org/10.1109/RO-MAN46459.2019.8956424

Young, J. E., Hawkins, R., Sharlin, E., & Igarashi, T. (2009). Toward Acceptable Domestic Robots: Applying Insights from Social Psychology. *International Journal of Social Robotics*, *1*(1), 95–108. https://doi.org/10.1007/s12369-008-0006-y

Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and Challenges in Human–Robot Interaction. *International Journal of Social Robotics*, *7*(3), 347–360. https://doi.org/10.1007/s12369-014-0267-6

Zwickel, J. (2009). Agency attribution and visuospatial perspective taking. *Psychonomic Bulletin & Review*, *16*(6), 1089–1093. https://doi.org/10.3758/PBR.16.6.1089

CHAPTER 2
ACTION CHAINS AND INTENTION UNDERSTANDING IN 3-6-YEAR-OLD CHILDREN

Cinzia Di Dio[1*], **Laura Miraglia[1*]**, Giulia Peretti[1], Antonella Marchetti[1] & Giacomo Rizzolatti[2]

[1]Department of Psychology, Università Cattolica del Sacro Cuore, Largo Gemelli 1, 20123 Milano, Italy; [2]Department of Neuroscience, Università di Parma, Via Volturno 39, 43100 Parma, Italy.

[*]Equal contribution

Under review on The Proceedings of the National Academy of Sciences (PNAS)

Keywords

Mirror neurons | Motor chains | Motor acts | Intention understanding | Social cognition.

Abstract

In intentional behavior, the final goal of an action is crucial in determining the entire sequence of motor acts. Neurons have been described in the inferior parietal lobule of monkeys, which besides encoding a specific motor act (e.g., grasping), have their discharge modulated by the final goal of the intended action (e.g., grasping-to-eat). Many of these “action-constrained” neurons have mirror properties responding to the observation of the motor act they encode, provided that this is embedded in a specific action. Thanks to this mechanism, the observers have an internal copy of the whole action before its execution, and may, in this way, understand the agent’s intention. This chained organization of motor acts exists in humans. In the present study, we recorded EMG from the mylohyoid (MH) muscle in children 3 to 5 years old in order to assess whether this mechanism is present at this early age. The results confirmed this hypothesis, with two main differences with respect to data previously reported in older children (6 to 9 years old). First, during the observation condition, no difference in MH muscle activation was found between the grasping-to-eat and grasping-to-place actions in the reaching phase. In contrast, a clear difference was present in the execution of the two actions. Thus, young children are able to modulate their intentional motor mechanism when acting, but unable to understand other’s intentions at the action outset. Second, the observation of the grasping-to-eat motor act, relative to grasping-to-place, showed greater activation in younger than in older children, suggesting poor inhibitory motor control and fostering evidence on early imitative learning.

Introduction

The “chained” organization of motor acts is a fundamental aspect of action organization of primates. This type of organization was first described in the inferior parietal lobule (IPL) of the macaque monkey by Fogassi et al. (Fogassi et al., 2005). The authors found that in IPL there is a set of neurons, which they named action-constrain neurons, that encode a specific motor act (e.g., grasping). Their defining feature is that their discharge is modulated by the action in which the motor act is placed (e.g., grasping-to-eat, grasping-to-place, etc.) (Bonini et al., 2011; Fogassi et al., 2005; Rozzi et al., 2008). It has been proposed that this “chained” organization of motor acts underlies a fundamental aspect of action execution: fluidity. Indeed, fluidity requires a close connection between the motor acts forming the whole action so that its execution can occur without interruption.

A large number of IPL grasping neurons show mirror properties (see, 4), firing both during the execution of a motor act as well as during the observation of the same motor act. Most interestingly, in many action-constrained neurons with mirror property, the discharge intensity during action observation was modulated by the action in which the motor act was embedded. In this way, the activation of IPL action-constrained mirror neurons provides information not only on the observed motor act (e.g., grasping), but also allows one to predict the final goal of the observed action (i.e., grasping-to-eat or gasping-to-place) and, therefore, to infer the agent intentions (Cattaneo et al., 2007; Fogassi et al., 2005).

There is strong, albeit indirect, evidence, that a similar motor-act chained organization is also present in humans. Cattaneo and colleagues (Cattaneo et al., 2007) recorded the electromyography (EMG) from the mylohyoid (MH) muscle - that is involved in mouth opening - in children aged 6 to 9 years, who were instructed to grasp a piece of food or a piece of paper and to bring it to the mouth (grasping-to-eat) or to a container (grasping-to-place), respectively. The results showed consistently greater activation of the muscles involved in mouth opening during the gasping-to-eat action, relative to the grasping-to-place one. Furthermore, the MH muscle activation started very early, already during the reaching phase, in the gasping-to-eat action relative to the grasping-to-place one. A clear difference in the MH activation between grasping-to-eat and grasping-to-place actions was also found when children observed the experimenter performing the two actions. As in action execution, also during action observation the activation of the MH was greater and started earlier in the gasping-to-eat than in grasping-to-place action. Thus, the action-chain mechanism appears to be involved in the selection of motor acts for both action execution and understanding the intentions of others.

Based on these findings, in the present study we examined whether children aged 3-to-6 years (preschoolers) are able to organize intentionally their actions and to understand the agent intentions behind the observed actions. To this aim, the activation of MH muscle was recorded while children were either executing or observing two actions: 1) reach out to pick up a piece of food with the aim of eating it and 2) reach out to pick a piece of paper with the aim of placing it in a container.

Materials and Methods

Participants

Eighteen (18) preschool neuro-typically developing children were recruited from two kindergartens in Milan and Trecate, Italy (40% female; mean age = 4.31 years; SD = .76; age range = 3.08-6.01 years). Informed consent was obtained from all children's legal guardian(s) in line with the Declaration of Helsinki and its revisions, as well as in accordance with the requirements of the ethics committee, Committee of the Department of Psychology (CERPS), Università Cattolica del Sacro Cuore, Milan, which approved the study.

Procedure

We investigated the motor activation of the mouth-opening mylohyoid (MH) muscle during two different actions, i.e., grasping to eat and grasping to place, in two experimental conditions: observation and execution. Children were assessed in both experimental conditions. In the eating-observation condition, children were instructed to carefully observe the researcher using her right hand to grasp a piece of food from a touch-sensitive plate, bringing it to her mouth, and eating it. In the placing-observation condition, children observed the researcher using her right hand to grasp a piece of paper from the touch-sensitive plate and place it into a container situated on the experimenter's right shoulder. Both actions were repeated 30 times. In the execution condition, children carried out the actions of grasping a piece of food to eat and a piece of paper to place into the container located on the child's shoulder. Children repeated the actions 30 times.

Children were seated at a table in front of the researcher and, according to experimental conditions, all trials started with the experimenter or participant's hand resting on the start button and, throughout the duration of the experiment, the activation of the participant's MH muscle was recorded using surface electrodes. Prior to the experiments, children underwent a training session where they performed grasping and eating the food or placing the piece of paper into the container, depending on the presented stimulus. A brief training session was also conducted for the observation condition. The conditions were presented in random order. No verbal instructions were provided during the trials, except for occasional prompt as “pay attention” or “look at the experimenter.”

To mark on the EMG recordings the beginning and the end of each action and the contact with the stimuli, a button and a touch-sensitive plate were placed on the table and connected to the EMG apparatus. The release of the hand from the button signaled the beginning of the action, and pushing the button again indicated the end of the action. The grasping of the piece of paper or the food was detected through the touch-sensitive plate and marked as the starting moment (T0).

Materials

For the recordings of the mouth-opening mylohyoid (MH) muscle, we used a wireless electromyograph (EMG; DueLite, OT Bioelettronica SRL, Turin, Italy). The MH muscle was recorded using two surface bipolar electrodes (CDE - Bipolar electrodes diameter 15mm with concentric connector). The two electrodes were placed under the child’s chin for MH muscle recording. EMG was recorded continuously throughout the experiment. The interval between trials was according to the compliance of children. The signal was amplified 1.000x, sample frequency of 2048 Hz, controlled by the EMG signal software, and stored for offline filtering (bandpass: 10-500 Hz) and further analysis.

Signal Processing and Definition of Epochs

Trials were discarded whenever the participant spoke, swallowed, or coughed contaminating the recordings, and even when the child was not showing enough attention to the tasks. We discarded contaminated trials by using video recordings.

In all experiments, the start button and the touch-sensitive plate on which the food/paper was placed, were connected to the EMG apparatus. The actions were thus divided into three epochs: reaching the stimulus, grasping it, and bringing it to the mouth or into the container placed on the shoulder. The release of the start button signaled the beginning of the reaching phase (T-2). The contact with the plate on which the stimulus was put signaled the grasping phase (T0). The release from the plate signaled the end of the grasping epoch and the start of the bringing phase (T+2). All recordings were aligned on the moment of paper or food lifting from the touch-sensitive plate.

All recordings were filtered (30-400 Hz). The instants of signal activation were identified in each device's recordings. Then, we calculated the amplitude of the MH muscle signal by considering 2 seconds before (reaching phase) and 2 seconds after (bringing phase) T0, dividing the signal into epochs of 0.2 seconds.

Data Analysis

GLM analysis was carried out by using the EMG activation in the three epochs of action as a dependent variable. First, for each child in each condition, the median of each trial was calculated, and the standard deviation (SD -2, SD +2) was used to normalize the data. In all experiments, within-subjects ANOVAs for repeated measures were carried out with 2 levels of condition (observation, execution), 2 levels of action type (eating, placing), and 3 levels of epoch (reaching, grasping, and bringing). In the GLM, the epochs of the action were identified as the moment in which the researcher/participant released the button (reaching – T-2), touched the plate (grasping – T0), and pushed the button (bringing – T+2).

Post hoc analyses were Bonferroni corrected. Significance levels were set at p = 0.05. A Greenhouse-Geisser (G-G) correction was applied whenever the sphericity assumption was violated. Additionally, correlation analyses (Spearman rho) were carried out to evaluate the relationship between the MH muscle activation and children's age.

Results

Main Analysis: Differences between conditions.

The activity of the mylohyoid (MH) muscle was recorded in eighteen (n=18) neurotypical preschool children during the observation and the execution of grasping-to-eat and grasping-to-place actions. In the observation condition, children were instructed to observe carefully the experimenter, while the experimenter was grasping a piece of food located on a touch-sensitive plate, bringing it to the mouth, and eating it (grasping-to-eat observation condition) or while the same experimenter was grasping a piece of paper from the same touch-sensitive plate and placing it into a container positioned on the experimenter's right shoulder (grasping-to-place observation condition). In the execution condition, the children actively carried out the two actions described above. Namely, they grasped the food and brought it to their mouth to eat (grasping-to-eat execution condition) and placed the piece of paper into the container positioned on their shoulder (grasping-to-place execution condition). In both conditions, the children's MH activity was recorded using surface electrodes. Two devices, a button and a touch-sensitive plate, were used to signal the release of the hand from the table and the contact of the hand with the stimuli, respectively. These signals enable the subdivision of the observation and execution tasks into three epochs (reaching, grasping, and bringing).

For each child, the EMG of MH muscle was recorded, rectified, and averaged separately in the three epochs (see above) for the two actions (grasping-to-eat and grasping-to-place) and the two conditions (observation and execution), and used as the dependent variable in the analysis. A GLM analysis was then carried out to assess an anticipation effect of object grasping, with 2 levels of condition (observation, execution), 2 levels of action type (eating and placing), and 3 levels of epoch (reaching, grasping, and bringing) as within-subject factors. Post-hoc comparisons were Bonferroni corrected and the Greenhouse-Geisser correction was used when the sphericity assumption was violated.

The results revealed a main effect of condition, F(1, 1426.2) = 16.63, p = .003, partial-η2 = .65, δ = .95, indicating that, regardless of action type (eating or placing), the activity of the MH muscle was greater when children executed than when observed the actions, Mdiff = 6.9, SE = 1.69, p = .003. A main effect of action type was also found, F(1, 876.39) = 25.63, p < .001, partial-η2 = .74, δ = 99, showing that, regardless of condition (observation or execution), the increase in MH activity was greater during grasp-to-eat action compared to grasp-to-place action, Mdiff = 5.41, SE = 1.07, p < .001. In addition, a main effect of epoch was found, F(1.99, 157.12) = 10.79, p < .001, partial-η2 = .55, δ = .98, indicating that independent of action-type or condition, the grasping phase (i.e., grasping the food or the piece of paper) was significantly higher compared to the reaching phase (i.e., reaching toward the food or the piece of paper), Mdiff = 3.93, SE = .82, p = .003. No significant difference was found between the grasping phase and the bringing phase, as well as between the reaching phase and the bringing phase. Finally, a significant three-way interaction was found between condition, action type, and epoch F(1.78, 86.28) = 5.98, p = .013, partial-η2 = .4, δ = .78. The three-way interaction mainly stems from the absence vs. presence of differences between eating and placing actions in the observation and execution conditions. The following section better outlines these findings.

MH Muscle Activation during Action Observation.

Pairwise post-hoc comparisons showed a significant increase in MH activation during action observation in the grasping phase of the grasping-to-eat action relative to the same phase in the grasping-to-place, action Mdiff = 3.97, SE = 1.00, p < .05. This effect is shown in Fig. 1, left.

The time course of the activations in the two conditions is illustrated in Fig,1, right. In both grasping-to-eat and grasping-to-place conditions, the activation of the MH muscle slightly increased during the reaching phase, peaked when the child observed the experimenter grasping the food (T0), and subsequently decreased. During the observation of the eating action, a significant increase in MH activation was observed at T0, while only a small increase was found in this phase during the observation of the placing action.

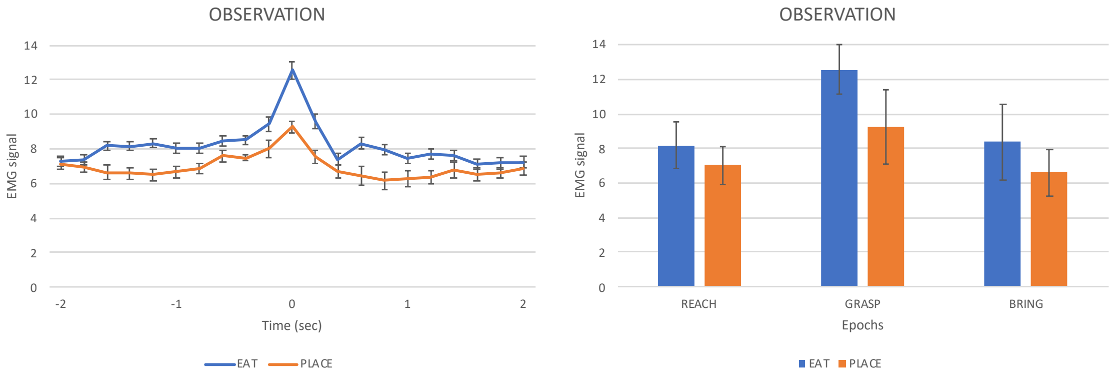


Figure 1. *Left figure:* Time course of the rectified EMG activity of the MH muscle during the observation of the grasp-to-eat action (blue) and the grasp-to-place (orange) actions. The curves are aligned to the moment of stimulus lifting from the touch-sensitive plate (T0). *Right figure:* EMG signal amplitude of the MH muscle during the eating and placing actions in the observation condition across the three epochs (reach, grasp, and bring).

MH Muscle Activation during Action Execution.

Pairwise post-hoc comparisons showed a significant increase in MH muscle activation during the execution of the eating action compared to the placing action for all three epochs (reaching, grasping, bringing). There was a significant difference in MH muscle activation between the execution of eating and the execution of placing actions already during the reaching phase, Mdiff = 6.98, SE = 3.16, p < .05. This increase became more pronounced during the grasping phase, Mdiff = 6.68, SE = 2.53, p < .05, and persisted during the bringing epoch, Mdiff = 14.15, SE = 2.44, p < .001. These results are illustrated in Fig.2, left.

Fig. 2, right, shows the time course of the median EMG signal of the MH muscle in the two conditions (grasping-to-eat and grasping-to-place). In the eating condition, the activation of the children's MH muscle started to increase several milliseconds before the hand grasped the food (T0), continued to rise during the grasping epoch and, as expected, reached its peak when the children started to open their mouth.

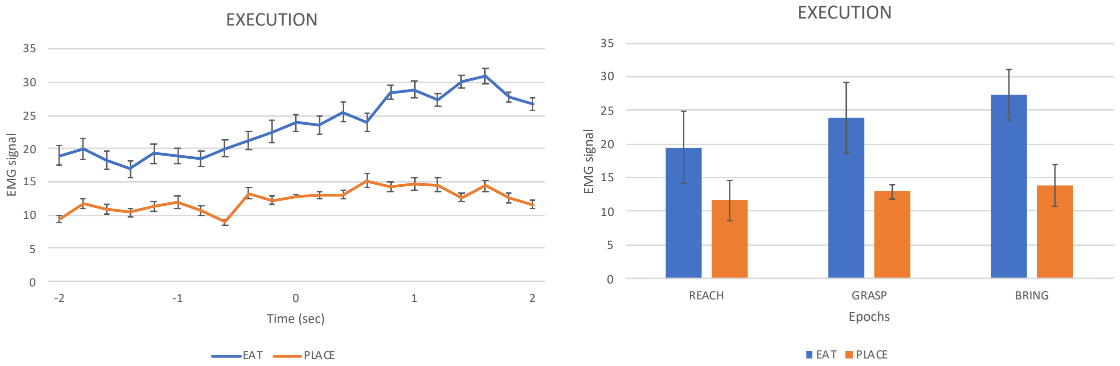


Figure 2. *Left figure:* Time course of the rectified EMG activity of the MH muscle during the execution of the grasp-to-eat action (blue) and the grasp-to-place (orange) actions. The curves are aligned to the moment of stimulus lifting from the touch-sensitive plate (T0). *Right figure:* EMG signal amplitude of the MH muscle during the eating and placing actions in the execution condition across the three epochs (reach, grasp, and bring).

Age Correlations.

To investigate whether MH muscle activation changed as a function of the children's age, Spearman correlations (one-tail) were examined between children's age (months) and the activation of EMG of the MH muscle in conditions (observation, execution), action types (eating, placing), and at the three epochs (reaching, grasping, bringing). The results showed a significant negative correlation between the children's age and the activation of the MH muscle during the grasp-to-eat phase in the observation condition, rho = -.56, p < .05. That is, as shown in Figure 3, the MH muscle showed significantly greater levels of activation in younger children, in comparison to the older ones, when they observed the experimenter grasping the food with the intention to eat it.

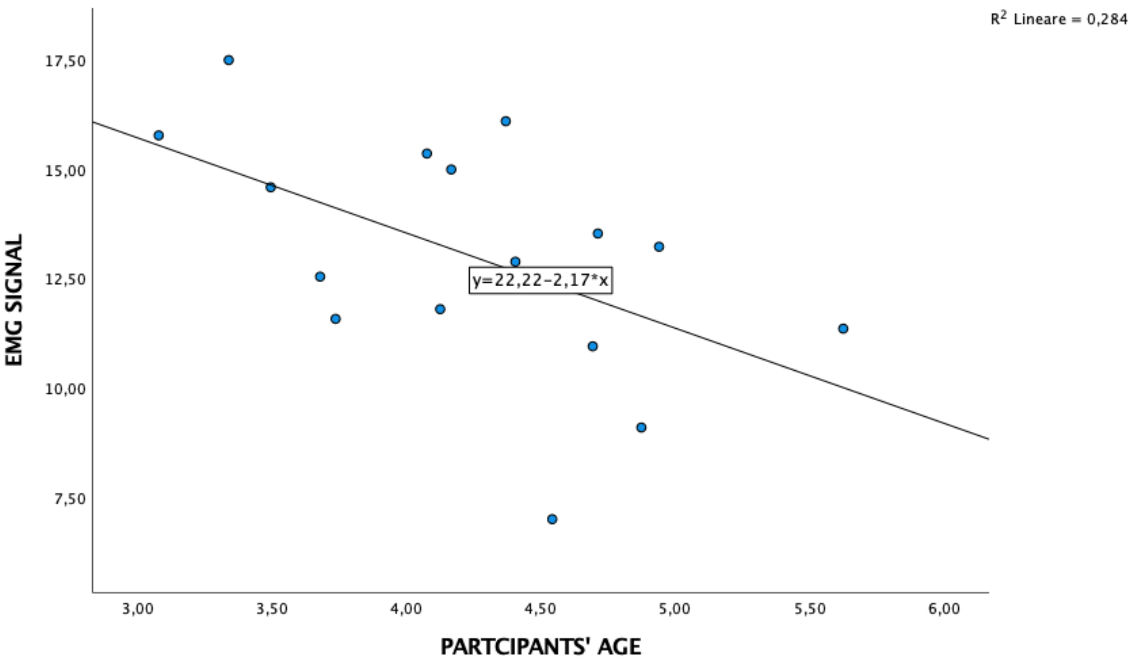


Figure 3. Spearman correlations between children's age and rectified EMG activation of MH muscle during the observation of the grasp-to-eat action (T0).

Discussion

The ability to temporally organize motor acts according to the end goal of an action is a key step in achieving the fluidity of action typical of adult motor behavior. Similarly, the ability to understand the intention underlying others' actions is vital for the development of social competencies. As for the latter, there is ample evidence that the basic aspects of both goal and intention understanding are mediated by the mirror mechanism (Fogassi et al., 2005; Gallese et al., 1996; Rizzolatti et al., 1996; Rizzolatti & Sinigaglia, 2007, 2010). This mechanism transforms sensory representations of others' behavior into one's own motor representations of the same behavior, thus allowing one to understand experientially the goal as well as, through the "action constrained" mirror neurons, the agent's intention.

In this study, we examined whether the ability to organize actions intentionally and understand the intentions behind the observed actions is already present at preschool age. To this purpose, the activation of MH muscle was recorded in 3-to-6 years old children, while they were executing or observing two actions: 1) reaching to grasp a piece of food with the aim of eating it and 2) reaching to grasp a piece of paper with the aim of placing it into a container. The results showed that preschoolers could select the appropriate motor chain intentionally during action execution. Specifically, during the execution of the grasping-to-eat action, a significant increase in activation of the mylohyoid (MH) muscle was found already during the reaching phase, several milliseconds before the child actually grasped the food. The muscle activation persisted during the grasping phase and peaked when the children opened their mouths. No similar activations were found during grasping-to-place action. These results most likely reflect a physiological mechanism whereby, during the preparation of goal-directed action, there is activation of the corresponding motor chain before the start of the actual movement, being the selection of the "chain" driven by the intention to perform a given action. These findings are in line with previous data collected in older children (6 to 9 years old) (Cattaneo et al., 2007) and demonstrate that the ability to select motor acts intentionally is already present in preschool children.

In contrast with action execution, during the observation of grasping-to-eat, the chain mechanism was activated later, when the experimenter initiated the actual grasping movement. As a matter of fact, no significant difference was found between grasping-to-eat and grasping-to-place actions in the reaching phase. It is important to note, in this respect, that in older children aged 6 to 9 years old, an activation of the MH muscle was found already 100 ms before stimuli grasping and the activity in the reaching phase was significantly higher in the grasping-to-eat than in grasping-to-place conditions (Cattaneo et al., 2007). These data indicate that while older children are able to both organize their actions intentionally (execution) and to understand the intentions of others (observation) rather early, preschoolers show a delayed activation most likely because their intention coding mirror mechanism is not sufficiently developed to understand others' intentions.

As to this finding, it is worth comparing the behavior of typically developing preschoolers in the present study with that of children with autistic spectrum disorder (ASD) tested in Cattaneo et al. (Cattaneo et al., 2007), who used the same paradigm as the present one. The results showed a complete lack of EMG activation in the MH muscle during the reaching phase of grasping-to-eat action execution. Similarly, the ASD children did not show any activation of the MH muscle during the observation of the experimenter grasping the food to eat. Thus, while preschoolers show a delayed ability to use their intentional mechanism to understand the intentions of others and yet have an adequately functioning anticipatory motor mechanism underlying the organization of action, children with ASD lack the latter as well and thus are necessarily incapable of understanding the intentions of others experientially (see below).

In the present article, we used several times the expression "understanding actions of others". But what does this mean exactly? The first answer that comes to mind derives from philosophy. It implies that the person observing the action of another person has some knowledge of the beliefs, desires, and intentions of the person executing that action. This knowledge allows the observer to infer the reasons behind the observed behavior. However, if we take a closer look at the commonly accepted meaning of "understand", we see that it does not exclude the possibility of understanding an observed action without necessarily having a knowledge of the mental states that have motivated it. Imagine, for example, a person reaching for a glass of beer and someone asking you if you have understood what that person is doing. Your answer would most likely be: "This person is picking up that glass". You might even say, "This person is going to pick up the glass in order to drink from it." Your answer is most likely correct, even if you are not aware of the reasons that motivated the person to reach for the glass. In this example, understanding the action means to identify the goal of the motor act (pick up the glass), and to infer the possible intention underlying it (to drink). In order to clarify the meaning of "understand", it has been suggested to call for a full understanding of the action if the notion of understanding includes knowledge of the mental states of the agent, and of basic understanding if the observer identifies the goal of the motor act and infers the intention behind the action without knowing the reasons that led to its execution (Rizzolatti & Sinigaglia, 2023).

The understanding of the others' intention (basic understanding) described by Cattaneo et al. (Cattaneo et al., 2007) is most likely mediated by the intentional mirror mechanism. As discussed above this mechanism is poorly developed, or even absent, in younger children. This finding appears to be in line with the data obtained in ASD children by Boria et al. (Boria et al., 2009) where ASD and typically developing (TD) children had to decide the "what" and the "why" of hand actions presented on a computer screen. To make an example, they saw a pair of scissors and a hand grasping them. The grasping hand could be positioned on the scissors handle, as one does for using the scissors, or on the center of the scissors, as one does for moving them. Both groups of children responded correctly to the "what" question, stating that the hand was grasping an object, scissors in our example; however, the ASD children failed to indicate the "why" of the observed gesture, most of the errors stemming from the "move" condition, i.e., the one showing an unusual action associated with scissors. In a further experiment, the same two groups of children saw pictures showing a hand grasping an object as in the previous experiment but, this time, within a context suggesting either the typical use of the object or its explicit placement into a container. Here, children with ASD performed like TD children correctly indicating the agent's intention. In summary, the present data and the previous data discussed above suggest that the capacity to understand the "what" of a motor act is a very early and robust acquisition. In contrast, the capacity to understand the intention of an action, even the basic one, is a later acquisition based on the context in which the motor act is performed and, later, on the mirror intention chains.

A further important finding that allows us to better clarify preschool-age data described here is that the activation of the MH muscle at the actual grasping time of the stimulus (time 0) during the observation of the experimenter's grasping action was substantially greater in the younger children. Beyond understanding actions, mirror neurons facilitate imitation (Rizzolatti, 2005), a well-known phenomenon in developmental psychology that allows children to learn from and interact with the world around them (Giudice et al., 2009; Heyes, 2001; Meltzoff & Marshall, 2018). As children develop, their imitation skills become increasingly sophisticated, moving from imitating facial expressions and other movements to imitating more complex actions and behaviors and understanding the intentions of others. The typical example is represented by the imitation of tongue protrusion originally described by Meltzoff and Moore (Meltzoff & Moore, 1977). These overt manifestations of imitative behaviors then tend to decrease and even disappear with age (see, 13). The increase of the MH muscle during the observation of the grasping action in the youngest children of the present study is most likely due to an immature frontal inhibitory control that gradually develops during childhood till early adulthood (Adleman et al., 2002; Bunge et al., n.d.; Rubia et al., 2006). This is also in line with our findings concerning the grasp-to-place conditions. The observed, albeit low, activation of the MH muscle during both the execution and observation of the placing action reinforces the idea that preschoolers have not yet developed an efficient frontal inhibitory control in the prefrontal lobe on the mirror neurons coding motor acts.

Author Contributions

CDD, GR, and AM designed the experiment. LM and GP performed data acquisition and analysis. CDD, GR, AM, LM, and GP interpreted the results. CDD, LM, and GP drafted the paper; GR and CDD revised and drafted the paper. All authors participated in reviewing and approving the manuscript.

Competing Interest Statement

The authors declare no competing interest.

Classification

Social Science and Neuroscience.

Acknowledgments

We would like to thank Dr Luigi Cattaneo for his invaluable contribution to the preparation of the research material and the set-up of the work. We thank Dr Lorenzo Borsa for his assistance in the use of the EMG signal encoding software. We would like to thank Sister Antonietta Occhetta, Director of the Scuola dell’Infanzia Paritaria Fratelli Russi, and the school La Locomotiva di Momo, for their deep involvement in organizing and supporting the study. We would also like to thank all the teachers who actively participated in the study. Finally, we would like to thank all the families who made themselves available for the realization of this work.

References

Adleman, N. E., Menon, V., Blasey, C. M., White, C. D., Warsofsky, I. S., Glover, G. H., & Reiss, A. L. (2002). A Developmental fMRI Study of the Stroop Color-Word Task. NeuroImage, 16(1), 61–75. https://doi.org/10.1006/nimg.2001.1046

Bonini, L., Ugolotti Serventi, F., Simone, L., Rozzi, S., Ferrari, P. F., & Fogassi, L. (2011). Grasping Neurons of Monkey Parietal and Premotor Cortices Encode Action Goals at Distinct Levels of Abstraction during Complex Action Sequences. The Journal of Neuroscience, 31(15), 5876–5886. https://doi.org/10.1523/JNEUROSCI.5186-10.2011

Boria, S., Fabbri-Destro, M., Cattaneo, L., Sparaci, L., Sinigaglia, C., Santelli, E., Cossu, G., & Rizzolatti, G. (2009). Intention Understanding in Autism. PLOS ONE, 4(5), e5596. https://doi.org/10.1371/journal.pone.0005596

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. E. (n.d.). Immature Frontal Lobe Contributions to Cognitive Control in Children: Evidence from fMRI.

Cattaneo, L., Fabbri-Destro, M., Boria, S., Pieraccini, C., Monti, A., Cossu, G., & Rizzolatti, G. (2007). Impairment of actions chains in autism and its possible role in intention understanding. Proceedings of the National Academy of Sciences, 104(45), 17825–17830. https://doi.org/10.1073/pnas.0706273104

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal Lobe: From Action Organization to Intention Understanding. Science, 308(5722), 662–667. https://doi.org/10.1126/science.1106138

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. Brain, 119(2), 593–609. https://doi.org/10.1093/brain/119.2.593

Giudice, M. D., Manera, V., & Keysers, C. (2009). Programmed to learn? The ontogeny of mirror neurons. Developmental Science, 12(2), 350–363. https://doi.org/10.1111/j.1467-7687.2008.00783.x

Heimann, M. (1989). Neonatal imitation, gaze aversion, and mother-infant interaction. Infant Behavior and Development, 12(4), 495–505. https://doi.org/10.1016/0163-6383(89)90029-5

Heyes, C. (2001). Causes and consequences of imitation. Trends in Cognitive Sciences, 5(6), 253–261. https://doi.org/10.1016/S1364-6613(00)01661-2

Meltzoff, A. N., & Marshall, P. J. (2018). Human infant imitation as a social survival circuit. Current Opinion in Behavioral Sciences, 24, 130–136. https://doi.org/10.1016/j.cobeha.2018.09.006

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of Facial and Manual Gestures by Human Neonates. Science, 198(4312), 75–78. https://doi.org/10.1126/science.198.4312.75

Rizzolatti, G. (2005). The mirror neuron system and its function in humans. Anatomy and Embryology, 210(5–6), 419–421. https://doi.org/10.1007/s00429-005-0039-z

Rizzolatti, G., & Craighero, L. (2004). THE MIRROR-NEURON SYSTEM. Annual Review of Neuroscience, 27(1), 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. Cognitive Brain Research, 3(2), 131–141. https://doi.org/10.1016/0926-6410(95)00038-0

Rizzolatti, G., & Sinigaglia, C. (2007). Mirror neurons and motor intentionality. Functional Neurology, 22(4), 205–210.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. Nature Reviews Neuroscience, 11(4), Article 4. https://doi.org/10.1038/nrn2805

Rizzolatti, G., & Sinigaglia, C. (2023). Mirroring Brains: How We Understand Others from the Inside. Oxford University Press.

Rozzi, S., Ferrari, P. F., Bonini, L., Rizzolatti, G., & Fogassi, L. (2008). Functional organization of inferior parietal lobule convexity in the macaque monkey: Electrophysiological characterization of motor, sensory and mirror responses and their correlation with cytoarchitectonic areas. European Journal of Neuroscience, 28(8), 1569–1588. https://doi.org/10.1111/j.1460-9568.2008.06395.x

Rubia, K., Smith, A. B., Woolley, J., Nosarti, C., Heyman, I., Taylor, E., & Brammer, M. (2006). Progressive increase of frontostriatal brain activation from childhood to adulthood during event-related tasks of cognitive control. Human Brain Mapping, 27(12), 973–993. https://doi.org/10.1002/hbm.20237

CHAPTER 3
SHARED KNOWLEDGE IN HUMAN-ROBOT INTERACTION (HRI)

Laura Miraglia[1], Cinzia Di Dio[1,2], Federico Manzi[1,2], Takayuki Kanda[3,4], Angelo Cangelosi[5], Shoji Itakura[6], Hiroshi Ishiguro[4,7], Davide Massaro[1,2], Peter Fonagy[8], Antonella Marchetti[1,2].

[1]Research Unit on Theory of Mind, Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy; [2]Research Unit in Psychology and Robotics in the Lifespan (PsyRoLife), Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy; [3]Human-Robot Interaction Laboratory, Department of Computer Science, Kyoto University, Kyoto, Japan; [4]Advanced Telecommunications Research Institute International, IRC/HIL, Keihana Science City, Kyoto, Japan; [5]School of Computer Science, The University of Manchester, Manchester, UK; [6]Centre for Baby Science, Doshisha University, Kyoto, Japan; [7]Department of Systems Innovation, Osaka University,
Toyonaka, Japan; [8]Research Department of Clinical, Educational and Health Psychology, UCL, London, UK.

Published on International Journal of Social Robotics, July 2023.

Keywords

Human-Robot Interaction (HRI) | Natural Pedagogy Theory | Shared Knowledge | Social cognition | attachment style.

Abstract

According to the Theory of Natural Pedagogy, object-directed emotion may provide different information depending on the context: in a communicative context, the information conveys culturally shared knowledge regarding the emotional valence of an object and is generalizable to other individuals, whereas, in a non-communicative context, information is interpreted as a subjective disposition of the person expressing the emotion, i.e., personal preference. We hypothesized that this genericity bias, already present in infants, may be a feature of human communication and, thus, present at all ages. We further questioned the effects of robotic ostensive cues. To explore these possibilities, we presented object-directed emotions in communicative and non-communicative contexts under two conditions: adult participants (N = 193) were split into those who underwent the human-demonstrator condition and those who underwent the robot-demonstrator condition, i.e., a human actor or a robot displayed object-directed emotions, respectively. Questionnaires further assessed the participants' attachment style and mentalization ability. The results showed that (1) Natural Pedagogy Theory applies to humans across the lifespan; (2) Shared knowledge depends on the contexts (communicative vs. non-communicative) and who is sharing the information (human or robot); and (3) robotic ostensive cues trigger participants' attention, conversely, in their absence, participants do not turn the robot into a communicative partner by not assigning it a communicative intention due to a difficulty in reading the robot's mind. Taken together, our findings indicate that robotic ostensive cues may ease the human-robot interaction (HRI), which is also biased by the human attachment style.

The study was preregistered in Open Science Framework, OSF on September 9, 2021 (Registration DOI https://doi.org/10.17605/OSF.IO/9TWY8).

Introduction

Natural Pedagogy Theory

The evolutionary success of our species depends crucially on the social transmission of knowledge both contemporaneously and throughout historical time. Hence, one of the first challenges a human being faces is learning about and from the world around him/her. Children are able to draw information relevant to their behavior by simply observing the reactions of adults and gaze direction toward an object or event (Moses et al., 2001; Mumme & Fernald, 2003), and by decoding emotional information and discriminating between facial (Baldwin & Moses, 1996) and vocal expressions (Mumme & Fernald, 2003). This form of social cognition emerges at approximately one year of age when infants begin to engage with others in various types of joint attentional activities, such as gaze, social reference, and gestural communication, which generate cultural learning that enables the acquisition of language, discursive skills, tool-use practices, and other conventional activities (Isernia et al., 2019; Tomasello, 1999).

Csibra and Gergely (Csibra & Gergely, 2009; Csibra & György, 2006) framed out the Natural Pedagogy Theory which posits Pedagogy as a specific type of communication that enables rapid and efficient social learning that – similarly to all types of social learning (imitation, emulation, etc.) – conveys generalizable knowledge that is valid beyond the actual situation (Csibra & György, 2006). Thus, it is important to consider the distinctive nature of Pedagogy both as a particular type of social learning and as a particular type of communication. Csibra and Gergely's theory is grounded on the Gricean notion of ostensive communication, which postulates that an essential feature of human communication is the expression and recognition of intents (Grice, 1991). Ostensive communication is achieved through the production of ostensive signals, stimuli, or cues that indicate a communicative intention towards an addressee. Ostensive cues typically lead the addressee to feel recognized as a subject (Fonagy & Allison, 2014), encouraging more rapid knowledge acquisition (Luyten et al., 2020) and allowing the establishment of epistemic trust (Fisher et al., 2021; Fonagy & Allison, 2014). Importantly, it is plausible that secure attachment acts as a guarantee of the authenticity of knowledge (Fonagy et al., 2007), as the child is more likely to attend to the known and trusted adult indicating and naming new objects or showing whether the object is good or bad through social referencing (Baldwin & Moses, 1996; Tomasello, 1999).

Ostensive cues and social interaction

Human infants are highly sensitive to social cues (Baldwin, 1993; Baldwin & Moses, 1996; Flom et al., 2007; Mumme & Fernald, 2003; Tomasello, 1999), i.e., behavioral cues – such as eye contact, infant-directed speech, turn-taking contingent discourse, calling an infant by name, etc. – which indicate a clear communicative intention of an agent. This has led to a growing interest in how social cues, such as object-directed emotional expressions, can be an important source of social information. The caregiver's ostensive cues not only cause the infant to interpret the adult's action as indicative of a communicative intention to transfer relevant knowledge but also generate attachment security through a sensitive and contingent response (Fonagy et al., 2007). Social cues act differently in terms of preparing the observer to obtain certain types of object information and should be distinguished from non-communicative cues concerning the expected effects (Csibra & Gergely, 2009; Csibra & György, 2006). In this respect, in communicative contexts, i.e., when the addressee is engaged through ostensive cues, a genericity bias is generated, in other words, the information conveyed is processed as generalizable to other individuals and valid beyond the present situation (Csibra & Gergely, 2009; Egyed et al., 2013; Marno et al., 2014; Yoon et al., 2008). In this sense, ostensive cues make it possible to convey generally shared knowledge. Infants' sensitivity to ostensive signals triggers an automatic predisposition in the child to receive new and relevant information, made manifest by the communicative intention of the adult through ostensive communication (Csibra, 2010; Gergely, 2007). Within these communicative contexts, ostensive cues (such as gaze shifting, head movement, and pointing) generate an expectation of generic content in the addressee: unless the context or other cues specify otherwise, children interpret the information they receive as generic rather than episodic (Csibra & Gergely, 2009; Egyed et al., 2013; Vorms, 2012). Consequently, Natural Pedagogy theorists have argued that an ostensive or communicative context generates a genericity bias whereby the addressee expects to be taught something generalizable and focuses his or her attention on the intrinsic characteristics of the object referent (Csibra & Gergely, 2009). For instance, Yoon and colleagues (Yoon et al., 2008) showed that when 9-month-old infants were introduced to an object in a communicative context, they remembered generic properties of the object (such as its identity); but when the same object was presented in non-communicative contexts, they were more likely to remember its location, that is context-specific properties. In line with these results, Marno and colleagues (Marno et al., 2014) also found that, in a communicative context, adult participants preferentially encoded the object's identity at the expense of its location, showing that communicative cues modulate attention to and encoding of the properties of an object in adults as well. Moreover, a recent study by Okumura et al. (Okumura et al., 2020) reported that, although both attentional cues (such as a beep) and ostensive cues affected infants' gaze-following, only ostensive cues facilitated their referential object learning. In line with these results, studies that have explored the effects of ostensive cues on infants' tendency to follow others' gaze toward objects, further showed that children were more likely to follow the agent's gaze if it was preceded by ostensive cues (Senju & Csibra, 2008), even when the agent was a robot (Okumura et al., 2013). In the experiments by Okumura et al. (Okumura et al., 2013), 12-month-olds watched videos in which a human or a robot looked toward an object. Their aim was to examine whether robots can influence infants' learning and, given the empirical evidence that has demonstrated the importance of verbalizations in establishing joint attention in infant-adult interactions (Parise et al., 2007), the authors added ostensive verbal signals while the robot gazed at an object. Results showed that when the robot's gaze was accompanied by ostensive verbal cues, children not only followed the direction of the robot's gaze but also paid preferential attention to the object when the ostensive cue was present. Based on this evidence, Natural Pedagogy theorists have argued that children encode information differently depending on whether it is presented in a communicative context compared with a non-communicative context.

Shared Knowledge

Within the conceptual framework of Shared Knowledge (Csibra & Gergely, 2009), in communicative contexts, children would assign an object-centered interpretation to individuals' object-directed emotions. Namely, the addressee of ostensive communication would focus on the intrinsic characteristics of the object and this kind of interpretation would allow children to a) act in an emotionally consistent way not only toward the referent in the here and now, but also in the future situations, and b) expect other people to share the same emotional disposition and act accordingly toward the same type of referent (Egyed et al., 2013). Ostensive signals increase the likelihood that the information provided will be generalized to other circumstances or interactions (Schröder-Pfeifer et al., 2018). To exemplify the concept, Egyed et al. (Egyed et al., 2013) provided the example of a snake: the parent who sees their child approaching a snake will show an expression of fear towards it to warn the child of the danger. The adult intends to communicate to the child that the snake is dangerous to approach. By addressing the child with ostensive signals, the child will assign to the referent an object-centered interpretation generalizable to future situations and other individuals, i.e., an awareness that snakes are dangerous. On the other hand, when the expression of fear is observed in a non-communicative context, infants would assign a person-centered interpretation to the snake that would lead them to not generalize the emotional disposition as applicable to other individuals but would interpret the object-directed emotion as an emotional attitude of that person (e.g., my parent is afraid of snakes) (Egyed et al., 2013). Previous works have suggested that children, even at very early ages, are able to flexibly assign person- and object-centered interpretations to the display of referential emotions depending on whether they are shown in a communicative or non-communicative context (Egyed et al., 2013; Träuble & Bätz, 2014). For instance, Egyed and colleagues (Egyed et al., 2013) found that 18-month-old infants flexibly assign person- and object-centered interpretations according to the context in which the emotion was displayed. The experiment consisted of an actress displaying positive versus negative emotions toward two objects differing in their shape and color; then, the same or another actress made a request for one of the objects. After being addressed in an ostensive communicative manner, infants were more likely to choose the object with a positive valence in response to the unknown actress' request. On the other hand, when the object-directed emotion was displayed within a non-communicative context (i.e., infants were not directly addressed), infants did not generalize the object-directed emotion when responding to the different actress's object request. The results suggested that 18-month-olds interpret expressions of emotion toward an object communicated in an ostensive way as revealing general valence information about the object that is also relevant to and shared by other people. In other words, infants assigned to the object-directed emotion an object-centered interpretation. In contrast, when the same emotion expression is displayed in a non-communicative context, infants' interpretation is person-centered, i.e., infants interpret the object-directed emotion as a person-specific attitude or a personal preference (she likes it/she does not like it) and episodic.

Aim of the study

The aim of the present study is twofold: 1) to investigate whether the Shared Knowledge assumption is a feature of human communication and thus it is found in adulthood; 2) to evaluate whether ostensive cues acted upon by a robotic agent may lead to effects beyond mere attentional arousal and whether the Shared Knowledge assumption persists in human-robot interaction (HRI) as well. These questions were inspired by the increasing use of robotic agents in educational settings, which demands an effort in understanding the mechanisms underlying HRI. Numerous studies have contributed to our understanding of how people interact with robots in educational contexts (Baxter et al., 2017; Belpaeme et al., 2018; Breazeal et al., 2016; Cangelosi & Schlesinger, 2018; Di Dio, Manzi, Itakura, et al., 2020; Di Dio, Manzi, Peretti, et al., 2020b; Kanda et al., 2004; Kont & Alimardani, 2020). These studies have shown that the attribution of mental abilities and psychological traits to robots (Di Dio, Manzi, Peretti, et al., 2020b) (for a review, see also (Marchetti et al., 2018)) and the human-likeness, play a significant role in establishing trust and facilitating human-robot interaction (Di Dio, Manzi, Peretti, et al., 2020a; Manzi, Peretti, et al., 2020; Manzi, Di Dio, et al., 2021; Manzi, Sorgente, et al., 2021; Vinanzi et al., 2019; Złotowski et al., 2015). Some fundamental mechanisms of social cognition, such as eye gaze (Manzi, Ishikawa, et al., 2020; Okumura et al., 2013, 2020) and joint attention (Chevalier et al., 2020), have been studied using humanoid robots, but little is actually known about the effect of ostensive cues on the conveyance of relevant information and related generalization processes acted by a robot.

Although there is little work in the state of art investigating the hypotheses of Natural Pedagogy in adults, the promising results found by Marno and colleagues (Marno et al., 2014) suggest that communicative and non-communicative contexts do not exclusively exert their effects on infants. Based on these assumptions, we hypothesized that the Shared Knowledge assumption described above reflects a feature of human communication, and if so, it should persist in adulthood. In addition, we wondered whether the Shared Knowledge assumption is restricted only to human ostensive cues. That is, what happens when a robot ostensively engages a person through eye contact and greetings? What effects do ostensive cues act by a robot exert on adult participants? Our study investigated these questions by directly comparing conditions in which a human or a robot displayed object-directed emotions in both communicative and non-communicative contexts. To investigate these hypotheses, one-hundred and ninety-three (193) Italian adult participants (age range = 18-61 years) were involved in the study. We have, therefore, developed a paradigm inspired by the work of Egyed, Kiràly, and Gergely (Egyed et al., 2013) to test whether the Shared Knowledge assumption persists into adulthood and whether this phenomenon is activated when a robotic agent acts as a communitive partner. We expected to replicate the results of the original work regarding the persistence of the genericity bias in adulthood. With respect to the robot condition, we hypothesized that robot ostensive communication acted by a robot that goes beyond mere attentional arousal and might influence participants' likelihood of sharing the positively valenced object. We have modified the original paradigm (Egyed et al., 2013) to adapt it to adult participants and online administration. In order to ensure optimal control over the actions and behaviors of both the human and robot participants in this initial study, we opted for a video-based version of the interaction, where several parameters could be properly manipulated and controlled. Assuming a significant effect of the robot agent on the participant's behavior, it would be conceivable to evaluate the behavior in a more ecological setting. We recorded video clips representing humans and robots acting as the demonstrator and requester (the one performing the object-directed emotion) and the requester (the one making the request to share). We split the sample into those who underwent the robot-demonstrator condition and those who underwent the human-demonstrator condition. The actor (human or robot) displayed two different emotions, one with positive and the other with negative valence, toward two different unfamiliar objects. We involved two social robots, namely QT Robot and NAO, to play the role of demonstrator and requester in the experimental condition. Crucially, during the familiarization phase, the demonstrator displayed emotions in a communicative (the human or the robot ostensively greeted the participant) or non-communicative context (the human or the robot acted as if alone); in the test phase, participants saw the requester make a request by reaching his/its hand (the requester could be the same, a different person/robot or another agent depending on the demonstrator's agency) and, subsequently, chose which object sharing with the requester. After the Shared Knowledge task, we administered two questionnaires assessing participants' attachment style and mentalization ability, the Attachment Style Questionnaire (ASQ; Feeney et al., 1994) and the Reflective Functioning Questionnaire (RFQ; Fonagy et al., 2016) respectively. Participants' attachment style and mentalization ability were assessed because they are constructs intrinsically linked to ostensive communication. Finally, we administered the Attribution of Mental States questionnaire (AMS-Q; Miraglia et al., 2023), a tool that assesses the degree of mental anthropomorphism of nonhuman agents.

Methods

Participants

One-hundred and ninety-three (193) Italian adult participants (Mean age = 27.98 years, SD = 8.89, age-range = 18-61 years) took part in the study. Inclusion criteria for all participants were age of majority and being a native Italian speaker. See Table 1 for details of the demographic characteristics of the sample. The participants were recruited on Prolific for 5.50$ per hour. The platform allows participants to be selected based on nationality and other characteristics of interest (e.g., absence of special medical conditions).

Participants were informed about the experimental procedure, the measurement items, and the materials. All participants gave written informed consent in line with the Declaration of Helsinki and its revisions and in accordance with the requirements of the ethics committee of the Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy, which approved this study.

Table 1 – Sample socio-demographic characteristics

Sociodemographic characteristics		
Age, mean ± SD		27.98 ± 8.89
Gender	N (%)	
Male		97 (50.3%)
Female		96 (49.7%)
Residence	N (%)	
Northwest Italy		56 (29.0%)
Northeast Italy		43 (22.3%)
Centre Italy		46 (23.8%)
South Italy		23 (11.9%)
Sicily and Sardinia		25 (13.0%)
Educational level	N (%)	
Middle school		2 (1.0%)
High school		96 (49.7%)
Graduate school		86 (44.6%)
Postgraduate school		9 (4.7%)
Employment status	N (%)	
Student		96 (49.7%)
Employed		74 (38,3%)
Unemployed		17 (8.8%)
Other		6 (3.1%)

Procedure and task

General procedure

In the current study, we modified the paradigm used by Egyed and colleagues (Egyed et al., 2013) to adapt it to adult participants and online administration. Participants were assessed under two experimental conditions: the sample was split into those who underwent the robot-demonstrator condition and those who underwent the human-demonstrator condition. In both the human and robot-demonstrator conditions, the experiment began with a familiarization phase in which object-directed emotion displays were presented in a communicative context, i.e., the demonstrator ostensively engaged with participants through eye contact and greetings, or non-communicative context, i.e., the demonstrator acted as if alone, without looking into the camera nor greeting participants. This was followed by a test phase in which the requester made a request by extending his/its arm toward participants and asking them to give an object. The object that is positively valenced by the demonstrator is referred to as the target object. As in the original work (Egyed et al., 2013), we varied the identity of the requester, who might be the same person who showed the expressions of referential emotions or a different person; in addition, we also varied the genus of the requester which could be a human or a robot. We involved two social robots, i.e., QT Robot and NAO (Figure 1), to play the role of demonstrator and requester in the experimental condition.

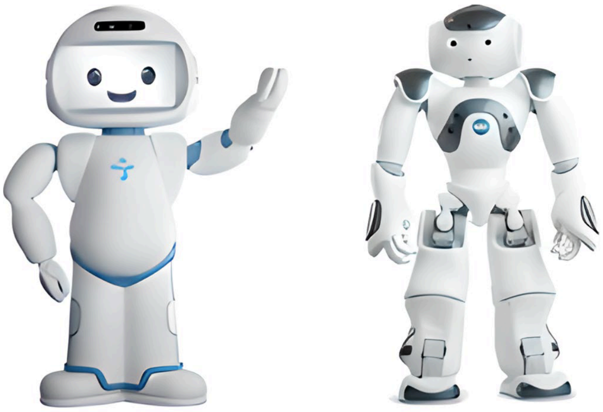


Figure 1 - QT Robot (on the left) and Nao (on the right).

We, therefore, created six experimental conditions: 1) communicative-context/same-person condition; 2) communicative-context/different-person condition; 3) communicative-context/different genus condition; 4) noncommunicative-context/same-person condition; 5) noncommunicative- context/different-person condition; and 6) noncommunicative-context/different genus condition. All conditions were administered in random order and were semi-balanced by agent (human-human, human-robot) and by role (demonstrator, requester). All conditions are illustrated in Figure 2 and 3.

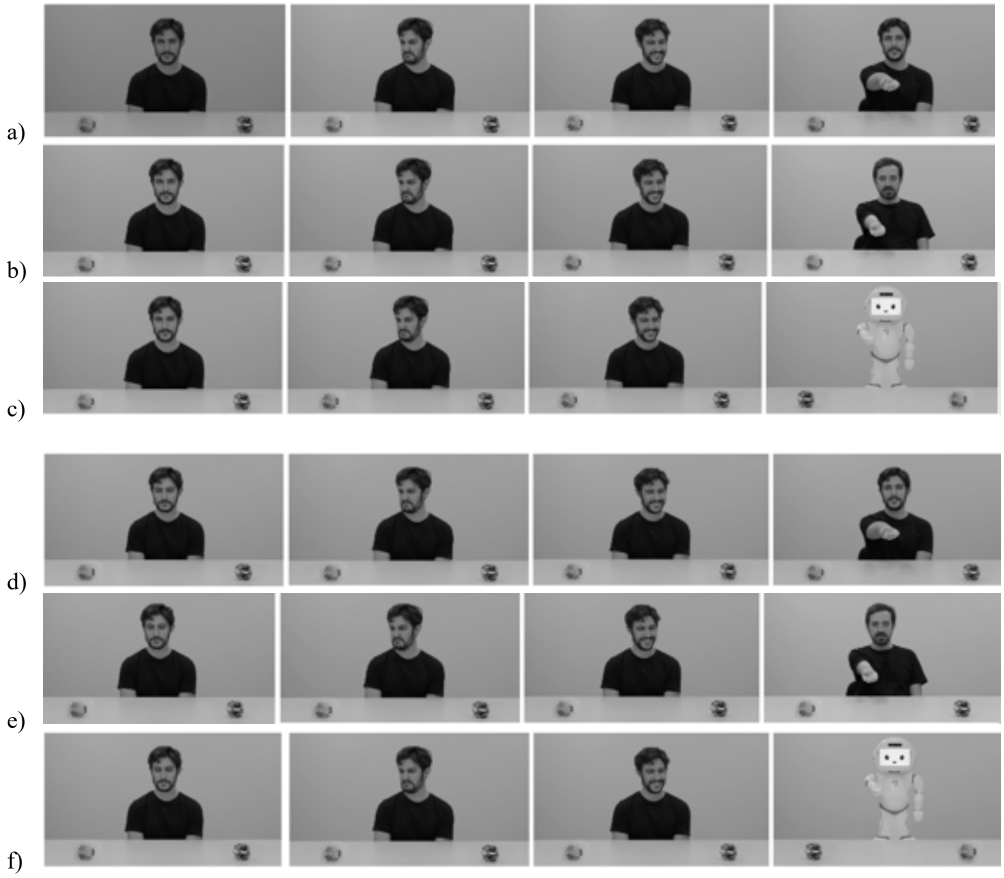


Figure 2 - Human-demonstrator conditions: a) communicative context, same person; b) communicative context, different person; c) communicative context, robot; d) non-communicative context, same person; e) non-communicative context, different person; f) non-communicative context, robot.

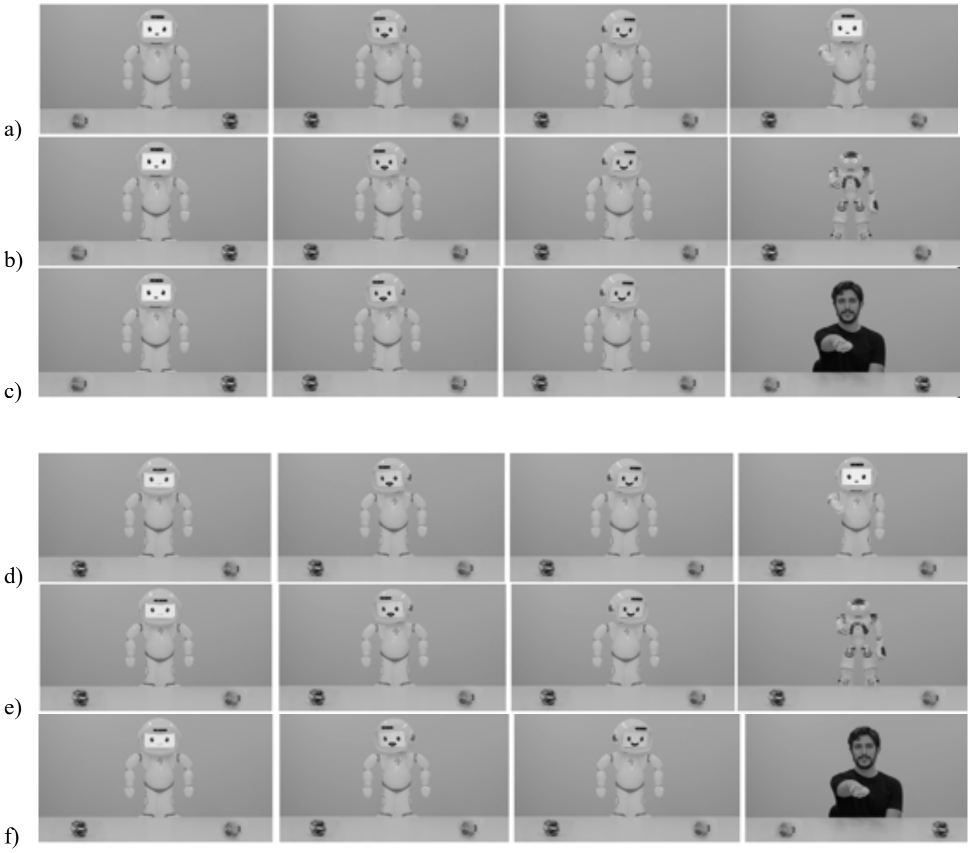


Figure 3 - Robot-demonstrator conditions: a) communicative context, same robot; b) communicative context, different robot; c) communicative context, human; d) non-communicative context, same robot; e) non-communicative context, different robot; f) non-communicative context, human.

After the Shared Knowledge task, participants were administered the following tests: the Attribution of Mental States questionnaire (AMS-Q) (Miraglia et al., 2023), the Reflective Functioning Questionnaire (RFQ) (Fonagy et al., 2016) (Italian version: (Morandotti et al., 2018)), and a short version of the Attachment Style Questionnaire (ASQ) (Feeney et al., 1994) (Italian version: (Fossati et al., 2003)). The questionnaires were administered in random order.

Experimental conditions: Shared Knowledge task

The design of the Shared Knowledge task was a 2x3x2 repeated measures mixed model, with 2 levels of context (communicative, non-communicative), 3 levels of requester (same identity – i.e. demonstrator and requester were the same person/robot; different identity – within agency, i.e., if the demonstrator was a human the requester was another human; if the demonstrator was a robot the requester was another robot; other identity – between agency, i.e., if the demonstrator was human, the requester was a robot and vice-versa) as the within-subject factors, and 2 levels of demonstrator (human, robot) as the between-subject factor.

The sample was initially split into two groups (between-subject factor): those who underwent the robot-demonstrator condition and those who underwent the human-demonstrator condition. Within each group, the participant watched six short video clips with a 24-second duration each (each frame of the video has the same duration), showing different conditions. Each condition differed in the type of context (communicative vs non-communicative, i.e., the demonstrator gazing toward and verbally engaging the participant prior to emotions display vs. a non-engaging approach); the demonstrator's agency (human vs. robot); and requester: same agent (i.e., demonstrator and requester were the same person/robot), different identity (demonstrator and requester had the same agency – human or robot – but a different identity), and finally, other agent (the demonstrator and the requester could be human and robot respectively, or robot and human, thus counterbalanced by role). The experimental condition was as follows:

- Familiarization phase: in this initial phase, a human or robotic demonstrator displayed an object-directed emotion, expressing joy toward one object and then turning toward the other, presenting an emotional expression of disgust (the objects are described in Stimuli). This sequence was repeated a second time. Before showing the emotions, in the communicative context, the demonstrator ostensively addressed the participants through eye contact (looking into the camera) and smiling while greeting them, saying, "Hi! Pay attention". In the non-communicative context, the demonstrator never interacted with the participants: the human or robotic actor never looked at or talked to participants either before or during her object-directed expressions of emotions.
- Test phase: In this subsequent phase, the human or robotic requester communicatively addressed the participants using ostensive signals (looking, smiling, greeting); he/it then displayed a hand request gesture (reaching out and placing his/its hand between the two objects with the palm facing upward), and said, "Give me one of them!". Throughout the test phase, the requester would only look at the camera and never at the objects. At this point, participants had to select which object they would like to share with the requester.

Stimuli

Two unfamiliar objects with different colors, different shapes, and similar affordance properties for both humans and robots were used (about 6.18 inch). Their left-right position on the table and the demonstrator's emotion associated with them were counterbalanced among conditions (Figure 4).

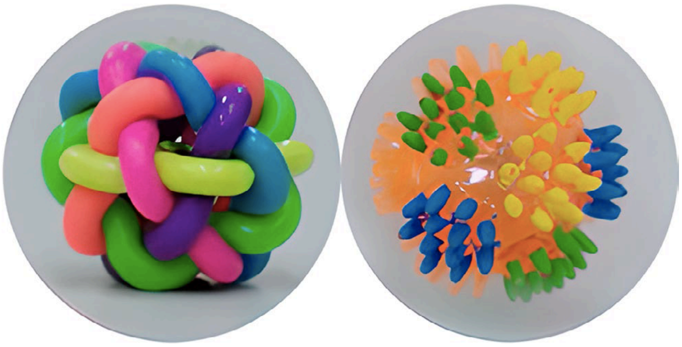


Figure 4 – Unfamiliar objects presented in the Shared Knowledge task: object A on the left; object B on the right.

Control conditions

Before starting with the experimental conditions (see above), control conditions were administered to evaluate object preferences, robot gender recognition, and emotion labeling. More specifically, prior to carrying out the experimental conditions, the participants had to express their liking of the two objects, as well as indicate the gender of the robot. Additionally, to ensure that the demonstrators' expressed emotions were clearly recognizable, participants were asked to view pictures of the human and robot demonstrators while expressing emotions of joy and disgust. Participants might choose from six different emotions, specifically: joy, anger, surprise, sadness, disgust, and fear.

Correlated assessments

Besides the Shared Knowledge task, the protocol included the Attribution of Mental States (AMS), the Reflective Functioning Questionnaire (RFQ), and a short version of the Attachment Style Questionnaire.

The Attribution of Mental States questionnaire (AMS-Q) (Miraglia et al., 2023). AMS-Q is a 23-item questionnaire that evaluates the attribution of mental and sensory states to pictures of a human stimulus (female or male). The tool measures the degree of mental anthropomorphization of the non-human agents (e.g., animals, robots, inanimate objects, paranormal entities, and even God) by comparing the scores obtained from the human stimulus with those obtained from the non-human stimuli. The AMS-Q consists of three subscales: AMS-NP, which reflects the attribution of epistemic mental states (e.g., beliefs, thoughts, inferences), well-being states, and positive emotions; AMS-N, which includes the attribution of mental states that belong to the semantic field of deception (e.g., tell a lie) and negative emotions; and AMS-S which refers to sensory states (e.g., hear, smell). Participants were asked to rate each item according to a 5-point Likert scale ranging from 1 (No, not at all) to 5 (Yes, very much). The scoring was calculated by averaging the items for each factor. The AMS-Q has been used in previous work (Di Dio et al., 2018; Di Dio, Manzi, Itakura, et al., 2020; Di Dio, Manzi, Peretti, et al., 2020b; Manzi, Peretti, et al., 2020) and has been shown to be a consistent measure in the attribution of mental states to both humans and robots. The questionnaire was administered twice with a human and a robot image as stimuli in random order. The reliability of the scale was excellent, with a Cronbach alpha coefficient reported of .96 and .91, respectively.

Reflective Functioning Questionnaire (RFQ) (Fonagy et al., 2016). The brief version of the RFQ comprises two subscales, assessing the degrees of uncertainty (RFQ_U) and certainty (RFQ_C) about mental states. The Italian brief version (Morandotti et al., 2018) is composed of 8 items that are scored by the participant on a 7-point Likert scale (ranging from "completely disagree" to "completely agree"). As a result, the low agreement reflects hypermentalizing, while some agreement reflects adaptive levels of certainty about mental states. The internal consistency of the sample test was acceptable, with a Cronbach alpha coefficient reported of .71.

The Attachment Style Questionnaire (ASQ) (Feeney et al., 1994; Fossati et al., 2003). ASQ is a 40-item self-report questionnaire, designed to measure five dimensions of adult attachment: Confidence in Self and Others (8 items), Discomfort with Closeness (10 items), Relationships as Secondary (7 items), Need for Approval (7 items), and Preoccupation with Relationships (8 items). Each item is rated on a 6-point scale, ranging from 1 (totally disagree) to 6 (totally agree). In the current study, we administrated three out of the five subscales were administered, i.e., Trust, which reflects a secure attachment orientation; Need for Approval, which reflects respondents' need for acceptance and confirmation from others; and Concern for Relationships, which involves an anxious and dependent approach to relationships. In the current study, the Cronbach alpha coefficient was .80, showing good internal consistency of the sample test.

Data Analysis

A General Linear Model (GLM) analysis was performed to assess the impact of context (communicative vs non-communicative) and requester (same requester, different requester, another agent) on participants' choice of target objects under two conditions: human-demonstrator condition and robot-demonstrator condition. The proportion of congruent responses, i.e., when participants chose to share the object with positive valence, was the dependent variable. The participants' object preference was then used as a covariate in the 2x3x2 repeated measures GLM, with 2 levels of context (communicative, non-communicative), 3 levels of requester (same identity; different identity, other identity) as the within-subject factors; and 2 levels of demonstrator (human, robot) as the between-subject factor. The Greenhouse-Geisser correction was used for violations of Mauchly's Test of Sphericity ($p < .05$). Post-hoc comparisons were Bonferroni corrected.

Furthermore, a GLM analysis was used to assess whether the participants discriminated between the human and robot's mental states, whereas independent binomial logistic regressions were carried out to assess possible predictive effects of the participants' reflective functioning skills and attachment style on responses in the Shared Knowledge task.

Results

Object preference

Participants were asked to express their liking for the two objects. They showed a significant preference for object A: 48.2% responded "like" for object A vs 14.5% for object B. For object B response distribution mainly fell between "neutral" (29.5%) or "like a little" (33.7%). Object preference was controlled both by design through randomization of objects' location and associated emotions (see Methods above), and by statistics, i.e., by including the variable "object preference" as a covariate in the GLM analysis carried out to examine participants' responses in the Shared Knowledge task.

Robot gender

As the actors involved in the study were men, we assessed the participants' perception of the robot's gender to ensure the gender match. Overall, the robot (QT robot) employed in the study as the demonstrator has been correctly identified as male (78.8%). About 20% answered "don't know", showing that some people do not consider the robot to belong to a specific gender. These data are in line with the results of a preliminary pilot study also evaluating people's identification of the robots' gender.

Emotion recognition

The emotions expressed by humans and robots were correctly recognized by most participants. Those who did not correctly name the emotions still correctly indicated the positive or negative valence of the observed emotion (i.e., joy, disgust, fear, sadness, surprise, and anger). The proportions of emotion recognition are given in Table 2.

Table 2 – Proportions in Emotion Labelling in the Emotion Recognition Task

Joy		N (%)	T-test (*p*)
	Human 1	168 (87.0%)	<.001
	Human 2	175 (90.7%)	<.001
	QT Robot	188 (97.4%)	<.001
Disgust		N (%)	T-test (*p*)
	Human 1	171 (88.6%)	<.001
	Human 2	153 (79.3%)	<.001
	QT Robot	178 (92.2%)	<.001
Fear		N (%)	T-test (*p*)
	Human 1	2 (1.0%)	<.001
	Human 2	0 (0%)	<.001
	QT Robot	3 (1.6%)	<.001
Sadness		N (%)	T-test (*p*)
	Human 1	3 (1.6%)	<.001
	Human 2	1 (0.5%)	<.001
	QT Robot	7 (3.6%)	<.001
Surprise		N (%)	T-test (*p*)
	Human 1	17 (8.8%)	<.001
	Human 2	16 (8.3%)	<.001
	QT Robot	5 (2.6%)	<.001
Anger		N (%)	T-test (*p*)
	Human 1	15 (7.8%)	<.001
	Human 2	39 (20.2%)	<.001
	QT Robot	1 (0.5%)	<.001

Main Analysis: Shared Knowledge task

The binomial analysis first revealed that the proportion of congruent responses was significantly above the chance level for all conditions (p <.001), indicating that the object with positive valence was more likely to be shared with the requesters independently of context (communicative vs. non-communicative), requester's identity (same, different, other), and demonstrator's genus (human, robot). Additionally, by introducing "object preference" as a covariate in the GLM analysis below, the results further showed no substantial correlations between the participants' responses in the Shared Knowledge task and object preference (p >.05), thus indicating that object preference did not influence participants' choices (see also analysis of covariates below). Values of the GLM with and without covariate are reported in Table 3.

Table 3 – Scores of robot demonstrator and human demonstrator conditions with and without object preference as a covariate

Demonstrator	**Context**	**Requester**		Pairwise comparisons without covariate		Pairwise comparisons with covariate	
				Mdiff	***SE***	***Mdiff***	***SE***
Robot	Communicative	Same robot	Different robot	.07	.05	.07	.05
			Human	**.13***	**.05**	**.14***	**.05**
		Different robot	Same robot	-.07	.05	-.07	.05
			Human	.06	.05	.06	.05
		Human	Same-Robot	**-.13***	**.05**	**-.14***	**.05**
			Different robot	-.06	.05	-.06	.05
	Non-Communicative	Same robot	Different robot	**-.16***	**.05**	**-.16***	**.05**
			Human	.01	.05	.01	.05
		Different robot	Same robot	**.16***	**.05**	**.16***	**.05**
			Human	**.17***	**.05**	**.16***	**.05**
		Human	Same robot	-.01	.05	-.01	.05
			Different robot	**-17***	**.05**	**-.17***	**.05**
Human	Communicative	Same person	Different person	.05	.05	.05	.05
			Robot	-.01	.05	-.01	.05
		Different person	Same person	-.05	.05	-.05	.05
			Robot	-.06	.05	-.06	.05
		Robot	Same person	.01	.05	.01	.05
			Different person	.06	.05	.06	.05
Human	Non-communicative	Same person	Different person	**.13***	**.05**	**.13***	**.05**
			Robot	.08	.05	.09	.05
		Different person	Same person	**-.13***	**.05**	**-.13***	**.05**
			Robot	-.04	.05	-.04	.05
		Robot	Same person	-.08	05	-.09	05
			Different person	.04	05	.04	05

The main results of the GLM analysis related to the Shared Knowledge task showed a significant interaction between demonstrator and context, F(1, 189) = 38,81, p < .001, partial-η2 = .17, δ = 1, and demonstrator and requester, F(2, 188) = .85, p < .05, partial-η2 = .06, δ = 90, suggesting a difference in the effectiveness of the ostensive cues and in the role played by human and robot in the processes of shared knowledge. Also, a significant three-way interaction was found between context, demonstrator, and requester, F(2, 188) = 38,81, p < .05, partial-η2 = .05, δ = .80.

First, under the human-demonstrator condition, pairwise comparisons showed that participants were more likely to share the target object (i.e., the positively valenced object) with the same person that acted as the demonstrator in the non-communicative context than in the communicative context, Mdiff=.17, SE=.05, p<.001. These data partially support the results of the original work ((Egyed et al., 2013); Figure 6). Also, a significant difference was found between sharing with the same person acting both as demonstrator and requester rather than with different human acting as a requester in the non-communicative context, Mdiff=.13, SE=.05, p < .05.

Secondly, under the robot-demonstrator condition, pairwise comparisons showed that participants were more likely to share the target object with the same robot that acted as a demonstrator in a communicative than a non-communicative context, Mdiff=.26, SE=.05, p<.001. Furthermore, the target object was more likely to be shared with a human in the communicative than non-communicative context, Mdiff=.13., SE=.05, p<.05. Within the communicative context, the target object was more likely to be shared with the same robot that acted as demonstrator than with the human, Mdiff= .14, SE=.05, p< .05; whereas, in the non-communicative contexts, it was more likely to be shared with the other robot than with both the same robot, Mdiff=.16, SE=.05, p<.05, and the human, Mdiff= .17, SE= .05, p<.05 (Figure 5).

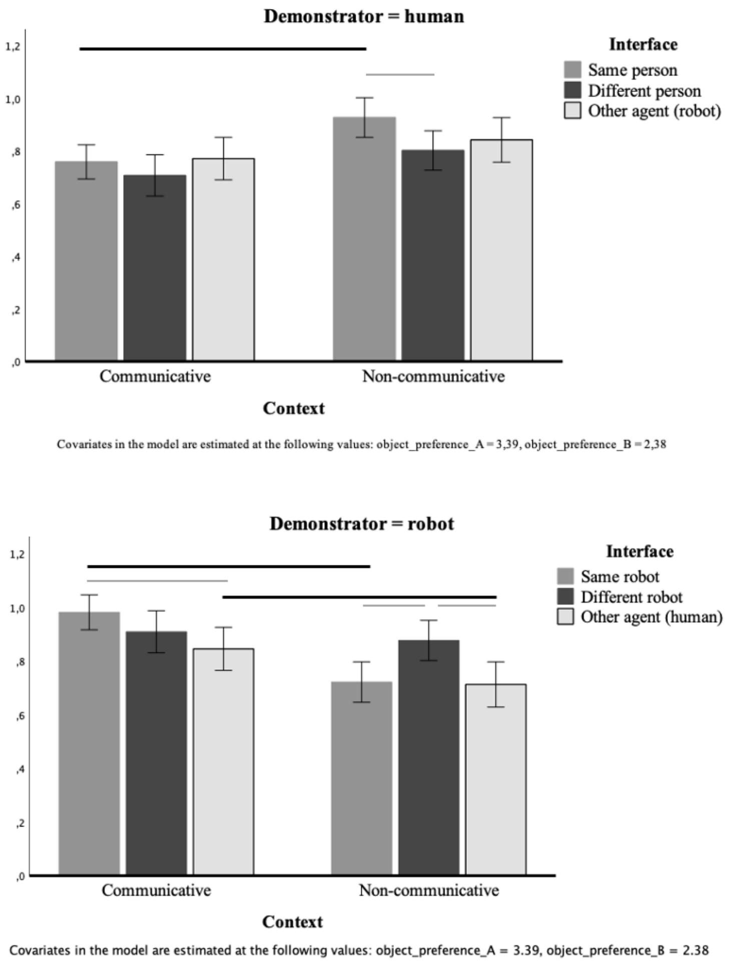


Figure 5 - interaction effect divided by the demonstrator's agency (human, robot), highlighting the differences between requesters within and between communicative contexts.

Also, in the communicative context, pairwise comparisons showed that participants were more likely to share the target object with the robot (either the same or different) when the robot played as the demonstrator than with the human (either the same or different), Mdiff= .22, SE= .05, p<.001; Mdiff= .20, SE= .06, p<.001, respectively. In contrast, in the non-communicative context, participants were more likely to share the target object with the same human as the requester than with the same robot acting as both demonstrator and requester, Mdiff= .21, SE= .05, p<.001, and to share with the human when the robot was the demonstrator and with a human when the robot was the demonstrator than with the robot when the demonstrator was human.

AMS-Q

To assess whether the human and robot were perceived as distinct entities from a mental content and sensory attributes perspective, a 2x2 GLM analysis was carried out, with two levels of AMS (mental states, sensorial states) and two levels of agent (human, robot). The results showed a main effect of AMS, F(1, 191) = 58.70, p < .001, partial-η2 = .24, δ = 1, indicating a greater attribution of sensory than mental states, and a main effect of agent, F(1, 191) = 2548.12, p < .001, partial-η2 = .93, δ = 1, indicating that the robot scored significantly lower than the human in states attribution. A significant interaction between AMS and agent, F(1, 191) = 63.8, p < .001, partial-η2 = .25, δ = 1, also showed that – for the robot – sensory states attribution was substantially greater than mental states attribution, Mdiff=.43, SE=.05, p<.001. This difference was not present for human, p>.05.

Logistic regression

Before running the binomial logistic regressions to assess possible predictive effects on participants' responses, we carried out Pearson's correlation analysis examining the relations between the Shared Knowledge task and participants' reflective functioning skills and attachment style. The analysis yielded a relation between the non-communicative/other-agent condition when the robot was the demonstrator, and Certainty (RFQ_C) about the mental states of self and others, r(97) = 0.22, p

<.01. A significant negative relationship was found between the non-communicative context/other- agent condition when the human was the demonstrator and the subscale of ASQ, namely, Need for Approval, r(96) = -0.27, p <.01; while, in the communicative context, the specular condition was moderately associated to the Concern about relationships of ASQ, r(96) = 0.21, p <.05.

Based on these results, we carried out binomial logistic regression on three conditions of the Shared Knowledge task with Bonferroni adjustments (p values <.016 considered significant) that correlated with the RFQ and ASQ. The model included five independent variables (RFQ_C, RFQ_U, ASQ-Trust, ASQ-Need for Approval, and ASQ-Concern about relationship). Linearity of the continuous variables with respect to the logit of the dependent variable was assessed via the Box- Tidwell procedure: all continuous independent variables were found to be linearly related to the logit of the dependent variable. We ran three independent logistic regression models for each condition, to outline possible predictive effects of the participants' reflective functioning skills and attachment style on responses in the Shared Knowledge task.

1) Robot demonstrator – human requester, non-communicative context. When the robot was the demonstrator, in the non-communicative context/other agent (i.e., when the requester was a human) condition, there was one standardized residual with a value of

-3.28 standard deviations and the associated case was deleted from the analysis. Thus, binomial regression was performed again. The full model containing all predictors was statistically significant, $\chi2(5, N = 96)$, 14.43, $p = .013$, indicating that the model is able to distinguish between those who gave a coherent answer versus those who gave an incoherent answer. The model correctly classified 76.0% of cases, indicating the correct identification of the coherent answer (i.e., the positive valence of the object conveyed by the robot). As shown in Table 4, only two of the independent variables made a statistically significant contribution to the model: RFQ_C and ASQ-Need for approval.

2) Human demonstrator – robot requester, non-communicative context. Similarly, when the human was the demonstrator, in the non-communicative context-other agent (i.e., when the requester was a robot) condition, there were two standardized residuals with a value of -4.59 and -3.88 standard deviations, which were discarded. Once again, the binomial regression was performed, and the assumption of linearity was not violated. The logistic regression model was statistically significant $\chi2(5, N = 94)$, 18.32, $p = .003$. The model correctly classified 88.3% of cases, indicating the correct identification of the coherent answer (i.e., the positive valence of the object conveyed by the human). In this condition, the predictor variable of sharing the positive valence conveyed by the human was the Attachment Style Questionnaire's need for approval, recording an odds ratio, Exp(B), of .77 (Table 4).
3) Human demonstrator – robot requester, communicative context. Lastly, when the demonstrator was acted by a human, in the communicative context-other agent (i.e., when the requester was a robot) condition, three standardized residuals with values above 2.5 were eliminated as these were clear outliers. The new binomial logistic regression showed that the full model containing all predictors was statistically significant, $\chi2(5, N = 93)$, 14.25, $p = .014$. The model correctly classified 82.8% of cases, indicating the correct identification of the coherent answer (i.e., the positive valence of the object conveyed by the human). Also in this condition, two of the independent variables made a statistically significant contribution to the model: ASQ-Need for approval and ASQ-Concern about relationships (Table 4).

Table 4 – Logistic regression predicting the likelihood of sharing object

		B	SE	Wald	*df*	*p*	Exp(B)	95% CI for EXP(B)	
								Lower	Upper
Robot_dem, Non-comm, Human_req	ASQ_NeedApp	-.15	.06	5.93	1	.015	.86	.77	.97
	RFQ_C	-.18	.07	6.15	1	.013	.83	.72	.96
Human_dem, Non-comm, Robot_req	ASQ_NeedApp	-.26	.9	8.76	1	.003	.77	.65	.92
Human_dem, comm, Robot_req	ASQ_NeedApp	-.12	.06	3.71	1	.05	.89	.79	1
	ASQ_CAR	.17	.06	7.37	1	.007	1.18	1.05	1.33

Note: ASQ_NeedApp = Need for Approval (ASQ); ASQ_CAR = Concern about Relationships (ASQ); FRQ_C = Certainty about the mental states of self and others.

Discussion

The present study aimed to assess whether the Shared Knowledge assumption, driven by Natural Pedagogy Theory, is a general feature of human communication and therefore persists into adulthood. Additionally, this study aimed at better understanding human-robot interaction, and, for this purpose, we investigated whether the robot, when using ostensive signals, prepares the addressee for the intent to communicate generalizable information. To this end, we developed a paradigm inspired by Egyed et al.'s work (Egyed et al., 2013), in which participants, after having observed the agent's positive or negative emotions toward two objects, had to decide which object to be shared with a requester. Generally, the results of the present study showed 1) that the process of Shared Knowledge previously evaluated in children persists into adulthood, and 2) a fairly different pattern of behavior, when the demonstrator was human or robot, i.e., the positively valenced object (target object), was shared differently depending on whether the demonstrator or requester was a human or a robot, as well as on the contexts in which the demonstrator delivered the information (i.e., communicative vs. non-communicative).

Human-demonstrator condition

Under the human-demonstrator condition, the results generally support the conclusions of the original work (Egyed et al., 2013), thus generalizing the paradigm carried out in presence of young children to adults. Specifically, we found that the target object was more likely to be shared with the same person than with another in the non-communicative contexts. Additionally, we further found that the tendency to share with the same person in the non-communicative context is even greater than sharing with the same person in the communicative context (this condition was absent in the original work). According to Egyed and colleagues (Egyed et al., 2013), our data would support the Shared Knowledge assumption of Natural Pedagogy Theory, which assumes that in a non-communicative context, children do not generalize the agent-specific attributions as applying to other individuals. Moreover, in the absence of ostensive cues, children assign a person-centered interpretation, whereby they interpret the received emotional information as valid only in relation to the referent in the current episodic situation. This idea is further supported by data from our study showing a lack of differences in object-sharing between requesters specifically in the communicative context: under the human-demonstrator condition, the target object was almost equally shared with the same person, with a different person, or even with the robot. As postulated by Natural Pedagogy Theory, communicative contexts place the addressee in an attentional state and prepare him or her to receive a subsequent communication containing information specifically relevant to him or her that should be remembered and encoded with other knowledge relevant to social situations (Csibra & Gergely, 2009; Egyed et al., 2013; Fonagy & Allison, 2014). Crucially, the ostensive cues that typically lead the infant to feel recognized as a subject (Fonagy & Allison, 2014), appear to exert their effects even on adults. Our results, consistent with Marno and colleagues' study (Marno et al., 2014), seem to confirm that ostensive signals have effects beyond simple attentional arousal but prepare the addressee for generalizable knowledge in adult communication as well. Additionally, ostensive cues facilitate the relationship between demonstrator and requester as they trigger the epistemic trust that allows the addressee of object-directed emotion to trust the authenticity of the shared information. Thus, it is plausible to claim that ostensive communication triggers a sense of trust in the person conveying the information as a benevolent, cooperative, and reliable source of cultural information.

Robot-demonstrator condition

The results paint a quite different picture when the robot was the demonstrator. Opposite to what is described above, the target object was more likely to be shared with the same robot in the communicative context than in the non-communicative context. In the communicative context, the target object was shared equally with the same robot and with a different robot and was less likely to be shared with the human requester. When the robot displayed an object-direct emotion, the human requester appeared as unprivileged as participants were less inclined to consider information received as generalizable to humans, but conversely applicable to any other robots. These results seem to suggest that the information the robot conveys might be considered "robot-specific". A possible explanation for this finding lies in the Theory of Natural

Pedagogy that the expectation of learning generalizable knowledge is driven by members of the same social group (Csibra & Gergely, 2009, 2011; Csibra & György, 2006) and, as a matter of fact, robots are not perceived as belonging to the same social group as humans (as evidenced by the data of the AMS-Q). A similar tendency was observed in the non-communicative context, in which the target object was less shared with the human requester. In contrast to the human-demonstrator condition, in the non-communicative context, participants did not generalize the object positively valenced by the robot to the same robot - as postulated by Natural Pedagogy Theory - but rather generalized the target object to the different robot. These findings suggest that when dealing with robots, Natural Pedagogy assumptions on the genericity bias are no longer valid. Some other processes would be – as per our data – in place.

Overall, the data collected in the robot-demonstrator conditions bring out the crucial role of ostensive cues in human-robot interactions. We asked whether and how robotic ostensive cues may influence interactions with humans. Although participants were not inclined to generalize the object-directed emotion displayed by the robot to the human requester, in the communicative context participants "listened to the robot" by paying attention to the expressed preference and sharing it with the same or different robot. In contrast, in the non-communicative contexts, participants, not being ostensively engaged by the robot demonstrator, tended to share more of the target object with the different requesting robot, which importantly always began the interaction by communicatively addressing the participants. It is, therefore, possible to hypothesize that when a robot does not communicate ostensively, people may not attribute a communicative intention to the robot and, consequently, do not consider it a communicative partner (Arita et al., 2005). The human-robot interaction apparently relies on a fundamental feature of human communication, namely the attribution of a communicative intention (Grice, 1991), that is generally afforded by ostensive cues (e.g., direct eye contact, direct speech, calling one's own name, or contingent response) (Russell, 1940; Vorms, 2012). When robots communicate ostensively, the addressee attributes a communicative intention to the robot (Itakura et al., 2008): verbalizations and gaze behavior facilitate the interpretation of the robot's actions as communicative acts specifically directed at the addressee, leading the addressee to turn the robot into a communicative partner. By attributing communicative intentions, the addressee may consider the information and the communicator's beliefs, views, and attitudes toward an object, even if conveyed by a robot. Our results are consistent with previous studies with children (Arita et al., 2005; Itakura et al., 2008; Okumura et al., 2013), in which the role of robotic ostensive cues has been found to be important, e.g., a robot that displays ostensive signals can facilitate the acquisition of information and learning (Okumura et al., 2013), and intention attribution (e.g., Itakura et al., 2008). Furthermore, according to the Natural Pedagogy theory, communicative signals play a primary role in conveying relevant information because the addressee recognizes the agent's actions as communicative acts, understands the intention to communicate, and feels involved as the recipient of the communication (Csibra & Gergely, 2009, 2011).

Attachment and mentalization ability in the Shared Knowledge task

Participants' attachment styles and specific reflective functioning processes were evaluated to explain the participants' behavior and choices in the Shared Knowledge task. Regression analyses showed that participants' attachment style and reflective functioning predicted their responses in the Shared Knowledge task. Ostensive cues, such as eye contact, accurate turn-taking, and appropriate contingent responsiveness (in time, tone, and content), used by the responsive caregiver to communicate consistent and clear emotional responses, increase the likelihood of a secure child-parent attachment. At least in infancy, ostensive cues can be viewed from a developmental perspective because they trigger a basic epistemic trust in the caregiver as a benevolent, cooperative, and reliable source of cultural information that facilitates the rapid learning of shared knowledge without the need to critically scrutinize its validity or relevance (Corriveau et al., 2009; Fonagy et al., 2007). Conversely, insecure attachment creates epistemic uncertainty (Fonagy & Allison, 2014) and the child constantly tests the trustworthiness of the information delivered by the caregiver. In this sense, attachment bonds serve as a guarantee of the authenticity of knowledge. Our data showed that insecure attachment – resulting in the need for approval and the attitudes of anxiety and dependence on relationships (Fossati et al., 2003) – predicted a less eagerness to share with the robot in the conditions in which the human was the demonstrator. This was independent of context (communicative or non-communicative). Also, those more in need of approval were less likely to share with the human in the conditions in which the robot was the demonstrator, this time only in the non-communicative context. The data first inform us that, generally, when the demonstrator and requester are of different entities, a greater need for approval results in less probability to share. A fine-grained analysis of the data further suggests that this is especially true when the relationship is first established with a human (demonstrator), and one must subsequently share with a robot. This dynamic (i.e., sharing less with the robot when the human was the demonstrator) may tentatively suggest a human-centric approach to relationships. The latter observation is particularly relevant as Csibra and Gergely emphasize that Shared Knowledge should be protected from deliberate distortion by individuals who do not share the same "genetic material". Indeed, when first engaged by a human, those most in need of approval could be regarded as either skeptical of the relationship with the robot or not want to “betray” the newly constructed relationship with the human demonstrator by tending to share with the other genus the human’s least favorite object. This would also be the reason that those who are most in need of approval and who were initially approached by a robot were less likely to share the target object with the human, especially in conditions where the robot did not engage them via an ostensive cue.

We also found a negative predictive power of the reflective functioning subscale certainty of one’s own and others’ mental states and sharing the target object with a human when the robot demonstrator did not engage participants via an ostensive cue. That is, the greater the participants’ certainty with respect to their own and others’ mental contents, the lower the likelihood of sharing with the other genus in a non-communicative context. This could possibly mean that participants cannot generalize the information delivered by an informant whose mind is, by its nature, opaque. These would, overall, be consistent with the idea proposed above that in the robot non-communicative context condition, in which participants were less likely to share the target object with a human than with a different robot. The regression data enrich this observation by suggesting that this phenomenon is mediated by confidence in one’s own mental abilities, which may be better applied when the mind to be read is human rather than robotic, whose content is unknown. Put another way, good mentalistic skills need to be nurtured by an understanding of the other's mind for "safe" sharing to occur, even more so if the genus acting as the mediator of the relationship (robot) has an opaque mental content and is not conducive to relational engagement with communicative cues.

Concluding remarks and limitations

In line with the theory of Natural Pedagogy, ostensive cues play a primary role both in human-human interaction and in human-robot interaction and make it possible to efficiently convey information because the addressee assigns to the human or robot demonstrator’s actions a communicative intent. Ostensive cues seem to generate the genericity bias in adults as well, namely, the information conveyed in a communicative context is interpreted as generic and extendable to other individuals; this is not the case in non-communicative contexts where information is considered episodic and personal dispositions. In sum, we have demonstrated that, just like in infants, ostensive cues modulate the attention and information encoding as an object- or a person-centered in adults as well; potentially configuring the Shared Knowledge assumption as an inherent part of human communication rather than specific to certain age groups.

Our results further suggest that also non-human ostensive cues elicited a similar attribution of a communicative intention to the addresser. These findings provide new evidence that robotic ostensive cues play a distinct role in human-robot interaction, allowing the robot to become an effective communicative partner. The crucial point is that the robot must first be considered a social agent with a relational intention. If the communication is not introduced by ostensive cues, the addressee does not consider the robot a communicative partner and does not pay attention to the information the robot wants to convey; hence, the genericity bias is not applied any longer. Moreover, the knowledge is shared by the members of the same social group and, as the AMS-Q data shows, robots and humans belong to two different genera. These differences outline an "inter-agent" discriminative attitude toward the robot. This is also evident from the fact that participants tend to attribute more inter-individual differences between people than between robots, which are perceived to be fundamentally the same precisely because what one robot likes is generalizable to any other robot (but not to humans).

Although our study contributes to human-robot interaction research and also provides input for practical use, some limitations need to be considered. Firstly, we did not investigate participants' familiarity with robots. Secondly, the Shared Knowledge tasks were administered online. We were fully aware that if we had shown a real robot, the physical embodiment might have produced an enhanced effect for participants and eased the influence of the robot's ostensive cues on the affective evaluation of the object and on sharing it. It is important to note that physical and social embodiment are inherently interconnected. From a fundamental perspective, physical embodiment refers to the space occupied by the robot and its ability to move and perceive its surrounding environment. When a second agent is introduced, social interaction also comes into play, even if there is no direct communication between the two parties. It might be worth exploring this in a more ecological context. That said, the fact that our results replicate the findings of the original study allows us to predict with a good degree of certainty the persistence of Shared Knowledge in adulthood and the effectiveness of the robot's ostensive cue if the paradigm of the present study were administered in the presence. Future studies are needed to examine the extent of ostensive cues in human-robot interaction and also to clarify whether the effects found will persist even when people are already familiar with a robot and perceive it within the relationship, such as a household robot. Such research may guide future directions for humanoid robot design in the field of social robotics and lead to new learning strategies.

Data availability

The dataset generated during and/or analyzed during the current study is available from the corresponding author upon reasonable request.

Ethics statement

All procedures performed in this study involving human participants were in accordance with the 1964 Declaration of Helsinki and its subsequent revisions or comparable ethical standards. The study was approved by the Ethics Committee of the Università Cattolica del Sacro Cuore of Milan, Department of Psychology (date: May 27, 2021; No. 49-21). Informed consent was obtained from all individual participants included in the study.

Author Contributions

All authors contributed to the study's conception and design. Material preparation, data collection, and analysis were performed by CDD, LM, and FM. The first draft of the manuscript was written by CDD and LM, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

The authors have no relevant financial or non-financial interests to disclose.

Acknowledgments

We thank Ermanno Bonazzi and Nicolò Di Dio for helping prepare and produce the experimental videos.

References

Arita, A., Hiraki, K., Kanda, T., & Ishiguro, H. (2005). Can we talk to robots? Ten-month-old infants expected interactive humanoid robots to be talked to by persons. Cognition, 95(3), B49–B57. https://doi.org/10.1016/j.cognition.2004.08.001

Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. Developmental Psychology, 29(5), 832–843. https://doi.org/10.1037/0012-1649.29.5.832

Baldwin, D. A., & Moses, L. J. (1996). The Ontogeny of Social Information Gathering. Child Development, 67(5), 1915. https://doi.org/10.2307/1131601

Baxter, P., Ashurst, E., Read, R., Kennedy, J., & Belpaeme, T. (2017). Robot education peers in a situated primary school study: Personalisation promotes child learning. PLOS ONE, 12(5), e0178126. https://doi.org/10.1371/journal.pone.0178126

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

Breazeal, C., Dautenhahn, K., & Kanda, T. (2016). Social Robotics. In B. Siciliano & O. Khatib (Eds.), Springer Handbook of Robotics (pp. 1935–1972). Springer International Publishing. https://doi.org/10.1007/978-3-319-32552-1_72

Cangelosi, A., & Schlesinger, M. (2018). From Babies to Robots: The Contribution of Developmental Robotics to Developmental Psychology. Child Development Perspectives, 12(3), 183–188. https://doi.org/10.1111/cdep.12282

Chevalier, P., Kompatsiari, K., Ciardo, F., & Wykowska, A. (2020). Examining joint attention with the use of humanoid robots-A new approach to study fundamental mechanisms of social cognition. Psychonomic Bulletin & Review, 27(2), 217–236. https://doi.org/10.3758/s13423-019-01689-4

Corriveau, K. H., Harris, P. L., Meins, E., Fernyhough, C., Arnott, B., Elliott, L., Liddle, B., Hearn, A., Vittorini, L., & de Rosnay, M. (2009). Young children's trust in their mother's claims: Longitudinal links with attachment security in infancy. Child Development, 80(3), 750–761. https://doi.org/10.1111/j.1467-8624.2009.01295.x

Csibra, G. (2010). Recognizing Communicative Intentions in Infancy. Mind & Language, 25(2), 141–168. https://doi.org/10.1111/j.1468-0017.2009.01384.x

Csibra, G., & Gergely, G. (2009). Natural pedagogy. Trends in Cognitive Sciences, 13(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. Philosophical Transactions of the Royal Society B: Biological Sciences, 366(1567), 1149–1157. https://doi.org/10.1098/rstb.2010.0319

Csibra, G., & György, G. (2006). Social learning and social cognition: The case for pedagogy. Attention and Performance, 21, 249–274.

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), 215824401880987. https://doi.org/10.1177/2158244018809874

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

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P., Massaro, D., & Marchetti, A. (2020b). Come I bambini pensano alla mente di un robot. Il ruolo dell'attaccamento e della Teoria della Mente nell'attribuzione di stati mentali a un agente robotico [How children think about the robot's mind. The role of attachment and Theory of Mind in the attribution of mental states to a robotic agent]. Sistemi Intelligenti, 1, 41–56.

Egyed, K., Király, I., & Gergely, G. (2013). Communicating Shared Knowledge in Infancy. Psychological Science, 24(7), 1348–1353. https://doi.org/10.1177/0956797612471952

Feeney, J. A., Noller, P., & Hanrahan, M. (1994). Assessing adult attachment. In Attachment in adults: Clinical and developmental perspectives (pp. 128–152). Guilford Press.

Fisher, S., Guralnik, T., Fonagy, P., & Zilcha-Mano, S. (2021). Let's face it: Video conferencing psychotherapy requires the extensive use of ostensive cues. Counselling

Psychology Quarterly, 34(3–4), 508–524. https://doi.org/10.1080/09515070.2020.1777535

Flom, R., Lee, K., & Muir, D. (2007). Gaze Following: Its Development and Significance.

Fonagy, P., & Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. Psychotherapy, 51(3), 372–380. https://doi.org/10.1037/a0036505

Fonagy, P., Gergely, G., & Target, M. (2007). The parent?infant dyad and the construction of the subjective self. Journal of Child Psychology and Psychiatry, 48(3–4), 288–328. https://doi.org/10.1111/j.1469-7610.2007.01727.x

Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y.-W., Warren, F., Howard, S., Ghinai, R., Fearon, P., & Lowyck, B. (2016). Development and Validation of a Self-Report Measure of Mentalizing: The Reflective Functioning Questionnaire. PLOS ONE, 11(7), e0158678. https://doi.org/10.1371/journal.pone.0158678

Fossati, A., Feeney, J. A., Donati, D., Donini, M., Novella, L., Bagnato, M., Acquarini, E., & Maffei, C. (2003). On the Dimensionality of the Attachment Style Questionnaire in Italian Clinical and Nonclinical Participants. Journal of Social and Personal Relationships, 20(1), 55–79. https://doi.org/10.1177/02654075030201003

Gergely, G. (2007). Learning'about'versus Learning'from'other Minds: Human Pedagogy and its Implications. In P. Carruthers (Ed.), The Innate Mind: Foundations and the Future. Oxford University Press, Usa.

Grice, P. (1991). Studies in the Way of Words: Harvard University Press.

Isernia, S., Baglio, F., d'Arma, A., Groppo, E., Marchetti, A., & Massaro, D. (2019). Social Mind and Long-Lasting Disease: Focus on Affective and Cognitive Theory of Mind in Multiple Sclerosis. Frontiers in Psychology, 10, 218. https://doi.org/10.3389/fpsyg.2019.00218

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(5), 519–532. https://doi.org/10.1080/15250000802329503

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 (Special Issues on Human-Robot Interaction), 19, 61–84. https://doi.org/10.1207/s15327051hci1901&2_4

Kont, M., & Alimardani, M. (2020, September 6). Engagement and Mind Perception within Human-Robot Interaction: A Comparison between Elderly and Young Adults.

Luyten, P., Campbell, C., Allison, E., & Fonagy, P. (2020). The Mentalizing Approach to Psychopathology: State of the Art and Future Directions. Annual Review of Clinical Psychology, 16(1), 297–325. https://doi.org/10.1146/annurev-clinpsy-071919-015355

Manzi, F., Di Dio, C., Di Lernia, D., Rossignoli, D., Maggioni, M. A., Massaro, D., Marchetti, A., & Riva, G. (2021). Can You Activate Me? From Robots to Human Brain. Frontiers in Robotics and AI, 8, 633514. https://doi.org/10.3389/frobt.2021.633514

Manzi, F., Ishikawa, M., Di Dio, C., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2020). 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. (2020). 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

Manzi, F., Sorgente, A., Massaro, D., Villani, D., Di Lernia, D., Malighetti, C., Gaggioli, A., Rossignoli, D., Sandini, G., Sciutti, A., Rea, F., Maggioni, M. A., Marchetti, A., & Riva, G. (2021). Emerging Adults' Expectations About the Next Generation of Robots: Exploring Robotic Needs Through a Latent Profile Analysis. Cyberpsychology, Behavior, and Social Networking, 24(5), 315–323. https://doi.org/10.1089/cyber.2020.0161

Marchetti, A., Manzi, F., Itakura, S., & Massaro, D. (2018). Theory of Mind and Humanoid Robots From a Lifespan Perspective. Zeitschrift Für Psychologie, 226(2), 98–109. https://doi.org/10.1027/2151-2604/a000326

Marno, H., Davelaar, E. J., & Csibra, G. (2014). Nonverbal communicative signals modulate attention to object properties. Journal of Experimental Psychology: Human Perception and Performance, 40(2), 752–762. https://doi.org/10.1037/a0035113

Miraglia, L., Peretti, G., Manzi, F., Di Dio, C., Massaro, D., & Marchetti, A. (2023). Development and validation of the Attribution of Mental States Questionnaire (AMS-Q):

A reference tool for assessing anthropomorphism. Frontiers in Psychology, 14, 999921. https://doi.org/10.3389/fpsyg.2023.999921

Morandotti, N., Brondino, N., Merelli, A., Boldrini, A., De Vidovich, G. Z., Ricciardo, S., Abbiati, V., Ambrosi, P., Caverzasi, E., Fonagy, P., & Luyten, P. (2018). The Italian version of the Reflective Functioning Questionnaire: Validity data for adults and its association with severity of borderline personality disorder. PLOS ONE, 13(11), e0206433. https://doi.org/10.1371/journal.pone.0206433

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for Referential Understanding in the Emotions Domain at Twelve and Eighteen Months. Child Development, 72(3), 718–735. https://doi.org/10.1111/1467-8624.00311

Mumme, D. L., & Fernald, A. (2003). The Infant as Onlooker: Learning From Emotional Reactions Observed in a Television Scenario. Child Development, 74(1), 221–237. https://doi.org/10.1111/1467-8624.00532

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., & Itakura, S. (2013). Can infants use robot gaze for object learning?: The effect of verbalization. Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems, 14(3), 351–365. https://doi.org/10.1075/is.14.3.03oku

Okumura, Y., Kanakogi, Y., Kobayashi, T., & Itakura, S. (2020). Ostension affects infant learning more than attention. Cognition, 195, 104082. https://doi.org/10.1016/j.cognition.2019.104082

Parise, E., Cleveland, A., Costabile, A., & Striano, T. (2007). Influence of vocal cues on learning about objects in joint attention contexts. Infant Behavior and Development, 30(2), 380–384. https://doi.org/10.1016/j.infbeh.2006.10.006

Russell, B. (1940). An Inquiry Into Meaning and Truth. Routledge.

Schröder-Pfeifer, P., Talia, A., Volkert, J., & Taubner, S. (2018). Developing an assessment of epistemic trust: A research protocol. Research in Psychotherapy: Psychopathology, Process and Outcome, 21(3). https://doi.org/10.4081/ripppo.2018.330

Senju, A., & Csibra, G. (2008). Gaze Following in Human Infants Depends on Communicative Signals. Current Biology, 18(9), 668–671. https://doi.org/10.1016/j.cub.2008.03.059

Tomasello, M. (1999). The Human Adaptation for Culture. Annual Review of Anthropology, 28(1), 509–529. https://doi.org/10.1146/annurev.anthro.28.1.509

Träuble, B., & Bätz, J. (2014). Shared function knowledge: Infants' attention to function information in communicative contexts. Journal of Experimental Child Psychology, 124, 67–77. https://doi.org/10.1016/j.jecp.2014.01.019

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: Biological Sciences, 374(1771), 20180032. https://doi.org/10.1098/rstb.2018.0032

Vorms, M. (2012). A-not-B Errors: Testing the Limits of Natural Pedagogy Theory. Review of Philosophy and Psychology, 3(4), 525–545. https://doi.org/10.1007/s13164-012-0113-4

Yoon, J. M. D., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. Proceedings of the National Academy of Sciences, 105(36), 13690–13695. https://doi.org/10.1073/pnas.0804388105

Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and Challenges in Human–Robot Interaction. International Journal of Social Robotics, 7(3), 347–360. https://doi.org/10.1007/s12369-014-0267-6

CHAPTER 4

DEVELOPMENT AND VALIDATION OF THE ATTRIBUTION OF MENTAL STATES QUESTIONNAIRE (AMS-Q): A REFERENCE TOOL FOR ASSESSING ANTHROPOMORPHISM.

Laura Miraglia[1*†], Giulia Peretti[1*†], Federico Manzi[†12], Cinzia Di Dio[12], Davide Massaro[12], Antonella Marchetti[12]

[†]These authors have contributed equally to this work and share the first authorship.

[1]Research Unit on Theory of Mind, Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy; [2]Research Unit in Psychology and Robotics in the Lifespan (PsyRoLife), Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy.

Published on Frontiers in Psychology, April 2023.

Keywords

Attribution of Mental States Questionnaire (AMS-Q) | Theory of Mind (ToM) | Mentalization | Mental states | Factor analysis | Validation | Adults.

Abstract

Attributing mental states to others, such as feelings, beliefs, goals, desires, and attitudes, is an important interpersonal ability, necessary for adaptive relationships, which underlies the ability to mentalize. To evaluate the attribution of mental and sensory states, a new 23-item measure, the Attribution of Mental States Questionnaire (AMS-Q), has been developed. The present study aimed to investigate the dimensionality of the AMS-Q and its psychometric proprieties in two studies. Study 1 focused on the development of the questionnaire and its factorial structure in a sample of Italian adults (*N* = 378). Study 2 aimed to confirm the findings in a new sample (*N* = 271). Besides the AMS-Q, Study 2 included assessments of Theory of Mind (ToM), mentalization, and alexithymia. A Principal Components Analysis (PCA) and a Parallel Analysis (PA) of the data from Study 1 yielded three factors assessing mental states with positive or neutral valence (AMS-NP), mental states with negative valence (AMS-N), and sensory states (AMS-S). These showed satisfactory reliability indexes. AMS-Q's whole-scale internal consistency was excellent. Multigroup Confirmatory Factor Analysis (CFA) further confirmed the three-factor structure. The AMS-Q subscales also showed a consistent pattern of correlation with associated constructs in the theoretically predicted ways, relating positively to ToM and mentalization and negatively to alexithymia. Thus, the questionnaire is considered suitable to be easily administered and sensitive for assessing the attribution of mental and sensory states to humans. The AMS-Q can also be administered with stimuli of nonhuman agents (e.g., animals, inanimate things, and even God); this allows the level of mental anthropomorphization of other agents to be assessed using the human as a term of comparison, providing important hints in the perception of nonhuman entities as more or less mentalistic compared to human beings, and identifying what factors are required for the attribution of human mental traits to nonhuman agents, further helping to delineate the perception of others' minds.

Introduction

The ability to mentalize (Fonagy, 1989; 1991; Fonagy & Bateman, 2008; Fonagy & Luyten, 2009), also called Theory of Mind (ToM; Perner & Wimmer, 1985; Premack & Woodruff, 1978; Wellman, 2020; Wellman et al, 2001; Wimmer & Perner, 1983), is a human-specific ability that allows attributing mental states – intentions, thoughts, desires, and emotions – to themselves and others to explain and predict behavior (Astington & Baird, 2005; Frith & Frith, 1999; 2006; Gopnik & Wellman, 1992; Tomasello, 1999; Tomasello et al., 2005; Wellman, 1992). Mind reading abilities are a crucial function of social cognition that enables engagement in human interactions and promotes adaptation in everyday social contexts (Mull & Evans, 2010). In daily life, the ability to mentalize allows people to function socially by distinguishing between accidental and intentional behavior, desires and reality, truth and deception (Bellagamba et al., 2012), and to reach goals, including understanding, predicting, or controlling another's behavior, as well as being able to understand the perspective of others, feel sympathy or compassion, and provide help (Batson et al., 1997; Davis et al., 1996; Galinsky et al., 2005; Goldstein et al, 2014; Waytz et al., 2010). So, nearly all children and adults consistently use their mind-reading skills for everyday social purposes. In this sense, the nature of social behaviors is rarely neutral and more often are behaviors that require prosocial or antisocial use of ToM skills (Arefi, 2010; Ronald et al., 2005). For these reasons, Ronald et al. (2005) proposed the expressions *nice Theory of Mind* and *nasty Theory of Mind* to distinguish prosocial and antisocial ToM abilities (Happé & Frith, 1996), identifying nice ToM in behaviors such as cooperating, comforting, considering others' feelings, and nasty ToM, which involves an intact mentalizing ability but used to manipulate, outwit, or tease others (Happé & Frith, 1996).

Mentalization skills, necessary for children's social functioning (Astington, 2003) and emotion regulation (Greenberg et al., 2017), develop from early dyadic relationships with mothers (Fonagy et al., 1991; 2007; Meins et al., 2002; 2012; Nelson, 2005; Slaughter et al., 2009), within which infants experience mental states through maternal language that contains references to the mental sphere (Beeghly et al., 1986; Symons et al., 2006; Slaughter et al., 2009; Giovannelli et al., 2020). However, words for mental states are not immediately understood by infants because of their abstract and invisible form (Slaughter et al., 2009). The development of mental states vocabulary begins at approximately 2 years of age within conversations with mothers who explicitly label children's mental states for them (Bartsch & Wellman, 1995; Bretherton & Beeghly, 1982; Slaughter et al., 2009). In this regard, a large body of research reveals that mothers' tendency to talk about emotions, desires, and beliefs and to make verbal references to their children's mental experiences provides relevant input into children's emerging mentalistic vocabulary (Meins et al., 2002; 2012; Nelson, 2005; Symons et al., 2006; Slaughter et al., 2009; Taumoepeau & Ruffman 2006; 2008). Later, mothers' tendency to make verbal references fades, and by age 4/6, children develop an awareness that others may have mental states different from their own (Wimmer & Perner, 1983). The development of a mentalistic vocabulary allows children to reflect on and understand their own and others' mental states, assuming that the other is structurally endowed with a mind capable of possessing internal mental states. Thus, from childhood, the attribution of mental states to others becomes an ongoing process that occurs constantly and continuously throughout the lifespan to understand, explain, and reduce uncertainty about people's behaviors. Importantly, we also make inferences about nonhuman agents' internal states to approach and interact with them (Di Dio et al., 2018; Martini et al., 2016; Waytz et al., 2010), i.e., non-anthropomorphic living entities (e.g., animals), anthropomorphic non-living entities (e.g., robots), non-living non-anthropomorphic entities (e.g., objects), and even God (Abell et al., 2000; Di Dio et al., 2018; Gervais, 2013; Giménez-Dasí et al., 2005; Harris & Koenig, 2006; Heider & Simmel, 1944; Manzi et al., 2020b; 2021b,c; Marchetti et al., 2018; Ramsey & Hamilton, 2010; Wigger et al., 2013; Wellman, 2017). As noted by Waytz and colleagues (2010) perceived similarity between self and another individual increases as one considers their mental state. At the same time, the characteristics of an agent, animate or inanimate, influence the perception of its mind. For instance, dogs are ascribed special mental properties due to some of their species-specific sensory characteristics – e.g., the sense of smell that allows the perception of an object closed in a sealed box – that is much more developed than in humans (Di Dio et al., 2018). In addition, regardless of religious background, preschoolers attribute qualities such as omniscience to the mind of God, thus perceiving God's mind at a higher epistemic level than humans' minds (Nyhof & Johnson, 2017; Di Dio et al., 2018). Several studies have focused also on the attribution of minds to robotic agents (for a review, see Thellman et al., 2022) and observed that adults are more inclined to ascribe greater mental states to robots characterized by human-like physical features (Airenti, 2015; Bartneck et al., 2009; Dario et al., 2001; Fink, 2012; Gray & Wegner, 2012; Kiesler et al., 2008; MacDorman et al., 2005; Manzi et al. 2020b; 2021a,b,c; Krach et al., 2008; Thellman et al., 2017; Wiese & Wiese, 2020; Złotowski et al., 2015). This tendency has also been found in children over the age of five, who are likely to attribute more mental states to robots with more human-like features; in contrast, younger children tend to anthropomorphize by giving less importance to the human aspect of the robotic agent (Di Dio et al., 2020 a,b; Manzi et al., 2020b). Attributing mental states and consequently perceiving an agent as more or less mentalistic has important implications on how one will interact with it because mind perception implies moral status (Gray, Gray, Wegner, 2007; Waytz, Cacioppo, Epley, 2010; Waytz et al., 2010). In fact, ascribing mind has consequences for both the perceiver and the perceived (Waytz et al., 2010), to the point of making it relevant to evaluate the perception of the minds of different entities as compared to humans.

The present study aimed to validate a new and agile measure, the Attribution of Mental State Questionnaire (AMS-Q) – already widely used in studies with children (Di Dio et al., 2018, 2020a,b; Manzi et al., 2020b; Peretti et al., 2023) and adults (Manzi et al., 2021c) – which assesses the attribution of mental and sensory states primarily to human. However, to the authors' knowledge, there is no currently validated measure to compare the mental traits of human and nonhuman agents to evaluate the level of mental anthropomorphization of nonhuman agents, including living and nonliving entities. The AMS-Q aims to fill this void, as its originality lies in comparing the attribution of mental states between human and nonhuman agents by also administering pictures of nonhuman agents as stimuli. In this sense, the human picture is used as a baseline to assess, through comparison, the level of mental anthropomorphization of nonhuman agents (e.g., animals, inanimate things, and even God). The general purpose of the current study was to validate the AMS-Q on human stimuli across two Italian samples and then to show its sensitivity in capturing differences in the attribution of mental and sensory traits between human and nonhuman agents through an example of the applicability of the questionnaire in which an image of a dog and a robot were administered as nonhuman agent stimuli in addition to the human baseline. This research consisted of two main studies preceded by a preliminary study aimed to develop and generate an initial item pool based on the previous version of the AMS-Q (Manzi et al., 2017, 2020; Di Dio et al., 2018) and on a wide corpus of literature. Study 1 investigated the structure of the questionnaire, whereas Study 2 focused on confirming the factor structure and aimed to investigate the construct validity of the questionnaire by investigating its convergent and divergent validity. The rationale, design, and hypotheses of each study are outlined in more detail in the following sessions.

Study hypotheses

Congruent with theoretical formulations postulating that people intuitively think about others' minds in distinct dimensional representations (Gray et al., 2007; Malle, 2019), we expected at least a two-factor model of the AMS-Q, with scales that distinctly assessed mental states attribution and sensory states attribution. We investigated the reliability and validity of the AMS-Q in two samples of Italian adults. Exploratory Factor Analysis and (multi-group) Confirmatory Factor Analysis (CFA) were used to investigate the factor structure of the questionnaire. Two different groups were recruited for the exploratory (*N* = 378) and confirmatory (*N* = 271) analysis.

The convergent validity of the AMS-Q was investigated by administering the Reading the Mind in the Eyes test (ET; Baron-Cohen et al., 2001; Italian version: Vellante et al., 2013), an advanced Theory of Mind test to evaluate the correspondence between the semantic definition of mental state and the image of the eye-region displayed on the screen. Differing from other measures that assess the individual's mental abilities, the Eye Test explicitly evaluates the ability to *attribute* mental states to others. However, the ET assesses mental states predominantly related to the emotional sphere. To overcome this limitation, we also included a second measure: the Multidimensional Mentalizing Questionnaire (MMQ; Gori et al., 2021), which aims to assess core aspects of mentalization, including the cognitive sphere. The MMQ, in fact, is a self-report measure, which assesses mentalization on four central axes (cognitive-affective, self-other, outside-inside, and explicit-implicit). We expected the AMS-Q subscales to be significantly positively correlated with the ET and the MMQ. We also correlated the AMS-Q and the Toronto Alexithymia Scale (TAS-20) (Italian version: Bressi et al., 1996), to test for divergent validity. TAS-20 is a self-report scale designed to evaluate the level of alexithymia, i.e., the inability to describe and/or distinguish one's own emotions (Westwood et al., 2017) (for a detailed description of scales, see Methods: Measures section). Negative correlations between AMS-Q and alexithymia were expected, as this dimension indicates poor awareness of emotions and feelings and mind-blindness.

Finally, the discriminant validity of AMS was investigated by testing its ability to differentiate between the attribution of mental and sensory states toward different entities. For this purpose, in addition to images of humans, we administrated two other stimuli: a picture of a robot (non-living entity) and a dog (living non-human entity). We assumed that the AMS-Q would be able to capture differences in terms of attributions of mental and sensory states between the human agent and the other two entities, allowing us to assess the level of mental anthropomorphism attributed to the agents examined.

Scale development: Item generation

Several sources were used in generating the initial item pool: the psychological lexicon of Lecce and Pagnin (2007); Slaughter and colleagues' (2009) theoretical model of mental verb categorization resulting from communicative exchanges between mother and child; and Martini and colleagues' work (2016). The initial item pool also included the mentalistic verbs of the earlier version of the AMS scale, which has been widely used in research with children (Di Dio et al., 2018; 2019; see also, Di Dio et al., 2020a; 2020b; Manzi et al., 2017, 2020) and adults (Manzi et al., 2021c).

Mentalistic vocabulary has been selected to encompass different categories of mental states: a) volition (i.e., nouns, verbs, adjectives, or adverbs referring to states of desire or intention); b) cognition (i.e., nouns, verbs, adjectives, or adverbs referring to mental acts of thought, intellect, or reasoning); and c) disposition (nouns, verbs, adjectives, or adverbs referring to states of preference or affect) (Slaughter at al., 2009). We also included a category referring to sensory states (e.g., smell, listen, look, taste, etc.) in the initial item pool. The resulting 69 mentalistic verbs and expressions were administered to fifty (50) Italian speakers, aged 18+ years (48.7% females; *Mean age* = 35.36; *SD* = 13.89). Participants were recruited through a mailing list built by the research team over time. Included in the email to the participants was an invitation letter and a link to access the online task on the Qualtrics platform. All participants participated on a voluntary and anonymous basis. They received no compensation for participating in the study.

To produce a valid factorial analysis, we asked participants to choose five words for each mental verbs category, after having looked at pictures depicting specific human characters (i.e., "Select five words/expressions that, according to you, are most representative to describe the image"). Each participant looked at five stylized, black-and-white images of human beings administered in random order: a woman, a man, a girl, a boy, and an infant.

Descriptive frequency analysis revealed the mentalistic expressions or verbs most selected by participants. This word evaluation method was chosen to provide a holistic approach to assessing the attribution of mental states in order to refine the items and to provide a questionnaire that can be representative of the concept of mental and sensory states related to human beings. Subsequently, the 26-item questionnaire was administered to a convenient sample of 22 (14 women and 8 men) Italian adults to investigate comprehensibility. This sample provided feedback on the clarity of item content and instructions, as well as on the images used. Items that were deemed odd or ambiguous were considered for rephrasing or exclusion. We decided to leave out two items that were defined as highly vague. The questionnaire was finally reduced to 24 items. In addition, overly detailed images of humans were discarded in favor of two black silhouettes because, especially the facial features drawn, seemed to suggest a state of mind that might could influence the attribution of emotional states.

Method and Materials

Participants

The construction sample (Study 1) included 378 Italian adults (54.2% female; *Mean age* = 30.6; *SD* = 12.23; age-range = 18-65 years). Sociodemographic characteristics of the construction sample are reported in Table 1. The validation sample (Study 2) included 271 Italian adults (55.4% female; *Mean age* = 26.1, *SD* = 8.09; age-range = 19-60 years). Sociodemographic characteristics of the validation sample are reported in Table 1.

All the participants were recruited on Prolific platform and rewarded with 6.35£ per hour. Written informed consent was obtained from all participants after a full explanation of the study procedure, in line with the Declaration of Helsinki and its revisions. The local Ethics Committee of the Department of Psychology, Università Cattolica del Sacro Cuore, Milan, approved the experimental protocol of all studies involved in the current research.

Table 1 – Sociodemographic characteristics of the construction and validation samples.

Sociodemographic characteristics	**Construction Sample**	**Validation Sample**
	N = 378	*N* = 271
Age, mean ± SD	30.6 ± 12.23	26.1 ± 8.09
Gender	N (%)	N (%)
Male	173 (45.8%)	121 (44.6%)
Female	205 (54.2%)	150 (55.4%)
Residence	N (%)	N (%)
North Italy	236 (62.6%)	222 (81.9%)
Centre Italy	64 (17%)	17 (6.3%)
South Italy	52 (13.8%)	20 (7.4%)
Sicily and Sardinia	25 (6.6%)	8 (3.0%)
Outside Italy	-	4 (1.5%)
Educational level	N (%)	N (%)
Middle school or below	9 (2.4%)	2 (0.8%)
High school	171 (45.2%)	169 (62.4%)
Graduate school	176 (46.5 %)	95 (35%)

Postgraduate school	22 (5.8%)	5 (1.8%)
Employment status	N (%)	N (%)
Student	175 (46.3%)	171 (63.1%)
Employed	140 (37%)	73 (27%)
Unemployed	24 (6.3%)	4 (1.5%)
Other	39 (10.3%)	23 (8.5%)

Procedure

Data were collected through an online survey hosted on the Qualtrics platform from November 2021 to January 2022.

With respect to Study 1, after the participants provided some sociodemographic information (age, gender, residence, occupation, and level of study), they completed the first version of the Attribution of Mental States Questionnaire in response to a male and female silhouette image evocative of human mentalistic traits (Figure 1). The items were randomized to avoid possible response bias by question order.

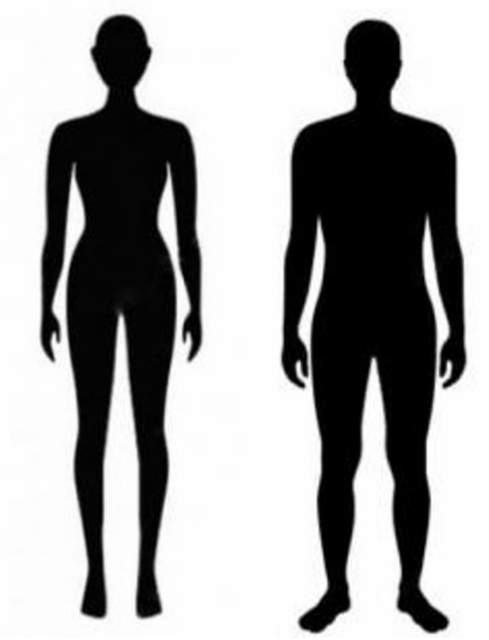


Figure 1 – Stimuli for Study 1: silhouettes of a woman and a man.

With respect to Study 2, participants completed a sociodemographic survey and the refined version of the AMS-Q in response to the male or female human silhouette. Participants completed the AMS-Q two more times with a robot and a dog picture as stimuli. The stimuli (Figure 2) were presented in random order. Finally, to test external validity, we correlated the questionnaires with validated tasks of Theory of Mind, mentalization ability, and alexithymia. All items were randomized to avoid participants' responses may be affected by question order.

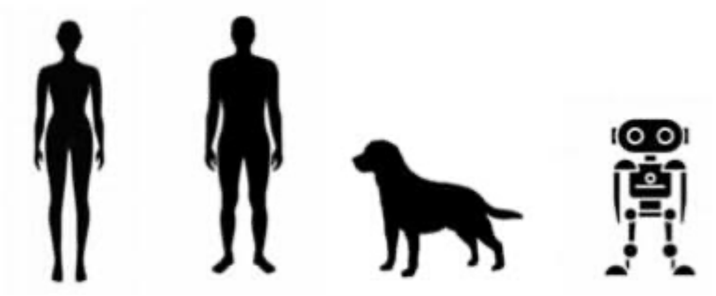


Figure 2 – Stimuli for Study 2: silhouettes of a woman, a man, a dog, and a robot.

Measures

All the participants in the construction sample were administered a sociodemographic questionnaire assessing age, sex, residence, school attendance, current job, and the pool of 24 items composing the AMS-Q developed in the previous steps. Participants were asked to rate each item according to a 5-point Likert scale ranging from 1 (*No, not at all*) to 5 (*Yes, very much*). Participants were informed that they would have had to evaluate one of the two silhouettes' images of human beings (i.e., "According to you, can a human being [*mental/sensory ability, e.g., think/taste*]?").

All the participants in the validation sample were administered the sociodemographic survey and a battery of questionnaires including the 24-item AMS-Q, the Multidimensional Mentalizing Questionnaire (MMQ), and the Italian version of the Reading the Mind in the Eyes Test (ET), and the Toronto Alexithymia Scale (TAS-20).

Reading the Mind in the Eyes Test (ET). The Reading the Mind in the Eyes Test (ET; Baron-Cohen et al., 2001; Italian version: Vellante et al., 2013) was administered to measure Theory of Mind and the attribution of mental states. Participants were randomly presented with a series of 36 photographs of the eye region of 19 actors and 17 actresses. Each photo was surrounded by four single-word mental state descriptors, e.g., bored, angry, happy. One of these descriptors targeted the mental state depicted in the photo, and the others were foils. The ET is based on a four-alternative forced-choice paradigm, with 25% correct guess rate. Participants were instructed to choose which of the four descriptors best describes what the person in the photo is thinking or feeling. The score on the test is the number of descriptors correctly identified by the participants, i.e., the number of mental states correctly identified. The maximum score is 36. In the validation sample, the internal reliability was acceptable (α = 0.52). As reported

Vellante et al. (2013), there is some agreement that Cronbach's coefficient alpha is a poor index of unidimensionality, in fact, the reliability of the Eyes test was rarely reported in past studies or obtained only acceptable values (Harkness et al., 2010; Voracek & Dressler 2006).

Multidimensional Mentalizing Questionnaire (MMQ). Multidimensional Mentalizing Questionnaire (MMQ; Gori et al., 2021) is a self-report measure that consists of 33 items, covering the different core aspects of mentalization on four different axes: (1) cognitive-affective; (2) self-other; (3) outside-inside; and (4) explicit-implicit. It permits a multidimensional assessment, with scores on the positive (reflexivity, ego-strength, and relational attunement) and negative (relational discomfort, distrust, and emotional dyscontrol) subscales, as well as an overall MMQ score, by summing all the items after having reversed those included in the negative subscales. The response format was on a five-point Likert scale from 1 (not at all) to 5 (a great deal). In the current study, internal reliability was good (α = 0.80).

Toronto Alexithymia Scale (TAS-20). Toronto Alexithymia Scale (TAS-20; Italian version: Bressi et al., 1996) is a self-report scale comprising 20 items rated on a five-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). It includes three subscales that measure three main dimensions of alexithymia: (1) difficulty in identifying feelings and distinguishing between feelings and bodily sensations in emotional activation, (2) difficulty in the verbal expression of emotions, and (3) externally oriented thinking. Taking the reversed items into account, the scores of the three scales were calculated. Internal reliability in the validation sample was good (α = 0.83).

Data Analysis

Study 1

In order to determine the dimensionality of the scale and sort out unsuitable items, we carried out an explanatory factor analysis using IMB SPSS Statistics version 27 and Jamovi statistical software version 2.5. A Principal Components Analysis (PCA) and a Parallel Analysis (PA; Horn, 1965) were carried out on the 24-item. PA is an adaptation of the Kaiser criterion eigenvalue > 1 (Kaiser, 1960), and minimizes the tendency to identify a greater number of factors due to sampling error. PA uses the 95th percentile of the distribution of eigenvalues generated from uncorrelated data and, therefore the number of factors extracted is considered to be "beyond chance".

Prior to performing PCA, the adequacy of the correlation matrix for factor analysis was assessed with Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) test. Adequacy of the correlation matrix is suggested by a significant Bartlett's test ($p < 0.05$) and a KMO index > 0.70. To examine the factor structure that underpins the AMS questionnaire, the PCA was carried out via oblique rotation (Promax) as the factors were presumably related to each other rather than independent. Delta was set to 0. Only items with a loading ≥ 0.30 (Hair et al., 1998) on a single factor were considered for further analyses. The solution revealed through PCA was further supported by the results of the PA.

Then, we investigated the internal consistency of the questionnaire and the presence of problematic items (i.e., items for which the Cronbach alpha improved). No items were removed and the version of the questionnaire with all 24 items was selected as it reported excellent reliability ($\alpha > .95$).

Study 2

The factor structure of the AMS-Q was subjected to a Confirmatory Factor Analysis (CFA) to confirm the three-factor model revealed in Study 1. To perform the analyses, Jamovi statistical software version 2.5 was used. Multi-group CFA was carried out using JASP team (2020). In order to evaluate the goodness-of-fit of the factor structure, we used the $\chi2/df$ ratio. A model in which $\chi2/df$ is ≤ 3, is considered acceptable. Furthermore, Hu and Bentler's guidelines (1999) for fit indices were used to determine whether the expected model fitted the data. The following fit indices were used: a) the Comparative Fit Index (CFI), with values ≥ 0.90 indicating a good fit (Bentler, 1990; Fan, Thompson, & Wang, 1999; Hu & Bentler, 1999); b) the Tucker Lewis Index (TLI), with values ≥ 0.90 indicating a reasonable fit of the model (Byrne, 1994); c) the Root Mean Square Error of Approximation (RMSEA), with values between 0.05 and 0.08 indicating the adequacy of the model (Browne & Cudeck, 1993), and values ≤ 0.05 indicating evidence of absolute fit (Lai & Green, 2016); and d) the Standardized Root Mean Square Residual (SRMR), with values ≤ 0.08 indicating an adequate fit (Hu & Bentler, 1999; Schermelleh-Engel et al., 2003).

Moreover, a multigroup CFA was performed to test invariance across gender of the final factor structure. We tested for configural invariance to assess whether the same number of factors is extracted across groups.

The validity of AMS-Q was assessed by correlating (Pearson *r*) the AMS-Q factors with theoretically related measures, namely the ET and the MMQ subscales to establish construct (convergent) validity. Second, we repeated the correlations between AMS-Q and TAS-20, to examine the discriminatory power of the measure and divergent validity.

Finally, to assess the discriminant validity of the AMS-Q we administered a picture of a living non-human agent (a dog) and a non-living non-human agent (a robot) in addition to the human stimuli. A repeated-measures GLM analysis comparing AMS-Q scores on human – i.e., the baseline –, dog, and robot stimuli was conducted to investigate the ability of the AMS-Q to discriminate between the attribution of mental and sensory states toward different entities. Comparison between the baseline and the two stimuli examined allows us to assess the level of mental anthropomorphism attributed to the dog and the robot. Greenhouse-Geisser correction for violations of the Mauchly sphericity test, $p < 0.05$, was used in the GLM analysis. All post hoc comparisons were Bonferroni corrected.

Results

Study 1

Exploratory Factor Analyses. A Principal Components Analysis (PCA) was carried out to explore the factors structure of the 24 items. The correlation matrix was suited for factor analysis (Bartlett's test of sphericity = 6320.2, *df* = 276, *p* = .000; KMO = 0.95). The PCA yielded three components with eigenvalues over 1, explaining 49.7%, 7.8%, and 5.9% of the variance, respectively. Altogether, the extracted factors explained 63.4% of the total variance. Since Parallel Analysis (PA) is the most accurate method for component extraction (Hubbard, 1987; Zwick, 1986), we proceeded by carrying out a PA on AMS data to confirm the structure previously found. The results of the analysis showed three components with eigenvalues exceeding the corresponding criterion values for a randomly generated data matrix of the same size (24 variables x 378 respondents). Thus, the questionnaire structure obtained from the PCA was confirmed by the results of the PA (Table 2). The inspection of the scree plot (Figure 3) also revealed that the three-factor solution was the most appropriate.

The first extracted factor had thirteen items with rotated loadings ranging from 0.32 to 0.83 (>.30; Anderson & Black, 1998), assessing the attribution of knowledge states (beliefs, thoughts, inferences) and non-epistemic mental states such as planning, feelings, and positive emotions such as joy; consequently, it was labeled "Mental states with neutral or positive valence" (AMS-NP). The second extracted factor had seven items concerning the semantic field of deception (lying, pretending, making a joke) and related emotions with negative valence such as sadness, fear, and anger, which loaded strongly (between 0.59 and 0.91) on Factor 2. This factor can be named “Mental states with negative valence” (AMS-N). Finally, Factor 3 was composed of four items clearly associated with the attribution of sensory states with strong loadings between 0.69 and 0.91, which was accordingly called “Sensory States” (AMS-S). Individual item loadings on the retained components and the Cronbach’s alphas for each factor are listed in Table 3.

Table 2 – Comparison of eigenvalues from PCA and criterion values from parallel analysis

Factor	Actual eigenvalue from PCA	Criterion value from PA	Decision
F1: AMS-NP	11.926	11.409	accept
F2: AMS-N	1.880	1.344	accept
F3: AMS-S	1.409	0.862	accept

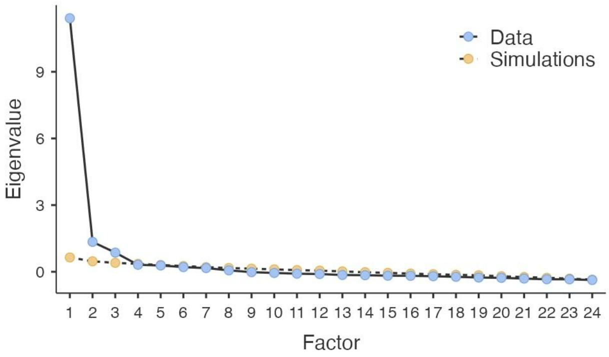


Figure 3 – Scree plot: Eigenvalues for study 1 factor analysis.

Table 3 - Study 1 pattern matrix presenting loading factors for each item, percent of explained variance, and Cronbach's alphas for each factor of the final factors.

AMS items	Factor 1	Factor 2	Factor 3
	AMS-NP	AMS-N	AMS-S
Learn	.728		
Think	.747		
Remember	.547		
Make a decision	.609		
Understand	.811		
Tell a lie		.856	
Dream	.535		
Imagine	.829		
Make a joke		.593	
Pretend		.741	
See			.828
Feel hot or cold		.743	
Taste			.638
Hear			.910
Smell			.857
Have fun	.569		
Love	.779		
Be happy	.798		
Be sad		.885	
Be scared		.910	
Get angry		.851	
Have the intention to do something	.702		
Want to do something	.668		
Make a wish	.320		
% of explained variance	49.69%	7.84%	5.87%
Cronbach's alpha	.93	.92	.88

Note: Extraction method: Principal Component Analysis. Rotation method: Promax with Kaiser Normalization.

a. Rotation converged in 6 iterations.

Reliability. The AMS-Q had excellent internal consistency, with a Cronbach alpha coefficient of 0.95. Partial alpha coefficients indicated that the three-component solution had satisfactory internal consistency (Factor 1 $\alpha = 0.93$; Factor 2 $\alpha = 0.92$; and Factor 3 $\alpha = 0.88$). There was no relevant change (neither diminishment nor improvement) in overall reliability if any of the items were deleted.

Study 2

Confirmatory Factor Analysis. Confirmatory Factor Analysis **(**CFA) was conducted on the three-factor model. First, we checked Bartlett's sphericity test to ensure inter-item correlation ($\chi 2$ 3258.86, *df* = 325, *p* = .000) and the Kaiser–Meyer–Olkin (KMO = .93) for the sample adequacy.

Although the three-factor solution fitted the data well ($\chi 2/df$ = 2.27; CFI = 0.89; TLI = 0.87; SRMR = 0.06; RMSEA = 0.07 [CI] = 0.061–0.076), coefficient R^2 was suboptimal (R^2 of 0.17) for item no. 2 (i.e., "*think*"), suggesting that the item's variance was poorly represented by the common factor. However, for the promising indices reported in Table 4 and because the item is representative of attribution that would otherwise be lost, we decided not to remove it. Nevertheless, we decided to remove item no. 12 ("*feeling hot or cold*") as it loaded moderately on two factors: respondents may possibly perceive this item as either a sensory state or a discomfort condition. Dropping out item no. 12 would then maximize the quality of responses.

Although most indices reached the recommended cut-off values (SRMR = 0.06; RMSEA = 0.07), the model could be improved, since inspection of modification indices (MI) >10 suggested that correlations between the errors of some pairs of items should be included in the model. CFA was re-run, and the goodness-of-fit indices indicated a satisfactory fit of the three-factor model. Indices with and without correlations between items are given in Table 4 (see also Figure 4).

The final version of the AMS-Q and the scoring is given in Appendix 1 and Appendix 2, respectively.

Table 4 – Goodness-of-fit indices generated by the Confirmatory Factor Analysis (CFA) with and without modification indices.

	Recommended value	Value obtained without MI	Value obtained with MI
$\chi 2/df$	≤ 3.00	2.27	1.87
CFI	≥ 0.90	.89	.93
TLI	≥ 0.90	.87	.92
SRMR	≤ 0.08	.063	.056

RMSEA	≤ 0.08	.069 ([CI] = 0.061–0.076)	.057 ([CI] = 0.048–0.065)

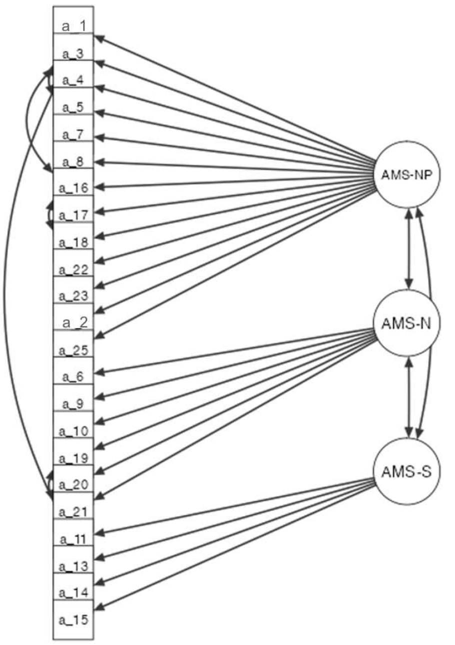


Figure 4 - Graphical summary of the CFA obtained from the 23-item of the Attribution of Mental States (AMS-Q) (N = 271).

Factor Structure Across Gender. Next, to investigate the efficacy of the model across gender, separate multi-group CFAs were carried out for women (*N* = 150) and men (*N* = 121). The CFA on the refined and fully unconstrained model indicated an adequate fit (see Table 5), suggesting factorial invariance across gender. The indexes were in line with the recommended cut-off values.

Table 5 – Goodness-of-fit indices generated by the multigroup CFA across gender.

	Recommended value	Value obtained
χ2/*df*	≤ 3.00	1.87
CFI	≥ 0.90	.87
TLI	≥ 0.90	.85
SRMR	≤ 0.08	.08
RMSEA	≤ 0.08	.08

Correlations. Table 6 lists correlations of the AMS-Q subscales with convergent and divergent measures. The validity of the AMS-Q was tested through Pearson correlations with theoretically related measures, namely the ET and the MMQ to test convergent validity, and the TAS-20 to test divergent validity. As shown, AMS-Q subscales correlated significantly and in the expected direction with the ET, MMQ, and TAS-20:

Convergent validity. All AMS-Q factors correlated significantly and in the hypothesized direction with the Eyes-test, *r* (AMS-NP)= .17, *p* <.01; *r* (AMS-N)= .18, *p* <.01; *r* (AMS-S)= .15, *p* <.05. Thus, the AMS-Q dimensions were correlated with convergent measures of ToM, configuring the AMS-Q as a questionnaire capable of assessing the attribution of mental states. Consistent with expectations, AMS-Q subscales correlated positively with measures of mentalizing abilities: Reflexivity scale of MMQ, *r* (AMS-NP)= .25, *p* <.01; *r* (AMS-N)= .32, *p* <.01; *r* (AMS-S)= .20, *p* <.01, and Relational Attunement scale of MMQ, *r* (AMS-NP)= .25, *p* <.01; *r* (AMS-N)= .18, *p* <.01; *r* (AMS-S)= .18 p <.01. AMS-NP and AMS-S correlated positively with the Ego-strength dimension of the MMQ, *r* (AMS-NP)= .16, *p* <.01; *r* (AMS-S)= .15, *p* <.05. As expected, no significant correlations were found with the other three factors of the MMQ – namely Relational Discomfort, Distrust, and Emotional Dyscontrol, *p* >.05 – as they refer to failures and distortions of mentalization abilities that are reflected in relationships and interpersonal difficulties, which are dimensions that AMS-Q does not evaluate.

Divergent validity. AMS-Q subscales were inversely correlated with the TAS, as expected. In particular, AMS-NP negatively correlated with Difficulty Identifying Feelings scale, *r* = .-14, *p* <.05, and Difficulty Describing Feelings scale of the TAS-20, *r* = .-13, *p* <.05. AMS-NP and AMS-N correlated negatively with External Oriented Thoughts, *r* = .-22, *p*<.01; *r* = .-12, *p* <.05.

Table 6 – Pearson's correlations between measures.

	AMS-NP	AMS-N	AMS-S
ET	**.171****	**.179****	**.154***
MMQ-F1	**.247****	**.323****	**.203****
MMQ-F2	**.164****	.104	**.154***
MMQ-F3	**.246****	**.177****	**.183****
TAS-DDF	**-.138***	-.042	-.099
TAS-DIF	**-.135***	.000	-.082
TAS-EOT	**-.216****	**-.120***	-.105

N = 271

*$p < .05$
**$p < .01$
Note: ET = Reading the Mind in the Eyes Test. MMQ = Multidimensional Mentalizing Questionnaire: MMQ-F1 = reflexivity; MMQ-F2 = ego-strength; MMQ-F3 = relational attunement. TAS-20 = Toronto Alexithymia Scale: TAS-DDF = difficulty identifying feelings; TAS-DIF = difficulty describing feelings; TAS-EOT = external oriented thinking.

Discriminant Validity. The GLM analysis with three levels of *AMS-Q factors* (AMS-NP, AMS-N, AMS-S) and three levels of *entity* (human, dog, robot) as within-subjects factors, was conducted to evaluate the impact of different stimuli on participants' scores on the AMS-Q. A main effect was found for the entity, *F*(1.68, 1315.85) = 1949.58*, p* < .001, partial-η2 = .89, δ = 1, indicating differences in participants' mental states attribution toward the three different entities. Specifically, post hoc comparisons (Bonferroni corrected) showed participants' tendency to ascribe greater mental states to the human than both the dog, M*diff* = .49, *SE* = .03, *p* < .001, and the robot, M*diff* = 2.22, *SE* = .04, *p* < .001. The dog also scored higher than the robot, M*diff* = 1.72, *SE* = .04, *p* < .001. The results also revealed a main effect of the interaction between entity and AMS-Q factors (Figure 5), *F*(3.27, 49.03) = 221.32*, p* < .001, partial-η2 = .45, δ = 1, indicating that humans scored higher on the attribution of knowledge states and positive emotions (AMS-NP) compared with both the dog, M*diff* = .63*, SE* = .04, *p* < .001, and to the robot, M*diff* = 2.30*, SE* = .05, *p* < .001. Respondents still attributed more negative value mental states (AMS-N) to humans than to dog, M*diff* = 1.18, *SE* = .04, *p* < .001, and robot, M*diff* = 2.67, *SE* = .05, *p* < .001. However, participants attributed greater positive (AMS-NP) and negative (AMS-N) value mental states to the dog compared to the robot, M*diff* = 1.67, *SE* = .05, *p* < .001; M*diff* = 1.50, *SE* = .05, *p* < .001. Finally, although more sensory states (AMS-S) were attributed to the human than to the robot, M*diff* = 1.69, *SE* = .06, *p* < .001; the dog was the highest scoring entity in attributing sensory states both compared to the robot, M*diff* = 2.01, *SE* = .05, *p* < .001, but also compared to the human M*diff* = .32, *SE* = .04, *p* < .001, pointing out the great sensitivity of the questionnaire to capture mental and sensory differences between different entities. Pairwise comparisons are listed in Table 7.

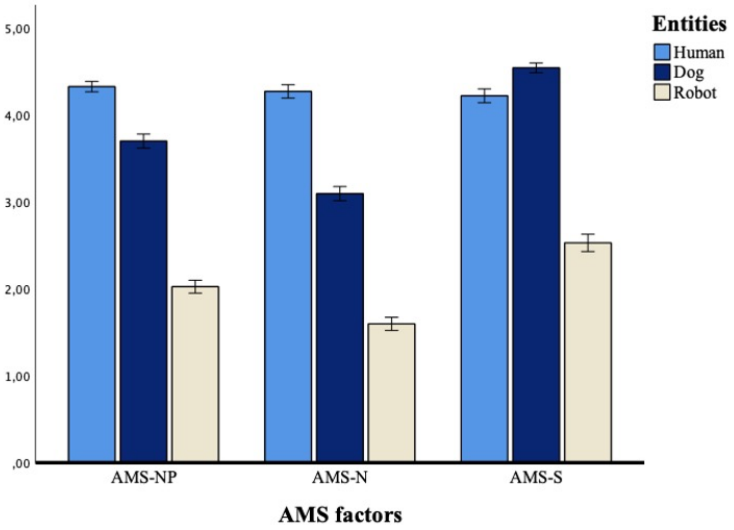


Figure 5 - Differences among stimuli in the attribution of mental and sensory states.

Table 7 – AMS-Q differences in the attribution of mental and sensory states to a human, a dog, and a robot.

Entity	**AMS-NP**		**AMS-N**		**AMS-S**	
	M	*SD*	M	*SD*	M	*SD*
Human	**4.32**	.03	**4.27**	.04	4.22	.04
Dog	3.70	.04	3.09	.04	**4.54**	.04
Robot	2.02	.04	1.60	.04	2.53	.05

Discussion

The aim of the present study was to develop and validate a new questionnaire measuring the attribution of mental states to humans, the Attribution of Mental States Questionnaire (AMS-Q), across two Italian samples. In the current study, we aimed to provide a questionnaire validated with human stimuli that can be used as baseline for comparing the attribution of human mental and sensory traits to different entities – including living entities (e.g., animals, plants, etc.) and anthropomorphic and nonanthropomorphic nonliving entities (e.g., robots, objects, etc.) – and to assess the level of mental anthropomorphism attributed to them.

In Study 1, we found that a 24-item version of the questionnaire had excellent psychometric properties (α = 0.95) and a three-factor structure. Exploratory Factor Analysis revealed that Factor 1 (Mental states with neutral or positive valence – AMS-NP) is composed of thirteen items: seven items concerning the attribution of epistemic mental states (beliefs, thoughts, inferences), three items concerning feelings, states of well-being, and positive emotions (love, have fun, and be happy), and three items concerning planning and volitional mental states (have the intention to do something, have a will to do something, and expressing a desire). Six items loaded on Factor 2 (Mental states with negative valence – AMS-N), three of which involved the attribution of cognitive mental states that belong to the semantic field of deception (i.e., tell a lie, deceive, and make a joke) and three related emotional states (be sad, angry, and afraid). Four items assessing the attribution of sensory states (i.e., hear, smell, look, and taste) are loaded on Factor 3 (Sensory states – AMS-S). This factor structure was confirmed in a new independent sample in Study 2 via Confirmatory Factor Analysis (CFA). Factors had high internal consistency and sufficient convergent and divergent validity. To further strengthen the structure of the questionnaire we revised the three-factor model by excluding one item (*"feeling hot or cold"*) and including correlations among errors for some pairs of items. The final version of the questionnaire included 23 items and maintained excellent internal consistency (α = 0.93). The abovementioned three-factor structure was also demonstrated through the multigroup CFA across gender. The goodness-of-fit indices were adequately close to support the three-factor model, which was also demonstrated by the strength of the factor loadings. This further means that females and males interpret the items in the same way and that the factor loadings are stable across groups.

As a reflection of this, the structure was also consistent from a theoretical perspective. The model revealed three factors that distinguished between the attribution of mental states (AMS-NP and AMS-N) and the attribution of sensory states (AMS-S). This was partially consistent with research showing that people intuitively think about other minds in terms of agency (the ability to plan and act) and experience (the ability to perceive and feel) (Gray et al., 2007; Gray et al. 2011). The AMS questionnaire clearly distinguished the dual nature of the mentalistic lexicon, with mental states on one side and sensory states on the other. In Gray et al. (2007, 2011) the "experience" dimension of mental perception also includes the ability to feel fear, pain, pleasure, joy, etc. However, in the AMS questionnaire, these emotional states loaded into either the first or second factor (mental states). Also, in the present model, positive and negative emotions appeared to be at the opposite poles of a continuum of prosocial and antisocial use of mentalization ability. The former seems to be the emotional reactions resulting from prosocial behavior. On the other hand, negatively valenced emotions are loaded along with mentalistic verbs reflecting behaviors that require antisocial use of ToM abilities. Behaviors such as lying and deception fall under the concept of *Nasty ToM* (Happé & Frith, 1996) and are characterized by an intact but distorted mentalization ability in the domain of antisocial behavior (McEwen et al. 2007; Lonigro et al., 2014; Ronald et al. 2005). The fact that Factor 1 and Factor 2 items did not load on the same factor points to the possibility that mentalistic language distinguishes behaviors that on the value level are perceived as positive or neutral (e.g., thinking) or negative (e.g., pretending). Similarly, AMS-Q reflects real life, in which few social situations are neutral and the ability to grasp the intentions, beliefs, desires, and emotions of others can be used in prosocial or antisocial ways. Indeed, people consistently use their mind-reading abilities to understand and even control another's behavior by manipulating, teasing, or other antisocial purposes (Arefi, 2010). Likewise, mentalizing abilities can offer help and cooperation, care about others, and consider their feelings. AMS-Q is thus able to capture the nuances of social behaviors that require the use of ToM, effectively distinguishing between "nice" and "nasty" ToM behaviors and their emotional consequences.

The AMS-Q demonstrated promising convergent validity as evidenced by correlations with validated measures of Theory of Mind and mentalization skills. The convergent validity of the AMS-Q was tested with the Eyes Test (ET; Baron-Cohen et al., 2001) which is considered to be an established measure of mentalization as it assesses adults' ability to recognize the mental state of others using just the expressions around the eyes, which are key in determining mental states. The ET goes beyond merely assessing mentalizing abilities but assesses the extent to which people attribute mental states to others. This specificity made ET the ideal measure to correlate with AMS-Q because, although they have different purposes, both are based on the assessment of the attribution of mental states. As we expected, the AMS-Q subscales were significantly correlated with the ET, which means that the questionnaire is a valid tool that measures the attribution of mental states. Also, we found significant positive correlations with some scales of the Multidimensional Mentalizing Questionnaire (MMQ; Gori et al., 2021), a tool that assesses several core aspects of mentalization that, although all interrelated, concern relatively distinct capacities, such as cognitive-affective, self-other, outside-inside, and explicit-implicit. The Reflexivity, Ego-Strength, and Relational Attunement subscales refer to "positive" and functional components of mentalization (Gori et al., 2021) and are correlated with AMS factors as they focus on understanding others, acquiring their perspective, and being able to tune into the emotional and cognitive states of others and deeply understand their experiences. These are necessary components of mentalization and subsequent attribution of mental states. Conversely, we found no correlations with Relational Discomfort, Distrust, and Emotional Dyscontrol subscales since they refer to failures and distortions and evaluate manifestations of non-mentalizing states, which are not specifically assessed in the AMS-Q. Furthermore, as predicted, negative correlations were found between the AMS-NP and AMS-N and the construct of alexithymia; conversely, no correlations were found between the AMS-

S and TAS-20 subscales. This is consistent as high scores in alexithymia indicate a difficulty in recognizing and attributing mental states; also, the AMS-S subscale assesses the ability to attribute sensory states while alexithymia can be defined as the inability to experience and identify emotions and reveals uncertainty about the emotional states of others and oneself.

In line with previous results with children and adults (Di Dio et al., 2018, 2020a,b; Manzi et al., 2020b, 2021c), the present data also showed that AMS-Q can discriminate the attribution of internal states to humans from nonhuman agents. In fact, the AMS-Q was able to differentiate between the entities used in the present study: human, dog, and robot. The GLM analysis indicated a significant difference in the attribution of mental and sensory states, resulting in greater attribution to humans than to robots and dogs, except for sensory states, where the dog was the highest-scoring entity. As a matter of fact, dogs have more developed senses (e.g., smell) than humans and this finding further enhances the sensitivity of the AMS to pick up on differences in the attribution of states, reflecting reality. Instead, the robot, contrary to the human and the dog, was perceived as an entity with low psychological and sensory skills. Overall, these results are in line with previous studies (Di Dio et al. 2018, 2020a,b; Hackel et al., 2014; Martini et al., 2016; Manzi et al., 2020b) reporting that different agents, or even the same agent with different characteristics (e.g., different types of robots; Manzi et al., 2020b), can evoke different – although diminished – attributions of human mental traits. Importantly, the tendency to attribute mental states to robots is also determined by factors such as people's age, motivation, cultural background, and attitude toward robots, as well as the behavior, appearance, and identity of the robot (Marchesi et al., 2019; Thellman et al., 2022). Likewise, in a recent study, Manzi and colleagues (2021c) showed that humans are particularly sensitive to the design of robots in terms of attribution of mental qualities; in fact, even when robots differ slightly in their physical appearance, the dissimilar design evokes different mental properties. Consistently, previous studies in which the AMS-Q was administered (Di Dio et al., 2018; 2019; 2020a; Manzi et al., 2020a) have shown that children attribute qualitatively different internal states to humans compared to robots, highlighting the sensitivity of the AMS-Q in capturing these differences. Moreover, correlational studies with the AMS-Q have identified those factors, i.e., the age (Di Dio et al., 2020b; Manzi et al., 2020b) and the human likeness (for a review, see Marchetti et al., 2018) can influence the perception of the minds of robotic agents. In this framework, the AMS-Q stands as a valuable questionnaire that can capture the individuals' ability to evaluate the level of mental anthropomorphism of nonhuman entities, including animals (Urquiza-Haas & Kotrschala, 2015), inanimate things (e.g., robot: Di Dio et al., 2019; 2020a; Manzi et al., 2020a; 2021c), paranormal entities (Gray et al., 2007), and even God (Di Dio et al., 2018); and provides interesting suggestions with respect to which factors may evoke different attributions of mental states. Therefore, the perception of the minds of living and nonliving beings has important implications. For instance, as Gray and colleagues (2007) have pointed out, there is a strong connection between the perception of mind and morality, such that attributing less mind to an entity also reduces its moral status, consequently affecting how people interact with that entity or agent. For example, the way people perceive and attribute mental states to others can lead to helping and praising or, conversely, denigrating and hurting. It may be concluded that the attribution of human mental traits (or the opposite dementalization) is predictive of attitudes (Urquiza-Haas & Kotrschala, 2015) and involve moral (Gray et al., 2007; Manzi et al., 2020c) and social evaluation processes (Kteily et al., 2016). Another advantage of the AMS-Q is that its data can be used flexibly in a variety of ways, as it allows for the investigation of the attribution of human mental states to nonhuman agents in order to assess the level of mental anthropomorphism. It may help explain the belief in God, the humanization of pets, and the attribution of responsibility to computers; and finally, it is a useful measure to identify which factors and conditions contribute to the increase or decrease in the process of mental anthropomorphizing. In conclusion, Dennett (1996) claimed that each mind is defined as such by the eye of the beholder, this is because it is individual perceptions that answer the question "what kind of things have a mind". However, the AMS-Q has shown to be able to capture not only whether things have more or fewer minds but to explore their dimensions, capturing "nice" and "nasty" attributes and their emotional consequences.

Conclusions and limitations

Important conclusions can be drawn from the current study. The Attribution of Mental State Questionnaire (AMS-Q) has shown good psychometric properties; the rapid and easy administration of the measure allows a comprehensive assessment of the attribution of mental and sensory states to human and the comparison with nonhuman entities. Moreover, this research has highlighted the sensitivity of the AMS-Q in distinguishing between mental and sensory states, positive (or neutral) and nasty attributes and their emotional correlates, and in discriminating among agents in terms of mental states. The AMS-Q can be usefully adopted in research whose goal is to identify possible differences in the attribution of mental and sensory states between entities, using the human stimuli as baseline: the theoretical framework proposed here can provide important suggestions in the perception of nonhuman entities as more or less mentalistic comparable to humans. The AMS-Q may also provide insight into the possible difference between age groups and the factors required for human mental traits to be attributed to nonhuman agents, further helping to delineate the perception of others' minds.

The study has some limitations that need to be acknowledged. Although the entire sample was of adequate size, there are significant differences in age, suggesting that the youngest may have greater weight in the analysis. In addition, our sample drew only from a nonclinical population. It is worth noting a gender difference in levels of mentalization, with females having higher mentalization abilities than males (Dimitrijevic et al., 2018; Focquaert et al., 2007). This gender effect could affect anthropomorphic attribution and thus the outcome of the questionnaire. However, this bias does not seem to compromise the structure of the questionnaire presented in the article. This was also confirmed by the multigroup CFA: most indices were close to the recommended cutoff values. However, replication with larger samples would allow higher levels of certainty regarding the underlying three-factor structure.

Another limitation is to have used only two stimuli (dog and robot) to assess discriminant validity. However, our findings are supported by previous studies that indicate the sensitivity of AMS-Q to grasp differences in the attribution of mental and sensory states. It is important to note that the images we used were given as an example and were selected as representing the categories of living and non-living entities. AMS-Q is thought to be administered with a variety of stimuli, from animals to inanimate things, to paranormal entities, and even God. Thus, in future studies, stimuli different from those reported in this study can be administered, depending on the focus of the research question, always keeping human stimuli as baseline to assess the anthropomorphization of non-human agents.

Despite the above limitations, for the present time, the AMS-Q seems well-positioned to fill the void in mental states attribution measures and appears to have the potential as a reliable and psychometrically valid questionnaire for research applications, worthy of further empirical investigation. Although future research with AMS-Q involving different clinical samples and investigating structure stability over time is needed, the results of the studies reported in this article provide preliminary evidence for its reliability and validity and highlight possibilities for its broader application.

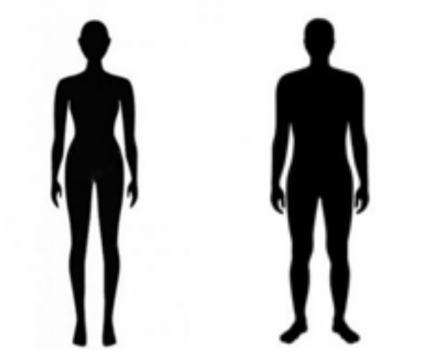
Appendix 1 – **Attribution of Mental States Questionnaire (AMS-Q)**

Answer the following questions using the scale provided: 1 No, not at all; 2 Yes, a little; 3 Yes, quite a bit; 4 Yes, a lot; 5 Yes, a lot.

No, per nulla *Not, at all*	1	2	3	4	5	Sì, moltissimo *Yes, a lot*

Secondo te, l'essere umano può […]?

In your opinion, can human beings [...]?



AMS-NP	AMS-N	AMS-S					
Imparare *Learn*			1	2	3	4	5
Pensare *Think*			1	2	3	4	5
Ricordare *Remember*			1	2	3	4	5
Decidere *Make a decision*			1	2	3	4	5
Capire *Understand*			1	2	3	4	5

Sognare *Dream*			1	2	3	4	5
Immaginare *Imagine*			1	2	3	4	5
Divertirsi *Have fun*			1	2	3	4	5
Voler bene *Love*			1	2	3	4	5
Essere felice *Be happy*			1	2	3	4	5
Avere intenzione di fare qualcosa *Have the intention to do something*			1	2	3	4	5
Avere voglia di fare qualcosa *Want to do something*			1	2	3	4	5
Esprimere un desiderio *Make a wish*			1	2	3	4	5
	Dire una bugia *Tell a lie*		1	2	3	4	5
	Fare uno scherzo *Make a joke*		1	2	3	4	5
	Far finta *Pretend*		1	2	3	4	5
	Essere triste *Be sad*		1	2	3	4	5
	Avere paura *Be scared*		1	2	3	4	5
	Arrabbiarsi *Get angry*		1	2	3	4	5
		Udire *Hear*	1	2	3	4	5

Annusare *Smell*	1	2	3	4	5
Guardare *See*	1	2	3	4	5
Gustare *Taste*	1	2	3	4	5

Appendix 2 – **Scoring: Attribution of Mental States Questionnaire (AMS-Q)**

The questionnaire is administered with images of a human silhouette (see above) representing baseline. Female and male images are administered across participants in random order, so that half of the participants will be presented with a female silhouette and half with a male silhouette. If the protocol foresees the presentation of a real human being, it is suggested to use a picture of that specific character instead. For example, if the participants are requested to attribute mental states to an experimenter, the silhouette image can be replaced with the picture of the experimenter. The questionnaire can also be used alongside other stimuli (living or non-living, e.g., images of God, an animal, an object, etc.). The items within the questionnaire must also be randomized.

Calculate the average of the items:

Mental states with neutral or positive valence (AMS-NP): AMS_1; AMS_2; AMS_3; AMS_4; AMS_5; AMS_6; AMS_7; AMS_8; AMS_9; AMS_10; AMS_11; AMS_12; AMS_13.

Mental states with negative valence (AMS-N): AMS_14; AMS_15; AMS_16; AMS_17; AMS_18; AMS_19.

Sensory states (AMS-S): AMS_20; AMS_21; AMS_22; AMS_23.

Finally, as a practical suggestion, creating a single index, AMS-Q scores obtained when evaluating mental states attribution of non-human entities may be subtracted to AMS-Q scores for the human entity (using the silhouette image as in this validation or the image of a specific human being involved in a particular experimental design). Positive scores would highlight greater attributes to the entity as compared to the human being, whereas negative scores would pinpoint the opposite.

Availability Statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics Statement

The studies involving human participants were reviewed and approved by Commissione Etica per la Ricerca in Psicologia, CERPS (Università Cattolica del Sacro Cuore, Milano). The patients/participants provided their written informed consent to participate in this study.

Author Contributions

All authors contributed to the study conception and design, commented on the initial versions, read, and approved the final manuscript. FM conceptualized the scale. CDD and LM secured ethical approval. LM, GP, and FM performed material preparation and data collection. LM carried out the statistical analysis. CDD suggested important improvements to the methodology. LM and GP wrote the first draft of the manuscript.

Funding

This research was funded by Università Cattolica del Sacro Cuore (D.1).

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.

Supplementary Material

The Supplementary material for this article can be found online at:
https://www.frontiersin.org/articles/10.3389/fpsyg.2023.999921/full#supplementary-material

References

Abell, F., Happé, F., and Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. Cognitive Development, 15(1), 1-16. DOI: 10.1016/S0885-2014(00)00014-9

Airenti, G. (2015). The cognitive bases of anthropomorphism: from relatedness to empathy. International Journal of Social Robotics, 7(1):117-127. DOI: 10.1007/s12369-014-0263-x

Arefi, M. (2010). Present of a casual model for social function based on theory of mind with mediating of Machiavellian beliefs and hot empathy. Social and Behavioral Sciences, 5:694-697. DOI: 10.1016/j.sbspro.2010.07.167

Astington, J. W. (2003). "Sometimes necessary, never sufficient: False-belief understanding and social competence", in Individual differences in theory of mind: Implications for typical and atypical development, Eds B. Repacholi & V. Slaughter (Psychology Press), 13-38.

Astington, J. W., and Baird, J. A. (2005). Why language matters for theory of mind. New York: Oxford University Press.

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. J Child Psychol Psychiatry, 42:241-251. DOI: 10.1111/1469-7610.00715.

Bartneck, C., Kanda, T., Mubin, O., Al Mahmud, A. (2009). Does the design of a robot influence its animacy and perceived intelligence?. International Journal of Social Robotics, 1(2):195-204. DOI: 10.1007/s12369-009-0013-7

Bateman, A., and Fonagy, P. (2010). Mentalization based treatment for borderline personality disorder. World Psychiatry: Official Journal of the World Psychiatric Association (WPA), 9(1):11–15. DOI: 10.1002/j.2051-5545.2010.tb00255.x

Beeghly, M., Bretherton, I., and Mervis, C. B. (1986). Mothers' internal state language to toddlers. British Journal of Developmental Psychology, 4(3):247-261. DOI: 10.1111/j.2044-835X.1986.tb01016.x

Bellagamba, F., Laghi, F., Lonigro, A., Pace, C. S. (2012). Re-enactment of intended acts from a video presentation by 18- and 24-month-old children. Cognitive Processing, 13(4):381–386. DOI: 10.1007/s10339-012-0518-0

Bentler, P.M. (1990). Comparative fit indexes in structural models. Psychological Bulletin, 107(2):238-246. DOI: 10.1037/0033-2909.107.2.238

Bressi, C., Taylor, G., Parker, J., Bressi, S., Brambilla, V., Aguglia, E., Allegranti, I., Bongiorno, A., Giberti, F., Bucca, M., Todarello, O., Callegari, C., Vender, S., Gala, C., Invernizzi, G. (1996). Cross validation of the factor structure of the 20-item Toronto Alexithymia Scale: an Italian multicenter study. Journal of psychosomatic research, 41(6):551-559. DOI: 10.1016/s0022-3999(96)00228-0

Bretherton, I., and Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. Developmental psychology, 18(6):906-921. DOI: 10.1037/0012-1649.18.6.906

Browne, M. W., and Cudeck, R. (1993). “Alternative ways of assessing model fit”, in Testing structural equation models, Eds. K. A. Bollen and J. S. Long (Newbury Park, CA, Sage), 136-162.

Byrne, B. M. (1994). Structural Equation Modelling with EQS and EQS/Windows: Basic Concepts, Applications, and Programming. Sage.

Choi-Kain, L. W., and Gunderson, J. G. (2008). Mentalization: ontogeny, assessment, and application in the treatment of borderline personality disorder. Am J Psychiatry, 165(9):1127-35. DOI: 10.1176/appi.ajp.2008.07081360

Dario, P., Guglielmelli, E., and Laschi, C. (2001). Humanoids and personal robots: Design and experiments. Journal of Robotic Systems, 18(12):673-690. DOI: 10.1002/rob.8106

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, 1-14. DOI: 10.1177/2158244018809874

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. 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. DOI: 10.3389/fpsyg.2020.00469

Dimitrijevic, A., Hanak, N., Altaras Dimitrijevic, A., Jolic Marjanovic, Z. (2017). The mentalization scale (MentS): A self-report measure for the assessment of mentalizing capacity. Journal of Personality Assessment, 100(3):268-280. DOI: 10.1080/00223891.2017.1310730

Fan, X.B., Thompson, & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes, Structural Equation Modeling: A Multidisciplinary Journal, 6:(1)56-83. DOI: 10.1080/10705519909540119

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, May 21-23, 2012, 54-59, München, Munich, Germany: Technische Universität. DOI: 10.1109/ARSO.2012.6213399

Focquaert, F., Steven, M. S., Wolford, G. L., Colden, A., Gazzaniga, M. S. (2007). Empathizing and systemizing cognitive traits in the sciences and humanities. Personality and Individual Differences, 43(3):619-625. DOI: 10.1016/j.paid.2007.01.004

Fonagy, P. (1989). On tolerating mental states: theory of mind in borderline patients. Bull Anna Freud Centre, 12:91-115.

Fonagy, P. (1991). Thinking about thinking: some clinical and theoretical considerations in the treatment of a borderline patient. Int J Psychoanal, 72:639–56.

Fonagy, P., and Bateman, A. (2008). The development of borderline personality disorder—a mentalizing model. J Pers Disord, 22(1):4–21. DOI: 10.1521/pedi.2008.22.1.4

Fonagy, P., Gergely, G., Target, M. (2007) The parent-infant dyad and the construction of the subjective self. J Child Psychol Psychiatry, 48(3-4):288–328. DOI: 10.1111/j.1469-7610.2007.01727.x

Fonagy, P., and Luyten, P. (2009). A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. Dev Psychopathol., 21(4):1355-81. DOI: 10.1017/S0954579409990198

Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y-W., Warren, F., Howard, S., et al. (2016). Development and Validation of a Self-Report Measure of Mentalizing: The Reflective Functioning Questionnaire. PLoS ONE 11(7):e0158678. DOI: 10.1371/journal.pone.0158678

Fonagy, P., Steele, H., Moran, G., Steele, M., Higgitt, A. (1991). The capacity for understanding mental states: the reflective self in parent and child and its significance for security of attachment. Infant Mental Health Journal, 13:200-217. DOI: 10.1002/1097-0355(199123)12:3<201.

Frith, C. D., and Frith, U. (1999). Interacting minds-A biological basis. Science, 286:1692-1695.

Frith, C. D., and Frith, U. (2006). The neural basis of mentalizing. Neuron, 50(4):531-534. DOI: 10.1016/j.neuron.2006.05.001

Gervais, W. M. (2013). Perceiving minds and Gods: How mind perception enables, constrains, and is triggered by belief in Gods. Perspectives on Psychological Science, 8:380-394. DOI: 10.1177/1745691613489836

Giménez-Dasí, M., Guerrero, S., & Harris, P. L. (2005). Intimations of immortality and omniscience in early childhood. European Journal of Developmental Psychology, 2:285-297. DOI: 10.1080/17405620544000039

Giovanelli, C., Di Dio, C., Lombardi, E., Tagini, A., Meins, E., Marchetti, A., Carli, L. (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

Gopnik, A., and Wellman, H. M. (1992). Why the child's theory of mind really is a theory. Mind & Language, 7(1-2):145-171. DOI: 10.1111/j.1468-0017.1992.tb00202.x

Gori, A., Arcioni, A., Topino, E., Craparo, G., Lauro Grotto, R. (2021). Development of a new measure for asessing mentalizing: the Multidimensional Mentalizing Questionnaire (MMQ). J. Pers. Med.,11:305. DOI: 10.3390/jpm11040305

Gray, H. M., Gray, K., and Wegner, D. M. (2007). Dimensions of mind perception. Science (New York, N.Y.), 315(5812):619. DOI: 10.1126/science.1134475

Gray, K., Jenkins, A. C., Heberlein, A. S., and Wegner M. W. (2011). Distortions of mind perception in psychopathology. PNAS, 108(2):477-479. DOI: 10.1073/pnas.1015493108

Gray, K., and Wegner, D. M. (2012). Feeling robots and human zombies: mind perception and the uncanny valley. Cognition, 125(1):125-130. DOI: 10.1016/j.cognition.2012.06.00

Greenberg, D. M., Kolasi, J., Hegsted, C. P., Berkowitz, Y., and Jurist, E. L. (2017). Mentalized affectivity: a new model and assessment of emotion regulation. PLoS ONE 12(10):e0185264. DOI: 10.1371/journal.pone.0185264

Hackel, L. M., Looser, C. E., & Van Bavel, J. J. (2014). Group membership alters the threshold for mind perception: The role of social identity, collective identification, and intergroup threat. Journal of Experimental Social Psychology, 52:5-23. DOI: 10.1016/j.jesp.2013.12.001

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., Black, W. C. (1998). Multivariate data analysis (5th ed.). New York, NY: Macmillan.

Happé, F., and Frith, U. (1996). Theory of mind and social impairment in children with conduct disorder. British Journal of Developmental Psychology 14(9984):385-398. DOI: 10.1111/j.2044-835X.1996.tb00713.x

Harkness, K. L., Jacobson, J. A., Duong, D., Sabbagh, M. A. (2010). Mental state decoding in past major depression: Effect of sad versus happy mood induction. Cognition and Emotion, 24:497-513.

Harris, P. L., and Koenig, M. A. (2006). Trust in testimony: How children learn about science and religion. Child Development, 77:505-524. DOI: 10.1111/j.1467-8624.2006.00886.x

Heider, F., and Simmel, M. (1944). An experimental study of apparent behavior. The American Journal of Psychology, 57(2):243-259. DOI: 10.2307/1416950

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. Psychometrika 30:179-185.

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1):1-55. DOI: 10.1080/10705519909540118

Hubbard, R., and Allen, S. J. (1987). A cautionary note on the use of principal components analysis: supportive empirical evidence. Sociological Methods & Research, 16(2):301-308. DOI: 10.1177/0049124187016002005

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20(1):141-151. DOI: 10.1177/001316446002000116

Kiesler, S., Powers, A., Fussell, S. R., Torrey, C. (2008). Anthropomorphic interactions with a robot and robot-like agent. Social Cognition, 26:2169-181. DOI: 10.1521/soco.2008.26.2.169

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. PloS one, 3(7):e2597. DOI: 10.1371/journal.pone.0002597

Kteily, N., Hodson, G., and Bruneau, E. (2016). They see us as less than human: Metadehumanization predicts intergroup conflict via reciprocal dehumanization. Journal of Personality and Social Psychology, 110(3):343–370. DOI: 10.1037/pspa0000044

Lai, K., & Green, S. B. (2016). The problem with having two watches: Assessment of fit when RMSEA and CFI disagree. Multivariate behavioral research, 51(2-3):220-39. DOI: 10.1080/00273171.2015.1134306.

Lecce, S., and Pagnin, A. (2007). Il lessico psicologico. La teoria della mente nella vita quotidiana. Il Mulino (ed). Bologna, Italy.

Lonigro, A., Laghi, F., Baiocco, R., Baumgratner, E., (2014). Mind reading skills and empathy: evidence for nice and nasty ToM behaviours in school-aged children. J Child Fam Stud, 23:581–590. DOI 10.1007/s10826-013-9722-5

Luyten P., Fonagy P., Lowyck B., Vermote R. (2012). Assessment of mentalization. In: Bateman A., Fonagy P., editors. Handbook of mentalizing in mental health practice, 43-65. Washington, DC: American Psychiatric Association.

MacDorman, K. F., Minato, T., Shimada, M., Itakura, S., Cowley, S., Ishiguro, H. (2005). Assessing human likeness by eye contact in an android testbed. In Proceedings of the XXVII annual meeting of the cognitive science society, 21-23. Stresa, Italy.

Malle, B. F. (2019). How many dimensions of mind perception really are there?. In A.K. Goel, C.M. Seifert, & C. Freksa (Eds.), Proceedings of the 41st Annual Meeting of the Cognitive Science Society; pp. 2268-2274. Montreal, QB: Cognitive Science Society.

Manzi, F., Di Dio, C., Di Lernia, D., Rossignoli, D., Maggioni, M., Massaro, D., Marchetti, A., Riva G. (2021a). Can you activate me? From robots to humans' brain. Frontiers in Robotics and AI, 8:633514. DOI: 10.3389/frobt.2021.633514

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 Preschoolers, Paper, in Proceedings of the Workshop on Adapted intEraction with SociAl Robots, (Cagliari, 17-17 March 2020), CEUR Workshop Proceedings, Aachen 2020:2724 20-27

Manzi, F., Massaro, D., Di Lernia, D., Maggioni, M., Riva G., Marchetti, A., (2021c). Robots are not all the same: young adults' expectations, attitudes and mental attribution to two humanoid social robots. Cyberpsychology, Behavior, and Social Networking, 24(5):307-314. DOI: 10.1089/cyber.2020.0162

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 XXX Congresso Nazionale, Associazione Italiana di Psicologia, Sezione di Psicologia dello Sviluppo e dell'Educazione (Messina, 14-16 September 2017), 65-66. Italy: Alpes Italia srl. Available online at: http: //hdl.handle.net/10807/106022

Manzi, F., Peretti, G., Di Dio, C., Cangelosi, A., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., and Marchetti, A. (2020b). A Robot Is Not Worth Another: Exploring Children's Mental State Attribution to Different Humanoid Robots. Front. Psychol., 11:2011. DOI: 10.3389/fpsyg.2020.02011

Manzi, F., Sorgente, A., Massaro, D., Villani D., Di Lernia, D., Malighetti, C., Gaggioli, A., Rossignoli, D., Sandini, G., Sciutti, A., Rea, F., Maggioni, M., Marchetti, A., Riva, G. (2021b). Emerging adults' expectations about next generation of robots: Exploring robotic needs through a latent profile analysis. Cyberpsychology, Behavior, and Social Networking, 24(5):315-323. DOI: 10.1089/cyber.2020.0161

Marchesi, S., Ghiglino, D., Ciardo, F., Perez-Osorio, J., Baykara, E., and Wykowska, A. (2019). Do We Adopt the Intentional Stance Toward Humanoid Robots? Front. Psychol. 10:450. DOI: 10.3389/fpsyg.2019.00450

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. 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

McEwen, F., Happé, F., Bolton, P., Rijsdijk, F., Ronald, A., Dworzynski, K., et al. (2007). Origins of individual differences in imitation: Links with language, pretend play, and socially insightful behaviour in two-year-old twins. Child Development 78(2):474-492.

Meins, E., Fernyhough, C., de Rosnay, M., Arnott, B., Leekam, S. R., Turner, M. (2012). Mind-mindedness as a multidimensional construct: Appropriate and nonattuned mind-related comments independently predict infant-mother attachment in a socially diverse sample. Infancy, 17(4):393-415. DOI: 10.1111/j.1532-7078.2011.00087.x

Meins, E., Fernyhough, C., Wainwright, R., Das Gupta, M., Fradley, E., Tuckey, M. (2002). Maternal mind–mindedness and attachment security as predictors of theory of mind understanding. Child development, 73(6):1715-1726. DOI: 10.1111/1467-8624.00501

Mull, M. S., and Evans, E. M. (2010). Did she mean to do it? Acquiring a folk theory of intentionality. Journal of Experimental Child Psychology, 107(3):207–228. DOI: 10.1016/j.jecp.2010.04.001

Nelson, K. (2005). Language pathways into the community of minds. In J. W. Astington & J. A. Baird (eds), Why language matters for theory of mind, 26-49. New York: Oxford University Press.

Nyhof, M. A., and Johnson, C. N. (2017). Is God just a big person? Children's conceptions of God across cultures and religious traditions. British Journal of Developmental Psychology, 35:60-75. DOI: 10.1111/bjdp.12173

Peretti, G., Manzi, F., Di Dio, C., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2023). Can a robot lie? Young children's understanding of intentionality beneath false statements. Infant and Child Development, e2398. https://doi.org/10.1002/icd.2398

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(3):437–471. DOI: 10.1016/0022-0965(85)90051-7

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

Ramsey, R., and Hamilton, A. F. D. C. (2010). Triangles have goals too: understanding action representation in left aIPS. Neuropsychologia, 48(9):2773-2776. DOI: 10.1016/j.neuropsychologia.2010.04.028

Ronald, A., Happé, F., Hughes, C., Plomin, R. (2005). Nice and nasty theory of mind in preschool children: Nature and nurture. Social Development, 14(4):664–684. DOI: 10.1111/j.1467-9507.2005.00323.x

Schermelleh-Engel, K., Moosbrugger, H., and Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. Methods of Psychological Research, 8(2):23-74.

Slaughter, V., Peterson, C. C., and Carpenter, M. (2009). Maternal mental state talk and infants’ early gestural communication. Journal of Child Language, 36(5):1053-1074. DOI: 10.1017/S0305000908009306

Symons, D. K., Fossum, K. L. M., and Collins, T. K. (2006). A longitudinal study of belief and desire state discourse during mother–child play and later false belief understanding. Social development, 15(4):676-692. DOI: 10.1111/j.1467-9507.2006.00364.x

Taumoepeau, M., and Ruffman, T. (2006). Mother and infant talk about mental states relates to desire language and emotion understanding. Child Development 77:465-81. DOI: 10.1111/j.1467-8624.2006.00882.x

Taumoepeau, M., and Ruffman, T. (2008). Steppingstones to others’ minds: Maternal talk relates to child mental state language and emotion understanding at 15, 24 and 33 months. Child Development 79:284-302. DOI: 10.1111/j.1467-8624.2007.01126.x

Thellman, S., Silvervarg, A., and Ziemke, T., (2017). Folk-Psychological Interpretation of Human vs. Humanoid Robot Behavior: Exploring the Intentional Stance toward Robots. Front. Psychol. 8:1962. DOI: 10.3389/fpsyg.2017.01962

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

Tomasello, M. (1999). The cultural origins of human cognition. Cambridge, MA: Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. Behavioral and Brain Sciences, 28(5):675-691. DOI: 10.1017/S0140525X05000129

Urquiza-Haas, E. G., and Kotrschal, K. (2015). The mind behind anthropomorphic thinking: attribution of mental states to other species. Animal Behaviour, 109:167-176. DOI: 10.1016/j.anbehav.2015.08.011

Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C., & Preti A. (2013). The "Reading the Mind in the Eyes" test: Systematic review of psychometric properties and a validation study in Italy. Cognitive Neuropsychiatry, 18(4):326-354. DOI: 10.1080/13546805.2012.721728

Voracek, M., and Dressler, S. G. (2006). Lack of correlation between digit ratio (2D:4D) and Baron-Cohen's "Reading the Mind in the Eyes" test, empathy, systemising, and autism-spectrum quotients in a general population sample. Personality and Individual Differences, 41:1481-1491.

Waytz, A., Cacioppo, J., Epley, N. (2010). Who sees human? The importance and stability of individual differences in anthropomorphism. Perspect. Psychol. Sci. 5(3):219-232. DOI: 10.1177/1745691610369336

Waytz, A., Gray, K., Epley, N., Wegner, D. M. (2010). Causes and consequences of mind perception. Trends in Cognitive Sciences, 14:383-388. DOI: 10.1016/j.tics.2010.05.006

Wellman, H. M. (1992). The child's theory of mind. The MIT Press.

Wellman, H. M. (2017). The development of theory of mind: Historical reflections. Child Development Perspectives, 11:207-214. DOI: 10.1111/cdep.12236

Wellman, H. M. (2020). Reading Minds: How childhood teaches us to understand people. Oxford, UK: Oxford University Press.

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

Westwood, H., Kerr-Gaffney, J., Stahl, D., Tchanturia, K. (2017). Alexithymia in eating disorders: Systematic review and meta-analyses of studies using the Toronto Alexithymia Scale. Journal of psychosomatic research, 99:66-81. https://doi.org/10.1016/j.jpsychores.2017.06.007

Wiese, E., and Weis, P. P. (2020). It matters to me if you are human-Examining categorical perception in human and nonhuman agents. International Journal of Human-Computer Studies, 133:1-12. DOI: 10.1016/j.ijhcs.2019.08.002

Wigger, J. B., Paxson, K., and Ryan, L. (2013). What do invisible friends know? Imaginary companions, God, and theory of mind. International Journal for the Psychology of Religion, 23:2-14. DOI: 10.1080/10508619.2013.739059

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

Złotowski, J., Proudfoot, D., Yogeeswaran, K., Bartneck, C. (2015). Anthropomorphism: opportunities and challenges in human-robot interaction. International Journal of Social Robotics, 7(3):347-360. DOI: 10.1007/s12369-014-0267-6

Zwick, W. R., and Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. Psychological Bulletin, 99(3):432-442. DOI: 10.1037/0033-2909.99.3.432

Chapter 5
General Conclusion

In an attempt to answer the question "What is a robot?", Isaac Asimov wrote *"A robot is a robot. Gears and metal, electricity and positrons, mind and iron! Human-made! If necessary, human-destroyed!"* (I. Asimov, *I, Robot*, 1950). Asimov succinctly captures the essence of what a robot is and provides a concise yet thought-provoking definition of a robot, highlighting its components, human origin, and the ethical considerations surrounding its existence and control. In a broader context, this quote captures the essence of what defines a robot as a technological creation with the potential for both great utility and significant ethical and societal issues. Today encounters with increasingly sophisticated social robots occur in real-life settings. As a consequence, the HRI discipline raises questions about how individuals perceive, communicate with, and interact alongside artificial agents. This is where the present thesis aims to get started. The core aim has been the study of social cognition. The investigation focused on the embodied components and cognitive aspects that underlie the understanding of others' behavior and how they are intertwined across the developmental span. These mechanisms, which are responsible for interpersonal relationship processes, have been investigated in both human-human and human-robot interactions. To this end, the assessment of embodied and cognitive components has been translated to the interactions with robotic agents. This step has been critical in informing the theories that support Human-Robot Interaction (HRI) in terms of identifying the psychological factors necessary and/or sufficient to support the between humans and robots.

Specifically, embodied cognition and social metacognition enable humans to make sense of each other's behaviors. On one hand, the well-known mirror system enables us to internally simulate and directly experience others' states, offering experiential insights into their sensations, actions, emotions, and intentions. On the other hand, people engage with one another by explicitly thinking about others' mental contents through abstract representations, using these metarepresentations to interpret behaviors. These sophisticated psychological mechanisms provide the foundation for understanding and responding to others appropriately and contingently. Almost logically, these abilities are posited to underpin interactions with robotic entities. Throughout this thesis, I have endeavored to explore embodied cognition and social metacognition in the context of human-robot interaction across different stages of life. Furthermore, during my Ph.D. journey, I have also argued that studying social cognition through the lens of HRI yields profound insights into human social cognition.

The first study, "Action Chains and Intention Understanding in 3-6-Year-Old Children," is part of a broader investigation into embodied cognition that aims to explore the mechanisms underlying interactions with robotic agents, both psychologically and physiologically. This study is a pioneering effort to indirectly study the activation of IPL neurons with mirror properties in preschool children. Remarkably, there are no studies in this age group that physiologically examine whether preschoolers can intentionally organize their actions and understand the intentions of others behind observed actions. The ability to understand the intention underlying actions of others is crucial in terms of social cognition since it represents a direct and experiential way to access the sense of individuals' behaviors. Through the record of the opening-mouth muscle activity, we examined the ability to organize actions intentionally and understand the intentions behind the observed actions in 3- to 6-year-old children. Our results showed that preschoolers could select the appropriate motor chain intentionally during action execution, reflecting a physiological mechanism whereby, during the preparation of goal-directed action, there is an activation of the corresponding motor chain driven by the intention to perform a given action (e.g., grasp to eat). Whereas, during the observation of a given action, the chain mechanism shows a delayed activation, that is the muscle activation was found just a few milliseconds before the actual grasp compared to older children (6- to 9-year-old) where the muscle activity was evident 100 milliseconds before. As a result, preschoolers show a delayed activation most likely because their intention coding mirror mechanism is not sufficiently developed to understand others' intentions. The results of the study are of great interest not only because they contribute to fill a gap in the literature on the embodied understanding of intentions, but also because they provide a solid starting point for an exploration of the mechanisms that govern child-robot interactions.

While the state of the art in HRI often emphasizes the importance of the kinematics of the observed robot actions, it is widely acknowledged that effectively managing complex relational dynamics depends on correctly interpreting the behavior of others. This is also true for interactions with robotic entities, where understanding their behaviors means analyzing all the components that define the action itself. In essence, the individual must understand *what* the robot is doing (the goal of the action), *why* it is performing a particular action (the intention behind the action), and *how* the action is executed, thereby conveying information about the agent’s emotional state. This study, which delves into the “what” and “why” of actions – two key dimensions for understanding each other’s behavior – provides a fascinating starting point for investigating the embodied mechanisms that can be explored to facilitate successful interactions between preschoolers and robots. Subsequent research will further illuminate how embodied mechanisms may contribute to children’s engagement with robotic entities.

The second study, “Shared Knowledge in Human-Robot Interaction (HRI),” investigated the role of robotic ostensive cues – such as eye contact, turn-taking, and appropriate contingent responsiveness – in communication between adults and humanoid robots. Drawing from developmental psychology, ostensive cues trigger a basic epistemic trust in the caregiver as a benevolent, cooperative, and reliable source of cultural information. Trust furthers the acquisition of shared knowledge without requiring rigorous scrutiny of the validity or relevance of the information conveyed. Thus, ostensive cues prepare the addressee of communication to receive new and relevant information. We found that ostensive cues play a primary role both in human-human and human-robot interactions. Interestingly, even when exhibited by robots, ostensive cues elicited a similar attribution of communicative intent in the addressee, making robotic social signals central in communication between humans and robots. We highlighted the importance of ostensive cues for effective human-robot communication, emphasizing that considering the robot as a social agent with communicative intentions is essential for engaging in successful interactions. This finding is important in light of HRI as it provides crucial input for practical applications. Specifically, our results support the safe integration of robotic agents in various educational contexts, spanning from infancy to elderhood. We have highlighted the critical role of ostensive communication when a robot conveys information to another individual. Thanks to ostensive cues, such as gaze shifting and using the individual’s name, a basic epistemic trust is developed, making the robot a reliable source of information and an effective communicative partner. This opens up possibilities for their efficient inclusion in educational settings. However, as our results show, humans do not place complete trust in what the robot conveys, or at least, robots are perceived as less reliable sources of information compared to humans. Robots can serve as mediators, tutors, and supportive figures in educational relationships; but the role they play in educational settings remains complementary and does not replace humans’ efforts.

The third study, “Development and Validation of the Attribution of Mental States Questionnaire (AMS-Q): A Reference Tool for Assessing Anthropomorphism,” attempts to fill a void. From a theoretical standpoint, the ability to infer and attribute mental states is a cornerstone of social cognition, underpinning smooth interactions. Furthermore, attributing a mind to another entity gives it the status of a social agent, a prerequisite for meaningful interaction. Methodologically, it becomes essential to have a tool capable of measuring the attribution of mental states to non-human agents, including robots. Attributing a mind, and consequently thinking about and understanding the mental states of other entities in social situations, involves the social metacognitive capacity to navigate complex social interactions, as it allows individuals to interpret social cues, understand the perspectives of others, and adjust behavior accordingly. It contributes effectively to communication, empathy, and social problem-solving. Moreover, we found that the design of robots is critical in the attribution of mental states: even when the robots differ slightly in their physical appearance, the dissimilar design evokes different mental properties. Additionally, we pointed out a strong connection between the perception of mind and morality. The attribution of human mental traits or its opposite, *de-mentalization*, significantly predicts and impacts attitudes, moral judgments, and social interactions with entities or agents. For example, the way people attribute mental states to others can lead to helping and praising or, conversely, denigrating and hurting. For all these reasons, the Attribution of Mental State Questionnaire (AMS-Q) was developed and validated in an Italian population. The AMS-Q measures the attribution of mental states to non-human agents, thereby assessing the level of mental anthropomorphism. In addition, administering the AMS-Q in conjunction with other questionnaires may help to examine various social and demographic factors, such as age, gender, and attitudes toward technology, that may influence the attribution of mental states. This is important for determining idiosyncratic attitudes with regard to the mental capacities of robotic agents and helps experts in the field to understand in depth the nature of human behavior in connection with such agents, even in the function of variables and factors mentioned above.

In conclusion, the three studies discussed in this thesis offer various insights. The lifespan perspective has allowed me to observe how the prevalence of embodied mechanisms underlies the understanding of the other's behavior during infancy, supporting imitation processes. The subsequent development of cognitive skills fits into the matrix of mechanisms that enable understanding of the other through a more thoughtful analysis of the mental content of others on a contextual basis. The identification of the embodied mechanisms and cognitive processes that govern the understanding of others' behavior has enabled the identification of some key factors that are suitable to support effective and functional interactions between human and robotic agents. First, for interactions with humanoid robots to occur, the robot must be perceived as a social agent, which requires the attribution of a mind, i.e., ascribing mental states to robots. This virtuous circle is in turn influenced by the design of the robot, people’s technological attitudes, and the human interactive partner’s socio-demographic factors. At the same time, social signals, a crucial component in HRI, shape the perception of robots as effective communicative and trustworthy partners. Trust, shaped by attachment bonds and prior caregiver experiences, is a psychological component of considerable importance in establishing successful interactions with robots. Within this framework, social cognition stands as the privileged way to access the realm of robotic agents. Future research directions will delve deeper into the role of embodied cognition to uncover whether unconscious and automatic mirroring occurs during social interactions between humans and robots, leading to an *internal simulation* and *direct experience* of humanoids.

Publications not included in this thesis

Di Dio, C., Manzi, F., Miraglia, L., Gummerum, M., Bigozzi, S., Massaro, D., Marchetti, A. (2023). Virtual agents and risk-taking behavior in adolescence: the twofold nature of nudging. Scientific Reports; 13(1):1-11. doi:10.1038/s41598-023-38399-w

Valle, A., Cavalli, G., Miraglia, L., Bracaglia, E. A., Fonagy, P., Di Dio, C., Marchetti, A. (2023). The Risk-Taking and Self-Harm Inventory for Adolescents: Validation of the Italian Version (RTSHIA-I). Behavioral Sciences; 13(4):1-17. doi:10.3390/bs13040321

Rizzato, M., Di Dio, C., Miraglia, L., Sam, C., D'Anzi, S., Antonelli, M., Donelli, D. (2022). Are You Happy? A Validation Study of a Tool Measuring Happiness. Behavioral Sciences. (8):295-312. doi:10.3390/bs12080295

Marchetti, A., Miraglia, L., Di Dio, C. (2020). Toward a Socio-Material Approach to Cognitive Empathy in Autistic Spectrum Disorder. Frontiers in Psychology. 10: 1-4. doi:10.3389/fpsyg.2019.02965

Curriculum Vitae

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

In 2020, I became a Ph.D. candidate at the Unit on Theory of Mind within the Department of Psychology, focusing my research on psychological and psychophysiological mechanisms involved in human-robot interaction, under the supervision of Professor Cinzia Di Dio.

In 2019, I graduated with an MSc in Pedagogy from the Università Cattolica del Sacro Cuore in Milan. My thesis, entitled "In Your Shoes: Cognitive and Affective Empathy in the Autistic Spectrum and Conduct Disorder - A State of the Art Review," was supervised by Professor Antonella Marchetti.

In 2017, I graduated from the Università degli Studi di Milano Bicocca with a BSc in Science of Education. My thesis, entitled "Conflicting Families: Children in parental alienation - from case study to critical reflection", was completed under the supervision of Professor Laura Formenti.

My professional experience includes working as an educator in a residential context designed for adolescents temporarily separated from their families. Later, I transitioned into the role of a school educator specializing in working with children with disabilities. I also took on the role of coordinator of a one-year project aimed at supporting foreign women.

I am a member of the Italian Association of Psychology (AIP) and the International Society for the Study of Behavioral Development (ISSBD).