Kaspar Explains: An Educational Platform Using Causal Explanations to Support Children with Autism with Visual Perspective Taking

Patrick Holthaus[1*], Gabriella Lakatos[1], Marina Sarda Gou[2], Sílvia Moros[1], Luke Jai Wood[1], Ben Robins[1], Hugo Araujo[3], Christine Augusta Ekua deGraft-Hanson[4], Mohammad Reza Mousavi[3], Farshid Amirabdollahian[1]

[1*]Robotics Research Group, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, United Kingdom.
[2]Department of Computer Science, University of Sheffield, Western Bank, Sheffield, S10 2TN, United Kingdom.
[3]Department of Informatics, King's College London, Strand, London, WC2R 2LS, United Kingdom.
[4]Garston Manor School: Specialist School for Autism, Learning Difficulties and Speech and Language, Horseshoe Ln, Watford, WD25 7HR, United Kingdom.

*Corresponding author(s). E-mail(s): p.holthaus@herts.ac.uk;
Contributing authors: g.lakatos@herts.ac.uk; m.sardagou@sheffield.ac.uk; s.moros@herts.ac.uk; l.wood@herts.ac.uk; b.robins@herts.ac.uk; hugo.araujo@kcl.ac.uk; head@garstonmanor.herts.sch.uk; mohammad.mousavi@kcl.ac.uk; f.amirabdollahian2@herts.ac.uk;

Abstract

The *Kaspar Explains* project aimed to assess the efficacy of utilising *causal explanations* on a social humanoid robot to address challenges in instructing autistic children. In a scoping retrospective study which looked back at nearly 18 years of research with the Kaspar robot in children's education, we first identified Visual Perspective Taking (VPT) as a challenge that children with autism often encounter and where causal explanations proved to have a strong potential to support the children. In the context of a local special needs school, we then created appropriate scenarios within the VPT domain and developed a formal causal model to support the children with common misconceptions about visual perspectives. This model was implemented in an interactive scenario for the Kaspar robot. Our approach featured formative and summative evaluation cycles. Together with teachers, parents, and children, we initially assessed the interactive games regarding their suitability for children with autism. We then evaluated the general feasibility of our explanation engine with healthy adults, **($n = 20$)**, before a summative evaluation study with children with autism **($n = 10$)** at our partner school. The study design allowed all the children access to both the intervention and control phases of the study by deploying an ECE-CEC design where E represents causal explanation phases and C represents control phases without explanations. Comparing the number of correct actions in both groups, there were statistically significant differences in favour of the intervention group **($p = 0.04$)**. The study shows that causal explanations can help children better understand and retain different aspects of VPT. This study contributes to advancing innovative and productive educational tools tailored for autistic children. In our future studies, we aim to apply causal explanations to other interactive educational scenarios and areas of difficulty within the curriculum to further assist pupils in reaching their desired educational outcomes.

1 Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that mainly affects communication and social interaction skills. It is often characterised by the difficulty in establishing and maintaining relationships with peers, family members, and other individuals [1]. Recent advancements in technology have opened up new opportunities for individuals with ASD to improve their social skills. In particular, humanoid social robots have shown promise as tools that can provide a controlled, safe and non-threatening environment where children with ASD can practice and enhance their social interaction and communication skills [2–4]. In this paper, we aim to address some of these challenges by expanding the communicative capabilities of the humanoid social robot Kaspar, offering causal explanations, which aim to support the specific needs and abilities of children with autism [5].

Causality and causal explanations are believed to support children in consistently predicting event outcomes [6]. Moreover, children with autism particularly benefit from causality to understand the world around them and the context of social interactions with others [5]. This article hence explores the usage of causal models in human-robot interaction to support children with autism in trusting a humanoid robot and understanding its explanations.

Since 2005, the Kaspar robot[1] has been used to work with children with autism to help break their social isolation by acting as a social mediator with great success [2, 7]. This paper looks into potential application domains of cause-and-effect training for children with autism, specifically looking for children to engage in multi-modal interaction with the robot, taking the initiative to explore the robot's autonomous reactions

[1] https://www.herts.ac.uk/kaspar

Fig. 1: Kaspar discourages certain tactile behaviours.

and its impact. Literature suggests that problems with verbal skills and eye gaze in children with autism create the need for the sense of touch to replace these important and most effective ways of communication [8]. Being a *tactile* robot, Kaspar can provide feedback that encourages appropriate tactile interaction and discourages inappropriate interaction [9]. With the help of a caregiver (parent, teacher, or therapist), the robot provides a platform for children to understand the causes and effects of their actions. The caregiver's role is to add and/or reinforce the appropriate feedback when needed, and to present further cognitive learning opportunities for the child when possible and according to the child's abilities.

The interaction often starts with simple tactile exploration where touching different parts of the robot will cause different reactions and movements, e.g. touching a hand or an arm causes the robot to raise the hand and say "This is my hand" or turn its head to the direction of the touched body part and say "This is my arm"; touching the side of the head will cause the robot to play some sounds, touching the upper torso causes the robot to laugh out loud (saying "Ha, ha, ha"), and touching the sole of the feet causing the robot to smile and say "This is nice, you are tickling me". Suppose the child interacts in any inappropriate tactile behaviour, e.g. hitting the robot or taking other forceful actions. In that case, the robot provides feedback by turning its head and torso away, having a *sad* expression (Figure 1) whilst covering its face with its hand and saying, "Ouch, you are hurting me".

Normally, within the above interaction, the caregiver is in the loop in defining and deciding the interactions based on observations, personal experiences, and protocols for behavioural modification [9]. Our research question is *whether these open-ended parts of the interaction scenario can be complemented with additional causal explanations to contribute to the scenario diversity and result in a better learning outcome.*

In order to arrive at the experiment setup, we therefore begin with a retrospective study (Sect. 2) where we identify situations in which Kaspar would frequently use causal explanations. This is then used to develop educational games and derive a formal causal model. In Section 3, we present related work relevant to the chosen scenario of visual perspective-taking and causal explanations, which is required to explain how we model causality within that scenario. Our approach then consists of a formative-summative evaluation method, described in Section 4, where we introduce a formal causal model for such situations and describe a pilot study, with the chance to gather formative feedback regarding the overall suitability with a healthy adults population

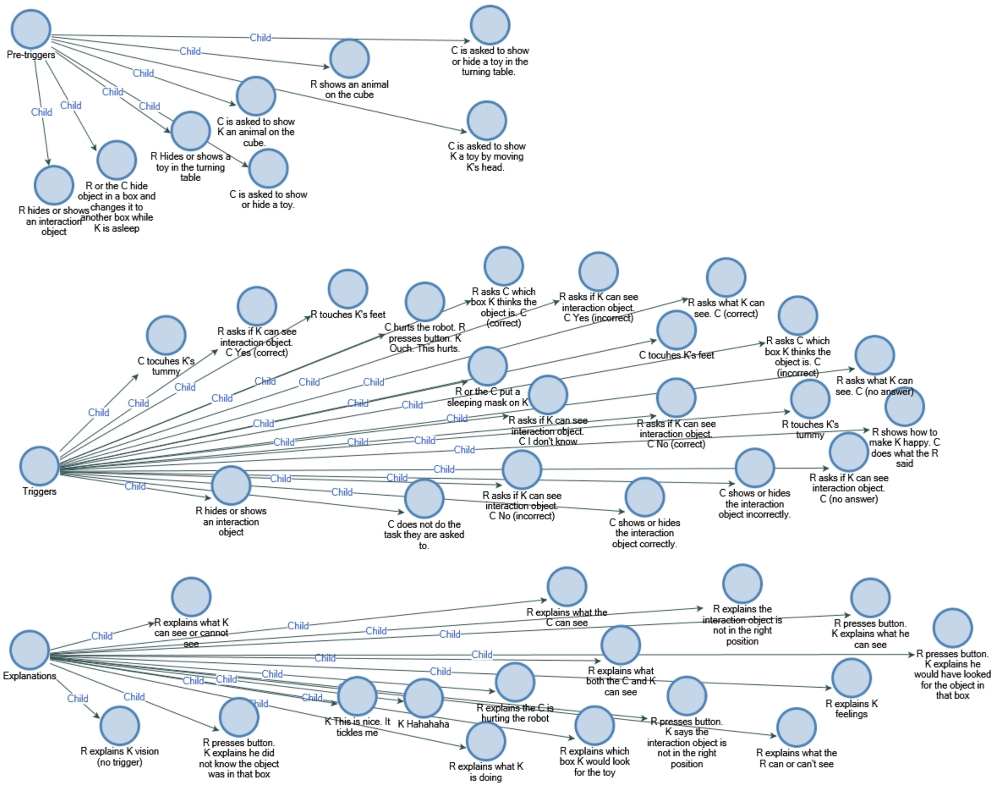


Fig. 2: Graph of relations extracted from retrospective video studies.

($n = 20$). A proof-of-concept experiment involving children ($n = 10$) and teachers at a special needs school then evaluates our developed interaction in a summative way.

Building on previously explored aspects, which individually discuss the pre-study [10] to identify situations where causal explanations play a prominent role in child-robot interactions, the development and pre-testing of the causal model [11] and initial evaluation [12], this article offers a comprehensive perspective on the 'Kaspar Explains' project. It not only synthesises insights from these prior works but also significantly advances them by presenting a holistic view, expanded analysis including further details, and a focused discussion of the project's impact.

2 Explanation extraction from existing studies

As an initial step, we determined a domain in which causal explanations can optimally support children with autism. For that, we observed existing video studies from interaction experiments with Kaspar that had ethical permission for re-analysis under protocol numbers COM/SF/UH/02069 and SPECS/SF/UH/4654(1). Insights obtained from this retrospective study rely on studies conducted during the Horizon 2020 project BabyRobot (Grant Agreement Number 687831) and videos that have been acquired in earlier studies with the Kaspar robot [13]. From these videos, we extracted interactive scenarios featuring explanations. The goal of the retrospective study was to investigate cases of causal explanation that have been used during robot-researcher-child triadic interactions to identify various types of interactions with the potential to utilise causal explanations as well as to provide a data basis for a formal causal explanation engine. We identified and assessed 20 such videos from varying interaction topics and formalised a set of cause-effect-based interaction scenarios. Our study protocol, further detailed in [10], allowed for independent verification of the episode identification with confirmed inter-rater reliability. These are presented by the graph of relationships identifying pre-trigger, trigger and explanations in the child-robot-researcher interactions analysed (Fig. 2).

From a matrix coding query, we found that the most common causal relationship is that a *child shows or hides an object* (trigger) *followed by the researcher pressing a button for Kaspar to explain what he can see* (explanation), with 165 instances. Using this graph of relations, combined with the frequency of triggers and explanations, we identified Visual Perspective Taking (VPT) (see Sect. 3) as an appropriate domain for exploring causal explanations in the interaction between children with autism and Kaspar. Hence, we developed four different games that could benefit from the richness of explanations in this domain (see Sect. 4.1). Moreover, the pre-trigger, trigger, and explanation structure was used to inform the development of the causal model (Sect. 4.3).

3 Related work

This section first introduces existing works that successfully apply robot-assisted therapy to support children with autism to demonstrate the feasibility of using humanoid educational robots in this context. We then briefly introduce the concept of visual perspective-taking as the identified application scenario and how children with autism have a different understanding of visual perspective than neurotypically developing children. To support our approach, we detail how causal explanations can offer valuable pathways to knowledge for children with autism before outlining how humanoid educational robots have the potential to combine interactive behaviours with a causal explanation engine to help children with autism better understand visual perspectives.

3.1 Robot-assisted therapy for children with autism

Support and feedback can help build confidence and motivation in children with ASD and can provide a foundation for further improvement [14]. By using humanoid social robots, children can practice their social interaction skills in a safe and controlled environment, without fear of negative consequences or judgement. Caregivers (e.g. therapists, teachers, and parents) can build on the interest displayed by children with ASD towards the robots and use them as mediator tools, tailoring the interaction to the specific needs of the children at any given time [2, 15, 16].

Therefore, the use of humanoid social and educational robots should be understood as a mediator tool for researchers and educators; they are to be used for improving interactive skills in children with ASD [17]. By providing opportunities for social interaction and practice, giving feedback and support [18], and creating a non-threatening and non-judgmental environment, humanoid social robots can play an important role in helping children with autism [13]. For example, a robot can be used for music therapy [19]. Accordingly, in our application domain, the robot will provide positive reinforcement and encouragement for successful attempts at VPT, or alternatively, constructive feedback – as in causal explanations – for areas that need improvement.

3.2 Visual perspective taking

Visual Perspective Taking (VPT) skills are an important aspect of social interaction and communication. They relate to the ability to see the world from another person's perspective, taking into account what they see and how they see it [20]. VPT refers to a person's understanding that other people might have a different line of sight than themselves, and to the understanding that two people viewing the same item from different points in space may see different things. VPT skills have two different levels, which are typically developing in succession [21]. VPT level 1 (VPT1) is the ability to understand what another individual can and cannot see, i.e. if an object is occluded from their view. VPT level 2 (VPT2) is the more advanced ability to understand that two or more people looking at the same object from different positions might not see the same thing [21].

Children with ASD often struggle with VPT [22]; this can impact their ability to understand and respond to the perspectives of others. As a result, some social interactions may prove challenging for children with ASD. However, recent research has shown that humanoid social robots can help autistic children improve their VPT skills [23]. Our retrospective study (Sec. 2) also identified Visual Perspective Taking (VPT) as a set of skills where children with autism are frequently supported by causal explanations, indicating that children benefit from them when improving these skills. Building on such previous studies that attest the usefulness of humanoid robots in VPT learning, our approach (Sect. 4.1), therefore, encompasses the design of scenarios where children learn about the different levels of visual perspective in interaction with an educational humanoid robot with a focus on support provided via *causal explanations*.

3.3 Causal explanations

Explanations are a key topic in robotics, as they play a vital role in building trust and enabling successful human-robot interactions [24], with explanations consistently ranked among the top preferred responses from robots. Likewise, causal explanations also help to make robots more predictable and explainable [25]. Similar to our context, approaches in fields such as explainable AI, employ models (e.g. decision trees [26]) to explain their own decision-making processes in a way that is understandable and interpretable for humans [27]. Our approach differs from this as we aim to provide an automatic explanation as to why the human in a human-robot interaction is not engaging as expected. Our work is based on the theory of actual causality [28], where, in a given scenario leading to an outcome, the events are analysed in order to find causes. This is in contrast with type-level causality, where general causal rules governing a system are sought. To our knowledge, our approach to causal explanation generation [11] is the first to apply the theory of causality to human-robot interactions. To provide some context about the use of actual causality, we summarise some of the most prominent related work below, which have, among others, been captured in a recent survey on relevant approaches that utilise the above notion of causality [29].

Leitner's theory of causality considers the temporal order as well as the non-occurrence of events [30]. It also provides a search-based, on-the-fly causality assessment that does not require the counterexamples to be generated in advance. In our proof-of-concept study, however, we could abstract away from the relative timing of events, since whether a particular event occurred before or after other events does not impact what we consider causes.

Causal analysis can also be used to explain counterexamples in hardware verification [31]. The proposed algorithm is implemented in the IBM RuleBase PE tool. Moreover, there are other theories of causality that are not concerned with a particular scenario; these include analysis of time series such as Granger's causality [32], which studies the possibility of a time series predicting another. These notions are in clear contrast with the notion employed in our work, where we start with a concrete scenario that comprises events leading to an effect. Our choice is justified by our context and, furthermore, by the possibility and scalability of its mechanisation. Moreover, Chockler et al. [33] employ a notion of responsibility (degree of causality [34]) to improve the quality of abstraction refinement by producing more efficient counterexamples. Besides the continuous aspects, our approach incorporates the modelling of platform (hardware), controllers (software) and environment into a single model that considers a high-level abstraction of the system. Incorporating a notion of responsibility is one of the directions for our future work to rank the explanations provided.

Particularly in the VPT context, demonstrating cause and effect is paramount in helping children and learners understand how their actions, positioning, or the environment influence what others can see or perceive. This understanding is critical for developing skills related to social cognition and problem solving. For instance, demonstrating how a change in position (cause) alters what is visible from another perspective (effect) makes abstract concepts of perspective tangible. In this work, we employ a model of causality to map causal relations between events in the interaction between children and Kaspar. This model enables the automatic provision of clear and concise explanations, which, in turn, allows children to more easily understand how others may see or experience situations differently.

4 Method

This section explains our approach to investigating whether the social educational robot Kaspar can be programmed to provide causal explanations to support children with ASD as they work on their VPT. We present a formative-summative evaluation where we have generated relevant scenarios (Sect. 4.1) and established a formal causal model (Sect. 4.3) to evaluate both in a formative cycle, looking at their general suitability with healthy adults when implementing them in a Kaspar robot. Eventually,

Table 1: Summary of educational games, their type of VPT and difficulty

Game	1: Bring animals	2: Animal cube	3: Turn Kaspar	4: Turntable
VPT type	I	II	I	I & II
Difficulty	Medium	Difficult	Difficult	Difficult

the entire system is evaluated in a summative cycle with an experiment involving children with autism at a special needs school (cf. Sect. 4.4). The two-stage evaluation approach combines research and expertise in robot-assisted therapy for children and formal causal models with the experience and situatedness at Garston Manor School, a specialist school for autism, learning and speech and language difficulties.

4.1 Scenario generation in formative cycle

From the situations identified in the retrospective study, we developed and fine-tuned four interactive games for the children to learn about and train VPT as summarised in Table 1. These games were designed to cover the different aspects of visual perspective as outlined in Section 3, in ascending order of difficulty. All games are administered in three consecutive trials to foster learning and reinforce explanations.

Game 1: *Bring toys into Kaspar's field of view.* This scenario is inspired by an earlier game described in [13], where Kaspar asks children to show him a specific plush animal from a selection of animals that are scattered around the room. Causal explanations when Kaspar could not see the animal have been identified and added to the game. Some of the explanations include "I cannot see it because you are holding it too low.", see Section 4.3. This game aims to develop VPT Type I, implicitly asking the child to assess where the limits of Kaspar's field of view are.

Game 2: *Show animals on the cube.* This scenario is also inspired by an earlier game [13]. Children are given a six-sided cube that has one animal on each side, where they are required to understand that Kaspar sees the animal on the side of the cube that is opposite from what they see themselves, a Type II VPT skill. The game was modified so that Kaspar can provide causal explanations such as: "I can see a picture, but you are holding it the wrong way around, can you turn it so I can see the animal?".

Game 3: *Turn Kaspar's head to see toys.* In this game, Kaspar asks the child to turn the robot's head towards one of the animals placed in the back of the room behind the child, c.f [13]. Again, appropriate causal explanations were identified and added to the game, for example: "I cannot see it yet. You didn't move my head to the right position. Please move my head a bit more so I can see it." Like Game I, this game asks the child to assess but also actively change the robot's field of view, looking at VPT Type I in a more actively engaging and challenging way.

Game 4: *Show toys on a turning table.* In this game, derived from [23], a physical separator device (turntable) is placed on the table between Kaspar and the child as depicted in Figure 3. The separators allow to create different positions where, at specific points in time, different animals can be seen either by the child or by Kaspar or by both. Several of the animals in the room can be placed on the turntable, and the researcher could move the turntable to different positions and ask the child questions about the visibility of the objects. In addition, Kaspar can request to see a specific

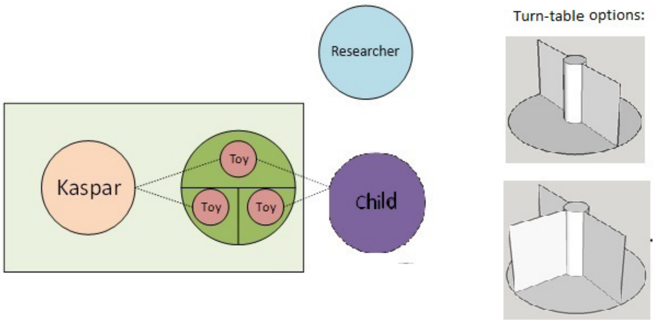


Fig. 3: Room set-up in Game 4 "*What can we see?*"

animal, and then the child would need to move the turntable to the correct position to make it visible to Kaspar. This game is classed as a VPT Type II exercise.

All of the games follow a simple dialogue structure that involves Kaspar and the child, beginning with the robot's instruction and ending with a commendation or explanation based on the child's response action. The following example provides the dialogue structure of the turntable game (Game IV):

1. Instruction: Kaspar says
   "Now, we will put some animal pictures on the turning table. My friend will move it into different positions to show or hide them. Then, we will find out who can see the animals".
2. Researcher moves the turntable
3. Question: Kaspar says "What animal do you think I can see?"
4. Response action: Child gives animal name.

5E. Explanation of incorrect action: Kaspar says "This is not the animal that I see."
5C. Commendation of correct action: Kaspar says "Well done!"

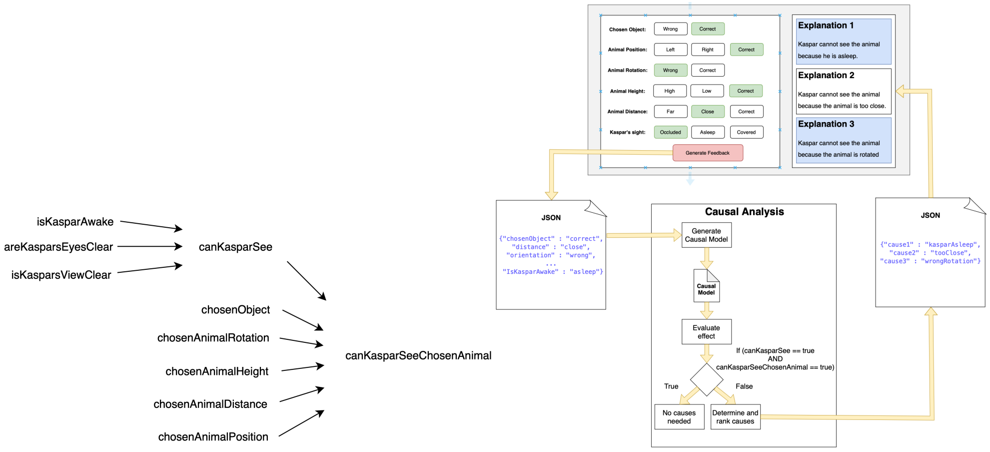
Similarly, the second part of Game IV follows the below dialogue:

1. Instruction/question: Kaspar says
   "I like to see the [animal], can you please turn the table so I can see it?"
2. Response action: Child moves the turntable

3E. Explanation of incorrect action: "That is not the animal I can see because the wall is in front of it."
3C. Commendation of correct action: "Well done. I can see the [animal]." [Sound of the animal]

Together with the teachers, three pupils and their parents, we confirmed the understandability of explanations and their applicability to the games to explore the utility of causal explanation and to provide an update to causal explanations generated, before the full validation verification in an interactive study.

4.2 Explanation generation with a formal causal model in formative cycle



(a) Causal model (b) Causal engine with decision process

Fig. 4: Kaspar Explains architecture overview consisting of a causal model and an engine that determines explanations based on causes indicated by the researchers.

In parallel to developing the interactive games, we used the scenarios extracted from the retrospective study and the derived games to develop an initial causal model. This model captured the most common interactions that can occur during the games and comprises the potential causes (such as mistakes) and the possible effects. For example, the fact that Kaspar's view is obstructed or that a picture is placed too much to the right are independently sufficient to allow for Kaspar to not see the object.

The initial causal model was then refined in several workshops by examining further hypothetical scenarios and mistakes that did not occur in the retrospective study. For example, a possible scenario is one where Kaspar's eyes are covered. For the case that a child would not notice this fact (by incorrectly answering a question about Kaspar's ability to see any animal), we added an explanation of why this would not be correct. Our final model, depicted in Figure 4a, consisted of the following rules and was used for all four interactive games.

- $\mathcal{F}_{canKasparSee}() = isKasparAwake = correct \wedge$
  $areKasparsEyesClear = correct \wedge$
  $isKasparsViewClear = correct$
- $\mathcal{F}_{canKasparSeeChosenAnimal}() = canKasparSee \wedge$
  $chosenAnimal = correct \wedge$
  $chosenAnimalPosition = correct \wedge$
  $chosenAnimalRotation = correct \wedge$

Table 2: List of causes ordered by their magnitude. Pre-triggers contributing to a single cause are indicated alphanumerically.

Rank	**Cause**
1	Kaspar is not able to see.
	(a) Kaspar is asleep.
	(b) Kaspar's eyes are closed.
	(c) Kaspar's eyes are covered.
2	The presented animal is incorrect.
3	The animal's latitude (left/right) is incorrect.
4	The animal's altitude (high/low) is incorrect.
5	The animal's distance (close/far) is incorrect.
6	The animal's rotation is incorrect.

$chosenAnimalHeight = correct \wedge$
$chosenAnimalDistance = correct$

This set of rules determines Kaspar's ability to see the correct object; it considers the robot's to see altogether (i.e. that he is not wearing a blindfold, not asleep, and his view is not obstructed) and also, whether the relevant object is being shown correctly concerning, for instance, its position, rotation, and distance. Based on this causal model, we developed a causal analysis engine that, in all four scenarios, automatically determines causal explanations that relate to causes identified by the researcher during an interaction with Kaspar. An overview of this engine is depicted in Figure 4b.

It is noteworthy that some scenarios might require multiple explanations; for example, if Kaspar is asleep and the chosen animal is too far away. Our causal analysis selects an explanation based on a pre-determined ranking of causes and presents the top explanation (e.g. Kaspar cannot see the animal because he is asleep) to the child. The ranking was determined by the research team assessing the magnitude of the cause with respect to the type of VPT as depicted in Table 2.

The top-ranked explanation is repeated twice in case the child does not correct the mistake to accommodate for problems in understanding. If the causes remain the same, the next explanation (e.g. that Kaspar cannot see the animal because the picture is too far away) on the above list is given subsequently. As a fallback option, the system also allows researchers to skip to the next explanation sooner. A detailed description of the causal model and engine can be found in our earlier publication [11].

4.3 Formative assessment of causal explanations

To assess the general understandability of the explanations listed in Table 2 and thus their suitability for the summative evaluation in school, we validated them in a formative study involving healthy adults. The study has been approved by the University of Hertfordshire's ethics committee for studies involving human participants, protocol number: SPECS/SF/UH/04944. Participants were provided with an information sheet describing the study. Implied consent was obtained at the beginning of the survey, giving participants the option to withdraw from the study at any time.

We asked 20 adult participants to watch all the videos of Kaspar providing all possible explanations in one of the selected scenarios and then rate each explanation using the Explanation Satisfaction (ES) scale [35]. This questionnaire evaluates the key attributes of explanations such as whether they are understandable, satisfying, sufficiently detailed, complete, informative about the interaction, useful, accurate, and trustworthy. Confirming that the explanation scored well in these attributes allowed us to assess the suitability of each explanation provided by an autonomous system. We used "what Kaspar can see" as the construct for the ES scale, which was shown to the participants for each video. We have additionally employed the Negative Attitude towards Robots Scale (NARS) [36] to calibrate the obtained results against potential biases against robots. That allows us to later compare the current study with future studies targeting different user groups, such as children. No other data has been collected. In total, we have shown 16 videos (available at: https://bit.ly/ke-vs1-videos) to participants that contain all possible explanations for the variables of the causal network (Fig. 4a) of the games identified in [10] and described in Section 4.1. More detail and an overview of the videos and descriptions of the utterances that Kaspar uses can be found in our earlier publication [11].

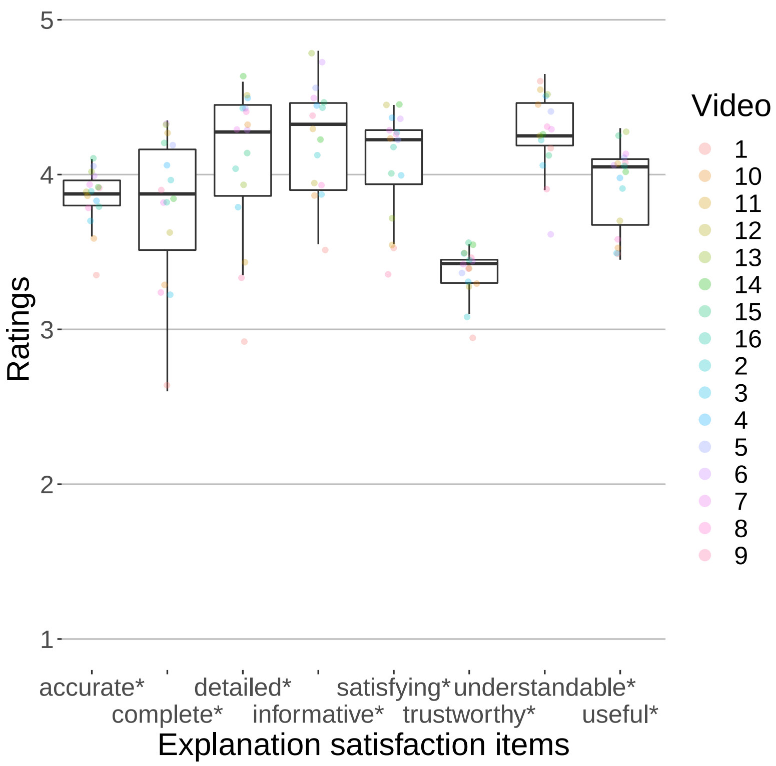


Fig. 5: Results of the ES scale (5-point Likert scale) grouped by explanation property. Coloured points indicate the mean values of the individual videos. Asterisks mark items significantly greater than the average value on the scale (3).

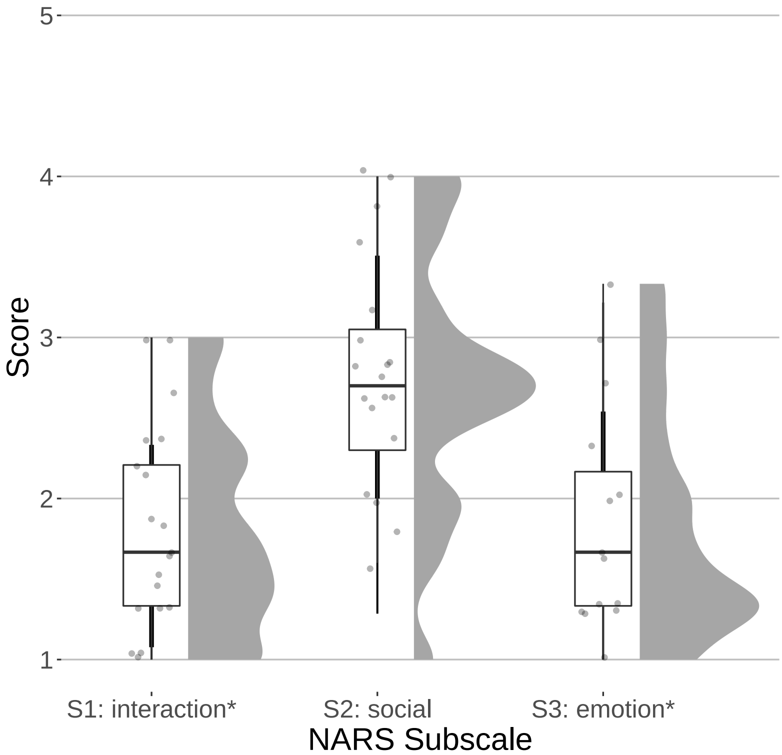


Fig. 6: Results of the NARS scale (5-point Likert scale) grouped by subscale. Points indicate the scores of the individual participants. Asterisks mark items significantly lower than the average value.

Because participant ratings were not normally distributed, we used the non-parametric one-sample Wilcoxon rank-sum test [37] to test whether ratings on the ES scale were greater than the neutral value, i.e. larger than the scale average (3), to see whether they are rated positively. Results attest that, when averaging across all the videos, each of the explanations is rated significantly above this neutral value (all $p < 0.001$), cf. Figure 5. Likewise, ratings across the explanations are rated above neutral for each of the videos (all $p < 0.001$). Participant ratings on NARS, as depicted in Figure 6, attested a low negative attitude towards robots with mean values for $\overline{S1} \approx 1.78$ (interaction subscale), $\overline{S2} \approx 2.7$ (social subscale), and $\overline{S3} \approx 1.48$ (emotion subscale). S1 and S3 are rated significantly below the neutral value (both $p < 0.001$), whereas S2 can not be reliably distinguished from neutral ($p \approx 0.053$).

These results confirm that, with healthy adults who are not negatively biased against robots, the explanations that the system can generate are beneficial to relate cause and effect. Participants consistently rate them as accurate, complete, sufficiently detailed, satisfying, understandable, useful to their goals, and informative about the interaction. Although explanation trustworthiness is rated lower than the other properties, the results also indicate that our explanations help to determine when to trust the robot. Knowing that adults find the generated explanations useful confirms that they could have the potential to help autistic children and enables us to use them in our summative evaluation in school.

4.4 Summative assessment of Kaspar's interactions mediated via the games

The educational games (see Section 4.1) and the causal model (see Section 4.3) were implemented and subsequently integrated into the Kaspar robot for a summative assessment of our approach. The evaluation process compared the scenarios with and without causal explanation at a local special needs school as previously reported [12]. The summative evaluation study was approved by the University of Hertfordshire's ethics committee for studies involving human participants, protocol number: SPEC-S/SF/UH/04944. Informed consent was obtained in writing from all parents of the participating children.

A sample of 10 children with ASD, selected based on the advice of their teacher, participated in three sessions with the robot where they engaged in various VPT tasks as part of the games. Sessions were distributed across different days in a two-week time window, leaving at least one day between each of the three experimental sessions for every child to allow us to observe short-term learning effects. The games presented were identical across all sessions. However, during some sessions, the robot offered additional constructive feedback by providing causal explanations related to VPT (explanation phase E), while other sessions served as the control where children experienced a typical interaction with Kaspar but without explanations (control phase C). We used a crossover design to evaluate the effectiveness of causal explanation, where children were randomly assigned to one experimental group that either experienced the sessions in order control - explanation - control (CEC) or explanation - control - explanation (ECE).

During the games, children were asked to respond to a total of 24 questions that were asked by Kaspar. Games 1-4 were presented in order of ascending difficulty [20], considering VPT Type I, e.g. out-of-sight positions and line-of-sight blockers, which are considered easier to understand (Games 1 and 3), to the more difficult VPT Type II tasks like understanding of different perspectives on the same object (Games 2 and 4). A total of 30 child-robot interactions (three sessions per child) were video-recorded and coded, identifying the correctness of the children's response in each VPT task, potentially following causal explanations, the incorrect answers (after a question and potentially after an explanation), rectifications (after an explanation or without explanation) and the total number of questions. The total number of mistakes and correct responses are grouped by experimental condition (control and explanation) and summarised in Table 3.

The ratio of correct actions (RC) over the total number of actions (both correct and incorrect) was taken as a suitable parameter for our analysis. In order to obtain this value, we used the equation $RC = \frac{c}{c+i}$, with c = the total number of correct actions and i = the total number of incorrect actions. The ratio of correct actions grouped

Table 3: Mistakes and correct responses for children's actions in educational games with Kaspar by experimental condition.

	Control (C)	**Explanation (E)**
Correct actions	238 (70%)	323 (82%)
Mistakes	100 (30%)	71 (18%)
Total actions	338	394

Table 4: Mistakes and correct responses for children's actions in educational games with Kaspar per session in each experimental group.

	Session	**Correct actions**	**Mistakes**	**Total actions**
CEC	1	56 (51%)	54 (49%)	110
	2	115 (79%)	30 (21%)	145
	3	89 (79%)	23 (21%)	112
	Total	**260 (71%)**	**107 (29%)**	**367**
ECE	1	100 (78%)	29 (22%)	129
	2	93 (80%)	23 (20%)	116
	3	108 (90%)	12 (10%)	120
	Total	**301 (82%)**	**64 (18%)**	**365**

by experimental condition is presented in Figure 7. Additionally, Table 4 details the totals and percentages of correctness and mistakes per session, i.e. in the order that children in CEC or ECE interacted with the robot.

We have performed a one-way ANOVA to identify higher-level effects between the two experimental groups, which showed a significant difference between the ECE and CEC groups for RC ($F(1, 28) = 4.461, p = .04, \eta^2 = .14$). An independent sample t-test was then used to find the underlying differences; the comparison of the two conditions C and E in the first session thereby revealed that there was a significant difference between the children who received causal explanations (who had a higher ratio of correct actions) and the children in the control session in their ratio of correct actions ($t(8) = -4.199, p = .003, 95\%$ CI, -0.43 to -0.13, Cohen's d $= 2.66$). Performing the same analysis to compare the two conditions in the second session resulted in ($t(8) = -.027, p = .979$) and again in the third session ($t(8) = -1.206, p = .262$), which indicates that after the first session, there were no more statistically significant differences between the two groups. However, we can observe a reduction of the p-value in session 3, showing that the differences between groups increased after session 2.

These results indicate a notable enhancement in the children's abilities when the robot offered causal explanations. Most notably, the ECE group showed a higher degree of correctness across all trials than the CEC group. This indicates that the ECE group, which had more sessions with explanations, performed better because of the higher total of explanations received by the robot. The biggest effect can be observed in the first session, when individuals either received an explanation or did not receive any explanation. The fact that this difference disappeared in the second session further strengthens our observation that explanations support the children's understanding because, in this session, all children have received explanations at least

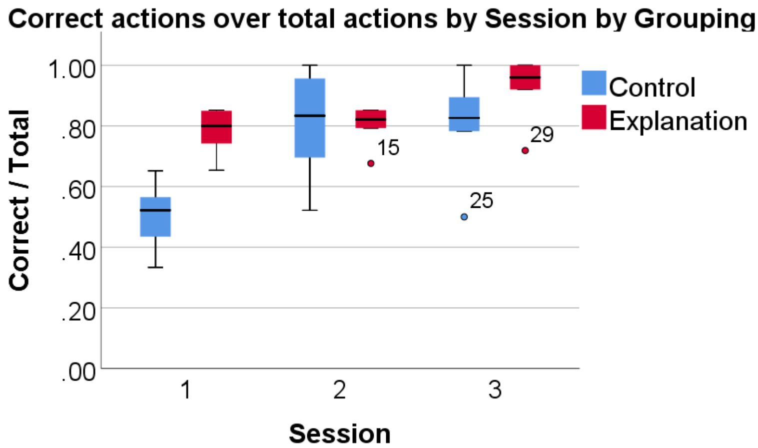


Fig. 7: Ratio of correctness for children's actions in educational games with Kaspar for each session of the trial grouped by experimental condition. Children experience the sessions in order CEC (blue-red-blue) or ECE (red-blue-red).

once. Our findings attest a significant role of causal explanations in fostering the children's understanding and suggest that employing a robot as an instructional tool for teaching VPT to autistic children can be particularly effective, especially when accompanied by causal explanations.

5 Discussion

This project focused on employing causal relations as the key ingredient in providing explanations during the interaction between the Kaspar robot and children with autism. Extensive analysis of previous studies has revealed the development of VPT skills as a crucial area where explanations might help particularly. We have consequently constructed and evaluated a model that builds on the theory of actual causation [28] to automatically provide accurate explanations for children's mistakes when interacting with the Kaspar robot. A simple interface for researchers thereby enables us to decouple the identification of children's mistakes from the production of a correct explanation, delegating the reasoning process to the causal engine. This allows for a more consistent form of feedback to reliably reinforce children with autism's learning. Section 4.3 shows that a causal engine that mathematically determines actual causality appears to be an appropriate model to support understanding of VPT. Explanations provided by Kaspar can thereby aim to enrich the interactions and improve children's learning.

While the scenarios and games we developed (cf. Sect. 4.1) are specifically tailored to address children's VPT skills, we believe that for more complex interactions, alternative causal explanations should be automatically ranked. More sophisticated models should, for example, take the brevity of explanations into account or children's understanding and reasoning capabilities, especially in cases of multiple simultaneous errors. While we have shown that generic rankings for explanations can already provide a workable solution, ranked explanations should be further empirically validated concerning each child's individual learning stage.

Moreover, this work is, to our knowledge, the first use of a rigorous methodology for determining causality in interaction between robots and children with special needs. This has provided us with some useful insights relating to the computational causality field. First, it demonstrated the clear effectiveness of applying causality as a tool for explanations to children. Secondly, although determining actual causality (i.e. using Halpern-Pearl's theory [28]) is computationally intractable [38], our methodology involved the use of mathematical proofs, which meant the assessment was very efficient; this is essential for smooth robot interactions. Lastly, another useful insight is the need for a ranking-based causal assessment for when multiple sufficiently independent causes have been found. This is somewhat supported by quantitative notions of causality such as the degree of responsibility [34] and, to a lesser extent, harm [39]; incorporating such notions may be a worthwhile venue for future work, specially for more complex (educational) scenarios.

The results presented in the summative evaluation (cf. Sect. 4.4) provide convincing evidence in favour of the use of causal analysis and explanations in interactions with Kaspar to support children's understanding of VPT. The difference between the experimental groups when receiving explanations for the first time and the disappearance of that effect when children have experienced at least one explanation, as well as the difference between children who receive higher total numbers of explanations, both indicate a direct learning effect of adding explanations to Kaspar's behaviour. Moreover, the increased variance of correctness in sessions without explanations could mean that some children begin to strongly rely on expectations and are either negatively affected by the lack of explanations or have understood the skills already after being exposed to the explanations only once.

The findings of this project have important implications for the design and implementation of interventions aimed at improving VPT in autistic children. The use of robots in this context can provide a more engaging and interactive experience for children, which could lead to better outcomes [23]. Following the results of the here presented study, researchers and practitioners may want to consider commonly using causal explanations when using robots for improving VPT in autistic children.

However, it should be noted that this study has some limitations. First, the sample size was relatively small, which limits the generalizability of the results. Second, the study only investigated the short-term effects of the robot intervention, and it is unclear whether the observed improvements in VPT would persist over longer periods. Therefore, future studies with larger sample sizes and longer follow-up periods are needed to further investigate the efficacy of robot-explained VPT interventions.

6 Conclusion

In this article, we presented a detailed analysis of how automatically generated causal explanations can support children with autism's understanding and learning. We first described the identification of an appropriate application domain in visual perspective-taking VPT where causal explanations offer a substantial value for children with ASD via a retrospective study looking at prior interactions with an educational robot. We then detailed the formulation of learning games together with the implementation of a causal model to generate relevant and understandable explanations within these scenarios. After a formative evaluation of the scenario and model, we lay out a summative study to test the feasibility of our approach in a school setting. There, we could confirm the positive effects of causal explanation generated by an educational robot on children with autism's learning and retaining of visual perspective taking. This was evidenced by children making fewer errors in interactions where the robot gave explanations and them maintaining their performance in following interactions where no such explanations were presented.

We believe a corpus of causal explanations provides a suitable knowledge base for expanding interaction scenarios towards individual therapeutic needs. We are hence keen to continue this work towards more complex scenarios and more autonomy in interaction, thus, allowing the creation of interaction scenarios with specific educational goals using the causal model. We also see potential for future work that further investigates how causal explanations promote additional trust in educational robots like Kaspar to further foster autistic children's learning.

Compliance with ethical standards

The authors have no conflicts of interest to declare that are associated with this article.

Principles of ethical and professional conduct have been followed in the here presented studies, which have been approved by the University of Hertfordshire's Health, Science, Engineering and Technology Ethics Committee with Delegated Authority (ECDA) under protocol numbers COM/SF/UH/02069, SPECS/SF/UH/4654(1), and SPECS/SF/UH/04944.

Informed consent was obtained in writing from all adult participants and parents of the participating children. Participants did not receive monetary compensation and were given the option to withdraw from the studies at any time without giving reasons.

Data availability

Data collected in the formative study can be made available upon reasonable request. However, all data collected during child-robot interactions is protected and is not available due to data privacy laws and the nature of consent given.

Acknowledgements

We would like to thank all families, teachers, and researchers who participated in our study. The study has been kindly supported by the UKRI Trustworthy Autonomous Systems (TAS) Hub (EP/V00784X/1) and the UKRI TAS Node on Verifiability (EP/V026801/2). We would like to further acknowledge that the retrospective study relies on data gathered in the Horizon 2020 project BabyRobot (687831).

References

[1] Wing, L. *The Autistics Spectrum: Aguide for parents and professionals* (Constable and Company Limited, 1996).

[2] Robins, B. & Dautenhahn, K. Kaspar, the social robot and ways it may help children with autism–an overview. *Enfance* **2018**, 91–102 (2018).

[3] Scassellati, B., Admoni, H. & Matarić, M. Robots for use in autism research. *Annual review of biomedical engineering* **14**, 275–294 (2012).

[4] Saleh, M. A., Hanapiah, F. A. & Hashim, H. Robot applications for autism: a comprehensive review. *Disability and Rehabilitation: Assistive Technology* **16**, 580–602 (2021).

[5] Rutherford, M. & Subiaul, F. Children with autism spectrum disorder have an exceptional explanatory drive. *Autism* **20**, 744–753 (2016).

[6] Legare, C. H. & Clegg, J. M. in *The development of children's causal explanations* 65–74 (Routledge, 2014).

[7] Robins, B., Dautenhahn, K., Boekhorst, R. T. & Billard, A. 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**, 105–120 (2005).

[8] Caldwell, P. *Getting in Touch: Ways of Working with People with Severe Learning Disabilities and Extensive Support Needs/c by Phoebe Caldwell* (Pavilion Publishing/Joseph Rowntree Foundation, 1996).

[9] Robins, B., Dautenhahn, K. & Dickerson, P. Sam Ge, S., Khatib, O., Cabibihan, J.-J., Simmons, R. & Williams, M.-A. (eds) *Embodiment and cognitive learning–can a humanoid robot help children with autism to learn about tactile social behaviour?* (eds Sam Ge, S., Khatib, O., Cabibihan, J.-J., Simmons, R. & Williams, M.-A.) *Social Robotics: 4th International Conference, ICSR 2012, Chengdu, China, October 29-31, 2012. Proceedings 4*, 66–75 (Springer, 2012).

[10] Sarda Gou, M. *et al.* Young Chong, N., Foster, M. E. & Venture, G. (eds) *Towards understanding causality - a retrospective study of scenarios that trigger explanations in interactions between a humanoid robot and autistic children.* (eds Young Chong, N., Foster, M. E. & Venture, G.) *International Conference on Robot and Human Interactive Communication (RO-MAN 2022)* (IEEE, Naples, Italy, 2022).

[11] Araujo, H. *et al.* in *Kaspar Causally Explains: Causal Explanation in an Assistive Robot for Children with Autism Spectrum Disorder* (eds Cavallo, F. *et al.*) *International Conference on Social Robotics (ICSR 2022)*, Vol. 13818 of *Lecture*

Notes in Computer Science 85–99 (Springer Cham, Florence, Italy, 2022).

[12] Sarda Gou, M. *et al.* Kyung, K.-U., Nakadai, K. & Barakova, E. (eds) *Kaspar Explains: The Effect of Causal Explanations on Visual Perspective Taking Skills in Children with Autism Spectrum Disorder.* (eds Kyung, K.-U., Nakadai, K. & Barakova, E.) *International Conference on Robot and Human Interactive Communication (RO-MAN 2023)* (IEEE, Busan, Korea, 2023).

[13] Zorcec, T., Robins, B. & Dautenhahn, K. Kalajdziski, S. & Ackovska, N. (eds) *Getting engaged: Assisted play with a humanoid robot kaspar for children with severe autism.* (eds Kalajdziski, S. & Ackovska, N.) *ICT Innovations 2018. Engineering and Life Sciences*, 198–207 (Springer International Publishing, Cham, 2018).

[14] Wainer, J., Robins, B., Amirabdollahian, F. & Dautenhahn, K. 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**, 183–199 (2014).

[15] Huijnen, C., Lexis, M. & De Witte, L. Robots as new tools in therapy and education for children with autism. *International Journal of Neurorehabilitation* **4**, 1–4 (2017).

[16] Wood, L. J. *et al.* Kheddar, A. *et al.* (eds) *The iterative development of the humanoid robot kaspar: An assistive robot for children with autism.* (eds Kheddar, A. *et al.*) *Social Robotics: 9th International Conference, ICSR 2017, Tsukuba, Japan, November 22-24, 2017, Proceedings 9*, 53–63 (Springer, 2017).

[17] Alabdulkareem, A., Alhakbani, N. & Al-Nafjan, A. A systematic review of research on robot-assisted therapy for children with autism. *Sensors* **22**, 944 (2022).

[18] Kim, S., Hirokawa, M., Matsuda, S., Funahashi, A. & Suzuki, K. Smiles as a signal of prosocial behaviors toward the robot in the therapeutic setting for children with autism spectrum disorder. *Frontiers in Robotics and AI* **8**, 599755 (2021).

[19] Feng, H., Mahoor, M. H. & Dino, F. A music-therapy robotic platform for children with autism: A pilot study. *Frontiers in Robotics and AI* **9**, 855819 (2022).

[20] Hamilton, A. F. d. C., Brindley, R. & Frith, U. Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition* **113**, 37–44 (2009).

[21] Flavell, J. H. The development of knowledge about visual perception. *Nebraska Symposium on Motivation* **25**, 43–76 (1977).

[22] Yirmiya, N., Sigman, M. & Zacks, D. Perceptual perspective-taking and seriation abilities in high-functioning children with autism. *Development and psychopathology* **6**, 263–272 (1994).

[23] Lakatos, G. *et al.* Robot-mediated intervention can assist children with autism to develop visual perspective taking skills. *Paladyn, Journal of Behavioral Robotics* **12**, 87–101 (2020).

[24] Wachowiak, L. *et al.* *When do people want an explanation from a robot?*, HRI '24, 752–761 (Association for Computing Machinery, New York, NY, USA, 2024). URL https://doi.org/10.1145/3610977.3634990.

[25] Brandao, M., Canal, G., Krivić, S. & Magazzeni, D. *Towards providing explanations for robot motion planning*, 3927–3933 (IEEE, 2021).

[26] Mahbooba, B., Timilsina, M., Sahal, R. & Serrano, M. Explainable artificial intelligence (xai) to enhance trust management in intrusion detection systems using decision tree model. *Complexity* **2021**, 6634811 (2021).

[27] Xu, F. *et al.* *Explainable ai: A brief survey on history, research areas, approaches and challenges*, 563–574 (Springer, 2019).

[28] Halpern, J. Y. *Actual Causality* (The MIT Press, 2016).

[29] Baier, C. *et al.* Bansal, N., Merelli, E. & Worrell, J. (eds) *From verification to causality-based explications.* (eds Bansal, N., Merelli, E. & Worrell, J.) *Proceedings of International Colloquium on Automata, Languages and Programming (ICALP 2021)* (2021).

[30] Leitner-Fischer, F. & Leue, S. Giacobazzi, R., Berdine, J. & Mastroeni, I. (eds) *Causality checking for complex system models.* (eds Giacobazzi, R., Berdine, J. & Mastroeni, I.) *International Workshop on Verification, Model Checking, and Abstract Interpretation*, 248–267 (Springer, 2013).

[31] Beer, I., Ben-David, S., Chockler, H., Orni, A. & Trefler, R. J. Explaining counterexamples using causality. *Formal Methods Syst. Des.* **40**, 20–40 (2012). URL https://doi.org/10.1007/s10703-011-0132-2.

[32] Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society* 424–438 (1969).

[33] Chockler, H., Grumberg, O. & Yadgar, A. Ramakrishnan, C. R. & Rehof, J. (eds) *Efficient automatic STE refinement using responsibility.* (eds Ramakrishnan, C. R. & Rehof, J.) *International Conference on Tools and Algorithms for the Construction and Analysis of Systems*, 233–248 (Springer, 2008).

[34] Chockler, H. & Halpern, J. Y. Responsibility and blame: A structural-model approach. *Journal of Artificial Intelligence Research* **22**, 93–115 (2004).

[35] Hoffman, R. R., Mueller, S. T., Klein, G. & Litman, J. Metrics for explainable ai: Challenges and prospects. *arXiv preprint arXiv:1812.04608* (2018).

[36] Nomura, T., Kanda, T. & Suzuki, T. Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. *Ai & Society* **20**, 138–150 (2006).

[37] Woolson, R. F. Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials* 1–3 (2007).

[38] Halpern, J. Y. *Actual causality* (MiT Press, 2016).

[39] Beckers, S., Chockler, H. & Halpern, J. A causal analysis of harm. *Advances in Neural Information Processing Systems* **35**, 2365–2376 (2022).