Predictable Robots for Autistic Children—Variance in Robot Behaviour, Idiosyncrasies in Autistic Children's Characteristics, and Child–Robot Engagement

BOB R. SCHADENBERG and DENNIS REIDSMA, University of Twente
VANESSA EVERS, University of Twente and Nanyang Technological University
DANIEL P. DAVISON, JAMY J. LI, and DIRK K. J. HEYLEN, University of Twente
CARLOS NEVES and PAULO ALVITO, IDMind
JIE SHEN and MAJA PANTIĆ, Imperial College London
BJÖRN W. SCHULLER, University of Augsburg and Imperial College London
NICHOLAS CUMMINS, University of Augsburg and King's College London
VLAD OLARU and CRISTIAN SMINCHISESCU, Institute of Mathematics of the Romanian Academy
SNEŽANA BABOVIĆ DIMITRIJEVIĆ and SUNČICA PETROVIĆ, Serbian Society of Autism
AURÉLIE BARANGER, Autism Europe
ALRIA WILLIAMS and ALYSSA M. ALCORN, University College London
ELIZABETH PELLICANO, University College London and Macquarie University

Predictability is important to autistic individuals, and robots have been suggested to meet this need as they can be programmed to be predictable, as well as elicit social interaction. The effectiveness of robot-assisted interventions designed for social skill learning presumably depends on the interplay between robot predictability, engagement in learning, and the individual differences between different autistic children. To better

This work was made possible through funding from the European Union's Horizon 2020 research and innovation program under grant agreement no: 688835 (DE-ENIGMA).
Authors' addresses: B. R. Schadenberg, D. Reidsma, D. P. Davison, J. J. Li, and D. K. J. Heylen, University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands; emails: {b.r.schadenberg, d.reidsma, d.p.davison, j.j.li, d.k.j.heylen}@utwente.nl; V. Evers, University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands and Nanyang Technological University, 50 Nanyang Ave, Singapore 639798, Singapore; email: v.evers@utwente.nl; C. Neves and P. Alvito, IDMind, Polo Tecnologico de Lisboa, Lt 1, 1600-546 Lisbon, Portugal; emails: {cneves, palvito}@idmind.pt; J. Shen and M. Pantić, Imperial College London, South Kensington Campus Exhibition Rd, South Kensington, London SW7 2AZ, United Kingdom; emails: {jie.shen07, maja.pantic}@imperial.ac.uk; B. W. Schuller, University of Augsburg, Universitätsstraße 2, 86159 Augsburg, Germany and Imperial College London, Exhibition Rd, South Kensington, London SW7 2BX, United Kingdom; email: schuller@informatik.uni-augsburg.de; N. Cummins, University of Augsburg, Universitätsstraße 2, 86159 Augsburg, Germany and King's College London, Strand, London WC2R 2LS, United Kingdom; email: nicholas.cummins@informatik.uni-augsburg.de; V. Olaru and C. Sminchisescu, Institute of Mathematics of the Romanian Academy, Calea Griviţei 21, Bucharest, Romania; emails: {vlad.olaru, cristian.sminchisescu}@imar.ro; S. B. Dimitrijević and S. Petrović, Serbian Society of Autism, Gundulicev venac street 38, Belgrade 11000, Serbia; email: suncica.petrovic@yahoo.com; A. Baranger, Autism Europe, Rue Montoyer 39, Brussels 100, Belgium; email: Aurelie.baranger@autismeurope.org; A. Williams and A. M. Alcorn, University College London, Gower St, London WC1E 6BT, United Kingdom; emails: {alria.williams, a.alcorn}@ucl.ac.uk; E. Pellicano, University College London, Gower St, London WC1E 6BT, United Kingdom and Macquarie University, Balaclava Rd, Macquarie Park NSW 2109, Australia; email: liz.pellicano@mq.edu.au.

1073-0516/2021/08-ART36
https://doi.org/10.1145/3468849

understand this interplay, we report on a study where 24 autistic children participated in a robot-assisted intervention. We manipulated the variance in the robot's behaviour as a way to vary predictability, and measured the children's behavioural engagement, visual attention, as well as their individual factors. We found that the children will continue engaging in the activity behaviourally, but may start to pay less visual attention over time to activity-relevant locations when the robot is less predictable. Instead, they increasingly start to look away from the activity. Ultimately, this could negatively influence learning, in particular for tasks with a visual component. Furthermore, severity of autistic features and expressive language ability had a significant impact on behavioural engagement. We consider our results as preliminary evidence that robot predictability is an important factor for keeping children in a state where learning can occur.

CCS Concepts: • **Human-centered computing** → **Interaction design theory, concepts and paradigms**; • **Social and professional topics** → *People with disabilities*; • **Computer systems organization** → Robotics;

Additional Key Words and Phrases: Predictability, variability, autism spectrum condition, human-robot interaction, engagement, individual differences

ACM Reference format:
Bob R. Schadenberg, Dennis Reidsma, Vanessa Evers, Daniel P. Davison, Jamy J. Li, Dirk K. J. Heylen, Carlos Neves, Paulo Alvito, Jie Shen, Maja Pantić, Björn W. Schuller, Nicholas Cummins, Vlad Olaru, Cristian Sminchisescu, Snežana Babović Dimitrijević, Sunčica Petrović, Aurélie Baranger, Alria Williams, Alyssa M. Alcorn, and Elizabeth Pellicano. 2021. Predictable Robots for Autistic Children—Variance in Robot Behaviour, Idiosyncrasies in Autistic Children's Characteristics, and Child–Robot Engagement. *ACM Trans. Comput.-Hum. Interact.* 28, 5, Article 36 (August 2021), 42 pages.
https://doi.org/10.1145/3468849

1 INTRODUCTION

Autism Spectrum Condition (hereafter referred to as "autism") is a neurodevelopmental condition that is characterised by difficulties in social communication and interaction (so-called "social features") and by restricted, repetitive behaviour and interests (so-called "non-social features") [3]. Whether these features are considered disabling for an individual can depend in part on the extent and nature of support provided by others [97]. This support can include both helping the individual child or young person to develop skills and strategies (for example, to understand situations and communicate their needs) and adapting the environment to enable the child to function and learn within it. In the context of social skill learning, experiencing discomfort due to dealing with unpredictability is problematic as it prevents children from being in a state where they are ready to learn. Incorporating a robot in social skill learning might be helpful in that it can provide a highly predictable manner of learning social skills, as we can systematically control the predictability of the robot's behaviour [128]. Indeed, the predictability of a robot is a commonly used argument for why robots may be promising tools for autism professionals working with autistic children [e.g., 34, 35, 43, 67, 120, 141]. Contemporary robot-assisted interventions have also shown that they can maintain engagement for at least a month, whilst operating autonomously [22, 123], and there are positive indications that such interventions can lead to learning [22, 35, 123, 138] and the generalisation thereof to different contexts [123].

The predictability of robots may make it easier for autistic children to engage in learning in a robot-assisted intervention and maintain this engagement. However, robots cannot behave fully predictably as well as provide meaningful learning, as the learning gains may not generalise to the less predictable world of people [1]. Moreover, a fully predictable robot would perpetuate in repetitive behaviour [33], limiting its long-term usefulness, and would also be unable to autonomously respond to the (unpredictable) dynamics of real-world settings [23]. Both make large scale deployment of such a robot difficult. Thus, there is a tradeoff between making the robot more predictable and providing the child with meaningful learning content with a tractable robot-assisted intervention, which requires a degree of unpredictability. This unpredictability stems from (a) adding novel content, which has not yet been learned, (b) making the learned skill's approximate—at least in part—what happens in real-world settings with humans, to facilitate generalisability to such highly unpredictable environments which autistic children need to learn to deal with, and (c) implementing complex robot behaviours (e.g., responsive actions to the environment), for which it may be difficult to discern the cause of its behaviour and thereby prevent children from learning to predict the robot's behaviour. Thus, on the one hand, we want autistic children to be in a state where learning can occur, where the children show high levels of engagement with the learning material. This may require very predictable interactions. On the other hand, we want them to learn skills that are meaningful in the human world, which comes with a degree of unpredictability. A balance needs to be struck then, where the robot is sufficiently predictable to maintain engagement while still providing learning content that is representative of human-human interaction.

What constitutes "sufficiently predictable" is likely to differ between autistic children as they are a notoriously heterogeneous group [60]. Some children may be better equipped to deal with unpredictability than others. For these children, predictability of the robot's behaviour may not be as important, and the focus can lie on optimising learning. Furthermore, children's reactions to the robot's unpredictability may be very different, and they may find different aspects of unpredictability problematic [54]. This makes it difficult to generalise findings on predictability and autism to autistic children working with robots.

The effectiveness of robot-assisted interventions designed for social skill learning presumably depends—in part—on the *interplay between robot predictability, engagement in learning, and the individual differences between different autistic children.* To better understand this interplay, we report on a study where autistic children participated in a robot-assisted activity, where we manipulated the variance in the robot's behaviour as a way to operationalise predictability, and measured the children's individual characteristics as well as their engagement related behaviours. The robot-assisted activity was developed by the European project "DE-ENIGMA"—which funded the current study—and revolves around playing with the basics of recognising facial expressions of emotion. In the remainder of this article, we will first elaborate on the concepts of engagement and of predictability in Section 2. This literature forms the basis upon which we base our research questions and hypothesis (Section 3), and our methods (Section 4). In Section 5, we present the results of our analysis of the robot's predictability, and two facets of engagement, namely behavioural engagement and visual attention. We conclude the article with a discussion on how we interpret the results of our study and what this means for the interplay between robot predictability, the two facets of engagement, and the individual differences in Section 6 and conclude on our research questions in Section 7.

The contribution of this article is threefold. First, we provide a literature-based explication of predictability as it relates to **human–robot interaction (HRI)**. Second, we have operationalised the predictability of a robot in a way that it can be assessed in real-life scenarios and provide measures that can be used to compare the predictability of different robots. Finally, we report on new analyses into how a robot's predictability influences two facets of engagement in a dataset of 27 autistic children.

2 BACKGROUND

2.1 Autism, Predictability, and Atypical Predictive Processing

Our senses are constantly dealing with ambiguous sensory information, trying to make sense of it all. According to Bayesian accounts of perception, the human brain is constantly generating predictions on what sensory information is expected in order to resolve this ambiguity and attempts to minimise the error between the incoming sensory information and the prediction [7, 24, 49, 63, 94]. This is referred to as predictive processing, or predictive coding. When a prediction does not match with the observed sensory information, additional cognitive effort is required to resolve the mismatch and to identify its cause. When these mismatches happen too often, it can lead to a person being overwhelmed or anxious [102]. Autistic individuals are believed to frequently experience these mismatches between predicted and observed sensory information due to a decreased influence of predictions on perception, which is hypothesised in current Bayesian accounts of autism to be at the heart of their condition [31, 81, 104, 135], [see 100, for a review]. According to the Bayesian accounts of autism, the non-social features are the result of trying to maintain a predictable environment through either self-generated sensory information—which can easily be predicted—or through enacting control over their environment. The social features of autism, on the other hand, are explained by the highly unpredictable nature of social environments, which can therefore be difficult to deal with. The Bayesian accounts of autism are not the only theories that relate prediction difficulties with autistic features. For instance, the Empathizing-Systemizing theory [8, 9], posits that the social features of autism arise from a below-average empathy, whereas the non-social features stem from above-average systemizing—identifying lawful patterns in the information. When faced with information that does not strictly adhere to discernible laws (i.e., unpredictable environments), such as in social interactions, autistic individuals struggle to make sense of it. This theory, however, is very descriptive in nature, and does not fully specify the (different) computational mechanisms underlying autism.

A decreased influence of predictions on the perception in autistic individuals is supported by various lab studies [6, 45, 53, 80, 105, 142], although support is not always found [17, 83, 88, 103, 140]. Nevertheless, literature has shown that, in practice, offering a more predictable environment is invaluable to facilitating learning by providing an environment that puts the child at ease by not having to deal with the discomfort resulting from unpredictability and requires fewer cognitive resources to be processed. Current educational practices, therefore emphasise the need for structure at schools (e.g., through the TEACCH approach [92]), so that autistic children know what to expect during the day, increasing their engagement in learning [18, 87, 99]. All things considered, there is compelling evidence that predictability is important to autistic children.

A construct that may be similar to a preference for predictability—or the intolerance of unpredictability—of autistic children is that of ***intolerance of uncertainty*** (**IU**). IU is defined as "the tendency to react negatively on an emotional, cognitive, and behavioral level to uncertain situations and events" [42], and consists of two components, namely a "desire of predictability" and "uncertainty paralysis" [13]. Knowing what will happen in the future allows us to increase the odds of a desirable outcome, or brace for future adversity. However, this requires knowledge on the probability, timing, and nature of the future event, which is often not available as the future is intrinsically uncertain. Anticipating the future can therefore induce anxiety, as the uncertainty reduces our ability to efficiently and effectively prepare for the future. Individuals with heightened levels of anxiety are believed to have an excessive response to uncertainty [56], IU is considered to be a cognitive vulnerability factor in a broad range of emotional disorders like general anxiety disorder, social anxiety, and depression [20, 44, 64]. IU has also been studied with autistic individuals, who obtain higher IU scores than the general population [15, 21, 95]. Furthermore, the higher levels of IU might explain the high levels of anxiety that are consistently reported in autistic children [131, 144], as IU was found to mediate the relation between autism and anxiety [15, 145]. However, the authors caution that a different causal story cannot be ruled out, given the non-experimental data used for the mediation analysis [15].

2.2 Predictability and Its Operationalisation as Variance in a Robot

Incorporating robotic technology in interventions for autistic children has been proposed to meet the need for predictability of these children [32, 34]. Robots are programmable and can therefore be designed to be highly predictable. However, regardless of a robot's programming, it will be difficult to correctly predict the robot's behaviour when meeting a robot for the first time. What is the robot going to do or say first? When is it going to greet me? How will it greet me? Without having prior knowledge of the robot, a person is hard pressed to correctly answer these questions. This example highlights that predicting robot behaviour is a learning process and that the robot is initially unpredictable. Through interacting with the robot, people become more familiar with its behaviour, improving their ability to predict its behaviour [40]. However, this requires one to perceive and learn the *structural regularities* in the robot's behaviour—also referred to as invariance detection [51, 52]. These structural regularities hold predictive value in that they can be used to generate predictions regarding the robot's behaviour in the future. For autistic children, detecting structural regularities (i.e., invariant structures) may be challenging [61]. *Variance* in robot behaviour can therefore lead to interactions with a robot that are too complex to be learned, and therefore predicted, or it can take the child longer to learn and predict. Designing a robot to be highly predictable then might entail programming robot behaviour with low variance such that it has high structural regularities that can be easily inferred from its behaviour.

In short, we define the predictability of a robot as an attribute that refers to "the degree in which a person can quickly and accurately learn to predict a robot's future behaviour". For the purpose of this article, we operationalise the design of predictable robots as decreasing or increasing the variance in their behaviour, where more variance decreases the predictive value of the robot's behaviour and thus decreases its predictability.[1]

2.2.1 Types of Robot Variance. Variance in robot behaviour can stem from any of the robot's expressive modalities, as well as how those modalities are used within the interaction dynamics. There are several types of variance that we consider in this article, namely variance in speech, motion, topic, and time:

- Speech variance refers to variability in the words that the robot uses (diction) and how those words are spoken (prosody). Diction variance can be operationalised as using different sentence grammars and/or different words that have similar meaning. Prosody variance relates to using different prosody and/or different voice actors to change intonation, stress, rhythm, and pitch of the sound. While the use of prerecorded speech can minimise speech variance, using models for natural language generation to generate dynamic speech may influence the variance in the robot's speech significantly.
- Motion variance refers to variability in the robot's motions, such as facial expressions and gestures. A robot could consistently use the same static animation to communicate a certain intent, which would keep the motion variance low. In contrast, when using models to generate dynamic motions, such as inverse kinematics, each robot motion can be unique in its trajectory, increasing variance in the robot's motions.
- Temporal variance refers to variability in the timing of robot behaviour, both within actions (e.g., timing of a motion trajectory) as well as the timing between actions. While this type of variance is not a type of variance that is actively designed for, temporal delays in robot behaviour are likely to occur as the result of a lack of computational power, or due to

[1] For more information on robot predictability, and different operationalisation thereof, see Schadenberg et al. [128], or see Chapter 6 in Schadenberg [124] for a more extensive elaboration.

the physics of the robot's motors, increasing temporal variance. Such temporal delays can disrupt the interaction of an autistic child [47].

— Topic variance refers to variability in the use of different actions that address different topics at a particular point in an interaction. For example, at the first step of an interaction, increased topic variance could involve a robot saying either a greeting or a comment about a bird, as opposed to consistently saying a greeting. Types of topic variance related to a robot can include (1) when a robot responds to its internal state (e.g., battery level, or its (faulty) perception) which the observer does not know about; (2) when a robot responds to an external event (e.g., a person walking in on the ongoing session), or item (e.g., a clock) in the environment; and (3) when a robot responds with an action that is not legible (i.e., understandable) to the observer (e.g., randomly saying "beep beep").

There are other sources of variance (e.g., variance in robot appearance, or morphology), but we do not utilise these sources in our study for introducing unpredictability. Finally, in addition to these types of variance, a robot can combine modalities into one action. For example, a robot capable of facial expressions of emotion can more clearly communicate these expressions when they are combined with short, emotional non-speech expressions [125]. This can also lead to variance when different combinations, intended to express the same multimodal intent, are perceived as different robot actions. For autistic children, processing multimodal information may be more difficult than for typically developing children [25, 59, 98]. Furthermore, autism professionals report that presenting multimodal information might cause an information overload [65].

2.2.2 Reducible and Irreducible Unpredictability. While prioritising the minimisation of the types of robot variance mentioned above can lead to highly predictable robots, doing so severely restricts the types of applications for robots for autism. An example of such prioritisation of predictability would be programming it to do the same thing, in exactly the same manner, again and again, resulting in a very limited repertoire of robot behaviours with the potential for endless repetition. For such robots, autistic children may quickly learn to predict their behaviour because there are only a few behaviours to predict. While such robots may play a role in the broader scope of robots for autism [39], a *degree* of unpredictability is unavoidable for most of the robot roles envisioned by autism professionals [1, 66] and researchers [19, 37, 38, 122]. For instance, when the robot's role is that of a trainer or educator, where its task is to model, teach, and/or practice a targeted skill and to provide feedback to the child. When the robot is positioned as a social interaction partner, it will have to perform novel behaviour and in general have a larger range of behaviours, in order to deal with the complexity, dynamics, and unpredictability of interacting with a human. The resulting unpredictability of the robot's behaviour is *irreducible*, as it is required for the robot to perform its primary task. That is, even when the robot is carefully designed in terms of its predictability, irreducible, and unpredictability is unavoidable and does not lead to design considerations. For a robot that is positioned as a tutor, an example of this would be the unpredictability that can stem from introducing learning content, which is the bare minimum it needs to do.

There are also technical reasons why a robot is not always behaving predictably, in particular for robots that behave autonomously, as the technical systems that allow the robot to perceive, reason, and/or act, can each introduce variance. The robot may respond to what it is perceiving, which can be faulty, or the observer can be unaware of what the robot is responding to. The robot's reasoning may be too complex, or faulty, to be understood by the observer. Or, the intent the robot tries to convey may simply not be understood by the observer from the robot's behaviour. In all of these examples, the resulting robot behaviour is unexpected and was not predicted. As the resulting unpredictability is (to a certain extent) inherent to using technical systems, we also consider this to be irreducible unpredictability. In part, this can be solved by utilising a Wizard-of-Oz paradigm, where a person remotely controls the robot and creates the illusion of interacting with an autonomous robot, as this reduces the need for certain technical systems. However, for large-scale and long-term use of robots, the use of a Wizard-of-Oz system is not practical and requires that the robot can operate autonomously [23, 122]. The autism professional using the robot is too busy paying attention and interacting with the child to also control the robot, and we deem it unlikely that robot-assisted interventions provide enough value to warrant paying an additional person to control the robot.

In contrast to irreducible unpredictability, there are also aspects of robot interaction design that we consider to be *reducible*. These are instances where adding one feature may improve a certain aspect of the interaction, but may decrease the robot's predictability. They are design choices, rather than a necessity for the robot to perform its primary task. For instance, common design principles for achieving long-term interactions for typically developing children include using intelligent tutoring systems to provide the optimal challenge and keep children motivated [110, 126], adding variance to speech to reduce boredom due to endless repetition of a robot's behaviour [27, 73, 77], or personalising speech [85, 137]. For a tutoring robot, these features are desirable, but are not necessarily required for the robot to perform its primary function. In the case of robots for autistic children, choosing whether to implement such non-essential features that increase unpredictability is a balancing act. When the robot becomes too unpredictable it may have negative effects on the interaction and could even negate the intended effect of feature.

2.3 Engagement and Autistic Children Engaging with Robots

The benefit of providing a predictable environment for autistic children in which they can learn is that it may be easier to engage with learning as they do not have to deal with the discomfort resulting from unpredictability. Engagement is a necessary prerequisite for learning [91], where higher engagement results in more opportunities for cognitive and social skill learning [48, 55].

2.3.1 Engagement Definitions and Measures. Azevedo [5] discuss how the concept of engagement in learning has been (mis)interpreted in various ways in the literature leading to different definitions and meanings of this concept. Within the field of HRI, engagement is often used as an outcome measure to say something about the quality and length of a participant's interaction with the robot. Engagement is also a concept used for the development of robots that can detect various stages of the user's engagement, such as the intention to engage, being engaged, or being disengaged. An often used definition of engagement in HRI literature is that of Sidner et al. [130], who define engagement as "a process by which individuals in an interaction start, maintain and end their perceived connection to one another". Other definitions emphasise that engagement is an affective process formalised as the degree to which an individual wants, or chooses, to engage with a system [e.g., 14, 96]. While there is no consensus on a single definition for engagement, it is generally viewed as a multi-dimensional concept, including a behavioural, cognitive, and affective component [28, 48]. Note that these components of engagement are overlapping and can sometimes be difficult to disentangle [134].

We are interested in engagement with a robot as it relates to the children's state in which learning can occur, since this is what we are trying to achieve with robot-assisted interventions. In such a context, BEHAVIOURAL ENGAGEMENT refers to the child's participation in learning activities and involves on-task behaviour. Overall, studies on the engagement in learning of autistic children mostly relate to behavioural engagement [72], often referred to as "social engagement" when the task is to engage with another person. As with the definition of engagement, there are also various approaches to measuring the behavioural engagement of autistic children in their interaction with a robot. It can be directly assessed through observing the behaviour of the child, annotating the level of behavioural engagement on a macro-behavioural level. For instance, Kim et al. [74] developed a compliance-based coding scheme for measuring (behavioural) engagement, where the speed of the autistic child's reaction to instructions or requests is indicative of the child's level of behavioural engagement. In this coding scheme, spontaneous engagement is the highest level of behavioural engagement, in contrast to a child refusing to comply to the robot or adult's request and walking away, which is the lowest level of behavioural engagement. Other studies report on micro-behavioural interactions, which are used to code the type of engagement, to get a deeper insight into how autistic children engage [e.g., 78, 127]. The choice for a certain measurement of behavioural engagement appears to be influenced by the purpose of the study and the type of data being gathered.

Next, AFFECTIVE ENGAGEMENT is about the child's (inferred) interest in the learning activity and how much the child enjoys it. It is generally assessed through observing the autistic child's emotions from which the underlying affect is inferred in terms of valence and/or arousal [e.g., 74, 116, 117]. Correctly inferring the affect from the emotional expressions of autistic children can be difficult, however, as they can produce unique and unusual facial expressions, including blends of incompatible emotions that are not seen in typically developing children or children with Down syndrome [148]. Also, the vocal intonation of autistic children can be atypical when expressing emotions [86, 90, 101]. Notwithstanding, valence and arousal can be successfully annotated for autistic children with a sufficient agreement between coders [74, 117]. Furthermore, using a machine learning approach trained on data of autistic children, valence and arousal can be detected using facial, body-pose, and audio features, and heart rate [116].

Lastly, COGNITIVE ENGAGEMENT refers to the quantity and quality of the child's psychological investment in learning (i.e., use of cognitive effort in order to understand). The cognitive engagement of autistic children is difficult to measure, as the current measures for cognitive engagement [5] overlap with the other components of engagement [93], or can be too complex to be used by autistic children, such as self-report questionnaires. Task-evoked pupillary responses have long been associated with attentional engagement [10] and cognitive activity [62, 71], as well as emotional arousal [16], and have been used by researchers to measure the cognitive/affective engagement of autistic children [e.g., 50]. However, measuring pupil dilation requires carefully controlled experiments and experiment environment to control for other factors that influence pupil dilation, such as the pupillary reflex to changes in illumination [11]. This makes pupil dilation difficult to use in real-world settings where the illumination cannot be fully controlled. A concept that is more easily measured—and is related to cognitive engagement—is that of attention, which is often viewed as a necessary component for basic forms of engagement to occur [29]. Attention has both a covert and overt component, where overt visual attention is relying on the gaze fixation on a certain location, and covert attention involves cognitive processes for paying attention to something without the movement of the eyes [147]. Indeed, gazing at a particular object is not always indicative of the person paying attention to that object [107]. Nonetheless, measuring visual attention through gaze is a commonly used proxy for cognitive engagement in the field of HRI [4, 112], and is also used with autistic children engaging with robots [e.g., 4, 69, 70, 78, 139].

Clearly, the various measures and components of engagement also show overlap. For example, as Miller [93] noted, gazing at a certain location may also indicate affective engagement, as people look more at what they like [89]. Altogether, engagement (the concept as a whole) is a fusion of behavioural, affective, and cognitive components of a person's involvement with a robot. As each component of engagement can result in learning, considering all three components together can provide a richer characterisations of a child's engagement than any single component. Sometimes all three components of engagement are combined into one bespoke measure for engagement [e.g., 68, 69, 132]. For example, through measuring social signals such as eye gaze, vocalisations, smiles, spontaneous interactions, and imitation, which can then be converted into an engagement score by adding one point of engagement for each social signal that is present in a certain segment.

2.3.2 Autistic Children Engaging with Robots. Scassellati et al. [122] discuss that studies on robots for autism often report positive effects of the robot on the engagement of autistic children [122], as shown by increases in positive affect [30, 76], communication [75, 143], and attention [34, 133, 139]. Importantly, the engagement that is observed is often social in nature and is directed not only at the robot, but also at other people near the robot [36, 43, 46, 75, 79, 114]. The latter is significant, because from a pedagogical point of view, it does not necessarily matter whether the child interacts with the robot or whether the robot elicits interaction between the child and adult, as learning can occur in either case. Robots then are uniquely positioned to assist autistic children in learning socials skills as they can be both highly predictable in their behaviour, and therefore deliver learning content in a manner that is easier to process and more comfortable to them, as well as elicit social interactions with the child through which the skills can be learned.

2.4 Individual Differences in Autism

Autistic children can vary widely from one another. Autism is a condition that causes atypicalities in cognitive, emotional, behavioural, and social functioning; however, such atypicalities are manifested differently in both quality and quantity [60]. For instance, some autistic children may have exceptional intelligence, whereas others have a severe intellectual disability [57]. Some speak fluently, while others never develop spoken language. In light of the Bayesian accounts of autism, some autistic children should be better equipped to deal with unpredictability than others. As we are looking to balance the robot's predictability, what constitutes "sufficiently predictable" is therefore likely to vary between autistic children. Indeed, Goris et al. [54] reported that, when autistic traits are measured in typically developing adults, these correlate significantly with preferences for predictability. Furthermore, autistic individuals' IU—as we mentioned earlier, a concept that shows similarities with a desire for predictability — is also known to vary between autistic people [15, 21].

Importantly, individual differences also affect how and to what extent autistic children interact with robot. The types of interaction that autistic children spontaneously initiated in a robot-assisted activity correlated with individual factors [127]: Those with stronger language ability, social functioning, and lower autistic features, initiated more functional interactions towards the robot (e.g., talking to it), in contrast to visual and tactile exploration of the robot's materials (i.e., touching and stroking specific parts of robot, such as its hands). The child's autistic features have also been reported to correlate strongly with the level of behavioural engagement of autistic children in a robot-assisted activity [78, 117].

In summary, children's reactions to the robot's unpredictability may be very different, and they may find different aspects of unpredictability problematic. For autistic children who are better equipped to deal with unpredictability, the predictability of the robot's behaviour may not be as important, and the focus can lie on optimising the learning content. In general, the heterogeneity between autistic children makes it difficult to generalise findings, but it also stresses that there is no one-size fits all solution to keeping them engaged in robot-assisted interventions whilst providing meaningful learning.

3 PROBLEM STATEMENT

To make an informed decision on how to position and design the robot's behaviour in terms of its predictability, we need to better understand the interplay between robot predictability, engagement, and the individual differences between different autistic children. The aim of our study was to investigate this interplay, where we specifically looked at *behavioural engagement* and *visual attention* (as a proxy for cognitive engagement) in relation to the robot's predictability. We measured these two facets of engagement through manual annotations of observable child behaviours (see Section 4.7). In our study, autistic children engaged in a robot-assisted activity that was about the basics of recognising emotional facial expressions. Over the course of four sessions, the children interacted with a robot that was either low or high in variance of its behaviour, which was how we operationalised robot predictability (see Section 2.2). We aimed at addressing the following research questions:

Research question 1: How does variance in the robot's behaviour affect the autistic child's engagement?
- To what extent does variance in the robot's behaviour affect the autistic child's *behavioural* engagement?
- To what extent does variance in the robot's behaviour affect the autistic child's *visual* attention?

We hypothesised that initially, in the first session, there should be no difference between the low and high variance conditions, as the robot's behaviour is novel and still has to be learned. However, based on the importance of predictability to autistic children, we expected that over sessions the behavioural engagement and visual attention of autistic children should increasingly diverge between the low-variance robot compared with the high-variance robot in favour of the low-variance condition. That is, we expected an interaction effect between robot predictability on behavioural engagement and on visual attention over sessions, but no main effects. For visual attention, this means that we expected the children to increasingly look less towards the robot (the source of the unpredictability) and more to non-activity-related locations that provided little unpredicted sensory input, such as the walls.

Our second research question is related to the individual differences between autistic children.

Research question 2: How do individual differences between autistic children influence their engagement in the activity in relation to the variance in the robot's behaviour?
- To what extent do autistic children's *autistic features* moderate the relation between the two facets of engagement and robot predictability?
- To what extent do autistic children's *expressive language ability* moderate the relation between the two facets of engagement and robot predictability?
- To what extent do autistic children's *IU* moderate the relation between the two facets of engagement and robot predictability?

Based on the preliminary findings of Goris et al. [54], Rudovic et al. [117], and Schadenberg et al. [127], we hypothesised that autistic children with higher autistic features should be less behaviourally engaged and pay less visually attention to activity-relevant locations (main effects), and their behavioural engagement should be more strongly and negatively affected than those with lower autistic features (interaction effect). Our hypothesis was similar for autistic children with lower expressive language ability. For autistic children who were more sensitive to unpredictability, as measured through their IU, we expected that they should respond more strongly to more variance in the robot's behaviour (interaction effect).

This study is complementary to another study which will be published elsewhere (hereafter referred to as the *complementary study*) and is the first study that assesses the claim on the benefits of robots being highly predictable. That complementary study is about micro behavioural analysis of autistic children in light of robot predictability. In the study reported in the current article, we used the same study design and data collection as the other study, but focused on different research questions and conducted different analyses to address those questions.

Table 1. Participant Characteristics

	Condition	
	Low-variance	*High-variance*
n (sex)	12 (5 female)	12 (3 female)
Age (years:months)		
M (*SD*)	8:8.92 (1:8.70)	8:4.42 (1:8.49)
Range	6:10–11:7	6:10–11:4
ADOS-2[a] Calibrated Severity Score		
M (*SD*)	6.08 (1.78)	6.25 (1.60)
Range	4–10	4–10
CARS-2[b]		
M (*SD*)	27.83 (4.09)	29.17 (6.56)
Range	20.5–33.0	21.5–38.5
SCQ[c]		
M (*SD*)	22.25 (8.72)	25.50 (5.79)
Range	8–37	17–33
Bespoke Scale of Expressive Language		
M (*SD*)	2.92 (0.79)	2.42 (1.24)
Range	2–4	0–4
IUSC-S[d]		
M (*SD*)	2.64 (1.04)	3.19 (1.16)
Range	1.42–4.00	1.00–4.92

[a]**ADOS-2** [84].
[b]**CARS-2** [129].
[c]**SCQ** [118].
[d]**Intolerance of Uncertainty Scale for Children—Simplified (IUSC-S)**.

4 MATERIALS AND METHODS

4.1 Participants

Autistic children from the United Kingdom were recruited from a special education institution in the Greater London area. In total, 27 children were recruited of whom 24 (8 girls) were included in the analysis. These were autistic children with limited spoken communication and high support needs. For the three children who were excluded from the analysis, participation was discontinued due to administrative error (1 girl), or due to elevated anxiety during the session (2 boys).[2] All included participants had previously received an independent clinical diagnosis of autism according the ICD-10 [146], DSM-IV-TR [2], or DSM-V [3]. All children were assessed by the **Autism Diagnostic Observation Scale-second edition (ADOS-2)** [ADOS-2, 84], the **Childhood Autism Rating Scale-second edition (CARS-2)** [CARS-2, 129], the **Social Communication Questionnaire (SCQ)** [SCQ, 118], and a bespoke scale of expressive language ability. In addition to receiving a clinical diagnosis of autism, all children scored above the autism cutoff on ADOS-2 (4 or higher). The participants' characteristics can be viewed in Table 1.

This study was reviewed and approved by the ethics committee of University College London, Institute of Education (REC 1175). For all children, parental consent was obtained prior to their

[2]Both were in the high-variance condition (see Section 4.3). For one boy, the activity was inducing too much anxiety—he did not get past the introduction in session one. The other boy repeatedly needed to be calmed in the second session, and eventually refused to go on. It was then decided to stop the experiment for this child.

Fig. 1. Robokind's humanoid robot R25 called "Zeno" or "Milo".

participation in our study. Our experimental protocols followed the ethical standards laid down in the 1964 Declaration of Helsinki.

4.2 Materials

The robot-assisted activity that was developed for the DE-ENIGMA project consisted of a social robot, a tablet, and one laptop. The social robot that was used is Robokind's humanoid robot R25 called "Zeno" (see Figure 1). It has five degrees of freedom in its face and two in its neck, making it capable of expressing various, recognisable, facial expressions of emotion [119, 125]. The robot also showed a number of bodily gestures, such as waving, cheering, or dancing using its the seven degree's of freedom in its body. A 9.7-inch Android tablet was used by the participant to provide answers to the tasks by selecting (part of) a picture of the robot, or choosing what action the robot should do. In turn, the robot would autonomously respond to the participant's choice. The interaction was recorded through four high-resolution webcams, and one wide-angle webcam, all with audio. In addition, high quality audio recordings were obtained through a dedicated microphone and 3D video recordings from a Microsoft Kinect. Note that we only used the data from the webcams for our analyses.

4.3 Experiment Design

We used the data collected in the complementary study. We will briefly summarise the experiment design here. The between-participants study involved two conditions, where the independent variable was the robot's behavioural variance, which was either low or high. The children were randomly assigned to one of the two condition and remained in this condition throughout the experiment. In the *low-variance condition*, the robot's behaviour showed minimal variance. This means that the verbal and physical behaviour for a certain action was always the same. For example, the robot would always say "Hi, my name is Zeno" and wave with its right arm as a way of greeting the child. In contrast, in the *high-variance condition*, we implemented speech, motion, temporal, and topic variance, to increase the behavioural variance of *all* of the robot's actions. Thus, the robot displayed behavioural variance throughout the interaction. In this condition, each of the robot's actions had four variations that differed the aforementioned types of variance, excluding the additional actions that were designed to introduce topic variance. This was done as follows:

— Speech variance: variability in the words that the robot uses (diction) and how those words are spoken (prosody).
— Motion variance: variability in patterns in the movements of the robot's face and arms during emotional facial expressions and robot actions.
— Temporal variance: variability in time offsets between the child pressing a button and the robot responding to the button press.
— Topic variance: the use of different actions that address different topics at a particular point in an interaction. Note that in the low-variance condition, the robot actions for topic variance were responsive actions in that the robot responded to an event (e.g., hearing a noise). In the high-variance condition, there would be no observable event that explained the robot's action.

The variant actions have been designed to display *unimodal* variance compared with their invariant counterparts. That is, for each robot action, the action can show variance on either speech or motion—not both modalities. An intent may translate to a combination of both speech and motion actions, but these will always be shown sequentially as two unimodal actions, and will not be combined to create a multimodal stimulus. For instance, when greeting the child, the robot would first say "Hello, my name is Zeno", which was followed by a wave. We consider this as a single robot action. For examples of behaviour variants, see Table A in Appendix A.

The choice for which variant to display for an action was determined through an algorithm. The action variant was chosen semi-randomly, where it would pick one of the four variants, excluding the antecedent variant for that action (if any). The latter was to prevent variants to be shown twice in a row. For actions with topic variance, a different selection mechanism needed to be used, as the variance for these actions related to whether the action is contingent or non-contingent on (internal or external) events in the environment. We therefore opted to use the Wizard-of-Oz paradigm, where another experimenter was controlling the robot (the wizard), without being visible to the child. The wizard was responsible for selecting the topic variant actions. To standardise the number of topic variance actions, the wizard would be notified through the wizard's control interface when it was time for the robot to perform a topic variant/invariant action. At this time, the wizard would look for one of the topic variant actions that was congruent with the condition the child was in. In the low-variance condition, the wizard would select topic variant actions that were a response to an observable event. For instance, the wizard could have the robot say "what was that noise?" in response to noise outside of the experiment room. For the high-variance condition, the wizard would ensure that the event the robot would respond to was not perceivable. In this case, the wizard would have the robot respond to noise when there was no noise to be heard.

4.4 Experimental Setup

The study took place at the children's school, in one of the offices. This room was converted into an experiment room. A picture of the experimental setup can be seen in Figure 2. The robot stood on a table facing the child and acted partly autonomously and was partly controlled by the wizard. This person was sitting in the same room behind a room divider and could view the interaction through a webcam. The wizard was responsible for selecting the correct task within the activity, selecting the topic variance actions, and for responding to any unscripted interactions through using a preset of robot actions such as having the robot say "no", "yes", and "I don't know". Within each task, the robot behaved autonomously, although the wizard could interrupt these behaviour at any time when the situation demanded it. The child would sit in front of the robot and next

Fig. 2. The experimental setup. The school's staff member who accompanied the child would sit behind the child in the corner of the room. The red circles outline the camera's that were used for annotating our engagement measures, which also included the camera from which this image was taken.

to the adult who was presenting the DE-ENIGMA activity. The recording equipment was placed behind and next to the table on which the robot stood. All computers and laptops were also placed behind the room divider and were managed by another researcher.

4.5 The DE-ENIGMA Activity

As we mentioned in Section 3, this study is complementary to another study. We use the same procedure and data collection that is presented in there. For clarity, we describe the DE-ENIGMA activity here as well, and the experiment procedure in the next subsection.

The participating autistic children engaged in a robot-assisted activity that was developed by the DE-ENIGMA project which revolves around playing with the basics of recognising facial expression (see [82] on how the activity was developed and assessed). In the current study, the activity was merely used to provide an interaction and structure the interactions between the child, robot, and the adult to ensure that the interactions in the two experimental conditions are comparable and offer similar opportunities for interaction. The activity itself was the same for both conditions.

The DE-ENIGMA activity starts with an introduction where the robot says, hello. Next, the robot would display various behaviours to attracts the child's attention, and to let the child get more familiar with the various motions and sounds the robot makes. For instance, the robot would show various facial expressions, do a dance, or play a nursery rhyme. After this introductory phase, the main body of the activity starts, namely playing several/all of the DE-ENIGMA games.

There were four games (see Figure 7 in Appendix B for a flow diagram and images of each of the DE-ENIGMA games), where each game builds on the previous game in complexity. In the first game, the children could explore the facial features (mouth, eyes, and eyebrows) of the robot. The tablet would show the robot's face and highlight the three facial features. When the child touches any of the features, the robot would move the features and label them. This way, children could freely explore the facial features that were of interest to them. In the second game, the robot would prompt the child to find one of the features and point them out using the tablet. When the child provided an incorrect answer, the robot would prompt them to try again. For correct answers, the robot would provide positive feedback and move the prompted facial feature. The third game was similar to the first game, only the child would instead explore facial expressions of emotions (happiness, sadness, anger, and fear). The robot would explain what and how the robot's facial features move for the selected emotion. Finally, for the fourth and final game, the robot would prompt the child to identify one of the four emotions in a similar fashion as the second game. Here the tablet would show four images of the robot, each with a facial expression of emotion. When the child had played with the robot for around 15 minutes, the robot would prompt the adult to end the session. After doing so, the robot would say goodbye.

Each game consisted of four opportunities for the child to make a selection, excluding instances where the child was prompted by the robot again. After the four opportunities the game was over, the tablet would display a blank screen, and the wizard would need to select the next course of action. There were also several "choice points" during the activity, where the child could choose a game or certain robot behaviours. These choice points allowed the adult to defer any requests for one of the games or certain behaviours to a choice point, rather than denying the child's request or interrupting the ongoing game. Moreover, by structuring the choice points and limiting them to two minutes, it allowed us to control it as much as possible with respect to following the same programme of content and the same order for each child. During the choice point, the child could choose from any content that they had already experienced (i.e., both games and generic robot actions). The child could make the choice through a "choice board", which contained icons of the available options. The icons can be added or removed from the choice board, as each was stuck on the board with velcro. The board and the icons on it were managed by the adult.

The games were designed to allow autistic children with limited receptive or limited expressive language to complete the games. For instance, the robot would use simple language where each speech action would consist of only a few words, and the tablet allowed the children to respond to the robot non-verbally.

4.6 Experiment Procedure

The autistic children would engage in the robot-assisted activity individually, once per day, for four to five sessions. The fifth session was only applicable to children who did not finish the activity in four sessions. Each session lasted around 15–20 minutes. The sessions were scheduled on consecutive school days as much as possible. Due to weekends and the children's schedule some variability was inevitable. For the low-variance condition there was an average of 0.36 days in between sessions (SD = 0.68), and for the high-variance condition this was 0.64 days (SD = 1.22).

Each child was assigned to one of three adults who would lead the sessions. The adult assigned to the child would also remain with that child for each of the child's sessions. The adult was tasked with augmenting the robot's instructions, supporting the children in using the tablet, giving feedback, and responding to their communicative overtures. Theirs was a supportive role, as the robot delivered the majority of the instructions and feedback. The children were often accompanied by a school staff member, who would sit in the back of the experiment room. They were asked not to participate in the activity unless they thought there was an issue such as when the child showed anxiety.

The content of each day of participation was scheduled to be delivered in a specific order and was identical across both conditions. When the children met the robot for the first time, the robot was covered by a blanket when they walked in the experiment room. The session would start with the robot being uncovered, introducing itself, and showing what movements it could do, some facial expressions, and what it sounded like. The goal here was to let the child get comfortable with the robot, what it looked like, and how it behaved. If the child liked any of the actions, they could be repeated by the wizard. When the child appeared comfortable, the adult would suggest to transition to the main activity—which the wizard then started—finishing the introduction. In subsequent sessions, the introduction was similar but shorter. After the introduction, the DE-ENIGMA activity described above would start. During the activity, the children would engage in the games that were set for that day, as well as in “choice points”, where the child could choose one of the games or specific robot behaviours they liked. As described in Schadenberg et al. [127], autistic children often make requests to the robot and denying the child’s request could disrupt the interaction. These choice points allowed the adult to defer any requests for one of the games or certain behaviours to a choice point, rather than denying the child’s request or interrupting the ongoing game. Moreover, by structuring the choice points and limiting them to two minutes, it allowed us to control it as much as possible with respect to following the same programme of content and the same order for each child. During the choice point, the child could choose from any content that they had already experienced. The child could make the choice through a “choice board”, which contained icons of the available options. This board and the icons on it were managed by the adult, and the wizard executed the child’s requests. For children who would or could not choose, the adult would first prompt the child by suggesting an activity or a robot action. If the child still did not express any preference, the adult ask the robot what to do next, upon which the wizard would select an activity. When the session was over, the robot would say goodbye, and the child went back to class.

4.7 Measures

4.7.1 Engagement Measures. We chose to operationalise and measure the autistic children’s engagement in terms of *behavioural engagement* and *visual attention* as these can be annotated as *patterns of manifest content*—child behaviours that are directly observable [108]. This is in contrast to scoring engagement holistically as *projective latent content* [108], as this would require the coder to employ subjective interpretations of the meaning of the behaviour, which is difficult without being familiar with how the participating children generally behave and understanding the meaning of their sometimes atypical and idiosyncratic behaviours.

Behavioural engagement. We measured behavioural engagement through expert annotations, using the coding scheme described in Table 2 on segments of 5 seconds. With this coding scheme, we make the distinction between autistic children being behaviourally engaged or disengaged with the HRI using a 5-point ordinal scale that denotes the amount of behavioural engagement.

While we believe that stimming behaviour—a self-stimulatory behaviour marked by a repetitive action or movement of the body—or fidgeting can be indicative of behavioural disengagement, it was problematic to annotate. Stimming behaviour provides sensory input for one of the senses, preventing the child from using this modality for engaging in the activity. However, whilst stimming, the child can still engage in the activity—and potentially learn—through using other modalities. For example, a child may be rubbing their hands while speaking to the adult or robot about the activity. Because we do not distinguish between modalities for annotating behavioural engagement, we decided to code all stimming behaviours that did not prevent the child from engaging in the activity and interaction were coded as passive. When the stimming was all consuming and prevented the child from engaging with the activity, we coded it as disengagement.

Visual attention. The experiment procedure and the activity did not allow us to directly measure cognitive engagement, which is why we opted to measure this through annotating the children’s visual attention—a proxy for cognitive engagement. Visual attention relates to the extent the participant paid overt visual attention to the ongoing activity and interaction, and tells us

Table 2. Coding Scheme for Annotating the Observed Behavioural Engagement of an Autistic Child

Level	*Meaning*	*Description*	*Example*
−2	Fully behaviourally disengaged	The child exhibits non-task related behaviour for most or all of the segment.	Child is playing around with objects that does not involve the task or interaction with the robot or adult. Child is engaging in stimming behaviours that prevent the child from partaking in the task. Child indicates wanting to stop with the interaction, e.g. through asking whether the game or session is finished, or turning the tablet to the adult prior to the game's conclusion. Talking with the adult about something unrelated to the activity or the ongoing triadic interaction, such as children saying they are hungry.
−1	Partly behaviourally disengaged	The child exhibits non-task related behaviour for some of the segment.	
0	Passive	Child does not behaviourally engage with the task, nor does the child show engagement in other activities.	Child is seemingly listening to the adult or robot, but does not communicate back. Child is looking at the task material, but does not physically interact with it. Also, echolalic and undirected vocalisations were are included on this level, as well as stimming behaviours that do not prevent the child from partaking in the task. Covering ears due to auditory sensitivity.
1	Partly behaviourally engaged	The child behaviourally engages with the task through interacting with the adult, with the robot directly or through the tablet, or other task materials for some of the segment.	

(Continued)

Table 2. Continued

Level	*Meaning*	*Description*	*Example*
2	Fully behaviourally engaged	The child behaviourally engages with the task through interacting with the adult, with the robot directly or through the tablet, or other task materials for most or all of the segment.	Pressing a button on the tablet, requesting facial expressions of the robot, choosing an activity, talking with the adult about the robot, dancing with the robot. Also includes non-verbal communication with adult, such as sharing enjoyment, or social references after the robot did something.

more on what or whom the children were engaged with. This was also measured through expert annotations by coding the children's gaze direction, but on segments of 2.5 seconds. From these annotations, we calculated the percentage of time spend looking at a certain direction for each session.

The coding scheme that we used included the following gaze directions: "robot", "tablet", "activity materials", "adult", "school staff member", "elsewhere", or "mixed". The latter was used to annotate instances where the child did not focus their gaze at any one point. The activity materials refers to any materials that were used in the activity, which primarily was the visual choice board for selecting the next task. And elsewhere refers to any gaze direction that was not in any of the other categories and thus *was not related to the activity*. Gaze direction such as the walls, the floor, the cameras, or the room divider. In contrast, the locations "robot", "tablet", "activity materials", and "adult" can be considered "activity-relevant locations". For our analysis of visual attention, we specifically look at the gaze directions towards the robot, and elsewhere, as our hypotheses relate to these locations. The other locations were annotated to enable better interpretation of the data.

The annotations were done at two levels. The first (primary) level represents the gaze direction where the child spent looking for the majority of the segment (1.25 seconds or more). The second (secondary) level was optional and could be used to annotate a secondary gaze direction during the segment which lasted at least 0.75 seconds, but no more than 1.25 seconds. This excludes brief glances, where the child's gaze would not stay on one direction.

4.7.2 Individual Factors. The complementary study also measured several individual differences between the autistic children. In the current study, we use a subset of those measures, namely the CARS-2 for measuring autistic features, a bespoke scale of expressive language, and an adapted version of the IUSC-parent report form [IUSC, 26].

The IUSC questionnaire was adapted from the complementary study to accommodate to children with limited spoken language, as we believed for many of them the IUSC questions were too difficult for parents to answer about their child. In the remainder of this article, we refer to this adapted version of the parent report of the IUSC as the *IUSC-S*. The adapted questions of the IUSC-S can be seen in Appendix C, Table 7, as well as principal component analysis of this adapted questionnaire. Based on this analysis, we excluded three questions for computing the IUSC-S scores.

The CARS-2 [129] is a 15-item autism screening and diagnostic tool and was administered to obtain a general measure of characteristics of autism. It was completed based on direct behaviour observation by a professional as well as reports from parents, teachers, or caretakers. The measure was completed by the adult who worked with the specific child. The total score on the CARS-2 reflects the severity of autistic features with scores of 15.0–29.5 indicating minimal-to-no evidence, 30.0–36.5 is mild-to-moderate severity, and 37.0 and higher is severe autistic features.

The bespoke scale of expressive language measures the spoken language ability of the child. The adult who gave the sessions rated the child's expressive language after the last session. This scale ranged from 0–4, where 0 means "no words", 1 is "some vocalisations or word approximations", 2 is "single words", 3 is "simple sentences" (two to three words), and 4 is "more complex speech, including complex sentences". The bespoke score reflects the level of expressive language that was generally used by the child during the sessions.

4.8 Annotation Procedures

As only a few children participated in five sessions, we only annotated the videos from the first four sessions. These sessions were annotated in terms of behavioural engagement levels and visual attention locations using the ELAN transcription software,[3] developed by the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands.

4.8.1 Behavioural Engagement Annotation. To determine whether a child is engaged or disengaged requires taking the situational context into account. For example, the child looking away may be part of the current activity, a response to the adult, or indicative of disengaging from the interaction. To preserve the situational context, we annotated 1 minute out of every 2 minutes, excluding only instances where there were technical problems with the system. We started the annotations from the moment the adult said hello to the robot and ended when the robot had said goodbye. This resulted in the annotation of 1,660 minutes of video recording, of which 848 minutes were in the low-variance condition and 812 minutes were in the high-variance condition. During initial testing of the coding scheme for behavioural engagement, we noted that the autistic children sometimes had brief behavioural disengagement episodes of a couple of seconds. We therefore, divided each minute into segments of 5 seconds, similar to Kim et al. [74] and Simpson et al. [132].

A single main coder annotated all the recordings, amounting to 9,947 segments. To calculate the inter-rater reliability of these annotations, a second coder annotated a randomly selected session for each participant. This amounted to 25% of the recordings being dual-coded, which also contained 25% of all the segments. To determine the agreement between the two coders, Cohen's κ statistic was used. There was good agreement between the two coders for behavioural engagement (Cohen's κ = .72, 95% CI [.70, .75], $p < .001$).

To gain insight into the nature of the coder disagreements, we inspected the confusion matrix for behavioural engagement (see Table 9 in Appendix D). The main coder annotated more instances of disengagement (approximately 14%), which were coded as passive by the secondary coder. Note that the main coder was not blind to conditions, but the secondary coder was. Given the agreement between the coders, we do not think it likely that this influenced our results. All in all, we deem these values high enough to continue our analysis on the basis of the main coder.

4.8.2 Visual Attention Annotation. For visual attention, we annotated the same minutes as for behavioural engagement. However, annotating segments of 5 seconds proved too long, as there were often more than three gaze directions. This made it difficult to code with our annotation scheme, which accounted for two gaze directions. The segments for visual attention therefore lasted 2.5 seconds to ensure that there were generally fewer than three gaze directions per segment.

[3] https://tla.mpi.nl/tools/tla-tools/elan/.

Again, a single main coder annotated all the recordings and second coder annotated a randomly selected session for each participant. The main and secondary coder were the same annotators as for the annotation of behavioural engagement. For visual attention, this amounted to a total of 19,855 segments. The dual-coding resulted into 25% of the segments being dual-coded. There was very good agreement for the primary gaze direction annotations (Cohen's κ = .87, 95% CI [.86, .89], $p < .001$) and good agreement for the additional, secondary gaze direction annotations on the (Cohen's κ = .69, 95% CI [.66, .73], $p < .001$). We deem these values high enough to continue our analysis on the basis of the main coder. Furthermore, inspection of the confusion matrices (see Tables 10 and 11 in Appendix D) showed no biases on the coder disagreements for one of the two coders.

4.9 Data Analysis

For the analysis of both behavioural engagement and visual attention, we used growth models with a maximum likelihood estimation method. These were modelled using the statistical program "R" [109], version 3.6.3, with the "nmle" package [106], and analysed with two-tailed tests and a 95% confidence level. Growth models allow us to take the multi-level nature into account, where the behavioural engagement scores/visual attention locations for each of the sessions are level-1 variables, the child is a level-2 variable, and the adult leading the session a level-3 variable. The adults were randomly assigned to the children regardless of the condition that was assigned to the child. The adult is therefore a crossed effect. For the visual attention analysis, we removed annotations where the child looked at the school staff member, as they were not always present in each session. Additionally, we removed annotations coded as "mixed", due to the uncertainty regarding the direction of the child's gaze.

As recommended by Raudenbush and Bryk [111], we first make a "basic" model and then add in variables as appropriate. In this article, we consider the model that includes random coefficients, the condition, and the session (time) as the basic (growth) model. We expected that the children will have different intercepts, as some children will be behaviourally more engaged than other children, or prefer looking at a certain gaze location more. Furthermore, we also expected that the slopes will be different between children, where one child loses interest faster than other children, resulting in difference in behavioural engagement and visual attention. The random coefficients account for these random effects in our models. Next, we further explored to what extent the individual factors improved the model fit of the basic conditional growth model. Finally, we investigated whether the different adults leading the sessions influenced the two facets of engagement of the children by adding the adult as a factor.

4.10 Manipulation Check

The study protocol had a flexible activity selection and session duration so as to support each autistic child's individual preferences. This means that the variability introduced by the robot differed per session and per participant. In turn, this means that we cannot be certain that the robot displayed high amount of variance in the high-variance condition, and vice versa for the low-variance condition. To check to what extent our manipulation of robot variance succeeded, we calculated the following variables:

(1) Average number of robot actions per minute. These should be similar between conditions, and serves as a baseline to put items (2) and (3) into perspective.
(2) Average number of unique robot actions per minute. These are unique *within* a session. Of the total number of robot actions, the high-variance conditions should have a more

unique robot actions per minute. This reflects the larger pool of unique actions that were implemented in the high-variance condition.

(3) Average number of novel robot actions per minute. These are actions that the child has not seen before in either the current or previous sessions. While the content of the interaction differed per session, there are robot actions that are used throughout the sessions, such as giving praise. Therefore, this number should drop over sessions in both conditions, but the high-variance condition should introduce more novel actions per minute than the low-variance condition in all sessions.

(4) The cumulative average number of repetitions per robot action. This number reflects how often the robot displayed a certain robot action, either in the current session or previous sessions. The higher the number, the more opportunities the child had to learn to predict this action. In the high-variance condition, this large amount of unique actions should result in fewer repetitions per action than in the low-variance condition.

For the manipulation check, we consider all *variants* of robot actions as being unique actions. For the variables that were calculated per minute, we excluded the time when there was a technical difficulty. To assess to what extent there is a difference between the two conditions on each of these four items, we conducted four mixed ANOVAs, where the condition is a between-subject variable and session a within-subject variable. Furthermore, we assessed each outlier and consider to remove them from the analysis. This was done by placing the outlier in context of what happened during the session as well as relating the value of the outlier to the values of the other condition.

5 RESULTS

5.1 Manipulation Check

On average the children engaged in the DE-ENIGMA activity for 16 minutes and 13 seconds (SD = 3 min, 1 s) in the low-variance condition. For the high-variance condition, the average was 15 minutes and 39 seconds (SD = 1 min, 23 s). There was no significant difference in the time spent in the activity between the conditions ($F(1, 22) = 3.30$, $p = .08$, $\eta_p^2 = .13$, η_p^2 90% CI[.00, .34]), over sessions ($F(3, 66) = 2.55$, $p = .06$, $\eta_p^2 = .10$, η_p^2 90% CI[.00, .20]), nor an interaction effect ($F(3, 66) = 0.22$, $p = .88$, $\eta_p^2 = .01$, η_p^2 90% CI[.00, .03]). Note that the variables reported below are all normalised to account for differences between children and sessions in the time that the session lasted.

The extent to which there was variance in the robot's behaviour per session can be seen in Figure 3 for both conditions. The average number of robot actions per minute for the low-variance condition was 5.43 (SD = 0.55) and 5.36 (SD = 0.82) for the high-variance condition. There was no significant difference in the average number of robot action per minute over sessions ($F(3, 66) = 0.58$, $p = .632$, $\eta_p^2 = .03$, η_p^2 90% CI[.00, .07]), between conditions ($F(1, 22) = 0.09$, $p = .763$, $\eta_p^2 < .01$, η_p^2 90% CI[.00, .12]), or an interaction effect ($F(3, 66) = 0.91$, $p = .441$, $\eta_p^2 = .04$, η_p^2 90% CI[.00, .10]). As expected, there was a significant difference between conditions in the average number of *unique* robot actions per minute ($F(1, 22) = 136.78$, $p < .001$, $\eta_p^2 = .86$, η_p^2 90% CI[.74, .90]), where this number was significantly higher in the high-variance condition (M = 3.80, SD = 0.63) than in the low-variance condition (M = 1.80, SD = 0.30). Similarly, the difference between conditions for the average number of *novel* robot action per minute was significant ($F(1, 22) = 363.44$, $p < .001$, $\eta_p^2 = .94$, η_p^2 90% CI[.89, .96]). The robot in the high-variance condition displayed a higher number of novel actions per minute than in the low-variance condition for each of the four sessions. For the average number of repetitions per action there was also a significant difference between the conditions ($F(1, 22) = 406.46$, $p < .001$, $\eta_p^2 = .95$, η_p^2 90% CI[.90, .96]). For each session, the high-variance condition had fewer repetitions per action than in the low-variance condition. Based on these tests, we conclude

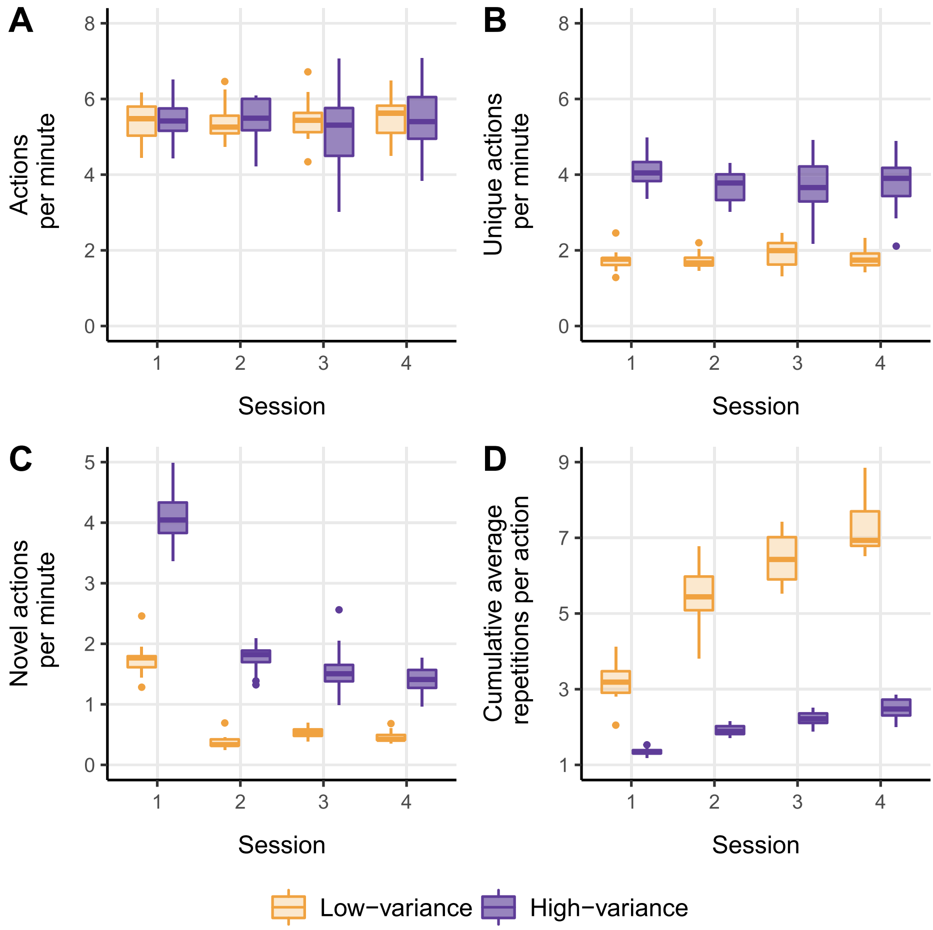

Fig. 3. Boxplots that show to what extent variance in the robot's behaviour was achieved. **(A)** shows the average number of actions the robot performed per minute, **(B)** shows the average number of unique actions performed by the robot actions per minute within a session, **(C)** shows the average number of actions that the robot performed for the first time per minute, and **(D)** shows the cumulative average number of repetitions per robot action.

that the robot's variance was significantly different between the conditions, as well as sufficiently large given the reported effect sizes.

While we consider the manipulation of the extent of variance in the robot's behaviour successful, there were four sessions of three children where the variance of the robot's behaviour was more akin to that of the other condition. In the low-variance condition, the first session of one child was only partially recorded in our logfiles due to technical issues. The data loss resulted in a shorter session recording, which in turn resulted in a higher number of unique and novel actions per minute, as well as fewer repetitions per action. Nonetheless, it was a regular session where the child was engaged for most of the time. It is therefore likely that the system performed as intended and produced low variance in the robot's behaviour. In the high-variance condition, two children had sessions where the robot showed few unique actions. For one child, this was the case in sessions 3 and 4, while for the other it was only for session 3. Note that for session 3, these instances are not statistical outliers and therefore are not shown as such in Figure 3. For all three

Table 3. Model Fit Measures for Each Model as They Increase in Complexity, as well as Chi-Square Statistic with 1 Degree of Freedom and Statistical Significance

Model	*df*	*AIC*	*BIC*	*Log-Likelihood*	χ^2	*p*
Basic growth models						
(A) Conditional	9	26.46	49.54	−4.23	-	-
(B) Conditional with quadratic trend	9	21.46	44.54	−1.73	-	-
(C) Three-level	12	27.45	58.22	−1.72	0.01	.999
Model B *with covariate*						
(D) CARS-2 score	10	13.86	39.50	3.07	9.60	.002
(E) Exp. lang. score	10	14.87	40.52	2.56	8.59	.003
(F) IUSC-S score	10	23.43	49.08	− 1.72	0.03	.867
Model D *with another covariate*						
(G) CARS-2 score and exp. lang. score	11	12.05	40.26	4.98	3.81	.051
Model G *with interaction between covariate and condition*						
(H) CARS-2 score * condition	12	14.05	44.82	4.98	<0.01	.961
(I) Exp. lang. score * condition	12	11.08	41.85	6.46	2.97	.085
(J) IUSC-S score * condition	13	15.14	48.48	5.43	0.91	.634

Each model is compared with the first model with lower complexity as reflected by the model's degrees of freedom. For the models with one or two covariates, the models are compared, respectively, to Model B and Model D. The models with an additional interaction effect between the covariate and the condition are compared with Model G.

sessions, the child was disengaged for most of the session, limiting the progression through the content. In turn, the robot only displayed a limited repertoire of its behaviours, which increased the chance of displaying actions that had already been displayed before. Given that this happened in the last two sessions, we cannot exclude that this happened due to their experiences in the first and second session. Moreover, the mean number of novel actions per minute and repetitions per action are in line with the high-variance condition. Therefore, we include these sessions in the main analysis.

5.2 Behavioural Engagement

The parameter estimates for each of the multi-level models are presented in Table 3. First, we fitted a basic conditional growth model (Model A) with random slopes and intercept. As fixed effects, this model contains the session, condition, and an interaction effect between the two. For the covariance structure, we used a first-order autoregressive covariance structure. To further explore the trend of behavioural engagement over sessions, we fitted a quadratic and cubic trend instead of the linear trend. The quadratic trend best fitted the change in behavioural engagement over sessions (Model B). Next, we investigated whether accounting for the adult who was leading the session improved the model (Model C). This required a three-level model, where the adult is a crossed effect. Compared with Model B, the three-level model did not significantly improve the model fit ($\chi^2(1) = 0.01$, $p = .999$). Thus, while the children received their sessions from one of three adults, this does not explain the variance in their behavioural engagement.

Next, we investigated whether individual differences in characteristics of the children could explain the variance in behavioural engagement. We took Model B, as it had the best model fit, and investigated whether the CARS-2 score, expressive language score, or IUSC-S score, improved the model fit. Adding the CARS-2 score as covariate significantly improved Model B ($\chi^2(1) = 9.60$,

Table 4. Parameter Estimates for the Growth Model on Behavioural Engagement with the CARS-2 Score (Model D), Bespoke Expressive Language (Exp. Lang.) Score (Model E), or both CARS-2 and Bespoke Expressive Language Score as Covariate (Model G)

	Model D: *CARS-2*		Model E: *Exp. Lang.*		Model G: *CARS-2 + Exp. Lang.*	
Variable	*b (SE)*	*95% CI*	*b (SE)*	*95% CI*	*b (SE)*	*95% CI*
Fixed effects						
Intercept	1.59 (0.33)	0.95, 1.59	−0.09 (0.20)	−0.47, 0.29	0.89 (0.41)	0.09, 1.69
Session^2	−0.01 (0.01)	−0.03, 0.00	−0.01 (0.01)	−0.03, 0.00	−0.01 (0.01)	−0.03, 0.00
Condition	−0.09 (0.14)	−0.37, 0.19	−0.08 (0.13)	−0.34, 0.18	−0.05 (0.12)	−0.31, 0.20
Ses^2:Cond	−0.01 (0.01)	−0.03, 0.01	−0.01 (0.01)	−0.03, 0.01	−0.01 (0.01)	−0.03, 0.01
CARS-2	**−0.04** (0.01)	−0.06, −0.02	-	-	**−0.03** (0.01)	−0.05, −0.01
Exp. Lang.	-	-	**0.20** (0.06)	0.07, 0.32	**0.13** (0.06)	0.01, 0.25
	SD	*95% CI*	*SD*	*95% CI*	*SD*	*95% CI*
Random effects						
Intercept	0.41	0.30, 0.56	0.33	0.24, 0.46	0.35	0.25, 0.50
Session	0.12	0.09, 0.17	0.12	0.09, 0.17	0.12	0.09, 0.17

Model D: *Behavioural engagement ~ Session^2 * Condition + CARS-2 + (Session | Participant).*
Model E: *Behavioural engagement ~ Session^2 * Condition + Exp. Lang. + (Session | Participant).*
Model G: *Behavioural engagement ~ Session^2 * Condition + CARS-2 + Exp. Lang. + (Session | Participant).*
The statistically significant parameter estimates for the fixed effects are in bold, excluding the intercept.

p = .002). Similarly, the expressive language score for expressive language significantly improved Model B ($\chi^2(1)$ = 8.59, p = .003). Adding the IUSC-S did not significantly improve Model B ($\chi^2(1)$ = 0.03, p = .867).

Based on the model fit of the models reported above, and whether they significantly improved the model fit, we consider Model D with the CARS-2 score as covariate, and Model E, which has the expressive language score as covariate, as the models that best explain the variance in behavioural engagement. However, to understand to what extent the CARS-2 score and the expressive language score explain the same variance in behavioural engagement, we fitted a model using both covariates (Model G). This did not significantly improve the model fit compared to the best model with only one covariate (Model D). Furthermore, while this model has a better fit than Model D, it is also more complex, as reflected by a higher BIC.

To investigate whether the covariates actually moderated the effect of the condition (robot predictability) on behavioural engagement, we further added an interaction effect between the covariate and the condition to Model G. For the CARS-2 score, this did not significantly improve the model fit of Model G ($\chi^2(1) < 0.01$, p = .961). Nor was the model fit improved when adding an interaction effect for the expressive language score ($\chi^2(1)$ = 2.97, p = .085), or the IUSC-S score ($\chi^2(1)$ = 0.91, p = .634).

The model parameters of all Model D, E, and G can be seen in Table 4. For Model D, the CARS-2 score was significant ($t(21) = -3.52$, p = .002). The condition was not significant ($t(21) = -0.64$, p = .529), nor was session^2 ($t(70) = -1.78$, p = .079), or the interaction between the condition and session^2 ($t(70) = -1.18$, p = .243). The relationship between the two conditions and behavioural engagement showed significant variance in intercepts across the children. In addition, the slopes significantly varied across children, and the slopes and intercepts were negatively and significantly correlated ($r = -.73$, 95% CI[−.88, −.45]).

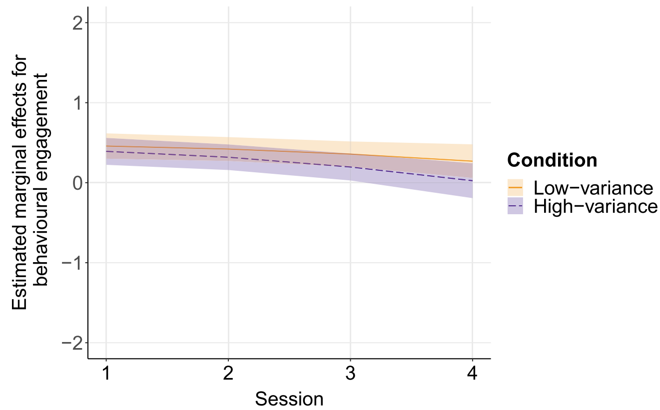

Fig. 4. The predicted values (marginal effects) for behavioural engagement of growth Model G, presented in Table 4.

The parameter estimates for Model E, including the expressive language score, show a similar trend to Model D. The covariate, expressive language score, was significant ($t(21) = 3.25$, $p = .004$). But neither the condition ($t(21) = -0.64$, $p = .530$), session^2 ($t(70) = -1.78$, $p = .079$), nor the interaction effect between the condition and session^2 was significant ($t(70) = -1.18$, $p = .243$). The random intercept and slopes showed a significant and negative correlation ($r = -.52$, 95% CI[−.78, −.12]).

The marginal means for behavioural engagement, estimated by Model G, can be seen in Figure 4. Model G, which includes the CARS-2 and expressive language score as covariates, shows that both covariates significantly contribute to predicting the variance in the behavioural engagement of the children. For the CARS-2 the parameter estimate is −0.03 ($t(20) = -2.51$, $p = .021$). This means that the model estimates that autistic children who scored higher on the CARS-2 were less behaviourally engaged. For the expressive language score, the parameter estimate is 0.13 ($t(20) = 2.18$, $p = .042$), which means that autistic children with more complex expressive language, with scores ranging from 0 to 4, were also more behaviourally engaged than those with less complex expressive language. The model shows that the condition was not significant, nor did the condition interact with the sessions.

5.3 Visual Attention

The visual attention over the sessions can be seen in Figure 5. As our hypotheses only relate to the "robot" and "elsewhere" gaze direction, we will not report on the other annotated gaze directions. Again, we fitted multi-level models, using random slopes, random intercepts, and first-order autoregressive covariance structure, to model the participants' visual attention with the robot and their visual attention elsewhere. The parameter estimates for the conditional growth model on visual attention *towards the robot* can be seen in Table 5. The relationship between the robot's variance and visual attention with the robot showed significant variance in intercepts across the children, but the slopes were non-significant. We investigated whether accounting for the differences between children, as measured by the individual factors, improved the model. In contrast to the previous results on behavioural engagement, this was not the case for any of the measures. The model that best fits the data is therefore a conditional growth model with random intercepts. The visual attention towards the robot significantly decreased over sessions ($t(70) = -2.91$, $p = .005$). This could be due to other factors that influence visual attention, such as a novelty effect that wears off, or boredom. There was no effect of condition ($t(22) = 0.56$, $p = .579$), nor was there an interaction effect between condition and session ($t(70) = -1.24$, $p = .219$).

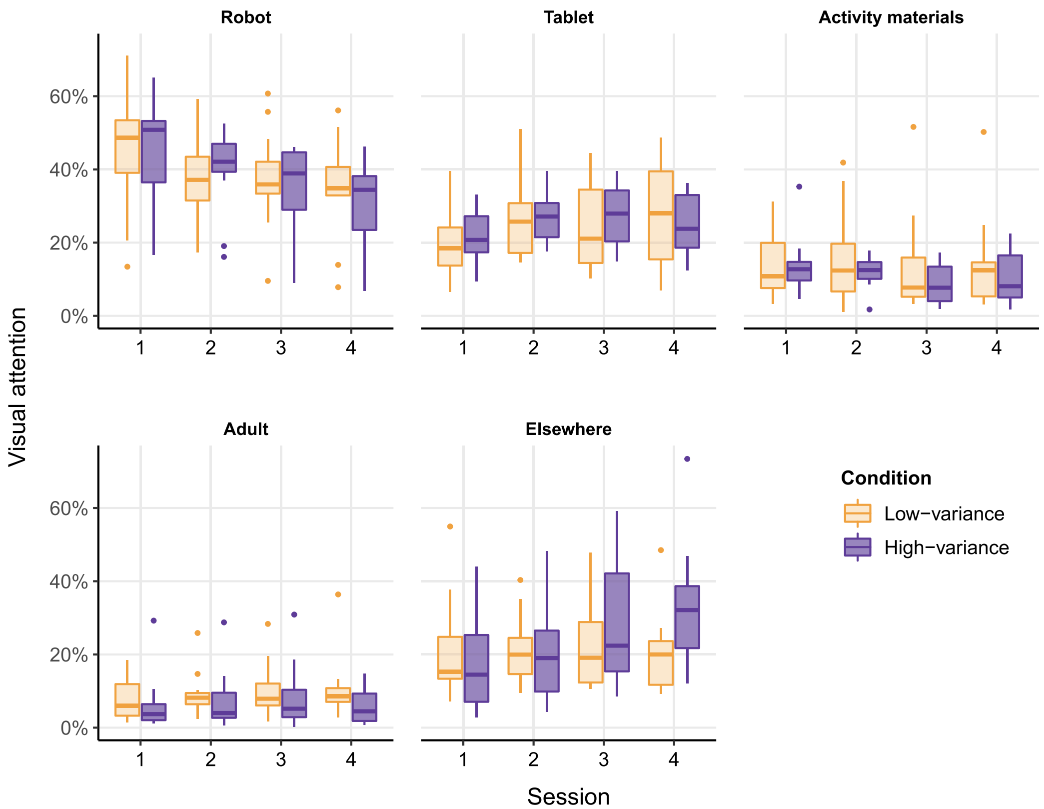

Fig. 5. The (unmodelled) percentage of time a participant spent looking at each of the annotated directions for each of the sessions (visual attention) and each of the conditions.

The model parameters for the children's visual attention *elsewhere* can also be seen in Table 5. We found no significant variance in the slopes across children. The variance in intercepts did significantly differ. Again, accounting for differences on the autism-specific measures did not improve the model. Therefore, the data was best described by a conditional growth model with random intercept. The visual attention elsewhere did not significantly decrease over sessions ($t(70) = -0.12$, $p = .908$), nor was there a significant difference between conditions ($t(22) = -1.75$, $p = .095$). There was, however, a significant interaction effect between the condition and session ($t(70) = 3.90$, $p < .001$). In the high-variance condition, children looked increasingly "elsewhere" over sessions compared with the low-variance condition. Thus, in contrast to the results for behavioural engagement, this shows an impact of robot predictability on the children's visual attention—which is indicative of engagement—to the robot-assisted activity.

6 DISCUSSION

The goal of our study was to investigate the interplay between robot predictability, the behavioural engagement and visual attention to the activity, and the idiosyncrasies therein between autistic children. To that end, we manipulated the variance in the robot's behaviour as a way to operationalise predictability, and measured the children's behavioural engagement and visual attention in a robot-assisted activity, as well as individual factors.

Table 5. Parameter Estimates for the Conditional Growth Models for Visual Attention towards the Robot and Elsewhere

	Towards the robot		Elsewhere	
Variable	*b (SE)*	*95% CI*	*b (SE)*	*95% CI*
Fixed effects				
Intercept	0.46 (0.04)	0.38, 0.55	0.22 (0.04)	0.14, 0.29
Session	**−0.03** (0.01)	−0.05, −0.01	−0.00 (0.01)	−0.02, 0.02
Condition	0.03 (0.06)	−0.09, 0.15	−0.10 (0.05)	−0.21, 0.00
Session:Condition	−0.02 (0.01)	−0.05, 0.01	**0.05** (0.01)	0.03, 0.08
	SD	*95% CI*	*SD*	*95% CI*
Random effects				
Intercept	0.10	0.07, 0.14	0.10	0.08, 0.15

Visual attention towards the robot ~ *Session * Condition + (1 | Participant).*

Visual attention elsewhere ~ *Session * Condition + (1 | Participant).*

The statistically significant parameter estimates for the fixed effects are in bold, excluding the intercept.

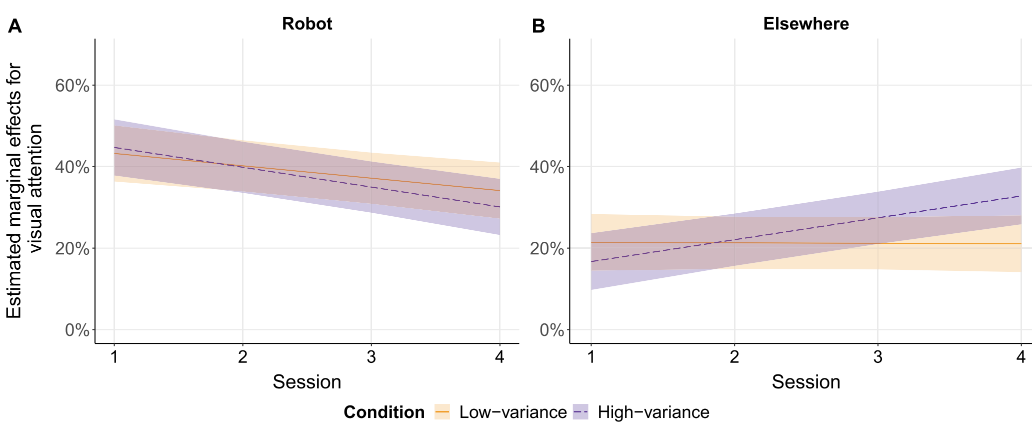

Fig. 6. The predicted values (marginal effects) for the visual attention towards the robot **(A)** and elsewhere **(B)** of the growth models presented in Table 5.

We found that for the less predictable robot, autistic children paid less *visual attention* to activity-relevant locations as the sessions progressed. Rather, the children started to pay more attention to locations where nothing moves (annotated as "elsewhere"), such as the walls, the cameras, or the room divider. In contrast to our predictions, however, the robot's predictability did not impact on *behavioural engagement*. The children continued to engage with the robot-assisted activity regardless of the robot's predictability. We did find, though, that *individual differences* in children's background characteristics were related to their behavioural engagement. Higher autistic features were related to less behavioural engagement of autistic children, while the children's expressive language ability was related to greater behavioural engagement. Visual attention, on the other hand, was not influenced by any of the individual differences measured. Finally, we found no evidence for a relation between the autistic children's IU and their response to the robot's predictability in terms of their behavioural engagement and visual attention.

In conclusion, it appears that the children continued engaging in the robot-assisted activity, but started to pay less visual attention over time to the activity-relevant locations when the robot was less predictable. Instead, the children started to pay more attention to locations that visually do not change (i.e., no changes in sensory information), such as the walls, or the floor—they look away from the activity. We believe this might be a coping strategy to minimise sensory input and deal with anxiety resulting from the inability to learn to accurately predict the robot's actions. Similarly to how stimming can be used by autistic children to generate predictable sensory information to deal with an overload of unpredictability [135]. For learning, paying less visual attention to the activity-relevant locations is problematic as it indicates that the child is less strongly engaged with learning tasks. In particular, this may impact long-term use of robot-assisted interventions, as in our study this effect got stronger over sessions. We may also expect that eventually the children may start to be less *behaviourally* engaged, when the encouragement from the adult starts to fail in motivating the child to engage with the robot and deal with the resulting unpredictability. However, as we did not measure the children's learning, we cannot draw a firm conclusion about whether the diminished visual attention also impacts the learning.

The result that individual differences influence the behavioural engagement of autistic children is in line with previous studies that found correlations between autistic features, measured through the CARS-2, and behavioural engagement [117, 127]. Similarly, Kostrubiec and Kruck [78] found correlations between autistic features, measured via the SCQ, and the proportion of prosocial behaviours in a robot-assisted intervention. Not only do individual differences seem to influence to what degree autistic children are behaviourally engaged in activities with robots, they are also correlated with how they behaviourally engage [127]. Researchers often report increased engagement, increased levels of attention, and novel social behaviours when incorporating a robot in the interaction [38, 113, 121, 122]. For such studies, it is then important to account for the individual differences and report these. But the relationship between the children's individual differences and behavioural engagement also raises a new question: Do the children's individual differences only influence their behavioural engagement directly, or do they *moderate* the relationship between the effect of the robot on behavioural engagement. Understanding this relationship would allow us to better determine which autistic children may benefit in particular from such robot activities or interventions. Lastly, these findings also suggest that the way autistic children interact with, or in the presence of, robots could potentially be indicative of their autistic features. This warrants further research, but might be particularly interesting for robots that are used for diagnosing autism.

Whether certain individual factors also moderate the effect of robot predictability on engagement remains unanswered by this study, as we found no evidence that this was the case. In particular, we had expected that those higher in IU would also be more strongly affected by the robot's predictability, given the similarity between IU and unpredictability. While other studies did find relationships between IU and autistic features, such as the presence of sensory sensitivities [95], repetitive motor behaviours, and insistence on sameness [145], we found no results in our study that support a relationship between IU and predictability. Similarly, in an earlier study with (typically developing) adults where we manipulated predictability only in terms of topic variance, we found no relationship either between IU and the social perception of the robot in terms of warmth, competence, and discomfort [128]. While unpredictability and uncertainty are often used interchangeably, suggesting conceptual similarity [56], our results suggest that greater caution is warranted when doing so. In the current study, we defined and manipulated robot predictability in terms of making it more or less difficult to learn to predict the robot's behaviour. However, the questionnaires on IU are more about uncertainty regarding events further into the future than the robot's next action(s) would be. The children knew that they were going to interact with "Zeno the robot" at a certain point during the day, that they would then play several games with the robot, and then go back to class—a fixed structure. In that sense, the children could predict the robot's behaviour on a more abstract level (i.e., the robot will perform actions related to the games), but not on precisely when, where, and how those robot actions would appear. Possibly, IU relates to an intolerance for situations without a clear structure on what is going to happen in a longer time span, resulting in the unpredictability of future events. Predictability, as we defined and manipulated it, related more to predicting sensory information in the immediate future. This distinction may also be relevant when considering the role of predictability and uncertainty in the non-social features of ASC, in particular in insistence of sameness (e.g., inflexible adherence to routines, or ritualised patterns) given that both IU and insistence of sameness refer to liking things to be predictable and a dislike of change [21].

To our knowledge, this is the first study that has operationalised and manipulated predictability in a real-world setting and showed how the extent of unpredictability can be quantified. The metrics in our manipulation check can be used to compare the degree of unpredictability between studies using robots. In general, robotic technology is uniquely position for investigating predictability in autistic children, as they allow us to carefully manipulate its predictability (unlike with humans), but they also elicit social interactions in autistic children. There is some preliminary work in trying to teach autistic children to deal with unpredictability [e.g., 58, 115]. To this end, robots may be particularly useful in that its unpredictability can be carefully increased in both intensity as well as in predicting different aspects of the environment (e.g., predicting robot actions, or predicting future events).

6.1 Limitations

In our study, we manipulated the predictability of the robot's behaviour through its variance. This ought to have made it more difficult to learn to predict the robot's behaviour. Through our manipulation check, we concluded that this manipulation was successful. However, possibly even in the high-variance condition, the robot's behaviour was not *problematically* unpredictable, as is the unpredictability of human behaviour for example. In practice, robots can be more unpredictable than we could manipulate in our study, as we needed to keep the conditions comparable and avoid confounding factors. In the future, robots will become more sophisticated and capable of levels closer to humanlike behaviour. In turn, so too does their ability to be unpredictable. Thus, robot predictability could affect the engagement of autistic children more strongly when they become more sophisticated and humanlike. In our study, we looked at *the ability to predict the robot's actions*, which is not the same as *the extent which one perceives the robot to be predictable* (attributed predictability) [128]. Even though a robot's behaviour is more difficult to predict, it can still be considered to be more predictable [41, 128]. Current measures for attributed predictability are too complex to be used for autistic children in our study, as they rely heavily on language and intellectual ability. When new measures become available, it would be interesting to assess the autistic children's attributed predictability in relation to different levels of variance in the robot's behaviour.

Note that there are several sources of variance that could not be controlled fully in our study. First, there is inherent variation in robot motion due to limits in reproducibility of motion by the robot's stepper motors, which may differ slightly with each operation. Second, we were not able to control for the affective quality and the attractiveness for engagement of each specific variant we chose, despite the fact that we selected ones with similar verbal and/or motion qualities. For example, children may have experienced more negative affect when the robot used the specific term "Time for dancing", which was present only in the high-variance condition. Additionally, for some sessions, technical difficulties resulted in unintended behaviours and form of behaviours, introducing some variability. These instances were not annotated, but the effects could possibly carry over to later in the session.

Finally, we also note that, as with most studies that concern robots and autistic children (see [12, 38]), our sample size was relatively low. This can negatively influence the generalisability of our results and prevents us from drawing strong conclusions. It is therefore important to take the reported margins of error into account when interpreting our results.

7 CONCLUSION

Predictability is important to autistic individuals, and robots have been suggested to meet this need as they can be programmed to be predictable, as well as to elicit social interaction. However, little was known about the interplay between robot predictability, engagement in learning, and the individual differences between autistic children. Here, we systematically manipulated the robot's predictability, and measured the behavioural and visual attention of the autistic children. Additionally, we also measured several individual factors, including the children's autistic features, expressive language ability, and IU. We found that the children will continue engaging in the activity behaviourally, but start to pay less visual attention over time to activity-relevant locations when the robot is less predictable. Instead, they increasingly start to look away from the activity. Ultimately, this could negatively influence learning. In particular for tasks with a visual component, where paying less visual attention leads to fewer opportunities for learning. Furthermore, we found that the severity of autistic features and expressive language ability had a significant impact on behavioural engagement. This finding is relevant for robots used for diagnosing autism and raises the question whether individual differences only directly influence behavioural engagement or whether they moderate the effect of a robot on the children's behavioural engagement.

We consider our results as preliminary evidence that robot predictability is an important factor for keeping children in a state where learning can occur. In particular, in long-term interactions with many sessions, our results indicate that the trend of paying less visual attention increases over time. As individual differences between autistic children were shown to have a significant impact on behavioural engagement, future studies should therefore be careful to account for these differences. Finally, our study indicates that predictability can be studied in real-life scenarios with real stimuli, rather than artificial stimuli in lab settings. We also showed how the degree of predictability can be quantified in a way that it can be used as a manipulation check to display the degree of unpredictability between conditions in a real-world setting.

Once the engagement of the children with the robot has been further clarified, future research should consider looking into the affective component of engagement to investigate whether a more predictable robot is more enjoyable to the children. Additionally, future research is needed to determine whether "higher quality" engagement also leads to more and faster learning. After all, increased engagement alone is insufficient to justify robots to assist in interventions for autistic children when it does not also lead to increased learning. In our study, we opted for a holistic approach to increasing the robot's unpredictability by implementing several types of variance. Future research should examine to what extent each types of variance influences autistic children. This would allow us to more carefully take the robot's predictability into account when designing its behaviour. Lastly, the goal of our study was to investigate how we should design robots to best aid autistic children in learning, specifically focusing on their need for predictable environments whilst taking the children's idiosyncrasies into account. Future research could also investigate the role of a robot's predictability in engaging typically-developing children in learning, in order to assess whether or not predictability is uniquely important to autistic children.

APPENDICES

A EXAMPLES OF THE VARIATIONS OF THE ROBOT'S BEHAVIOURS

Table 6. Some Examples of the Variations for Certain Robot Actions

Description	*Variant one (Default)*	*Variant two*	*Variant three*	*Variant four*
Introduction actions				
Greeting	"Hi, my name is Zeno." + wave with right arm	"Hello. I am Zeno the robot." + wave with left arm	"I am Zeno. Hello!" + wave with right arm	"Hi, I am Zeno. + wave with left arm"
Game actions				
Prompt eyes	"Find eyes."	"Choosing eyes."	"Where are eyes?"	"Find my eyes."
Resolve eyes	"You found my eyes!" + eyes open then close	"You found them!" + eyes left then right	"That's right" + eyes open then close	"You found... eyes!" + eyes left then right
Generic actions				
Praise	"Good job!"	"Well done!"	"Good working!"	"Excellent!"
Unsure	"I don't know"	"Sorry, don't know"	"Maybe?"	"Not sure"

Topic variant actions

	Low-variance condition		*High-variance condition*	
Description	*Triggering event*	*Action*	*Triggering event*	*Action*
Respond to noise	There was noise	"What's that noise?"	There was no noise	"What's that noise?"
Respond to child's presence	Child is out of view of the robot	"Where are you?"	Child is sitting in front of the robot	"Where are you?"
Respond to WiFi connection lost	-	-	Robot loses WiFi connection (unobservable internal event)	"Reconnecting. Reconnecting. Success."

In the low-variance condition, the robot would only perform the default variant of an action. In the high-variance condition, the robot would select one of the four variants of an action. For the topic variant actions, there is only one variant. These actions were triggered by the wizard, who made sure that the actions were a response to an observable event in the low-variance condition. In the high-variance condition, there was no observable event that could explain the action. There were also more types of topic variant actions in the high-variance condition, as some of those actions were plausible responses to internal events. The child had no way of observing these internal events, thus they only occurred in the high-variance condition.

B FLOW DIAGRAM OF THE DE-ENIGMA GAMES

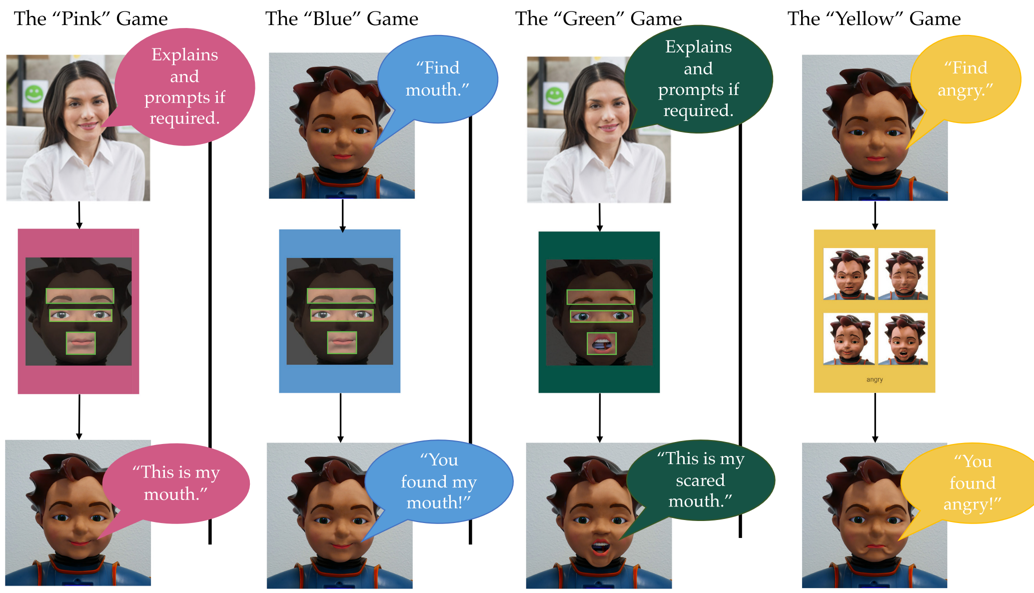

Fig. 7. Flow diagram of the four DE-ENIGMA games. First, the robot or adult would prompt the child with a question or to explore the options that were shown on the tablet. The child could then select one of the options, which was highlighted with the green square, or one of the four images in the Yellow Game. After selecting an option, the highlights on the tablet are removed. The robot then responded to the child's option by labelling the chosen option and moving the respective facial feature, or displaying the facial expression. In case of the Blue and Yellow Game, the robot would also evaluate the child's answer and provide positive feedback. Each of the sequences shown were repeated four times for each game (with different facial features/facial expressions).

C IUSC-S QUESTIONNAIRE

The IUSC-S questionniare can be seen in Table 7. To assess how many dimensions of intolerance of uncertainty the IUSC-S measured, we conducted a Principle Component Analysis (PCA) on 15 items with varimax rotation. The Kaiser-Meyer-Olkin (KMO) measure verified the sampling adequacy for the analysis (KMO = .65), and all KMO values for individual items were >.55. Given our sample size of 24, this is sufficient, but mediocre. Bartlett's test of sphericity indicated that correlations between items were sufficiently large for PCA ($\chi^2(105) = 312.61$, $p < .001$). Two components had eigenvalues above Kaiser's criterion of 1 and together explained 70.85% of the variance. Table 8 shows the component loadings for each item, where we highlighted items with loadings of .72 or higher and with a cross loading difference greater than .2, based on recommendations from Stevens [136]. Based on these component loadings, we exclude item 2, 7, and 11 for computing the IUSC-S scores.

Table 7. The IUSC-S, Parent-Report form, Adapted from the Complementary Study

Questions
1. Uncertainty makes my child's life intolerable.
2. My child's mind cannot be relaxed if he/she does not know what will happen tomorrow.
3. Uncertainty makes my child uneasy, anxious, or stressed.
4. Unforeseen events upset my child greatly.
5. It frustrates my child to not have all the information he/she needs in a situation.
6. Uncertainty keeps my child from living a full life.
7. When it's time to act, uncertainty paralyzes my child.
8. When my child is uncertain he/she cannot function very well.
9. Other children seem to be more certain than my child.
10. Uncertainty makes my child unhappy or sad.
11. My child always wants to know what the future has in store for him/her.
12. My child cannot stand being taken by surprise.
13. Uncertainty keeps my child from sleeping soundly.
14. My child tries to get away from all uncertain situations.
15. The ambiguities of life stress my child.

Table 8. The Component Loadings for the Items on The IU as Measured by the IUSC-S

Item	*Component 1*	*Component 2*
Item 6	**.90**	−.12
Item 4	**.87**	.20
Item 15	**.87**	−.09
Item 3	**.86**	−.25
Item 1	**.86**	−.12
Item 12	**.84**	.02
Item 10	**.84**	−.23
Item 5	**.79**	−.27
Item 14	**.78**	.27
Item 13	**.77**	.17
Item 9	**.74**	−.17
Item 8	**.73**	.11
Item 7	.70	.55
Item 11	.58	.58
Item 2	.48	−.63
Eigenvalue	9.16	1.47
% of variance	61.08	9.77
Cronbach's α	.96	.34

The component loadings that meet the recommendations of Stevens [137] are in bold.

D CONFUSION MATRICES FOR BEHAVIOURAL ENGAGEMENT AND VISUAL ATTENTION

Table 9. Confusion Matrix of the Annotations for *Behavioural* Engagement between the Main Coder and Secondary Coder

	Secondary coder					
Main coder	*−2*	*−1*	*0*	*1*	*2*	*Total*
−2	**188**	47	28	14	4	281
−1	6	**105**	53	15	1	180
0	7	20	**706**	60	2	795
1	9	17	79	**796**	57	958
2	2	4	3	60	**198**	267
Total	212	193	869	945	262	**2,481**

The number of annotations that both coders annotated similarly are in bold.

Table 10. Confusion Matrix of the *Primary* Annotations for *Visual Attention* between the Main Coder and Secondary Coder

	Secondary coder							
Main coder	*Robot*	*Tablet*	*Teaching Mats.*	*Adult*	*Assistant*	*Elsewhere*	*Mixed*	*Total*
Robot	**1,599**	23	10	16	0	60	27	1,735
Tablet	19	**1,070**	5	7	0	24	12	1,137
Teaching Mats.	7	2	**534**	5	0	13	4	565
Adult	12	8	6	**373**	0	28	19	446
Assistant	1	0	0	0	**10**	2	4	17
Elsewhere	28	40	11	16	0	**904**	19	1,018
Mixed	11	7	5	9	0	25	**55**	112
Total	1,677	1,150	571	426	10	1,056	140	**5,030**

The number of annotations that both coders annotated similarly are in bold.

Table 11. Confusion Matrix of the *Secondary* Annotations for *Visual Attention* between the Main Coder and Secondary Coder

	Secondary coder							
Main coder	*Robot*	*Tablet*	*Teaching Mats.*	*Adult*	*Assistant*	*Elsewhere*	*Mixed*	*Total*
Robot	**305**	16	7	13	1	23	16	381
Tablet	19	**122**	1	6	0	7	6	161
Teaching Mats.	10	1	**60**	4	0	5	3	83
Adult	13	7	3	**118**	0	7	9	157
Assistant	0	0	0	0	**7**	0	0	7
Elsewhere	17	8	4	11	0	**119**	10	169
Mixed	6	4	0	6	1	3	**52**	72
Total	370	158	75	158	9	164	96	**1,030**

The number of annotations that both coders annotated similarly are in bold.

ACKNOWLEDGMENTS

The authors and the DE-ENIGMA project are extremely grateful to the school that hosted this research, and the staff members, children, and families who generously gave their time for the study. We would also like to thank Meike Berkhoff, Hannah Viner, Audrey McMillion, Mutluhan Ersoy, Manual Milling, Alice Baird, Mihai Zanfir, Elisabeta Oneata, Vesna Petrović, Bridgette Connell, Lynn Packwood, and Cristina Fernández Álvarez de Eulate for their help with this study.

REFERENCES

[1] Alyssa M. Alcorn, Eloise Ainger, Vicky Charisi, Stefania Mantinioti, Sunčica Petrović, Bob R. Schadenberg, Teresa Tavassoli, and Elizabeth Pellicano. 2019. Educators' views on using humanoid robots with autistic learners in special education settings in england. *Frontiers in Robotics and AI* 6, November (Nov. 2019), 1–15. DOI: https://doi.org/10.3389/frobt.2019.00107

[2] American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). American Psychiatric Association, Arlington, VA. DOI: https://doi.org/10.1176/appi.books.9780890423349

[3] American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Author, Washington, DC. DOI: https://doi.org/10.1176/appi.books.9780890425596

[4] Salvatore M. Anzalone, Sofiane Boucenna, Serena Ivaldi, and Mohamed Chetouani. 2015. Evaluating the engagement with social robots. *International Journal of Social Robotics* 7, 4 (Aug. 2015), 465–478. DOI: https://doi.org/10.1007/s12369-015-0298-7

[5] Roger Azevedo. 2015. Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist* 50, 1 (Jan. 2015), 84–94. DOI: https://doi.org/10.1080/00461520.2015.1004069

[6] Joshua H. Balsters, Matthew A. J. Apps, Dimitris Bolis, Rea Lehner, Louise Gallagher, and Nicole Wenderoth. 2017. Disrupted prediction errors index social deficits in autism spectrum disorder. *Brain* 140, 1 (Jan. 2017), 235–246. DOI: https://doi.org/10.1093/brain/aww287

[7] Moshe Bar. 2007. The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences* 11, 7 (Jul. 2007), 280–289. DOI: https://doi.org/10.1016/j.tics.2007.05.005

[8] Simon Baron-Cohen. 2002. The extreme male brain theory of autism. *Trends in Cognitive Sciences* 6, 6 (Jun. 2002), 248–254. DOI: https://doi.org/10.1016/S1364-6613(02)01904-6

[9] Simon Baron-Cohen. 2009. Autism: The empathizing-systemizing (E-S) theory. *Annals of the New York Academy of Sciences* 1156, 1 (2009), 68–80. DOI: https://doi.org/10.1111/j.1749-6632.2009.04467.x

[10] Jackson Beatty. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin* 91, 2 (1982), 276–292. DOI: https://doi.org/10.1037/0033-2909.91.2.276

[11] Jackson Beatty and Brennis Lucero-Wagoner. 2000. The pupillary system. In *Proceedings of the Handbook of Psychophysiology* (2nd ed.), John T. Cacioppo, Louius G. Tassinary, and Gary G. Berntson (Eds.). Cambridge University Press, New York, NY, 142–162.

[12] Momotaz Begum, Richard W. Serna, and Holly A. Yanco. 2016. Are robots ready to deliver autism interventions? A comprehensive review. *International Journal of Social Robotics* 8, 2 (Mar. 2016), 157–181. DOI: https://doi.org/10.1007/s12369-016-0346-y

[13] Howard Berenbaum, Keith Bredemeier, and Renee J. Thompson. 2008. Intolerance of uncertainty: Exploring its dimensionality and associations with need for cognitive closure, psychopathology, and personality. *Journal of Anxiety Disorders* 22, 1 (Jan. 2008), 117–125. DOI: https://doi.org/10.1016/j.janxdis.2007.01.004

[14] Timothy W. Bickmore, Daniel Schulman, and Langxuan Yin. 2010. Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence* 24, 6 (Jul. 2010), 648–666. DOI: https://doi.org/10.1080/08839514.2010.492259

[15] Christina Boulter, Mark H. Freeston, Mikle South, and Jacqui Rodgers. 2014. Intolerance of uncertainty as a framework for understanding anxiety in children and adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 44, 6 (2014), 1391–1402. DOI: https://doi.org/10.1007/s10803-013-2001-x

[16] Margaret M. Bradley, Laura Miccoli, Miguel A. Escrig, and Peter J. Lang. 2008. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* 45, 4 (Jul. 2008), 602–607. DOI: https://doi.org/10.1111/j.1469-8986.2008.00654.x

[17] Ricarda Braukmann, Emma Ward, Roy S. Hessels, Harold Bekkering, Jan K. Buitelaar, and Sabine Hunnius. 2018. Action prediction in 10-month-old infants at high and low familial risk for Autism Spectrum Disorder. *Research in Autism Spectrum Disorders* 49, May 2017 (May 2018), 34–46. DOI: https://doi.org/10.1016/j.rasd.2018.02.004

[18] Linley C. Bryan and David L. Gast. 2000. Teaching on-task and on-schedule behaviors to high-functioning children with autism via picture activity schedules. *Journal of Autism and Developmental Disorders* 30, 6 (2000), 553–567. DOI: https://doi.org/10.1023/A:1005687310346

[19] John-John Cabibihan, Hifza Javed, Marcelo Ang, and Sharifah Mariam Aljunied. 2013. Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism. *International Journal of Social Robotics* 5, 4 (Nov. 2013), 593–618. DOI: https://doi.org/10.1007/s12369-013-0202-2

[20] R. Nicholas Carleton. 2016. Into the unknown: A review and synthesis of contemporary models involving uncertainty. *Journal of Anxiety Disorders* 39 (Apr. 2016), 30–43. DOI: https://doi.org/10.1016/j.janxdis.2016.02.007

[21] Paul D. Chamberlain, Jacqui Rodgers, Michael J. Crowley, Sarah E. White, Mark H. Freeston, and Mikle South. 2013. A potentiated startle study of uncertainty and contextual anxiety in adolescents diagnosed with autism spectrum disorder. *Molecular Autism* 4, 1 (2013), 31. DOI: https://doi.org/10.1186/2040-2392-4-31

[22] Caitlyn Clabaugh, Kartik Mahajan, Shomik Jain, Roxanna Pakkar, David Becerra, Zhonghao Shi, Eric C. Deng, Rhianna Lee, Gisele Ragusa, and Maja J. Matarić. 2019. Long-term personalization of an in-home socially assistive robot for children with autism spectrum disorders. *Frontiers in Robotics and AI* 6, November (Nov. 2019), 1–18. DOI: https://doi.org/10.3389/frobt.2019.00110

[23] Caitlyn Clabaugh and Maja J. Matarić. 2019. Escaping Oz: Autonomy in socially assistive robotics. *Annual Review of Control, Robotics, and Autonomous Systems* 2, 1 (May 2019), 33–61. DOI: https://doi.org/10.1146/annurev-control-060117-104911

[24] Andy Clark. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36, 3 (Jun. 2013), 181–204. DOI: https://doi.org/10.1017/S0140525X12000477

[25] Olivier Collignon, Geneviève Charbonneau, Frédéric Peters, Marouane Nassim, Maryse Lassonde, Franco Lepore, Laurent Mottron, and Armando Bertone. 2013. Reduced multisensory facilitation in persons with autism. *Cortex* 49, 6 (Jun. 2013), 1704–1710. DOI: https://doi.org/10.1016/j.cortex.2012.06.001

[26] Jonathan S. Comer, Amy K. Roy, Jami M. Furr, Kristin Gotimer, Rinad S. Beidas, Michel J. Dugas, and Philip C. Kendall. 2009. The intolerance of uncertainty scale for children: A psychometric evaluation. *Psychological Assessment* 21, 3 (2009), 402–411. DOI: https://doi.org/10.1037/a0016719

[27] Alexandre Coninx, Paul E. Baxter, Elettra Oleari, Sara Bellini, Bert P.B. Bierman, Olivier A. Blanson Henkemans, Lola Cañamero, Piero Cosi, Valentin Enescu, Raquel Ros Espinoza, Antoine Hiolle, Rémi Humbert, Bernd Kiefer, Ivana Kruijff-Korbayová, Rosemarijn Looije, Marco Mosconi, Mark A. Neerincx, Giulio Paci, Georgios Patsis, Clara Pozzi, Francesca Sacchitelli, Hichem Sahli, Alberto Sanna, Giacomo Sommavilla, Fabio Tesser, Yiannis Demiris, and Tony Belpaeme. 2015. Towards long-term social child-robot interaction: Using multi-activity switching to engage young users. *Journal of Human–Robot Interaction* 5, 1 (Aug. 2015), 32–64. DOI: https://doi.org/10.5898/JHRI.5.1.Coninx

[28] James P. Connell and James G. Wellborn. 1991. Competence, autonomy, and relatedness: A motivational analysis of self-system processes. In *Self processes and Development*. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, 43–77. Retrieved from https://psycnet.apa.org/record/1991-97029-002.

[29] Lee J. Corrigan, Christopher Peters, Dennis Küster, and Ginevra Castellano. 2016. Engagement perception and generation for social robots and virtual agents. In *Toward Robotic Socially Believable Behaving Systems*. Anna Esposito and Lakhmi C. Jain (Eds.), Intelligent Systems Reference Library, Vol. 105, Springer, International Publishing, Cham, 29–51. DOI: https://doi.org/10.1007/978-3-319-31056-5_4

[30] Cristina A. Costescu, Bram Vanderborght, and Daniel O. David. 2015. Reversal learning task in children with autism spectrum disorder: A robot-based approach. *Journal of Autism and Developmental Disorders* 45, 11 (2015), 3715–3725. DOI: https://doi.org/10.1007/s10803-014-2319-z

[31] Sander Van de Cruys, Kris Evers, Ruth Van der Hallen, Lien Van Eylen, Bart Boets, Lee De-Wit, and Johan Wagemans. 2014. Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review* 121, 4 (2014), 649–675. DOI: https://doi.org/10.1037/a0037665

[32] Kerstin Dautenhahn. 1999. Robots as social actors: Aurora and the case of autism. In *Proceedings of the 3rd Cognitive Technology Conference (CT'99)*. Vol. 359. M.I.N.D. Lab, San Fransisco, CA, 359–374. Retrieved from https://books.google.nl/books?id=QM--tgAACAAJ.

[33] Kerstin Dautenhahn. 2007. Socially intelligent robots: Dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362, 1480 (Apr. 2007), 679–704. DOI: https://doi.org/10.1098/rstb.2006.2004

[34] Kerstin Dautenhahn and Iain Werry. 2004. Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition* 12, 1 (2004), 1–35. DOI: https://doi.org/10.1075/pc.12.1.03dau

[35] Daniel O. David, Cristina A. Costescu, Silviu-Andrei Matu, Aurora Szentagotai, and Anca Dobrean. 2020. Effects of a robot-enhanced intervention for children with ASD on teaching turn-taking skills. *Journal of Educational Computing Research* 58, 1 (Mar. 2020), 29–62. DOI: https://doi.org/10.1177/0735633119830344

[36] Lorenzo Desideri, Marco Negrini, Massimiliano Malavasi, Daniela Tanzini, Aziz Rouame, Maria Cristina Cutrone, Paola Bonifacci, and Evert-Jan Hoogerwerf. 2018. Using a humanoid robot as a complement to interventions for

children with autism spectrum disorder: A pilot study. *Advances in Neurodevelopmental Disorders* 2, 3 (2018), 273–285. DOI:https://doi.org/10.1007/s41252-018-0066-4

[37] Joshua J. Diehl, Charles R. Crowell, Michael Villano, Kristin Wier, Karen Tang, and Laurel D. Riek. 2014. Clinical applications of robots in autism spectrum disorder diagnosis and treatment. In *Comprehensive Guide to Autism*. Vinood B. Patel, Victor R. Preedy, and Colin R. Martin (Eds.). Springer, New York, NY, 411–422. DOI:https://doi.org/10.1007/978-1-4614-4788-7_14

[38] Joshua J. Diehl, Lauren M. Schmitt, Michael Villano, and Charles R. Crowell. 2012. The clinical use of robots for individuals with autism spectrum disorders: A critical review. *Research in Autism Spectrum Disorders* 6, 1 (Jan. 2012), 249–262. DOI:https://doi.org/10.1016/j.rasd.2011.05.006

[39] Lucy Diep, John-John Cabibihan, and Gregor Wolbring. 2015. Social Robots: Views of special education teachers. In *Proceedings of the 3rd 2015 Workshop on ICTs for improving Patients Rehabilitation Research Techniques*. ACM, New York, NY, 160–163. DOI:https://doi.org/10.1145/2838944.2838983

[40] Anca Dragan and Siddhartha Srinivasa. 2014. Familiarization to robot motion. In *Proceedings of the 2014 ACM/IEEE International Conference on Human–Robot Interaction*. ACM, New York, NY, 366–373. DOI:https://doi.org/10.1145/2559636.2559674

[41] Katherine Driggs-Campbell and Ruzena Bajcsy. 2016. Communicating intent on the road through human-inspired control schemes. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 3042–3047. DOI:https://doi.org/10.1109/IROS.2016.7759471

[42] Michel J. Dugas, Kristin Buhr, and Robert Ladouceur. 2004. The role of intolerance of uncertainty in etiology and maintenance. In *Generalized Anxiety Disorder: Advances in Research and Practice*. R. G. Heimberg, C. L. Turk, & D. S. Mennin (Eds.), Guilford Press, New York, NY143–163.

[43] Audrey Duquette, François Michaud, and Henri Mercier. 2008. Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. *Autonomous Robots* 24, 2 (Feb. 2008), 147–157. DOI:https://doi.org/10.1007/s10514-007-9056-5

[44] Danielle A. Einstein. 2014. Extension of the transdiagnostic model to focus on intolerance of uncertainty: A review of the literature and implications for treatment. *Clinical Psychology: Science and Practice* 21, 3 (Sep. 2014), 280–300. DOI:https://doi.org/10.1111/cpsp.12077

[45] Louise Ewing, Elizabeth Pellicano, and Gillian Rhodes. 2013. Atypical updating of face representations with experience in children with autism. *Developmental Science* 16, 1 (Jan. 2013), 116–123. DOI:https://doi.org/10.1111/desc.12007

[46] David J. Feil-Seifer and Maja J. Matarić. 2009. Toward socially assistive robotics for augmenting interventions for children with autism spectrum disorders. In *Experimental Robotics*, O. Khatib, V. Kumar, and G. J. Pappas (Eds.), Vol. 54. Springer, Berlin Heidelberg, 201–210. DOI:https://doi.org/10.1007/978-3-642-00196-3_24

[47] Cindy Ferrara and Suzanne D. Hill. 1980. The responsiveness of autistic children to the predictability of social and nonsocial toys. *Journal of Autism and Developmental Disorders* 10, 1 (1980), 51–57. DOI:https://doi.org/10.1007/BF02408432

[48] Jennifer A. Fredricks, Phyllis C. Blumenfeld, and Alison H. Paris. 2004. School engagement: Potential of the concept, state of the evidence. *Review of Educational Research* 74, 1 (Mar. 2004), 59–109. DOI:https://doi.org/10.3102/00346543074001059

[49] Karl J. Friston. 2005. A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, 1456 (Apr. 2005), 815–836. DOI:https://doi.org/10.1098/rstb.2005.1622

[50] Morgan Frost-Karlsson, Martyna Alexandra Galazka, Christopher Gillberg, Carina Gillberg, Carmela Miniscalco, Eva Billstedt, Nouchine Hadjikhani, and Jakob Åsberg Johnels. 2019. Social scene perception in autism spectrum disorder: An eye-tracking and pupillometric study. *Journal of Clinical and Experimental Neuropsychology* 41, 10 (Nov. 2019), 1024–1032. DOI:https://doi.org/10.1080/13803395.2019.1646214

[51] Eleanor J. Gibson and Anne D. Pick. 2000. *Perceptual Learning and Development: An Ecological Approach*. Oxford University Press, New York, NY.

[52] Lakshmi J. Gogate and George Hollich. 2010. Invariance detection within an interactive system: A perceptual gateway to language development. *Psychological Review* 117, 2 (2010), 496–516. DOI:https://doi.org/10.1037/a0019049

[53] Judith Goris, Senne Braem, Annabel D. Nijhof, Davide Rigoni, Eliane Deschrijver, Sander Van de Cruys, Jan R. Wiersema, and Marcel Brass. 2018. Sensory prediction errors are less modulated by global context in autism spectrum disorder. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 3, 8 (Aug. 2018), 667–674. DOI:https://doi.org/10.1016/j.bpsc.2018.02.003

[54] Judith Goris, Marcel Brass, Charlotte Cambier, Jeroen Delplanque, Jan R. Wiersema, and Senne Braem. 2020. The relation between preference for predictability and autistic traits. *Autism Research* 13, 7 (Jul. 2020), 1144–1154. DOI:https://doi.org/10.1002/aur.2244

[55] Charles R. Greenwood. 1991. Longitudinal analysis of time, engagement, and achievement in at-risk versus non-risk students. *Exceptional Children* 57, 6 (May 1991), 521–535. DOI:https://doi.org/10.1177/001440299105700606

[56] Dan W. Grupe and Jack B. Nitschke. 2013. Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience* 14, 7 (Jul. 2013), 488–501. DOI: https://doi.org/10.1038/nrn3524

[57] Rebecca Grzadzinski, Marisela Huerta, and Catherine Lord. 2013. DSM-5 and autism spectrum disorders (ASDs): An opportunity for identifying ASD subtypes. *Molecular Autism* 4, 1 (2013), 12. DOI: https://doi.org/10.1186/2040-2392-4-12

[58] Victoria Hallett, Joanne Mueller, Lauren Breese, Megan Hollett, Bryony Beresford, Annie Irvine, Andrew Pickles, Vicky Slonims, Stephen Scott, Tony Charman, and Emily Simonoff. 2021. Introducing 'predictive parenting': A feasibility study of a new group parenting intervention targeting emotional and behavioral difficulties in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 51, 1 (Jan. 2021), 323–333. DOI: https://doi.org/10.1007/s10803-020-04442-2

[59] Francesca Happé and Uta Frith. 2006. The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders* 36, 1 (Jan 2006), 5–25. DOI: https://doi.org/10.1007/s10803-005-0039-0

[60] Francesca Happé, Angelica Ronald, and Robert Plomin. 2006. Time to give up on a single explanation for autism. *Nature Neuroscience* 9, 10 (Oct. 2006), 1218–1220. DOI: https://doi.org/10.1038/nn1770

[61] Annika Hellendoorn, Lex Wijnroks, and Paul P. M. Leseman. 2015. Unraveling the nature of autism: Finding order amid change. *Frontiers in Psychology* 6 (2015), 359. DOI: https://doi.org/10.3389/fpsyg.2015.00359

[62] E. H. Hess and J. M. Polt. 1964. Pupil size in relation to mental activity during simple problem-solving. *Science* 143, 3611 (Mar. 1964), 1190–1192. DOI: https://doi.org/10.1126/science.143.3611.1190

[63] Jakob Hohwy. 2013. *The Predictive Mind*. Oxford University Press, Oxford. DOI: https://doi.org/10.1093/acprof:oso/9780199682737.001.0001

[64] Ryan Y. Hong and Mike W.L. Cheung. 2015. The Structure of cognitive vulnerabilities to depression and anxiety. *Clinical Psychological Science* 3, 6 (Nov. 2015), 892–912. DOI: https://doi.org/10.1177/2167702614553789

[65] Claire A. G. J. Huijnen, Monique A. S. Lexis, Rianne Jansens, and Luc P. de Witte. 2017. How to implement robots in interventions for children with autism? A co-creation study involving people with autism, parents and professionals. *Journal of Autism and Developmental Disorders* 47, 10 (Oct. 2017), 3079–3096. DOI: https://doi.org/10.1007/s10803-017-3235-9

[66] Claire A. G. J. Huijnen, Monique A. S. Lexis, Rianne Jansens, and Luc P. de Witte. 2019. Roles, strengths and challenges of using robots in interventions for children with autism spectrum disorder (ASD). *Journal of Autism and Developmental Disorders* 49, 1 (Jan. 2019), 11–21. DOI: https://doi.org/10.1007/s10803-018-3683-x

[67] Bibi E. B. M. Huskens, Rianne Verschuur, Jan C. C. Gillesen, Robert Didden, and Emilia I. Barakova. 2013. Promoting question-asking in school-aged children with autism spectrum disorders: Effectiveness of a robot intervention compared to a human-trainer intervention. *Developmental Neurorehabilitation* 16, 5 (2013), 345–356. DOI: https://doi.org/10.3109/17518423.2012.739212

[68] Shomik Jain, Balasubramanian Thiagarajan, Zhonghao Shi, Caitlyn Clabaugh, and Maja J. Matarić. 2020. Modeling engagement in long-term, in-home socially assistive robot interventions for children with autism spectrum disorders. *Science Robotics* 5, 39 (Feb. 2020), eaaz3791. DOI: https://doi.org/10.1126/scirobotics.aaz3791

[69] Hifza Javed, Rachael Burns, Myounghoon Jeon, Ayanna M. Howard, and Chung Hyuk Park. 2019. A robotic framework to facilitate sensory experiences for children with autism spectrum disorder. *ACM Transactions on Human–Robot Interaction* 9, 1 (Dec. 2019), 1–26. DOI: https://doi.org/10.1145/3359613

[70] Hifza Javed, WonHyong Lee, and Chung Hyuk Park. 2020. Toward an automated measure of social engagement for children with autism spectrum disorder—a personalized computational modeling approach. *Frontiers in Robotics and AI* 7, April (Apr. 2020), 14. DOI: https://doi.org/10.3389/frobt.2020.00043

[71] D. Kahneman and J. Beatty. 1966. Pupil diameter and load on memory. *Science* 154, 3756 (Dec. 1966), 1583–1585. DOI: https://doi.org/10.1126/science.154.3756.1583

[72] Deb Keen. 2009. Engagement of children with autism in learning. *Australasian Journal of Special Education* 33, 2 (2009), 130–140. DOI: https://doi.org/10.1375/ajse.33.2.130

[73] C.D. Kidd and Cynthia Breazeal. 2008. Robots at home: Understanding long-term human–robot interaction. In *Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 3230–3235. DOI: https://doi.org/10.1109/IROS.2008.4651113

[74] Elizabeth Kim, Rhea Paul, Frederick Shic, and Brian Scassellati. 2012. Bridging the research Gap: Making HRI useful to individuals with autism. *Journal of Human–Robot Interaction* 1, 1 (Aug. 2012), 26–54. DOI: https://doi.org/10.5898/JHRI.1.1.Kim

[75] Elizabeth S. Kim, Lauren D. Berkovits, Emily P. Bernier, Dan Leyzberg, Frederick Shic, Rhea Paul, and Brian Scassellati. 2013. Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism and Developmental Disorders* 43, 5 (May 2013), 1038–1049. DOI: https://doi.org/10.1007/s10803-012-1645-2

[76] Elizabeth S. Kim, Christopher M. Daniell, Corinne Makar, Julia Elia, Brian Scassellati, and Frederick Shic. 2015. Potential clinical impact of positive affect in robot interactions for autism intervention. In *Proceedings of the 2015 International Conference on Affective Computing and Intelligent Interaction*. IEEE, 8–13. DOI: https://doi.org/10.1109/ACII.2015.7344544

[77] Jacqueline Kory and Cynthia Breazeal. 2014. Storytelling with robots: Learning companions for preschool children's language development. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 643–648. DOI: https://doi.org/10.1109/ROMAN.2014.6926325

[78] Viviane Kostrubiec and Jeanne Kruck. 2020. Collaborative research project: Developing and testing a robot-assisted intervention for children with autism. *Frontiers in Robotics and AI* 7, March (Mar. 2020), 1–16. DOI: https://doi.org/10.3389/frobt.2020.00037

[79] Hideki Kozima, Marek P. Michalowski, and Cocoro Nakagawa. 2009. Keepon. *International Journal of Social Robotics* 1, 1 (Jan. 2009), 3–18. DOI: https://doi.org/10.1007/s12369-008-0009-8

[80] Rebecca P. Lawson, Christoph Mathys, and Geraint Rees. 2017. Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience* 20, 9 (Jul. 2017), 1293–1299. DOI: https://doi.org/10.1038/nn.4615

[81] Rebecca P. Lawson, Geraint Rees, and Karl J. Friston. 2014. An aberrant precision account of autism. *Frontiers in Human Neuroscience* 8, May (May 2014), 302. DOI: https://doi.org/10.3389/fnhum.2014.00302

[82] Jamy J. Li, Daniel Davison, Alyssa M. Alcorn, Alria Williams, Snezana Babovic Dimitrijevic, Sunčica Petrović, Pauline Chevalier, Bob R. Schadenberg, Eloise Ainger, Liz Pellicano, and Vanessa Evers. 2020. Non-participatory user-centered design of accessible teacher-teleoperated robot and tablets for minimally verbal autistic children. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '20)*. 51–59. DOI: https://doi.org/10.1145/3389189.3393738

[83] Itay Lieder, Vincent Adam, Or Frenkel, Sagi Jaffe-Dax, Maneesh Sahani, and Merav Ahissar. 2019. Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia. *Nature Neuroscience* 22, 2 (Feb. 2019), 256–264. DOI: https://doi.org/10.1038/s41593-018-0308-9

[84] Catherine Lord, Michael Rutter, Pamela C. DiLavore, Susan Risi, Katherine Gotham, and Somer L. Bishop. 2012. *Autism Diagnostic Observation Schedule: ADOS-2.* Western Psychological Services, Los Angeles, CA. Retrieved from https://www.wpspublish.com/store/p/2648/ados-2-autism-diagnostic-observation-schedule-second-edition

[85] Nichola Lubold, Erin Walker, and Heather Pon-Barry. 2016. Effects of voice-adaptation and social dialogue on perceptions of a robotic learning companion. In *Proceedings of the 2016 11th ACM/IEEE International Conference on Human–Robot Interaction*. IEEE, 255–262. DOI: https://doi.org/10.1109/HRI.2016.7451760

[86] Hope Macdonald, Michael Rutter, Patricia Howlin, Patricia Rios, Ann Le Conteur, Christopher Evered, and Susan Folstein. 1989. Recognition and expression of emotional cues by autistic and normal adults. *Journal of Child Psychology and Psychiatry* 30, 6 (Nov. 1989), 865–877. DOI: https://doi.org/10.1111/j.1469-7610.1989.tb00288.x

[87] Gregory S. MacDuff, Patricia J. Krantz, and Lynn E. McClannahan. 1993. Teaching children with autism to use photographic activity schedules: Maintenance and generalization of complex response chains. *Journal of Applied Behavior Analysis* 26, 1 (Mar. 1993), 89–97. DOI: https://doi.org/10.1901/jaba.1993.26-89

[88] Catherine Manning, James Kilner, Louise Neil, Themelis Karaminis, and Elizabeth Pellicano. 2017. Children on the autism spectrum update their behaviour in response to a volatile environment. *Developmental Science* 20, 5 (Sep. 2017), e12435. DOI: https://doi.org/10.1111/desc.12435

[89] Lizzie Maughan, Sergei Gutnikov, and Rob Stevens. 2007. Like more, look more. Look more, like more: The evidence from eye-tracking. *Journal of Brand Management* 14, 4 (Apr. 2007), 335–342. DOI: https://doi.org/10.1057/palgrave.bm.2550074

[90] Joanne McCann and Sue Peppé. 2003. Prosody in autism spectrum disorders: a critical review. *International Journal of Language & Communication Disorders* 38, 4 (Jan. 2003), 325–350. DOI: https://doi.org/10.1080/1368282031000154204

[91] Linda McCormick, Mary J. O. Noonan, and Ronald Heck. 1998. Variables affecting engagement in inclusive preschool classrooms. *Journal of Early Intervention* 21, 2 (Jan. 1998), 160–176. DOI: https://doi.org/10.1177/105381519802100208

[92] Gary B. Mesibov and Victoria Shea. 2010. The TEACCH program in the era of evidence-based practice. *Journal of Autism and Developmental Disorders* 40, 5 (May 2010), 570–579. DOI: https://doi.org/10.1007/s10803-009-0901-6

[93] Brian W. Miller. 2015. Using reading times and eye-movements to measure cognitive engagement. *Educational Psychologist* 50, 1 (Jan. 2015), 31–42. DOI: https://doi.org/10.1080/00461520.2015.1004068

[94] D. Mumford. 1992. On the computational architecture of the neocortex. *Biological Cybernetics* 66, 3 (Jan. 1992), 241–251. DOI: https://doi.org/10.1007/BF00198477

[95] Louise Neil, Nora Choque Olsson, and Elizabeth Pellicano. 2016. The relationship between intolerance of uncertainty, sensory sensitivities, and anxiety in autistic and typically developing children. *Journal of Autism and Developmental Disorders* 46, 6 (Jun. 2016), 1962–1973. DOI: https://doi.org/10.1007/s10803-016-2721-9

[96] Heather L. O'Brien and Elaine G. Toms. 2008. What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology* 59, 6 (Apr. 2008), 938–955. DOI: https://doi.org/10.1002/asi.20801

[97] Mike Oliver. 2013. The social model of disability: Thirty years on. *Disability and Society* 28, 7 (Oct. 2013), 1024–1026. DOI: https://doi.org/10.1080/09687599.2013.818773

[98] Alexia Ostrolenk, Vanessa A. Bao, Laurent Mottron, Olivier Collignon, and Armando Bertone. 2019. Reduced multisensory facilitation in adolescents and adults on the autism spectrum. *Scientific Reports* 9, 1 (Dec. 2019), 11965. DOI: https://doi.org/10.1038/s41598-019-48413-9

[99] Mark O'Reilly, Jeff Sigafoos, Giulio Lancioni, Chaturi Edrisinha, and Alonzo Andrews. 2005. An examination of the effects of a classroom activity schedule on levels of self-injury and engagement for a child with severe autism. *Journal of Autism and Developmental Disorders* 35, 3 (Jun. 2005), 305–311. DOI: https://doi.org/10.1007/s10803-005-3294-1

[100] Colin J. Palmer, Rebecca P. Lawson, and Jakob Hohwy. 2017. Bayesian approaches to autism: Towards volatility, action, and behavior. *Psychological Bulletin* 143, 5 (2017), 521–542. DOI: https://doi.org/10.1037/bul0000097

[101] Rhea Paul, Amy Augustyn, Ami Klin, and Fred R. Volkmar. 2005. Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 35, 2 (Apr. 2005), 205–220. DOI: https://doi.org/10.1007/s10803-004-1999-1

[102] Martin P. Paulus and Murray B. Stein. 2006. An insular view of anxiety. *Biological Psychiatry* 60, 4 (Aug. 2006), 383–387. DOI: https://doi.org/10.1016/j.biopsych.2006.03.042

[103] Philip J. Pell, Isabelle Mareschal, Andrew J. Calder, Elisabeth A. H. von dem Hagen, Colin W.G. Clifford, Simon Baron-Cohen, and Michael P. Ewbank. 2016. Intact priors for gaze direction in adults with high-functioning autism spectrum conditions. *Molecular Autism* 7, 1 (Dec. 2016), 25. DOI: https://doi.org/10.1186/s13229-016-0085-9

[104] Elizabeth Pellicano and David Burr. 2012. When the world becomes 'too real': A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences* 16, 10 (2012), 504–510. DOI: https://doi.org/10.1016/j.tics.2012.08.009

[105] Elizabeth Pellicano, Linda Jeffery, David Burr, and Gillian Rhodes. 2007. Abnormal adaptive face-coding mechanisms in children with autism spectrum disorder. *Current Biology* 17, 17 (Sep. 2007), 1508–1512. DOI: https://doi.org/10.1016/j.cub.2007.07.065

[106] Jose Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team. 2020. *Nlme: Linear and Nonlinear Mixed Effects Models*. Retrieved from https://CRAN.R-project.org/package=nlme

[107] Michael I. Posner. 1980. Orienting of attention. *Quarterly Journal of Experimental Psychology* 32, 1 (Feb 1980), 3–25. DOI: https://doi.org/10.1080/00335558008248231

[108] W. James Potter and Deborah Levine-Donnerstein. 1999. Rethinking validity and reliability in content analysis. *Journal of Applied Communication Research* 27, 3 (Aug. 1999), 258–284. DOI: https://doi.org/10.1080/00909889909365539

[109] R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Technical Report. R Foundation for Statistical Computing, Vienna. Retrieved from http://www.r-project.org/

[110] Aditi Ramachandran, Chien-Ming Huang, and Brian Scassellati. 2017. Give me a break!. In *Proceedings of the 2017 ACM/IEEE International Conference on Human–Robot Interaction*. ACM, New York, NY, 146–155. DOI: https://doi.org/10.1145/2909824.3020209

[111] Stephen W. Raudenbush and Anthony S. Bryk. 2002. *Hierarchical Linear Models. Applications and Data Analysis Methods* (2nd ed.). Sage Publications, Thousand Oaks, CA.

[112] Charles Rich, Brett Ponsleur, Aaron Holroyd, and Candace L. Sidner. 2010. Recognizing engagement in human–robot interaction. In *Proceeding of the 5th ACM/IEEE International Conference on Human–Robot Interaction (HRI '10)*. ACM Press, New York, NY, 375. DOI: https://doi.org/10.1145/1734454.1734580

[113] Daniel J. Ricks and Mark B. Colton. 2010. Trends and considerations in robot-assisted autism therapy. In *Proceedings of the 2010 IEEE International Conference on Robotics and Automation*. IEEE, 4354–4359. DOI: https://doi.org/10.1109/ROBOT.2010.5509327

[114] Ben Robins, Kerstin Dautenhahn, René te Boekhorst, and Aude Billard. 2005. Robotic assistants in therapy and education of children with autism: Can a small humanoid robot help encourage social interaction skills? *Universal Access in the Information Society* 4, 2 (Dec. 2005), 105–120. DOI: https://doi.org/10.1007/s10209-005-0116-3

[115] Jacqui Rodgers, Anna Hodgson, Kerry Shields, Catharine Wright, Emma Honey, and Mark Freeston. 2017. Towards a treatment for intolerance of uncertainty in young people with autism spectrum disorder: Development of the coping with uncertainty in everyday situations (CUES©) programme. *Journal of Autism and Developmental Disorders* 47, 12 (Dec. 2017), 3959–3966. DOI: https://doi.org/10.1007/s10803-016-2924-0

[116] Ognjen Rudovic, Jaeryoung Lee, Miles Dai, Björn Schuller, and Rosalind W. Picard. 2018. Personalized machine learning for robot perception of affect and engagement in autism therapy. *Science Robotics* 3, 19 (Jun. 2018), eaao6760. DOI: https://doi.org/10.1126/scirobotics.aao6760

[117] Ognjen Rudovic, Jaeryoung Lee, Lea Mascarell-Maricic, Björn W. Schuller, and Rosalind W. Picard. 2017. Measuring engagement in robot-assisted autism therapy: A cross-cultural study. *Frontiers in Robotics and AI* 4, July (Jul. 2017), 36. DOI: https://doi.org/10.3389/frobt.2017.00036

[118] Michael Rutter, Anthony Bailey, and Cathrine Lord. 2003. *The Social Communication Questionnaire: Manual*. Western Psychological Services, Los Angeles, CA. Retrieved from https://www.wpspublish.com/scq-social-communication-questionnaire

[119] Michelle J. Salvador, Sophia Silver, and Mohammad H. Mahoor. 2015. An emotion recognition comparative study of autistic and typically-developing children using the zeno robot. In *Procedings of the 2015 IEEE International Conference on Robotics and Automation*. IEEE, 6128–6133. DOI: https://doi.org/10.1109/ICRA.2015.7140059

[120] Felippe Sartorato, Leon Przybylowski, and Diana K Sarko. 2017. Improving therapeutic outcomes in autism spectrum disorders: Enhancing social communication and sensory processing through the use of interactive robots. *Journal of Psychiatric Research* 90 (Jul. 2017), 1–11. DOI: https://doi.org/10.1016/j.jpsychires.2017.02.004

[121] Brian Scassellati. 2007. How social robots will help us to diagnose, treat, and understand autism. In *Robotics Research*. Sebastian Thrun, Rodney Brooks, and Hugh Durrant-Whyte (Eds.). Springer, Berlin Heidelberg, 552–563. DOI: https://doi.org/10.1007/978-3-540-48113-3_47

[122] Brian Scassellati, Henny Admoni, and Maja J. Matarić. 2012. Robots for use in autism research. *Annual Review of Biomedical Engineering* 14, 1 (2012), 275–294. DOI: https://doi.org/10.1146/annurev-bioeng-071811-150036

[123] Brian Scassellati, Laura Boccanfuso, Chien-Ming Huang, Marilena Mademtzi, Meiying Qin, Nicole Salomons, Pamela Ventola, and Frederick Shic. 2018. Improving social skills in children with ASD using a long-term, in-home social robot. *Science Robotics* 3, 21 (8 2018), eaat7544. https://doi.org/10.1126/scirobotics.aat7544

[124] Bob R. Schadenberg. 2021. *Robots for Autistic Children: Understanding and Facilitating Predictability for Engagement in Learning*. Ph.D. Dissertation. University of Twente, Enschede. DOI: https://doi.org/10.3990/1.9789036551649

[125] Bob R. Schadenberg, Dirk K. J. Heylen, and Vanessa Evers. 2018. Affect bursts to constrain the meaning of the facial expressions of the humanoid robot Zeno. In *Proceedings of the 1st Workshop on Social Interaction and Multimodal Expression for Socially Intelligent Robots* 30–39. Retrieved from http://ceur-ws.org/Vol-2059/paper4.pdf

[126] Bob R. Schadenberg, Mark A. Neerincx, Fokie Cnossen, and Rosemarijn Looije. 2017. Personalising game difficulty to keep children motivated to play with a social robot: A Bayesian approach. *Cognitive Systems Research* 43 (2017), 222–231. DOI: https://doi.org/10.1016/j.cogsys.2016.08.003

[127] Bob R. Schadenberg, Dennis Reidsma, Dirk K. J. Heylen, and Vanessa Evers. 2020. Differences in spontaneous interactions of autistic children in an interaction with an adult and humanoid robot. *Frontiers in Robotics and AI* 7 (Mar. 2020), 19. DOI: https://doi.org/10.3389/frobt.2020.00028

[128] Bob R. Schadenberg, Dennis Reidsma, Dirk K. J. Heylen, and Vanessa Evers. in press. "I see what you did there": Understanding people's social perception of a robot and its predictability. *ACM Transactions on Human–Robot Interaction* 10, 3 (in press), 1–27. DOI: https://doi.org/10.1145/3461534

[129] Eric Schopler, Mary E. Van Bourgondien, Glenna J. Wellman, and Steven R. Love. 2010. *The Childhood Autism Rating Scale (CARS-2)*. Western Psychological Services, Los Angeles, CA. Retrieved from https://www.wpspublish.com/store/p/2696/cars-2-childhood-autism-rating-scale-second-edition

[130] Candace L. Sidner, Christopher Lee, Cory D. Kidd, Neal Lesh, and Charles Rich. 2005. Explorations in engagement for humans and robots. *Artificial Intelligence* 166, 1-2 (Aug. 2005), 140–164. DOI: https://doi.org/10.1016/j.artint.2005.03.005

[131] Emily Simonoff, Andrew Pickles, Tony Charman, Susie Chandler, Tom Loucas, and Gillian Baird. 2008. Psychiatric disorders in children with autism spectrum disorders: Prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of the American Academy of Child & Adolescent Psychiatry* 47, 8 (Aug. 2008), 921–929. DOI: https://doi.org/10.1097/CHI.0b013e318179964f

[132] Kate Simpson, Deb Keen, and Janeen Lamb. 2013. The use of music to engage children with autism in a receptive labelling task. *Research in Autism Spectrum Disorders* 7, 12 (Dec. 2013), 1489–1496. DOI: https://doi.org/10.1016/j.rasd.2013.08.013

[133] Ramona E. Simut, Johan Vanderfaeillie, Andreea Peca, Greet Van de Perre, and Bram Vanderborght. 2016. Children with autism spectrum disorders make a fruit salad with probo, the social robot: An interaction study. *Journal of Autism and Developmental Disorders* 46, 1 (Jan. 2016), 113–126. DOI: https://doi.org/10.1007/s10803-015-2556-9

[134] Gale M. Sinatra, Benjamin C. Heddy, and Doug Lombardi. 2015. The challenges of defining and measuring student engagement in science. *Educational Psychologist* 50, 1 (1 2015), 1–13. https://doi.org/10.1080/00461520.2014.1002924

[135] Pawan Sinha, Margaret M. Kjelgaard, Tapan K. Gandhi, Kleovoulos Tsourides, Annie L. Cardinaux, Dimitrios Pantazis, Sidney P. Diamond, and Richard M. Held. 2014. Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences* 111, 42 (2014), 15220–15225. DOI: https://doi.org/10.1073/pnas.1416797111

[136] James P. Stevens. 2002. *Applied Multivariate Statistics for the Social Sciences* (5th ed.). Taylor & Francis, New York, NY.

[137] Caroline L. van Straten, Iris Smeekens, Emilia I. Barakova, Jeffrey C. Glennon, Jan K. Buitelaar, and Aoju Chen. 2018. Effects of robots' intonation and bodily appearance on robot-mediated communicative treatment outcomes for children with autism spectrum disorder. *Personal and Ubiquitous Computing* 22, 2 (Apr. 2018), 379–390. DOI: https://doi.org/10.1007/s00779-017-1060-y

[138] Dag Sverre Syrdal, Kerstin Dautenhahn, Ben Robins, Efstathia Karakosta, and Nan Cannon Jones. 2020. Kaspar in the wild: Experiences from deploying a small humanoid robot in a nursery school for children with autism. *Paladyn, Journal of Behavioral Robotics* 11, 1 (Jul. 2020), 301–326. DOI: https://doi.org/10.1515/pjbr-2020-0019

[139] Adriana Tapus, Andreea Peca, Amir Aly, Cristina A. Pop, Lavinia Jisa, Sebastian Pintea, Alina S. Rusu, and Daniel O. David. 2012. Children with autism social engagement in interaction with Nao, an imitative robot: A series of single case experiments. *Interaction Studies* 13, 3 (2012), 315–347. DOI: https://doi.org/10.1075/is.13.3.01tap

[140] Furtuna G. Tewolde, Dorothy V. M. Bishop, and Catherine Manning. 2018. Visual motion prediction and verbal false memory performance in autistic children. *Autism Research* 11, 3 (Mar. 2018), 509–518. DOI: https://doi.org/10.1002/aur.1915

[141] Serge Thill, Cristina A. Pop, Tony Belpaeme, Tom Ziemke, and Bram Vanderborght. 2012. Robot-assisted therapy for autism spectrum disorders with (Partially) autonomous control: Challenges and outlook. *Paladyn, Journal of Behavioral Robotics* 3, 4 (Jan. 2012), 209–217. DOI: https://doi.org/10.2478/s13230-013-0107-7

[142] Marco Turi, David C. Burr, Roberta Igliozzi, David Aagten-Murphy, Filippo Muratori, and Elizabeth Pellicano. 2015. Children with autism spectrum disorder show reduced adaptation to number. *Proceedings of the National Academy of Sciences* 112, 25 (Jun. 2015), 7868–7872. DOI: https://doi.org/10.1073/pnas.1504099112

[143] Joshua Wainer, Ben Robins, Farshid Amirabdollahian, and Kerstin Dautenhahn. 2014. Using the humanoid robot KASPAR to autonomously play triadic games and facilitate collaborative play among children with autism. *IEEE Transactions on Autonomous Mental Development* 6, 3 (Sep. 2014), 183–199. DOI: https://doi.org/10.1109/TAMD.2014.2303116

[144] Susan W. White, Donald Oswald, Thomas Ollendick, and Lawrence Scahill. 2009. Anxiety in children and adolescents with autism spectrum disorders. *Clinical Psychology Review* 29, 3 (Apr. 2009), 216–229. DOI: https://doi.org/10.1016/j.cpr.2009.01.003

[145] Sarah Wigham, Jacqui Rodgers, Mikle South, Helen McConachie, and Mark H. Freeston. 2015. The interplay between sensory processing abnormalities, intolerance of uncertainty, anxiety and restricted and repetitive behaviours in autism spectrum disorder. *Journal of Autism and Developmental Disorders* 45, 4 (2015), 943–952. DOI: https://doi.org/10.1007/s10803-014-2248-x

[146] World Health Organization. 1992. *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. World Health Organization, Geneva. Retrieved from https://apps.who.int/iris/handle/10665/37958

[147] Shu-Chieh Wu and Roger W. Remington. 2003. Characteristics of covert and overt visual orienting: Evidence from attentional and oculomotor capture. *Journal of Experimental Psychology: Human Perception and Performance* 29, 5 (2003), 1050–1067. DOI: https://doi.org/10.1037/0096-1523.29.5.1050

[148] Nurit Yirmiya, Connie Kasari, Marian Sigman, and Peter Mundy. 1989. Facial expressions of affect in autistic, mentally retarded and normal children. *Journal of Child Psychology and Psychiatry* 30, 5 (Sep. 1989), 725–735. DOI: https://doi.org/10.1111/j.1469-7610.1989.tb00785.x

Received November 2020; revised March 2021; accepted May 2021