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

Implementation of ethical constraints in the self-monitoring subsystem

Due Date: **31/03/2019**
Submission date: **31/03/2019**

Start date of project: **01/03/2014** Duration: **59 months**

Organisation name of lead contractor for this deliverable: De Montfort University

Responsible Persons: Kathleen Richardson and Mark Coeckelbergh

Revision: 3

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

I Executive Summary

Deliverable D7.3: Implementation of ethical constraints in the cognitive controller of the robot concerns the coding of ethics into robotic systems that would resemble the therapists ethics in similar circumstances. Psychotherapeutic training is a highly skilled practice taking many years of education and practice. If robots are to take on therapeutic roles then they must be able to exercise similar high standards. If robots are to be put to use in close proximity to children with 'supervised autonomy' then there should be robust safety measures implemented in machines so that they 'do no harm' to the child. At the end of the DREAM project, the cognitive controller remains more of a theoretical proposition rather than a fully functioning algorithm.

II Principal Contributor

Kathleen Richardson, De Montfort University

III Revision History

Version 1.0 (KR 01-02-2019)
Version 2.0 (KR 18-03-2019)
Version 3.0 (KR 29-09-2019)

IV Coding Ethics in DREAM's Robots

Deliverable D7.3: Implementation ethical constraints in the self-monitoring subsystem in a robot concerns the coding of ethics into robotic systems that engage directly with children with autism. This is the final ethics deliverable of the DREAM project. The ethical constraints are put into the robot in the form of a cognitive controller is envisioned as thus: the cognitive controller would have inputs from task T5.1, (child behaviours) and Work Package 4 (sensing and interpretation). The exact behaviours of interest were defined in task T1.3 and the challenge was to identify a mapping from the percepts provided by WP4 to these behaviours. These together would form the basis for assessing and quantifying the children's degree of motivation/engagement as well as their performance (task T5.2). Technically, both values will be presented on a continuous numerical scale with the exact meaning of the values assigned by the therapists. And from this to use the outputs to create an algorithms for the cognitive controller for robot behaviour selection (WP6).

Cognitive controllers are not a new ideas in robotics and have been developed in other platforms such as Willow Garage, OROCOS, Robotics Studio and Microsoft). Most of these CCs are specific to a particular platform and describe an interface to sensors and manipulators on mobile platforms. DREAM wanted to move beyond this, and in turn move on the capacity of the CC to act as more than an interface and to make it an interface and an agent of action.

If the purpose of the DREAM project is a to develop a robot that is sophisticated enough to identify and respond to the psychological dispositions of children with autism, read from facial expressions and bodily behaviour, there was equally the desire to marry this with something that would be able to control the actions of the system and so a cognitive controller was imagined to play this role. In traditional robot-autism therapy set-ups there is a Wizard of Oz (WoZ) style controller of the robot (Martelaro 2016). The Wizard of Oz is an apt name because it refers to the wizard behind the curtain, controlling events and making it seem as though something is being done by itself when it is not. In the 1939 film of the same name, the wizard of Oz turns out to be an ordinary man.

There are a number of ethical concerns about WoZ as it can lead vulnerable children to believe a robot is able to act or speak when it is not. It is not known if children with autism believe robots to be 'real', or in fact they treat them as sophisticated toys like typically developing children of their peer age groups. The evidence is unclear, particularly as the process of asking questions is limited in experimental settings where the robots are used. Children with autism are discouraged from poking the robot as it is seen as 'interfering' with the experiment. If there is no debrief or full disclosure to the child on cessation of the experimental process then the child could be left unclear about a robot's capability, particularly as WoZ is deliberately relying on illusion to work effectively. If the robot was increasingly playing a supervised autonomous role, and less WoZ, then it was necessary to imagine a new layer of safety in the form of a cognitive controller that would triggers the 'next steps' of actions in the robot based on input from the sensors about the child's disposition. While

DREAM envisaged that the cognitive controller will be platform independent and social interaction will be specified in a platform agnostic manner. And rather than controlling actuators and modules specific to a robot platform, the cognitive controller will set parameters in descriptions and representations that are common across all platforms and these robot non-specific commands will then be translated into robot-specific actions.

The aim to 'algorithms' to infer the child with autism psychological disposition is a difficult task (difficult for a human being also). The machine learning algorithm calculated what it considered to be the feeling of the child based on facial expression, happy and sad are two obvious examples but there could also be tired, distressed, excited, joyous, irritated, moody, unresponsive, lacklustre etc. There is obvious complexity in trying to design algorithms to accurately detect complex human expression that has taken hundreds of thousands (if not millions) of years to develop to its current level of complexity. A machine could try to frame a behaviour as one thing or another by mapping facial features and attributing a tag to it. Learning data could be feed into the algorithm before the experiments so that the program has some sense of happy or sad. Even this simple calculation is an ethical issue. What happens if the machine is wrong and triggers next step responses based on its calculation? Also the speed in which interpersonal social interaction occurs is significant, running in 'real-time', and social expression is subtle. People do not learn facial expressions abstractly but through billions and trillions of social interactions and a have a body capable of receiving and interpreting sensory information from others.

DREAM offered this rigorous testing and accountability of benefits of robots, while using a particular therapeutic method – ABA. ABA is an heir of behavioural psychology and based on the notion that behaviour and thoughts coexist together and that learning is predominately by rote and repetition (Skinner 1988). This is why many of the ABA practices involve identifying the 'missing' behaviour and targeting it with a series of therapeutic inputs from the therapist. In the DREAM project we identified turn-taking, joint-attention and imitation as the three target behaviours. ABA is often criticised by many as time-consuming and expensive with children often engaged in hours of therapy per week. The benefits of ABA therapy are known to be statistically beneficial, but parents and recipients of the treatment have been less enthusiastic about its benefits. Parents have been known to try several therapeutic treatments for their children, and some have developed new ones. Therapies with movement and horses are two that have been developed by parents frustrated with existing medically inspired therapeutic treatments.

DREAM aimed to develop advanced algorithms that facilitate the interpretation of sensory data in order to infer the child's psychological disposition and assess his or her behaviour, thereby allowing the robot to act autonomously and fulfil the clinical requirements through its own appropriately-conducted social behaviour. We refer to this next-generation approach as robot-enhanced therapy (RET) to distinguish it from current approaches to RAT. The results produced by both the signal processing units and the algorithms will also form the basis for a diagnostic tool to collect quantitative data on the patient. Since autonomous interaction is the goal, there is a need to answer new ethical questions regarding the use of robots with vulnerable children, as well as a need to ensure ethically-compliant robot behaviour. This is essential to gain the trust and acceptance of the clinical end-users.

The DREAM team ambitiously planned to develop a cognitive controller capable of being guided by 'ethical behaviour' but this was not achieved during the project. Instead, a cognitive controller that can constrain a self-monitoring subsystem remains a theoretical proposition. This deliverable then sets out some problems in designing a cognitive controller for a socially interactive and therapeutic robot. Social interaction is vastly complex and has taken hundreds of thousands of years to refine so it seems unlikely that such behaviour that are the result of complex evolutionary development are possible, or desirable to mimic in robots. Unlike automation that occurs in other areas of robotic development, when the action can be performed in another way producing the same outcome for example a washing machines does not wash like a human but it meets the end goal (Floridi 2010 p. 13). By contrast social interaction is not exclusively goal orientated (Bowlby 1978, Dykas and Cassidy 2013), and if the social basis of the interaction is not well understood, it could produce unintentional, even confusing effects for a child (Richardson 2018). While social interaction between people is learned through thousands, if not millions of micro engagements made up of speech, touch, sight and sound. The benefits of social interaction allow humans to get their needs, wants and desires met from self and others. In the absence of an ability to communicate one's needs, wants and desires to self and others there are potentially life threatening consequences (Dykas and Cassidy 2013; Kanner 1943; Runswick-Cole, Mallett and Timimi 2016).

V Complexities of Social Interaction and coding therapeutic expertise

Therapeutic practices require a body of knowledge and requisite skills but are developed because social interaction forms the basis of all human lived experience (Vygotsky 1986, Rogers 2003). Often therapists take years to train and become certified and able to practice a therapeutic technique such as ABA or person centred psychotherapy (Rogers 2003). Therefore a cognitive controller must be both able to recognise and respond appropriately to social interaction and emotional responses and simultaneously fine tune these responses through a therapeutic frame. The therapeutic frame of the DREAM project was ABA therapy. While behaviour maybe observable in body language and facial expression, there are also unspoken and implicit modes of communication that are subtle and not easily observable and further complicated by a complex cultural environment in which norms about age, class, sex and race maybe prescient. While in European societies communication rituals that devalue people on the basis of their sex, race or class may be frowned upon, there is still evidence they occur. Moreover, children traditionally have less political or economic power than adults, and are undergoing a developmental process into adulthood (Carey 1985, Bowlby 1978). There are some aspects of a child's existence that need careful consideration and engagement when therapists engage with them and sometimes the models applied to childhood experience can be up for question (Bowlby 1978, Vygotsky 1986). Take for example the Jean Piaget's claim of a spontaneous *egocentrism* in children (a child could only see something from his or her own perspective), was later challenged by experts in the field who believed Piaget overestimated the child's egocentric position instead showing children to be more than aware of other people (Kesselring and Müller 2010, Piaget 2001, Matusov and Hayes 2000; Richardson 2018) Likewise, experts see autism in different ways, some as a biological condition with difference in the brain and connected to the male sex (Auyeung et al. 2009, Baron-Cohen 1997, Baron-Cohen 2003) while others (Runswick-Cole, Mallett and Timimi 2016, Kanner 1943) view autism through a social frame, exploring how the biological condition is interpreted and shaped by culture.

ABA therapy was the one that researchers in DREAM project tried to reproduce in robots. ABA is rooted in the behaviourist theories of B.F Skinner that believes behaviour is shaped by reinforcement (good and bad) learning practices, and learning is constant repetition (Skinner 1988). ABA therapists will spend weeks, even years using repetition and reinforcement with children with autism to learn imitation, turn taking and joint attention. Three keys skills that are essential for social communication and learning and often impaired in children with autism. ABA therapy rests on a theory of learning and requires years of professional training. There was no intention to programme the theory of ABA into the machine only to get the robot to carry out actions that were similar to therapeutic behavioural methods. As already discussed in previous deliverables, social communication is complex and more than a sequence of signals, even if those signals are performed at the right moment. A cognitive controller would need to understand when each social action was optimal to perform, to respond to the child's responsiveness to the action, and then to engage in the next appropriate action, and so on throughout the therapeutic engagement. This was not achieved during the DREAM project and to date, a cognitive controller capable of carrying out complex, responsive and reciprocal actions with the children remains a theoretical problem.

If robots are to be put to use in close proximity to children with 'supervised autonomy' then there should be robust safety measures implemented in machines so that they 'do no harm' to the child. While much is known is about autism spectrum disorders (ASDs), so much is unknown. ASD is a complex developmental disorder impacting on a child's ability to get their attachment needs met through typical social interaction that is compartmentalised into several types of social behaviours including, turn-taking, joint attention, eye-contact, imitation, social touch and language development. Developmental difficulties in all these areas impact greatly on the child's ability to process complex social stimuli and respond to or instruct their parents and caregivers of their needs (APA 2013, Baron-Cohen 2001).

VI The Science Fiction of Ethical Robots

There is no shortage of discussion about 'ethics' paired with robots and or AI in contemporary culture (Anderson and Anderson 2011, ESPRC 2010, Winfield, Blum and Liu 2014, Wallach and Allen 2010). The European commission and many European nations have signposted themselves as most politically responsible for creating ethical and transparent systems. While this is notable, it does somehow rest on an assumption that ethics is something that can be transferred to technological artefacts, or in fact, a science fiction idea that has become bound up with contemporary robotics. Ethics has a branch of philosophy is one of the oldest traditions and arguably predates philosophy, when any rules were developed for supporting one form of behaviour over another, one could argue that ethics is integral to human cultures.

Moreover, some ethical systems have not translated well into the 21st century political life where there is a emphasis on equality, universal humanism and women's rights. The philosophy of Aristotle (2000) promoted ethical norms that singled out women, slaves and the lower classes as inferior to property owning citizens of ancient Greece. The term 'citizen' only referred to less than 6% of the population in ancient Greece now refers to all Europeans nationals including those humans that were once regarded as similar to other forms of property including land, animals or money. Moreover, there are ethical models based on religious beliefs or political values, as there are ethical models that are against all forms of war and militarism and those that are for war and militarism based on an assembling of ideas and coming up with different conclusions. While there is this diversity in ethical models, there is one thing that connects them together – they are all created by humans for living in human societies. The recipients of these ethical models are exclusively human. This is not to say that animals or other living beings do not have something approximate to ethics, but they have not openly shared any with humans, nor have we engaged in teaching Aristotle to trees or rabbits.

In this sense, 'ethical' robots is more likely to be a dead end for our actions because you cannot teach ethics to a system that is not conscious where learning becomes reducible to algebra and statistics. Only a sentient and conscious intelligence can absorb or make sense of ethical principles in a way that it makes them capable of acting in a different way (or continuing to engage in ethical behaviour). The automation of social interaction (Breazeal 2002), a new *mechanical sociality* (Richardson 2018) is put on a par with human social interaction when it should be classified in alternative ways because machines are not conscious. While there are numerous accounts of making ethical algorithms or robots it is still a theoretical issue that is a long way to becoming resolved (Wallach and Allen 2010, Anderson and Anderson 2011, Lin, Abney and Bekey 2011, Malle 2015). In trying to simulate ethics in algorithms and robots it might be forgotten that ethics is tied to consciousness and so that ethics and can only be practiced by forms of life that possess consciousness. It might be worth referring back to Josef Weizenbaum who created one of the first therapeutic machines Eliza using the Rogerian person-centred approach which meant turning the end of a statement back into a question for the client. This method produced an endless source of interest for some of Eliza's early using, an issue that concerned Weizenbaum (1976). He began to despair at a future where machines increasingly took over human functions and believed we were increasingly moving in a direction where calculation, rather than reason and judgement would become the order of the day.

VII Recommendations

- To reassess the rationale for an ethical cognitive controller
- To ensure that children with autism are not *othered* for the sake of producing theories in psychiatry or robotics.
- To involve other paradigms and ways of understanding autism when developing robot project proposals and research.
- To engage with other beliefs that are different from your own so that issues are discussed and explored from many angles.
- All robotic projects should involve adults with autism and parents of children with autism on their advisory boards to monitor project progress and give critical feedback to researchers.

VIII References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Anderson, M. and Anderson, S.L. eds., 2011. 'Introduction' In *Machine Ethics*. Cambridge University Press.

Aristotle, 2000. *The Politics*. London: Penguin.

Auyeung, B., Baron-Cohen, S., Ashwin, E., Knickmeyer, R., Taylor, K. and Hackett, G., 2009. Fetal testosterone and autistic traits. *British Journal of Psychology,* 100(1), pp.1-22.

Baron-Cohen, S., 2003, T*he essential difference: men, women, and the extreme male brain,* Basic Books, NY.

Bowlby, J., 1978, *Attachment and loss*. John Bowlby. Vol. 1, Attachment. Harmondsworth: Penguin, 1978.

Breazeal, C., 2002, *Designing sociable robots*. Cambridge, Mass: MIT Press.

Carey, S., 1985. *Conceptual change in childhood*. Cambridge, Mass: MIT Press.

DREAM, DOW, Development of Robot-Enhanced Therapy for Children with AutisM spectrum disorders. *Grant agreement* no: 611391.

DREAM, D2.1.1, 'Tasks for social robots on developing social skills (Wizard of Oz system) Development of Robot-Enhanced Therapy for Children with AutisM spectrum disorders. *Grant agreement* no: 611391. <Accessed 1.2.16 http://www.dream2020.eu/wp-content/uploads/2015/04/DREAM_Deliverable_2.1.1.pdf>

DREAM, D2.1.2, 'Tasks for social robots on developing social skills (Wizard of Oz system) Development of Robot-Enhanced Therapy for Children with AutisM spectrum disorders. *Grant agreement* no: 611391. <Accessed 1.2.16 http://www.dream2020.eu/wp-content/uploads/2015/11/DREAM_Deliverable_2.1.2.pdf>

D7.1, 'Robot Ethics Manual''Tasks for social robots on developing social skills (Wizard of Oz system) Development of Robot-Enhanced Therapy for Children with AutisM spectrum disorders. Grant agreement no: 611391. <Accessed 1.2.16 http://www.dream2020.eu/wp-content/uploads/2015/04/DREAM_Deliverable_D7.11.pdf>

Dykas, MJ & Cassidy, J., 2013, 'The first bonding experience: the basics of infant caregiver attachment' in Human bonding: the science of affectional ties, eds. C Hazam & Campa, The Guildford Press, New York.

ESPRC Robotics Guidelines, 2010, Principles of Robotics [Online] <Accessed 10.9.2015 https://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/ >

Floridi, L 2010. Ethics after the Information Revolution in Floridi (ed), The Cambridge Handbook of Information and Computer Ethics. Cambridge: Cambridge University Press.

Kanner, L., 1943, Autistic disturbances of affective contact. *Nervous Child, 2*, 217-250.

Kesselring, T and U Müller. 2011. The concept of egocentrism in the context of Piaget's theory. *New Ideas In Psychology*, 29 (2011), 327-345.

Lin, P., 2011, *Robot ethics: the ethical and social implications of robotics*. MIT press.

Lin, P., Abney, K. and Bekey, G., 2011, Robot ethics: Mapping the issues for a mechanized world. *Artificial Intelligence*, *175*(5), pp.942-949.

Lin, P., Abney, K. and Bekey, G.A., 2011, *Robot ethics: the ethical and social implications of robotics*. MIT press.

Malle, B.F., 2015. Integrating robot ethics and machine morality: the study and design of moral competence in robots. *Ethics and Information Technology*, pp.1-14.

Martelaro, N., 2016, March. Wizard-of-Oz Interfaces as a Step Towards Autonomous HRI. In 2016 *AAAI Spring Symposium Series*.

Matusov, E and R Hayes. 2000. Sociocultural critique of Piaget and Vygotsky. *New Ideas in Psychology*, 18 (2000) 215-239.

Meadows. T. 2017. ABA Practitioner Training Manual. Amazon Media.

Piaget. J. 2001. *The Language and Thought of the Child*. London and New York: Routledge.

Roger. C. 2003. *Client Centred Therapy: Its Current Practice, Implications and Theory*. London: Robinson Press.

Runswick-Cole, K, Mallett R and S Timimi. 2016. Introduction in *Re-thinking autism: diagnosis, identity and equality*. London: Jessica Kingsley Publishers.

Skinner, B.F. 1988. *About Behaviorism*. Random House, USA.

Vygotsky, L., 1986, *Thought and Language*. Translation newly revised and edited by Alex Kozulin. Cambridge, Mass: MIT Press.

Wallach, W. and Allen, C., 2010. Moral machines: Teaching robots right from wrong. Oxford University Press.

Weizenbaum, J., 1976, *Computer power and human reason: From judgment to calculation*. Cambride, Mass: MIT Press.

Winfield, A.F., Blum, C. and Liu, W., 2014, Towards an ethical robot: internal models, consequences and ethical action selection. In *Advances in Autonomous Robotics Systems* (pp. 85-96). Springer International Publishing.