Development of Robot-enhanced Therapy for Children with Autism Spectrum Disorders

Project No. 611391

DREAM

Development of Robot-enhanced Therapy for Children with Autism Spectrum Disorders

Agreement Type: Collaborative Project
Agreement Number: 611391

D2.1.2 Tasks for social robots on developing social skills (Wizard of Oz system)

Due Date: **01/09/2015**
Submission date: **08/09/2015**

Start date of project: **01/04/2014** Duration: **54 months**

Organisation name of lead contractor for this deliverable: **Babes Bolyai University**

Responsible Person: **Daniel David**

Revision: **1.0**

Project co-funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	**PU**
PP	Restricted to other programme participants (including the Commission Service)	
RE	Restricted to a group specified by the consortium (including the Commission Service)	
CO	Confidential, only for members of the consortium (including the Commission Service)	

Contents

Executive Summary

This deliverable D2.1.2 (Tasks for social robots on developing social skills (Wizard of Oz system) is based on the results from studies carried out in task T2.1, in which we have tested the effectiveness of RET using a Wizard of Oz system. The effectiveness of RET is being tested for: joint attention, imitation, and turn-taking skills. The principal results in this deliverable are the parameters and parameter values that characterize the child behaviors identified in deliverable D1.3. These provide the basis for classification of behaviors, i.e. behavior assessment, in work package WP5 and, in particular, this deliverable provides the training set for the learning process that maps sensory cues to classes of behaviors. In this deliverable (D2.1.2) we will present: the theoretical background, objectives, design, procedure, environmental setup, results from the experiments carried in task 2.1, conclusions and discussions. This is a short version of the deliverable, and it has two Appendixes (*Appendix 1* and *Appendix 2*).

Principal Contributors

The main authors of this deliverable are as follows (in alphabetical order)

Cristina Costescu, Babes-Bolyai University
Daniel David, Babes-Bolyai University
Anca Dobrean, Babes-Bolyai University
Silviu Matu, Babes-Bolyai University
Aurora Szentagotai, Babes-Bolyai University

Revision History

Version 1.0 (07-08-2015)
First draft, describing the methodology without the results.
Version 2.0 (30.08.2015)
Results are included.
Version 3.0 (07.09.2015)
Revising conclusions.
Version 4.0 (08.09.2015)
Final version.

Introduction

Autism Spectrum Disorder (ASD) is characterized by widespread abnormalities in social interactions and communication, as well as severely restricted interests, and highly repetitive behavior (American Psychiatric Association, 2013). The most challenging deficits for treatment are social impairments, expressed as social communication (verbal and nonverbal; Jones & Klin, 2013), and social-emotional reciprocity (i.e., difficulties in understanding and expressing thoughts, feelings, intentions, and preferences - Theory of Mind) (Baron-Cohen, 1997; Baron-Cohen, Lombardo, Tager-Flusberg, & Cohen, 2013; Bons et al., 2013). All this social deficits limit the possibilities of individuals with ASD to engage in meaningful reciprocal interactions (Zwickel, White, Coniston, Senju, & Frith, 2011).

Therefore one of the main goals of the interventions specially developed for children with ASD is to engage children in social interactions in order to improve their abilities of communication and social-emotional reciprocity.

Several studies explored the potential value of social robots as interactional partners in social tasks with children with ASD (Dautenhahn, Te Boekhorst, & Billard, 2004; Vanderborght et al., 2012; Kim et al., 2013), especially because the social information that needs to be learned by ASD is presented in a manner that is easily understood and clearly identifies the expected behaviors, issues that are important for a successful learning of social behaviors (Quirmbach, Lincoln, Feinberg- Gizzo, Ingersoll, & Andrews, 2009; Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012). Another possible explanations for children's' preferences for technological tools may be inferred from the *Empathizing-Systemizing (E-S) Theory* of psychological sex differences, proposed by Baron-Cohen, (e.g., Baron-Cohen, 2009; Baron-Cohen, Knickmeyer, & Belmonte, 2005), which states that human males have stronger systemizing tendencies (i.e., analyzing a system in terms of the rules that govern it, in order to predict its behavior) compared to females, who exhibit stronger empathizing tendencies (i.e., the drive to identify another's mental states and to respond to them appropriately) (D2.1.1: David et al., 2015). Therefore the choice for a robot-enhanced approach for psychological intervention in ASD is justified by several advantages derived from the above mentioned theories, such as the embodiment of the robot offers human like social cues, while keeping at the same time object-like simplicity. This issue is a really important one when addressing children with autism, especially because the complexity of human behaviors may not allow them to learn properly when interacting with people. Moreover, robots work like systems that have rules which can be governed, allowing the users to predict its behavior, and the operators to gradually increase the complexity of the tasks. Another issue is that the tasks developed in this paradigm can be repeated in similar formats without trainer fatigue.

Studies investigating the use of social robots for children with ASD addressed a wide range of social skills, using different designs of robots and different scenarios. Regarding the types of abilities investigated, the three major reviews in the field of using social robots for children with ASD suggest that the most investigated abilities in a child-robot interaction are the following: imitation, joint attention, gaze, eye contact, attention focus, communication, play, turn-taking, and emotional reactions (Cabibihan, Javed, & Aljunied, 2013; Diehl, Schmitt, Villano, & Crowell, 2012; Scassellati, Admoni, & Matarić, 2012). Despite a decade of research, the effectiveness of robot enhanced interventions in teaching social skills is not yet proven, because of the mixed results of the experimental studies and inconclusive data from qualitative studies. This may be due to the lack of methodological rigor in the research from this field.

In the following section we will summarize the most important studies that investigated the abilities that were also measured in our study. We have classified the investigated abilities in primary outcomes and secondary outcomes. Primary outcomes represent the performance of the child in the designed tasks, such as: (1) turn taking, (2) joint attention (response to JA and initiation of JA), and (3) imitation and they are supposed to underlie the clinical symptoms/signs of ASD. Secondary outcomes were measured in order to investigate the reactions of the children in the robot-child interaction, most of them measuring clinical symptoms/signs of ASD.

Turn-Taking (TT). Turn-taking skills are frequently targeted in social skills and play interventions for children with disabilities. Turn-taking refers to "smooth interchanges between communicative partners" (Goldstein, Kaczmarek, & English, 2002), and includes behaviors like: rapid turning, avoidance of overlaps, observance of tasks that are needed to be responded, and topic-relatedness (Stanton-Chapman & Snell, 2011). Successful turn-taking requires children to maintain their turn, which is often very challenging for children with ASD when interacting with peers or with adults. One of the studies that tried to investigate collaborative play in children with ASD when interacting with a robot showed that children were more entertained, seemed more interested in the game, and collaborated better during their second sessions of playing with a human than their first, which may be due to the children's intermediary play session with the robotic partner (Wainer, Dautenhahn, Robins, & Amirabdollahian, 2010). However, while the children seemed to see their robotic partner as being more interesting and more entertaining than their human partner, they seemed to work better with the human partner for solving the game. Pop, Pintea, Vanderborght, and David (2014) found similar results when investigating play skills, such as children with ASD did not perform better in a functional play task with the robot compared to an adult partner, however they seemed to enjoy better the interaction with the robot compared to the interaction with the human partner, spending a longer time in the collaborative task. Also in the case of, social verbal communication (i.e. contingent utterances and verbal initiations), there are some studies showing mixed results. For example, regarding verbal social communication, Kim et al. (2013) reported that the interaction with a social robot elicited speech directed toward an adult confederate, and also toward the robot. Similarly, robot-based activities (i.e. working with peers on programming the robots in an afterschool program) elicited social interactions in 12 years old high-functioning children with ASD (Wainer, et al., 2010). On the other hand, Huskens, Verschuur, Gillesen, Didden, and Barakova (2013) revealed that the number of self-initiated questions did not differ between a robot and a human trainer.

Joint Attention (JA). Regarding our second targeted ability, joint attention (JA), it is considered to have a pivotal role in the development of language and social skills for children with ASD (Murray et al., 2008; Whalen, Schreibman, & Ingersoll, 2006). Comparisons between children with ASD and other populations of children, both with and without disabilities showed that children with ASD display deficits in eye-gaze shifting deficits, in gestural JA, and show that they are less responsive to JA episodes (Charman, 2003; Loveland, 2007). Emerging research in robot enhanced interventions shows that robots can be used to elicit JA episodes (Robins, et a.,, 2004; Kozima, et a., 2007; Robins et al. 2009). These conclusions were also highlighted in a more recent study developed by Warren et al., (2013), where they found that, across a series of four sessions, five of the six children with ASD exhibited lower average levels of prompt in their ability to orient to JA prompts administered by the robot. On the other hand, on a larger sample of children with ASD (n=16) and when compared to a human condition children with ASD had a lower performance with the robotic agent during a JA elicitation task (Anzalone et al. 2014).

Imitation. In regard to the first ability investigated, imitation, it plays a very important role in development, laying the foundation for communication and language. There is considerable evidence for imitation deficiencies in ASD children (Williams, Whiten & Singh, 2004). There is also empirical data suggesting that repeated sessions of imitation leads to improvements in imitation skills and to enhanced social responsiveness in children (Field, Field, Sanders & Nadel, 2001). To what it concerns the robot enhanced tasks designed by different research groups to investigate the imitation ability, studies have shown that individuals with ASD might benefit from tasks that involve imitating robots in comparison to imitating humans and that participants showed preference for robot-like characteristics (e.g. Dautenhahn & Billard, 2002; Robins, Dautenhahn, Te Boekhorst, & Billard, 2004; Kozima, Nakagawa & Yasuda, 2007; Pioggia, et al., 2007; Robins, Dautenhahn & Dickerson, 2009). However, studies that had a more rigorous methodology involving also quantitative measures tend to contradict these results. For example, Duquette, Michaud and Mercier (2008) conducted single case experimental study involving four children with ASD and they found that imitation of body movements and of familiar actions are higher with when the interactional partner is a human compared to a robot as interactional partner. However, they found increased shared focused attention (visual contact, physical proximity) and more positive affects when interacting with robot. Similar results were found also in the study conducted by Tapus, et al. (2012) who present a series of 4 single subject experiments aimed to investigate if children with ASD show more social engagement when interacting with the robot, compared to a human partner in a motor imitation task. Their results suggest that for two of the children no differences were observed between the robot and human condition and for other 2 children, there were difference in favor of the robot for eye gaze and smile/laughter when compared to the human partner, but not for the imitative gestures.

Our objectives are to teach imitation, JA, and turn-taking behaviors during repeated sessions of interactive games using social robots. This training is expected to lay a foundation for developing a set of implicit rules about communication, rules that will be transferred to the interaction with human persons. The research questions is to identify to what degree social robots (using Wizard of Oz system) can improve JA, imitation skills, and turn-taking skills and whether or not these type of intervention provides similar or better gains that standard intervention.

Method

Participants

We have recruited 11 participants with ASD aged between 3 and 5 years old from Autism Transylvanian Association, out of which 7 participants were selected based on the inclusion criteria to follow the intervention sessions. The inclusion criteria were: a. children are diagnosed with ASD using DSM-V criteria by a psychiatrist, b. their diagnosed is confirmed by the Autism Diagnostic Observation Scale (ADOS) (Lord et al., 2000), adapted into Romanian by our group (David, Anton, Stefan, Mogoase, & Matu, 2010), c. they have difficulties in performing the targeted behaviors in the evaluation phase. The IQ scores were assessed with the Snijders- Oomen Non-verbal Intelligence Test Revised Version (SON-R 2½-7), (Tellegen, Winkel, Wijnberg-Williams, & Laros, 1998). The minimum IQ score was 73 and the maximum score was 103 for the assessed children. The ADOS was administered by two experienced clinical psychologists, certified by the National Board of Psychologists.

(See the detailed description of the participants in Appendix 1)

Design

We are using single case experiments in order to assess the effectiveness of robot enhanced interventions for children with ASD, which represent a valuable, common and evidence-based methodology for clinical research (Janosky, Leininger, Hoerger, & Libkuman, 2009; Riley-Tillman & Burns, 2009). A classic single-case alternative treatments design (Barlow & Hayes, 1979) is being used.

Environmental Setup

In both robot-enhanced treatment (RET) condition and standard human treatment (SHT) children are interacting directly with the robot/human. In the RET condition the robot is sitting on the table and in SHT condition the therapist is sitting on a chair behind the table. In the right part of the room the operator is controlling the robot's movements by using a Wizard of Oz paradigm. In both of the tasks there is the third person in the interaction, who mediates the interaction between the interactional partner (robot or human) and the child. The role of the mediator in the task is to connect different moments of the protocol and to provide the necessary prompt. The video cameras and sensors are placed in the experimental room, behind the robot, capturing the facial expressions, the gaze and the movements of the children, as they interact with the robot/human.

Procedure

We are implementing the tasks following the discrete trial format, a commonly used approach in early intervention programs for autism (Ingersoll, 2008). This approach targets several skills, which are later used to teach more complex behaviors (Ingersoll, 2008; Lovaas, Freitas, Nelson, & Whalen, 1967). The elements that characterize this approach are: the teaching environment is highly structured, behaviors are broken into discrete sub-skills, which are presented over multiple, successive trials; the child is taught to respond to the partner's discriminative stimulus (e.g. "Do like me!") through explicit prompting; prompt fading; and contingent reinforcement (see Ingersoll, 2008).

Following the single case experiment design each child will go through the next scenario:

- Baseline measurements (BM; for imitation/JA/turn-taking) for approximate 6 to 8 measurements, until a stable baseline level has been established;
- Robot-enhanced treatment (RET; for imitation/JA/turn-taking) for approximate 8 sessions;
- Standard human treatment (SHT; for imitation/JA/turn-taking) for approximate 8 sessions;
- RET or SHT depending on which of the treatments worked better for each child, for approximate 4 sessions.

Each session lasts between 5 and 20 minutes (depending on the task) and is being delivered daily. The order between RET and SHT are randomized to avoid order effects. The baseline sessions for each ability are delivered in interaction with the therapist and they respect partially the discrete trial format. The child is tested to see if he/she responds to the partner's discriminative stimulus, without explicit prompting, prompt fading or contingent reinforcement.

The turn-taking task was developed by using the 'Sandtray' (see description in the section below). On the large touchscreen was running a scenario that provided a collaborative activity which acts as an interaction medium for the interactional partner and the child. This setup enables the pair to work collaboratively on the sorting game that was developed in order to complete this task. During the turn-taking task on the screen appears an emotional expression (either sad or happy) and the child has to match this facial expression with one of the categories from the left or right of the screen (the sadness category or the happiness category). The chosen stimulus set is sad and happy faces of children selected from the NIMH-ChEFS data base developed by Egger et al., 2011.

The JA task consisted in the interactional partner using one of the following methods: gazing; gazing and pointing; gazing, pointing and vocalizing at different objects in order to induce JA responses. Two different objects are placed on the table that sits in front of the child.

The imitation task consists of four different parts. The first part is represented by functional imitation with objects, and there are 4 different movements and sounds that the child has to imitate: moving a car, drinking from a cup, moving a plane and smelling a flower. The second part of the imitation task is the symbolic imitation with objects, which has the same four movements as the functional imitation task, only that this time instead of the real objects the child and his interactional partner use a wood cylinder pretending that it is a real object. The third part of this task is imitation without objects and it consists in four types of arms movements that are accompanied by sounds. The following movements were tested: putting both hands on the head, alternative lateral arm swinging, eyes covering with both hands, waving one hand. The last part of the imitation task consisted in imitating four basic emotions: happiness, sadness, fear and anger. The emotions were illustrated by using hand gestures, head movements and sounds. The order of the phases is as listed above.

All the three tasks had the same structure in the intervention sessions. The structure of each task is presented in Table 1. The standard treatment for each ability (turn-taking, JA and imitation), is delivered by the therapist respecting the discrete trial format. In the RET sessions the discriminative stimulus is delivered by the robot NAO. The robot also provides the contingent reinforcement based on the child's answer.

Table 1. The structure of each task: turn-taking task, JA task, and imitation task.

Instruction
Provided by the interactional partner (robot or human) Dependent on the task - imitation task: "Do like me!" - turn-taking task: "Now it is your/my turn." - JA task: looking; looking +pointing; looking + pointing + saying "Look!"
Response
Provided by the child Dependent on the task - imitation task: moving arms/objects and making sounds in similar ways as the interactional partner - Turn-taking task: waiting with his/her hands still when it is the turn of the interactional partner - JA task: looking in the direction indicated by the robot
Consequence
Provided by the interactional partner (robot or human) Depending on the child's answer - If the answer is correct then the child receives positive feedback: "Well done!" - If the answer is incorrect then the child receives negative feedback: "Try again!"

In the intervention sessions each child receives the instruction twice; if the child answers correctly to both instructions, she/he receives the instruction one more time. If the child makes a mistake in at least one of the first two instructions, the third time when receiving the instruction the child receives a prompt (physical, verbal or indicative) in order to complete the task, and after that the child receives the instruction a fourth time. The child receives feedback from the interactional partner every time he/she executed an instruction, except when he/she receives a prompt to execute the instruction.

In order to avoid interference between different skills, the interventions for the three abilities are implemented sequentially for each child, meaning that we focus on one ability at a time and move to the next one only after ending the intervention for the previous one. The order of the target abilities was randomized across children.

All variables were manually coded using the program Elan – Linguistic Annotator, version 4.5 (Lausberg & Sloetjes, 2009). Variables were only assessed during the task and not during the introduction or demonstration phases. All the therapy sessions were recorded using two Kinect sensors (Microsoft Corporation) and three high resolution cameras (1280*960pixels) that were all connected to a workstation central. The coders were trained in data collection procedure. The training consisted in giving clear definitions of the dependent variables, in offering examples and non-examples for each category of behavior. Training continued until the inter-observer agreement reached 80% on two successive observations.

Platforms used for delivering the intervention

For this study we are using the main DREAM experimental platform – the humanoid robot NAO developed by Aldebaran Robotics (Gouaillier et al., 2009). NAO is 58-cm tall, has 5-kg in weight and 25 degrees of freedom for movements. It is equipped with a rich array of sensors: 2 cameras, 4 microphones, sonar rangefinder, 2 IR emitters and receivers, 1 inertial board, 9 tactile sensors, and 8 pressure sensors. NAO has various communication devices including LED lights, two loud-speakers, a voice synthesizer with language-specific intonation and pronunciation. However, the Romanian voice is not yet available on the NAO platform so that a pre-recorded human voice with sound processing effects was used in these experiments.

An additional technological tool integrated in this research was the electronic "Sandtray" developed by the team from Plymouth University (Baxter, Wood, & Belpaeme, 2012). Inspiration for this platform is drawn from the "sandbox" technique in child therapy where sand play is used to foster collaborative story-telling interactions between child and therapist (Bradway, 1999). The interaction platform described here uses a touchscreen as opposed to a sandbox. However, it allows social engagement through a collaborative interaction platform. The hardware consists of a 26-inch capacitive touchscreen and associated control server, upon which a series of pictures can be manipulated by dragging (on the part of the human partner), or simulated dragging (on the part of the robot partner). The touchscreen thus serves as a medium for collaboration.

Measurements

These measures were also presented in the intermediary version of this deliverable (D2.1.1: David et al., 2015).

Primary outcomes

In order to assess the child **turn-taking skills** we are using a sorting task game with facial expressions which is played on the Sandtray, while we are coding the following behaviors:

Turn-taking	1 – Child plays the game with the robot/adult and respects turns when playing with the robot/human 0 - Child does not wait his turn turns when playing with the robot/human
Turn-taking engagement: showing enthusiasm and eye contact	2 - Child shows enthusiasm, makes eye contact and respects turns when playing with the robot/human 1 - Child shows enthusiasm or makes eye contact when playing with the robot/human

In order to assess the child **JA** skills we are using the following behavioral grid:

Response to JA - looking	1 - Child reacts and turns his head immediately after the robot/human does it

	0- Child does not react/does something else
Response to JA –looking +pointing;	2 - Child reacts and turns his head immediately after the robot/human does it 1 - Child points immediately after the robot/human does it 0- Child does not react/does something else
Response to JA – looking + pointing + saying "Look!"	2 - Child reacts and turns his head immediately after the robot/human does it 1 - Child points or gives vocal instruction immediately after the robot/human does it 0- Child does not react/does something else
Initiation of JA episodes	2– Child tries to show something to the robot/human by integrating different ways of showing: using gaze, vocalization and pointing 1 - Child tries to show something to the robot/human by using only one behavior from different ways of showing: using gaze, vocalization or pointing 0- Child has no attempts to initiate any JA episode

In order to assess the child **imitation skills** (with or without using objects) we are using the following behavioral grid:

Imitation with objects- functional behavior	2 - Child imitates the functional behavior (movement and sound) 1 - Attempts of the child to imitate the movement OR sound 0 - Child does not react/does something else
Imitation with objects- symbolic behavior	2 - Child imitates the symbolic behavior (movement and sound) 1 - Attempts of the child to imitate the movement OR sound 0 - Child does not react/does something else
Imitation without objects	2 - Child imitates the behavior with movement and sound 1 - Attempts of the child to imitate the movement or sound 0 - Child does not react/does something else
Imitation of emotional gestures	2 - Imitation of the gestures made by the robot/human, gestures that refer to a specific emotion (anger, happiness, sadness and fear) which include hand movements, head movements and sound 1 - Attempts of the child to imitate the gestures (hand movement OR head movement) made by the robot that refer to a specific emotion 0 - Child does not react/does something else

Secondary outcomes

Besides these primary outcomes we also have some secondary outcomes that are relevant for every session with the robot or the therapist regardless of the task. Such as we have measured variables that are related to social interaction and communication (engagement, eye contact and verbal utterances); behavioral outcomes (stereotype behaviors, maladaptive behaviors and adaptive behaviors); cognitive outcomes (irrational and rational beliefs) and emotional outcomes (negative emotions, negative dysfunctional emotions, negative functional emotions and positive emotions).

We will present a detailed description of each variable that was measured as follows:

Secondary outcomes: social interaction and communication

Engagement in the task:

Rating	*Meaning*	*Description*
0	*Intense noncompliance*	*The child walked away from the place in which the robot/human interaction took place*
1	*Noncompliance*	*The child refused to comply with the experimenter's request to play with the robot/adult*
2	*Neutral*	*The child complied with instructions to play the game with the robot/adult after several prompts from the experimenter.*
3	*Slight interest*	*The child required two or three prompts from the experimenter before responding to the robot/adult.*
4	*Engagement*	*The child complied immediately following the experimenter's request to play with the robot/adult.*
5	*Intense engagement*	*The child spontaneously engaged with the robot/adult.*

Eye contact: looking at the upper region (not necessary at the eyes) of the robot/human for more than 3 seconds (measured in frequency).

Contingent utterances: verbal utterances (one word or a couple of words) that are in context, congruous with the interaction with the partner (e.g. yes-no responses, responses to the questions) (measured in frequency – the number of contingent utterances said by the child during the task).

Verbal initiations: verbal utterances (one word or a couple of words) that are in context, congruous with the interaction with the robot/human partner and adds a new information, including expansion, adding to the content of the robot/human utterances or introducing new related topics (e.g. ask some questions, makes references to their own personal experience; measured in frequency – the number of verbal initiations made by the child during the task).

Secondary outcomes: behavioural

Stereotypical behaviors: a repetitive or ritualistic movement (especially hand mannerisms), posture, or utterances (measured in frequency – the number of stereotype behaviors performed by the child during the task). Specific stereotypical behaviors that were coded for each participant of the study are listed in the description of the participants.

Adaptive behaviors – generally imply the skills needed for successful adaptation and in our study are measured through strategies that communicate an interest in the interactional partner and in searching different ways of communication: approach in order to get some help, orienting behaviors, and seeking comfort. For example, approach for help might be reflected by one of the following: trying to solve the problem by making statements and questions that are aimed at understanding the situation. Orienting behaviors might be reflected by: behaviors of orienting in the environment and looking to the experimenter. Seeking comfort might be reflected by: soothing/communication self-comforting, gesture, and seeking comfort/ contact.

Maladaptive behaviors – are those behaviors that interfere with effective adjustment (e.g., aggression and hostility). We have grouped them in three categories: behavioral distraction/avoidance, demands and aggression. Behavioral distraction/ avoidance could be indicated by: doing something else than focusing on the task, turning attention away from the task (e.g. shifting gaze, staring into space, laying his or her head on the table). Demands are reflected by: expressing requests to others to do something in a louder voice and with an imposing tone; socially inappropriate words. Aggression (direct and indirect) could be reflected by disruptive behaviors like: socially inappropriate actions directed toward the experimenter, or the robot (e.g. throwing objects, self-aggression, physically aggressive toward others or others' toys).

Secondary outcomes: cognitive

Rational/Irrational beliefs (definition by Dryden & DiGiuseppe, 2003).

1. Demandingness (irrational) *vs. preferences* (flexible but strong belief; rational). Demandingness refers to the tendency to make absolutist demands instead of (strong) wishes or preferences and is expressed in the form of "musts", "should", and "oughts".

2. *Awfulizing* (appraising an event as catastrophic or as the worst things that could happen; irrational) vs. *non- awfulizing* (evaluating an event in terms of badness, as extremely bad but not the worst it could happen; rational). Awfulizing refers to the tendency to evaluate events as being the worst in the world instead of evaluating them on a continuum of badness.

3. *Low frustration tolerance* (irrational) vs. *frustration tolerance* (rational). Low frustration tolerance refers to an individual's belief that he or she will not be able to endure a specific situation. Frustration tolerance implies that a situation might be appraised as extremely distressful and in need of change, but however endurable.

4. *Global evaluation of the self, others, and/or life (irrational) vs. non-global evaluation of the self, others, and/or life* (accepting and focusing on changing specific behaviors; rational). Global evaluation refers to instances in which individuals make generalized evaluations or denigrations (i.e., overgeneralizations) about themselves, others, or the entire world instead of accepting and focusing on conditionally accepting the human person/life and evaluating specific behaviors. All these beliefs were measured in frequency by analyzing the content of the children's speech during the task.

Secondary outcomes: emotional

Positive emotions: the child laughed or smiled while interacting with the robot/human (measured in frequency – the number of smiles or laughs performed by the child during the task).

Negative emotions: the child shows anger, sadness or fear while interacting with the robot /human (measured frequency – the number of facial expressions that express anger/sadness or fear performed by the child during the task). We have also made the distinction between dysfunctional negative emotions and functional negative emotion, based on whether the facial expression was and respectively, was not accompanied by disruptive behaviors.

Data analysis

The data-analysis plan includes the following steps:

1. Plotting and visual analysis of the data.

To facilitate interpretation, data from each phase was plotted along with trend and variability indicators, namely the mean line, the celeration line, and lines for ± 2 standard deviations (SD). Visual analysis is very useful in identifying broad patterns, extreme values, and trends that continue from one phase to the other. Sometime, higher (or lower) values in one condition as compared to the other might be interpreted straight forward, especially when there is low variability and the measures being conducted have clear clinical interpretation. However, this approach might be vulnerable to subjectivity coming from the clinician (or the persons conducting data analysis) and thus more rigorous criteria are need. Some approaches derived from statistical process control, make use of trend and variability indicators (e.g., celeration or OSL regression lines, lines based on sigma units or standard deviations) in order to asses more accurately the results that are being plotted (see Nugent, 2010 for a review). For example, the more scores in one condition go far beyond the range of variability in another condition, the more likely is that relevant difference are present between the two conditions. One guiding rule of interpreting the results (that we also implemented here, Gottman, & Leiblum, 1974) states that if at least two scores in one condition are above (or below) the two standard deviations line of another condition, than it might be considered that there are relevant differences between the two conditions / interventions. Despite the fact that such a rule is easy to be applied, there are many situations where it might be unusable, for example, when high variability is present in one of the two conditions being analyzed (or both), or in situations when a large number of values in one condition are above (or below) the values in another conditions (and the difference has clear clinical relevance) but not above the two standard deviation line. Thus, to overcome the limitations of visual analysis, we also employed statistical procedures.

Follow-up sessions were also plotted when data was available. However, this data was not included in the statistical analysis presented.

2. Statistical analysis for testing trends in data or changes in trends.

To check for possible trends in baseline data as well as to compare the effect of the two interventions we endorsed the C statistics. This test is regarded as sensitive in identifying trends in data coming from single case research as well as to changes in trends that might be due to the effects of experimental manipulations (Janosky, Leininger, Hoerger, & Libkuman, 2009; Jones, 2003). The C value divided by its error is compared to the normal distribution (the test is one tailed only; see Jones, 2003) and thus the probability for obtaining a certain pattern due to random chance is calculated. The test does not indicate the direction of the pattern and one must rely on the visual interpretation of the data to identify the direction.

Also, consideration should be given to changes in patters, as the test is sensitive to this situations.

3. Parametric comparisons between conditions.

We used non-parametric statistical analysis in order to compare the effect of the different phases on the target outcomes. We used statistical procedures based on rank differences, which have few assumptions for the data and are considered to be adequate for analyzing single case research data (see Janosky, et al., 2009). Specifically we computed a Kruskal Wallis omnibus non-parametric test comparing ranks in all three conditions (BM, SHT and RET), and if the results of this test proved to be significant, we continued with pair-wise comparisons with Mann Whitney U test for the difference of ranks. We choose to conduct an omnibus test first as we aimed to reduce the risk of Type I error due to the large number of significance tests we conducted.

4. Computation of effect sizes.

Finally, given that the non-parametric tests we used rely only on rank differences, we decided to compute a parametric indicator for the magnitude of the differences between different conditions. Thus, for each comparison we computed the *Cohen's d* measure of effect size, a commonly used indicator across clinical literature. When presenting the results, this indicator was reported only for statistically significant differences. However, at the end of the results for each child and each ability, a graph presenting all the effect size values is presented in *Appendix 2* (effect sizes plotted as red bars indicated effects sizes coming from comparisons what where statistically significant). The values of *d* are commonly interpreted in the following manner: $\leq .20$ is not relevant; between .20 and .50 is a small effect size, between .50 and .80 is medium effect size; $\geq .80$ is a large effect size. However, as one can see further, most of the effect sizes that reach statistical significance in these experiments had values much larger than the .80 threshold.

Results and Conclusion

(See the detailed description of the results in Appendix 2)

Turn-taking (TT)

Participant no 1: S.P. Both primary outcomes related to turn-taking (TT) and secondary outcomes are presented. After 8 sessions in each intervention conditions, the decision for which intervention will be carried on in the follow-up sessions was based on the *TT* Engagement score, which was coded online.

Both SHT and RET proved to be effective for this child on primary outcomes. Despite a small advantage for RET in comparison to SHT, no significant difference were found between the two conditions on turn-taking skills. On secondary outcomes, both interventions led to more frequent eye contact and higher levels of positive emotions. However, RET was also superior to both BM and SHT in relation adaptive behaviors.

Participant no 2: V.E.

Because of practical reasons, only 6 sessions were performed as baseline measurements for this child. There was a descendent data trend in BM for TT engagement which reached statistical significance ($p < .05$). This suggests that is very unlikely that the results for the two interventions were affected by an ascending trend line in BM that was not measured.

For this child, only RET was more effective than BM on one of the primary outcomes, namely, showing enthusiasm trough eye contact. On secondary outcomes, both conditions led to increases in eye-contact and verbal utterances, with a significant difference on eye contact, favoring RET. However, both also led to increases in irrational beliefs, but this time scores in RET were higher than scores in SHT. In the RET condition the child expressed more frequent stereotype and maladaptive behaviors as compared to BM, but also more frequent adaptive behaviors, as compared to both BM and SHT. This pattern of increases in both disruptive and adaptive behaviors was also seen in imitation experiments and might be explained by slower responses in the robot which could generate some form of frustration for the children.

Participant no 3: L.C.

For L.C. only SHT proved to be effective (superior to BM) for one of the primary outcomes, namely engagement in turn-taking play. RET showed lower scores on primary outcomes, as compared to both SHT and RET. Both interventions led to lower performances on engagement, and RET performance was lower even when compared to SHT. Yet, interaction with the robot led to significantly higher levels of positive emotions as compared to both other conditions, indicated that despite the low level of performance, the child enjoyed interacting with the robot.

Is worth mentioning that comparing with other participants, this child showed better levels of performance on primary outcomes, starting with BM, indicating a possible ceiling effect (little space for improvement was available).

Participant no 5: O.R.

Both interventions proved to be effective for this child, showing higher performances than BM. However, RET was also more effective than SHT. On secondary outcomes, both interventions lead to increased eye contact (similar between the two conditions). Yet, both interventions also led to more frequent stereotype behaviors, but this effect was somewhat compensated in RET, where adaptive behaviors were more frequent as well. The level of adaptive behaviors in RE was also higher than SHT.

Participant no 7: A.A.

Because of practical reasons no follow-up sessions were performed for this child. He exited the study just after ending the intervention phases. It is worth mentioning that this child had in general very low levels of performance (often the lowest level on the scale) in almost all BM sessions. This patter was visible for both primary outcomes, but also for secondary outcomes related to social interaction and communication.

Both interventions were effective on turn-taking engagement (leading to significantly higher scores than BM), but not for enthusiasm expressed through eye contact. Also, both interventions led to more frequent eye contact. Some negative results were also observable, especially for SHT, where negative emotion and negative dysfunctional emotions were more frequent as compared to BM. Also, negative dysfunctional emotions were more frequent in RET. SHT however, had a benefic effect on positive emotions, as the child expressed more positive affect as compared to BM. No statistically significant differences were observable between SHT and RET.

Overall, results for TT experiments indicate that SHT and RET were similar across primary outcomes (turn-taking skills). Most of the children benefited to a similar extent from both SHT and RET when comparing to BM. Participant no. 3, L.C., expressed lower performance in interaction with the robot, while participant no. 5, O.R., benefited more from the technology enhanced sessions as compared to the standard treatment. Is worth noting that participant no. 3, L.C. was the only child that had higher performance levels for primary outcomes, starting with BM, and less space for improvement was available for this child. This suggests that children with prior TT skills would benefit less form RET (this pattern was replicated across the other skills as well).

A consistent pattern emerged for eye contact. Both interventions led to better performance on this variable, as compared to baseline level. However, is worth noting that the effect sizes were much larger in RET (in some cases two to three times larger) which led, for some of the children, to significant differences between the two interventions, favoring RET. A similar pattern with the results in imitation experiments was visible for behavioral outcomes. SHT did not have any specific effect on this variables, but RET had a contrasting one. On one hand, RET lead to some increases in stereotype and maladaptive behaviors, while at the same time it also led to increases in adaptive behaviors. In fact, the effect on adaptive behaviors in RET was higher (and it reached statistical significance in most of the children) than the one in SHT. Therefore, interventions for ASD children integrating a robotic agent should dedicate special attention to possible problematic behaviors that might emerge in the interaction with the robot, and try to correct them. Some variations are visible on the other outcomes, but no consistent pattern can be identified.

Taken together, the results suggest similar benefits form RET and SHT on TT abilities. However, RET condition was also associated with more frequent eye contact.

General message for TT:

- For primary outcomes, RET seems to be as good as SHT, and both interventions were effective in improving TT skills (above BM for most children). Negative results for RET were present in a single case, where the child had good TT skills starting with BM (the intervention was less needed in this case).
- For secondary outcomes, RET is better than SHT in behavioral activation. However, this activation was related to both adaptive and maladaptive behaviors and thus clinicians should be aware (and use it as an opportunity) to reinforce the adaptive behaviors and decrease the maladaptive behaviors, once they are elicited.

Joint attention (JA)

Both primary outcomes related to joint attention (JA) and secondary outcomes are presented below. After 8 sessions in each intervention conditions, the decision for which intervention will be carried on in the follow-up sessions was based on an overall score for response to JA which was computed as the average performance in response to all JA triggers (head, head + point, head + point + vocal), which is also presented for each child. Overall performance for response to JA is a composite index for which psychometric properties are not known. Thus, this score cannot be used to assess the overall efficacy of the interventions being tested here.

Participant no 1: S.P.

It seems that none of the interventions were effective for increasing JA skills for this child. RET had lower performance scores than SHT on pointing and on the overall response to JA. Both conditions led to less frequent verbal utterances when compared to BM, while RET also led to lower levels of engagement in comparison to both BM and SHT. A positive effect was visible for RET on adaptive behaviors which were more frequent in comparison to both other conditions. A ceiling effect was present for some of the primary outcomes (response to head and pointing, response to head, pointing and vocal instruction), meaning that the child had very high levels of performance in BM, and little space for improvement was available on this outcomes.

Participant no 2: V.E.

None of the interventions were effective for V.E, for any of the JA skills. RET however showed lower scores than SHT and BM on JA prompted by, head movement, pointing and vocal cues. RET led also to lower engagement when compared to BM. Yet, the child expressed more frequent eye-contact in RET as compared to BM. A pattern seen previously was also present here. In the robot condition, the child expressed more frequent stereotype behaviors, but also more frequent adaptive behaviors, and these effects emerged in comparison to both SHT and BM. Intriguingly, the child reported more frequent rational beliefs in BM as compared to both interventions.

A ceiling effect could be observed in BM on almost all primary outcomes, with the exception of initiation of JA episodes. This suggests that little space for improvement was available for these outcomes.

Participant no 3: L.C.

L.C showed better performance on the overall response to JA index, in SHT, when compared to BM and RET. However, the positive effect was not present on any of the specific JA skills. Comparing SHT with BM on secondary outcomes, SHT was effective in increasing engagement and reducing maladaptive and stereotype behaviors. SHT led also to increases in positive emotions. On the other hand, RET was more effective than BM only on reducing stereotype behaviors, and increasing positive emotions. RET was inferior to SHT on engagement, but superior on positive affect.

Participant no 4: D.M.

D.M benefited from both interventions, as he showed higher performances on almost all outcomes, as compared to BM. Also, both interventions led to higher engagement and more frequent eye contact, as well as to lower levels of maladaptive behaviors. The only difference that emerged between the two interventions was related to positive emotions which were more frequent in RET.

Participant no 5: O.R

This child benefited from both interventions, as he expressed higher performance on JA response to head movement, as compared to BM. SHT was better than BM and RET on the overall performance index and also it overcomes RET on the head movement response. Both interventions lead to more frequent eye-contact, but SHT had a positive effect on engagement as well. The level of engagement in SHT was in fact higher than both BM and RET. However, RET was better for eye contact, as both SHT and BM had lower frequencies on this variable. Finally, SHT had also a positive effect on maladaptive behaviors, which decreased in comparison to BM.

Conclusions for JA experiments

Results for JA were mixed. First of all, comparing SHT with BM on primary outcomes we found that only one child has consistently benefited from this intervention (participant no. 4, D.M.). There are several positive results for SHT under overall response to JA performance, but this is an additive (mathematical) effect, coming from small improvements across different abilities, which did not change significantly in the first place. Thus, the efficacy of SHT for those cases where this effect emerged should be regarded with some caution. For secondary outcomes, a consistent pattern was related to improvements in engagement in the SHT condition (compared to BM), a positive result that was found in three out of the five children. The result was also positive for a fourth child (participant no. 1, S.P., a large effect size which did not reach statistical significance) but for the fifth child, the comparison on this variable indicated the reverse pattern). Another consistent pattern was also present for maladaptive behaviors. Four out of the five children expressed less maladaptive behaviors in SHT as compared to BM, and for three of them this comparison yielded statistically significant differences. For other variables, the results seem mixed, with some children benefiting from SHT, while others did not. Also, in some cases, SHT lead to negative effects.

The comparison between RET and BM, indicated that participant no. 4., D.M., also benefited from this intervention on several primary outcomes. A pattern that was present in most of the children was related to eye contact, which increased in interaction with the robot (this was also visible in the imitation experiments). RET lead to a reduction in maladaptive behaviors only for one child, but on the other hand, two other ones expressed an increase in adaptive behaviors, as compared to BM. Also, more positive emotions were expressed by two children, an effect which was not present in SHT, indicating that some children might enjoy more interacting with the robot, while other will express the same enjoyment with both the human mediator and the robotic agent. For other outcomes, results were mixed, some children benefiting from the RET intervention, while other did not, or did worse in this condition.

Finally, when comparing SHT and RET, few differences were apparent. No child that experienced more gains from SHT than RET on at least two variables. There are two children for which RET lead to an increase in adaptive behaviors, and two children for which the interaction with the robot lead to more positive affect than the one with the human mediator. However, gains on these variables should be treated with some caution, as these results are not very consistent.

Is important to mention that ceiling effect (meaning that the performance in baseline was close to maximum level of performance) was present for several primary outcomes, for several children, suggesting that most of them were highly functional on this ability. When space for improvement was available (see for example, participant no 4, D.M. across all outcomes, or participant no. 5, O.R., for response on head movement) RET and SHT were equivalent, and most of the times superior to BM. This pattern suggests that children with the highest need of improvement benefited from both interventions to a similar extent. Higher gains for SHT as compared to RET were present in high functioning cases. Thus, it seems that children with JA deficits could benefit from RET to a similar extend as from SHT.

General message for JA:

- For primary outcomes, most of the times, and especially for children with lower JA skills, RET seemed to be as good as SHT.
- For secondary outcomes, RET was better than SHT on some variables (e.g., adaptive behaviors, positive emotions).

Imitation skills

Both primary outcomes, related to imitation, and secondary outcomes are presented below. After 8 sessions in each intervention conditions, the decision for which intervention will be carried on in the follow-up sessions was based on an overall imitation score computed as the average performance on all imitation outcomes (primary), which is also presented for each child. Overall imitation performance is a composite index and its psychometric properties are not known. Thus, this score cannot be used to assess the overall efficacy of the interventions being tested here.

Participant no 1: S.P.

Because of practical reasons, only 6 sessions were performed as baseline measurements for this child. There was a descendent data trend in BM for the overall imitation performance, but it did not reach statistical significance ($p > .05$). This suggests that most likely the effects of the two intervention conditions were not confounded with an increasing baseline trend that was not identified.

Results showed that none of the two interventions led to increases in imitation performance for this child. The child showed lower scores during RET sessions as compared to both BM and SHT for all primary outcomes, with the exception of imitation with objects, were all three conditions were similar. Is important to have in mind that SHT was not superior to BM on any of the primary outcomes, although a large effect size was present for imitation without objects. The maximum score (and the maximum average score) the child could have received on any of these outcomes was 2. S.P had high levels of performance starting with the first baseline measurements, and in general had a relatively stable pattern (with the exception of symbolic imitation, which decreased in the absence of any reinforcement). Thus, for this child is likely that a sealing effect was present for primary outcomes, meaning that little space for improvement was available from the beginning.

Results for secondary outcomes were mixed. Child's performance was better in the SHT condition for engagement as compared to RET, but similar to BM. For eye contact and verbal utterances, performance in RET was better than both SHT and BM. The child expressed more stereotypical and maladaptive behaviors in the RET condition as compared to both SHT and BM. Stereotype behaviors were also higher for SHT as compared to BM, and the same pattern was observed for maladaptive behaviors, where a large effect size was present but which did not reach statistical significance. However, the child expressed more adaptive behaviors in RET as compared to the other phases. For emotional and cognitive outcomes, no significant differences were found.

Participant no 2: V.E.

Results for primary outcomes related to imitation indicated significant descendent trend lines for baseline measurements, as the child started from the top values in the rage of possible values for imitation and descendent along BM. SHT seemed to be superior to RET on several outcomes, but these results emerged only when no differences between SHT and BM were found. When SHT was superior to BM, SHT and RET were equivalent. All these indicate that the child had a good previous level of imitation skills and none of the treatments were effective for primary outcomes, with the exception of symbolic imitations, where both SHT and RET were similar and superior to BM.

For secondary outcomes, no differences were found between BM and SHT. However, RET was superior to both other conditions for eye contact, verbal utterances, and adaptive behaviors. It was also superior to BM for positive emotions. However, despite positive results on adaptive behaviors, the child expressed more stereotype behaviors in the RET condition as compared to both BM and SHT.

Participant no 3: L.C.

No statistically significant differences emerged between SHT and BM on any of the primary outcomes, despite the fact that several large effect sizes favored SHT. The child had lower performance on imitation without objects when comparing RET with both BM and SHT. Also, the child had lower performance on imitation of emotion in RET when compared to SHT. The additive effects on these variable lead to lower performance in RET on the overall index of imitation, in comparison to SHT. Several descending trend lines were found across all the conditions, despite the fact that the level of performance at the beginning of BM was as the middle of the range of possible values. All these results suggest that the child did not benefit from any of the interventions for imitation abilities.

Results for secondary outcomes suggest that the child benefited more on eye contact and positive emotions from RET. This result emerged when compared to both SHT and BM. Is worth to mention that higher effect sizes were present for both adaptive and maladaptive when comparing RET with SHT, both variables showing higher values in RET. Such a pattern emerged also for other children.

Participant no 4: DM

Because of practical reasons, no follow-up sessions were performed for this child. The child left the study at the end of the 16 intervention sessions with SHT and RET. This child expressed a high level of imitation performance in BM, thus conclusions regarding the efficacy of any of the intervention conditions need to be made with caution. The child expressed lower levels of performance in SHT and RET as compared to BM for imitation with objects. Also, BM was superior to RET for imitation of emotion and for overall imitation performance. This difference on overall imitation performance is an additive effect of all other primary outcomes for which the child did better in BM as compared to RET, but a similar pattern was present for SHT as well.

For secondary outcomes, results indicated that the child expressed more verbal utterances and more positive affect in the RET condition as compared to both BM and SHT.

A curious result is that the level of maladaptive behaviors was higher in SHT and RET as compared to BM.

Participant no 6: S.V.

Due to practical reasons, only six BM sessions were performed with this child. No significant trend was found in BM (p > .05) and the child showed high levels of performances. We continued the intervention session with RET to see how the child would react to this innovative intervention and if the level of performance could be maintained.

For primary outcomes this child had similar performance across all conditions, with the exception of imitation of emotions, were better performance was identified for SHT when compared to BM. A ceiling effect was present across primary outcomes, as the child expressed top performance in almost all sessions, including BM. The sole exception was also imitation of emotions were BM started from lower levels, but also reached top performance in the final session. Despite the fact that is hard to assess the efficacy of the two interventions due to good prior imitation skills, both conditions maintained this level of performance (it was actually higher, almost always at the top level performance, but it could not reach statistical significance).

For social and communication secondary outcomes, the child showed more verbal utterances in the SHT and RET conditions as compared to BM. Is worth mentioning that the effect for the comparison between RET and BM was more than four times larger than the effect for the comparison between SHT and BM. Also, the child made more frequent eye-contact during RET as compared to both SHT and BM. On the other hand, for behavioral outcomes, in the RET condition the child also expressed more stereotype behaviors when compared to both SHT and BM, and more stereotype behaviors when compared to SHT. Similar to pattern seen in other children, adaptive behaviors were also more frequent in RET in contrast to BM. Finally, positive emotions were more frequent in SHT and RET as compared to BM, but the effect size for RET was four times higher than for SHT.

Conclusions for imitation experiments

Results for imitation skills (primary outcomes) showed that for several children RET was not as good as SHT. On the other hand, SHT was not superior to BM, and thus it seems that none of the interventions was effective for enhancing imitation skills. However, any conclusion based on these results is very limited, as a ceiling effect in BM was present in four out of five children, suggesting that little room for improvement was available in the first place. The fact that children started with a good level of performance in BM might be explained by the fact that all children were going through a classical therapeutical program (imitation skills are often targeted by such programs), outside of what was offered in the present research. Thus, is also possible that the children have already learned to imitate human behavior and this interfered with their response to the robot behaviors, which is qualitatively different (more simple and predictable than the one of a human agent). Another possible explanation for the fact the robot had a lower performance compared to BM or SHT, like some previous studies have shown, is the fact that in our task the robot provided limited social clues, meaning that the it did not have any extra interactions with the children besides the regular ones (the same in every session), did not use the name of the child, and it did not provide extra information regarding his story, etc. Previous research has shown that these social clues can improve the performance of the participants interacting with the robot.

What can been in this data is the fact that when the children started with lower levels of performance in BM (lower prior ability/functioning on this ability) and no ceiling effect (or a lower one) was present, most of the times RET and SHT seemed to be similar. For example, participant no 2, V.E., and participant no. 3, L.C., both expressed lower BM levels for imitation with objects and symbolic imitation, and in this cases SHT and RET were similar, and at least equivalent to BM. There are also few cases where this result was not present, however, in those situations that are most clinically relevant (BM data shows that the child requires improvements on a specific ability), both SHT and RET seem to be equivalent. Is true that the interventions do not always lead to improvements as compared to BM, but this might be due to the above mentioned reasons, or other alternative explanations need to be investigated. Yet, the most meaningful data coming from imitation experiments, suggest that when the child would benefit from some form of intervention, SHT and RET have similar effects. This results was also present for other abilities, and might indicated a possible predictor for RET outcomes.

Comparing BM and SHT, these conditions were almost identical for all outcomes, both primary and secondary, and the few differences do not indicate any pattern.

Looking at the effects of RET on secondary outcomes, it seems that this technologically enhanced intervention had a positive impact on verbal utterances and eye contact, and the effects were consistent for most of the children when comparing with both BM and SHT. Yet, no such effect was present for engagement (also a secondary outcome related to social interaction). One possible explanation for this lack of effect on the engagement variable is that in order to keep the experiments and task rigorous, we designed them in a highly structured manner, and this could have negatively affected the engagement of the child in interacting with the robot. In fact, some of the children learned the steps of the intervention and they knew every single behavior of the robot that was planned. The fact that they were able to predict all the behaviors of the robot could have also contributed to the decreased level of engagement.

An intriguing pattern that we observed, was related to behavioral outcomes. Some of the children expressed an increase in stereotype and maladaptive behaviors in interaction with the robotic agent, and significant differences emerged when comparing with both BM and SHT. Yet, the same children also expressed increases in adaptive behaviors (again, this increase in adaptive behaviors emerged in comparison with both SHT and RET). To what it concerns the stereotype and maladaptive behaviors, there are several possible explanations for their increase. The first one is that the robot was performing its behaviors slower than the therapist did (e.g. grasping objects) and this could have led to some frustration, which in turn was expressed as disruptive behaviors. Another possible explanation could be that some of the children had some oppositional behaviors and they started interacting with the robot in a way that was not appropriate (e.g. pulling the robot's hand). For the interaction with the therapist they had already learned that "it is inappropriate". Regarding the increases in adaptative behaviors in the RET sessions, this may be due to the novelty of the instrument that was introduced in their therapeutical sessions (i.e., the robot). Children are used to SHT, and they don't feel the need to share the experience with the mediator during the SHT session. On the other hand they seek comfort and want to share what they see with the mediator during the RET sessions, and thus more adaptive behaviors are promoted (see the description of the behaviors under this variable). The effect on behavioral outcomes could be regarded as a non-specific behavioral activation, in which both adaptive and maladaptive behaviors are increased. This could be regarded as a possible advantage of RET, giving to the therapists the possibility to reinforce adaptive behaviors, and thus increasing their frequency, while correcting maladaptive behaviors.

Finally, the majority of the children expressed more positive emotion when comparing RET to either SHT or BM. This suggests that indeed they enjoyed interacting with the robot partner, despite the fact that they did not initiate imitative behavior.

General message for imitation:

- For primary outcomes, both SHT and RET were not effective for developing imitation skills, in comparison with baseline level, while SHT had some advantages over RET. Most of the results seem to be affected by ceiling effect, and when such an effect is not present, RET and SHT are equivalent. Overall, results suggest that for high functioning ASD children, RET might not work for improving imitation as compared to SHT.
- For secondary outcomes, most of the time RET and SHT are equally effective, with some differences on:
    - Eye contact and verbal utterances, where RET is better than SHT;
    - Behavioral activation, where RET has larger effects than SHT. However, the activation is related to both adaptive and maladaptive behaviors and thus the clinicians should be aware and use this effect as an opportunity to reinforce the adaptive behaviors and decrease the maladaptive behaviors once they are activated.

General discussion and conclusions

Overall, the results of the experiments being presented in this deliverable show mixed results for the efficacy of the robot enhanced intervention (RET), especially for primary outcomes. These results are very important because they can help us to understand under what conditions robots can be implemented in the therapy of ASD, and where the human therapist should be the main actor.

We think that the general messages of the results that should be considered in the new phases of the projects is as follows:

For primary outcomes:

- Turn-taking: RET seems to be as good as or even better than SHT, especially for children with lower levels of prior skills.
- Joint Attention: RET seems to be as good as SHT.
- Imitation: RET seems less effective than SHT. However, SHT is not more effective than baseline level. For children with lower levels of prior skills, SHT and RET are similar and sometimes are better than BM. Overall, it suggests that for high functioning ASD children RET might not work for imitation, as compared to SHT.

Secondary outcomes:

- Most of the time RET seems to be as good as SHT and even better for:
    - Eye contact (in imitation task); this suggests that the ASD children in our study were interested in the robot mediator.

- Verbal utterance (in imitation task); this might be regarded as an additional proof for ASD children's interest and willingness to interact with robots.
- Positive emotions (in imitation and joint attention task).
  Some cautions should be given to the above mention positive results for RET, as they might be altered by a timing effect. As the robot had slower movements that the therapist, is possible that at least some of the effects on these variables are due to longer time periods in which the children could have expressed eye contact, verbal utterances and positive effect. However, the fact that a similar pattern (higher scores in RET as compared to SHT and BM) was not present for negative emotions, indicated that only positive social interactions were prompted by the robotic agent.
- There seems to be a stronger behavioral activation in RET as compared to SHT. However, activation is related to both adaptive and maladaptive behaviors (in imitation and turn-taking tasks and only adaptive behaviors in joint attention task) so the clinicians should be aware (and use it as an opportunity) to reinforce the adaptive behaviors and decrease the maladaptive behaviors once they are activated. Some possible explanations for the effects on stereotype and maladaptive behaviors are related to the fact that in comparison with the human mediator the robot required more time to give the instructions and respond to children's behavior, while technical issues generated sometimes even longer delays. These could have generated some frustration for the children and make them reacted trough stereotype and maladaptive behaviors. On the other hand, the robot could have attracted children's interest and made them want to share the experience with the mediator and/or with the therapist, which was expressed through adaptive behaviors.

Having said that, the readers should have in mind that much of the previous research was focused on children's reactions to robotic agents, and few studies have used a rigorous clinical research methodology (David et al., 2014). The studies described in this deliverable are some of the most extensive and rigorous clinical studies investigating the efficacy of RET in ASD children. However, more research is needed to understand which children could benefit the most and under which conditions. Moreover, future investigations should also focus on additional research questions, beside the one asking if RET is more effective or not than standard treatment. For example, one such question is related to whether the effects of RET appear faster than those of the therapist mediated intervention. Although by the present analysis we moved the field from case studies and exploratory tasks to clinical single case experiments, a new level is needed to answer such aspects, namely the level of randomized clinical trials. Also, improvements in the robot's capability to deliver the intervention are required in order to exclude variable that could interfere with the efficacy of the intervention.

As a final conclusion, we can say that RET is a promising approach that could have at least a similar efficacy (or be even better) as classical interventions for a large variety of outcomes in the case of children with ASD. Indeed, positive responses for RET are visible in most of the children included in our studies, but the efficacy of the RET vs. SHT needs to be seen nuanced, as presented above.

References

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Anzalone, S. M., Tilmont, E., Boucenna, S., Xavier, J., Jouen, A. L., Bodeau, N., ... & Michelangelo Study Group. (2014). How children with autism spectrum disorder behave and explore the 4-dimensional (spatial 3D+ time) environment during a joint attention induction task with a robot. *Research in Autism Spectrum Disorders*, *8*(7), 814-826.

Barlow, D. H., & Hayes, S. C. (1979). Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of Applied Behavior Analysis, 12*(2), 199-210.

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

Baron-Cohen, S. (2009). Autism: the empathizing–systemizing (E-S) theory. *Annals of the New York Academy of Sciences*, 1156(1), 68-80.

Baron-Cohen, S., Knickmeyer, R. C., & Belmonte, M. K. (2005). Sex differences in the brain: implications for explaining autism. *Science*, *310* (5749), 819-823.

Baron-Cohen, S., Lombardo, M., Tager-Flusberg, H., & Cohen, D. (Eds.). (2013). *Understanding Other Minds: Perspectives from developmental social neuroscience*. OUP Oxford.

Baxter, P., Wood, R., & Belpaeme, T. (2012). A touchscreen-based 'Sandtray'to facilitate, mediate and contextualise human-robot social interaction. In *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on* (pp. 105-106). IEEE.

Bons, D., van den Broek, E., Scheepers, F., Herpers, P., Rommelse, N., & Buitelaaar, J. K. (2013). Motor, emotional, and cognitive empathy in children and adolescents with autism spectrum disorder and conduct disorder. *Journal of Abnormal Child Psychology*, *41*(3), 425-443.

Bradway, K. (1999). Sandplay with children. *Journal of Sandplay Therapy*, *8*(2), 9-12.

Cabibihan, J. J., Javed, H., Ang Jr, M., & Aljunied, S. M. (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), 593-618.

Charman, T. (2003). Why is joint attention a pivotal skill in autism?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1430), 315-324.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences, 16*(4), 231-239.

Dautenhahn, K., & Billard, A. (2002). Games children with autism can play with Robota, a humanoid robotic doll. *In Universal Access and Assistive Technology* (pp. 179-190). Springer London.

David D., Anton R., Stefan S., Mogoase C., & Matu S. (2010). *Adaptation of the Autism Diagnostic Observation Schedule*. Romania, Bucharest: O.S. Publishing.

David, D., Costescu, C., Dobrean, A., Matu, S &, Szentagotai (2015) D2.1.1 Tasks for social robots on developing social skills (Wizard of Oz system) (*Intermediary version) DREAM Project: Development of Robot-enhanced Therapy for Children with Autism Spectrum Disorders*. (www.dream2020.eu).

Diehl, J. J., Schmitt, L. M., Villano, M., & Crowell, C. R. (2012). The clinical use of robots for individuals with autism spectrum disorders: A critical review. *Research in Autism Spectrum Disorders*, *6*(1), 249-262.

Dryden, W., & DiGiuseppe, R. (2003). *Ghid de terapie rational emotiva si comportamentala.* Cluj-Napoca, Romania: Editura ASCR.

Duquette, A., Michaud, F., & Mercier, H. (2008). Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. *Autonomous Robots*, *24*(2), 147-157.

Egger, H. L., Pine, D. S., Nelson, E., Leibenluft, E., Ernst, M., Towbin, K. E., & Angold, A. (2011). The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS): a new set of children's facial emotion stimuli. *International Journal of Methods in Psychiatric Research*, *20*(3), 145-156.

Field, T., Field, T., Sanders, C., & Nadel, J. (2001). Children with autism display more social behaviors after repeated imitation sessions. *Autism*, *5*(3), 317-323.

Goldstein, H., Kaczmarek, L. A., & English, K. M. (2002). *Promoting social communication: Children with developmental disabilities from birth to adolescence*. Brookes Publishing Company.

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 39*(5), 693-705.

Gottman, J., & Leiblum, S. (1974). *How to do psychotherapy and how to evaluate it*. Holt, Rinehart, and Winston: New York.

Gouaillier, D., Hugel, V., Blazevic, P., Kilner, C., Monceaux, J., Lafourcade, P., ... & Maisonnier, B. (2009). Mechatronic design of NAO humanoid. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on* (pp. 769-774). IEEE.

Huskens, B., Verschuur, R., Gillesen, J., Didden, R., & Barakova, E. (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), 345-356.

Ingersoll, B. (2008). The social role of imitation in autism: Implications for the treatment of imitation deficits. *Infants & Young Children*, *21*(2), 107-119.

Janosky, J. E., Leininger, S. L., Hoerger, M. P., & Libkuman, T. M. (2009). *Single subject designs in biomedicine*. Springer Science & Business Media: London.

Jones, W. P. (2003). Single-case time series with Bayesian analysis: A practitioner's guide. *Measurement and evaluation in counseling and development, 36*(1), 28-39.

Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R., & Scassellati, B. (2013). Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism and Developmental Disorders, 43*(5), 1038-1049.

Kozima, H., Nakagawa, C., & Yasuda, Y. (2007). Children–robot interaction: a pilot study in autism therapy. *Progress in Brain Research, 164*, 385-400.

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Instruments, & Computers, 41*(3), 841–849.

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., & Schopler, E. (1989). Austism diagnostic observation schedule: A standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders*, *19*(2), 185-212.

Lovaas, O. I., Freitas, L., Nelson, K., & Whalen, C. (1967). The establishment of imitation and its use for the development of complex behavior in schizophrenic children. *Behaviour Research and Therapy, 5*(3), 171-181.

Loveland K. & Landry, S. (2007). Joint attention and language in autism and developmental language delay, *Journal of Autism and Developmental Disorders, 16*(3), 335–349.

Murray, D. S., Creaghead, N. A., Manning-Courtney, P., Shear, P. K., Bean, J., & Prendeville, J. A. (2008). The relationship between joint attention and language in children with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities*, *23*(1), 5-14.

Nugent, W. (2009). *Analyzing single system design data*. Oxford University Press: New York.

Pioggia, G., Sica, M. L., Ferro, M., Igliozzi, R., Muratori, F., Ahluwalia, A., & Rossi, D. D. (2007, August). Human-robot interaction in autism: FACE, an android-based social therapy. In *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium* on (pp. 605-612). IEEE.

Pop, C. A., Pintea, S., Vanderborght, B., & David, D. O. (2014). Enhancing play skills, engagement and social skills in a play task in ASD children by using robot-based interventions. A pilot study. *Interaction Studies, 15*(2), 292-320.

Quirmbach, L. M., Lincoln, A. J., Feinberg-Gizzo, M. J., Ingersoll, B. R., & Andrews, S. M. (2009). Social stories: mechanisms of effectiveness in increasing game play skills in children diagnosed with autism spectrum disorder using a pretest posttest repeated measures randomized control group design. *Journal of Autism and Developmental Disorders, 39*(2), 299-321.

Riley-Tillman, T. & Burns, M. (2009) *Evaluating educational interventions: Single-case design for measuring response to intervention*. The Guilford Press.

Robins, B., Dautenhahn, K., & Dickerson, P. (2009). From isolation to communication: a case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot. In *Advances in Computer-Human Interactions, 2009. ACHI'09. Second International Conferences* on (pp. 205-211). IEEE.

Robins, B., Dautenhahn, K., Te Boekhorst, R., & Billard, A. (2004). Effects of repeated exposure to a humanoid robot on children with autism. In *Designing a More Inclusive World* (pp. 225-236). Springer London.

Scassellati, B., Admoni, H., & Mataric, M. (2012). Robots for use in autism research. *Annual Review of Biomedical Engineering, 14*, 275-294.

Stanton-Chapman, T. L., & Snell, M. E. (2011). Promoting turn-taking skills in preschool children with disabilities: The effects of a peer-based social communication intervention. *Early Childhood Research Quarterly*, *26*(3), 303-319.

Tapus, A., Peca, A., Aly, A., Pop, C., Jisa, L., Pintea, S., ... & David, D. O. (2012). Children with autism social engagement in interaction with Nao, an imitative robot–A series of single case experiments. *Interaction studies*, *13*(3), 315-347.

Tellegen, P. J., Winkel, M., Wijnberg-Williams, B. J., & Laros, J. A. (1998). *Snijders-Oomen Niet-verbale Intelligentietest, SON-R 21/2-7: Verantwoording en handleiding* [Snijders-Oomen Nonverbal Intelligence Test, SON-R 21/2-7: Manual].

Vanderborght, B., Simut, R., Saldien, J., Pop, C., Rusu, A. S., Pintea, S., ... & David, D. O. (2012). Using the social robot probo as a social story telling agent for children with ASD. *Interaction Studies, 13*(3), 348-372.

Wainer, J., Dautenhahn, K., Robins, B., & Amirabdollahian, F. (2010). Collaborating with Kaspar: Using an autonomous humanoid robot to foster cooperative dyadic play among children with autism. In *Humanoid Robots (Humanoids), 2010 10th IEEE-RAS International Conference* on (pp. 631-638). IEEE.

Warren, Z. E., Zheng, Z., Swanson, A. R., Bekele, E., Zhang, L., Crittendon, J. A., ... & Sarkar, N. (2013). Can Robotic Interaction Improve Joint Attention Skills?. *Journal of Autism and Developmental Disorders*, 1-9.

Whalen, C., Schreibman, L., & Ingersoll, B. (2006). The collateral effects of joint attention training on social initiations, positive affect, imitation, and spontaneous speech for young children with autism. *Journal of Autism and Developmental Disorders*, *36*(5), 655-664.

Williams, J. H., Whiten, A., & Singh, T. (2004). A systematic review of action imitation in autistic spectrum disorder. *Journal of Autism and Developmental Disorders*, *34*(3), 285-299.

Zwickel, J., White, S. J., Coniston, D., Senju, A., & Frith, U. (2011). Exploring the building blocks of social cognition: spontaneous agency perception and visual perspective taking in autism. *Social cognitive and affective neuroscience, 6*(5), 564-571.