Computers in Human Behavior 148 (2023) 107877

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Early prediction of Autism Spectrum Disorders through interaction analysis in home videos and explainable artificial intelligence

Claudio Paolucci [a,**,1], Federica Giorgini [b,1], Riccardo Scheda [b,1], Flavio Valerio Alessi [a,1], Stefano Diciotti [b,c,*,1]

[a] Department of Philosophy and Communication Studies, University of Bologna, Bologna, Italy
[b] Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi", University of Bologna, Cesena, Italy
[c] Alma Mater Research Institute for Human-Centered Artificial Intelligence, University of Bologna, Bologna, Italy

ARTICLE INFO

Handling editor: Paul Kirschner

Keywords:
Attunement
Autism spectrum disorder
Early diagnosis
Explainable artificial intelligence
Interaction analysis
Semiotics

ABSTRACT

There is considerable discussion about the advantages and disadvantages of early ASD diagnosis. However, the development of easily understandable and administrable tools for teachers or caregivers in order to identify potentially alarming behaviours (red flags) is usually considered valuable even by scholars who are concerned with very early diagnosis. This study proposes an AI *pre-screening* tool with the aim of creating an easily administrable tool for non-competent observers useful to identify potentially alarming signs in pre-verbal interactions. The use of these features is evaluated using an explainable artificial intelligence algorithm to assess which of the proposed new interaction characteristics were more effective in classifying individuals with ASD vs. controls. We used a rating scale with three core sections – sensorimotor, behavioural, and emotional – each further divided into four items. By seeing home videos of children doing everyday activities, two experienced observers rated each of these items from 1 (highly typical interaction) to 8 (extremely atypical interaction). Then, a machine learning model based on XGBoost was developed for identifying ASD children. The classification obtained was interpreted through the use of SHAP explanations, obtaining an area under the receiver operating curve of 0.938 and 0.914 for the two observers, respectively. These results demonstrated the significance of early detection of body-related sensorimotor features.

1. Introduction

The prevalence of Autism Spectrum Disorders (ASD) in Europe is 12.2 per 1000 (one in 89) children (Salari et al., 2022). Many scholars have claimed that early diagnosis of ASD has proven to be crucial in achieving effective treatment (Gabbay-Dizdar et al., 2022), thereby improving the lives of ASD infants and their families (Elder et al., 2017; Franz & Dawson, 2019; Rotholz et al., 2017; van 't Hof et al., 2021; Volkmar, 2014). However, positions concerned by very early diagnosis, especially if observers are primed with specific confirmation factors (McCarty & Frye, 2020), have also grown in recent years. While remaining neutral in this important discussion, this work seeks to develop easily understandable and administrable tools for teachers, parents or caregivers, in order to identify potentially alarming behaviours (red flags), which is usually considered a very valuable achievement also by scholars who are concerned by very early diagnosis (Daniels & Mandell, 2014). Indeed, clear signs of impairments and atypicalities that can lead to ASD can be seen and read by looking at embodied and prelinguistic interactions between infants and caregivers when the toddler is between 9 and 18 months old (secondary

* Corresponding author. Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi", University of Bologna, Via dell'Università 50, 47521, Cesena, Italy.
** Corresponding author. Department of Philosophy and Communication Studies, University of Bologna, via Azzo Gardino 23, 40122, Bologna, Italy.
E-mail addresses: c.paolucci@unibo.it (C. Paolucci), federica.giorgini4@studio.unibo.it (F. Giorgini), riccardo.scheda2@unibo.it (R. Scheda), flaviovalerio.aless2@unibo.it (F.V. Alessi), stefano.diciotti@unibo.it (S. Diciotti).
[1] Stefano Diciotti and Claudio Paolucci contributed to the study conception and design. Data collection was performed by Claudio Paolucci and Flavio Valerio Alessi, material preparation and analysis were performed by Federica Giorgini and Riccardo Scheda. The first draft of the manuscript was written by all authors and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.The questionnaire and methodology for this study were approved by the Human Research Ethics committee of Alma Mater Studiorum-University of Bologna (Ethics approval number: 0173489).

https://doi.org/10.1016/j.chb.2023.107877
Received 23 January 2023; Received in revised form 15 May 2023; Accepted 10 July 2023
Available online 12 July 2023

intersubjectivity) (Gallagher & Hutto, 2008; Paolucci, 2020; Trevarthen & Hubley, 1978). Indeed, as several retrospective studies (Alonim et al., 2021) and reviews show (Zwaigenbaum et al., 2013), early symptoms and behaviours related to ASD can be seen long before the infant enters the linguistic phase.

It is difficult to provide a diagnosis before the age of two years and a half, which usually follows the observation of a linguistic skills development delay in children (van 't Hof et al., 2021).Worried by this and other issues, parents get concerned and seek assistance: later, they usually realize that they may have already noticed a number of other signals of non-typical behaviour. Furthermore, due to the difficulty in distinguishing non-adaptive behaviours such as a tendency toward isolation from shyness, and – more important – due to the lack of time spent together with the infant, also pediatricians sometimes fail to recognize warning signs (red flags) in children's early life (Elder et al., 2016).

Critical issues can also be found within screening tests such as the Autism Diagnostic Observation Schedule (ADOS) (Luyster et al., 2009), an observational screening test that composes the gold standard for ASD diagnoses with the Autism Diagnostic Interview – Revised (ADI-R) (Lord et al., 1994). These tests are typically used to screen the general community or a population that is already at risk, identify cases of ASD (sensitivity), and separate them from other conditions that have similar symptoms (specificity). As Lebersfeld and colleagues (Lebersfeld et al., 2021) showed in a recent systematic overview of studies administered using ADOS, this tool has an average sensitivity between 0.89 and 0.92 and an average specificity between 0.81 and 0.85. The test lasts from 40 to 60 min and consists of a series of highly structured activities, during which examiners elicit and assess the presence of specific behaviours that are natural to a neurotypical (NT) subject and that are usually lacking and/or deficient in ASD subjects. This test is administered in a controlled and laboratory setting, differently from tests such as the Modified Checklist for Autism in Toddlers (M-CHAT) (Robins et al., 2001), which is directly compiled by the caregivers of the examined subject, and do not require a specific medical competence.

Hence, ADOS mostly succeeds for two reasons: i) the observer is a highly competent subject, usually a neuropsychiatrist; ii) ADOS is a highly grammaticalized test carried out in a highly grammaticalized setting, i.e., a controlled laboratory situation and not "in the wild".

However, observers like caregivers, teachers or parents have difficulty identifying warning signs without any kind of expertise, even if they are the ones who spend most of the time with later ASD diagnosed infants. Furthermore, the dynamics of daily activities shared between caregivers and infants are not as structured and schematic as those that make up ADOS tests, with the consequence that even if caregivers had this kind of expertise, it would be difficult for them to apply it in daily, unplanned interactions. This implies that non-expert observers run the risk of not identifying potentially warning signs occurring from the earliest months of life. The work of our team tries to fill this gap, developing easily understandable and administrable tools for teachers, parents or caregivers, the non-competent observers who spend the majority of the time with future ASD children.

Summarizing, we can identify three main issues related to ADOS' administrational criteria and results:

i. ASD is usually diagnosed at 28 months (or later[2]), an age which has already surpassed the higher neuroplasticity window of 9–18 months, in which early interventions prove to be efficient on a cognitive, behavioural and emotional level (Franz & Dawson, 2019).
ii. A neuropsychiatrist is a competent observer. Basically, we can't use an ADOS test conducted by a neuropsychiatrist for every single baby.
iii. ADOS setting is not a real-life (in-the-wild test), where the infants interact with the people they usually interact with and do the things they usually do. Hence, we have an "ecological validity" issue (Lewkowicz, 2001) that has been discussed, also referring to the significant increase in ASD diagnoses in the last years.

In this paper, we propose a tool aiming to overcome these issues, with the ideal goal of creating a *pre-diagnostic* instrument for non-competent observers (teachers, caregivers, parents, etc.) useful to identify red flags in pre-verbal everyday interactions, as early as the 9–18-month age range (see (Paolucci, 2021b; 2021a)). Thus, the function of this tool is *not* to *diagnose* ASD – this is a task for neuropsychiatrists in ADOS contexts – but to build a reliable observation methodology in order to detect non-typical interactions, so that infants can be observed, monitored and, eventually, diagnosed at an earlier stage – a viable operation that can lead to important benefits. In other words, the aim is to build bridges between the pre-primary teachers, the families (who have babies), and the medical institutions (that have the competencies).

This contribution's tool was developed as part of the European Project NeMo ("NEw MOnitoring guidelines to develop innovative early childhood education and care (ECEC) teachers curricula", Grant Agreement: 2019-1-IT02-KA201-0633400)[3] aimed at increasing the quality of ECEC services, which are marred by the lack of uniform guidelines for the 0–6 age range, with the goal of ensuring greater inclusion of children with ASD in classrooms. The outcome of this project consists of a manual written by the University teams of the five European countries (Sweden, Spain, Italy, Slovenia, and Cyprus) involved in the project (*Nemoproject outputs* (n.d.)),[4] which addressees are pre-primary teachers, parents and caregivers, explaining the theoretical reasons and methodological criteria for understanding and administering the instrument.

Following the previous remarks on ADOS, our methodology and tool are based on three substitutions:

i. A substitution in the age of the infant, since it observes 9–18-month-old infants.
ii. A substitution of the neuropsychiatrist with a less competent observer, i.e., an Ordinary Observer (OA): a caregiver or a pre-primary teacher.
iii. A substitution of the laboratory and highly grammaticalized ADOS setting with a real-life, in the wild one, such as home videos shot through smartphones by caregivers.

The observer, after reading the manual – translated into five European languages (Italian, Spanish, Swedish, Slovenian, and Greek) – could detect if an interaction is typical or non-typical while watching homemade videos shot on smartphones by caregivers who interact with 9–18-month-old-infants (Fusaroli & Paolucci, 2011; Paolucci, 2012a, 2021a, 2021b; Luyster et al., 2009). To test the validity of our observation methodology, in this study, we make use of Explainable Artificial Intelligence (XAI) – a rapidly growing research field that refers to techniques to build human interpretable predictive models, i.e., algorithms that can explain the reasoning process and the specific piece of evidence that contributed to a decision (Jordan & Mitchell, 2015). Specifically, we sought to predict ASD diagnosis through interaction features extracted according to our methodology by experienced observers watching children's videos during their daily activities. We gave XAI interpretability through SHapley Additive exPlanations (SHAP) – a game-theoretic approach (Lundberg & Lee, 2017) to explain the output

[2] Indeed, ADOS Toddler Module is usually administered from 12 months on.

[3] https://site.unibo.it/nemoproject/en.

[4] Link to the page where you can read and download both the manual and the digital version of the tool in the different languages in which it has been translated https://site.unibo.it/nemoproject/en/outputs

of any machine learning model – to evaluate which of the newly proposed interaction features were more effective in classifying individuals with ASD compared to controls .

2. Materials and methods

2.1. Theoretical background and feature extraction through attuning assessment

Neurodevelopmental and developmental psychology (Onishi & Baillargeon, 2005; Reddy, 2003; Trevarthen, 1999, 2002) show how, since the first weeks of life, infants are able to adapt to other people's movements, recognize and express emotions and behaviours appropriately, and express expectations toward them. These perspectives are also shared by the embodied cognitive sciences (see Gallagher, 2017) which, contrarily to the classic view of the Theory of Mind model (Baron-Cohen, 1995), do not conceive social cognition as an inferential and disembodied process, instead considering it as an embodied and intersubjective phenomenon. This viewpoint places a strong emphasis on attunement, which is the capacity to modify one's sensorimotor, behavioural, and emotional expressions in response to how the other moves, acts, or feels (Gallagher, 2017; Stern, 1985). Thus, attunement is an intersubjective competence characterizing interactions that expresses the infants' desire of partaking in shared, affective dynamics and their desire of being recognized socially and emotionally. It is somehow comparable to a "love story", since we all love those who take into account how we move, the way we behave and feel, while distancing ourselves from those who do not.

According to DSM-5 (American Psychiatric Association, 2013), ASD can be seen as a disorder which features are expressed through persistent deficits in emotional, behavioural and social communication and interaction, and restricted, repetitive patterns of movement, behaviour, interests and activities. These traits become apparent in the early months of life through the lacking interactions parents and other caregivers try to have with their babies, even before linguistic impairments in the infants' development force parents to seek assistance. Indeed, ASD children compared with NT children show attuning problems such as a tendency not to be able to recognize, adjust, anticipate and respond appropriately to the movements of others (Gallese et al., 2013; Trevarthen & Delafield-Butt, 2013), to appropriately look for and respond to others' emotions (Hobson, 2014) and behaviours (Trevarthen, 2002), despite these features being intrinsically salient due to their pragmatic and affective value (Rochat et al., 2013). Neuroscientific studies show how ASDs present anomalies in the superior temporal sulcus (STS), an area highly active during social interactions and related to social rewards (Zilbovicius et al., 2006). Furthermore, ASDs present abnormalities in the mirror neuron system, which is active during observation, imagination and anticipation of others' gestures. (Gallese et al., 2013). Thus, ASDs seem[5] not to recognize social stimuli (such as emotions) as inherently salient, an impairment that manifests itself in their lack of competence in anticipating and coordinating to the gestures and expressions of others. That is, infants with ASD seem to be uninterested in being the object of other people's attention (Reddy, 2003), a crucial feature emerging from the affective and bodily proto-conversations (0–9 months) (Trevarthen, 1979) and exchanges of vocal expressions, eye contacts and vocalizations (Stern, 1985). The competence of shared action and attention emerges from this embodied ability to consistently anticipate and respond to the actions and expressions of others. In shared action and attention, crucial milestones for social, emotional and cognitive development emerging during secondary intersubjectivity (9–18 months) (Trevarthen & Hubley, 1978), infants direct others' attention towards stimuli creating a shared interactional scenario, engaging in playful sessions. This feature seems lacking in ASDs' interactions (Reddy, 2003).

Our approach interprets these results using the semiotic tradition's viewpoints on cognition (Fusaroli & Paolucci, 2011) and interactions (Lorusso et al., 2012), focusing on the semiotic account of social cognition and ASD provided by various studies (Paolucci, 2019; 2021a; 2021b). Indeed, Narrative Practice Semiotic Hypothesis (NPSH) (Paolucci, 2019; 2021a) offers a theoretical framework to explain infants' social cognition deficits characterizing ASD within an interactional perspective. According to semiotics (see Paolucci, 2012b, 2019, 2020, 2021a; Greimas, 1970, 1983; Lorusso et al., 2012), an inter-action whatsoever (a "doing together") has always a regular structure developing in four steps: (1) a "contract" phase (something worth acting for), (2) a "competence" phase (the "knowing-how"), (3) a "performance" phase ("the doing") and (4) a "sanction" phase (the acknowledgement). For instance, when a caregiver feeds his/her baby, there is something worth acting for, there is a mutual attunement in order to know how to do that together, there is a performance and a mutual judgement (smiles, gazes, mutual posture etc.) according to this performance. In another situation, a caregiver can ask the infant to look for his hidden favorite toy (contract), the infant accepts this role and starts exploring the environment in order to build the competence needed to find the toy (competence), looks for the toy (performance) and gets feedbacks on his/her actions from the caregiver (sanction). This is why, in our observation methodology, video-recordings have been segmented looking for those four meaningful moments and analyzed accordingly. Indeed, ASDs' apparent lack of interest towards socio-emotional stimuli can be read as a sanction problem, expressed through their tendency to set different systems of values, since ASD infants do not *seem* to care to be sanctioned by the others in a positive way and they usually carry on in their own business. For instance, in "meal" situations, they almost did not *seem* to perceive that the caregiver was getting tired, angry or disappointed because of their behaviour. Of course, we are not telling that they do not care, we are simply telling that they do not *seem* to care from the point of view of a typically developing adult caregiver or researcher observing the interaction (see Jaswal & Akhtar, 2018).[6]

If "culture" is grounded on the idea that 'others' have their own social logics, ASD persons look like people from a different culture interacting inside a culture that does not seem to be their own one, but it is. And not only they look so, but they *feel* so, experimenting the very same difficulties, obstacles and troubles of who is a "stranger" in his own place. Their great attention to detail, their preference for repetition and sameness, their restricted interests, their love for inanimate objects and for stable regularities are inherently meaningful signs for people with ASD, and not just, as they have often been conceived, inappropriate behaviours to be treated away. It is the way ASD people sense-make, trying to build a meaningful world they can handle. Indeed, Autism Spectrum Disorders are characterized by different ways of perceiving and moving, as well as particular emotional-affective aspects. Evidence ranges from hypo- and hyper-sensitivities, over difficulties with the timing, coordination, and integration of movement and perception, painfulness of certain stimuli, muscle tone differences, rigid posture, movement, attention, and saliency problems. All this leads to differences in bodily, affective and emotional coordination and attuning to others during social interactions, if compared to typically developing people (De Jaegher, 2013). Consequently, ASDs and caregivers' interactions do not follow a shared pattern of actions: they do not seem to share the

[5] We highlight the verb seem, for we do not assume, as the Social Motivation Theory (Chevallier et al., 2012) states, that, due to neurocerebral anomalies, ASDs are not interested in social interactions, stimuli and sanctions. Rather, we state that ASDs seem to be driven by a different way of perceiving and enacting sociality, as confirmed by the fact that musical therapies, based on activities aimed at strengthening social cooperation and coordination through rhythmic dynamics, can enhance ASD infants' cognitive, sensorimotor, social and emotional competence (Trevarthen, 2002).

[6] We underline "seem" here, for the very same reasons stressed in note 5.

same system of values through which coordinate to, anticipate and coherently respond to one another's actions in view of the occurring goal, showing a clear disattunement in the "doing together".

During an interaction, many different dimensions are intertwined and co-dependent in the dynamics of exchanges between people. Starting from the two main categories provided by DSM-5 to identify ASD – that is social affect and restricted, repetitive patterns of movement, behaviour, interests and activities (American Psychiatric Association, 2013)– and building on previous remarks about sensorimotor, behavioural and emotional abnormalities observed in ASD children, we sought to reduce the plethora of items on the most frequently used ASD screening tests – such as ADOS (Luyster et al., 2009), ADI-R (Lord et al., 1994), M-CHAT (Robins et al., 2001), Infant-Toddler Checklist (ITC) (Wetherby et al., 2004, 2008), Checklist for Early Signs of Developmental Disorders (CESDD) (Dereu et al., 2010) – focusing on the key signs which could prevent the development of a typical interaction.

In particular, we tried to analytically distinguish the main areas involved in a caregiver-infant interaction through a simplified system in which the observer only needs to look at the attunement between the child and the caregiver during their interaction. Thus, our idea of attunement upon which the domains structuring our methodology (see Sensorimotor dimension: A – the bodies; Behavioural dimension: B – the doing; Emotional dimension: C – the feeling) is not aimed at analysing or considering infants' specific cognitive competences and skills. Instead, following a semiotic perspective, our methodology focuses on their general capacity and willingness to manage the meaning production and recognition through the various phases of the practices in which they are involved (e.g. Does the infant respond to his/her caregivers' movements, invitations and expressions? If so, does s/he do that appropriately, considering the ongoing interactional dynamics?). Thus, our methodology aims at analysing and evaluating infants' competence to manage sense-making processes through interaction. In this vein, our methodology is consistent with the methodological criteria and aims of ADOS-2, the goal of its activities being "not to test specific cognitive abilities or other skills, but to present tasks that are sufficiently intriguing so that the child or adult being assessed will want to participate in social interchanges" (Lord et al., 2012).

As our addressee is an ordinary observer, in order to reduce the heterogeneity and complexity of the searchable signs identified in the diagnostic screening tests (see *Introduction*), we have operated with a view to simplification. Simplification means that all signs must be *summed up in a small number of things to look for* that a caregiver can easily evaluate. The hardest part of this work has been removing all of the semiotic technicalities that have been used in order to accomplish that and ending up with something that can be told in a very simple and clear way: *if the infant attunes to the caregiver, he/she is essentially a typical-developing infant; if not, the child needs to be monitored*, as children who do not tune in to their caregivers during interactions often receive a diagnosis of ASD or other neurodevelopmental disorders later. The main aspect that makes the system very simple is that one only needs to observe the attunement between the infant and the caregiver during their interaction.

So, what can be attuned in an interaction? We have identified three main dimensions of attunement: A) the bodies; B) the doing; C) the feelings, a *sensorimotor*, a *behavioural,* and an *emotional* dimension (See Table 1).

Table 1
Description of the features.

SENSORIMOTOR DIMENSION	BEHAVIOURAL DIMENSION	EMOTIONAL DIMENSION
A-The bodies	B -The doing	C-The feelings
A1-The space	**B1**-The doing together	**C1**-The feeling together
A2-The body of the other (bodily attunement)	**B2**-The mutual gaze while doing together	**C2**-The emotional gaze
A3-The infant's own body	**B3**-Joint attention	**C3**-The facial expression
A4-Degree of attention to the motor sanction of the caregiver	**B4**-Degree of attention to the behavioural sanction of the caregiver	**C4**-Degree of attention to the emotional sanction of the caregiver

2.1.1. Sensorimotor dimension: A - The bodies

The four items structuring this first dimension aim at analysing infant-caregiver interaction on its sensorimotor features. Indeed, infants later diagnosed with ASD present a lack of motor control i.e. they seem not able to coordinate and balance the movements between limbs, trunk and head (Teitelbaum et al., 1998). This first dimension presents four signs useful to identify sensorimotor anomalies, which prevent the development of an attuned dynamic of interaction. As far as the bodies are concerned, *a typical, attuned interaction is like a good dance*. What do we usually do when we dance? In dancing, your body attunes to the body of the other in a harmonious way, and your body adapts to what the other is doing. The contrary is also very easy to understand: when there is disharmony between two people – for instance when lovers are angry after struggling – each body moves with its own separate instructions. For instance, i) she is on one side of the couch with the telephone, ii) he is on the other side with the remote control (or vice versa). If during a doing together, the body of the infant seems to move with its own instructions without attuning to the body of the other, the interaction may not be typical, and this could be a sign of possible future ASD impairment. Of course, this misattunement can have many causes, but in our study, we look for neither causes nor motivations, only for meaningful signs used to detect if something is potentially worrisome. Thus, it is important to stress that the way bodies behave during an interaction is extremely revelatory and questions our ordinary distinction between the body and the mind, and the correlated idea that ASD involves mainly mindreading and communication problems (Paolucci, 2019, 2020). ASD also involves the bodies, and the way bodies behave is an extremely meaningful dimension for semioticians to fathom (Fontanille, 2004; Pennisi, 2021). In order to maximize its revelatory power, as far as the bodies are concerned, our methodology looks for four different things:

A1 - The space. This category considers the distance between subjects, the moving towards/away from each other, the way the child moves in space, and measures the typicality with which the infant approaches – or moves away from – caregivers or other infants. Research has noted that ASD children often interact differently with others regarding personal space than children with typical development. Children with ASD have a tendency to stay too close to others or maintain an excessive distance. In addition, ASD children may not seem to notice the presence of others nearby and may also seem to resist physical proximity or being touched actively. Furthermore, children with ASD usually prefer to interact with toys or other objects, as opposed to people, and thus inhabit space accordingly (see (Gessaroli et al., 2013)).

A2 - The body of the other (Bodily Attunement). This category assesses the extent to which the child seems to adapt his/her body to caregivers or other children during physical encounters. For example, the infant may turn his/her whole body toward the direction indicated by the caregiver or may physically react to the caregiver's voice calling him/her (ASD infants sometimes seem deaf and do not respond to their name). In general, infants with typical development (TD) adapt their bodies to the caregiver's movements as in a dance, in which one's body attunes to the movements of the other body. In contrast, ASD infants generally fail to adapt their posture and body movements to align with another person's movements (Trevarthen & Delafield-Butt, 2013). ASD children often interact in a way that appears rigid, controlled, inattentive, and inflexible (Teitelbaum et al., 1998). This reaction could be a potential warning sign; in fact, the child fails to adapt his/her posture and body movements to align with the other person's movements if s/he seems more interested in coordinating his/her movements to play (usually alone) with his/her toys, and/or if s/he avoid the other's persons attempts to engage with him/her in a reciprocal fashion of bodily dynamics. These “unattuned” bodily interactions may appear as if the infant resists physical interaction or is anxious or unsure of his/her role in the situation.

A3 - The infant’s own body. This category measures the child’s overall body posture style and movement style, even in non-interactive situations. Indeed, infants are usually able to coordinate their motor movements and posture, balancing head, trunk, hands, arms, and legs movements to start or continue any type of activity (e.g., spreading their arms as they crawl to reach for a toy or another body). In contrast, since the first weeks after birth infants later diagnosed with ASD show sensorimotor deficiences, which are reflected in poor motor control and intersubjective coordination. (Posar & Visconti, 2022). A potential warning sign may be present if the child makes repetitive body movements, which often take the form of hand-waving/flapping, rubbing, rocking, or pacing. These movements are known as “stimming” (self--stimulating). Furthermore, ASD children often assume a stiff and rigid posture, sometimes even while engaged in stimming behaviours. Motor disorders such as difficulty in crawling, lack of integration between the upper body (which is generally looser) and the lower body (generally stiffer) may also be present. ASD infants may also show weaker muscular tone. The presence of some of these signs is evaluated by giving them a high number on the scale (see The Tool and the Rating Procedure).

A4 - Degree of attention to the motor sanction of the caregiver. Sanction should be interpreted as any kind of evaluation (positive or negative) produced by the caregiver with words, actions, sounds and gestures that are used to reinforce the infant’s actions and reactions. A sanction is usually used to motivate, stimulate and help the infant to direct his/her performance. Regarding “the bodies” this category measures the child’s attention and anticipation of the caregivers’ body movements. Particular attention should be paid to the ‘end’ of an action or when an action requires a *specific* reaction from the infant. This is most frequently observed when (but not limited to) the infant prepares his/her own body to be picked up or hugged by the caregiver. Unlike more general and open-ended behaviours that are measured in A2, a sanction requires a specific bodily reaction from the infant. Whereas TD infants naturally observe their caregivers’ movements and adapt their posture and body movements to prepare themselves, a potential warning sign could occur if the infants show a markedly reduced ability to react appropriately to the actions of others. Furthermore, TD infants often mimic caregivers’ bodily actions even outside of strictly interactive contexts, whereas these mimic behaviours are often absent or significantly reduced in ASD infants.

2.1.2. Behavioural dimension: B - The doing

In terms of doing, this is probably the most obvious and easy-to-see dimension of the system. ASD infants usually mind their own business and do not seem they care about interacting with caregivers. Of course, we are not saying they don’t care – they probably do – but they do not seem to, perhaps because it is difficult for them to interact properly, so they prefer to drop (Paolucci, 2021b). When you are not good at something you usually do not want to do it because it reminds you your own inadequacy. The “doing” dimension is also divided into four items:

B1 – The doing together. This category measures the child’s degree of spontaneous participation in successfully shared activities. To note a potentially alarming case, it is important to assess whether the child is able to participate fluently in activities that are not strictly planned and/or structured. Thus, this criterion measures the child’s performance in interactive contexts, such as the ability to play his or her role within a game or interactive task, paying particular attention to situations in which the task/game suddenly changes or a new element is introduced to which the child must fluently adapt. As we have seen (see Introduction), this way of conceiving the behavioural competence of the infant is at the core of ADOS’ activities and evaluation criteria.

B2 – The mutual gaze while doing together. This category measures the frequency and style with which the child makes eye contact with his/her caregiver or another infant during a shared activity. These are natural behaviours used by infants for communicative and pragmatic purposes. If the infant avoids to seek out and/or respond to the other’s gaze repeatedly, infrequently, or only a few times, this can be referred to as a potential alarm situation (Reddy, 2003). For instance, reduced or absent eye contact and a noticeable lack of attention to other people’s faces are both indicators of possible alarm situations. Thus, if during a game or cooperative task, the child pays significantly more attention to objects than to other people and/or seems to avoid eye contact, this could be an indication of a possible ASD disorder.

B3 – Joint attention. This category measures the extent to which the attention of the child and caregiver seem to “synchronize” with each other during a shared game or task. For example, if the caregiver solicits the child’s attention to focus on a toy in order to participate in a shared activity, the child will look at the toy and is likely to invite the caregiver to play together. Conversely, a potential warning sign might be present if the child shows difficulty in attuning his or her attention to that of the caregiver and/or communicates less frequently with others, both verbally and nonverbally, during shared tasks and play. In such cases, infants look like as if they are “in their own world,” which may manifest in a marked reduction in communicative, exploratory, and eye-gazing behaviours. It is important to pay attention to both the rise and fall of something new, the infant’s exploratory behaviour and his or her communication with the caregiver.

B4 – Degree of attention to the behavioural sanction of the caregiver. This category measures the degree to which the infant perceives, is aware of, and can react appropriately to context-relevant actions and/or gestures made by the caregiver. Similarly to A4, we are looking at how well the infant pays attention to the caregiver’s behaviour in relation to engagement in shared games and tasks and how much attention he or she pays to the caregiver’s reactions and evaluations of his or her own behaviours. For example, during a meal, if the infant refuses to eat and the caregiver keeps insisting that he or she eat, how much does that affect the infant’s behaviour? A potentially alarming situation could occur if the child does not understand the overall meaning of an action, game, or task, if he or she does not understand unspoken instructions or other subtleties present in interactions, and/or if he or she does not achieve the desired outcome after a series of encouragements, instructions, and motivations. In addition, children with ASD or developmental disorders are often less responsive to gestures that make other children feel good and aid interaction, such as positive words and gestures (e.g., pointing, thumbs up, or pats on the back), which can hinder their ability to learn and form social bonds (Reddy, 2003).

2.1.3. Emotional dimension: C - The feelings

As for emotions, in an ordinary interaction, behaviours and feelings change according to the change in others’ emotions. If someone gets angry, a person interacting with him/her questions the anger and maybe changes his/her behaviour and mood accordingly. ASD children usually do not question this type of behaviour or have difficulty doing so (Hobson, 2014). We are not saying that children should be happy when the caregiver is happy or sad when he is sad. This is not attunement at all. Attunement does not mean feeling the same emotion: it is neither empathy nor emotional contagion. It simply means taking the other person’s emotions into consideration. For example, imagine a TD child who notices that his/her parent is angry and decides to attune to him/her by continuing to disobey because s/he wants to. In this case, the infant takes his/her parent’s emotions into consideration and decides to attune in his/her own way according to his/her will. Instead, ASD children may simply disregard any change in the caregiver’s emotions. This can be quantified through four different elements that we need to evaluate.

C1 – The feeling together. This category measures how the infant and caregiver adjust their emotional states in response to each other (see Gallese et al., 2013; Stern, 1992). Pay close attention to whether or not the infant becomes happy when the caregiver is happy and/or is able to adjust this happiness if the caregiver subsequently shows subtle signs of disappointment. Indeed, usually, both the infant and caregiver continually adapt and adjust their emotional states in response to those shown by the other in a spontaneous, fluid, and dynamic way. It is necessary to keep in mind that the caregiver who uses so-called "infant talk" often has the power of grasping the infant's attention, so do not overestimate the infant's capacity for emotional regulation if the caregiver introduces a sudden change from "normal-talk" to "infant-talk" that captures attention.

C2 – The emotional gaze. This category measures the frequency with which children and caregivers make eye contact outside of task-related contexts. Children frequently and spontaneously make eye contact with caregivers or other children and adults, even outside of play- and task-related situations, in a way that seems natural and spontaneous. This eye contact usually has a communicative function and contributes to the overall quality of the interaction (Trevarthen, 1979). A potentially alarming situation could occur if the child seems disinterested in meeting another person's gaze or communicating through eye contact, and/or even seems to avoid it. Keep in mind that if you are evaluating a video recorded by a human being who does not appear in the recording, often the infant will appear to look directly at or just above the camera if he or she makes eye contact with the person recording the video.

C3 – The facial expressions. This category measures the extent to which the child imitates or reacts spontaneously to caregivers' facial expressions. Instead of observing a general emotional state as in C1, in this category we pay more attention to how the child's facial expression (e.g., smile, laughter, frown, surprise) corresponds to that expressed by the caregiver, as well as how the child's own expression changes in response to that expressed by the caregiver (e.g., does the child become sad if the caregiver suddenly appears sorry?). Children seem naturally and spontaneously predisposed to reflect the emotional expressions of their caregiver. In contrast, a potentially alarming situation could arise if infants were more prone to remain unaware of the meaning behind the caregiver's facial expression and how they should react emotionally in response to it (Hobson, 1986).

C4 – Degree of attention to the emotional sanction of the caregiver. This category measures how attentive the child generally seems to be to the caregiver's emotional "requests." An alarming situation would occur if the infants were inclined to show less interest in the emotional states of others and thus not respond to prompts or feel emotions when the caregiver would like them to. For example, when interacting with ASD children, the caregiver may appear to be continually trying to elicit emotional states in the child that are not being met (e.g., trying to make the child feel arousal), or the caregiver may feel visible frustration when the child does not respond contextually to his/her emotional state (e.g., anger at the child's misbehaviour) and the child may continue to go about his or her own needs.

2.2. The tool and the rating procedure

Based on these domains articulating the interactions between infants and caregivers, we built a tool, which items and rating system aim at individuating signs of potential concern. Indeed, one of the main problems in ASD screening tests concern in its questions and rating criteria. For instance, the questionnaire-based M-CHAT (Robins et al., 2001), despite being widely used and efficient for the diagnosis of ASD even at an early stage (see (Kamio et al., 2014)), has several problems related to the polar structure of the answers (yes/no), which implies that the signs and anomalies of ASD may be only absent or present, while instead occurring more or less frequently and/or appropriately (Barbaro & Dissanayake, 2013). Thus, in order to bring forth the ecological approach implied by our observational methodology, which emphasizes the prominence of the occurring situation and contextual variabilities of the behaviours, we associated each item of the three dimensions to a rating scale from 1 to 8.

In particular, as detailed in the manual using video examples, 1 stands for a very typical interaction (high level of attunement) while 8 stands for a very atypical interaction (low level of attunement). Depending on the severity of the condition and the number of anomalies detected, we divide each field into a range of possible concerns – where 1–2 means no concern, 3–4 means light concern, 5–6 means mild concern, and 7–8 means severe concern. This approach has the major advantage of sterilizing the observers' emotions and point of view. Indeed, the observers do not make any diagnosis: they simply evaluate a behaviour. The observers also have the possibility not to rate one (or more) of the features if the home video they are watching is not explicit on that particular aspect (too short, unclear etc.) or if they feel unsure or incapable of doing so. However, in the first case, they are invited to insert a "not readable" value in the sheet, while, in the second case, they are encouraged to get acquainted with the videos and with the system, trying to read the manual again and train themselfs with new material. It will be the system that will later tell whether the child recorded in that particular interaction is behaving typically or not (see Results).

The training of the observers was performed using hundreds of home videos of infants-caregivers interactions collected before this study began (Paolucci, 2021b) (see Videos and Participants). They were home videos of i) children who were later diagnosed with ASD when they were 9–18-months-old and ii) typically developing children who were used as a control group.

2.3. The pilot study

In this pilot study, we evaluated our methodology's potential to prove a high level of sensitivity and specificity to ASD indications and behaviours when applied by a competent observer. The two observers of this study are part of the working group created for NeMo Project, which have an expertise in ASD studies. They were asked to rate the videos (see Videos and Participants) using our methodology without knowing the condition of the infants. Individuating sensorimotor, behavioural and emotional impairments the observer would be able to understand that he/she is in front of a potential case of ASD. This contribution describes the results of our pilot study.

2.3.1. Videos and participants

We recruited a total of 32 children with a diagnosis of ASD (10 individuals, 2 females) and TD children (22 individuals, 10 females). At the recruitment stage, the children were between 18 months and 11 years old, while all the collected videos show them interacting with a caregivers during their 9–18 months period. All participants were recruited after the project dissemination through social media and word of mouth. The caregivers interested in participating have followed a standard procedure elaborated by the staff under the supervision of the University of Bologna data protection officer (DPO) and AIAS's DPO (Support the Project! (n.d.)).

ASD and TD children were sex-matched ($\chi^2 = 1.9$, p-value $= 0.16$). A total of 67 home videos were collected (28 for the ASD group; 39 for the TD group); for each child, we collected 1.86 ± 1.3 (mean $\pm$ standard deviation) videos for the ASD group and 1.77 ± 1.0 for the TD group. Each video was made by parents or caregivers when the children were aged 9–18 months. They were recorded using smartphone cameras during their daily activities to cover the different aspects related to the three dimensions of interaction. The only criterion given to the parents and caregivers for shooting or selecting the video to send us, as also stated in our manual (see Introduction), was that it had to represent a scenery in which the infant plays or interacts with his/her caregiver and/or other infants in a standard, daily situation (eating, bathing, playing, etc.), in which mutual gestures, eye-gazes, bodily movements, behaviours could be observed. Two experienced observers independently scored the different categories by watching these videos without knowing the children's actual diagnosis.

2.3.2. Ethics

The study was approved by the local Ethics Committee. Participants sent an e-mail expressing their willingness to participate, attaching a signed "Informative consent for data processing form" for each adult and minor in the video (no data concerning sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction were collected). To join the project each participant had to provide his/her consent to the processing of his/her personal data and his/her children's data. The participant could optionally provide his/her consent to the dissemination of the data, if these are found to be particularly significant for the aims of the project. Also, s/he could optionally provide his/her consent to the storing of the data after the conclusion of the project, for future reuse of the data in similar projects with the same aims and purposes. We created a private and protected cloud space for each participant and sent a link to the database where access was possible only through an automatically generated passcode for each participant. The videos were then associated with a pseudonym and analyzed by the other staff members. After the analyses, a researcher running the platform contacted participants asking about their children's condition, thus obtaining a self-reported diagnosis from the primary caregiver.

2.3.3. Measurement of the inter-observer agreement

We measured the inter-observer agreement of the features using the linearly weighted kappa statistic (Altman, 1999). We adopted the classes of interpretation of the kappa statistic proposed by Landis and Koch (Landis & Koch, 1977) for descriptive purposes. Accordingly, a kappa value below 0 indicates poor agreement, between 0 and 0.20 a slight agreement, between 0.21 and 0.40 a fair agreement, between 0.41 and 0.60 a moderate agreement, between 0.61 and 0.80 a substantial agreement and between 0.81 and 1 an almost perfect agreement (see Table 2).

2.3.4. ASD vs. TD classification through an XAI approach

We built a machine learning model for ASD/TD children classification based on the eXtreme Gradient Boosting (XGBoost) tree-based classifier – a scalable end-to-end tree boosting system that has been widely used to achieve cutting-edge performance on a variety of recent machine learning challenges (Chen & Guestrin, 2016). We fed the machine learning model by all attunement features and sex of children. We trained and tested the ASD vs. TD binary classification task using nested cross-validation (CV) (Mueller & Guido, 2017). Cross-validation is a machine learning technique which allows to have stable average performances in the case of small samples-sized datasets (Géron, 2017). Indeed, dividing a small dataset into a fixed training set and a fixed test set implies statistical uncertainty around the estimated average test error. Alternative procedures to standard train-test split method such as CV enable to use all of the samples of the dataset in the estimation of the mean test error (Goodfellow et al., 2016). These procedures are based on the idea of repeating the training and testing computation on different randomly chosen subsets or splits of the original dataset. In particular in this study we performed a nested CV loop with a stratified child-based group data splitting scheme to examine the unbiased generalization performance of the trained model and, at the same time, perform hyperparameters optimization (See Fig. 1) (Mueller & Guido, 2017).

In particular, this strategy involves nesting two k-fold CV loops where the inner loop is used for optimizing model hyperparameters, and the outer loop gives an unbiased estimate of the performance of the best model. The stratified child-based group data splitting scheme in the inner and outer CV prevents data leakage, i.e., that videos belonging to the same child may be included, at the same time, in the training/validation/test sets. Briefly, we start by splitting the dataset into k folds (outer CV); one fold is kept as a test set of the outer CV, while the other k-1 folds (the training set of the outer CV) are split into k inner folds (k-1 for training and the kth for validation). Specifically, we used a 5-fold nested CV because it offers a favorable bias-trade-off (Hastie et al., 2013; Lemm et al., 2011). The models' hyperparameters are chosen from the hyperparameter space through a grid search based on the average performance of the model over the validation sets of the inner folds. In particular, we varied the gamma hyperparameter of the XGBoost in the set {0, 1, 2}, maximum depth in {4, 5, 6, 7, 8}, minimum child weight in {1, 2, 3, 4, 5}, and maximum delta step in {1, 3, 5, 7}. Once the best combination of hyperparameters that maximized the area under the Receiver Operating Characteristic (ROC) curve in the validation sets of the inner CV has been found, the model with that combination of hyperparameters is re-trained on the outer training set and tested on the test set kept out from the outer CV. This procedure is repeated for each fold of the outer CV. Before training each XGBoost classifier, both in the inner and the outer CV, we firstly applied feature imputation, i.e., we replaced missing features with the average value of that feature in the set. Secondly, the set was standardized, i.e., each feature was rescaled to have zero mean and unit variance. For each iteration of the inner and outer CVs, these transformations were applied to the training, test and validation sets using Python scikit-learn transformers, thus not using test data in any way during the learning process, – preventing any form of peeking (Diciotti et al., 2013). Performance was quantified in terms of the area under the ROC curve (AUC) in the test sets of the outer CV. The point of the ROC curve with minimum distance from the performance of the ideal observer (0,1) was also computed for both observers.

Since the performance may vary depending on how the data are split in each fold of the CV, we repeated the nested CV procedure ten times and took the average and standard deviation of the results from all repetitions to get a final model assessment score. Since we were also interested in explaining the model predictions, we adopted SHapley

Table 2
Descriptive statistics of the features measured by both observers. [a] SD: standard deviation.

Feature	Observer #1		Observer #2		Weighted kappa statistic
	TD group [mean±SD[a] (% unassigned)]	ASD group [mean ± SD (% unassigned)]	TD group [mean ± SD (% unassigned)]	ASD group [mean ± SD (% unassigned)]	
A1	1.54 ± 1.2 (10.2%)	4.42 ± 2.3 (25.0%)	1.63 ± 1.2 (10.3%)	3.84 ± 2.6 (32.1%)	0.81
A2	1.54 ± 0.9 (5.1%)	5.64 ± 1.77 (21.4%)	1.63 ± 1.0 (5.1%)	4.91 ± 2.0 (21.4%)	0.85
A3	1.92 ± 1.4 (2.6%)	6.36 ± 1.7 (0.0%)	1.95 ± 1.4 (2.6%)	5.82 ± 2.1 (0.0%)	0.89
A4	2.03 ± 1.4 (12.8%)	6.34 ± 1.8 (17.9%)	2.06 ± 1.6 (12.8)	5.23 ± 2.2 (21.4%)	0.81
B1	1.54 ± 1.1 (5.1%)	4.96 ± 2.7 (14.3%)	1.54 ± 1.2 (5.1%)	5.21 ± 2.6 (14.3%)	0.95
B2	1.65 ± 1.3 (5.1%)	5.48 ± 2.36 (3.6%)	1.68 ± 1.5 (5.1%)	5.61 ± 2.4 (0.0%)	0.90
B3	1.76 ± 1.4 (5.1%)	5.20 ± 2.6 (10.7%)	1.68 ± 1.4 (5.1%)	5.10 ± 2.4 (10.7%)	0.94
B4	1.63 ± 1.4 (5.1%)	5.11 ± 2.6 (3.6%)	1.66 ± 1.5 (2.6%)	5.28 ± 2.6 (10.7%)	0.94
C1	1.51 ± 1.0 (10.3%)	6.12 ± 2.1 (7.1%)	1.34 ± 0.95 (10.3%)	6.08 ± 2.3 (7.1%)	0.91
C2	1.65 ± 1.6 (5.1%)	5.88 ± 2.3 (10.7%)	1.65 ± 1.6 (5.1%)	5.88 ± 2.2 (7.1%)	0.98
C3	1.53 ± 1.3 (7.7%)	6.00 ± 2.2 (7.1%)	1.50 ± 1.3 (12.8%)	6.23 ± 2.2 (7.1%)	0.94
C4	1.63 ± 1.3 (10.3%)	6.24 ± 2.2 (10.7%)	1.53 ± 1.1 (12.8%)	6.00 ± 2.2 (10.7%)	0.93

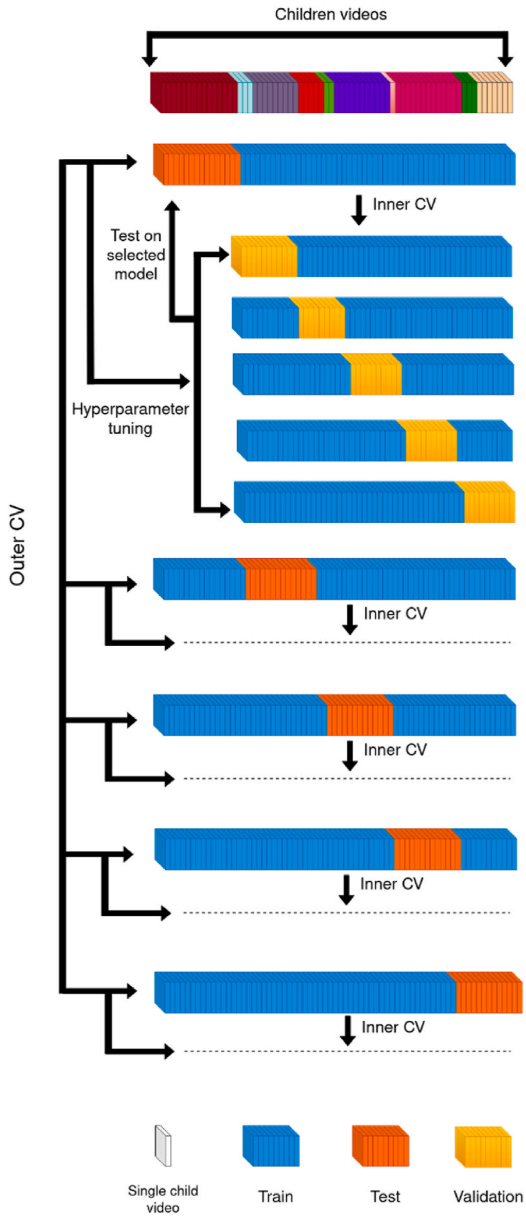

Fig. 1. Five-fold nested CV with a stratified child-based group data splitting scheme.

Additive exPlanations values (SHAP) (Lundberg & Lee, 2017) – a framework for interpreting predictions based on game theory by assigning to each feature an importance value for every sample. Accordingly, for each model of the outer CV, SHAP values were computed to produce an average (and standard deviation) of the feature importance explanation for the final model.

The training, validation, and test of the XAI models were carried out using a custom code in Python language (v. 3.8.8) using the following modules: matplotlib v.3.4.1, numpy v.1.21.3, pandas v.1.2.3, scikit-learn v.1.1.dev0 (Pedregosa et al., 2011), seaborn v.0.11.2, and xgboost v.1.4.2. The total computation time for the training, validation, and test was about 40 min on all cores of a Linux workstation equipped with a 4-core (4 threads) Intel i7-7500U CPU and 8 GB RAM.

3. Results

3.1. Measurement of the inter-observer agreement

The weighted kappa statistic of the inter-observer agreement was almost perfect, ranging between 0.82 and 0.96 (see Table 2).

3.2. ASD/TD classification through a machine learning approach

The model trained with features extracted by observers #1 and #2 reached an AUC of 0.938 and 0.914, respectively – thus indicating an excellent performance of the classification models (see Fig. 2). The point of the ROC curve with minimum distance from the performance of the ideal observer (0,1) corresponds to a sensitivity = 0.89, specificity = 0.86 for observer #1 and sensitivity = 0.85, specificity = 0.86 for observer #2 (see Fig. 2). In the global feature plot, each feature's global importance is assumed as the mean absolute SHAP value for that feature over all the given samples expressing the average impact on model output magnitude.

In Fig. 3, we showed the ranking of the feature importance, i.e., the SHAP strength (the absolute value of the SHAP values) for both observers.

As an example, in Fig. 4, we plotted the beeswarm plots for the models trained in the first repetition. These plots are designed to show an information-dense summary of how the principal features in the dataset impact the model's output. For each video, the given explanation is represented by a single dot on every single feature. The SHAP value of that feature defines the x position of the dot, and dots "pile-up" along each feature row to show density. Colors are used to display the original value of a feature. In other words, from these plots, we can see the value of SHAP for each sample and each feature. The SHAP value of a given sample (a dot point in the graph) takes on a different color (from blue to pink) whether it has a high or low value, while it takes a different position in the graph (from the base value of SHAP to the right or left), based on its impact on the decision of the model. Therefore, if we observe a value of SHAP colored in pink and located to the right, the feature has considerably affected the model's decision. In particular, for observer #1, high values of feature A2 move the model's output toward the ASD group, while for observer #2, a low value of feature C1 moves the model's output toward the TD group. For both observers, the model gave considerable importance to the sex of the children, and if the sex is male (coded as '1') the model's output moves toward the ASD group.

4. Discussion

In this study, we predicted ASD classification through interaction features extracted according to our observational methodology by experienced observers watching children's videos during their daily activities (Paolucci, 2021a). We found that the inter-observer agreement was almost perfect between the two observers. Then, we trained a machine learning model with features extracted by observers #1 and #2, reaching excellent AUC in discrimination ASD vs. TD children (Fig. 2). Then, we gave XAI interpretability through SHapley Additive exPlanations (SHAP) to evaluate which of the newly proposed features were effective in classifying individuals with ASD compared to healthy individuals. For observer #1, the features A3 and A2 that both concern "The bodies" and the sensorimotor dimension, were the most crucial in classifying ASD vs. TD subjects, while observer #2 individuated feature A3 as the third most salient. Maybe this "supremacy" and heuristics of the body – if compared to the doing and the emotions – in order to reveal ASD in prelinguistic infants during secondary intersubjectivity is worth further reflection.

As previously stated (see Introduction), the classical cognitive sciences thought of cognition as a representational mental faculty, in which perception used to be conceived as the input and action as the output. Indeed, according to this model, cognition was grounded on largely

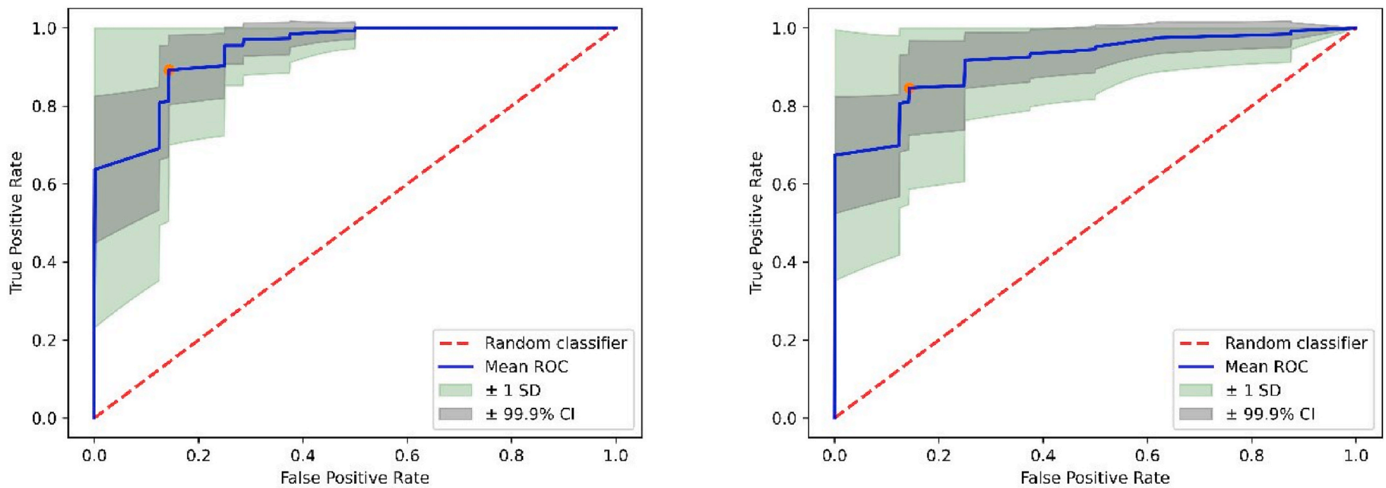

Fig. 2. ROC curves in the test set of the outer CV.

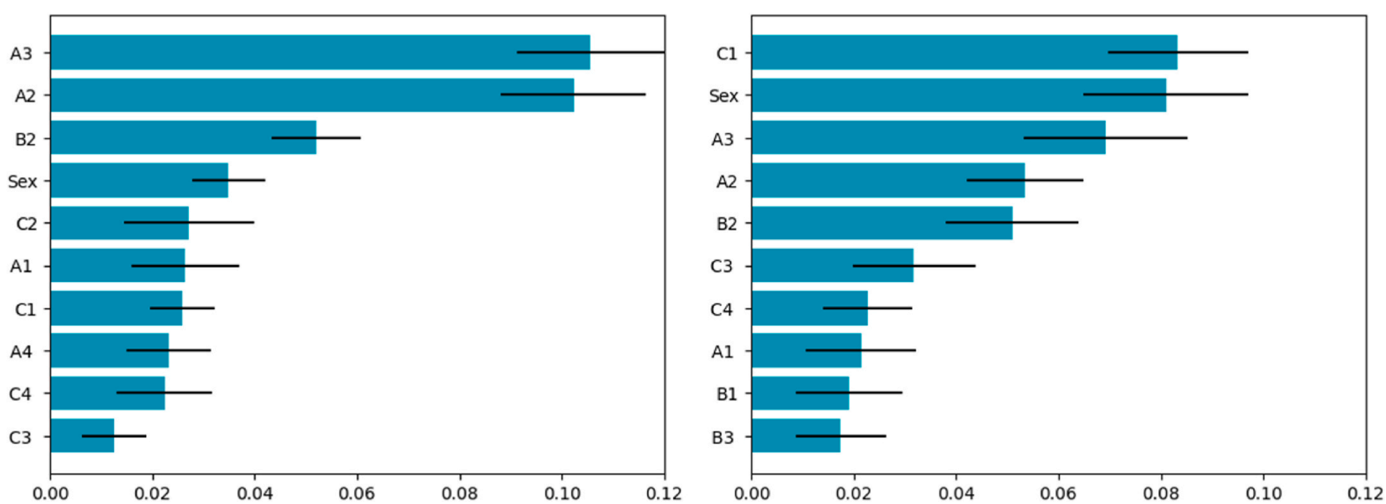

Fig. 3. Ranking of the SHAP strength.

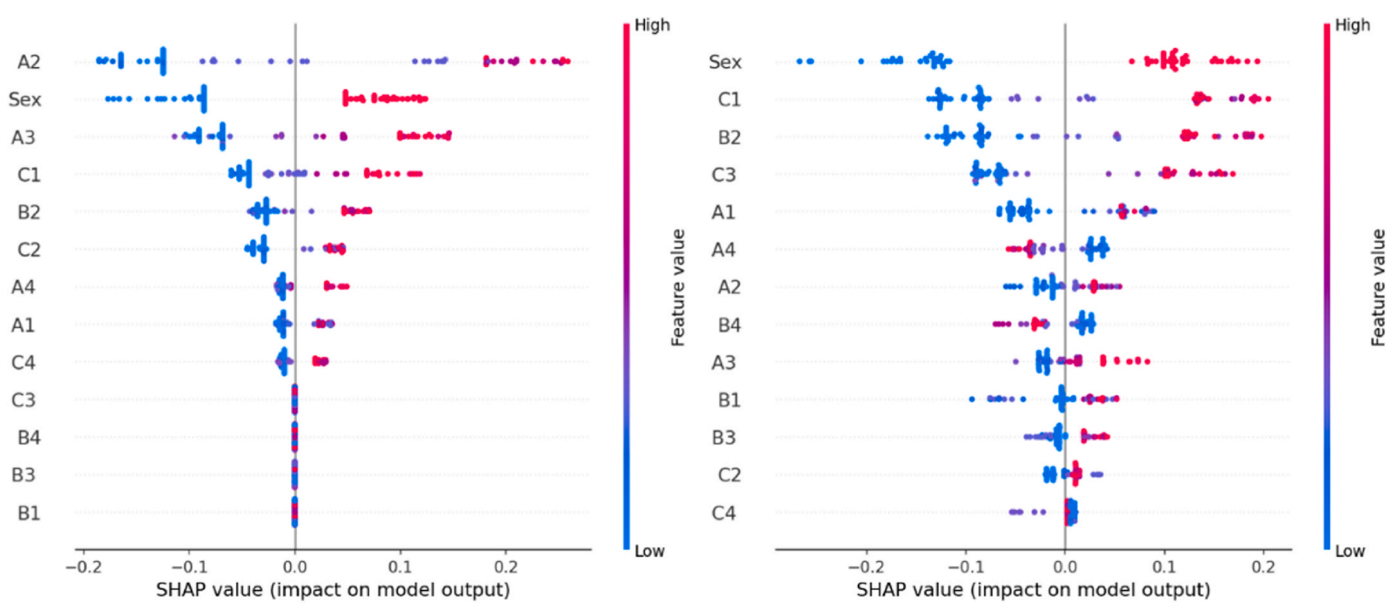

Fig. 4. SHAP beeswarm plot.

disembodied mental computations, in which the body certainly played a role, albeit a secondary one, if compared to the centrality of the mind and its meta-representational abilities. The classical way of thinking, diagnosing, and understanding autism spectrum disorders has been strongly influenced by this framework: ASD has been considered a Theory of Mind disorder. In other words, ASD was regarded as a disturbance of the metarepresentational capacity of the constitutive type of mind that enables us to make cognitive sense of the actions of others. Indeed, classical and impactful studies, such as those by Baron-Cohen (Baron-Cohen, 1995, 2000; Baron-Cohen et al., 1985) have linked the inability of ASD children to pass metatransactional tests of false belief well into adulthood (and beyond) precisely to a deficit in their social cognition skills. However, as cognitive sciences gradually abandoned the so-called "sandwich model", which identified cognition with the processing of mental representations while increasingly considering perception and action as constituent parts of cognition, the centrality of the theory of mind for ASD began to weaken. On the other hand, the problematic nature of this centrality fits perfectly with the following finding: typical-developing children themselves do not pass theory-of-mind tests until about age 3–4 years, whereas – as this study also shows – we are capable of discriminating between population groups with ASD and control groups much earlier with excellent performance. Thus, the AI analysis presented here shows that the body-related signs are the most effective in discriminating between ASD and the control group. In particular, attunement to the other's body (A2) and the signs related to owning one's body (A3) seem to be extremely revealing of a possible future typical autism spectrum disorder, even in very young children, and also in contrast to the other signs from the other two dimensions. This centrality of the body and, in particular, the centrality of attunement with the other's body also seems to be the least related to the cultural variability of interaction and individual enculturation practices that are often linked to social action during caregiving practices. For example, we know well how "looking into each other's eyes" is a typical mode of interaction in our societies (Ochs et al., 2004), but it is not as important, for example, in Chinese culture or some Latin American communities. It should be noted that the most important features are not exactly the same between the two observers, and in future studies that are already underway, we will use multiple observers to understand better whether these twelve features can be reduced without significantly losing performance acuity. However, the XAI model is clearly pointing us in the right direction: aiming for a simplification of the system that preserves its performance and robustness while at the same time indicating which of the core signs may better reveal a potential ASD case than others.

This was made possible by the SHAP approach, which belongs to game theory: features can be seen as players playing a cooperative game to have a specific prediction, and through the so-called Shapley values (Lundberg & Lee, 2017), we can compute the importance of each feature. SHAP method is better than other explanation methods (like LIME, DeepLIFT) because it satisfies three desirable properties for an explanation model (Lundberg & Lee, 2017): local accuracy, meaning that we can also have local explanations and not only global interpretations; missingness, so that a missing feature does not contribute to the explanation at all; and consistency, which says that if a model changes so that the marginal contribution of a feature value increases or stays the same (regardless of other features), the Shapley value also increases or stays the same.

However, this study has several limitations. The first concern relates to the observer. Two competent observers have led this pilot study, but the ideal aim of the methodology is to be efficient and apt for non-competent observers. For ethics and privacy reasons, we could not show videos to people that were not researchers of the working group of NeMo Project. Nonetheless, this studio shows how the methodology, if correctly used, can give fruitful results: this is why it needs to be accompanied by the results of other studies led in the NeMo Project, in which the observation methodology has been tested by pre-primary teachers directly in interaction with infants (see Implications). The second concern is related to the sample size, consisting of a relatively small number of children. Indeed, in order to generalize our results have to be tested with a larger amount of video analyses, a process which is currently underway. Thirdly, a disproportion between male and female children is denoted by a small number of females (i.e., 12 out of 32 total children). This condition is frequently encountered in the study of ASD and is partly due to genetic reasons and partly because the clinical scales used to date are mainly made on male subjects, thus resulting in females being underdiagnosed. Therefore, the results might be more calibrated to males than both sexes. Eventually, the most rated features are not completely consistent with one another. The common features statistically evaluated as salient by both observers are A2 and A3, respectively relating to infants' capacity to adapt his/her body to caregivers or other children during physical encounters (A2) and overall body posture style and movement style, even in non-interactive situations (A3). The prominence of bodily features seems to confirms Teitelbaum and colleagues' study (Teitelbaum et al., 1998), in which, through the analyses of recorded home videos, they identified anomalies in sensorimotor and bodily movements as red flags for ASD disorders since the very first months of life. However, despite the inconsistency in the overall balance between the most rated features, the results generated by our AI system prove to be highly sensitive and specific for individuating potential red flags. More observers will perform future analyses, in order to determine if it will be possible to individuate a more consistent number of features rated statistically, possibly to create an even simpler, yet as sensible and specific set of items.

In conclusion, our results suggest that examining videos of children engaged in their daily activities through an explainable machine learning algorithm allows the validation of the effectiveness of some of the constituent features of the proposed scale in classifying ASDs.

4.1. Implications

This pilot study suggests that, despite all of the aforementioned limitations, our methodology is sensitive and specific for identifying potential indicators of infant development impairment. The results of the cross-observation between observers and the machine learning system seem therefore promising. Of course, there is still a lot of work to be done: in the next studies, we will expand the corpus of analyzed videos and increase the number of observers, which is currently still insufficient to guarantee a generalizable validity of the study. At the same time, disseminating the methodology among pre-primary teachers through the manual and instructing them in the evaluation of interactions could bring new data to the research. Thanks to the NeMo project, we have already offered training courses to teachers in five European countries (Spain, Cyprus, Italy, Sweden and Slovenia) who will soon inform us of the results of their observations. A second relevant point is that the methodology shown could be particularly useful for refining a certain sensitivity to the fundamental dimensions of an interaction between children and caregivers. The goal should be to accustom observers to a methodology that, in the version we propose, currently appears promising in identifying red flags for future possible cases of neurodevelopmental disorders, since teachers and caregivers should be able to spot warning indicators. In this sense, the methodology's high identification rate makes it a useful tool whose margin of specificity and error is yet to be verified in future studies.

Credit author statement

Stefano Diciotti and Claudio Paolucci contributed to the study conception and design. Data collection was performed by Claudio Paolucci and Flavio Valerio Alessi, material preparation and analysis were performed by Federica Giorgini and Riccardo Scheda. The first draft of the manuscript was written by all authors and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

The research has been funded by the European Commission with the project NeMo: Project identifier: 2019-1-IT02-KA201-063340, Project coordinator: Alma Mater Studiorum-Università di Bologna, Via Zamboni 33, 40126 Bologna, IT.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

References

Alonim, H., Lieberman, I., Schayngesicht, G., & Tayar, D. (2021). *A retrospective study of prodromal variables associated with autism among a global group of infants during their first fifteen months of life* (Vol. 7).

Altman, D. G. (1999). *Practical statistics for medical research.* Chapman & Hall/CRC.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders DSM-5* (5th ed.). American Psychiatric Association Publishing.

Barbaro, J., & Dissanayake, C. (2013). Early markers of autism spectrum disorders in infants and toddlers prospectively identified in the Social Attention and Communication Study. *Autism: Int. J.Res.Pract., 17*(1), 64–86. https://doi.org/10.1177/1362361312442597

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind.* https://doi.org/10.7551/mitpress/4635.001.0001

Baron-Cohen, S. (2000). Theory of mind and autism: A review. In *International review of research in mental retardation* (Vol. 23, pp. 169–184). Academic Press. https://doi.org/10.1016/S0074-7750(00)80010-5.

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

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends Cognit. Sci., 16*(4), 231–239. https://doi.org/10.1016/j.tics.2012.02.007

Daniels, A. M., & Mandell, D. S. (2014). Explaining differences in age at autism spectrum disorder diagnosis: A critical review. *Autism: Int. J.Res.Pract., 18*(5), 583–597. https://doi.org/10.1177/1362361313480277

De Jaegher, H. (2013). Embodiment and sense-making in autism. *Front. Integr. Neurosci., 7*. https://www.frontiersin.org/articles/10.3389/fnint.2013.00015.

Dereu, M., Warreyn, P., Raymaekers, R., Meirsschaut, M., Pattyn, G., Schietecatte, I., & Roeyers, H. (2010). Screening for autism spectrum disorders in Flemish day-care centres with the checklist for early signs of developmental disorders. *J. Autism Dev. Disord., 40*(10), 1247–1258. https://doi.org/10.1007/s10803-010-0984-0

Diciotti, S., Ciulli, S., Mascalchi, M., Giannelli, M., & Toschi, N. (2013). The "peeking" effect in supervised feature selection on diffusion tensor imaging data. *Am. J. Neuroradiol., 34*(9). https://doi.org/10.3174/ajnr.A3685. E107–E107.

Elder, J. H., Brasher, S., & Alexander, B. (2016). Identifying the barriers to early diagnosis and treatment in underserved individuals with autism spectrum disorders (ASD) and their families: A qualitative study. *Issues Ment. Health Nurs., 37*(6), 412–420. https://doi.org/10.3109/01612840.2016.1153174

Elder, J. H., Kreider, C. M., Brasher, S. N., & Ansell, M. (2017). Clinical impact of early diagnosis of autism on the prognosis and parent–child relationships. *Psychol. Res. Behav. Management, 10*, 283–292. https://doi.org/10.2147/PRBM.S117499

Fontanille, J. (2004). *Figure del Corpo per una semiotica dell'impronta.* Meltemi.

Franz, L., & Dawson, G. (2019). Implementing early intervention for autism spectrum disorder: A global perspective. *Pediat. Med., 2*. https://doi.org/10.21037/pm.2019.07.09, 44–44.

Fusaroli, R., & Paolucci, C. (2011). *The external mind: An introduction* (Vol. 28). An Introduction.

Gabbay-Dizdar, N., Ilan, M., Meiri, G., Faroy, M., Michaelovski, A., Flusser, H., Menashe, I., Koller, J., Zachor, D. A., & Dinstein, I. (2022). Early diagnosis of autism in the community is associated with marked improvement in social symptoms within 1–2 years. *Autism, 26*(6), 1353–1363. https://doi.org/10.1177/13623613211049011

Gallagher, S. (2017). *Enactivist interventions: Rethinking the mind*. Oxford University Press.

Gallagher, S., & Hutto, D. (2008). Understanding others through primary interaction and narrative practice. In J. Zlatev, T. P. Racine, C. Sinha, & E. Itkonen (Eds.), *The shared mind: Perspectives on intersubjectivity* (pp. 17–38). John Benjamins Publishing Compan.

Gallese, V., Rochat, M. J., & Berchio, C. (2013). The mirror mechanism and its potential role in autism spectrum disorder. *Dev. Med. Child Neurol., 55*(1), 15–22. https://doi.org/10.1111/j.1469-8749.2012.04398.x

Géron, A. (2017). *Hands-on machine learning with scikit-learn and tensorflow: Concepts. Tools, and Techniques to build intelligent systems* (2nd ed.). O'Reilly Media, Inc.

Gessaroli, E., Santelli, E., di Pellegrino, G., & Frassinetti, F. (2013). Personal space regulation in childhood autism spectrum disorders. *PLoS One, 8*(9), Article e74959. https://doi.org/10.1371/journal.pone.0074959

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT Press.

Greimas, A. J. (1970). *Du sens. Essais sémiotiques. Editions du Seuil*.

Greimas, A. J. (1983). *Du sens II. Essais sémiotics*. Editions du Seuil.

Hastie, T., Tibshirani, R., & Friedman, J. (2013). *The elements of statistical learning data mining, inference, and prediction* (2nd ed.). Springer.

Hobson, R. P. (1986). The autistic child's appraisal of expressions of emotion. *Child Psychol. Psychiatr. Allied Discip., 27*, 321–342. https://doi.org/10.1111/j.1469-7610.1986.tb01836.x

Hobson, R. P. (2014). Autism and emotion. In *Handbook of autism and pervasive developmental disorders* (4th ed., pp. 332–354). John Wiley & Sons, Inc.

van 't Hof, M., Tisseur, C., van Berckelear-Onnes, I., van Nieuwenhuyzen, A., Daniels, A. M., Deen, M., Hoek, H. W., & Ester, W. A. (2021). Age at autism spectrum disorder diagnosis: A systematic review and meta-analysis from 2012 to 2019. *Autism, 25*(4), 862–873. https://doi.org/10.1177/1362361320971107

Jaswal, V. K., & Akhtar, N. (2018). Being versus appearing socially uninterested: Challenging assumptions about social motivation in autism. *Behav. Brain Sci., 42*, e82. https://doi.org/10.1017/S0140525X18001826

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kamio, Y., Inada, N., Koyama, T., Inokuchi, E., Tsuchiya, K., & Kuroda, M. (2014). Effectiveness of using the Modified Checklist for Autism in Toddlers in two-stage screening of autism spectrum disorder at the 18-month health check-up in Japan. *J. Autism Dev. Disord., 44*(1), 194–203. https://doi.org/10.1007/s10803-013-1864-1

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159. https://doi.org/10.2307/2529310

Lebersfeld, J. B., Swanson, M., Clesi, C. D., & O'Kelley, S. E. (2021). Systematic review and meta-analysis of the clinical utility of the ADOS-2 and the ADI-R in diagnosing autism spectrum disorders in children. *J. Autism Dev. Disord., 51*(11), 4101–4114. https://doi.org/10.1007/s10803-020-04839-z

Lemm, S., Blankertz, B., Dickhaus, T., & Müller, K.-R. (2011). Introduction to machine learning for brain imaging. *NeuroImage, 56*(2), 387–399. https://doi.org/10.1016/j.neuroimage.2010.11.004

Lewkowicz, D. J. (2001). *The Concept of Ecological Validity: What Are Its Limitations and Is It Bad to Be Invalid?, 2*, 437–450.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord., 24*(5), 659–685. https://doi.org/10.1007/BF02172145

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Luyster, R., Gotham, K., Bishop, S., & Guthrie, W. (2012). *ADOS-2 – autism diagnostic observation Schedule* (2nd ed.). Western Psychological Services, 2nd ed.

Lorusso, A. M., Paolucci, C., & Violi, P. (2012). *Narratività problemi, analisi, prospettive*. Bologna University Press.

Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. ArXiv:1705.07874 [Cs, Stat] http://arxiv.org/abs/1705.07874.

Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., Bishop, S., Esler, A., Hus, V., Oti, R., Richler, J., Risi, S., & Lord, C. (2009). The autism diagnostic observation schedule—toddler module: A new module of a standardized diagnostic measure for autism spectrum disorders. *J. Autism Dev. Disord., 39*(9), 1305–1320. https://doi.org/10.1007/s10803-009-0746-z

McCarty, P., & Frye, R. E. (2020). Early detection and diagnosis of autism spectrum disorder: Why is it so difficult? *Semin. Pediat. Neurol., 35*, Article 100831. https://doi.org/10.1016/j.spen.2020.100831

Mueller, A. C., & Guido, S. (2017). *Introduction to machine learning with Python: A guide for data scientists.* O'Reilly Media.

Ochs, E., Kremer-Sadlik, T., Sirota, K. G., & Solomon, O. (2004). Autism and the social world: An anthropological perspective. *Discourse Stud., 6*(2), 147–183. https://doi.org/10.1177/1461445604041766

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

Paolucci, C. (2012a). Per una concezione strutturale della cognizione: Semiotica e scienze cognitive tra embodiment ed estensione della mente. In M. Graziano, & C. Luverà (Eds.), *Bioestetica, bioetica, biopolitica. I linguaggi delle scienze cognitive.* Corisco edizioni.

Paolucci, C. (2012b). Social cognition between theories, simulations and narratives: The narrative practice semiotic Hypothesis. *Intellectica. Revue de l'Association pour la Recherche Cognit., 58*, 173–195. https://doi.org/10.3406/intel.2012.1103

Paolucci, C. (2019). Social cognition, mindreading and narratives. A cognitive semiotics perspective on narrative practices from early mindreading to Autism Spectrum Disorder. *Phenomenol.Cognitive Sci., 18*(2), 375–400. https://doi.org/10.1007/s11097-018-9575-x

Paolucci, C. (2020). A radical enactivist account of social cognition. In A. Pennisi, & A. Falzone (Eds.), *The extended theory of cognitive creativity. Interdisciplinary approaches to performativity* (pp. 59–74). Springer.

Paolucci, C. (2021a). *Cognitive semiotics. Integrating signs, minds, meaning and cognition* (Vol. 24). Springer.

Paolucci, C. (2021b). NeMo project's observation methodology for embodied interactions and autism spectrum disorders. *Reti, saperi, linguaggi*. https://doi.org/10.12832/102767, 2021.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine learning in Python. *Machine Learning in Python, 6*.

Pennisi, A. (2021). *Che ne sarà dei corpi? Spinoza e i misteri della cognizione incarnata*. Il Mulino.

Posar, A., & Visconti, P. (2022). Early motor signs in autism spectrum disorder. *Children, 9*(2). https://doi.org/10.3390/children9020294. Articolo 2.

Reddy, V. (2003). On being the object of attention: Implications for self-other consciousness. *Trends Cognit. Sci., 7*, 397–402. https://doi.org/10.1016/S1364-6613(03)00191-8

Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *J. Autism Dev. Disord., 31*(2), 131–144. https://doi.org/10.1023/a:1010738829569

Rochat, M. J., Veroni, V., Bruschweiler-Stern, N., Pieraccini, C., Bonnet-Brilhault, F., Barthélémy, C., Malvy, J., Sinigaglia, C., Stern, D. N., & Rizzolatti, G. (2013). Impaired vitality form recognition in autism. *Neuropsychologia, 51*(10), 1918–1924. https://doi.org/10.1016/j.neuropsychologia.2013.06.002

Rotholz, D. A., Kinsman, A. M., Lacy, K. K., & Charles, J. (2017). Improving early identification and intervention for children at risk for autism spectrum disorder. *Pediatrics, 139*(2), Article e20161061. https://doi.org/10.1542/peds.2016-1061

Salari, N., Rasoulpoor, S., Rasoulpoor, S., Shohaimi, S., Jafarpour, S., Abdoli, N., Khaledi-Paveh, B., & Mohammadi, M. (2022). The global prevalence of autism spectrum disorder: A comprehensive systematic review and meta-analysis. *Ital. J. Pediatr., 48*(1), 112. https://doi.org/10.1186/s13052-022-01310-w

Stern, D. (1985). *The interpersonal world of the infant*. Basic Books.

Stern, D. (1992). *Diary of a baby. What your child sees, feels and experience*. Basic Books.

Teitelbaum, P., Teitelbaum, O., Nye, J., Fryman, J., & Maurer, R. G. (1998). Movement analysis in infancy may be useful for early diagnosis of autism. *Proc. Natl. Acad. Sci. U.S.A., 95*(23), 13982–13987. https://doi.org/10.1073/pnas.95.23.13982

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa, & A.c. Di (Eds.), *Before speech: The beginning of human communication*. Cambridge University Press.

Trevarthen, C. (1999). Musicality and the intrinsic motive pulse: Evidence from human psychobiology and infant communication. *Musicae Scientiae, 3*(1_suppl), 155–215. https://doi.org/10.1177/10298649000030S109

Trevarthen, C. (2002). Autism, sympathy of motives and music therapy. *Enfance, 54*(1), 86–99. https://doi.org/10.3917/enf.541.0086

Trevarthen, C., & Delafield-Butt, J. T. (2013). Autism as a developmental disorder in intentional movement and affective engagement. *Front. Integr. Neurosci., 7*, 49. https://doi.org/10.3389/fnint.2013.00049

Trevarthen, C., & Hubley, P. (1978). Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year. In A. Lock (Ed.), *183–229Action, gesture and symbol: The emergence of language*. Academic Press.

Volkmar, F. R. (2014). Editorial: The importance of early intervention. *J. Autism Dev. Disord., 44*(12), 2979–2980. https://doi.org/10.1007/s10803-014-2265-9

Wetherby, A. M., Woods, J., Allen, L., Cleary, J., Dickinson, H., & Lord, C. (2004). Early indicators of autism spectrum disorders in the second year of life. *J. Autism Dev. Disord., 34*(5), 473–493. https://doi.org/10.1007/s10803-004-2544-y

Wetherby, A. M., Brosnan-Maddox, S., Peace, V., & Newton, L. (2008). Validation of the Infant-Toddler Checklist as a broadband screener for autism spectrum disorders from 9 to 24 months of age. *Autism: Int. J.Res.Pract., 12*(5), 487–511. https://doi.org/10.1177/1362361308094501

Zilbovicius, M., Meresse, I., Chabane, N., Brunelle, F., Samson, Y., & Boddaert, N. (2006). Autism, the superior temporal sulcus and social perception. *Trends Neurosci., 29*(7), 359–366. https://doi.org/10.1016/j.tins.2006.06.004

Zwaigenbaum, L., Bryson, S., & Garon, N. (2013). Early identification of autism spectrum disorders. *Behav. Brain Res., 251*, 133–146. https://doi.org/10.1016/j.bbr.2013.04.004