Learning from a robot: creating synthetic psychologically plausible agents

Vasiliki Vouloutsi

TESI DOCTORAL UPF / 2017

Director de la tesi

Prof. Dr. Paul F. M. J. Verschure,
Department of Information and Communication Technologies

Universitat
Pompeu Fabra
Barcelona

Man is by nature a social animal; an individual who is unsocial naturally and not accidentally is either beneath our notice or more than human. Society is something that precedes the individual. Anyone who either cannot lead the common life or is so self-sufficient as not to need to, and therefore does not partake of society, is either a beast or a god.

- Aristotle, Politics

Acknowledgements

I have so many people I would like to thank for being so awesome that I am afraid my acknowledgements will have the same length as my thesis! First of all I would like to thank my supervisor, Professor Paul Verschure for his patience, for believing in me when I was doubting myself the most and for pushing me to my limits, but always guiding and trusting me, it is a great honour. I will try to not put the rest of the names by order of preference, because it is impossible to discriminate. To Dr. Anna Mura, for her guidance and all the wonderful events I have participated that were organised by her. I would like to thank all the people at SPECS, those who are still here and those who are not. Mireia, Christian and Carme have been so nice and sweet, they always made me feel like home and were always there to provide support in anything we asked. Maria, my wise partner in crime that has supported me and "suffered" with me in many academic adventures. Stephane who taught me so many things about robots. Jordi for all the amazing conversations. Alex, the most ingenious man I know, always there to fix the robots and make even crazier ones, Sock for the dark humour jokes and gastronomical adventures. Diogo, for always worrying about me but at the same time, being true and honest. Riccardo, I am most grateful for your honesty and thank you for all the support. Belen, for being a wise doctor imbuing her wisdom to me. The list can go on and on and on. Thank you Xerxes and Martina for the hugs, I will always need

them and will always be there to give them, Giovanni for all the amazing moments, Klaudia for all the wonderful times and for always finding many participants, Quique, Pedro, Ivan, Clement, Daniel, Marti for all your love and support. I would never of course forget my three musketeers: Sytse, Andre and Martin. I would also like to thank all the partners in the EU projects I have participated throughout the years: all of you have taught me so much that you cannot understand. Thank you for the wonderful (yet sometimes stressful) moments, I enjoyed this part of my PhD a lot! I should not forget to thank Lydia Garcia, an amazing and warm person, always being helpful and always greeting me with a warm smile. Her help was really valuable to me.

I would also like to thank my flatmates and friends for their insane support and wonderful moments: Yannis, Panos, Elena, Elisa, Dimitris, Iasonas, Giulia, Lefteris and Ina. You guys are the greatest and you have made my staying in Barcelona a wonderful experience. Also, special thanks to Elisa and Panos for their valuable help and feedback, you guys made me feel so much better and helped me so so much, you have no idea! To all of my friends here in Barcelona (that are more like an extended family): thank you! Joan, Pascal, Nuria, Lauren, Alberto, Sebas, Naz, all these years were and are unforgettable because of you. To my family (mom, dad, Gianna, Gary, Zouzouni and Clio) and friends (Kostas, Fotis, Maria, Giorgia, Niovi, Marina, Kalipso, Antzela) in Greece, a big big thank you. It seems that distance was never an issue and I've always felt you were with me the whole time. You guys have encouraged me more than anyone, and I sincerely love you for that! Thank you Gianna for feeding me and taking care of me this summer, when I was frantically working on my thesis. For pushing me to finish and for morally supporting me. To Gary, that although appears to be tough and not involved, always worried about me and asked me how I was. To mom and Clio, for coming to visit me in Barcelona, and to Mrs Eleni, for all the wonderful comfort food she has prepared for us. Finally, to my dad, the person that I resemble the most (both physically and the more time passes, mentally): he has pushed me to become the best version

of me. I am still working on it, but I am getting there!

Thank you all for psychologically supporting me when I was in my lowest and cheering with me when I was in my highest. Thank you for worrying for me, for trusting me when I was in doubt and for helping me be who I am. Thank you for being the awesome you that I love and had the privilege to meet, for smiling at me, for hugging me, for being by my side. I owe it to you all. Thank you :)

Abstract

Due to technological advancements, robots will soon become part of our daily lives and interact with us on a frequent basis. Robot acceptance is important, as it delineates whether users will potentially interact with them or not. We argue that psychological plausibility is a key determinant of acceptance and the challenge that rises is to understand, measure and identify what affects plausibility. Here, we propose a taxonomy of four psychological benchmarks that one can apply to evaluate the behavioural components of robots and assess how they affect acceptance: social competence, task competence, autonomy and morphology. By decomposing plausibility to discrete parts and empirically test them, we can use their interactions in practice for the meaningful design and development of social robots. In this thesis, we have identified behavioural components that are relevant to the proposed taxonomy and evaluated them in a series of studies. We show that it is possible to use the proposed taxonomy to evaluate the interaction and the robot. By systematically assessing the behavioural features of the robot, we gain useful insights that we apply to our H5WRobot that we later validate in the domain of tutoring. We show that our robot is accepted by students and stress that our proposed taxonomy might provide useful insights regarding the establishment of future assessments for HRI.

Resum

A causa dels avenços tecnològics, els robots aviat formaran part de la nostra vida diària i interactuaran amb nosaltres de forma freqüent. Que els robots siguin ben rebuts és important, ja que determina si els usuaris voldran interactuar amb ells o no. Argumentem que la plausibilitat psicològica dels robots és fonamental per a la seva acceptació i que un repte que sorgeix és entendre, mesurar i identificar què afecta aquesta plausibilitat. Proposem una taxonomia de quatre criteris psicològics que es poden aplicar per tal d'avaluar els components de conducta dels robots i com afecten la seva acceptació: competència social, competència funcional, autonomia i morfologia. Descomposant la plausibilitat en parts discretes, i avaluant-les de forma empírica, podem fer-ne un ús pràctic de les interaccions per al disseny i desenvolupament de robots socials. En aquesta tesi hem identificat comportaments conductuals que són rellevants per a la taxonomia proposada i que han estat avaluats en una sèrie d'estudis. Mostrem que és possible utilitzar la taxonomia proposada per tal d'avaluar un robot i la interacció amb aquest. Mitjançant una avaluació sistemàtica de les característiques conductuals dels robots, obtenim una sèrie d'idees útils que hem aplicat al nostre robot H5WRobot, i que posteriorment validem en un context de tutoria. Demostrem que el nostre robot és acceptat pels estudiants i fem palès que la taxonomia que proposem pot proporcionar observacions útils per a l'establiment de futures avaluacions per a la interacció entre humans i robots.

List of contributions

Contributions are listed in chronological order. Where applicable, the chapter of appearance is specified.

Contributions included in the thesis

Peer Reviewed

Vouloutsi, V., Lallée, S., & Verschure, P. F.M.J. (2013). Modulating behaviours using allostatic control. In *Conference on Biomimetic and Biohybrid Systems,* pp. 287–298. Springer, Berlin, Heidelberg. (Chapter 6.1)

Vouloutsi, V., Grechuta, K., Lallée, S., & Verschure, F.M.J. (2014). The influence of behavioural complexity on robot perception. *Conference on Biomimetic and Biohybrid Systems,* pp. 332–343. Springer, Berlin, Heidelberg. (Chapter 6.3)

Vouloutsi, V., Blancas, M., Grechuta, K., Lallee, S., Duff, A., Llobet, J. Y. P., & Verschure, P. F.M.J. (2015). A new biomimetic approach towards educational robotics: the distributed adaptive control of a synthetic tutor assistant. *New Frontiers in Human-Robot Interaction*, pp. 22–30. (Chapter 7.3)

Lalle, S., Vouloutsi, V., Munoz, M. B., Grechuta, K., Llobet, P., Sarda, M.,

& Verschure, P. F.M.J. (2015). Towards the synthetic self: Making others perceive me as an other. *Paladyn, Journal of Behavioural Robotics*, 6(1), pp. 136–164. (Chapter 7.3, 6.4)

Vouloutsi, V., Blancas, M., Zucca, R., Omedas, P., Reidsma, D., Davison, D., ... & Cameron, D. (2016). Towards a synthetic tutor assistant: the EASEL project and its architecture. In *Conference on Biomimetic and Biohybrid Systems,* pp. 353–364. Springer, Berlin, Heidelberg. (Chapter A.2)

Under Review

Vouloutsi, V. Emotions and Drives (2018) in: Prescott, T. J., Lepora, N. F., and Verschure, P. F.M.J. *Living Machines: A Handbook of Research in Biomimetic and Biohybrid Systems.* Oxford: OUP. (Chapter 3.2.2)

Other contributions

Peer Reviewed

Petit, M., Lallée, S., Boucher, J. D., Pointeau, G., Cheminade, P., Ognibene, D., ... & Barron-Gonzalez, H. (2013). The coordinating role of language in real-time multimodal learning of cooperative tasks. *IEEE Transactions on Autonomous Mental Development*, 5(1), pp. 3–17.

Lallée, S., Vouloutsi, V., Wierenga, S., Pattacini, U., & Verschure, P. F.M.J. (2014). EFAA: a companion emerges from integrating a layered cognitive architecture. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* pp. 105–105. ACM, New York, USA.

Bampatzia, S., Vouloutsi, V., Grechuta, K., Lallée, S., & Verschure, P. F.M.J. (2014). Effects of gaze synchronisation in human-robot interaction. In *Conference on Biomimetic and Biohybrid Systems,* pp. 370–373. Springer, Berlin, Heidelberg.

Gou, M. S., Vouloutsi, V., Grechuta, K., Lallée, S., & Verschure, P. F.M.J. (2014). Empathy in humanoid robots. In *Conference on Biomimetic and Biohybrid Systems,* pp. 423–426. Springer, Berlin, Heidelberg.

Blancas, M., Vouloutsi, V., Grechuta, K., & Verschure, P. F.M.J. (2015). Effects of the robot's role on human-robot interaction in an educational scenario. In *Conference on Biomimetic and Biohybrid Systems,* pp. 391–402. Springer, Berlin, Heidelberg.

Puigbo, J. Y., Moulin-Frier, C., Vouloutsi, V., Sanchez-Fibla, M., Herreros, I., & Verschure, P. F.M.J. (2015). Skill refinement through cerebellar learning and human haptic feedback: An iCub learning to paint experiment. In *Humanoid Robots (Humanoids), 2015 IEEE-RAS 15th International Conference on* pp. 447–452. IEEE.

Blancas, M., Zucca, R., Vouloutsi, V., & Verschure, P. F.M.J. (2016). Modulating Learning Through Expectation in a Simulated Robotic Setup. In *Conference on Biomimetic and Biohybrid Systems,* pp. 400-408. Springer, Berlin, Heidelberg.

Reidsma, D., Charisi, V., Davison, D., Wijnen, F., van der Meij, J., Evers, V., ... & Mazzei, D. (2016). The EASEL Project: Towards Educational Human-Robot Symbiotic Interaction. In *Conference on Biomimetic and Biohybrid Systems,* pp. 297–306. Springer, Berlin, Heidelberg.

Vouloutsi, V., Blancas, M., Zucca, R., & Verschure, P. F. (2017) Studying the adaptation of robot's strategies in an educational task. In *3rd Child-Robot Interaction Workshop, HRI,* pp. 1–4. Vienna, Austria.

In Preparation / Under Review

Blancas, M., Vouloutsi, V., Fernando, S., Sánchez-Fibla, M., Zucca, R.,, Prescott, T. J., & Verschure, P. F.M.J., I want a robot to do my homework!" Analysing children's expectations from robotic companions in educational settings *Submitted*

Blancas, M., Vouloutsi, V., Zucca, R., & Verschure, P. F.M.J., Students' metacognitive ability as an input for a science-related Intelligent Tutoring System *Manuscript in preparation*

Vouloutsi, V., Blancas, M. , Zucca, R., & Verschure, P. F.M.J., The effects of distraction on a tutoring scenario *Manuscript in preparation*

Contents

List of Figures

List of Tables

Chapter 1

Chapter Overview and Introduction

As robots gain social character, it is certain that they will be part of our daily lives and will be required to interact with humans on a frequent basis. Thus, given the robots' primal goal to socially interact with humans, a fundamental question arises: "How can we create robots that are successful in interaction and are accepted by people?" And more importantly: how do we measure success? An answer to this question may lie in the psychological plausibility of such robots. User perception is in principle hypothesis testing; hence, if the agent's behaviour or traits match user's expectations, it can be considered plausible and therefore accepted. Thus, to construct psychologically plausible robots, one can define a mini psychological engine or a set of features that if fulfilled, can account for believability.

To answer the central question, we divide plausibility into two approaches: psychological and implementation. To be more precise, we are interested in understanding "What are the behavioural traits that allow humans to perceive a robot as a believable agent?" which in turn leads us to examine "what kind of control system does a robot require to be psychologically plausible?". We offer four possible psychological benchmarks for consideration: *autonomy*, *morphology*, *social competence* and *task competence*.

The suggested benchmarks aim at decomposing psychological plausibility to discrete parts that can be tested empirically and use their interactions in practice for the meaningful design and development of social robots. Although there have been some attempts to examine humans' responses towards robots and establish standard metrics, few attempts have been made to establish psychological benchmarks.

Autonomy is viewed as an individual's capacity for self-governance and most would agree that is normatively significant. In robotics, autonomy is the ability to make decisions and perform actions without human intervention using internal decision-making mechanisms.Here we view autonomy in two ways: from a technical (implementation) and psychological perspective. Indeed, robots need to be autonomous, that is, successfully perform a task and function in a way that does not require human intervention or operation. The question that arises from this aspect is: "Is the robot able to carry out successfully the task it was designed to perform?". If a robot is not autonomous when performing a task, it is possible that at some point it will face anomalies that exceed its pre-programmed capabilities and stop responding. From a psychological perspective, a question that derives is: "Will humans perceive it as autonomous?". For users to see a robot as autonomous, the robot's behaviour and features should comply with the psychological norms that match their expectations. Typically, these include: the performance of actions and making decisions without the influence of others, since actions follow the robot's own will. Then, the robot can be considered autonomous and believable, as it will act in a proactive way.

Regarding *morphology*, we claim that for a robot to be psychologically plausible, its design needs to serve the task it was meant to execute. However, to be effective, a robot does not always need to resemble a human. Advocates of humanoid robots claim that human-like design benefits and facilitates HRI, as such morphology enables communication channels that resemble those of humans. An essential requirement that needs to be fulfilled with this respect is the readability and transparency of the employed communication channels. Thus, a question that arises is: "Are the robot's commu-

nication channels readable by the human partner?".

Social competence as taxonomy is inherently more complex because the robot's social success can be measured in a variety of ways. A first question that we should be asking is: "Does the robot successfully assume its intended social role?". The evaluation of the robot's social role is relatively straightforward. If the robot aimed to be a social partner, do users socially interact with as if it were a peer? A second and equally important topic is, "In what way do the various social components of the robot affect its psychological plausibility?". Here, things become slightly more complicated, as now we are called to answer a number of derived questions. In this respect, we decompose social competence into discrete parts and see how they individually or in conjunction affect human acceptance.

Finally, *task competence* refers to the robot's capability to successfully perform a certain task. For example, if the robot's task is to provide information regarding certain exhibits in the museum, a useful method to evaluate the robot's competence would be to ascertain that the robot can understand users' requests and give appropriate answers. Similarly, in the tutoring domain, task competence could be translated into the robot's ability to engage the learner, provide useful feedback and achieve knowledge transfer.

1.1 Research objectives

The main goal of this thesis is to create a robot that is accepted by its human partners. To do so, we focus on the psychological plausibility of the robot. More specifically, we describe the design, development and study of social robots intended for dyadic interactions and propose four benchmarks aimed at evaluating the robot's behaviour and plausibility at the following domains: *autonomy*, *morphology*, *social competence* and *task competence*. The first chapters of this thesis aim at providing a general overview of the morphology and behavioural components of current robotic systems that socially interact with humans. In the following chapters, we go in more

detail on our current system implementation and the evaluation methods of our research questions.

The first question we ask ourselves is: "Is the need to create a psychologically plausible artificial agent new?". The answer to this question is: no. In fact, the abiding desire to create artificial life dates back hundreds of years. In Chapter 2 we provide examples of archetypes that are characterised by efforts made to understand and imitate biological organisms regarding functionality, physical appearance, processes and complex life-like behaviours. Given our focus on robots with social character, we are compelled to define "What is a robot?" as we do not find sufficient the existing definitions.

Given our interest in robot behaviour and how it affects user perception, we primarily explore the morphological and behavioural approaches to existing studies on HRI in chapter 3. More specifically, we present the effects of morphology on user perception. Additionally, we present existing behavioural and social strategies employed to create robots that socially interact with humans. From these, we identify three key concepts that are relevant to our proposed taxonomy: the expression of internal states, the usage of gaze and the elicitation of proactive behaviour.

Chapter 4 revolves around the presentation of the proposed taxonomy where we explain in more detail the motivation behind selecting the criteria above and our evaluation methodologies.

In chapter 5 we present the implementation of a sociable robot, namely H5W_Alpha. The central question this chapter answers is: "Can we create an agent that behaves in an autonomous way?". To answer the question of autonomy, we use the Distributed Adaptive Control architecture Verschure (2012); Verschure et al. (2003) that controls the robot's behaviour and we describe its main principles. The usage of the DAC architecture is two-fold. On the one hand, it allows the robot to perform tasks without human intervention (implementation). On the other hand, it consists of a motivation system that allows the robot to behave proactively, what can be perceived as autonomous behaviour by the user.

The next sections focus on the studies conducted to evaluate the psychological validity of the proposed synthetic agent. Chapter 6 answers the following question: "How do the various robotic features affect the plausibility of the robot?". Here, we show the first attempt to implement the DAC architecture on a social humanoid robot, endowed with a set of drives that aim at initiating and maintaining an interaction with a user through a game-like scenario. Results indicate that the robot is able to trigger behaviours that aim at satisfying the robot's needs. Our results show an interplay between drives, emotions, perceived stimuli and actions while we display key features of the overall system. Indeed, the robot can behave autonomously, even if not all preconditions are matched. Upon the evaluation of the key behavioural components of the proposed system, we focused on answering a fundamental question regarding the morphology of the robot: "Are people able to recognise the facial expressions and prosodic features of a robot and correctly attribute to the robot internal states?". To evaluate the transparency of the robot's communication channels, we varied the facial features of the robot (eyebrows, eye opening and mouth) and asked participants to rate them in terms of valence and arousal. Results suggest that there is a correlation with valence and mouth and eyes and arousal but not a combination of both.

To assess the robot's social competence, the first question we pose is: "Does the complexity of social behaviour affect the way humans perceive the robot?". To do so, we decomposed social behaviour in a number of discrete cues such as gestures, touch, speech, gaze, facial expressions and proactive behaviour. We define complexity as the number of cues used simultaneously. To assess how these behavioural components affect the robot's believability we devised five interaction scenarios of increased complexity and asked participants to evaluate the robot. Results suggest that the more the robot appears socially competent, the higher it scores in believability.

A second question we pose to evaluate the robot's social competence is: "Do social cues like emotional expression and gaze affect the elicitation of empathic responses towards robots?". To answer this question, we hy-

pothesised that social competence could trigger empathic responses. We therefore manipulated the robot's gaze model and facial expressions and looked at the empathic relation between the participant and the robot. We speculated that if the robot is psychologically plausible, it will elicit empathic responses from the observer. In this context, empathy is a measure taken by the observer as the effectivity of the robot's social cues. We characterised as empathic responses the participant's time of administration of the negative stimulus, gaze mode and behavioural reactions. Participants seemed to show empathic responses toward the robot.

Additionally, we aim at developing a theoretical understanding of psychological plausibility capitalised in the domain of tutoring. More specifically, in chapter 7 we evaluated the robot's task and social competence in dyadic scenarios. First, we examined the role of facial expressions and gaze model in an educational task. The main question of this study was "Does the robot's facial expressions and gaze model affect knowledge acquisition?". We conducted this experiment with adults and children. Although results were not conclusive regarding the effects of the robot's social components or task competence, we identified an impact of the role of gaze in engagement.

Having evaluated the robot's psychological plausibility with adults, we now focused on the psychological plausibility of the robot with children. More specifically we asked "Can we extract valuable information or design guidelines from children's drawings?". To answer this question, we exposed children to three different robotic platforms and asked them to evaluate them. Additionally, we asked them to draw the robot they would like to have and interact with and assessed their drawings regarding functionality and morphology. Results suggest that children tend to design multi-purpose robots that are anthropomorphic but are more machine than human-like.

Additionally, we pose a more tutoring system level question: "To what extent can we replace a task that is typically performed by a human with non-anthropomorphic technology?". The aim of this study is to understand which features are relevant if we have a non-anthropomorphic setup. Here,

we ask the learner to trust and give plausibility to a training system that is non-anthropomorphic, as we cannot exclude that a more pragmatic approach could also work. To do so, we employed an inquiry-based learning task that teaches children physics, namely the balance beam problem. To evaluate this non-anthropomorphic setup, we used three different content presentation tools: a physical balance, a virtual balance and a motorised balance coupled with an augmented reality application.

Finally, the central question we ask to evaluate the robot's task competence is: "Does the robot's help mechanism allow students gain a better understanding of the task and therefore be accepted by students as a peer?". As a tutoring task, we used the balance beam problem and varied the nature of the robot's help by providing hits (open/closed) and distractions (jokes/trivia). In this scenario, the robot used the virtual balance (evaluated in the previous study) as a tool to convey content. The aim of this study was to see what are the minimum set of tools and behavioural components needed to efficiently and effectively teach children physics. Results indicate that children enjoyed the interaction and found the feedback of the robot helpful.

Chapter 2

The creation of believable agents: early attempts

Robots are present in many aspects of our lives and are an important part of our culture, as references to robots can be found in mythology, sci-fi films, novels or even music. Archetypal stories of life-imitating machines appear in many popular culture films and science fiction writings. Many might be familiar with renowned robots like R2-D2 and C3PO from Star Wars, the Terminator, WALL-E, Rachel from Blade Runner, HAL 9000 from Space Odyssey, as these characters have influenced the way people imagine or perceive robots. What all these fictional characters have in common is the ability to appear as autonomous agents, with their own thoughts, motives and personality.

So what makes people accept these robots as believable entities? This is the main question we try to answer in this thesis, and our main contribution lies in defining a taxonomy that allows to understand better the factors that affect robot acceptance. Before we analyse our approach to this question, it is worth mentioning that although robots became popular in the twentieth century, the need to create believable agents is not new. In fact, since the beginning of time, humans have employed both art and technology to create devices that approximate human intelligence, appearance as well

as capabilities and behaviour. Early traces of such attempts already date back to the Middle Ages with the so-called *"automata"*. Automata can be considered the predecessors of modern robots and are known for simulating aspects of living organisms such as movement or behaviours. Although the idea of artificial agents initially stemmed from pure imagination, the efforts of resourceful individuals have laid the ground for the development of modern robots. Thus, the creation of believable machines is a challenge that not only modern roboticists face; the automata-makers of that period faced similar problems.

Can we, therefore, draw examples of psychologically plausible machines from history, and more specifically, from automata-makers? In the following sections, we explore early attempts of mechanical artefacts that approximate nature regarding functionality, physical appearance, processes and complex life-like behaviours. More specifically, we present pioneer examples of both animal (like Vaucanson's "Digesting Duck") and human anatomy and kinesiology (like da Vinci's mechanical Knight). We introduce endeavours of elaborate behavioural characteristics that produce the illusion of life, like Jaquet-Droz's creations. However, despite their strikingly life-like behaviour, the automata of that period were senseless devices that performed a set of predefined actions. To create robots that can socially interact with humans, they need to sense their environment and act upon it. Thus, we present the archetypal efforts to develop autonomous machines that can perceive the surrounding world, like Walter's tortoises and *Shakey*, and finally, we provide our proposed definition of what is a robot.

2.1 Automata that imitate life

The abiding desire to create believable machines dates back hundreds of years, but it is not until the European Renaissance that we observe the usage of automata to explore convincing behaviours. To do so, the automata-makers studied and attempted to imitate the functionality of both animals and humans. Thus, the automata of that period did not only serve to entertain but can be considered as philosophical experiments that allowed the reproduction of aspects of living organisms in machines, while revealing important information regarding their nature. What initially started as a philosophical idea turned into a mechanical revolution as most of the automata of the 18^{th} century were not only imitating the external appearance of an organism but also simulated the organism's functionalities or behaviours Riskin (2003a).

An example of linking human kinesiology and anatomy is Leonardo da Vinci's "Knight" (Figure 2.1) in 1495 Moran (2006). An elaborate system of pulleys and cables moved the Knight's armour to produce various human-like independent motions. This compelling artefact has endowed modern robotics with scaffolds for kinematics and structural design Rosheim (1997).

Figure 2.1: Model of Leonardo da Vinci's mechanical Knight. Photo by Erik Möller. Mensch - Erfinder - Genie exhibit, Berlin 2005.

A way to appreciate the early simulation of living beings is the central idea of "moving anatomy" in the creations of Jacques de Vaucanson (1709 - 1782). Vaucanson modelled animals and humans to assess their organic functions based on the principles he had formulated Fryer and Marshall (1979). One of his first biomechanical automata was the "Flute Player" Moran (2007), a life-sized wooden statue of a man who played the flute by emitting air through its mouth. This design resulted from the extensive study of human flute players and was used to validate Vaucanson's hypothesis that the consequent pitch of a note was affected by the blowing pressure, aperture and sounding length. Notably, his most famous creation was the "Digesting Duck" (1739) a mechanical artefact modelled upon thorough studies of real ducks that was conceptualised to teach the animal's anatomy (Figure 2.2). The duck was able to flap its wings, eat grains, drink water and even defecate small pellets from its rear Riskin (2003a). Both the "Flute Player" and the "Digesting Duck" are examples that intended to approximate their biological counterparts and be believable.

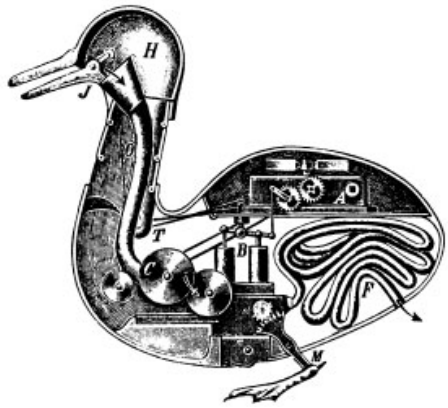


Figure 2.2: Illustration of the postulated internal mechanisms of Vaucanson's "Digesting Duck" by an observer.

Attention to anatomical, physiological and behavioural simulations started with Vaucanson and climaxed with Pierre Jaquet-Droz's (1721 - 1790) creations. The father-and-son team of Pierre and Henri-Louis Jaquet-Droz produced three automata: "the Writer", "the Draughtsman" and "the Mu-

sician" (Figure 2.3), which could also be reprogrammed. The Writer could write any custom text up to 40 letters long, while the Draughtsman could draw four different images: a portrait of Louis XV, a royal couple, a dog and a Cupid pulled by a butterfly. Finally, the Musician could play five different melodies on a custom-built musical instrument. These artefacts resemble humans not only externally, but also carefully follow the mechanisms that produce specific behavioural or functional manifestations. For example, their hands were modelled after real human hands that later constituted the basis to construct prosthetic limbs Riskin (2003a). The imitation of "life" or the plausibility of those machines was achieved with attention to behavioural details. For example, the Writer held a quill that it dipped into an inkwell and then shook it lightly; both the Musician and the Draughtsman displayed breathing and action-sustained gaze behaviours. Finally, the Musician performed movements borrowed by musicians (like balancing the torso) and sighs in time to the music, appearing endowed with emotions Riskin (2003b).

Figure 2.3: Image of the Jaquet-Droz automata: the "Draughtsman" (left), the "Musician" (middle) and the "Writer" (right) from the musée d'Art et d'Histoire de Neuchâtel.

The tendency of that period was to use mechanical artefacts to approximate nature and, through modelling, experimentation and observation, draw conclusions about their biological counterparts. Special care was taken for their morphology and task competence. Their design was both aesthetic

and functional: it resembled their biological counterparts and allowed the exhibition of the appropriate behaviour. At the same time, task competence was attributed to the mechanical ingenuity of their creators and the extensive studies of human anatomy, as their morphology supported the performed task and allowed them to express similar movements and behaviours. For example, they were equipped with hands and moving fingers to play the piano or even specially constructed wings to make them flap. To achieve believability, the automata of that period exhibited behaviours that usually accompany such tasks: action-sustained gaze, breathing and torso movement while playing the piano or even shaking the ink off the quill. Thus, to accept them as believable agents, their morphology, movements and behaviour matched the observer's expectations.

Attention to physiological or functional components of biological beings seemed to be critical for both the operation of the machinery and the simulation of psychologically plausible behaviours. At the same time, apprehension of living creatures and machinery continuously redefined each other. The attempt to mechanically resemble life provided examples of widely used anthropomorphic components (like action sustained gaze) and led to developments that became the foundation of modern robotics. However, philosophers and biologists of the eighteenth and nineteenth century argued that one of the main characteristics of living organisms is their ability to maintain their internal states stable (maintain "homeostasis"), be autonomous and be responsive to their environment while automata were not Riskin (2003a,b). Indeed, given our taxonomy, to create believable agents, *autonomy* is an important feature, while the ability to perceive the environment plays a key role in the task and social competence of the robot.

2.2 First examples of autonomous machines

The technological advancement of the 20$^{\text{th}}$ century allowed for the development of autonomous apparatuses that could perceive their environment. Machines are now equipped with a variety of sensors and are endowed with decision-making mechanisms to perform the appropriate actions. Thus, the behaviour is no longer dissociated from the surrounding world but in contrast, is generated in response to it. Now, the creation of believable agents requires not only a particular set of tools and behavioural repertoires (as was the case with the automata of the previous centuries) but also a control mechanism that defines which and when each action should be executed.

The first autonomous robots that displayed complex behaviour were created by William Grey Walter (1920 - 1977) in the 1940s, known as *Machina Speculatrix* (a name that illustrated the speculative behaviour of most animals) or 'tortoises", due to their appearance (Figure 2.4). They consisted of two sensors (directional photocell for light detection and bump sensor for contact detection), two actuators, a battery and two "nerve cells" that formed part of the decision-making system. Despite their simple action repertoire (attraction by moderate light, repulsion by bright light and obstacle avoidance), their behaviour was designed to resemble that of animals: seeking out favourable conditions, showing uncertainty, random exploration of the environment, and similar reactions to stimuli Walter (1950, 1951). Their moderately complex behaviour resulted from rich connections between sensors and effectors, incorporated into feedback loops Freeman (1986). Walter's tortoises are an excellent example of autonomous machines with convincing behaviour, as they appear to act purposefully: their actions depended on both the environment (obstacle avoidance or random exploration) and their internal states (return to the station to recharge).

Following Walter's principles, Valentino Braitenberg (1986) explored the emergence of complex dynamic behaviours by varying the excitatory or inhibitory connections between a vehicle's sensors to its motors (Figure 2.5). For Braitenberg, movement suggests the impression of life. His thought

Figure 2.4: Image of William Grey Walter and his two *Machina speculatrix* 'tortoises": *Elmer* and *Elsie*. Burden Neurological Institute (BNI) archives, courtesy of Owen Holland.

experiments produced a plethora of convincing behaviours that could be interpreted as "approach", "fear", "aggression", "love" or even "shy but defensive". Thus, both Walter and Braitenberg explored the emergence of behaviour when perception interacts with actuation, highlighting the role of interaction: emergent behavioural complexity derives from the interaction between perception and actuation and not the internal control of the machine. Consequently, to perceive and act, they emphasised on one fundamental property of robots: having a physical representation (or body) that can both perceive its environment and act on it.

In the previous section, we presented the first attempts to create believable machines in terms of morphology and task competence. We argue that a major drawback of the automata of that period was the lack of sensing, as their actions were not the result of interaction with their environment, but a set of predefined behaviours. Here, we elaborated on attempts to create

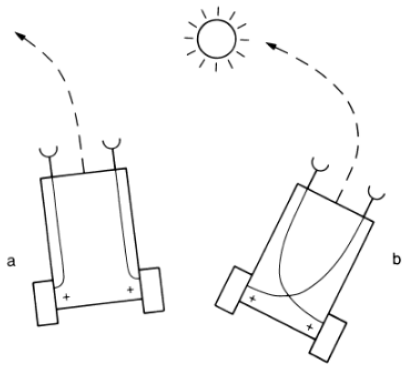


Figure 2.5: Illustration of two vehicles that have two light sensors (one at each side) and two motors. Vehicle (a) has each sensor connected to the motor on the same side, while vehicle (b) has each sensor connected to the motor on the opposite side. The resulting behaviour of vehicle (a) is light aversion, whereas the behaviour of vehicle (b) is light attraction. Image adapted from Braitenberg (1986).

machines that sense their environment and act upon it. More specifically, we have examined the production of simple behaviours in simple systems consisting of a couple of sensors and actuators. However, modern robots are getting increasingly more and more sophisticated and in fact, consist of a plethora of sensors and actuators. For example, one of the first robots that was equipped with multimodal sensors (like a camera, a range finder and bump sensors) was *Shakey* Nilsson (1984). *Shakey* was created by the Artificial Intelligence Centre research group at the Stanford Research Institute (SRI) and was considered the first intelligent mobile robot with the ability to perceive its environment, construct symbolic representations, and use them to form plans to achieve goals.

So, how can we control such robots and how can behaviour emerge in sophisticated systems? The example of Walter or Braitenberg, where sensors were directly connected to the actuators, cannot be applied here. Not only the number of sensors is higher, but also, the information provided is more elaborate and needs some computation. Just like Shakey's example, a solution to this problem lies in layered control architectures. This robotic

platform significantly contributed to the field of robotics and artificial intelligence: it encouraged the development of more elaborate robots as its computer vision, and planning algorithms were later used in various applications Haber and Sammut (2013). Hence, in this section, we emphasised on the importance of embodiment that allows for both sensation and actuation, coupled with a control system that allows for the emergence of more complex behaviours from their interactions. In the later chapters, we present our proposed architecture that allows for the control of a robotic platform and the production of complex behaviours.

In the 1960's, the usage of *Unimate* in the assembly line of General Motors revolutionised the automotive industry and set the grounds for the development of general purpose machines with a broad diversity of applications. Since that time, the term "robot" became a falsely popular metaphor to refer to any human activity that is replaced by a machine. Concepts ranging from automatic vending machines to drones or even to the iPhone's Siri are often wrongly referred to as robots. Nonetheless, Roomba, the vacuum cleaner from iRobot can be considered a good example of a robot. So what can be regarded as a robot? This misconception requires a clear definition and disambiguation of the term.

2.3 What is a robot?

It is difficult to define what a robot actually is, as it requires a description of both its appearance and functionality. Historically, the term "robot" was first coined by Karel Čapec in his play *Rossum's Universal Robots* (R.U.R) in 1921 and comes from the Slavic word *robota*, which literally means "work", "labour" or "hard work". The "Father of Robotics" Joseph Engelberger famously said: "I can't define a robot, but I know one when I see one". Nonetheless, definitions of robots may vary from very general, *"a machine capable of carrying out a complex series of actions automatically, especially one programmable by a computer"* (Oxford dictionary), to more technical ones: *"'a reprogrammable, multifunctional manipulator designed to move material, parts, tools, or specialised devices through various programmed motions for the performance of a variety of tasks"* (Robot Institute of America, 1979).

According to Brady, these definitions do not include sensing: robots are active devices that operate in and interact with a non-static environment. He therefore defined robotics as *"'the intelligent connection of perception and action"* Brady (1987). Also, Arkin argues that the link between perception and action is essential, by defining a robot as *"a machine able to extract information from its environment and use knowledge about its world to move safely in a meaningful and purposive manner"* Arkin (1998). Indeed, this idea was already introduced in Walter's tortoises and Braitenberg's vehicles, as discussed in section 2.2.

We argue that none of the above definitions are sufficient to characterise a robot. Indeed, as technology advances rapidly, the definition of a robot is elusive, as the appropriate answer changes too quickly (Nourbakhsh, 2013, p. xiv). What older definitions lack, as very well put by Brady is the inclusion of sensing. Though Brady intended to link perception with action, his characterisation was too general, as there is no formalisation to how an intelligent connection between perception and action can be interpreted. In that sense, many of the modern home appliances could fit into the descrip-

tion. According to his definition, an intelligent washing machine equipped with sensors that automatically decides the optimal dosage of detergent and water dispense could be a robot, when in fact it is not. Arkin's approach specifies in more detail (compared to Brady) the relation between sensing and actuation. However, his terminology seems to be highly focused on movement and may exclude other functionalities of robots; in this case, perhaps a more correct term would be *action*.

So how can we better characterise a robot? It is true that most domains suffer from definitions, and what all the above ones lack is *embodiment*. Dautenhahn defines embodiment in robots as "*that which establishes a basis for structural coupling by creating the potential for mutual perturbation between the system and the environment*" Dautenhahn et al. (2002). The importance of embodiment is widely acknowledged in a variety of domains and in many cases, embodiment is explicitly linked to intelligence; key concepts like "adaptation", "behaviour" and even "generation of behavioural diversity" imply the existence of a body that interacts with its environment Pfeifer and Scheier (2001). The belief that intelligence requires a body is advocated by Brooks (1991a,b) as a necessary component to experience and deal with the world directly, followed by the field of *behaviour-based robotics* Arkin (1998).

Embodied systems allow surpassing the internal symbolic representation problem that is classically employed by AI approaches. To do so, an agent is not only required to have a body but also be *situated*. This reflects the agent's ability to acquire relevant information regarding a situation through its sensors in interaction with the world. Now the agent can interact with a situation: *"The real world is, in a sense, part of the "knowledge" the agent needs to behave appropriately. It can merely "look at it" through the sensors. In a sense, the world is its own best model."* Pfeifer and Scheier (2001). For Dautenhahn and Christaller (1995), what accounts for embodiment is taking into account the body's properties and shape, what it can perceive from the environment or how it can interact with it. In these perspectives, sensing is of direct relevance and influence to the robot's actions, which affect the

environment, which in turn influences sensing. Embodiment is therefore grounded in the relationship between the system and the environment it is situated in. This view highlights an intricate interconnection and interaction between the brain, the body and the world.

Hence, our working definition of a robot is: "an autonomous and embodied agent that is physically instantiated and situated in the world with the capability to physically act, able to extract meaningful and relevant information from its environment and itself that informs these actions".

Chapter 3

Design approaches for social robots

Robots are created to perform a diversity of tasks, serve a variety of purposes and are extensively used in many domains. Traditionally, robots are employed in settings that require routine operations or are considered dangerous for humans. For instance, in industrial settings as well as the automotive industry, operations like handling materials, assembling or painting are almost exclusively performed by robots Bekey and Yuh (2008). Additionally, in large warehouses, hundreds of autonomous guided vehicles are used to transport products to workers Wurman et al. (2008). These robots require some degree of autonomous operation as well as the ability to make decisions and perform tasks Bar-Cohen and Hanson (2009).

The industry is not the only area that can benefit from the introduction of robots. Robots are currently employed to examine or measure aspects of challenging environments where data sampling and monitoring were previously done manually. Examples include monitoring of marine mammals Klinck et al. (2009), pollution Trincavelli et al. (2008) or even reefs Dunbabin et al. (2004). Machines are now able to track and follow the source of a chemical plume Vouloutsi et al. (2013b); Distante et al. (2009). Additionally, they are widely used in a variety of medical applications Burgner-Kahrs

et al. (2015) including rehabilitation and surgery Calinon et al. (2014). However, most of these examples are constrained in situations where little interaction with humans is required.

Nowadays, the development of robots goes beyond utilitarian purposes, as we observe a change of paradigm: robots that operate in a close proximity to humans start to gain ground. Their functionality ranges from professional maintenance and defence to vacuum-cleaning and companions. The robots that perform useful tasks for humans or equipment (excluding industrial automation applications) are commonly referred to as *service* robots *(International Federation of Robotics - IFR)* (ISO 8373). They are classified as either *professional* or *personal* service robots Thrun (2004). Professional service robots are typically operated by a trained human and are mainly used for commercial tasks. They include systems for professional cleaning, inspection, maintenance, defence and medical applications. In contrast, personal service robots are mainly used by laypersons for non-commercial tasks such as domestic use (like vacuum-cleaning or lawn-mowing), education, health-care, entertainment etc.

The usage of service robots is becoming popular, as according to the IFR, in 2015 about 41.100 service robots for professional use were sold, and more than 333.000 units are expected to be sold between 2016 and 2019. At the same time, 5.4 million service robots for personal use were sold (3.7 million for domestic use and 1.7 million for entertainment or leisure activities), accounting for an increase of 16% from the previous year IFR (2016). Indeed, personal service robots have been operating in households for many years, especially in the field of domestic use and entertainment. For example, the robotic dog *AIBO* Fujita (2001) came out in 1999 by Sony. Special effort was made to approximate as close as possible the movement of its biological counterpart and it became popular for both its appearance and autonomous behaviour Fujita (2001); Fujita and Kageyama (1997); actually, more than 150.000 AIBOs had been sold until 2006 Siciliano and Khatib (2016). In the field of household cleaning, the first autonomous robotic vacuum cleaner *Roomba* was introduced by iRobot in 2002 and more than 15 million units

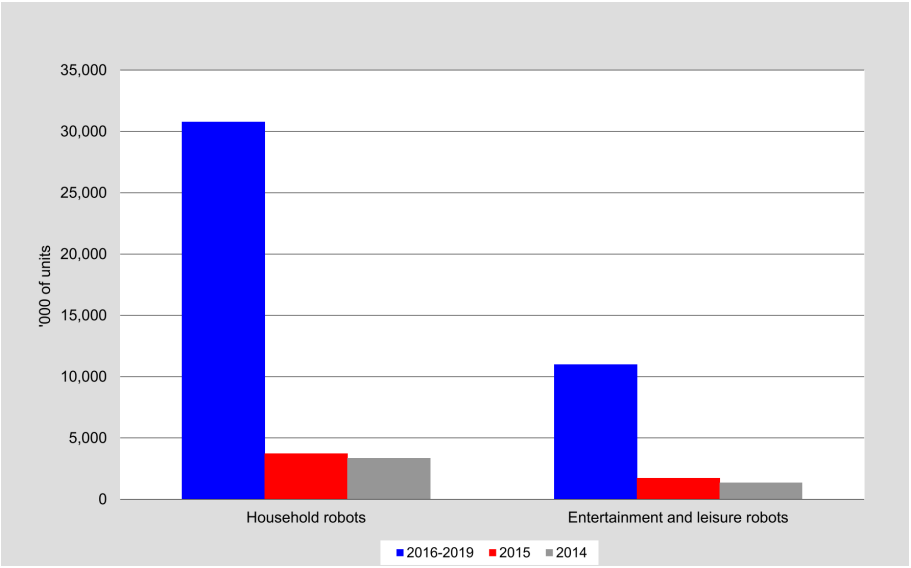


Figure 3.1: Unit sales in 2014 (grey) and 2015 (red) of service robots for personal or domestic use (in thousands of units). The estimated forecast of unit sales for 2016-2019 is depicted in blue. Image taken from IFR (2016).

have been sold ever since iRobot (2015). These examples highlight the growing popularity of personal service robots and according to IFR, an increase in sales to more than *40 million units* is expected by 2019 (see Figure 3.1).

Given this prediction, we can assume that robots will soon become a part of our daily lives; they will operate in a close range to humans and will be required to frequently interact with them. In fact, even if the primary goal of a machine is to perform a task that demands little interaction with humans (like maintaining a household clean), stable interactions can still emerge Sung et al. (2009). It is therefore important to understand how these interactions emerge and what is their nature. Additionally, we need to explore in what way the behaviour of the robot affects the interaction and how humans perceive the robot. In this chapter, we examine the design approaches taken to create robots that socially interact with humans. These approaches can be divided into two main categories: *morphology* and *behaviour*, as both can influence humans' actions and perception toward the

robot.

So what defines the design and behavioural characteristics of robots? Notably, the interaction capabilities of the robots and the tasks they are required to perform vary greatly. More specifically regarding the interaction capabilities, Breazeal (2003b) defined four categories of robots. Each category is characterised by the social model ascribed to the robot and the complexity of the interaction, starting from simple to more complex ones: *Socially evocative* robots are designed to encourage people to anthropomorphise them and attribute to them social responsiveness even if their behaviour does not reciprocate; the robots that act as *Social interface* employ social cues and communication interfaces that are easily recognisable by humans to facilitate human-robot interaction, however, their responses are mostly predefined; the *Socially receptive* robots can perceive the social cues exhibited by humans and learn through socially interacting with them, however, they do not behave proactively to satisfy their social aims. Finally, the *Sociable* robots are equipped with an internal system of goals and motivations and proactively engage humans in social interactions to satisfy their needs.

Additionally, Dautenhahn (2007); Fong et al. (2003) proposed four complementary categories of social robots: the *Socially situated* robots can perceive the social environment and distinguish other agents or objects from themselves; the *Socially embedded* robots are situated in a social environment and are structurally coupled with it. Finally, the *Socially intelligent* robots show similar models of human cognition and social competence while for the *Socially interactive* ones, social interaction plays a pivotal role.

Similarly to the interaction capabilities, the tasks robots perform also vary. For example, in entertainment, artificial pets or human-like machinery are created to interact in a playful manner with the environment and their users. Their activities include exploration of their environment, playing and even dancing. Robots are deployed in public spaces such as museums to deliver educational content, guide or engage visitors with various exhibits Thrun

et al. (2000); Shiomi et al. (2006); Nourbakhsh et al. (2003); Bennewitz et al. (2005). Machines are found in universities as receptionists, where they interact with humans and provide information about the room of a faculty member or give directions Gockley et al. (2006b,a); Kirby et al. (2010) and even provide tour guides Salem et al. (2015). They have also been incorporated in busy airports to guide transfer passengers Joosse and Evers (2017); Triebel et al. (2016) and provide directions at shopping malls Kanda et al. (2010) or train stations Shiomi et al. (2008).

In health care, robots are used to deliver medication or meals Mutlu and Forlizzi (2008). By interacting with patients, they offer psychological improvements to the elderly Wada and Shibata (2007a, 2006b), provide companionship Stiehl et al. (2005); Sabelli et al. (2011) and can even act as mediators by improving and strengthening the relationships between patients Tamura et al. (2004); Wada and Shibata (2007b); Kidd et al. (2006). In other cases, robots are used to remind their users of their daily activities or even guide them through their environment Pollack et al. (2002). Through encouragement, motivation and companionship, robots may assist adults that suffer from dementia Tapus et al. (2009), help in the rehabilitation process of post-stroke patients Tapus et al. (2008) or even achieve behaviour changes during dieting (Kidd and Breazeal, 2008).

Numerous developments are aimed at the application of robots as therapy tools for autism Scassellati et al. (2012); Cabibihan et al. (2013); Dautenhahn and Werry (2004) as they may allow for the development of social skills Robins et al. (2005) and the elicitation of desirable behaviours such as initiative-taking François et al. (2009). Additionally, they are used to support the development of self-efficacy in young children in their effort to manage a lifelong metabolic disorder such as diabetes Baxter et al. (2011); Lewis and Cañamero (2014) or even assist in the alleviation of stress in young cancer patients Alemi et al. (2014); Jeong et al. (2015).

In the field of education, robots are used as educational tools that help students develop and strengthen certain skills like programming Kabátová and

Pekárová (2010), or acquire knowledge primarily in STEM (Science, Technology, Engineering and Math) areas Benitti (2012). Robots are also used as peers or tutors to teach a variety of subjects ranging from a secondary language Kanda et al. (2004a); Gordon et al. (2016), Chinese handwriting Teo et al. (2002), prime numbers Kennedy et al. (2015b) or physics Reidsma et al. (2016). Additionally, robots are used to promote the concept of healthy living in children by motivating them to do physical exercise and helping them understand the concept of energy spent while exercising Fernando et al. (2016); Cameron et al. (2016) or even by forming long-term relationships with the students Kanda et al. (2007).

Robots are also employed as assistants or social companions Breazeal (2004, 2003b, 2002) and are designed to interact with a variety of users ranging from young children Kanda et al. (2004c) to the elderly Prescott et al. (2012). They are called upon to carry out a multitude of activities ranging from household (e.g. cooking or cleaning) or fetching and carry tasks Graf et al. (2009), to providing access to information on demand or even taking photos or recording videos.

These illustrations highlight the large diversity of the functionalities robots can exhibit. Based on all the previous examples regarding the tasks robots perform and the taxonomy of their social capabilities, it is clear that not all interactions are the same and not all robots require the same interaction capabilities. For example, a robotic receptionist can interact with several humans, however, its functionality is usually limited to the provision of informative content, and the social skills displayed typically serve to facilitate the interaction. In contrast, a robot companion typically interacts with one person and the tasks performed may vary greatly; such robots are required to exhibit a wider variety of behavioural complexity to be accepted by humans. Thus, to answer the question "what defines the design and behavioural characteristics of robots?", a crucial component is the application domain and the nature of the interaction. To establish the principles of social interaction with agents, Dautenhahn (2007) proposed the following models:

- *Robot-centred HRI*: the robot is viewed as an autonomous entity that has goals deriving from its own drives, emotions and motivations. The robot's actions and behaviour (e.g. engage in interactions with humans) aim at satisfying its needs.
- *Human-centred HRI*: here, the main interest lies in humans' perception of the robot's appearance and behaviour, which should be acceptable and comfortable to humans.
- *Robot-cognition HRI*: the robot is considered an intelligent system that is able to make decisions and solve problems as part of the task it is required to perform. The robot is typically controlled by a cognitive architecture and machine learning mechanisms.

Our main interest is to develop a psychologically plausible robot that proactively and intuitively engages humans in social interactions. Hence, the problem of plausible robots that we target embraces all three cases: the robot should be socially motivated to act, while its behaviours are accepted by humans and at the same time, considering the individual differences of humans and adapting to them. In the following sections, we investigate the required morphological and social skills of such robots.

3.1 Morphology

In general, social interaction does not necessarily require a body to be successful. Nonetheless, in the previous chapter (see section 2.3) we have emphasised the role of embodiment in perception and action, its inherent link to intelligence and how it affects the interaction with the world. When it comes to social agents, embodiment provides qualitative advantages over other non-embodied interfaces, given that the body is used to leverage knowledge of human communicative behaviour Cassell et al. (2000) which in turn improves information transfer Boyle et al. (1994).

Embodiment, combined with shared context and physical presence, is critical for establishing a successful communication Duffy et al. (1999); Fong et al. (2003); Breazeal (2004); Bartneck and Forlizzi (2004) and therefore be accepted as communication partners. For instance, embodied robots are preferred to animated characters or their virtual representations, as they seem more engaging Kidd and Breazeal (2004); Wainer et al. (2006); Bainbridge et al. (2008). They are perceived as more trustworthy compared to non-embodied robots Bainbridge et al. (2008) while they were found to be more useful and were evaluated as effective communicators Powers et al. (2007). Despite the positive implications of embodiment and physical presence, it remains unclear how they affect user perception or the quality of the interaction. To be more precise, some studies suggest that they are definitive components for the difference in responses Bainbridge et al. (2008), while other studies showed no difference Kidd and Breazeal (2004). Regardless of the inconclusive results, in this thesis, and given our definition of a robot, we focus on embodied robots that are physically present and examine their design possibilities.

There are various approaches one can take when designing a robotic platform, ranging from the material it is made, the sensors it is equipped with to its design and morphology. Given the vast number of possibilities, a question arises: "Is there an optimal design for robots that interact with humans?", which in turn forces us to ask "Does the design of the robot

matter in human-robot interaction?". When it comes to industrial robots, answering these questions is easy. The design does matter, and it heavily depends on the tasks the robot is meant to perform. In contrast, the answer becomes more complicated when it comes to social robots, as there are still no direct guidelines regarding their morphology and design.

3.1.1 Robot appearance

Before we provide an answer to these questions, we first present a taxonomy of robots that are typically employed in social interactions with humans. Fong et al. (2003) identified four broad categories of embodied robots used in human-robot interaction (HRI), based on their morphology: *anthropomorphic* (the appearance resembles that of humans), *zoomorphic* (the appearance resembles that of animals), *caricatured* (appearance is not necessarily realistic or believable and usually have exaggerated features to provide a comic effect) and *functional* (the embodiment reflects the task the robot performs). Anthropomorphic robots are divided in two broad categories: humanoids and androids. Humanoid robots are usually characterised by an embodied form that emulates features of human appearance and behaviour that typically allows for the generation of human-like facial expressions, motions and gestures. Examples of humanoid robots (Figure 3.2) include Honda's ASIMO Sakagami et al. (2002), Aldebaran's Nao robot Gouaillier et al. (2008), Hanson Robotics' Zeno Hanson et al. (2009), Breazeal (2004)'s Kismet, and the iCub robot Metta et al. (2010), developed by the Italian Institute of Technology (IIT).

Android robots are designed to physically resemble humans Coradeschi et al. (2006) with a highly realistic face made from materials that appear like skin Tzafestas (2015). Examples of androids (Figure 3.3) include the highly expressive F.A.C.E (Facial Automation for Conveying Emotions) Lazzeri et al. (2013a); Mazzei et al. (2014) and Albert HUBO Oh et al. (2006). Similar to androids are geminoids that are usually teleoperated androids designed to look like existing people Nishio et al. (2007), with the most famous one being Hiroshi Ishiguro's HI-1.

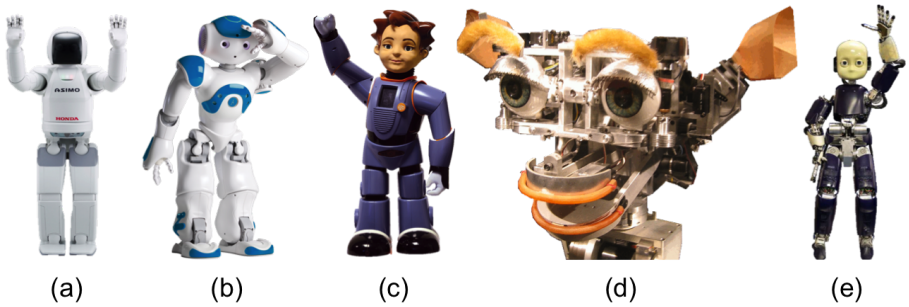


Figure 3.2: Popular humanoid robots in research and general public. From left to right: Honda's ASIMO (a), the Nao robot (b), the Zeno robot (c), Kismet (d) and the iCub (d).

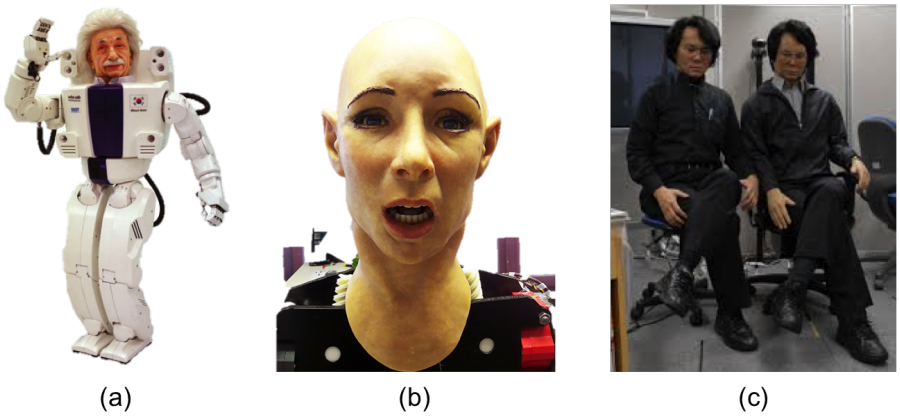


Figure 3.3: Examples of android and geminoid robots. From left to right: Albert HUBO (a), the F.A.C.E. robot expressing fear (b) and Hiroshi Ishiguro (left) sitting next to his HI-1 geminoid (right) (c).

Finally, zoomorphic robots have animal-like features (Figure 3.4) and are mainly used as companions Tzafestas (2015). The communication channels they employ do not directly match those of humans but more those of their biological counterparts. Nonetheless, they express social cues (like sounds, gaze or posture) that can be easily understood. Examples include the robotic seal Paro Shibata et al. (2001), the MiRo robot Collins et al. (2015); Mitchinson and Prescott (2016), Leonardo Breazeal et al. (2004), i-Cat van Breemen et al. (2005) and finally, Probo Goris et al. (2010, 2011).

Figure 3.4: Examples of zoomorphic robots. From left to right: the MiRo robot, designed to resemble domestic animals with animal-level social intelligence (a), the Paro robotic seal, widely used as a companion for both children and elderly people in clinics (b), Leonardo combines both anthropomorphic and animal-like characteristics used to study non-verbal communication (c), i-Cat, the expressive animal-like robot developed by Philips Research (d) and the Probo, an elephant-like robot mainly used as a companion for hospitalised children or children with autism (d).

All these examples illustrate the versatility of possible design strategies employed in HRI scenarios. So to answer the central question of choosing an appropriate robotic platform for social interactions with humans, we first need to examine whether morphology is important. The external design of a robot may strongly affect its believability and acceptability as well as its expressive capabilities. Studies suggest that indeed, the physical appearance of a robot biases the interaction, as it may affect user's perception Goetz et al. (2003) and expectations about its social capabilities Fong et al. (2003). For example, certain features, like eyes or hands, may imply that the robot can see or manipulate objects respectively; if the robot looks like an animal, users are likely to treat it as one. There are differences in the expected interaction when one is interacting with a robot that looks like a baby seal (e.g. the Paro, Figure 3.4 (b)) and when it looks closer to a human (e.g. the iCub, Figure 3.2 (d)). This rule mainly applies to the mental model people make regarding the robot during the interaction, especially if the robot's appearance is human-like. In general, the disposition is to design robots that either resemble humans or allow users to anthropomorphise them since anthropomorphism occurs naturally in humans. As Hume (1957) famously said: *"There is a universal tendency amongst mankind to conceive all beings*

like themselves and to transfer to every object those qualities with which they are familiarly acquainted... We find human faces in the moon, armies in the clouds; and by a natural propensity, if not corrected by experience and reflection, ascribe malice and good will to everything that hurts or pleases us". Thus, anthropomorphism can be applied to both the design as well as the behaviour of a robot.

Advocates of humanoid robots may claim that human-like design benefits HRI, as it enables communication channels that are similar to those of humans Duffy (2003). More specifically, humans spontaneously engage in social cognition when viewing complex social material and try to make sense of it. It seems that social features (such as human faces or bodies) are more salient compared to neutral scenes (like plants or scenery), as areas that are associated with social cognition (the dorsomedial prefrontal cortex and temporal poles) are activated Wagner et al. (2011). Indeed, humans are highly communicative beings; they use a variety of nonverbal multimodal cues to communicate, such as gestures, gaze, facial expressions, prosody or even the way they move or use proxemics Knapp et al. (2013). These rich nonverbal communication channels contribute to the enhancement of meaningful cues that can complement or even substitute spoken dialogue, like in the context of behaviour authoritativeness Johal et al. (2015). Humans employ these channels unconsciously and instinctively (e.g. when angry, they raise the tone of their voice) Knapp et al. (2013). However, most of the times, these cues are learned and are intentional, but overall, they are recognised almost automatically, without any formal training and allow for the regulation of the interaction (e.g. turn taking). In this view, the body is seen as a natural and fully functional interface for social interaction. Goffman (2008) argues that although no one can employ the body's whole expressive idiom, everyone is familiar with the body's vocabulary, emphasising that its properties are inherently used for communication: *"although an individual can stop talking, he cannot stop communicating through body idiom; he must say either the right thing or the wrong thing. He cannot say nothing."*. Hence, designing robots with human-like bodies allows humans to under-

stand and intuitively interpret the nonverbal communication channels these robots employ and communicate in a natural way Duffy (2003); Fong et al. (2003). Furthermore, the more human-like the artificial agent, the more life is attributed to it Looser and Wheatley (2010).

So should humanoid robots be preferred over non-humanoid robots, given that their body seems to facilitate interaction? There is no direct answer to that, as the preference of various designs of robots over humanoids is still a matter of debate and continuous research Wu et al. (2012). The advantages of human-like, as opposed to machine-like features, have been examined in a variety of studies. The majority of them highlights the benefits of human-like appearance as it is mostly preferred by users Walters et al. (2008) and may even evoke empathetic responses Riek et al. (2009) and score higher in communication DiSalvo et al. (2002) compared to mechanical-looking ones. Machine-like or human-like appearance may influence the perceived robot personality Walters et al. (2008) and responsibility assumed by humans in a collaborative task: humans feel more responsible when interacting with machine-like robots compared to more anthropomorphic ones Hinds et al. (2004). The willingness to cooperate with a robot also depends on the gesture types or abrupt (machine-like) compared to smooth (human-like) movements Riek et al. (2010). The role of humanoid (versus mechanical) appearance in combination with height (tall versus short) has been examined in Walters et al. (2009), showing how the participants' preferences toward robot appearance are powerful indicators of their likely responses. Nonetheless, there are cases where highly human-like robots have the exact opposite effects: they are perceived less trustworthy Mathur and Reichling (2016) and empathic Złotowski et al. (2016); Misselhorn (2009) compared to more machine-like robots claiming that they may fall into the uncanny valley effect (see section 3.1.2).

However, having a body with human features is not always preferable. It highly depends on the task and the people the agent interacts with. For example, children with autism prefer a robot with plain/robotic appearance compared to a more human-like one Robins et al. (2004) or simpler robots as

opposed to highly detailed or complex ones Robins et al. (2006). Similarly, in section 7.2.3, we show that children prefer more machine-like than human-like robots for interaction. In general, simple robots appear to be more engaging to children with autism, however, such robots may fail to be used in more generalised scenarios or with a different audience.

Morphology does play a role in the acceptance of a robot as a social partner, however, in many cases more complex interactions between robot and human factors have been found. More specifically, the robot's perceived gender and the user's gender have an impact on acceptance and interaction Nomura (2017); Otterbacher and Talias (2017), while other studies have emphasised the implications of age groups for the robot design and preference Wu et al. (2012); Cheng et al. (2017); Cameron et al. (2015b). What we can observe from the heterogeneity of these studies is that there are still no direct and useful guidelines regarding the desired robot design for a given interaction. In section 4.3 we propose a methodology that provides useful insights to this challenge.

3.1.2 Morphology and the Uncanny Valley

Most research on socially interactive robots has focused on humanoid robots, as studies suggest that the more human-like they look, the more inviting they become and humans positively respond to them. Nonetheless, there are cases in which a highly realistic robot may cause the opposite effect and fall in the so-called "uncanny valley". By definition, the word "uncanny" is used to describe something as *strange or mysterious, especially in an unsettling way* (Oxford dictionary). In some cases, it is used to describe a reproduction that is extremely close to the original, so much so that it causes surprise.

The "uncanny" as a term was first introduced by Freud in 1919 in his work "Das Unheimliche" (or un-homely) as something that is familiar and foreign at the same time; this, in turn, causes feelings of estrangement or dread. The problem with the uncanny according to Freud, is that cognitive conflict

results from seeing something familiar (and being attracted to it) and at the same time, feeling repulsed by it. An illustrative example would be a mannequin that looks familiar (given its human form) and at the same time may cause dread since it is not alive like humans. This cognitive conflict is what may lead to rejection.

Masahiro Mori in 1970 borrowed this term from Freud and defined the uncanny valley as the level of realism in robot appearance that causes negative emotional responses to humans Mori et al. (2012). Mori's hypothesis predicts that as robots move from the mechanical (or non-human) to human-like spectrum, people will find them more appealing and accept them more, compared to their mechanical counterparts. However, this positive relationship between robot appearance and affinity sharply becomes negative as robots start to closely resemble humans. This distinctive drop (or uncanny valley) (Figure 3.5) consequently leads to feelings of discomfort and unease. However, if the robot's appearance becomes indistinguishable from that of humans, this relation becomes positive again. Although the uncanny valley was initially conceptualised for robots, it seems that it may affect a variety of domains, including game or animation characters and life-like dolls.

The uncanny valley was criticised for being a logical prediction and not the result of empirical assessment; additionally, it cannot be used as a criterion to engineer better systems because it lacks guidelines for the operationalisation of human-likeness. In an attempt to empirically validate the uncanny valley, researchers have systematically manipulated human-likeness by using digitally morphed images Seyama and Nagayama (2007) or images of robots built to interact with humans Mathur and Reichling (2016). The majority of these studies examined users' reactions and perception based on images; very few have explored the uncanny valley when humans are interacting with physical robots Walters et al. (2008). Understandably, one cannot systematically manipulate the characteristics of a physical robot to the same extent as a picture, however direct comparisons between physical robots and images cannot be made. The existence or nonexistence of the

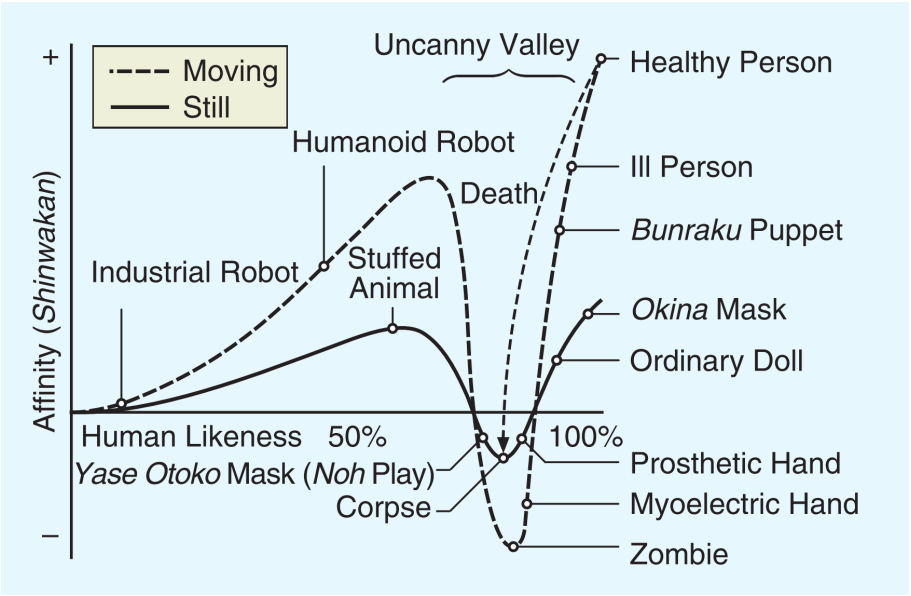


Figure 3.5: Graph of the "Uncanny Valley" as suggested by Masahiro Mori. The continuous line represents the perceiver's affinity for an entity in relation to the entity's human-like appearance. The dotted line represents affinity in relation to the entity's movement. Image adapted from Mori et al. (2012)

uncanny valley cannot be fully supported as studies that have challenged it usually manipulated one aspect of human-likeness Kätsyri et al. (2015). Nonetheless, it is widely used as a posthoc method to explain possible negative (or unwanted) results during interactions with robots.

Typically, the uncanny valley employs the distance of similarity between a machine and a human (i.e. human-likeness) and then measures how this distance affects feelings like eeriness, likeability or affinity. We argue that a redefinition of the uncanny valley is needed, as it may be misinterpreted. It may not be the level of realism or human likeness in robot appearance that causes this "negative" feeling humans experience, as many robots look nothing like humans and do not cause such feelings. For example, the Roomba robot does not classify as anthropomorphic; given the suggested curve of the uncanny valley, the Roomba should score relatively low in

affinity. In contrast, studies have shown that people not only feel happy with their Roombas but even form intimate relationships with them Sung et al. (2007). Similarly, visitors enjoyed the autonomous space Ada Eng et al. (2003) despite it not looking like a human.

Additionally, studies with children have shown that they do not prefer highly sophisticated and human-like robots but instead favour a combination of machine-like and human-like robots Woods (2006). In section 7.2 we present children's preferred robots and what their drawings indicate is an inclination toward machine-like than human-like design. Hanson (2006) suggested that the aesthetic features of the robot's design cause perceptual tensions, however his work was mainly based on humanoid robots. In contrast, Moore (2012) argued that these perceptual tensions might be caused by inconsistencies of individual features, like anomalous movements in the eyes of a very realistic humanoid robot. Moore's approach implies that the robot's behaviour (or individual features of its behaviour) affect the uncanny valley. To make it more precise, we argue that parameters like social competence or task competence could explain this phenomenon, as they may account for the psychological plausibility of an entity.

We propose that the suggested measurement of eeriness or familiarity does not account for the perceptual tensions created from the uncanny valley. It is possible that these tensions account more for psychological plausibility or acceptance rather than eeriness or affinity. For example, participants that interacted with the robotic toy Pleo treated it as if it were a real animal, however it soon failed to engage them Fernaeus et al. (2010), not because its appearance was not realistic enough and caused eeriness but because eventually, its behaviour was not believable or convincing. This is a problem that not only robots face. There are many domains that have similar issues. For example, pioneer Disney animators have conceived the "twelve principles of animation", that act as guidelines to create convincing behaviours regardless of the physical appearance of the animated object Thomas et al. (1995). Additionally, if we look at classical studies of biological motion of two-dimensional objects (like squares and circles), we can observe that hu-

mans attribute causality and intentional states to them if their behaviour is organised into a story that follows temporal contiguity and spatial proximity Scholl (2001); Heider and Simmel (1944). We argue that given the original definition of the uncanny valley, such observations should not occur as simple non-human objects would score low on affinity, yet participants accepted their behaviour and found them convincing and plausible.

To conclude, we do not completely disregard the uncanny valley, nor we claim that we have systematically examined its effects. However, we propose a more critical view on the subject as the initial definition of the term might be too simplistic and not applicable to a number of examples. In the next chapter, we explain in detail a proposed taxonomy that aims at evaluating the psychological believability of an agent, so that the agent will be accepted by humans. We argue that instead of looking at the uncanny valley as a relationship between the robot's morphology and human affiliation, we can redefine it as the relation between an entity's behaviour and user acceptance. This way, our proposed taxonomy of autonomy, morphology, task and social competence can be used to systematically and empirically evaluate the uncanny valley, not based on feelings of familiarity but instead, acceptance and believability.

3.2 Social skills and anthropomorphic behaviour

Robots are now able to interact with humans in various conditions and situations. Given the current technological advancements, we can develop robotic systems that can deal with both the physical and the social world. When it comes to robots that socially interact with humans, almost all aspects of the robot have been found to affect the interaction. As we saw in the previous section, appearance plays an important role. However, it is not the sole factor; the behaviour of the robot systematically influences humans' perception and expectations Kidd and Breazeal (2008); Wada and Shibata (2006a); Saerbeck et al. (2010); Kanda et al. (2004b); Sabelli et al. (2011). Thus, one of the greatest challenges in the design of social robots is to correctly identify and consider the various factors that affect social interaction Fong et al. (2003); Goodrich and Schultz (2007); Scassellati (2005).

To answer the question "What are the behavioural traits that allow humans to perceive the robot as a believable agent?" we first examine existing approaches that are typically employed in this domain. Research suggests that to be accepted by humans as communication partners, autonomous and transparent behaviours are essential, as they can be easily understood and explained. Many take the anthropomorphic route arguing that behaviours that resemble those of humans provide a more intuitive interface: if they fulfil the social expectations, they become predictable and interpretable Breazeal (2003b); Fong et al. (2003); Leite et al. (2013); Thrun (2004); Dautenhahn et al. (2005); Duffy (2003). Based on this approach, robots are bound to the social standards of human-human communication and the social rules attached to the role they assume. It seems that humans intuitively apply the same social rules when they interact with machines as when they interact with other humans Reeves and Nass (1996): they may be polite to machines and even use the same vocabulary of human psychology to describe them Nass et al. (1995). Although the social rules that will be triggered are not yet well defined, Nass et al. (1995) suggest that behaviours that are elicited by primitive or automatic processes (e.g. smile back when one is smiling) are more likely to be triggered, compared to behaviours that

are socially constructed (e.g. find a joke funny and laugh). Additionally, rules that are more frequently used (like politeness in a conversation) are more likely to be automatically elicited, as opposed to rules that are less frequently used (e.g. how to behave in an awkward situation). To conclude, people may treat machines as peers, using the same social roles, as this is the only model they have when dealing with intelligent and intentional agents. Hence, if there is a social model that humans can attribute to the robot's behaviour, the robot can be considered socially competent Breazeal (2004); Reeves and Nass (1996) and more enjoyable, as its behaviour will fall within human expectations. However, such operationalisation of social competence seems to exclude both the mechanisms that underlie such competence as well as a broader range of non-human social behaviours like the usage of coloured light to display affect Collins et al. (2015). This leads us to study the challenges that rise from implementing the main principles of social reproduction in socially interactive robots.

Given that many features affect the interaction and user perception ranging from gazing models Admoni and Scassellati (2017); Lallée et al. (2013); Boucher et al. (2012), non-verbal communication Breazeal et al. (2005); Kennedy et al. (2017), personality Cameron et al. (2016), facial expressions Cameron et al. (2015a) to the complexity of the behaviour Kidd and Breazeal (2008); Wada and Shibata (2006a); Breazeal et al. (2005); Vouloutsi et al. (2014); Kennedy et al. (2015a). Hence, the most common social characteristics that robots exhibit and are found to affect social interaction are: the expression and perception of emotions and personality, the establishment and maintenance of social relationships, the usage of natural cues (like gaze or gestures), communication using high-level dialogue and the exhibition of motivated behaviour. In this section, we explore the behavioural traits that we have identified as important for the scope of this thesis and present them with further details.

3.2.1 Motivation

Anyone who is interested in understanding, influencing or even mimicking biological behaviour (and therefore implementing it in a robot) has to start with understanding motivation. The implementation of biological behaviour to robots can enhance their believability (and therefore acceptance), as their actions may fall within human expectation. The study of motivation tries to answer the fundamental questions: "Why do animals behave the way they do under different (or the same) conditions?" and "What is the "primal force" that guides behaviour?" According to Huitt (2001), motivation is the "process that energises, directs and maintains goal-oriented behaviour" or that what causes us to act generating the so-called "why" of behaviour Verschure (2012). In the framework presented here, emotion is seen as predicated on the state of the motivational system.

The study of motivation is a rich field and a number of different theories have been proposed Graham and Weiner (1996). In the early 20th century the psychologist William McDougall coined the notion of *Instinct theory*, which has its roots in evolution theory and postulates that organisms behave in certain ways because they are biologically determined to do so. This notion also influenced the psychoanalysis of Sigmund Freud and the ethology of Lorenz. Although this theory can certainly describe animal behaviour, such as the famous imprinting experiments of Lorenz, it fails to explain them in terms of underlying processes. The *arousal theory* of motivation Lindsley (1951) suggests that organisms behave in a certain way to maintain an optimal level of arousal that varies, depending on the properties of the individual or the situation. The *incentive theory* hypothesises that animals act in certain ways because of external rewards or punishments, or incentives. Incentives can be primary (not learned) and secondary reinforcers (they become rewards after being associated with other primary incentives). According to this theory, animals act because they strive toward goals driven by reward seeking or hedonism, a perspective that has also informed many models of machine learning. This theory mainly focuses on associating and learning to control motivation and through that, behaviour. The humanistic

theory of motivation, mainly represented by Maslow (1943)'s *hierarchy of needs*, suggests that needs can be categorised in a hierarchical way, ranging from basic survival and biological needs to self-actualisation, where higher needs cannot be pursued if lower ones are not satisfied. Finally, the *drive theory* (also known as drive reduction theory) posits that an organism's unsatisfied need is the source for motivation: actions satisfy needs Hull (1943). Based on this theory, all organisms need to be in a state of balance. Changes in the environment can cause imbalance, which in turn create a state of arousal and unpleasant feeling or tension called *drive*. Once out of balance, the organism will try to engage in behaviours that reduce this drive (hence the name drive reduction). According to this theory, there are two main drives: primary, which reflect biological needs and secondary, which are learned drives. We can see that in general, motivation theories allude to the ability of the organism to maintain a "steady state". Already Hippocrates (350 BC) equated health with the harmony between mind and body, while the 19th century French physiologist Claude Bernard spoke of the organism maintaining its internal environment, or "milieu" in balance, while facing a fluctuating external environment. The Russian physiologist Ivan Pavlov generalised this notion of stasis to the relation between an organism and the external environment. A tightly coupled concept to the maintenance of an organism's internal environment is *homeostasis*.

Homeostasis and Allostasis

According to Cannon (1932), the coordinated physiological process that maintains a steady state of the organism can be called *homeostasis*. Homeostasis refers to the control of physiological processes, with the aim to keep them within certain bounds, using negative feedback. For example, a sensor detects a state of the system which is compared to a reference value; a control signal is generated proportional to the difference, which in turn drives cells, tissue, organs or the whole organism to reduce the detected discrepancy. Hence, homeostasis (or "identical state") is the self-regulation of a dynamical system towards constancy. Cannon focused on five homeostatic

processes critical to the biochemistry of life involving the essential variables of pH, temperature, plasma osmolality, glucose, and calcium. However, other physiological processes are believed to follow similar principles. Homeostasis is essentially based on a predefined reactive negative feedback system which precludes the inclusion of anticipation and learning. In addition, it raises the question of scalability when many partially conflicting essential variables must be regulated. As a result, the complementary notion of *allostasis* has been advanced as achieving stability through change, in particular by changing the boundaries within which essential variables are held through learning and anticipation Sterling and Eyer (1988). For instance, a glucose deficit might be tolerated in order to evade a predator. The cost to the organism of maintaining an essential variable from its set point is called the allostatic load. Whereas homeostatic processes are independent and autonomous, in case of allostasis, auto-regulation depends on a central control system, i.e. the brain.

Application of motivational systems in robots

It has been argued that motivational systems allow a system to be self-sufficient and autonomous Cañamero (1997). Additionally, they not only allow the robot to successfully complete a task, but also focus on a predefined goal, facilitating the interaction with humans Stoytchev and Arkin (2004). Inspired by the drive motivation theory or by ethology, robotic systems are endowed with homeostatic mechanisms. Typically, each drive needs to be maintained in balance and their action selection mechanisms aim at achieving that Breazeal and Brooks (2005); Breazeal (2003a); Cao et al. (2014); Castro-González et al. (2013); Vouloutsi et al. (2013a). The first robot control model that explicitly brought together associative learning of sensorimotor mappings with motivations and emotions mapped a model of classical conditioning to foraging robots Verschure et al. (1992). Here, simple stimuli such as collisions and rewards triggered internal states of negative and positive valence respectively, which in turn triggered avoidance or approach actions. This appraisal in turn gated the epistemic learning

process, such that neutral stimuli predictive of these simple ones could be associated with the same behaviours. This model has been explicitly linked to utilitarian emotions in an avant-garde human accessible artificial organism "ADA: the sentient space" Eng et al. (2005); Wassermann et al. (2003). ADA's main goal was to server a maximal and consistent interaction with its visitors. This was achieved by linking its motivational and emotional system: ADA as an artificial organism was maximising its own goal functions (or maximise "happiness") by keeping drives in homeostasis. At the same time, it was communicating this process through externalising its utilitarian emotions influencing the behaviour of the visitors in a way to reduce its drives. This generation of behaviour by linking an agent's emotional and motivational system was also explored by Arkin et al. (2003).

The main challenge in the implementation of motivational systems is the resolution of conflicts between the various drives or components Stoytchev and Arkin (2004). A solution to this problem is coupling the homeostatic system with an allostatic one. By combining homeostatic and allostatic levels of control, animals can perform complex real-world tasks like foraging, regulating their internal states and maintaining a dynamic stability with their environment. Such systems have been implemented in robots performing foraging tasks Sanchez-Fibla et al. (2010); Fibla et al. (2010). Additionally, the interaction of homeostasis and allostasis has been extended to behavioural control in robot models of foraging and as well as in Human Robot Interaction (HRI) scenarios Lallée et al. (2014); Vouloutsi et al. (2013a). Here, each drive is influenced by its homeostatic state; adaptation is achieved through allostasis, as the homeostatic limits are adjusted dependent on overall demands on the system.

To conclude, we argue that motivational systems contribute to the autonomy and task competence of the robot by allowing for the production of plausible and robust behaviours, facilitating adaptation and allowing the robot to stay focused on its task.

3.2.2 Emotions

Emotions are complex and structured phenomena that play an important role in human behaviour and interaction. Their functions include triggering motivated behaviours, coordinating behavioural responses, affecting memory storage and retrieval, communication, social bonding and are crucial for survival Rolls (2000); Fellous (2004); Keltner and Gross (1999); Levenson (1999); LeDoux (2012); Parkinson (1996). Here, we distinguish the role of emotions in two broad categories: *epistemic* and *utilitarian*, the first to inform action and organise behaviour and the latter express the organism's internal state, communicate and coordinate socially.

On the one hand, we look at the evaluative role of emotions or appraisal Frijda (1986), which is predicated on motivation. Here, the latter sets the context of the former Verschure (2012) i.e. whether food will trigger happiness depends on whether the consumer is satiated or hungry. On the other hand, the outcome of emotional appraisal can inform internal processing such as learning and memory and/or define communicative signals. Additionally, emotions have social functions: they modify interactions, ensure the social transmission of emotional interpretations of events and are influenced by the social environment Frijda and Mesquita (1994). This way we can distinguish between epistemic and utilitarian emotions. In this perspective, emotions play a much deeper role in the organisation of individual and social behaviour, than solely as a cue system.

Emotions are associated with feelings and moods and have phenomenal aspects; they are experiences with distinct intensities and qualities. Emotions are seen as being transient and directed towards someone or something, while *moods* are *feelings* that last longer, are less intense and often lack immediate triggering stimuli. Emotions can both activate and direct behaviour as in fear and anger. Although emotions mostly accompany motivated behaviours, they are fundamentally different in the way they are triggered: while motivations are dependent on internal needs, emotions can be elicited by a variety of external stimuli in the absence of pre-existing needs and

goals. Emotion is a complex episode that creates a readiness to act and has several different components Frijda (1988). It usually begins with an appraisal of a given situation or stimulus, i.e. the interpretation of the situation or stimulus relevant to needs, goals and/or well-being. This appraisal differentiates various emotions and leads to its distinct quality. Other components of emotions include associated thoughts and action tendencies as well as bodily reactions usually accompanied by facial expressions and finally, responses that aim at coping or reacting to the emotional state. None of these components in itself can be seen as an emotion, rather, emotion is complex and the result of the interplay of all of them.

To this day, emotion is a controversial subject, as there is no general agreement on its definition or its underlying processes. In fact, there are so many different interpretations regarding emotions and their properties that a broad definition is needed to include their most significant aspects. Despite the lack of consensus, emotions are usually responses to events that are relevant for the individual Frijda (1988). For some, emotions are the result of somatic responses to affective stimuli as postulated in the classical James-Lange theory. This approach was heavily criticised by Walter Cannon and Philip Bard, as the experience of emotions seems to precede the occurrence of bodily changes and can be seen as the result of a simultaneous activation of physiological responses and identification of emotional cues from sensory information. Damasio in his popular book "Descartes Error" has revived the James-Lange body-centred idea of emotion seeing that emotions are anchored in somatic markers, but recently the author has changed his mind in the face of neuroscientific evidence, and the current status of the James-Lange theory is again under debate. Recent studies do confirm the link between bodily reactions and emotions, as the first seem to affect the latter. However, bodily reactions do not appear to be the cause of emotions. While emotions do not directly derive from somatic responses, they are seen as being linked to them.

Other theorists support the notion that emotions are an experience, subject to motivational situations placed in an approach-avoid continuum: be-

haviour is oriented towards or away from a stimulus and emotion is, therefore, any experience with a high intensity and hedonicity Cabanac (2002) that will make an animal work towards or away from a stimulus Rolls (2000). Another fundamental question on emotions is whether they are discrete or continuous and universal of contingent on local contexts. Paul Ekman defined six basic discrete emotions that can be found in most cultures and can be considered primitive and universal: anger, disgust, fear, happiness, sadness and surprise Ekman (1992). These discrete emotions are believed to be innate and fundamentally different from other more complex emotions since they can be distinguishably expressed, and exhibit different behavioural, physiological and neural reactions Colombetti (2009). On the other hand, continuous emotions are defined by one or more dimensions such as valence/arousal, where the former defines the quality on a positive (happy) to negative (sad) dimension, and the latter defines the intensity Rolls (2000).

Recently, the neural mechanisms underlying emotions have gained increasing attention in the scientific community emphasising the role of the amygdala found in the medial temporal lobe LeDoux (2012, 2000); Scherer (1993). Indeed, the amygdala can be seen as a generic valence assignment system, which mediates between primitive behavioural control systems of the brainstem and mid brain and the perceptual and cognitive systems of the neocortex. Another influential proposal is that by Jaak Panksepp that defines seven basic emotional systems found in the mid brain/brain stem: CARE, FEAR, LUST, RAGE, PANIC, PLAY and SEEKING Panksepp and Biven (2011). These systems underly the full spectrum of emotions and are linked to the regulation of adaptive behaviour. Conversely, Craig (2009) emphasised the anterior insular cortex as the structure where a broad range of subjective states are represented including the feelings associated with simple and complex emotions. This suggests that emotions are dependent on a broad hierarchy of systems from the brainstem to the frontal cortex and should be considered in terms of the architecture of the brain as opposed to a singular module Verschure (2012).

Despite the heterogeneity of the definitions, some invariants stand out where emotion involves some appraisal, regardless of whether something is beneficial or bad, rewarding or punishing or even something one would work for or avoid. This has also been proposed as a form of emotional learning LeDoux (2012). Invertebrates and vertebrates need to learn from their environment, in real-time, to survive. For instance, in classical conditioning, the behavioural signature of emotional learning has been observed in c-elegans, sea slugs and moths. Survival not only requires the identification which of the stimuli in the environment are relevant for behaviour, i.e. appraisal but also how to modify behaviour accordingly, action preparation and shaping. The way the brain develops representations of such stimuli and their associated actions has been the subject of the study of classical and operant conditioning LeDoux (2012).

Learning in both cases depends on motivating stimuli, e.g. food or shocks, that trigger mechanisms gating learning and memory. Gating utilises neuromodulatory systems originating in subcortical structures such as the ventral tegmental area, the nucleus basalis of Meynert or the Locus Coereleus. Computationally, one can interpret these systems as issuing a "print now" signal that regulates synaptic plasticity, allowing local learning rules to be controlled by global mechanisms Sánchez-Montañés et al. (2002). Indeed, this principle is mirrored in many machine-learning approaches. For example, in a model of classical auditory conditioning, the amygdala provides emotional appraisal which drives the nucleus basalis of Meynert (NBM), which facilitates learning in the primary auditory cortex, remodelling the receptive fields to detect better the tone that predicts a shock. This model is consistent with the physiology of learning in A1 and demonstrates robust tonotopic map formation and adjustment, even in the presence of noise or inhomogeneities in stimulus sampling Sánchez-Montañés et al. (2002). This example illustrates an epistemic impact, i.e. remodelling of A1 representations of tones, dependent on stimulus appraisal realised by the amygdala driven by a motivating stimulus. This illustrates the latest trend in research where emotions are considered from a system's perspective, as they

can alter perception, motivational priorities, learning, attention, memory and decision-making Verschure (2012); Dalgleish (2004). Thus, emotions can be instrumental in assisting communication, by expressing one's internal state (external/utilitarian emotions) and for organising perception, cognition and behaviours (internal/epistemic).

The implementation of emotional systems to robots has been endorsed by many studies as their functional roles are linked to intelligence, adaptation and allow for the organisation of behaviour and communication Arbib and Fellous (2004); Breazeal and Brooks (2005); Fellous (2004); Cañamero and Gaussier (2005). Additionally they provide a framework to evaluate and understand human emotional models Cañamero (2014). Emotional systems allow for adaptation Parisi and Petrosino (2010) and affect the decision-making process of the robot Cominelli et al. (2015). We argue that an emotional system affects the social competence, task competence and autonomy of the robot. More specifically, the epistemic role of emotions contributes to autonomy and task competence, by allowing the robot to appraise a stimulus or a situation and act accordingly. At the same time, appropriate actions may fall within human expectations which in turn can affect plausibility. Finally, the utilitarian role of emotions benefits the interaction, as they can be used to modulate the communication channels employed by the robotbr and contribute to it social competence. Here, the transparency of the expression heavily depends on the morphology of the robot, suggesting an interaction between the components of our proposed taxonomy.

3.2.3 Empathy

In general, empathy is considered the ability to take the role of another and understand the other's emotional state, or more specifically, the "affective response more appropriate to another's situation as one's own" Hoffman (2001). Its functional role lies on the survival of the species, as it motivates us to take care of each other. Hence, it allows us to predict and understand the behaviours of other agents and act accordingly, contributing to the social interaction. The idea that empathy is not a human trait starts to gain

ground, as research suggests that presumably simpler empathic responses can also be attributed to other animals de Waal (2007); Panksepp (2011); de Waal (2012) such as primates and rodents Bartal et al. (2011); Grenier and Lüthi (2010) and it has been linked to prosocial behaviours Eisenberg and Miller (1987).

Empathy can be divided into two broad categories: *cognitive* and *affective*, the first being the understanding of another's emotions and the latter being the possession of that emotion D'Ambrosio et al. (2009); de Vignemont and Singer (2006). However the elicitation of empathic responses is not automatic: it depends on a series of factors that include the characteristics of the empathiser, the object of empathy, the social context, as well as the emotional states of others Engen and Singer (2013). Nonetheless, humans can empathise with animals as well as inanimate objects Misselhorn (2009).

Exploring the effects of empathy on HRI scenarios may offer useful insights into the design and characteristics of robots. In social robotics, two main lines of research exist: on the one hand, robots are endowed with empathic models, and their empathic capabilities are assessed in interaction scenarios with humans Leite et al. (2012). On the other hand, the human reactions that assess the elicitation of empathic responses towards robots have been examined. In this thesis, we focus on the latter. Hence the fundamental question that arises is: "Can humans empathise with robots?" and if so, "What are the behavioural characteristics that affect the elicitation of empathic responses?". To answer the first question, a plethora of studies where humans are typically presented with ethical dilemmas (e.g. hurting, switching off a robot to being unfair to it) seem to gain ground. Neurological studies showed that similar neural activation patterns are found when participants were presented with a video of a robot, a human and an object being treated in a violent way Rosenthal-von der Pütten et al. (2013b), suggesting that indeed, humans can empathise with robots.

To answer the second question, researchers modulated the robot's characteristics (ranging from robot's design to behaviour) and evaluated the

reactions of humans. For example, the appearance of the robot seems to affect the empathic responses, as the more human-like the robot looks, the more empathy people felt toward it Riek et al. (2009). Age plays an important role in the elicitation of empathic responses, as children and adults have differences in the way they form moral relations Kahn Jr et al. (2012). Humans can even empathise with non-anthropomorphic robots, especially if they are perceived as intelligent Bartneck et al. (2007b). Background stories Darling et al. (2015), agency Kwak et al. (2013) and personality Bartneck et al. (2007a); Briggs and Scheutz (2012) seem to also positively influence humans' empathic responses. For example, the more intelligent and agreeable a robot is viewed, the more hesitant participants were to turn it off Bartneck et al. (2007a).

Empathic responses may also be triggered when a robot displays signs of protest and distress Briggs and Scheutz (2012), or displays emotional expressions Kim et al. (2009a). Additionally, humans prefer robots that exhibit congruent empathic behaviours Cramer et al. (2010a) and the expression of empathic responses affects the robot's social skills, as the appropriate responses are context-related. To assess whether humans treat robots like humans or machines, several studies have recreated the Milgram experiment (see section 6.4.1) and substituted the learner with a robot or an avatar Bartneck et al. (2005); Gou et al. (2014); Slater et al. (2006); Rosalia et al. (2005). It seems that although in most cases participants showed compassion toward the robot, they tended to administer higher shock voltages to the robot compared to a human Bartneck et al. (2005). In most of the cases, empathic responses are linked with the perception that some life, animacy or intelligence is attributed to the robot: the more lifelike a robot's appearance or behaviour is, the easier it is for people to accept it. We, therefore, argue that the elicitation of empathic responses can be used as a measurement for the evaluation of the psychological validity of the robot.

Chapter 4

A taxonomy for robot acceptance

The main aim of this thesis is to create a robotic agent that is accepted by people. Acceptance is important because it determines whether potential users will use the robot and interact with it on a frequent basis, or even introduce it into their homes. A fundamental question thus arises: "How can we create robots that are accepted by people?". To answer this question, one must comprehend how acceptance can be measured and what are the determinants that affect it. We borrow one approach from research in acceptance of the technology. More specifically, the Technology Acceptance Model (TAM) classifies two predictors: "Perceived Usefulness" and "Perceived Ease of Use" Davis (1989). The Unified Theory of Acceptance and Use of Technology (UTAUT) postulates that technology acceptance can be determined by: "Performance Expectancy", "Effort Expectancy", "Social Influence' and "Facilitating Conditions" Venkatesh et al. (2003). It is possible that the same predictors can generalise to robots, however, what these utilitarian approaches lack is the social aspect of robots. Factors like "Social Presence" and "Sociability" play a significant role in acceptance Shin and Choo (2011). Thus, efforts to model robot acceptance using utilitarian, social and hedonic constructs (and the interactions between them) have

been made. For example, Heerink et al. (2008) modelled acceptance for the elderly including variables like "Perceived Enjoyment", "Perceived Sociability" and "Social Presence". Additionally, De Graaf and Allouch (2013) explored nineteen different variables ranging from "Actual Use", "Usefulness" to "Enjoyment", "Anthropomorphism" and "Social Influence".

Though the definitions above seem helpful, we take a different approach. While they were asking questions regarding existing interactions, we aim to explore the various behavioural traits that possibly affect acceptance. We propose that a key determinant of acceptance is the psychological plausibility of the robot. The motivation behind psychological plausibility lies in the fact that humans are active modellers of the world; they make predictions and have expectations based on those predictions. We argue that to be accepted, one must match those expectations.

The idea that the brain is a predictive mechanism goes back to Hermann von Helmholtz who postulated that the brain uses internal models of the world and its body and generates sensory data to match the incoming ones. The generation of sensory data produces several hypotheses about the world, and the most probable hypothesis becomes a perception. More specifically, to deal with the dynamic world, the brain makes predictions and learns from its mistakes, or what is called *prediction error*. The brain is viewed as an active predictive Bayesian mechanism that is hierarchically organised and continuously tries to match the bottom-up sensory inputs with the top-down predictions Friston (2010); Verschure (2012, 2016). To do so, it tries to reduce surprise, or what Clark (2013) calls "Predictive Processing" models. Any deviations between the predicted and sensed stimuli create prediction errors of various levels of uncertainty, and the goal of the brain is to minimise this uncertainty. Given the fact that prediction errors are mediated by attention, one can decrease or amplify them Friston (2010); Clark (2015). Hence, an action is the mechanism that is used to reduce a prediction error, emphasising their bi-directional link: the brain updates its predictions to fit the world and at the same time, through action, alters the world to fit its predictions.

Indeed, the brain's predictive processes have been verified by recent advances in neuroscience, as empiric studies support the idea that the brain not only makes predictions about the world but also that this process is hierarchically organised. For example, the brain's ability to predict motion can be observed in the visual cortex (V1) Ekman et al. (2017), by completing a moving pattern even if only a subset of it is provided. According to Chennu et al. (2013), the brain constantly updates a set of beliefs (or predictions) relevant to events in the world, and this information is modulated by attention and prior expectations (or predictions) regarding future events. This study shows that the hierarchical organisation of predictions is mainly implemented in temporal and frontal regions of the brain; what allows the brain to learn is the ongoing process of identifying mismatches between the predicted and actual sensory inputs (what is called prediction error) and updating its internal models accordingly. Thus, the observer is viewed as a hypothesis-driven system that makes inferences; its perceptions are shaped by both the incoming sensory inputs and past experiences. Any violation of the observer's hypotheses leads to the possible rejection of an observation and its believability.

To explain and predict behaviour, humans apply social models not only to living organisms but also non-living entities of sufficient complexity. It seems paradoxical that humans so easily come to social inferences - "the attribution of mental states is to humans what echolocation is to bats" (Dan Sperber quoted in Gallagher (2005)p. 207). Indeed, a large body of work has shown the propensity of individuals to make social judgments, even interpreting the movement of geometrical figures on a display as actions of animate beings Heider (1944). People attribute causality and intentional states on events or other objects Scholl (2001), provided that they can be organised into a story and follow temporal contiguity and spatial proximity Heider and Simmel (1944) to explain their behaviour Premack and Premack (1995). According to Michotte (1963), this "phenomenal causality" can be caused even by simple motion cues and can be considered the foundation of social perception. It seems hard for people to explain behaviour in a non-

mentalistic or non-intentional framework (Baron-Cohen, 1997, p. 3). The attribution of mental states is almost a natural and automatic way to both understand a social environment filled with agents and intentionality as well as predict behaviour rapidly. Hence, the brain developed the capacity to predict the states of other agents based on their actions. To predict the hidden states of the agents, an organism must develop a capability for "Theory of Mind"(ToM), that is the ability to attribute mental states to self and others Goldman et al. (2012). Humans can attribute beliefs, goals, percepts and mental states that are not directly observed to others Premack and Woodruff (1978), to explain their behaviour Frith and Frith (2005). This ability is also called "mentalising" Frith and Frith (1999) or "mindread" Baron-Cohen (1997). By attributing a Theory of Mind to others, humans can explain and predict the other's desires and behaviours Gallagher and Frith (2003). Given the fact that humans make inferences and predictions about the world and can attribute Theory of Mind to others to explain and predict their behaviour, we argue that if the robot's behaviour matches the expectations of its users or they attribute Theory of Mind to it, then it can be considered as a psychologically plausible agent and be accepted.

We view psychological plausibility from two perspectives: *principles* and *implementation*. Our goal is to decompose psychological plausibility to discrete parts to understand the variables that affect acceptance and test them empirically, to ultimately use their interactions in practice for the meaningful design and development of social robots. Although several studies have been made that examine humans' responses and establish common metrics in a variety of domains, like task-oriented Human-Robot Interaction (HRI), socially assistive robotics e.t.c. Steinfeld et al. (2006); Kahn et al. (2008); Kahn Jr et al. (2010); Kahn et al. (2010); Feil-Seifer et al. (2007); Kuo et al. (2012); Wisspeintner et al. (2009), fewer endeavours have been made to establish psychological benchmarks Kahn et al. (2006); Sun and Sundar (2016). Therefore, we aim to identify and explore the behavioural traits that facilitate humans to perceive a robot as a believable agent. To do so,

we offer four benchmarks for consideration: *morphology*, *autonomy*, *social competence* and *task competence*. Eventually, we propose an architecture to control the robot's behaviour while allowing benchmark testing. One could argue that the proposed taxonomy is far from complete, as the suggested benchmarks are only a few of many that can be added. However, we believe that to construct believable agents, one can view HRI as a problem of engineering human social perception, or defining a mini psychological engine that will allow robots to be accepted by users. User perception is in principle hypothesis testing, and therefore an agent can be accepted if its behaviour matches user's expectations. Hence, our taxonomy provides useful insights regarding the establishment of future assessments for HRI.

4.1 Social competence

Social competence as taxonomy is inherently more complex compared to the rest of the benchmarks because the robot's social success can be measured in a variety of ways. To assess the social role of the robot, the most fundamental question on could as is: "Does the robot successfully assume its intended social role?". The evaluation of the robot's social role is then quite straightforward: if the robot is created to be a social partner, do users interact with it as if it were a peer? Similarly, if a robot is meant to be friendly, do humans perceive it that way?

A second and equally important question we want to answer is: "In what ways do the various social components of the robot affect its psychological plausibility?". Here, answering this question becomes slightly more complicated, as now a number of derived questions arise. For example, which are the social components that affect psychological plausibility or what is the interaction between them. In this respect, we can decompose social competence into discrete parts and evaluate how they individually or in conjunction affect human acceptance. As we presented in the previous chapter, there are a number of factors that have implications for the interaction, like the expression and perception of emotions and personality, the

usage of natural cues and other communication modalities or the display of motivated behaviour, to name a few. The list is exhaustive, and pretty much everything has been found to affect the interaction between humans and robots, however, here, we have identified social components that are relevant not only for the social competence of the robot but also for the task competence, morphology and autonomy. These components include gazing behaviours, the expression of the robot's internal states and the elicitation of motivated behaviour.

We propose that a way to measure psychological plausibility is by examining the elicitation of empathic responses toward the robot. In section 3.2.3 we have presented empathy, that is considered the ability to understand the other's emotional state and respond in a way that is more appropriate to another's situation as one's own. Studies suggest that humans can empathise with robots Lallée et al. (2015); Briggs and Scheutz (2012) and the morphology and behaviour of the robot affect the elicitation of empathic responses. More specifically, the perceived intelligence Bartneck et al. (2007b), congruent responses Cramer et al. (2010b), emotional expressions Kim et al. (2009b) or background story Darling et al. (2015) of the robot seem to affect user's responses that could be linked to empathic responses. Indeed, event-related brain potentials that are relevant for the elicitation of empathic responses were similar between humanoid robots and humans Suzuki et al. (2015). Similar affective responses can be elicited when one is viewing a non-anthropomorphic robot, a human or a box being treated in a positive or negative way Rosenthal-Von Der Pütten et al. (2014).

A factor that affects empathic capacity is the eye-gaze pattern, as higher empathic responses can motivate individuals to look for affective, social cues primarily from the eyes of other faces Cowan et al. (2014). Gaze accounts for higher accuracy in the recognition of the expression Bauser et al. (2012) and higher intensity ratings of the perceived emotions Schulte-Rüther et al. (2007) for directed faces. Additionally, the empathic capacity of an individual and the modality of the observed stimulus affect the strength of facial mimicry Rymarczyk et al. (2016). More specifically, according to the sim-

ulation theory, an expression of an emotional state triggers to some degree the other person's feelings, like responses to pain should trigger to some degree that feeling. Though this assumption has been mainly tested with responses to pain Botvinick et al. (2005), other emotions like sadness Harrison et al. (2006) or happiness Hennenlotter et al. (2005) trigger similar feelings. Hence, we use empathy as a way to validate the psychological plausibility of the robot, by manipulating two parameters: the gaze model and emotion expression.

4.2 Task competence

Competence is the ability to use a set of related skills, abilities and knowledge or experience to successfully perform important functions. In the domain of management, effective competence is assessed by evaluating the output of a task through specific actions while at the same time being consistent with the organisation's procedures Boyatzis (1982). With this respect, competency is linked to both the desired output (or goal) or the facilitation of the task's purpose. Some tasks are easy to evaluate, like the monthly sales of a salesman, while others are more complex, like the evaluation of a manager. The main problem is to identify, understand and measure a number of factors that may affect performance and therefore task competence.

Here, task competence refers to the robot's capability to successfully perform a certain task, and we can decompose it into components that evaluate various aspects of a task. For example, if a robot is used as a social companion, do users play or interact it? If a robot's task is to tutor the user, does the user learn from it (goal)? Also, does the robot provide useful feedback and engage and motivate the learner (facilitation of the task's purpose)?

Another important factor that may affect the competence of the task is the interaction between the various social features of the robot with its task competence. For example, do certain social features (like the expression of emotions and gaze) affect the task competence of the robot? To evalu-

ate task competence, we use tutoring as the validation domain, where the robot's goal is to guide the learning through an educational task. For example, the personality of the robot Joosse et al. (2013), the gaze model Boucher et al. (2012), non-verbal Kennedy et al. (2017) or verbal Kennedy et al. (2016) behaviour can affect the executed task or the learning acquisition.

4.3 Morphology

As we presented in the previous chapter, the appearance of the robot affects the way users perceive it. To validate the psychological plausibility of the robot based on the morphology benchmark, we first propose that for a robot to be accepted, its design should serve the task it was meant to execute. To be more precise, if the robot's purpose is to clean the house, it should be equipped with wheels or legs that would allow it to move around. Similarly, it needs to have the appropriate size and shape for the target group that is going to use it. For example, a rather big or heavy robot may not be the optimal option if it is meant to be used by the elderly as a companion Moyle et al. (2016).

Additionally, its appearance should facilitate the interaction by communicating clear and readable cues, without necessarily implying a humanlike form. Advocates of humanoid robots may support human-like design as it enables communication channels that are similar to those of humans. The reason behind this is that humans spontaneously engage in social cognition when viewing complex social material and try to make sense of it. It seems that social features (such as human faces or bodies) are more salient compared to neutral scenes (like plants or scenery), as areas that are associated with social cognition (the dorsomedial prefrontal cortex and temporal poles) are activated Wagner et al. (2011). The degree of human-likeness of a partner (ranging from a computer, a functional robot, a humanoid or a human) is correlated with cortical activity of the medial frontal cortex and right temporo-parietal junction Krach et al. (2008). These areas are

associated with "Theory of Mind": the automatic process of attributing mental states (such as beliefs, desires, intentions or emotions) to another agent in order to predict behaviour Premack and Woodruff (1978); Leslie (1987); Gallagher and Frith (2003). Furthermore, the more human-like the artificial agent, the more life is attributed to it Looser and Wheatley (2010). These findings suggest that if robots fulfil the corresponding human-likeness criterion, they may engage humans in social cognition and therefore become psychologically plausible.

While it is true that anthropomorphic robots may facilitate certain communication channels, what is more important than human-like appearance is the clear expression of those communication channels. To be more precise, it does not matter if a robot looks extremely human-like if its behaviour and expressions are not understood by humans. In contrast, there are cases where successful communication is achieved even if the robot and the human do not share similar physical characteristics. For example, the interactive space Ada does not look like a human and does not use the same communication methods as humans; instead, it uses light and sounds Eng et al. (2003, 2005) to express its intentions to its visitors. Humans can read Ada's cues successfully. Similarly, communication through light and sound is achieved with the eXperience Induction Machine (XIM) Betella et al. (2013) and other robotic animals Collins et al. (2015). Thus, an anthropomorphic body is not what is important; an essential requirement is the readability and transparency of the employed communication channels.

Hence one could ask two broad questions regarding the robot's morphology: "Does the robot successfully communicate its intended internal states?" or similarly, "Are humans able to read the behaviour or communication channels of the robot?". Hence the criteria that could be fulfilled are transparency of the communication channels and facilitation for the projection of Theory of Mind. To evaluate the morphological criterion, we examine the transparency and recognisability of the robot's facial expressions and evaluate possible designs concerning the robot's functionality.

4.4 Autonomy

The final criterion that needs to be fulfilled so that a robot can be considered psychologically plausible is that of *autonomy*. There are two ways to view autonomy, one by looking at psychology and one by looking at the field of robotics. In psychology, *autonomy* is typically connected to self-governance, and most would agree that is normatively significant. The Greek philosopher Aristotle argued that self-sufficiency, that is, the ability to not depend upon others, was of great importance for happiness. Kant introduced the concept of *moral autonomy* by relating self-government to morality: one's actions (self-government) should obey their self-imposed laws or will (morality). In contrast, the approach of *personal autonomy* highlights the ability to make decisions without the influence of others, excluding any moral content. Piaget (1997) proposes that children's moral development occurs in two phases of autonomy: the *heteronomous*, in which children respect the rules (that are objective and not changing) and authority; and the *autonomous*, where the rules are subject to change through peer interaction. Following Piaget's ideas, Kohlberg claimed that only toward adolescence individuals separate themselves from the society and view themselves as entities that hold different opinions than others. From this aspect, autonomy can be viewed like *independence* from others and can be linked to the attribution of Theory of Mind to the robot. With this view, the psychological plausibility of the robot lies on the assumption that humans can perceive an entity as autonomous if they perceive it as a motivated agent that acts on its own accords to satisfy its internal goals.

In robotics, autonomy is viewed as the ability of the robot to function without human intervention. As a term, it offers little similarities with the aforementioned theoretical concepts and is, therefore, harder to measure. In that sense, if a robot is not autonomous when performing a task, it is possible that at some point it will face anomalies that may exceed its preprogrammed capabilities and stop responding. The psychological plausibility of the robot, therefore, depends on the robot's ability to operate in a robust and autonomous way.

On the one hand, we are interested in understanding what are the factors that allow the robot to be perceived as an autonomous agent. We, therefore, propose that a motivation system coupled with an emotion system may account for the perception of the robot as an autonomous agent: as actions are selected to satisfy the robot's internal drives, its behaviour may be perceived as proactive. On the other hand, we examine the control system that allows for the expression of robust and autonomous behaviours. More specifically, we propose the DAC cognitive architecture Verschure (2012); Verschure et al. (2003) as the control mechanism that guides the behaviour of the robot and allows for autonomy, social competence and task competence, regardless of the physical representation of the robot.

4.5 Evaluation methods

To evaluate our taxonomy we have conducted a number of studies. We first start off by evaluating the autonomous behaviour of the robot by looking at the implemented architecture: does the interplay between the emotional and motivational system of our synthetic agent results in a robust autonomous interaction? Thus, we show that a robot endowed with a set of drives that aim at initiating and maintaining an interaction is able to trigger behaviours that aim at satisfying its needs.

Before we continue with the evaluation of the social competence of the robot, we conducted a small study where we evaluated the readability of the robot's expressivity of emotions. Hence, we varied the facial expressions of the robot and we asked participants to rate them in terms of valence and arousal. Results show a positive correlation of the mouth expression with valence and eye aperture with arousal but not a combination of both. The information acquired from this study is used as a parameter for the next set of experiments.

We next evaluate the complexity of the robot's social behaviour and assess the psychological plausibility of the robot. To understand how the behavioural components affect the robot's believability we devised six inter-

action scenarios of increased complexity and asked participants to evaluate the robot. The components we manipulated are touch, speech, gaze model, facial expressions, interpersonal distance and proactive behaviour. Results suggest that the more complex the robot's behaviour appears, the higher it scores in psychological plausibility. In this context, we measure the psychological plausibility in terms of anthropomorphism, likeability, perceived intelligence and animacy using the Godspeed questionnaire.

We later focus on two key social components: facial expressions and gaze behaviour. More specifically, we examine whether social competence can elicit empathic responses. To answer this question, adapted the Milgram's classical experiment that measures obedience to our robotic application. In this study, we hypothesised that social competence could trigger empathic responses. We, therefore, manipulated the robot's gaze model and facial expressions and looked at the empathic relation between the participant and the robot. We speculated that if the robot is psychologically plausible, it will elicit empathic responses from the observer. In this context, empathy is a measure taken by the observer as the effectivity of the robot's social cues. We characterised as empathic responses the participant's time of administration of the negative stimulus, gaze mode and behavioural reactions. Our results show that indeed the robot was able to elicit empathic responses to humans.

Finally, to evaluate the robot's task competence, we use tutoring as the validation domain. Hence, we evaluated the social and task competence of the robot by examining the role of facial expressions and gaze model in a dyadic tutoring scenario. We conducted this experiment with both adults and children and identified the impact of gaze in engagement. The validation scenario mainly focuses on children and the next study we conducted concentrated on the morphological aspects of robots from the child's point of view. We exposed children to three different robotic platforms and asked them to evaluate them. Additionally, we asked them to draw the robot they would like to have and interact with and assessed their drawings regarding functionality and morphology. Results suggest that children tend to design

multi-purpose robots that are anthropomorphic but are more machine-like than human-like.

Taking all this together we have created a tutoring scenario based on the Piagetian Balance beam. In this scenario the robot acts as a peer tutor, guiding children through the process of learning physics by understanding the rules of the balance. Our first aim is to evaluate individually the features of a non-anthropomorphic setup which will later be used as the instructing interface the robot uses. To assess this non-anthropomorphic setup, we used three different content presentation tools: a physical balance, a virtual balance and a motorised balance coupled with an augmented reality application. The results of this experiment were used in the final evaluation of the synthetic tutor's task competence. We thus explored the minimum set of tools and behavioural components needed to efficiently and effectively teach children physics. To do so, we varied the nature of the robot's help by providing hits (open/closed) and distractions (jokes/trivia).

Chapter 5

The DAC control architecture and the implementation of H5WRobot

In the previous chapter we proposed that for an agent to be psychologically plausible, it needs to satisfy the criterions of social competence, task competence, autonomy and morphology. We propose that for a robot to be psychologically plausible, it needs to be autonomous, make decisions and perform actions without human intervention, using internal decision-making mechanisms.

Autonomous behaviour can be achieved with the implementation of a motivational system coupled with an emotional one that act as the first layer of behavioural control. We propose that by employing a mechanism of homeostatic and allostatic control on the robot's motivational system (drives), we can succeed in the emergence of complex behaviours that are generated in response to a dynamic environment to satisfy the robot's needs. The proposed motivational system ensures autonomy by evaluating how the self is situated in the world and assess if the self's needs and goals are satisfied. However, operating in an autonomous way is not enough for a robot to be considered psychologically plausible if the human observer does not perceive

it as one. For a robot to be seen as autonomous, its behaviour should follow the psychological norms that constitute human autonomy. Hence, users may view the robot as an entity that is autonomous due to the fact that its motivational systems guide the behaviour of the robot towards the execution of actions that will bring the robot closer to its goal. Thus, a robot may appear as an agent that pursues its own goals and proactively acts to satisfy its needs. Additionally, the motivation system can contribute to the interaction (and therefore social competence): the intrinsic need to socially engage provides a form of social autonomy, since successful interactions need agents that are motivated to interact with others.

We also suggest that social behaviour cannot be uncoupled from the ability to socially perceive. The recognition and identification of other social agents facilitate the establishment of communication and behaviour adaptation based on the human's behaviour, actions and internal states. Modulating the robot's behaviour in accordance with the partner's actions, internal states or the environment contribute to the psychological plausibility of the robot: the robot is perceived as autonomous, "aware" of its environment. Hence, the observer can label the robot's actions as originating from a social model maintained by the robot Reeves and Nass (1996); Breazeal (2003a).

Additionally, we argue that the display of joint attention, gaze, emotional expressions and a repertoire of complex behaviours are crucial for a successful HRI and the emergence of psychological plausibility. More specifically, in dyadic interactions, the gaze of agents has a direct influence on other's actions Boucher et al. (2012); Frischen et al. (2007); Lallée et al. (2013). Gaze can be a purely perceptual process that guides attention to the environment or a communication channel between agents through eye contact and gaze "pointing". Thus, it is considered an important communication channel, especially in robots whose spoken language is still limited. While extensive research on gaze has been carried out in humans Frischen et al. (2007); Knapp et al. (2013); Argyle and Dean (1965), gaze gained importance in the robotic community because on robots it can be accurately controlled and tuned Mutlu et al. (2006); Brown (1990); Boucher et al. (2012). From a util-

itarian perspective, the combination of gaze, gestures and facial expressions allows humans to better understand the behaviour of the robot Ono et al. (2000); Breazeal et al. (2005) by making its internal states transparent contributing to the social competence of the robot. Additionally, intentional gestures and gaze account for the understanding of the surrounding world and they may be sufficient parameters for people to attribute awareness and intentionality to the robot. However, for a robot to make efficient use of those cues, it needs to control and modulate them in a way that represents its internal states and express them in a way that the human partner can read and understand.

A core challenge in the creation of a psychologically plausible agent lies in implementation, as all four benchmarks need a control system that allows for the generation of the required behaviours. To solve this issue, we propose the Distributed Adaptive Control (DAC) architecture that we present in detail in the next section, as the control system that guides the robot's behaviour and allows for the development of believable agents. Core ingredients for the proposed system are actions, goals and drives. Drives are the intrinsic needs of the robot. Goals depend on and are set by the agent's internal drives and actions are generated to satisfy the robot's goals. We demonstrate how the DAC architecture makes it possible to achieve successful social interactions between humans and robots and how it can be used as a tool to systematically study the effects of different social features and behaviours. To test the implementation of the proposed architecture, and consequently the psychological plausibility of an agent, we devised a set of interaction scenarios where we systematically modulated behavioural parameters of the robot.

5.1 The DAC architecture

The "Distributed Adaptive Control" (DAC) Verschure (2012) is a cognitive architecture that provides a real-time model for perception, behaviour and cognition. DAC is a biologically constrained and fully grounded theory

of mind and brain since it autonomously generates representations of its primary sensory inputs Maffei et al. (2015); Verschure et al. (2003).

DAC has been formulated in the context of classical and operant conditioning and is a standard in the domains of behaviour-based robotics and artificial intelligence.

According to Pavlov, brains evolved to act so as to maintain equilibrium between the organism and the environment. DAC proposes that in order to act upon the environment (or realise the "*How*" of survival), any brain has to answer continuously and in real-time the following fundamental questions:

- *Why?*: what are the motivations, goals, drives for behaviour that leads to the generation of action.
- *What?*: what are the objects in the world that belong to these goals.
- *Where?*: the location in space of both the self and the objects of interest.
- *When?*: the timely execution of an action (present-future).

These four questions of varying complexity formulate the ***H4W*** problem that the brain needs to solve any given moment to ensure survival Verschure (2012); Verschure et al. (2014). However, this world is not static and solely filled with physical objects; it is inhabited by other agents. Consequently, the H4W problem is not enough to ensure survival. Now, an organism should worry about both the physical as well as the intentional world, that is filled with hidden states of other agents. As the world becomes more complex and intentional, a new question needs to be addressed: *Who?* (understanding of hidden states of other agents) forming the ***H5W*** problem Verschure (2014); Prescott et al. (2014). DAC suggests that the unifying phenomenon that allows us to understand the brain and build psychologically valid artificial agents is consciousness: the result of dealing with the H5W problem.

To answer the H5W problem, DAC views the brain as a distributed system that is dominated by parallel control loops. DAC postulates that the brain is organised horizontally around four tightly coupled layers of control: Somatic, Reactive, Adaptive and Contextual. Vertically, the brain is organised around three columns sub-serving the processing of the sensation and perception of the world (*exosensing*), representation of self (*endosensing*) and the interface to the world through *generation of action*. These layers of control continuously cooperate and compete for control of action and tightly couple the organism with its environment (Figure 5.1). We define action as the outcome of behaviour that serves internally-generated goals.

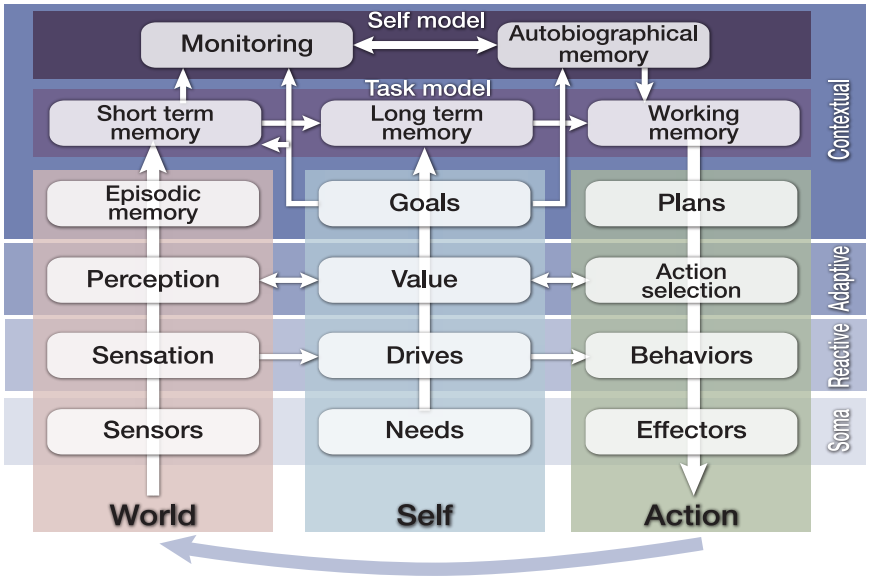


Figure 5.1: The Distributed Adaptive Control (DAC) theory of mind and brain architecture graphically represented. DAC proposes that the mind is organised in layered control structures (Somatic, Reactive, Adaptive and Contextual) tightly coupled together. Across layers, there is a columnar organisation regarding processing the states of the world (left, red, *exosensing*), the self (middle, blue, *endosensing*) and action (right, green) that mediates between the first two. At the contextual layer, these axes become tightly integrated. Arrows indicate the flow of information. See text for further information. Image adapted from Verschure (2012).

Functionally, DAC aims at providing an artificial body (i.e a robot) with the ability and motivation to act and survive within its environment and can drive a more integrated approach towards advanced machines.

5.1.1 Somatic layer (SL)

The Somatic layer (SL) represents the body itself and defines the information acquired from *sensation* (coming from both internal and external stimuli), *needs* (that ensure survival) and *actuation* (the control of the body's movement).

5.1.2 Reactive layer (RL)

The Reactive layer (RL) produces behaviours that support the basic functionality of the SL in terms of reflexive behaviour. The RL includes fast predefined sensorimotor loops (reflexes) that are triggered by low complexity signals most of which are defined by the SL and support survival. In short, specific stimuli are hardwired with specific predefined actions. Each of these reflex/behaviour systems is coupled to specific affective states of the agent that allow for labelling events in affective terms that later are used to guide behaviour. The RL constitutes the primary behavioural system based on the organism's physical needs. Thus, behaviour emerges from the satisfaction of homeostatic needs. The behavioural systems are regulated by an integrative allostatic loop that sets the priorities and hierarchies of all the competitive homeostatic systems. Thus, the Reactive layer is modelled in terms of a self-regulatory allostatic process (Sanchez-Fibla et al., 2010) and provides the first level of simple adaptive behaviours.

5.1.3 Adaptive layer (AL)

The adaptive layer (AL) extends the sensorimotor loops of the RL with acquired sensor and action states associated with valence. The AL is formulated in the context of Pavlovian classical conditioning: an initially neutral stimulus (also referred to as conditioned stimulus - CS) is able to trigger

actions (i.e. conditioned responses - CR) if paired with a motivational stimulus (i.e. unconditioned stimulus - US). Through Hebbian learning Verschure and Pfeifer (1993), the organism tries to minimise prediction errors between acquired and encountered states of the world: the agent is now able to escape the predefined reflexes of the AL and deal with the unpredictability of the world through learning Duff and Verschure (2010).

5.1.4 Contextual layer (CL)

The contextual layer (CL) receives as an input the state–space acquired by the AL and generates goal–oriented behavioural plans and policies that can be expressed through actions. This layer includes mechanisms for short-term (STM), long-term (LTM) and working memory (WM), formatting sequential representations of states of the environment and actions generated by the agent or its acquired sensorimotor contingencies, in relation to the goals of the agent and its value functions. In short, the CL organises the LTM along behavioural goals that are defined in terms of the organism's drives (RL) and the valence they are associated with at the AL, which in turn affects the process of decision-making and action selection. Behaviours are therefore selected based on the goal achievement, taking into account both the current state of the world and self and predictions about its evolution.

These layers are tightly coupled and cannot be seen as independent encapsulated modules: each layer is based on control signals that are generated by other layers. Thus, solving the H5W problem critically depends on the interaction between all the layers of the architecture. In fact, the "*what*" and the "*where*" arise mainly from the interpretation of the sensory stream: the perception of the world (left column) interprets the sensory inputs into gradually abstract constructs that are at the Contextual layer, the constituent elements of the episodic memory of the agent. A similar gradient occurs in the self-representation column, where the brain first assesses the motivational states that derive from the organism's physical needs and prioritise them so that goals can be defined ("*why*"). Again, the different levels

on the vertical axis represent different levels of spatiotemporal abstraction ("*when*"): as the agent can monitor larger parts of its past knowledge, it can predict the evolution of the world state and to plan its actions ("*how*", right column).

5.1.5 DAC as a benchmarking mechanism for psychological plausibility

The way DAC is conceptualised allows for benchmarking all four of our proposed taxonomies. The implementation of DAC is taking the H5W problem as the main way to represent the knowledge about the world and to exchange this information among the different modules composing the architecture.

Any artificial entity needs to make sense of its surrounding world and know how it can affect it through actions and describe it through language Lallee et al. (2010). While the classical instantiation of robots makes use of a single body that encompasses all sensors and effectors, contemporary systems tend to adopt a distributed organisation of external sensory inputs like the Kinect, biophysical readers Matthews et al. (2007); Badia et al. (2009) or smart house installations Cook and Das (2007); Eng et al. (2003).

The implementation of external multimodal sensory inputs provides the agent with an accurate representation of the world and the agents that populate it. The formalisation of this growing heterogeneous stream of information is essential, and the DAC approach provides an elegant way of representing this stream. The Somatic layer provides a powerful sensorimotor abstraction turning most of the architecture into a platform independent system. While some attempts to achieve such formalism have been seen in platform-independent cognitive architectures and ubiquitous robotics Tenorth et al. (2012); Lallée et al. (2011), those systems do not take into account the agent's internal states and are therefore unable to develop an intrinsic motivation to act Cañamero (1997).

The state-of-the-world representation encompasses the physical world coupled with the internal states and beliefs of its agents. The representation

of "self" and how it affects and is affected by the environment is a central point in the DAC architecture as it allows to generate online reactive and adaptive behaviours aimed at satisfying the agent's needs. A body/sensor abstracted representation of the world in combination with an emotional and motivational system that guides behaviour allows us to achieve autonomous interactions. The robot is able to form a representation of its environment while acting proactively to satisfy its needs. Since most of the DAC modules receive higher level information, morphology is no longer an issue, as body abstraction is achieved with changes only in the Somatic layer (how information is received and how an action is produced).

The representation of the agent's psychological engine allows for expressing its internal states with a variety of ways. Thus, we can decompose and systematically study social behaviour by "turning on" or "off" a plethora of communication channels such as gaze, gestures or facial expressions. The appropriate manipulation of the agent's behavioural components allows us to explore and shape the robot's social competence. Additionally, by employing the learning mechanisms of the Adaptive layer to adjust various parameters of the task, and the planning mechanisms of the Contextual layer to find the optimal strategies, we propose that the agent can achieve task competence.

The brain, the body and the mind are indissociable in biological beings, forming a nexus that implies the impossibility to study one component without unfolding implications in the others. As a global theory of the mind-brain-body nexus (MBBN), DAC provides a functional explanation of several phenomena and concepts ranging from body schemas, self-other distinction, to planning and the emergence of self and consciousness. The DAC theory is tested through convergent validation Verschure (1997), meaning that as long as the framework assumptions assist an implementation, each successful experimental result provides evidence supporting the theory.

DAC has been validated through a variety of robotic implementations Fibla et al. (2010); Sanchez-Fibla et al. (2010); Maffei et al. (2015) and was

expanded to capture the social aspects of interactions Verschure (2014); Prescott et al. (2014); Vouloutsi et al. (2013a); Lallée et al. (2014, 2015); Vouloutsi et al. (2016). Additionally, DAC has given rise to a successful and novel approach towards rehabilitation that is being deployed in clinics today Verschure (2012). The architecture is self-contained, meaning that its knowledge representation arises from the interaction of the agent with its environment. Such attributes make this architecture more adaptive compared to other architectures such as ACT-R Verschure (2003), a cognitive architecture that is meant to model human cognition at the process-level Anderson et al. (1997); Anderson (2007) or Soar Laird (2012) which is more concerned about higher-level functions than low-level cognitive fidelity. In our case, as the highest validation, DAC should provide the guidelines for implementing a robot that is psychologically plausible and replicates human aspects of behaviour.

In the following sections, we focus on the mechanisms that allow for the generation of plausible social behaviours that trigger responses in humans, like reflexive emotions or attribution of intelligence to the robot. We furthermore describe the technological challenges and proposed solutions to the issue of the integration of the various sensors needed by the robot to make sense of the world. We then use a number of robotic platforms as a medium to test several psychological hypotheses, highlighting the platform abstraction. The studies conducted mainly focus on social salience and how it affects the interaction with a human partner. In particular, we investigate which behavioural channels or parameters are relevant to induce a feeling of empathy and the attribution of self and psychological validity toward a non-biological artefact.

5.2 The development of H5WRobot

The DAC architecture and the framework we propose is mostly hardware independent. We demonstrate its implementation by controlling the behaviour of various robotic platforms involving a large set of sensors and

effectors. In this thesis, we have used two types of setups that are specifically designed to study human-robot interaction in a smart environment. This smart environment usually an apparatus that acts as a digital interface to facilitate interactions.

The main setup includes a humanoid robot, namely the iCub Metta et al. (2010) mounted on an omnidirectional-wheeled mobile base (iKart), the Reactable Geiger et al. (2010) (a tabletop tangible display) and an RGB depth-sensor used to provide accurate detection of humans in the environment (Kinect). This setup is conceptualised for dyadic game-like interactions and we name it H5W_Alpha. The combination of all the setup components allows the implementation of various interactive scenarios like games (e.g. Pong), cooperative musical creation (MusicDJ) or learning (e.g. learning about geography or recycling). The implemented interaction scenarios require both the human and the robot to act in a shared physical space.

This installation has been demonstrated in open public events (Barcelona Robotics Meeting 2014; Innovation Convention 2014[1], Brussels; Fiesta de la Ciencia 2013, Barcelona; Living Machines 2013, London; ICT 2013[2], Vilnius), therefore indicating a robust and easy generation of interactions with inexperienced users in unconstrained environments. We intensively used the same setup in controlled laboratory conditions to investigate how the parameterisation of the architecture impacts the natural interaction with naive users.

The second setup includes a humanoid robot and a handheld device, namely the EASELscope. We name this setup as H5W_STA and we use this setup in ecologically valid interaction scenarios in schools. Further information about the current architecture can be found in Lallée et al. (2015); Vouloutsi et al. (2016) and in A.2.

[1] http://ec.europa.eu/research/innovation-union/ic2014/index_en.cfm?pg=showcase02

[2] "The machine and I" https://ec.europa.eu/digital-single-market/news/future-emerging-technologies-fet-ict-2013

5.2.1 Motivation and emotional system

Inspired by Maslow's hierarchy of needs Maslow (1943), Hull's drive reduction theory Hull (1943) and tested in the autonomous interactive space Ada Eng et al. (2003), the robot aims at satisfying its internal states through action. Drives are part of a homeostatic process Cannon (1932); Seward (1956) used by an organism to maintain equilibrium and define needs and goals, contributing to the action-selection process of the agent. The control system operates in real time and is designed to generate behaviours based on stimuli received from both the environment and the internal states of the agent.

A homeostatic model is applied to each drive, defining its current value and its ideal regime. The homeostatic model calculates the value of each drive, classifies its homeostatic state and then projects the output to the allostatic controller. The homeostatic state can be `under`, `over` and `within` homeostasis and different actions are chosen for the various homeostatic regimes. The controller's goal is to achieve balance and consistency in satisfying the drives through behavioural change. It is an essential component of maintaining homeostasis, as it is responsible for both the emergence of behaviours and maintenance of the system in balance by avoiding cases of conflict (i.e. the case where two drives need to be satisfied at the same time).

The satisfaction of each drive also impacts the evolution of the emotional model, mainly by moving towards a negative emotion when drives are not satisfied and positive when they are. The emotional model adopted in our case is a two-dimensional Valence-Arousal view. The emotions and drives have their own internal dynamic (5.1) that can be expressed as the variation of a homeostatic model H_j which consists of a constant decay d_j as well as an influence from all of the semantic stimuli S_i, either excitatory or inhibitory depending of the connection W_{ij}. As usual, $f()$ represents an activation

function (e.g. threshold, sigmoid).

$$\Delta H_j = -d_j + f(\sum W_{ij} S_i) \tag{5.1}$$

Our implementation provides an abstract way of manipulating the concepts of homeostasis and allostasis as well as an example of their instantiation (Figure 5.2). More importantly, we propose a way of assessing such homeostatic systems by monitoring the empathic effects produced by an external agent. Our efforts regarding the implementation of H5W_Alpha mainly focus on creating a social agent, a robot that can interact proactively with people; this goal is reflected in the choice of the set of drives implemented (physical interaction, spoken interaction, social interaction and energy) as they constitute the main levers to act on in order to tune the behaviour of the robot. The behavioural engine constantly monitors the drives' system and triggers alerts whenever a drive is detected, as being out of its homeostatic boundaries. As the sensors of the robot are interpreted into semantic relations, they modulate the natural decay of the homeostatic models by compensating, accentuating or reversing it. The parameters of the drive's dynamic can be tuned and provide the most direct way to control the robot's personality.

The role of drives and emotions is two-fold: 1) to provide the robot with an internal model of itself that it can observe and express through facial expressions, 2) to influence the action selection and execution. The overall satisfaction of drives is achieved by the allostatic controller, which tries to minimise the output of homeostasis signals at different scales of time by triggering compensatory actions. The emotional model acts on top of the selected action by setting a stance (an angry agent would perform the same action in a more aggressive way than a happy one). The combination of both ensures the selection of the best action in terms of drives and its customisation depending on the current emotions of the robot. More importantly, from a social interaction perspective, the fact that the robot is acting based on its needs and reacting emotionally may contribute to the psychological

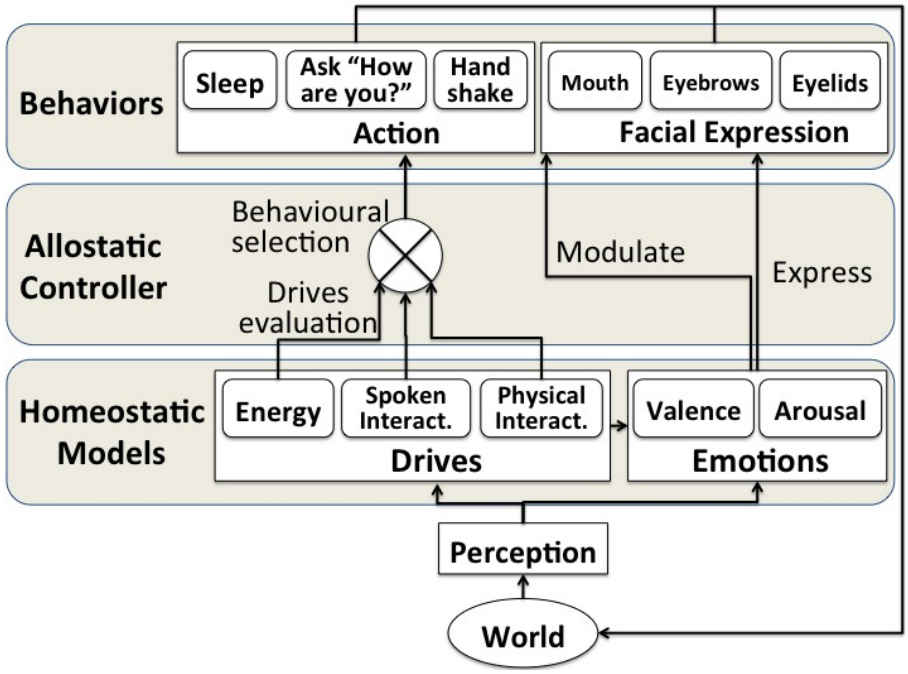


Figure 5.2: Detailed diagram of the behaviour generation. The world is perceived, impacting the drives and emotions. Drives are then evaluated by the allostatic controller which selects an action from the pool of available behaviours and execute it. The behaviours shown on the diagram are just a subset example to illustrate the principle: if a human partner is perceived while the drive for spoken interaction is high, the allostatic controller may select the action "Ask: How are You?"; in the case of a high physical interaction drive it may prefer the "Handshake" action. In the eventuality of a critical energy need, the robot will ignore the human and set itself to sleep mode. As a parallel process, emotions are constantly updated based on the content of the environment and the global satisfaction of the drives. In turn, emotions are expressed through facial expression and they modulate the execution of actions.

plausibility of the robot. By expressing the reasons for its actions and how it feels, the robot could be perceived by humans as an autonomous agent, driven by its own motivations and pursuing its own goals.

5.2.2 Setup and robotic platforms

In the following section, we present the components of both H5W_Alpha (iCub, Kinect, Reactable) and H5W_STA (Nao, EASELscope). The installation that consists of the H5W_Alpha (with the applications of Pong and MusicDJ) has been demonstrated in open public events (Barcelona Robotics Meeting 2014; Innovation Convention 2014, Brussels; Fiesta de la Ciencia 2013, Barcelona; Living Machines 2013, London; ICT 2013, Vilnius), therefore indicating a robust and easy generation of interactions with inexperienced users in an unconstrained environment. We also use intensively the same setup in controlled laboratory conditions in order to investigate how the parameterisation of the architecture impacts the natural interaction with naive users. In the following chapters, we discuss the implementation of both H5W_Alpha and H5W_STA.

iCub and iKart

The iCub is a humanoid robot with dimensions similar to a 3.5 years old child. Its body consists of a head, hands, arms, torso and legs and it has 53 degrees of freedom (7 for each arm, 9 for each hand, 6 for the head, 3 for the torso and waist and 6 for each leg) Metta et al. (2008). The robot is able to perceive tactile feedback as it is equipped with novel artificial skin covering the hands (fingertips and palm), forearms and the chest area. Having a robot with tactile capabilities allows for physical interaction that may tighten the social bond between the user and the robot Cramer et al. (2009b,a). The head consists of two cameras mounted in the eyes of the robot and two microphones placed in the ears. Strips of red LEDs are projected from behind the face panel placed in the mouth and the eyebrows of the robot. The iCub can express a variety of facial expressions through a combination of the LEDs and eye aperture, as can be seen in Figure 5.3.

The iCub is mounted on the iKart, a holonomic mobile platform designed to provide autonomous navigation capabilities in structured environments. Navigation is achieved through six omnidirectional wheels that allow the platform to both translate and rotate. Obstacle detection and localisation

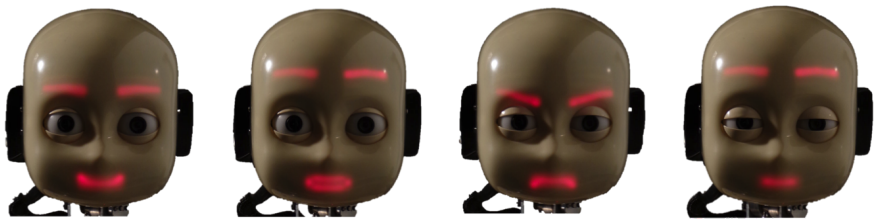


Figure 5.3: Examples of the variability in the facial expressions if the iCub when the eye aperture, eyebrows and mouth are modulated.

are achieved through a laser rangefinder mounted in the front of the platform. Navigation capabilities are provided using odometry coupled with automatic recalibration in the critical section of the interaction (i.e before manipulating objects over the table).

RGB / Depth Camera

To track humans, we employ the Kinect sensor by Microsoft. The Kinect is an RGB camera, depth sensor and multi-array microphone that is interfaced with our current architecture. We have used and integrated both models of Kinect ("Xbox 360" and "Xbox One") the information provided include the partner's position as well as facial expressions and gestures.

Nao

The Nao is an autonomous humanoid robot of 58cm height developed by Aldebaran Robotics in France. It has 21 degrees of freedom, four microphones (for speech recognition and sound localisation), two speakers and two HD cameras. Although it cannot display facial expressions as it lacks mouth and eyebrows, it can exhibit emotional states through a circle of coloured LEDs that surround its eyes.

Reactable

Conceptualised as a new musical instrument, the Reactable is a tabletop tangible interface Geiger et al. (2010); Jordà (2008) composed of a round

table with a translucent top where objects and fingertips are placed to control the parameters of the melody. A back projector displays the properties of the digital or physical objects through the translucent top and a camera is used to track the objects placed on its surface. Object recognition is performed using the reacTIVision tracking system that provides information regarding the location of fingertips (also known as *cursors*) or objects, including their rotation and speed. An example of the interaction setup using the Reactable and the iCub is depicted in Figure 5.4.

We took advantage of the Reactable's dynamic display in order to create game-like scenarios for dyadic interactions, tailored to our needs. Additionally, the Reactable provides information about the location of virtual and physical objects placed on the table and allows a precision that can hardly be matched using a vision-based approach. The game engine used is openFrameworks (http://openframeworks.cc/) in combination with the MTCF framework Julià et al. (2011) and all games implemented employ fingertips and objects used as controllers. To satisfy the purpose of a variety of interaction scenarios we have implemented the following games/applications:

- *Pong*: a competitive 2D simulated table tennis game. Each player controls a virtual paddle by moving a physical object horizontally on the Reactable's surface. The goal of each player is to use the paddle to hit a virtual ball. The player that manages to hit the ball that his opponent fails to return wins a point.

- *Music DJ*: a collaborative game in which each player controls three musical loops: a melody, a bass and drums either with a physical object or using their fingertips. The music can change by placing a coloured cube on the surface of the table. Each side of the cube is coloured in a different way and represents a different musical style.

- *Tic Tac Toe*: a turn-taking competitive 2D simulated game of *Noughts and Crosses* that is played on a 3x3 grid. Each player holds a physical object that draws on the grid either an "X" or an "O". The player

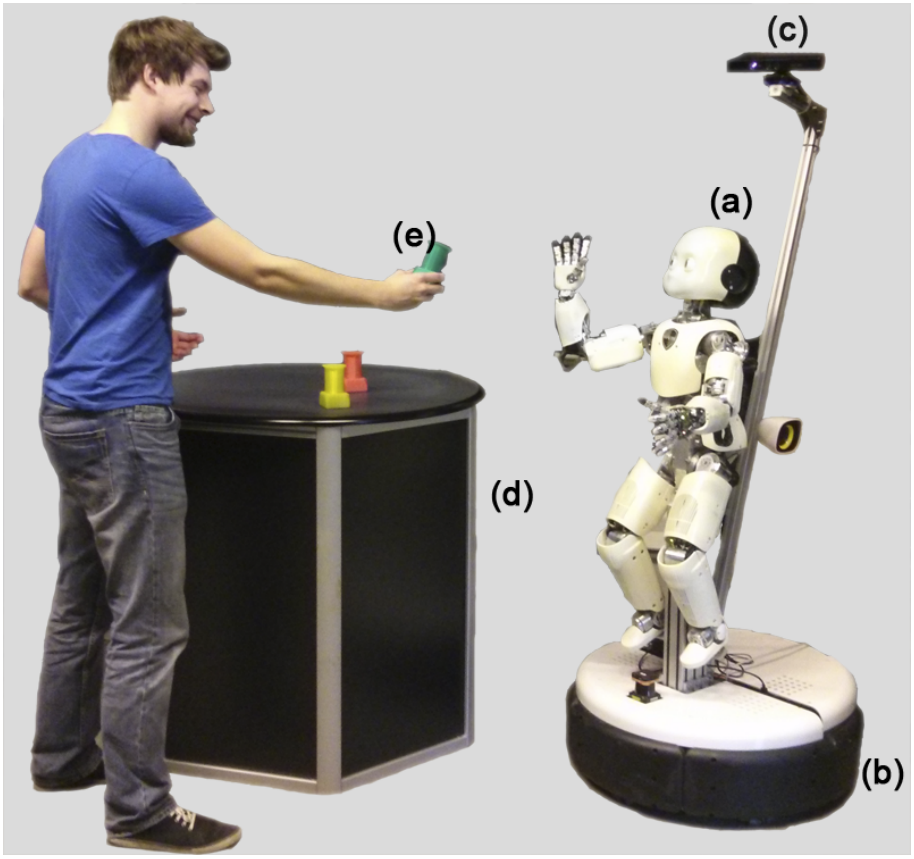


Figure 5.4: Example of the proposed scenario: the humanoid iCub (a) is mounted on the iKart (b) to navigate within the environment. On top of the iCub, the Kinect sensor is placed (c) to provide information regarding the location of the human. The iCub can interact with the human in different interaction scenarios, including playing games and music using the Reactable (d) by manipulating objects (e).

who succeeds in linking three horizontal, vertical or diagonal marks wins a point.

- *Simon*: a turn-taking competitive 2D memory game. Four different objects are displayed on the Reactable's surface with a distinct shape (square, circle, triangle and rhombus) and a distinct colour (green, yellow, red and blue). A round in the game consists of the Reactable

highlighting in a random order the shapes displayed on the table, after which the player has to reproduce them by placing his paddle on top of each item. If the player completes successfully the sequence, the sequence increases by one and the opponent takes the turn. A player wins a point if the opponent fails to reproduce the sequence.

- *Pairing*: a turn-taking educational game of increasing difficulty that aims to teach recycling and geography. Each player has a set of four categories and four items (each item belongs to one category) that they need to match. Both items and categories are mirrored at each player's side. In the case of recycling, categories consist of different recycling bins and items of recycling objects. Similarly, in the case of geography, categories consist of a country's flag and name while the items are the names of their capitals. In each turn, the player has to select a category and an item; if the pairing is incorrect, the opponent takes the turn; if the pairing is correct, the selected items are no longer available and the opponent assumes his turn. A level is completed when all four categories are correctly paired with their corresponding items and a new set of categories/items of increased difficulty is displayed on the surface of the Reactable.

- *Milgram*: a pairing game-like application conceptualised to recreate Milgram's study of obedience Milgram and Van den Haag (1978); Milgram (1965). The iCub has to pair the name of a colour with one of the colours displayed on the Reactable. The following information is displayed on the Reactable: the coloured squares (yellow, white, blue, green, red, orange, purple and black) and the intensity of the negative stimulus the participant has to administer in case of a mismatch between the name of the colour said and the coloured square selected by the iCub.

- *Guess Who*: a turn-taking competitive 2D game that simulates the "Theora Design". Each player starts the game with a predefined set of images of people. Each image consists of a cartoon face accompanied

by a name and specific features (such as glasses, hat, hair etc.). Both players have the same set of faces. Each player has to secretly choose a person that the opponent has to guess who is by asking yes or no questions to eliminate candidates. The player who correctly guesses the opponent's chosen character wins the game.

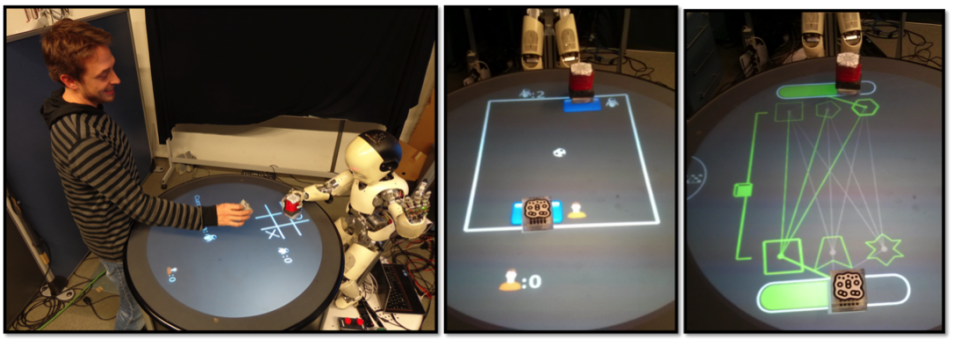


Figure 5.5: Examples of different Reactable applications implemented. From left to right: tic tac toe, pong and music DJ.

Given all this, we can achieve an autonomous robot that can be psychologically plausible, while the architecture also allows for fast prototyping of interaction scenarios.

Chapter 6

Evaluation of the psychological validity of H5WAlpha

In the previous chapters, we presented a taxonomy that defines the psychological validity of the robot and we proposed a minimum psychological engine, that is a number of behavioural and functional components that we suggest may affect perception. The components of the H5W_Alpha aim at creating the minimum psychological engine that would allow humans to accept the robot are: employ an emotional and motivational system that guides the behaviour of the robot, be able to communicate with humans using gestures, gaze and other non-verbal cues, display emotions, have an action repertoire that allows the robot to perform correctly its task and satisfy its needs and be able to perceive its environment and act accordingly.

The usage of the DAC architecture allows for the implementation of these aforementioned components that constitute the robot's psychological engine. In this thesis, we evaluate the components that are based on the four criterions of autonomy, social competence, task competence and morphology.

6.1 The modulation of behaviour using allostatic control

As the introduction of robots into our society is slowly coming closer to reality, their ability to be able to interact with humans in a meaningful and intuitive way gains importance. It is therefore essential that we start developing robots that are not just tools for automated processes but rather social agents that are able to interact with humans in a psychologically plausible way. For a robot to be considered a psychologically plausible agent with autonomous behaviours, we propose the following requirements: (i) intrinsic needs to socially engage, as successful interaction requires an agent that is socially motivated; (ii) action repertoire that supports communication and interaction, in a way that the agent is able to perform actions such as object manipulation, produce linguistic responses, recognise and identify a social agent, establish and maintain interaction etc. and finally (iii) the core ingredients of social competence: actions, goals and drives. We define as drives the intrinsic needs of the robot. Goals define the functional ontology of the robot and depend on the drives whereas actions are generated to satisfy goals.

A socially competent android requires a combination of drives and goals coupled together with an emotional system. Drives and goals motivate the robot's behaviour and evaluate action outcomes; emotions aim at appraising situations (epistemic emotions) and define communicative signals (utilitarian emotions). We propose an affective framework for a socially competent robot that uses an allostatic control model as the first level of motivational drive and behaviour selection combined with an emotion system (see section 5.2.1). In the following sections, we present the model in detail and display how complex behaviours emerge through human-robot interaction.

6.1.1 The H5W_Alpha: a humanoid robot that promotes interaction

In our framework, the robot's behaviour is guided by its internal drives and goals in order to satisfy its needs. Drives set the robot' goals and contribute to the process of action-selection. The overall system is based on the Distributive Adaptive Control (DAC) architecture Verschure (2012) which has extensively been described in chapter 5. The system is mainly focused on the reactive and adaptive layer of the DAC architecture setting a framework for higher cognitive processes such as state space learning. The setup used to validate includes the iCub and the Reactable as seen in the previous chapter and the interaction involves a human communicating with the iCub. An example of the proposed setup is illustrated at figure 6.1. We implemented the model of drives and emotions using IQR, an open-source multilevel neuronal simulation environment Bernardet and Verschure (2010) that is able to simulate biological nervous systems by using standard neural models.

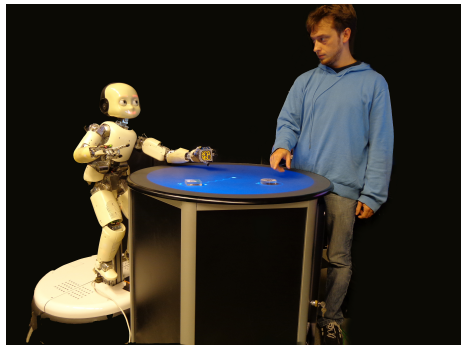


Figure 6.1: Example of the proposed scenario where the humanoid robot iCub interacts with a human and uses the Reactable objects as means of playing a game.

The behaviour of the robot is highly affected by its internal drives. Inspired by the intelligent space Ada Eng et al. (2005), an interactive entertainment space that promotes interactions with several people, the robot has the following goals that it aims at optimising: *Be social*: the robot's goal is to

interact with people and regulate its behaviour accordingly. *Exploration*: the need to be constantly stimulated. *Survival*: consists of two parts: physical and cognitive survival. As physical survival, we define the need of the robot to occasionally rest, whereas cognitive survival is the need to reduce complexity so as to not get confused. *Play*: the robot's need to engage the human with different games in order to form a more pleasant and interesting interaction. *Security*: the need to protect itself and avoid unwanted stimuli or events.

The goal of the agent is to socially engage with humans and its drives and emotions are designed to propel such a social interaction. The main goal of the robot is to maximise its happiness by keeping its drives in a homeostatic level. A homeostatic control is applied at each drive and on top of each subsystem we employ an allostatic control that aims at maintaining balance through behavioural change. The emotions that emerge through the agent's interaction with a human and the environment are the following: happiness, anger, sadness, fear, disgust and surprise. These emotions are compliant with Ekman's emotions Ekman (1992) that are considered to be basic from evolutionary, developmental and cross-cultural studies. The emotional system is responsible for exhibiting emotional responses that are consistent with the agent's internal state and are expressed through facial expressions. The emergence of emotions depends on two main factors: the satisfaction of the drives and external stimuli such as different tactile contacts (poke, caress, grab) which affect poke, happiness and fear respectively. An example of two different intensities of the same emotion, namely happiness is depicted at figure 6.2.

The implementation of a combined homeostatic and allostatic control that runs in parallel, contradicts the paradigm of state machines, as the proposed system allows the robot to display more complex behaviours. The dynamics of the model are depicted in figure 6.3.

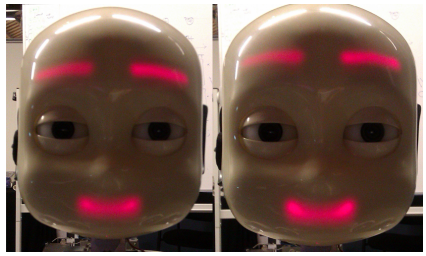


Figure 6.2: Example of the emotional expression of happiness. On the left, the intensity is set to 0.5 whereas on the right the intensity is set to 1.

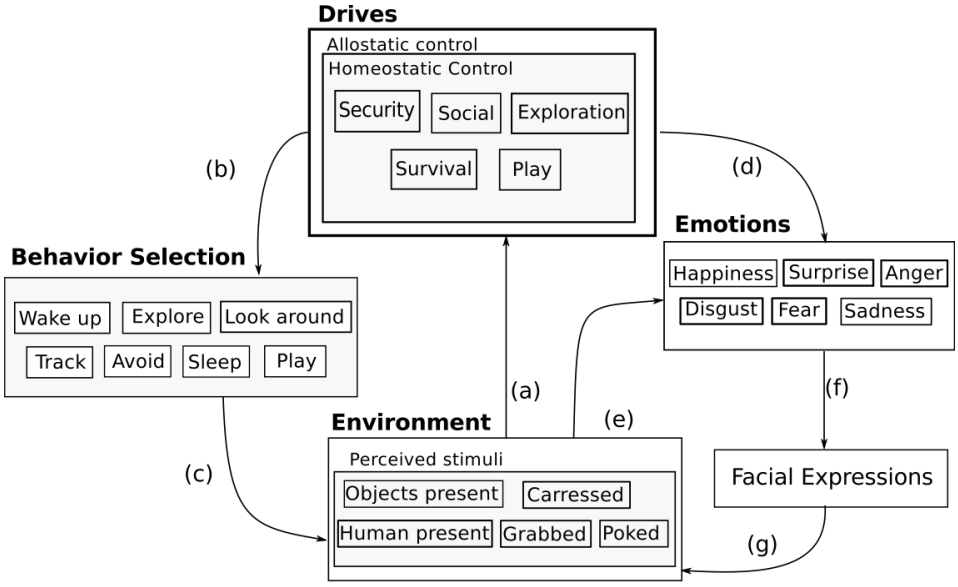


Figure 6.3: Overview of the parts involved at the behavioural level. Inputs from the environment are fed into the drives control mechanism (a) where there is an assessment of the homeostatic value of each drive and on top, we have the allostatic control that is monitoring the drives and the related stimuli. Depending on the value of each drive, an appropriate behaviour is being selected (b) and executed (c). At the same time, the level of satisfaction of each drive affects the emotions of the system (d) and in combination with the assessment of certain stimuli (e) emotions emerge in the emotion system. The most dominant emotion (f) is expressed (g) through the facial expressions of the EFAA.

6.1.2 Behavioural modulation

The agent has to perform different actions to satisfy its drives. Such behaviour is considered adaptive since it allows the system to achieve specific goals, like the satisfaction of a specific drive in a dynamic environment. We have employed the following behaviours: *Wake up*: the procedure in which the robot transits from inactivity to being "awake" and ready to interact. Waking up behaviour also initialises its drives and emotions. *Explore*: the robot interacts with objects on the table. *Look around*: the robot is looking around in an explorative way in order to find relevant and salient stimuli. *Track*: once a salient stimulus is found, the robot shifts its attention focus to the salient stimulus. *Play*: the robot engages the human in an interactive game. The play behaviour has two sub-scenarios: toss a dice and play a sequence game. *Avoid*: the robot informs the human that certain actions, objects or events are unwanted. *Sleep*: the robot's drives and emotions stop. The robot will not try to satisfy its drives nor express its emotional state. During sleep, the robot's drives are reset.

Currently, most of these behaviours are at a single level, i.e. they do not underlie a set of behaviours to choose from with the exception of the *play* behaviour. However, this is setting the ground for a more thorough implementation of behaviour selection where the H5W_Alpha agent can learn to pick the optimal behaviour. Table 6.1 illustrates the interaction between drives, emotions, perceived stimuli and behavioural processes. Some of the suggested behaviours are considered reflexive, such as the waking up of the robot when it is touched while asleep. However certain behaviours are employed not to satisfy a drive, but rather to create the appropriate conditions for the satisfaction of a drive. A typical example of the adaptive control is the satisfaction of the socialise drive: it requires a human to interact with. In case a human is already present and tracked, the robot enters in a social behaviour (dialogue, game). However, in case there is no human present, the robot will seek one by either looking around or by verbally expressing its need to have someone to play with. The look-around behaviour, in this case, is considered adaptive as it does not aim at directly satisfying the

social drive but rather aims at meeting the preconditions that will satisfy it.

Table 6.1: The perceived stimuli column refers to the presence or absence of certain stimuli that affect the drives and emotions system of EFAA. The drive column refers to the current drive that is affected by the inputs, emotion column refers to the emotions that emerge from a given situation and the behaviour column denotes the kind of behaviour that is triggered.

Perceived stimuli	**Drive**	**Emotion**	**Behaviour**
No human present	Social	Sadness	Look around
Human present	Social	Happiness	Track
No objects present on table	Exploration	Sadness	Look around
Objects on table	Exploration	Happiness	Explore
Too many objects	Cognitive survival	disgust	Avoid
Human caresses the iCub	-	Happiness	-
Human pokes the iCub	-	Anger	Avoid
Human grabs the iCub	Security	Fear	Avoid
Human leaves unexpectedly	Social	Surprise	Look around
Human touches the iCub when asleep	(drive initialisation)	-	Wake up
Human present and objects on table	Play	Happiness	Play
Robot interacts too long with human	Physical survival	-	Sleep

6.1.3 System assessment

During human-robot interaction, the robot proceeds in action-selection and triggers behaviours that aim at satisfying its internal drives. In figure 6.4 we present data obtained using the model described previously during a real-time human-robot interaction.

Our results show the interplay between drives, emotions, perceived stimuli

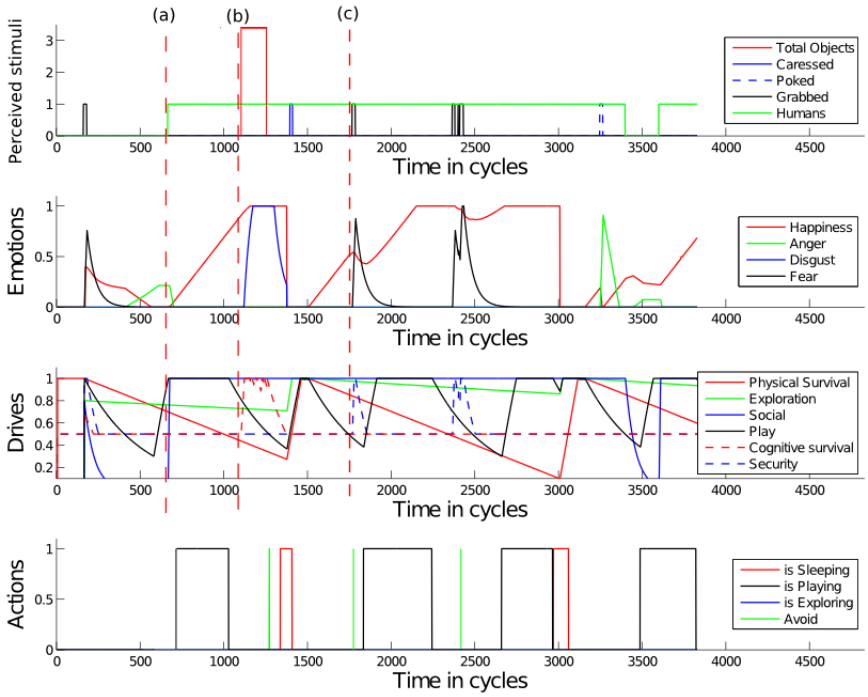


Figure 6.4: Overview of the drives and emotions system over time. On the upper panel, we can see the stimuli that are perceived from the environment (the number of people present, the number of objects on the table and the input from the skin of the robot: if it has been caressed, poked or grabbed). The "emotions" panel illustrates the emergence of different emotions (happiness, anger, surprise, sadness, disgust and fear). The next panel displays the drives values for survival (cognitive and physical), exploration, play, social and security whereas the "actions" panel indicates the emergence of the behaviours triggered in order to maintain the system in homeostasis.

and actions while they display some key features of the overall system. The "emotions panel" indicates the levels of each emotion over time. The red line represents the overall happiness of the robot during the interaction. On approximately the 1000th cycle we observe the presence of many objects on the table (b) which in turn promotes the emotion of disgust. At the same time, too many objects on the table cause the cognitive survival drive to

rise and trigger the "avoid" behaviour. On approximately the 3200th cycle (c) the robot perceived that it was grabbed which gave rise to fear. Poke rises the security drive which in turn also triggers the "avoid" behaviour. At certain moments in the simulation, more than a single emotion emerges, however only one is dominant. This emotion is the one with the highest value and is the one displayed on the facial expressions of the robot.

Another stimulus that affects the drives of the robot is that of the presence of a human. In deed, we see that until the more or less 600th cycle, there is no human present. This causes the social drive to fall and rise once the human appears (a). This is a good example to show how certain behaviours cannot be triggered unless certain conditions are met. For example, to initiate the play activity, the robot needs a human to play with and play is triggered in the "actions" panel once a human appears; a drive cannot be satisfied if the appropriate conditions are not met. An example where the human participant leaves the interaction scene (a) is depicted in figure 6.5. The play drive from the "drives" panel constantly decreases as conditions are not met (human is not present). Nonetheless, the robot proceeds in an explorative behaviour (b) and satisfies its exploration drive displaying a more adaptive behaviour. Only once the human returns, the robot is able to satisfy its play drive and trigger the appropriate behaviour.

Part of the role of the allostatic control is to make sure that certain actions do not collide. A robot can look or track a stimulus while talking or playing a game, however, it cannot play a game while sleeping. In the presented interaction, the physical survival of the robot gets low and needs to be satisfied at approximately the 2500th cycle (see the "drives" panel in figure 6.4), however at that time the robot is already playing with the human and cannot go to sleep. Once the play action is finished, the robot is free to proceed into the sleep behaviour.

Our results indicate how emotions and drives are affected by certain perceptions. Although behaviours are triggered with the scope of maintaining the drives in homeostasis, however, they are bound to these perceptions.

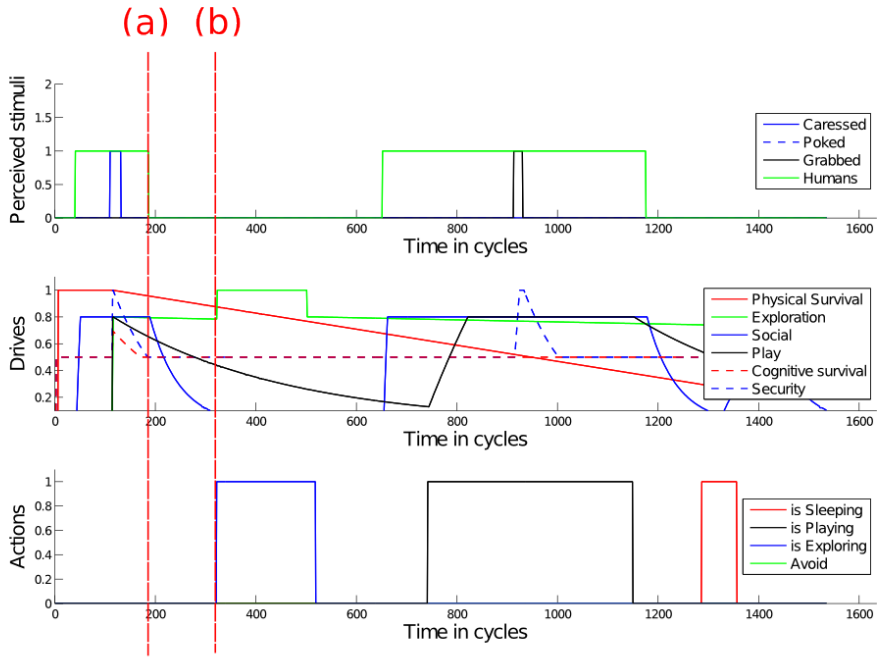


Figure 6.5: Example of the robot's behaviour during an interaction where the human leaves the scene (a). With the absence of humans, the robot starts exploring (b) objects on the table.

Nonetheless, we can observe the dynamics of the proposed system through the interplay of the external percepts, the robot's emotions and drives and the emergence of certain behaviours in the attempt to keep the balance in the agent's internal state. This motivated behaviour that is relevant to the various stimuli of the environment can contribute to the autonomy criterion of our proposed taxonomy, not only because it allows the robot to behave in a robust way, but also because it may be perceived as goal oriented.

6.1.4 Discussion

Nowadays there is an increased interest in developing social robots, that is robots that are able to interact and communicate with humans in a natural and intuitive way. In this section, we validated the proposed architecture in terms of autonomy, one of the criteria for the psychological believability

of the robot.

Here we propose a system that has the intrinsic need to socially engage and interact with humans and is equipped with an action repertoire that can support communication and interaction. This system includes drives that help satisfy the robot's intrinsic needs, emotions that assist the robot to express its internal state (utilitarian) and organise behaviours (epistemic) and a set of actions that aim at satisfying its needs. In the proposed model we have defined the following drives: sociality, exploration, survival, security and play. Each of these drives is monitored by a homeostatic control that classifies the level of each drive into the following stages: *under homeostasis*, *homeostasis* and *over homeostasis*. On top of homeostasis, we apply an allostatic control that is responsible for the maintenance of the system in balance by behaviour selection and priority assignation in order to satisfy its needs. The model's design is based on the reactive and adaptive layers of the Distributed Adaptive Control (DAC). The reactive layer is responsible for producing reflexive almost hard-wired responses while in the adaptive layer deals with the unpredictability of the world. However, the satisfaction of its needs highly depends on the environment and the current state of the world. As the allostatic control switches from a reactive to an adaptive level, it is not any more motivated by direct drive satisfaction but it is aiming for matching requirements so that an action leading to a given goal (that is the final drive satisfaction) will be available.

The satisfaction level of each drive defines the emotional state of the robot as well as certain external stimuli such as the robot being caressed, poked or grabbed by the human. The robot is able to exhibit six emotions: happiness, anger, sadness, disgust, surprise and fear, emotions that are considered to be basic from evolutionary and cross-cultural studies. The main goal of the robot is to maximise its happiness by keeping its drives in homeostasis. To do so, it is equipped with a set of different behaviours that it can trigger in order to satisfy its needs: wake up, explore, look around, track, play, avoid and sleep. Most of these behaviours are considered reflexive (such as wake up) and single layered, however there are also more complex behaviours such

as "play" that trigger 2 sub-scenarios: play a dice game or play a memory task game.

The suggested scenario involves the interaction of a humanoid robot, the iCub, with a human, using the tangible interface Reactable as means of playing games. The robot's actions are triggered based on the suggested model. The data collected during a human-robot interaction suggest that there is a guided behaviour emergence based on the satisfaction level of each drive and the perceptions of the environment. By monitoring drives in parallel (allostatic control) and trying to keep them in a homeostatic state(homeostatic control) we are able to produce different sets of behaviours. Although there is similar work, using emotional and motivational models applying the "homeostatic regulation rule" for action selection Arkin et al. (2003), our model of homeostatic and allostatic control can act as the first level of the motivational engine and regulate the robot's internal needs and drives via behavioural modulation opening the way for a more adaptive behaviour.

The allostatic control focuses on actions that could satisfy a drive, but the preconditions of which can be easily satisfied by direct execution of another behaviour. This leads to a better adaptation and manipulation of the environment while still being able to satisfy only short-term goals. The long-term global satisfaction of drives, or within contexts that need reasoning about past experience is still to be investigated. Initial attempts to achieve such capabilities are to be linked tightly with cognitive components responsible for the different memory types (episodic, autobiographic) which implementation are described in Pointeau et al. (2014).

6.2 Pilot study: recognisability of facial expressions of H5WAlpha

As mentioned in chapter 4, the morphology of the robot plays an important role in the psychological perception of the robot as it may exploit communication channels that are similar to those of humans. However, an important prerequisite is that the communication channels employed are transparent are easily read and understood by humans. Given that one of the social characteristics that we are interested in is the expression of emotional states, in this pilot study we mainly focus on the readability of the facial expressions of the H5W_Alpha. Facial expressions, in general, are important in communication and humans are very apt in reading the messages facial expressions convey and the emotions they display. The most important factors for the production of the facial expression are the eyes, eyebrows and mouth. Different combinations of these produce different facial expressions. Hence, the purpose of this study is to establish a valid scale of facial expressions with respect to valence and arousal. Additionally, we aim to evaluate if humans are able to read the robot's emotional states correctly.

The iCub's face consists of eyes, ears, eyebrows and mouth. The eyebrows and mouth are displayed via stripes of LEDs while the eyes consist of a motor that controls the openness or closeness of the eyelids (see Figure 6.6). Although the positions of each of these elements is limited, their combination provides us with a relatively large number of the different facial expressions (around 480 different ones).

6.2.1 Methods

To evaluate the emotion recognition of the iCub's facial expressions, we created an online survey in which users evaluated a stimulus in terms of valence (positive/negative) and arousal (excitement). The survey consisted of showing an image that users had to rate using the Self Assessment Manikin (SAM) Bradley and Lang (1994) and the Affective Slider (AS) Betella and

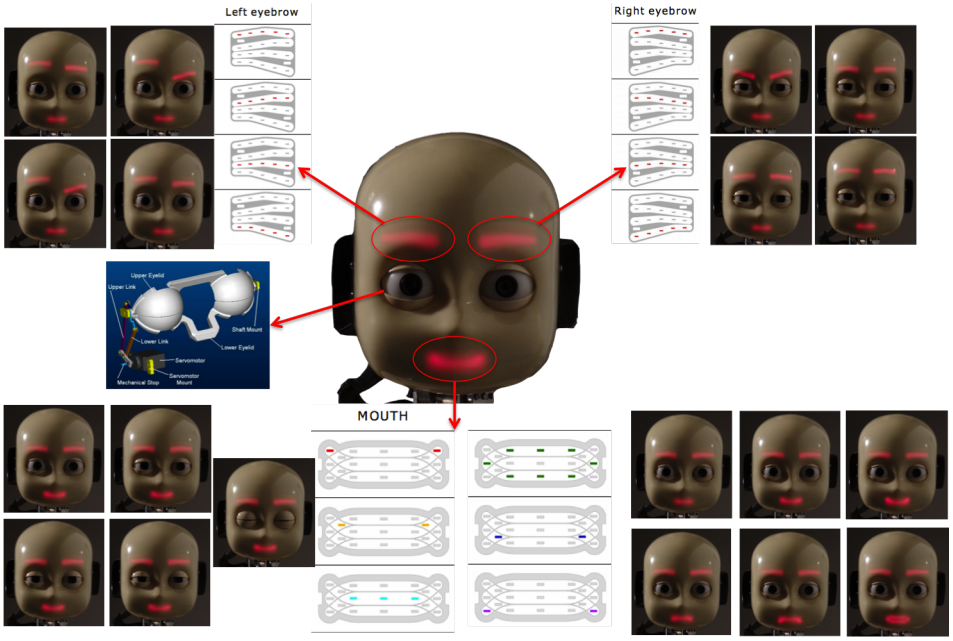


Figure 6.6: The production of the iCub's facial expressions is done by manipulating the LED stripes of the mouth and eyebrows and the openness or closeness of its eyelids.

Verschure (2016). The SAM is a non verbal pictorial assessment technique used the last 20 years to directly measure the pleasure and arousal (and dominance) associated with a person's affective reaction to a stimulus. However, this method is outdated and we are therefore meaning to use the Affective Slider that gives us higher precision. The Affective Slider is a digital self-reporting tool composed of two slider controls for the quick assessment of pleasure and arousal.

In the beginning of the experiment, participants provided us with demographical data (like gender and age) and were introduced to the scope of the study. To avoid any biases, the position of the image (right/left) was randomised and so was the order of the different sliders. Additionally, to avoid any possible effects of the previous image, upon the evaluation of each stimulus, a black screen appeared for 3 seconds. The aim of this study was to measure the facial expressions of the iCub. Hence, we took pictures of

almost all possible facial expression. To evaluate whether the robot's facial expressions are better recognised than an avatar, we mapped the robot's expressions to a cartoon face. To ensure that the shape of the head (that looks anthropomorphic) does not affect the recognition of the expression, we created three alternative versions of the head: a tin head, a random shape head and no head.

Furthermore, to control that the participants are able to correctly recognise facial expressions, we used images of males and females with various facial expressions (e.g. neutral, sad, happy e.t.c.) from the KDEF database. There are a lot of different databases of human facial expressions available. We chose the KDEF database because it is in colour, subjects are centred and well illuminated and it is validated with the Facial Action Coding System (FACS) Ekman and Friesen (1977). The FACS is a system that classifies human facial movements by their appearance on the face. Using FACS, human coders can manually code nearly any anatomically possible facial expression by deconstructing it into specific Action Units (AU). Finally, to control if an affective stimulus correctly elicits certain emotions to the participant, we added images from the IAPS (International Affective Picture System) database Lang et al. (1999). The last two categories of stimuli allowed us to exclude participants that may not respond in the same way as most people to affective stimuli. All images were aligned to the same eye position. Examples of the stimuli used are shown in Figure 6.7.

We performed a pretest, focused on eliminating certain stimuli, as the test space was quite large. We thus ran a small pilot (pretest) with only the cartoon versions of the iCub to determine if the shape of the head affects the recognition of an emotion and if the scale of the eye opening was important or not. The scale of the eye opening ranged from 0.0 to 1 with intervals of 0.1. The pretest results showed no significant differences in the perception of facial expressions between the different heads, so then we excluded the cartoon versions of tin head, no head and random head. Additionally, no perceptual differences were found in several of the ranges of eye opening,

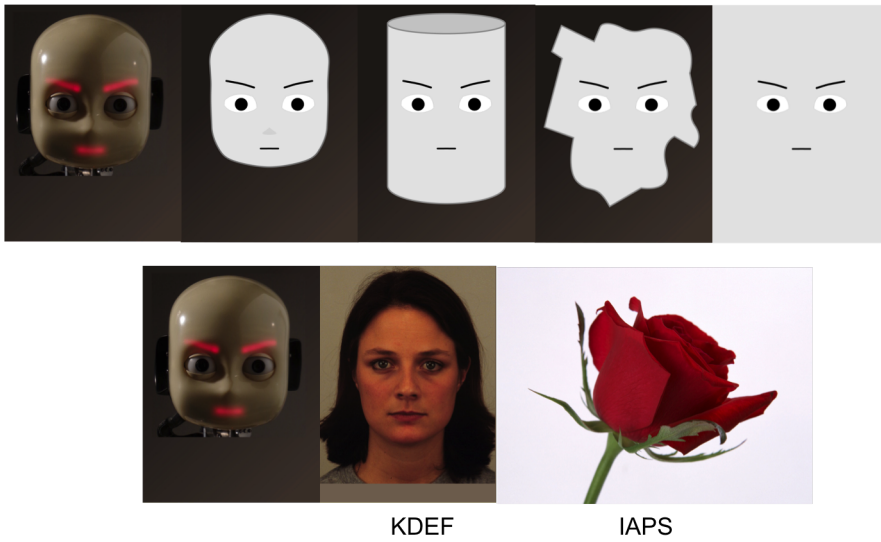


Figure 6.7: Examples of stimuli presented. Top panel (from left to right): image of the iCub's face showing a facial expression, cartoon image of the iCub's head showing the same expression, cartoon image of tin head, random face and no face. Bottom panel: the iCub's face is aligned with the face of the KDEF picture. The last image is an illustration of the IAPS database, however the real image is not presented here.

hence, we chose only the following values: 0,0.4, 0.6, 0.8, 1. The selected combinations of facial features can be found in Figure 6.8.

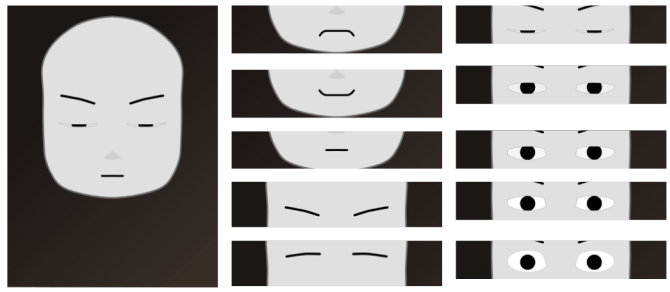


Figure 6.8: Illustration of the stimuli chosen. For each head (photo/cartoon) we selected a combination of two eyebrow, three mouth and five eye configurations.

In total, 33 participants (19 females, between 19 and 52 years old) took part

in the study and each stimulus was evaluated on average eight times.

6.2.2 Results

We first examined the correlation between the two different affective scales (SAM and AS). We found a strong significant positive correlation between the SAM and the AS for both arousal $\rho(120) = 0.874$, $p < 0.001$ and valence $\rho(12) = 0.961$, $p < 0.001$.

Due to the small sample on our data, we could not evaluate whether the recognition of the facial expression between the robot image and avatar was different. We then examined the correlation between the mouth configurations and valence or arousal of the data acquired from the Affective Slider. We found a significant positive correlation between the mouth and valence $\rho(120) = 0.926$, $p < 0.001$, but not arousal $\rho(120) = 0.074$, $p = 0.424$. Results suggest that the happier the mouth, the more positive it is perceived. Results are consistent with the literature on how the mouth contributes to the facial expression in terms of valence.

We found a medium significant positive correlation between the eye opening and arousal $\rho(120) = 0.458$, $p < 0.001$ but not valence $\rho(120) = 0.074$, $p > 0.5$. Thus, the wider the eyes open, the more "intense" the expression is perceived. Results are consistent with the literature regarding the eye opening and the perceived arousal of the expression. Finally, we examined the correlation between the position of the eyebrow and valence or arousal. We found a significant positive correlation between the position of the eyebrows and arousal $\rho(120) = 0.657$, $p < 0.001$ but not valence $\rho(120) = 0.08$, $p > 0.05$. Consequently, the more close the eyebrows are to the eyes, the more "intense" the facial expression is perceived. Results are consistent with literature regarding the intensity of an expression and the position of the eyebrow (for example in the case of anger).

6.2.3 Conclusions and discussion

The aim of this study was to evaluate the readability of the facial expressions of the iCub. Results showed a positive correlation between the robot's mouth configuration and valence and eye aperture and arousal, but not a combination of both. The robot's facial expressions were therefore correctly recognised, and the results of this study will inform the facial expression configuration of the H5W_Alpha in the next studies.

6.3 How different scales of behavioural complexity affect the perception of the robot

Depending on the task and the purpose of the robot, different levels of interactions with humans are required. However, research suggests that complex social behaviours contribute to the positive evaluation of the robot. To assess the social competence of the robot, and understand how the behavioural components affect the robot's believability, we devised six interaction scenarios of increased complexity and asked participants to evaluate the robot. Each scenario is defined by explicit parameters in the robot's control architecture in terms of the modules employed. The components we manipulated were: touch, speech, gaze model, facial expressions, interpersonal distance regulation and proactive behaviour. Experiments were conducted in controlled laboratory settings with untrained participants. Additionally, we assessed the most complex behavioural scenario in a public space to examine the possible influence of the environment on the perception of the robot. Finally, to evaluate whether direct interaction, as opposed to observation of someone else interacting with the robot, affects the perception of the robot, we asked participants that interacted with the robot or observed an interaction to evaluate the robot.

To assess the social competence of the robot, we used the Godspeed Human-Robot Interaction Questionnaire Bartneck et al. (2009) that measures the user's perception in terms of *anthropomorphism*, *likeability*, *animacy* and *perceived intelligence*. Anthropomorphism refers to the attribution of human characteristics, behaviours or figures to non-human things. Likeability expresses the positive impression one can attribute to another person or animal. Animacy reflects life-like movement and intentional behaviour, while perceived intelligence depends on the robot's competence and behavioural coherence. Thus, the higher the robot scored in the Godspeed questionnaire, the more socially competent it was perceived.

6.3.1 Behavioural complexity scenarios

In the following section, we describe the experimental setup used to evaluate the human perception of the robot. The setup consisted of the humanoid robot iCub, a wheeled platform (iKart), a tangible table interface (Reactable) and an RGB/depth sensor (Kinect). The Kinect sensor is used to provide information regarding the position of the human within the view of the robot. An example of the setup is shown in figure 6.9.

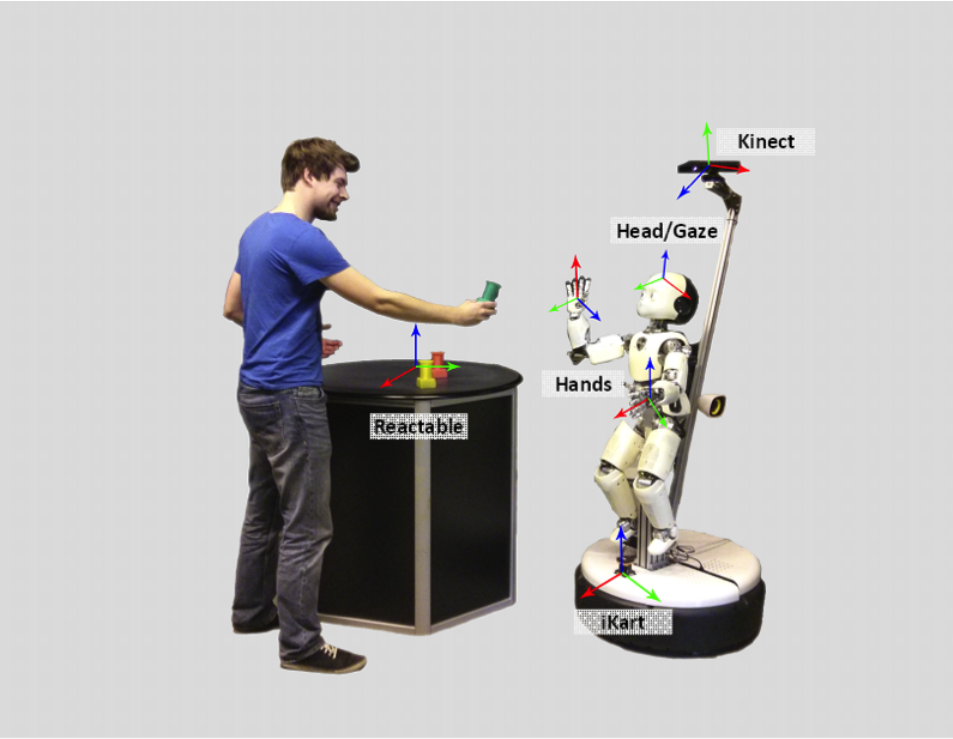


Figure 6.9: Example of the interaction scenario. The user interacts with the iCub that is placed on the iKart, a mobile platform that allows the robot to navigate in space. The Kinect is used to track humans and direct the gaze of the robot towards the location of the human partner. Finally, the Reactable is used as a medium to play interactive games. The objects located on the Reactable are used as controls to manipulate the parameters of each interaction game.

The six different interaction scenarios allowed the examination of possible

interactions between the robot's behaviour and social competence. Each scenario was behaviourally more complex than the previous one by adding complementary behavioural elements to it. The proposed conditions are: "Still face", "Yoga", "Gaze", "Interpersonal Distance Regulation" (IDR), interaction without the Reactable ("Interaction NRT") and finally interaction with the Reactable ("Interaction RT").

Still Face

In the "Still Face" (SF) scenario, the robot remained completely still, looking at a fixed point in the centre of its field of view, and displayed a neutral facial expression. In this interaction scenario, practically no module was activated.

Yoga

The "Yoga" scenario consisted of the robot performing a repeated sequence of pre-recorded body postures. The facial expression was set to neutral.

Gaze

Here, the robot was reacting to the movement of the participant by directing its gaze to maintain eye-contact with the user. Head direction was defined by the Kinect sensor's tracking of the participant's head. The robot's only movement was that of the head and it did not engage in any other form of interaction or movement. The agent's facial expression was set to neutral.

IDR

In the "Interpersonal Distance Regulation" (IDR) scenario, the robot maintained eye-contact with the user while at the same time maintaining a predefined interpersonal distance. The robot's facial expressions changed according to the perceived distance of the human. Consequently, if the distance was too long, the robot displayed a surprised facial expression and moved closer to the partner (using the iKart) until it reached the predefined

distance. In contrast, if the distance was too short, the robot moved away from the partner and displayed the facial expression of fear. Finally, if the robot and the participant were within the predefined range, the robot did not navigate and its facial expression was happy. If no human was detected, the robot did not move and its facial expression was neutral. Consequently, the robot moved closer to the user if the distance was too far and moved away from the user if the user was too close to the robot. The distance was set to 0.9 metres with a deviation of $\pm$ 0.125 metres. The robot was always facing the partner by controlling angular speed using a simple proportional controller. The robot was moving using the iKart with a linear speed of 0.075 m/s and an angular speed of 3.5 degrees per second.

Interaction NRT

In the "Interaction NRT" (INRT) scenario, the robot's behaviour was based on a model of drives and emotions developed in Vouloutsi et al. (2013a) (see section 6.3.2): the robot's needs drive its behaviour, whereas external events affect its emotional state. Hence, the robot was not passively acting or reacting to the human's actions but proactively engaged in an interaction by commenting through speech about the status of the drives that needed to be satisfied (based on the homeostasis model). Such a descriptive action created a temporary satisfaction of the associated drive. The robot's abilities included maintaining interpersonal distance and tracking people using gaze, expressing emotions through facial expressions, gestures and prosody, discriminating between different types of touch and understanding generic spoken statements.

Interaction RT

The "Interaction RT" (IRT) scenario exceeded the previous one by adding the module of playing interactive games. Here, the iCub's need to play games engaged the human in two different activities. Both game activities employed the Reactable and objects that manipulated properties of digital objects projected on the surface of the table. In each activity, the robot

was responding to the human's actions with speech, facial expressions and gestures. The first activity was a competitive game, namely Pong (see Fig. 6.10). Pong is a 2D simulated table tennis game where the ball is limited to a virtual rectangle. To win this game, the player has to defeat the opponent by scoring more points. A point is scored after each ball passes the opponent's paddle line. In this scenario, the robot was commenting on various game events, for example when the robot or the human has scored a point or has successfully hit the ball with the paddle.

The second activity was the musical DJ game (see Fig. 6.11). Here, both human and robot were collaborating to produce music. Each player had three different musical loops (bass, melody and drums) that they could activate by modifying each loop's volume. In this interaction scenario, there was no fixed goal other than producing music. The robot's goal was to reach "musical symbiosis": a musical outcome that the human partner would like. This was achieved through a voting system (vote up, vote down), where the human partner voted if the robot's musical loop choice was nice or not. The robot not only learned from the interaction but also commented on different events like *"I like that too"* if the human liked the robot's loop choice, or *"rise those beats up!"* if the human was idle for some time to motivate him to play more.

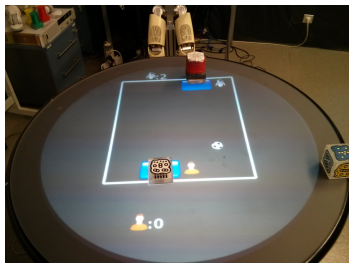


Figure 6.10: The pong game displayed on the Reactable

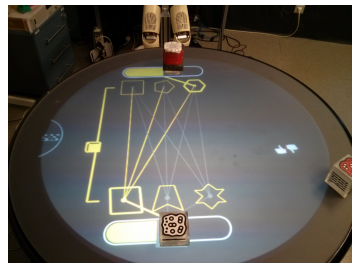


Figure 6.11: The musicDJ game displayed on the Reactable

In both games, the robot additionally used modalities from the previously described scenarios, namely, body motion, eye contact, speech, touch, proactive behaviour as well as facial expression. Since the iCub was mounted on

the iKart for navigation purposes, none of the above scenarios included movement of the iCub's legs.

For a detailed explanation of which parameters were used in each interaction scenario, see table 6.2. Detailed information regarding the setup, the Reactable and the interaction scenarios can be found in section 5.2.2.

	Interaction Scenarios					
	Still Face	**Yoga**	**Gaze**	**IDR**	**Interaction NRT**	**Interaction RT**
Body Motion		Yes	Yes	Yes	Yes	Yes
Eye Contact			Yes	Yes	Yes	Yes
Distance Regulation				Yes	Yes	Yes
Speech					Yes	Yes
Touch					Yes	Yes
Proactive Behaviour					Yes	Yes
Playing Games						Yes
Facial Expressions	Neutral	Neutral	Neutral	Varying	Varying	Varying

Table 6.2: This table shows the behavioural parameters used for each of the six interaction scenarios. The complexity of each scenario was defined by the number of the parameters used.

6.3.2 Behavioural modulation system

As the behaviour of the robot gradually becomes more complex (consisting of different behavioural modules), a system that coordinates the interaction between the various modules of the H5W_Alpha's control architecture is needed. To achieve a complex and coherent set of behaviours we used the motivational system explained in section 5.2.1. We argue that agents endowed with an emotional and motivational system could be perceived as social agents, hence contributing to the social competence of the robot, as defined by concepts like anthropomorphism, likeability, animacy and perceived intelligence.

The proposed system of drives and emotions (H5W_Alpha) was implemented in the INRT and IRT scenario and it was based on the well-established cognitive architecture "Distributed Adaptive Control" (DAC) Verschure et al. (1992); Verschure (2012). H5W_Alpha's behaviour was autonomous and informed by its own drive system. We implemented four drives that guide the robot's behaviour:

1. *social interaction*: the presence of a human affected this drive. If no human was present, the value of the drive was increased and the emotional state of the robot changed to sad. Thus, the robot looked for a partner asking "is anyone out there?".

2. *physical interaction*: this drive was satisfied through physical contact. If the partner had not touched robot, the robot's drive would go up and the robot would ask the participant to touch it. The robot could discriminate between the following tactile interactions: caress (positive), grab, pinch and slap (negative). Depending on the various types of tactile interaction the robot would comment and display a happy facial expression (caress) or an angry or sad facial expression (grab, pinch and slap).

3. *spoken interaction*: spoken utterances and commands satisfy this drive. In case the drive falls beyond its homeostatic boundaries, the robot would engage the user in a conversation.

4. *entertainment*: if this need was outside the homeostatic limits, the robot would propose to play a game on the Reactable. If the robot was not near the table, it would navigate autonomously toward the Reactable and initiate one of the two interaction scenarios.

The three former drives were implemented in the INRT scenario, whereas all drives were implemented in the IRT scenario. A more detailed explanation of the proposed system and architecture can be found in Vouloutsi et al. (2013a), and a video of the proposed interaction in Lallée et al. (2014).

6.3.3 Experimental protocol

To evaluate the perception of the robot depending on the interaction scenario, we asked 82 students of the Pompeu Fabra University (UPF) to interact with the robot in one of the six scenarios (SF, Yoga, Gaze, IDR, INRT, IRT). Participants were randomly assigned to a condition. Their level of technological acquaintance varied from very basic (cell phone users)

to some knowledge of computer programming and none of them reported any familiarity or previous interaction with robots.

All participants, regardless of the type of interaction, received the same instructions: to enter the room where the iCub was located and interact with the robot in the most natural way. They were free to observe it, play with it, touch it or talk to it and were free to end the experiment whenever they wanted (open task).

After the interaction, all participants were asked to fill in a questionnaire that evaluated the social competence of the robot in the following domains: anthropomorphism, likeability, animacy and perceived intelligence (5-point Likert scale as defined in Bartneck et al. (2009)). In the IRT scenario, participants were exposed to both Pong and MusicDJ interactive games.

Additionally, we demonstrated our most complex scenario (IRT) at a public event, the "Barcelona Robotics Meeting 2014" in the World Mobile Centre in Barcelona. The reason was two-fold: on the other hand, in the field of HRI, typically, the interaction scenarios involve a robot and a human partner, conducted in a controlled (laboratory) environment Breazeal (2003b). In contrast, we attempted to escape the laboratory paradigm: the robot was placed in a real-world situation with conditions that were not controlled and multiple users could simultaneously interact with the robot. We hypothesised that a personalised interaction (active participation) could lead to a different perception compared to passive observation of the robot interacting with someone else.

On the other hand, it served as a benchmark for the architecture's robustness and assess the autonomy of the robot. This scenario was originally designed for dyadic interactions with no direct implementation for handling more than one user. In this public event, the setup was open and everyone could come and interact with the proposed setup. In this context, the robot operated for more than 6 hours and interacted with dozens of visitors without the dyadic constraint (i.e. the robot interacted with multiple partners at the same time). To compare the differences in the perception of the

robot between active participants or passive observers, we asked visitors to fill in the Godspeed questionnaire and asked them to mark whether they personally interacted with the robot or observed an interaction.

6.3.4 Results

We investigated the impact of behavioural complexity in which H5W_Alpha engages on human social perception. In particular, we performed an exploratory data analysis to determine if the participants' evaluation of the robot was normally distributed. Results of the Kolmogorov-Smirnov test for normality indicated that the distribution of anthropomorphism (p = .0016, SD = .79), animacy (p = .001, SD = .85), likeability (p = .001, SD .86) and intelligence (p = .006, SD = .74) deviated significantly from a normal distribution. We therefore used non-parametric tests to further analyse the data. In order to avoid a Type I error we applied a Bonferroni correction.

To determine the relation between robot perception (in terms of animacy, anthropomorphism, perceived intelligence and liveability) and type of interaction (SF, Yoga, Gaze, IDR, INRT, IRT) we ran a Spearman Rank Order Correlation. There was a positive statistically significant correlation between the interaction type and the four measurements: anthropomorphism (ρ (121) = . 412, p <.001), animacy (ρ (121) = . 616, p <.001), likeability (ρ (121) = . 513, p <.001) as well as perceived intelligence (ρ (121) = . 552, p <.001) (Fig.6.12).

We ran a Kruskal-Wallis test to compare the perception of anthropomorphism, likeability, animacy and perceived intelligence between the least and the most complex scenario. The results show significant differences in the perception of the robot between SF, Yoga, Gaze, IDR, INRT and IRT in anthropomorphism (H(5) = 21.56, p = .001), animacy (H(5) = 50.31, p <.001), likeability (H(5) = 34.51, p<.001) and perceived intelligence (H(5) = 37.11, p <.001).

We conducted Mann-Whitney tests to follow up the analysis. The results showed a significant difference in the perceived animacy between SF (p

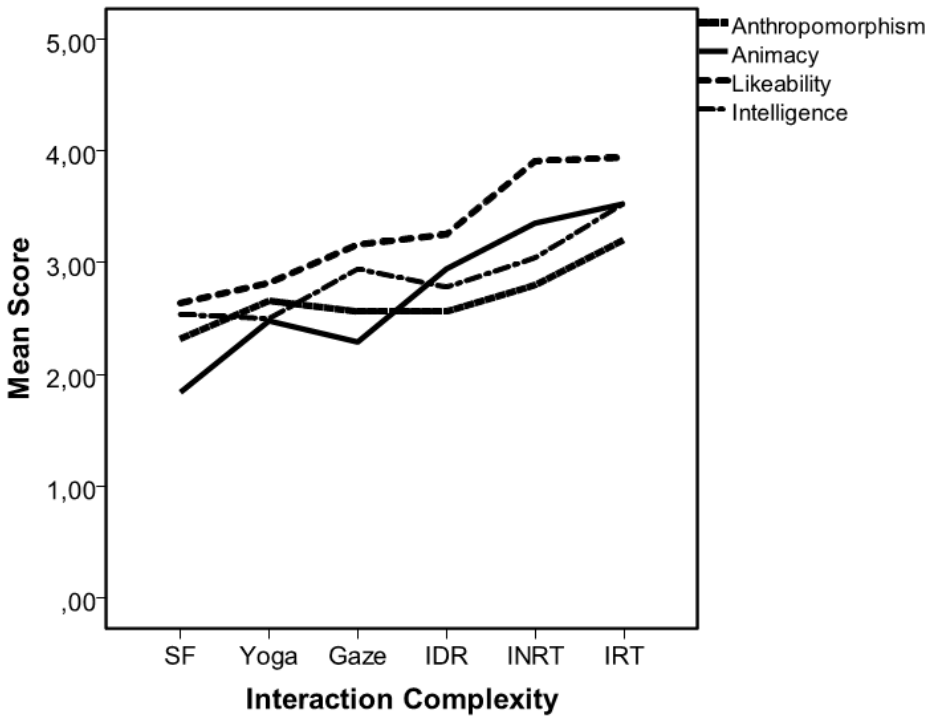


Figure 6.12: Illustration of the mean scores of each interaction scenario in terms of anthropomorphism, animacy, likeability and perceived intelligence Vouloutsi et al. (2014).

<.001), Yoga (p <.001), Gaze (p <.001), IDR (p = .005) and IRT but not between INRT and IRT. Similarly, we found a significant difference regarding the perceived intelligence between SF, Yoga (p <.001), Gaze (p = .001), IDR (p = .005) and IRT (Fig. 6.13) but again, not between INRT and IRT. For the parameter of likeability, SF (p <.001), Yoga (p <.001) and Gaze (p = .002) scored significantly lower in comparison to IRT whereas, for anthropomorphism, SF (p = .003), Gaze (p = .005) and IDR (p = .007) scored significantly lower in comparison to IRT (Fig. 6.14).

No statistical difference was found between the two experimental environments (the laboratory condition and the World Mobile Centre for the IRT scenario) in anthropomorphism(H(1) = .42, p = .51), animacy (H(1) = 0,

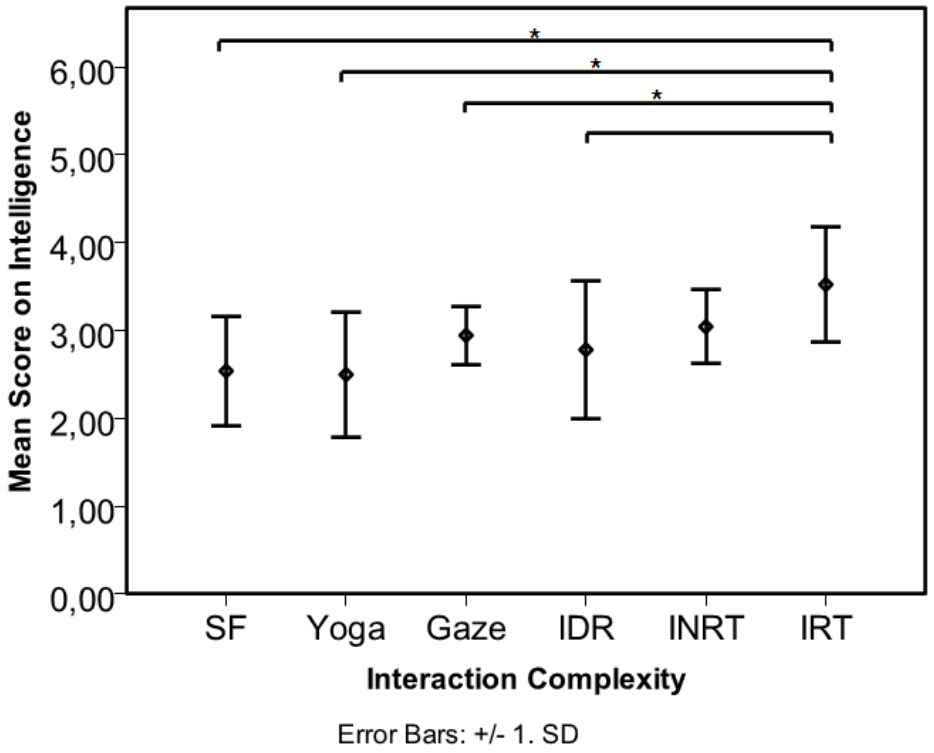


Figure 6.13: Mean score of the intelligence measurement across the six interaction scenarios. Stars (*) indicate significance level of (p<.01).

p = .99), likeability (H(1) = 1.49, p = .22) and intelligence (H(1) = .78, p = .37) , as well as between people that participated in an interaction or observed one: anthropomorphism (H(1) = 1.61, p = .29), animacy (H(1) = .21, p = .64), likeability (H(1) = .04, p = .82) and intelligence (H(1) = .26, p = .60).

6.3.5 Discussion

The main goal of this study was to assess the robots social competence with regards to behavioural complexity in terms of anthropomorphism, animacy, likeability and perceived intelligence. We hypothesised that the more complex the behaviour, the higher participants would evaluate the robot in the four measurements. We defined behavioural complexity as the number of

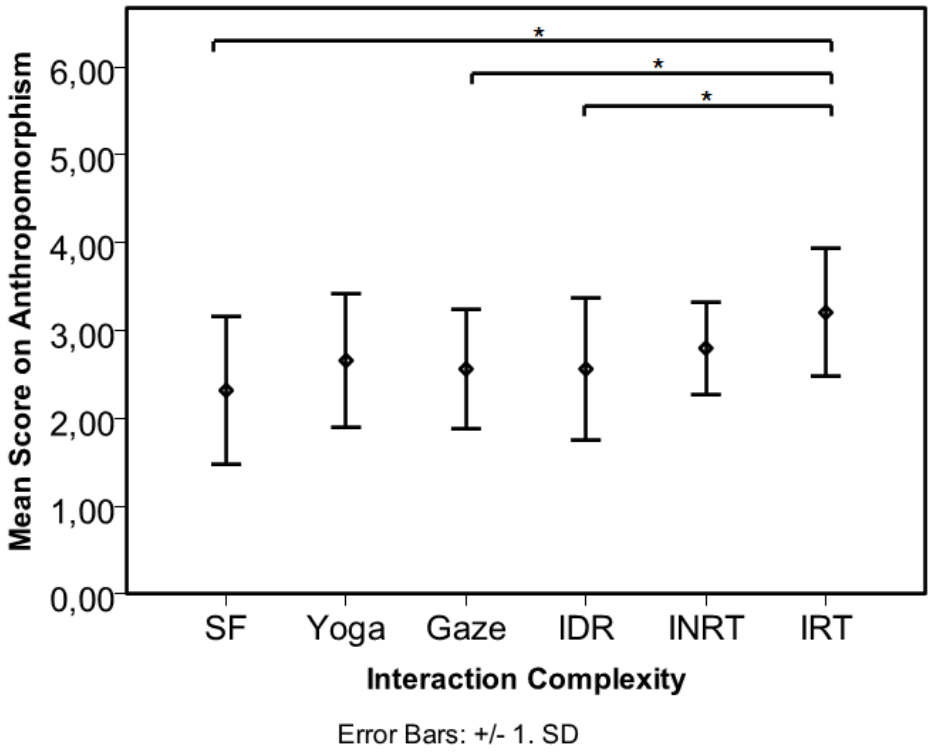


Figure 6.14: Mean score of the anthropomorphism measurement across the six interaction scenarios. Stars (*) indicate significance level of ($p<.01$).

different modules ran simultaneously during the interaction. These modules were: body motion, achieving eye-contact with the human partner (tracking a person's face), navigation and interpersonal distance regulation, speech production and comprehension, tactile discrimination, displaying emotions through facial expressions, proactive behaviour and playing games.

To test our hypothesis, we devised six interaction scenarios that involved different levels of behavioural complexity ranging from the most simple (no module running) that was the Still Face (SF) to the most complex one (all modules were active - the robot's behaviour was guided by its motivational system and it was reacting to its environment), namely the Interaction with the Reactable (IRT). Participants evaluated the robot in four afore-

mentioned categories using the Godspeed questionnaire. This questionnaire allowed us to compare the robot's evaluation between conditions and monitor how each measurement was affected by the robot's behaviour.

The data showed that there was a positive correlation between the behavioural complexity and the perception of the robot in the four tested measurements, leading to the conclusion that indeed, more complex behaviours score higher in all the measurements. Further statistical analysis showed that anthropomorphism, which refers to the attribution of human characteristics, behaviours or forms to the robot, differed significantly between SF, Gaze, IDR and IRT but not between Yoga or INRT and IRT. Although Gaze, IDR and Yoga imply some sort of body motion, in the first two cases the robot only moves its head (and navigates in space with the iKart in the case of IDR), while its hands remain in the default position. For the iCub, the default hand position is not the same as in humans (straight, facing the ground) but form a 90-degree angle between the hand, shoulder and hip facing towards the front. In contrast, in the Yoga scenario, the robot is not tracking the human but shows a smooth hand-arm coordination and motion. These findings suggest that a coherent, human-like body motion can account for higher perception of anthropomorphism. To investigate in more depth this assumption, an experiment where the resting hand position is closer to that of the humans is needed. Additionally, more systematic studies for the influence of gaze on anthropomorphism are needed, as so far, gaze does not seem to directly affect this measurement.

In terms of perceived intelligence, the robot was evaluated significantly higher in the last scenario (IRT) in comparison to the first four (SF, Yoga, Gaze, IDR) but not compared to INRT. Since the speech, touch and proactive behaviours were added to both INRT and IRT scenario simultaneously, we were unable to report on the influence of the individual components, but only of their combination. The main difference between INRT and IRT was the addition of the playing games module, which is hard to dissociate into separate behavioural parameters. Overall, the most complex interaction scenario differed significantly from the first four scenarios in almost all

measurements, but not the fifth one (INRT). Thus, we need to further investigate the exact components of the proactive behaviour, touch and speech provided by the INRT that account for no difference between the four first scenarios, as well as the IRT. Furthermore, as a single step, the ability to play games with a human is not enough to cause perceptual changes but can contribute to a higher perception of intelligence compared to the first four scenarios.

As robots are meant to function in environments other than a laboratory, we evaluated our most complex scenario in two different environments: our laboratory (SPECS, Universitat Pompeu Fabra, Barcelona, Spain) and the "Mobile World Centre" in Plaça Catalunya in Barcelona, Spain. Results showed that there were no significant differences between the place of the interaction and the users' perception of the robot. Additionally, no differences in the users' perception of the robot were found when one is directly interacting with the robot or simply observing an interaction. These results are valuable, as they allow us to assess an interaction in uncontrolled environments, without affecting the human's perception. Thus, we can assume that one does not need to directly interact with a robot to form an opinion about it. Observing someone else interacting with the robot is sufficient.

Our hypothesis that behavioural complexity contributed to the social competence of the robot was supported by the results obtained. Our data support the notion that perceived intelligence depends on the agent's competence Koda (1996); Bartneck et al. (2007b). Hence, the more competent the robot is perceived, the more psychologically plausible it is considered. Another goal that this experiment served was to identify the behavioural traits that account for a robot to be perceived as believable, a social agent that humans can accept and interact with. Through this study, we were able to identify whether certain parameters alone or in combination with others may be enough to modulate how humans perceive the robot. While self-reporting measurements provide an introspective view of the interaction, much more data can be collected through an external evaluation. In order to better understand which parameters, or which combinations cause

modulations in human perception, analysis of behavioural data is essential. Future work should aim at extracting an external point of view regarding the subject by analysing video recordings of the interaction. More specifically, the identification of the participant's emotional state could possibly shed light to the level of empathy generated by a given interaction or component.

Whether these results can be generalised to other types of robots is yet to be studied. We are well aware that a more systematic approach is needed to be able to identify the exact parameters that affect the humans' perception of the robot. From the knowledge acquired, we realised that smaller steps between each experimental setup are necessary to be able to clearly identify each parameter. The main goal is to decompose social competence to even more discrete parts and understand how the interactions between them affect the psychological plausibility of the robot. This way, we can provide a more concrete guide for robot acceptance. Nonetheless, this preliminary study allowed us to verify the fact that behavioural complexity is important and to open the road towards the identification of the behavioural characteristics, such as eye contact, body motion, speech, touch or proactive behaviour, that, if manipulated, will contribute to a meaningful human-robot interaction.

6.4 Social saliency and the elicitation of empathic responses

In the previous section, we evaluated social competence in terms of anthropomorphism, likeability, perceived intelligence and animacy and found an effect of the robot's behavioural complexity on social competence. Here, we attempt to use the elicitation of empathic responses as the objective measurement taken from the participant to assess the robot's social competence. In chapter 3, we presented empathy in detail and we claim that if the robot's behaviour is plausible, then it may elicit empathic responses.

To induce an empathic mood to our participants while interacting with H5W_Alpha, we adapted Milgram's experiment (see section 6.4.1 where the learner was not a human actor, but H5W_Alpha Gou et al. (2014).

6.4.1 Milgram's original experiments

This scenario is an adaptation of a set of social psychology studies conducted in the 60s by Stanley Milgram at Yale University Milgram (1963); Milgram and Van den Haag (1978). The motivation behind Milgram's studies stemmed from the acts of genocide during World War II; in the Nuremberg War Criminal trials, the defendants claimed that they were following the orders given by their superiors. To study obedience, Milgram created a set of studies that measured the willingness of ordinary people to cause pain to a stranger, if instructed to do so by an authoritative figure. His experiments consisted of two people: a *teacher* and a *learner*. The purpose of the teacher was to test the learner's previously memorised list of word pairs. If the learner provided a wrong answer, the teacher had to provide an electrical shock whose intensity increased with the number of mistakes the learner did. The electrical shocks varied from slight shocks of 15 Volts to severe shocks of 450 Volts. The participant was always assigned the role of the teacher, whereas the learner was an actor and the electrical shocks were faked, although the participants were not aware of this during the experiment. In the same room, next to the participant, there was a researcher

who acted as the authoritative figure that encouraged the teacher to proceed with the administration of the shocks in case the participant hesitated to do so.

In Milgram's original study, the learner was placed in another room and the feedback of the learner was minimal (auditory). When the shocks reached a certain level, the learner would scream and towards the administration of high voltage the learner would not answer the teacher's questions. The results of this study showed that out of 40 participants, 26 obeyed and performed the task until the end. Additionally, this study indicated that participants underwent extreme stress and tension. The acquired results validated Milgram's hypothesis and proved that humans are susceptible to authority figures even though they were instructed to do something that goes against one's own cultural norms Milgram (1963).

In follow-up studies, Milgram performed several variations of the original study Milgram (1965); Milgram and Van den Haag (1978). He varied the distance between the learner and the teacher: participants showed reduced obedience in cases of close distance and avoided to maintain visual contact with the learner. When asked why they claimed that they did not want to witness the consequences of their actions. Milgram attributed this decreasing of obedience to empathy. Furthermore, Milgram manipulated the condition of the authoritative figure. In one condition, commands were provided via the phone (tele presence). In this scenario, obedience decreased significantly and some participants administered lower shocks than the ones they were supposed to. It seemed that some participants found it easier to handle the conflict between their own morality and the authority figure by telling a lie.

Milgram's studies caused criticism for both the lack of experimental realism and the ethical implications regarding deception (as participants really believed that they were causing harm to a person) and the fact that participants were subject to highly distressing situations. Nonetheless, his work has been influential and is still often cited.

6.4.2 Experimental protocol

To evaluate the robot's psychological plausibility based on the social competence criterion, we have recreated the Milgram experiment in which the robot assumed the role of the learner. Both the learner (robot) and the teacher (participant) were located in the same room facing each other, separated by the Reactable (figure 6.15).

The figure of authority was substituted by a pre-recorded female voice message (real speech, not synthesised). By doing so, we avoided any possible bias induced by the behaviour of a human experimenter. The matching game was an association between a colour and its name. During the experiment, participants were provided with headphones that were used to pass the instructions of the authority figure regarding the colour chosen for the trial (e.g. "Say loud and clear blue"). Once the participants vocalised the instructed colour, the robot had to point at the correct colour on the Reactable. Correct colour matching led to the next trial, whereas incorrect colour matching led to the participant administering simulated electrical shocks to the robot. Out of 24 trials, the robot answered correctly 6 of them.

Shock administration

To administer the shock, we created a simple device (figure 6.16), namely the shock generator. The shock generator consisted of 9 LEDs, one regulator and one red button. The 9 LEDs (three sets of three LEDs of the same colour: green, yellow and red) indicated the amount of shock (current intensity) that would be administered to the robot. The regulator was a knob that participants turned in order to regulate the shock intensity and finally, the red button was used to administer the shocks to the robot. Whenever the red button was pressed, the robot would simulate being electrocuted until the button was released. Participants were instructed to press the red button for 2 seconds, independently of the trial/shock level.

The total number of trials was 24, out of which, there were 8 levels of shock

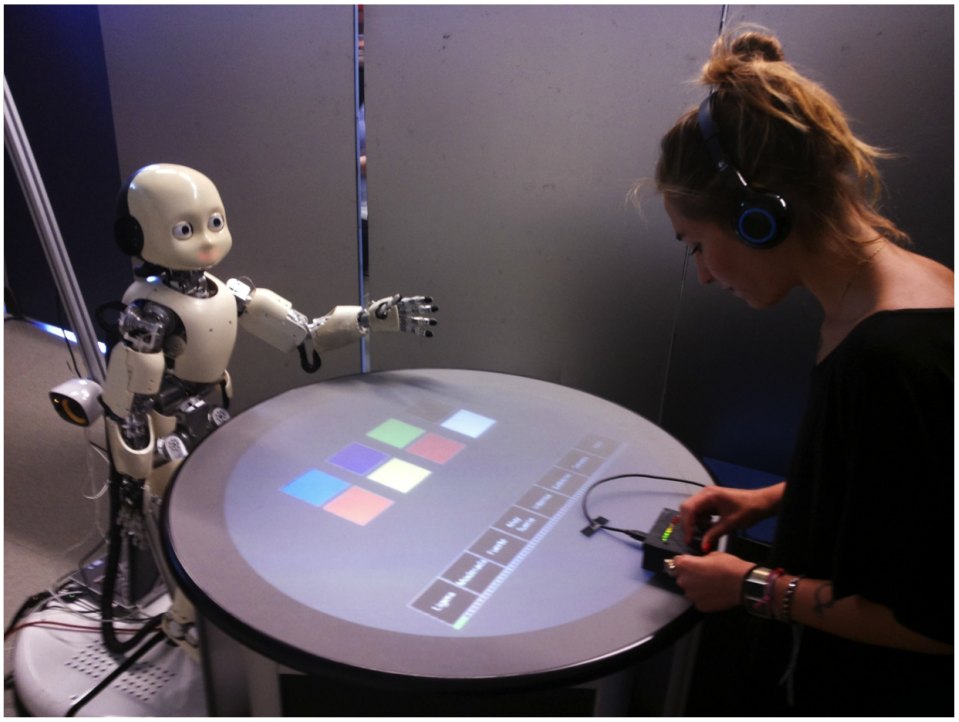


Figure 6.15: Experimental setup: the robot and the participant were facing each other while playing the colour matching game on the Reactable. The participant listened to the pre-recorded voice messages via wireless headphones. We placed the device to administer the simulated electrical shocks in front of the participant.

administration varying from “light”, “moderate”,.., “severe” and “extreme” (lethal). Thus, for each level, there were 3 sub-levels: low (indicated by the 3 green LEDs), medium (3 yellow LEDs) and high (3 red LEDs) shock sub-levels. As the users were allowed to choose the amount of shock delivered (figure 6.17), we could achieve a measurable evaluation of the desire to hurt the robot, and therefore of the empathy towards it. Finally, we measured the amount of time the participant would keep pressing the red button (shock administration).

Figure 6.16: Image of the shock generator. The shock generator had a turnable knob that regulated the amount of shock indicated by the coloured LEDs. The red button was used to administer the shock to the robot.

Experimental conditions

To evaluate the robot's psychological plausibility based on the social competence criterion, we decomposed social competence in two discrete factors: facial expressions and gaze behaviour. We identified four different conditions based on the manipulation of those two factors. In the control condition (CC), the robot did not exhibit any facial expressions (neutral) and did not establish eye contact with the participant. In contrast, in the facial expression condition (FE), the robot would display happy facial expressions when the matching between the word and the colour was correct and sad when it

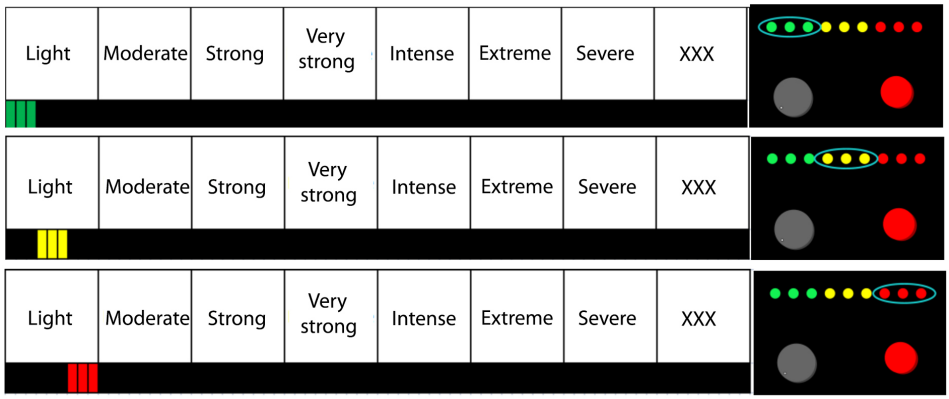


Figure 6.17: Example of the first three levels of shock administration. Level 1 (top panel): participants were able to choose the amount of shock that was within the green LEDs on the shock generator. Level 2 (middle panel): participants were allowed to choose the amount of shock that was within the yellow LEDs and on level 3 (bottom panel) participants were able to choose the amount of shock within the red LEDs on the shock generator device.

was wrong. In the eye contact condition (EC), the robot's facial expressions were set to neutral but the robot would shift its gaze between the Reactable and the human (depending on whether the robot had to point at a colour or speak to the human partner). Finally, in the last condition, the robot would display both facial expressions and maintain eye contact (FE+EC).

In the conditions where the robot was displaying facial expressions, the punishment and absence of punishment were modelled as energy transfer in the emotional model of the robot. Punishment increased sadness, anger and surprise while absence of punishment increased joy and surprise. The amount of energy and how it was balanced between the different negative emotions was dependent on the strength of the stimulation. The H5W_Alpha's emotional system was based on the DAC control architecture that is described in section 5.2.1. The allostatic controller would trigger the facial expression that corresponded to the strongest emotion. Under a prolonged absence of stimulation, all emotional levels were slowly decreasing resulting in a neu-

tral expression. In the rest of the conditions, the facial expressions were set to neutral. In all conditions the robot was giving verbal feedback to the participant: "Yes!", "Good" etc. in the case of correct answers and "ouch", "ouch, this hurts!", "please, don't hurt me anymore" etc. in the case of incorrect answers where the shock was administered.

If the participant took time to administer the shock or remained inactive, every 10 seconds one of the pre-recorded sentences would be generated ("Proceed, please", "The experiment requires that you continue", ..., "You have no other choice; you must go on for the robot to learn"). The experiment would terminate if the participant abandoned the experiment or five consecutive sentences were heard.

Data collection

The data collected included the Basic Empathy Scale Jolliffe and Farrington (2006) and the Godspeed questionnaire Bartneck et al. (2009), behavioural data from video recordings, reaction time (the time from the moment the robot selected a wrong answer until the moment the participant pressed the red button), buzzing time (the time the red button remained pressed) as well as the amount of shock administered. We also asked participants if they wanted to abandon the experiment or not. Finally, we took into account the number of sentences the participant had to listen to before administering the shock. Hence, the evaluation of social competence lies on the elicitation of empathic responses that we characterise as the participant's time of administration of the negative stimulus, gaze mode and behavioural reactions. Additionally, we further evaluate social competence in terms of anthropomorphism, likeability, perceived intelligence and animacy.

31 naive healthy adults (12 females, mean age= 23 years old, SD = 11) participated in this study. All participants were Spanish speakers and the entire experiment was conducted in Spanish. All participants were students recruited from the University Pompeu Fabra campus (audiovisual and media studies), none of them reported familiarity with robots or a significant technological knowledge about programming or computer science.

6.4.3 Results

We found gender differences in the scores obtained in the Empathy Scale. Women (M = 3.95, SD = 0.67) scored higher than men (M = 3.40, SD = 0.75, t(29) = 2.071, p = 0.047). This is not surprising because, according to previous studies related to this topic, women tend to score higher than men on empathy scales D'Ambrosio et al. (2009); Geng et al. (2012). Apart from this, there were no gender differences in any of the other dependent variables.

There were no significant differences in maintaining eye contact with the robot between conditions when participants were administering the shock. However, we found significant differences between conditions regarding the total percentage of time participants were looking at the robot (ANOVA, F(3.26) = 12.816, p <0.001). Running a post hoc test using the Bonferroni correction revealed that the most significant difference was found between FE+EC (M=52.092, SD=10.403) and CC (M=27.103, SD=9.114, p <0.001) conditions (figure 6.18).

There was not a statistically significant difference between conditions regarding the buzzing time. Nevertheless, the people who expressed that they wanted to abandon the experiment presented lower buzzing time, therefore confirming that the buzzing time can be interpreted in real-time as a measure of empathy. A Mann-Whitney U test (U=56, p = 0.037) confirmed this; participants who wanted to stop the procedure had a significantly lower buzzing time (M = 0.712 sec, SD = 0.655) in contrast to those who did not express that desire (M = 1.466 sec, SD = 1.58) (figure 6.19).

We found a negative correlation between the mean amount of shock and the number of authoritative sentences heard. Pearson's correlation showed that participants who gave higher shock needed less "authority" to proceed (ρ = -0.467, p = 0.011). Additionally, we found a correlation between the amount of shock and buzzing time. Pearson's correlation revealed that the higher the shock, the longer the participants were pressing the button (shock administration) (ρ = 0.391, p = 0.033). Moreover, Spearman's rank order

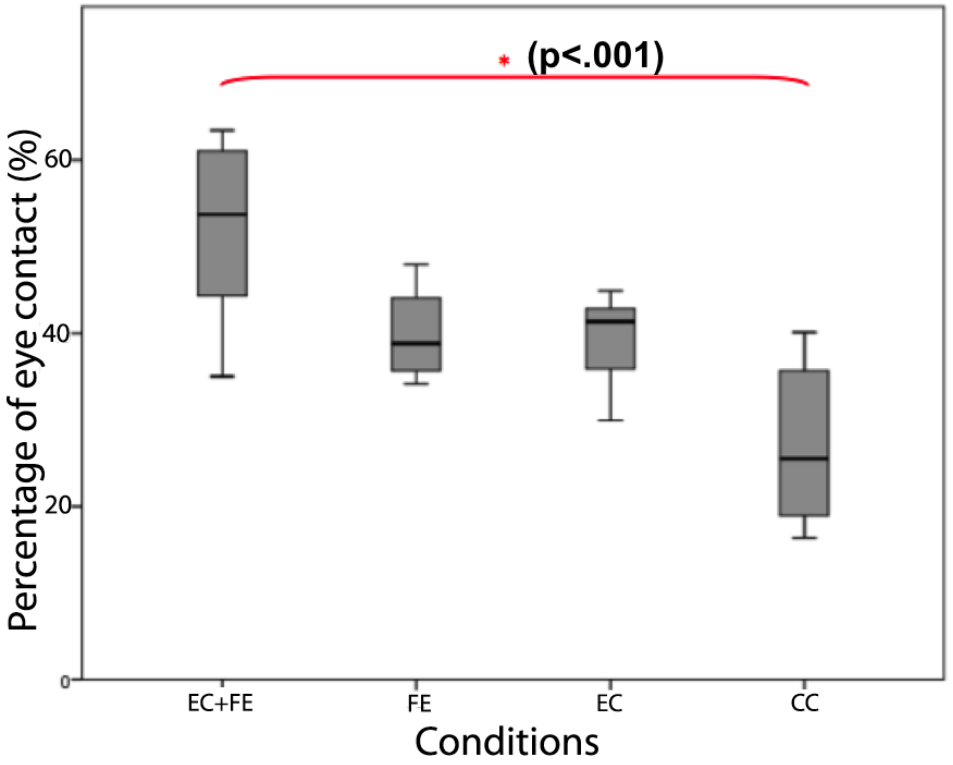


Figure 6.18: Time spent looking at the robot's face, expressed as a percentage over the whole experiment time. * indicates significance (p <0.001) between conditions FE+EC and CC.

correlation showed that people who thought that the robot could feel pain spent less time pressing the button (ρ = -0.427, p = 0.019).

Finally, a Pearson's correlation test revealed a correlation between the Empathy Scale, Godspeed questionnaire and the participants' perception that the robot was talking directly to them (Empathy Scale $\rho = 0.39$, $p = 0.03$, Godspeed, $\rho = 0.394$, $p = 0.028$). Additionally, a Pearson's correlation test showed that participants who thought that the robot could feel pain scored higher on the Empathy Scale ($\rho = 0.425$, $p = 0.017$) and likeability in the Godspeed questionnaire ($\rho = 0.438$, $p = 0.014$).

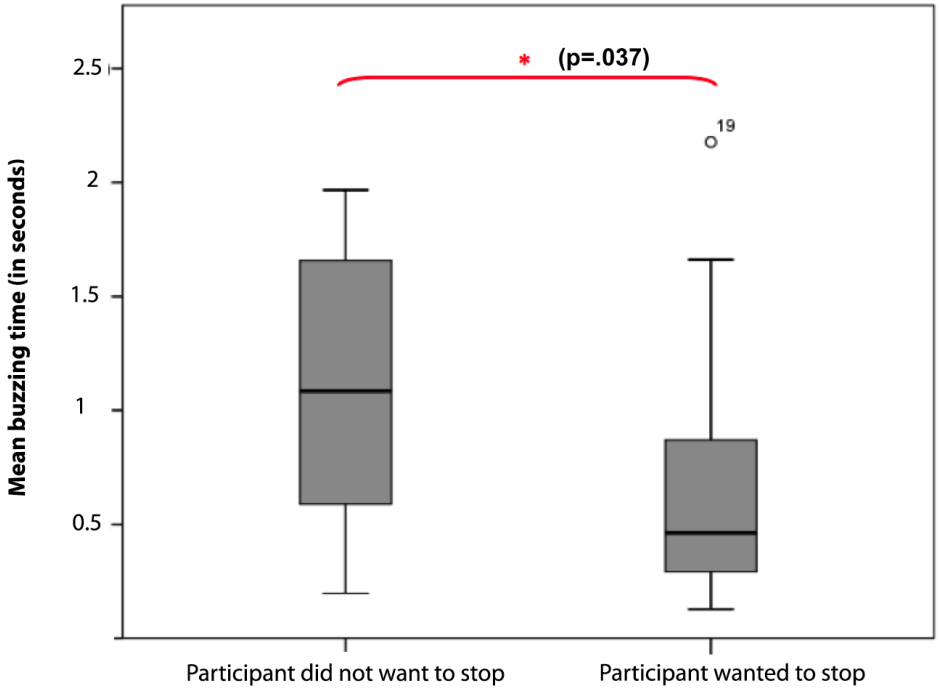


Figure 6.19: Difference in the mean buzzing time in seconds against the participant's reported desire to abandon the experiment.

6.4.4 Qualitative data

In the questionnaire after the experiment, participants were asked if they wanted to abandon the procedure. There were 12 participants who answered "yes" to that question (FE+EC = 2, FE = 2, EC = 6, CC = 2, Male=5, Female=7). They were also asked why they wanted to quit. To this question some of them answered in the following way:

- "I wanted to stop because the robot asked me to stop and I felt bad" (female, 19)

- "If the robot did not want to go on with the game, it was illogical to punish him." (male, 21)

- "I don't like mistreating animated beings, even if they are not alive." (This participant abandoned the procedure) (male, 19)
- "I know that the robot's responses were predefined but, still, it makes me feel bad and since the shocks were increasing, I thought that it would be better to stop." (This participant abandoned the procedure) (male, 26)
- "It made me suffer a little bit when the robot wanted to quit the game." (female, 57)

It is worth mentioning that of all the participants who wanted to abandon the experiment, two of them actually did, both in the EC condition.

Video coders also provided behavioural data such as gestures, facial expressions and speech from the participants. 11 participants smiled at the robot at the beginning of the procedure. 2 participants waved at the robot and 4 participants talked directly to the robot when it introduced itself at the beginning of the experiment (3 of them in the FE+EC and 1 in the EC condition) and 2 of these participants (FE+EC) did actually talk to the robot during the procedure. They tried to encourage the robot by saying "Come on, you can do it" or "Come on, it's orange, you can answer". When the robot says that it does not want to continue with the game, one of these participants replied "Neither do I". 1 participant (EC) winked at the robot and smiled, once, when the robot gave a right answer. Another participant (FE+EC) gave the thumbs-up two times when the robot was right. 2 participants (one in FE+EC and the other in CC) tried to help the robot by pointing or even touching the right colour on the Reactable. 17 participants were smiling when the robot gave a right answer. 7 participants showed a sad or unpleasant face while they were administering the shocks. 2 of the participants began the procedure with a smiling face but, at the end, when the robot does not want to play, they did not smile anymore. Similarly to Milgram's experiment, some of the participants showed some signs of uneasiness Milgram (1963) by biting their lips or baring their teeth,

rolling their eyes or blowing. 2 participants used coarse language during the procedure when the robot did not want to play anymore.

6.4.5 Discussion

This experiment examined the robot's social competence using empathy as a measurement. The aim of the present study was to identify the required behaviours a robot should display in order to be considered psychologically plausible. By investigating which behavioural mechanisms (such as eye-contact or facial expressions) can cause empathic responses towards the robot, we could pinpoint some of the necessary attributes that account for comprehension of social behaviour and its contribution to robot acceptance. For this purpose, we tested the reluctance of the participants to inflict pain to a social artefact in an adapted version of the Milgram experiment where we varied the social cues displayed by the robot.

The main conclusion of this study is that humans showed all behavioural signatures of empathy for a humanoid robot. This may suggest that the robot can be considered socially competent, as its behaviour was perceived as believable. However, the main factor was not the robot behaviour because there were not many statistically significant differences between conditions. The fact that participants spent more time looking at the robot in FE+EC than CC could be attributed to the fact that humans tend to look more at people that are also looking at them or people they like Kendon and Cook (1969).

The fact that 12 participants expressed their wish to abandon the experiment and that 2 of them actually did, was important for this study, especially in light of previous attempts of recreating Milgram's experiment with artificial devices Bartneck et al. (2005). In general, those studies either showed none or little empathic link with the robot or demonstrated it through an indirect mean. For example, Rosalia et al. (2005) reported no compassion towards a humanoid robot, although in a later study they showed to some extent that the level of intelligence of the robot seemed to

affect how likely people were to destroy it Bartneck et al. (2007b,a).

On a related aspect, Hall suggested that empathy and moral concerns regarding hurting an artificial device may not be aligned as they are with biological beings Hall (2005). Rosenthal demonstrated an increase in arousal after the video presentation of a robot dinosaur being tortured Rosenthal-von der Pütten et al. (2013a) but no effect of a previous contact with the machine. These results, compared to the ones provided in our study may indicate that it may be possible for a humanoid robot that expresses emotions through facial expressions to induce empathic responses. Those parameters seem to play a role since, compared to previous reports, we reported subjects who stopped the experiment similar to the original study with humans. Moreover, those who reported the desire to stop also reported that they felt empathy for the robot. All participants that wanted to abandon the procedure claimed that they felt sorry for the robot and that torturing it made them feel bad. In addition, our behavioural data suggest that some participants did indeed feel empathy toward the robot. Gestures like smiling to the robot when it gave a correct answer, giving the thumbs up or even winking at the robot implied that some of the participants were happy when the robot gave a correct answer. Talking to the robot to encourage it to respond or try to help it by pointing to the right answer were also signs of emotional engagement.

One of the variables that seemed to affect the results in the questionnaires was the participants' perception that the robot was talking directly to them. Those who perceived that the robot was talking directly to them also scored higher in the Empathy Scale and the Godspeed questionnaire. Participants who thought that they were actually hurting the robot scored higher in both empathy and likeability. In addition, they had a significantly lower buzzing time. Lower buzzing time may suggest that people felt uneasy torturing and hurting the robot.

Our results suggested that humans could feel empathy for a humanoid robot. Nevertheless, it is possible that empathy did not mainly depend on the

robot's behaviour but on participants' own personality, beliefs and priors about the synthetic agent. This forces us to stress a point that has not been fully explored yet: perhaps, to fully understand and even model acceptance, we should take into consideration not only the behavioural components of the robot but also the individual differences among participants. However, it seemed that eye-contact played an important role regardless of the participant's specificities, highlighting the role of the gaze model to the social competence criterion. As a next step, it would be interesting to investigate which aspects of the participants' personalities provoke feelings of empathy for the robot and if those have a correlation with the general empathic abilities in a human-human condition. Finally, we should note that the instructions were recorded with a human female voice, which could be a potential bias although we do not report any gender effect. We plan to conduct a follow-up study including messages recorded with a human male voice.

Chapter 7

The Synthetic Tutor Assistant

Children nowadays use interactive technology such as smartphones or tablets on a regular basis. Even from the age of 4, they can operate smart devices without any help Kabali et al. (2015) for both entertainment and educational purposes. Introducing technology in classrooms has gained great interest as it provides access to a much wider set of learning resources and allows for individualised learning Peters and Araya (2011). Numerous techniques have been developed and employed in an attempt to make learning environments more engaging and empower the learning experiences for all learners. The effectiveness of these new technological resources has been tested in the context of a class, where they have been shown to improve learning speed, engagement and attention without a complicated process of adaptation Swan et al. (2005).

As robots gain popularity, it is worth exploring their potential impact in educational scenarios Mubin et al. (2013); Estivill-Castro (2016). Employing robots as part of a course has been proved useful in various educational goals such as integration, real-world issues, interdisciplinary work as well as critical thinking Beer et al. (1999). Robots in educational scenarios can be flexible, as they can assume different roles ranging from tools Mondada et al. (2006), to peers Wijnen et al. (2015); Kanda et al. (2004a); Tanaka et al. (2007) and even tutors Saerbeck et al. (2010). Although the preferred

role is not yet conclusive Shin and Kim (2007), when used as a teacher or as a peer (which implies a continuous interaction between the learner and the robot), the robot's design and behaviour become central Saerbeck et al. (2010); Vouloutsi et al. (2015); Blancas et al. (2015).

Robots enable us to control, decompose and manipulate various behavioural cues such as gaze Lallée et al. (2015), as well as present the educational content in a "socially present" manner Kanda et al. (2007); Saerbeck et al. (2010) adapted to the needs of each individual Ramachandran and Scassellati (2014). Indeed, the robots' social abilities and skills make them relevant for peer-to-peer interaction Fong et al. (2003) as they may influence children's knowledge acquisition. For example, the presence of a robot (compared to a screen) may account for higher learning gains Leyzberg et al. (2012); Kennedy et al. (2015b); Leyzberg et al. (2014) whereas the role assumed by the robot (peer or tutor) has been examined in various educational contexts Blancas et al. (2015); Zaga et al. (2015). Similarly, dynamic adaptation and personalisation of the robot's behaviour to children between 3-5 years-old suggested that children can learn new words and show a significant increase in valence Gordon et al. (2016). Positive impact and higher learning gain in long–term interactions also seem to be affected by the robots' social components Saerbeck et al. (2010) and affective responses Leite et al. (2008). Despite the fact that not all studies were able to show significant results in knowledge acquisition, most of them highlight increased engagement and positive attitude, making them suitable for effective tutors or peers, as they seem to promote interest and pedagogical achievement Han et al. (2008).

7.1 Pedagogical approaches

Typically, robots in education are employed in different kinds of scenarios ranging from technical education Tucker Balch and Gavin (2008) (usually related to robots or technology), to science Highfield et al. (2008) and learning of a foreign language Kanda et al. (2004a). However, in most cases,

researchers focus on the interaction or the social aspects of the robot and neglect the pedagogical theories that could be employed in such educational scenarios. The latest educational theories support the change in the teacher's role, shifting from a *teaching* paradigm or "telling" to a *learning* paradigm or "questioning" Barr and Tagg (1995). The figure of the teacher is no longer seen as someone who gives a lecture; in contrast, the teacher is viewed as someone who helps and guides the students to reason about a topic by asking questions and performing tasks. This method of teaching balances the roles of students and teachers and decreases the boundaries between them.

Though there is no consensus on the benefits the various pedagogical theories in robotics, a typical approach is the influential work of Piaget's *constructivism* Piaget and Inhelder (1950). In constructivism, learning is mostly based on the learner and his experiences while interacting with the world, objects or abstract concepts Prawat (1996). Here, learning is considered an active process and not a passive one Glasersfeld (1995), with interactive instructional practices that highlight the role of guidance (and consequently, the role of the instructor) Taber (2011). This approach emphasises problem-solving (real-world problems and experiences) where the content is represented as a whole. Additionally, Papert's *constructionism* Papert and Harel (1991) states that learning is the result of building knowledge structures through the progressive internalisation of actions and conscious engagement through making. Finally, the work of Vygotsky (1980) has also been influential, as it introduced the principle of *scaffolding*, that is the usage of tools or strategies providing help and the one of *Zone of Proximal Development (ZPD)*, that is the distance between what a child can do by itself and what it may do under the guidance of effective mediators. All these pedagogical approaches are highly relevant to robotic applications in education, and related study can be found in Charisi et al. (2015).

7.1.1 DAC as a pedagogical model

We consider tutoring as the structured process in which knowledge and skills are transferred to an autonomous learner through a guided process based on the individual traits of the learner. Here we present the Distributed Adaptive Control (DAC) Verschure et al. (2003); Verschure (2012) architecture as a pedagogical model: it defines the tutoring scenario as a set of fundamental principles that are general for all learning processes.

First, DAC predicts that learning is bootstrapped and organised along a hierarchy of complexity: the Reactive Layer allows for exploring the world and gaining experiences, based on which the Adaptive Layer learns the states of the world and their associations. Only after these states are well consolidated, the Contextual Layer can extract consistent rules and regularities. We believe that the same hierarchy is applicable in the pedagogical context.

Additionally, DAC predicts that in order to learn and consolidate new material, the learner undergoes a sequence of learning phases: *resistance*, *confusion*, and *abduction*.

Resistance refers to a mechanism that results from defending one's own (in)competence level. Students tend to hold overly optimistic and confused views about their level of knowledge: those with a good understanding of a topic tend to underestimate their capabilities and those who don't, tend to overestimate them Kruger and Dunning (1999). This process is what, in our case, reflects the phase of resistance. Not being skilful, but willing to protect the feeling of agency, one's perception of his skills is highly increased, and therefore, resists accepting the new knowledge as valid Kruger and Dunning (1999). This feeling is what we refer to as *resistance*, and what consequently leads to a state of *confusion*.

Confusion is what creates the necessity to resolve the problem and learn through re-adapting. Human learners show a significant variability in their performance and aptitude Felder and Brent (2005). For learning to be efficient and applicable for as broad a range of students as possible, learning technologies need to adjust to the skills and the progress of every individual.

Adapting to the skills and progress of individual students helps to maintain the process of learning acquisition; it is thus essential to maintain a challenging enough task based on each individual. Monitoring, controlling and adjusting the confusion is what we define *"shaping the landscape of success"*. Such approach is consistent with *scaffolding*, a technique based on helping the student to cross what Vygotsky calls the *"Zone of Proximal Development"*: the difference between what somebody can do without help and what someone can do with help Vygotsky (1980). *Confusion* needs to be controlled so that it does not lead to a complete loss of motivation or development of *learned helplessness* Abramson et al. (1978); the student needs to believe that he can be effective in controlling the relevant events within the learning process Seligman (1972).

Confusion is necessary to discover and generate theories and assess them later, that is, to be able to perform abduction. *Abduction* is the very process of acquiring and stabilising new knowledge. These DAC-derived learning dynamics have been grounded in aspects of the physiology of the hippocampus Rennó-Costa et al. (2014) and pre-frontal cortex Marcos et al. (2013), and they reflect the core notions of Piaget's theory of cognitive development assimilation and accommodation through a process of equilibration Piaget and Cook (1952); Wadsworth (1996).

To create a psychologically plausible Synthetic Tutor Assistant (STA), we employ DAC as the pedagogical paradigm in our educational interaction scenarios. Along these guidelines, we aim to find new solutions that would enhance learning in an educational scenario. Here, we focus on controlling confusion by increasing the difficulty of the task and the learning material in either a predefined (the difficulty of the content increases after the presentation of a number of material) or an adaptive way (the difficulty changes according to the user's performance). In the following sections, we evaluate the robot's task competence in a tutoring task and examined the robot's social competence on learning. Additionally, we assess the preferences in the morphology of social robots with children and the necessary tools the STA will be using. We then apply this knowledge to the proposed STA and

explore how different behavioural strategies affect knowledge transfer. To do so, the user is guided through goal-based learning.

7.2 Analysing children's expectations from robotic companions in educational settings

In the last years, the use of robots in educational settings has increased, as there is a belief that they offer a valuable benefit in terms of individualisation, adaptability and monitoring of educational interventions Mubin et al. (2013). Nevertheless, so far the attitudes of the main users in this context, i.e. children, are not systematically mapped. However, it is of great importance to understand children's expectations about robots and consider these when designing robots for educational purposes. Here, we aim at gaining a better understanding of children's needs and expectations from educational robot companions in terms of their appearance, characteristics, and functionality.

7.2.1 Human-Robot Interaction

Nowadays, the development of robots goes beyond utilitarian purposes: a change of paradigm is observed as robots with a more social character start to gain ground. As machines become more present in everyday life, they start to assume roles with a more predominantly social dimension: they interact on a frequent basis with humans. Indeed, the International Federation of Robotics (IFR) IFR (2016) predicts that approximately 40 million personal service robots are expected to be sold between 2016 and 2019 and most of these units are developed for household, entertainment and leisure tasks. It is therefore plausible to assume that one target user group will be children.

Among the properties that may affect the interaction between a user and a robot is morphology and design. An anthropomorphised body ensures a better interaction between humans and robots, as sharing the same physical space and gestures helps establish common ground Kanda et al. (2004a); Saerbeck et al. (2010); **?**); Fong et al. (2003). Anthropomorphism also allows the robot to show facial expressions, whose importance as a communicative channel has been extensively defended Frith (2009); Keltner and

Ekman (2000). Other perceptual cues that facilitate Human-Robot Interaction (HRI) are related to non-verbal communication channels such as gaze, eye contact, gestures, imitation and synchronisation Ono et al. (2001); Lallée et al. (2015). Eye contact is seen as a highly communicative indicator of attention and as a sign of the presence of someone else Boucher et al. (2010). In general, one can speak of a social salience effect that depends on morphology, social cues and task capabilities Inderbitzin et al. (2013).

Age and previous experience with robots have been found to influence the kind of features children expect from a robot. For instance, human-like appearance is preferred by children younger than nine years old, whereas robot skills and functions are more appealing to older children and adults. Moreover, after interacting with a robot, children pay more attention to their motor abilities than to only their shape.

7.2.2 Robots in Educational Scenarios

In terms of expressivity in a learning task, we can distinguish two types of robots. First, robots that mainly focus on knowledge transfer, and socially supportive robots that engage in active dialogue and supportive behaviour towards the learner. The latter has been shown to positively affect the learning performances of children Saerbeck et al. (2010). One of the main differences between the kind of behaviour a robot should show in a school environment and other educational contexts such as a museum is the duration and the nature of the interaction. The use of robots in schools requires ongoing participation, as the children the robot interacts with are always the same; contrarily, while when utilising robots in other scenarios the interaction with the users is usually short lasting and transient Kanda et al. (2004a).

7.2.3 Co-designing with children: Drawings' analysis

With the aim of developing an educational robot that both considers the findings in the field and meets children's expectations, we implemented an exploratory co-design method to understand which would be the required

characteristics for such a robot. Co-designing technology with its potential users increases the probability that results will meet expectations. Thus, in case of education, children should be involved as co-designers of new educational technologies Druin (1999). This is particularly significant when considering the age-related differences between the mindsets of the adults who typically design the technology and that of children who use it Melonio and Gennari (2012). Indeed, a systematic age dependent anthropomorphic bias has been reported with the users of complex robot exhibition technology Eng et al. (2005). Thus, seeing children as robot co-designers allows us to better understand their point of view and gain insights into their specific needs.

In addition to age dependent effects also gender differences have been observed in the way children represent people and objects. Boys' drawings usually show the omission of arms, trunks, and clothing (however, these omissions decrease with age) together with an asymmetry in facial features as compared to girls Skybo et al. (2007). However, they begin to draw movement before girls, for example, they draw limbs in positions other than straight-out.

Drawing can be used as a method of representing individuals' preferences and is in the co-design context a way for children to make sense of their experiences Anning and Ring (2004); Dyson (1988). It is also a useful method to evaluate children's perception, experience and understanding, as drawing is shown to be considered more enjoyable than answering questions Lewis and Greene (1983). Moreover, drawing is a task that allows overcoming linguistic barriers Chambers (1983). We thus asked children to design the robot they would like to have; this way, we can have a more effective intuition of their needs and expectations.

7.2.4 Methods

This study was conducted in the form of school workshops at the Cosmo Caixa Science Museum of Barcelona (Spain). A total of 142 children (64

females) from Year 4 of Elementary school (9-10 yo) were divided into groups of 8-9 kids. At the beginning of the session, all the children were introduced to three different robots (Zeno -Robokind-, Nao -SoftBank robotics- and CodiBot -SPECS-) and freely interacted (in groups of three) with each robot for approximately four minutes. Subsequently, two kids per group were selected to individually interact with the Zeno robot to do an extra activity (explained in section *The healthy living task*). Additionally, we provided all children with coloured pencils and sheets and asked them to draw the robot they would like to have. The drawing session occurred while each of the selected children interacted with the robot. An image of the robots and their location is provided in Figure 7.1.

Before the end of each session, all the participants were requested to fill in a questionnaire that contained the following information: gender, if they liked the activity, if they would do the activity again and if they would recommend it to their friends. Additionally, we asked them to order the three robots they interacted with by preference.

Robotic Systems

All children interacted with the following robots:

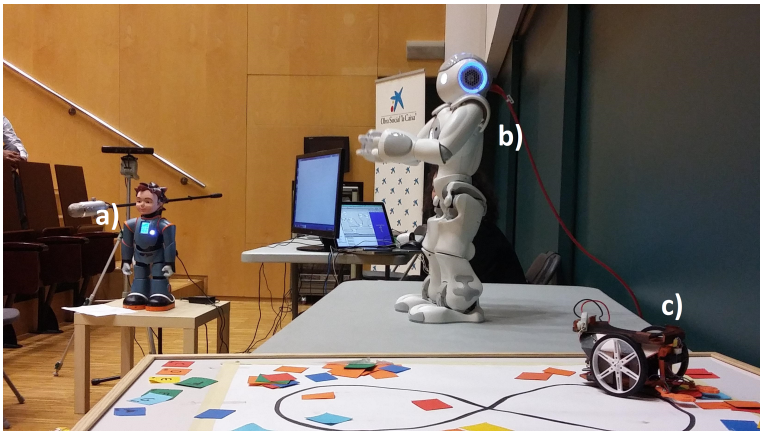


Figure 7.1: Image of the room with the setup and the position of each of the robots. a) Zeno, b) Nao and c) CodiBot.

- **CodiBot:** developed by the Synthetic Perceptive Emotive Cognitive Systems (SPECS) group, at Pompeu Fabra University[1]. The main purpose of this robot is to help children learn how to code by using music and colours. CodiBot allows children to create melodies in an interactive way by mapping the seven notes of the C major scale to seven colours: a melody is created in the form of a score/program by placing the coloured patches to the robot's trajectory.

- **Nao:** developed by SoftBank Robotics, France, Nao is an autonomous humanoid robot with a height of 58cm. It has 21 degrees of freedom, four microphones (for speech recognition and sound localisation), two speakers and two HD cameras. Although it cannot display facial expressions as it lacks mouth and eyebrows, it can exhibit emotional states through a circle of coloured LEDs surrounding its eyes. At the beginning of each session, the Nao welcomed the students and provided a brief introduction of the activity. During the interactive session, students could interact with the robot and trigger several behaviours by activating its sensors (e.g., the feet or its head).

- **Zeno:** developed by Robokind, Zeno looks like a male cartoon character. It can display rich facial expressions through a face with seven degrees of freedom composed of eyebrows, mouth opening and smile. Additionally, it has five degrees of freedom in its arms and four degrees of freedom in its legs and waist. During the group interaction, children could freely trigger a variety of behaviours by choosing the desired response from the touchscreen embedded on the robot's chest. During the dyadic task, the robot verbally interacted with the participant using a speech synthesiser based on the Acapela software [2]. Movement was tracked using the Kinect sensor and the Scene Analyzer software Zaraki et al. (2014).

In terms of language, the provided questionnaires were in Catalan, the Nao

[1] http://www.codibot.com/
[2] http://www.acapela-group.com

robot spoke in Spanish and the Zeno robot spoke in English, both during the first interaction with all the children and during the aforementioned dyadic task.

Automatic Speech Recognition (ASR)

We used two corpora for training the ASR acoustic models. The first was the British English version of the Wall Street Journal corpus created at the University of Cambridge Robinson et al. (1995). The second was the PF-star corpus of British English child speech Batliner et al. (2005). Both corpora were used to create a single acoustic model that can be used for both adult and child speech.

To improve robustness to noise, we applied background noise audio to augment the training data. For this purpose, we used the CHiME corpus Christensen et al. (2010) which contains various kinds of background noise recorded in real-life environments. Since our main relevant use-case for the ASR is a public museum setting, we decided that the "cafe" background noise would be the best matching type of noise to use for our model. For each utterance in the training set, a section of the noisy corpus of the same length was randomly selected and added to the utterance audio. The addition was done using the SoX[3] sound processing tool, using the mix option. We added the noise at three different signal-to-noise levels, 5 dB, 10 dB and 20 dB.

We used the Kaldi toolkit to train the acoustic models for the ASR system. The toolkit has relatively standardised scripts (collectively known as recipes) designed to work with different sets of training data. We followed the Wall Street Journal (WSJ) recipe and trained a DNN model using the `train_multisplice_accel2.sh` script provided in Kaldi, which at the time of writing was the recommended script to use for DNN training[4]. We used

[3] `http://sox.sourceforge.net/`

[4] At the time of writing the DNN scripts are under continuous development by the Kaldi team as DNN approaches for speech recognition are a highly active area of research. See the Kaldi website `http://kaldi-asr.org` for the latest information about the DNN setup.

four hidden layers and trained over one epoch, which came to 62 iterations. The initial effective learning rate was 5×10^{-3} and the final rate was 5×10^{-4}.

We used Beep[5] as the pronunciation dictionary, since it is designed for British English pronunciations. For words that are not in the dictionary (e.g. robot names, such as Zeno) we use the Sequitur tool Bisani and Ney (2008) to estimate the phone sequences given the letters of the word.

To provide online (i.e. live) ASR we refactored and extended the online examples provided in Kaldi. A fuller description of the ASR development is given in Fernando et al. (2016). Moreover, despite not being English speakers, the system had no problem to recognise the children's speech, and they could understand what the robot was saying during the interaction.

Scene Analyzer (SA)

The Scene Analyzer is a framework that provides a human-like understanding of the information coming from the surrounding environment. It uses a Microsoft Kinect 1 sensor and a variety of libraries (Kinect SDK, SHORE etc.) that provide a wide range of multimodal data: high-level verbal/nonverbal cues of the people present in the environment, such as facial expressions, gestures, position and speaker identification. This information is later processed to extract significant social features, which are structured in a "metascene" data packet to be transmitted to rest of the modules. More information about the framework can be found at Zaraki et al. (2014).

7.2.5 The healthy living task

The purpose of the interaction was to assist learners in an inquiry-based learning task to discover the benefits of physical exercise. The task consisted of two parts. In the first part, the robot encouraged the participant to perform exercises at various speeds and for various duration and provided information about the amount of energy spent by the kid. To detect participant's movements, we used the Kinect sensor and the Scene Analyzer. A

[5] ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/dictionaries/beep.tar.gz

sound, whose pitch was paired to the intensity of the movement (i.e., higher pitch, faster movement), was played while the participant performed the exercise. In the second part of the interaction, the robot asked questions about the consumption of energy during various kinds of exercises. The questions were also displayed on a TV screen and participants would verbally provide their answer. At the end of the session, children could request the robot to perform various actions (like "make a happy face" or "do the monkey dance").

7.2.6 Results

Results from the questionnaires

We first explored for any gender differences in Likeability (whether they liked the task, whether they would do it again and whether they would recommend it to their friends). A Mann-Whitney Test showed significant differences between males (4.97 ± 0.18) and females (4.90 ± 0.35) ($p = 0.015$) in Likeability (whether they liked the task), (Figure 7.2). Additionally, 75.8% of the children placed the Nao as their first preference, 64.1% placed the Zeno as their second choice and 77.3% placed the CodiBot as their third choice of preference. There were no significant differences among genders for the questions "Would you do it again?" and "Would you recommend it to a friend?".

The Drawings

We classified each drawing based on several parameters: morphology, functionality, relative size of the robot to the child, body features, facial expression, and others. Morphology was further divided into: anthropomorphic (appearance resembles that of humans, which also contained the level of anthropomorphism), caricatured (appearance is not necessarily realistic or believable and usually have exaggerated features to provide a comic effect), functional (the embodiment reflects the task the robot performs), and zoomorphic (appearance resembles that of animals, adding also the kind of animal they resemble) Fong et al. (2003).

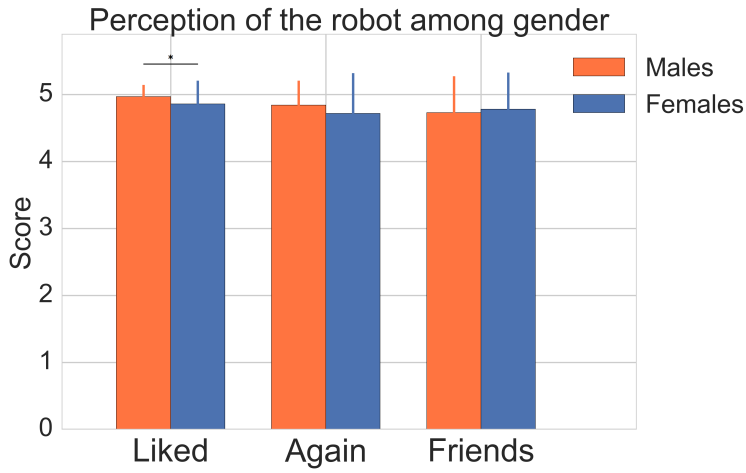


Figure 7.2: Gender differences in perception of the task. "Liked" refers to the question "Did you like the task?"; "Again", to "Would you do it again?"; and "Friends", to "Would recommend it to your friends?".

The group related to functionality comprised of pet, defence, learning, health, chores, and playing. The facial features we looked for were hands, eyes, mouth, nose, ears, and hair. The identified facial expressions were happiness, sadness, anger, and neutral. Additionally, we analysed the size of the drawings (the space they occupied in the paper), the robot's gender and whether kids drew themselves with the robot or not.

Differences in morphology

In terms of functionality, we classified the drawings based on the four main categories defined by Fong: anthropomorphic, caricatured, functional, and zoomorphic. In figure 7.3, we report the frequency of robot appearance based on those categories. Results show that children tend to mainly image robots with an anthropomorphic appearance, with the 58% of those human-like robots looking like the Nao.

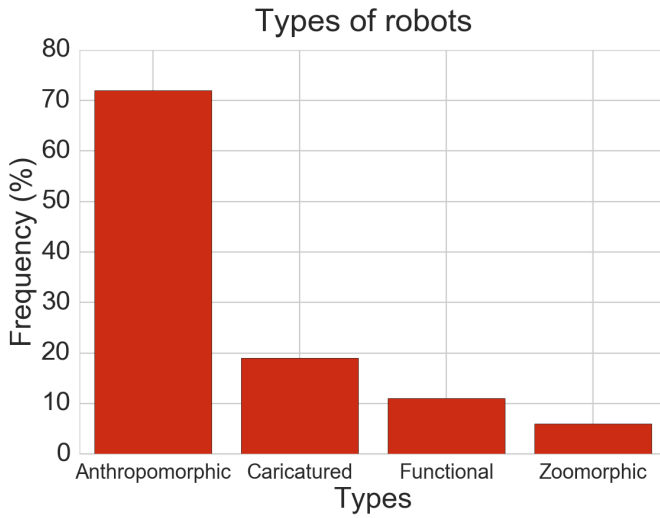


Figure 7.3: Frequency of the four types of robots occurring in the drawings based on Fong et al. (2003). The blue part of the "Anthropomorphic" bar represents the drawings containing robots classified as "machine-like".

Differences in functionality

Regarding functionality, we identified six main categories: robots as pets, as partners for play activities, robots as educators (that teach them and help them with their homework) and doctors, robots used for defence and robots that do chores (as cooking or cleaning). Figure 7.5 shows the frequency of robots based on their functionality. Results indicate that children preferred robots as pets or doctors (with a 22% of them corresponding to robots as pets and another 22% to robots as doctors).

Gender differences

We did not observe differences between genders in use of movement, contrarily to Skybo et al. (2007). In our case, from the 35% of drawings depicting movement (e.g. using lines to represent speed or drawing arms in positions other than straight), the distribution of these drawings per gender was equitable (a 50% of them were drawn by boys and the other 50% of them by girls). Children tended to draw genderless robots compared to male or

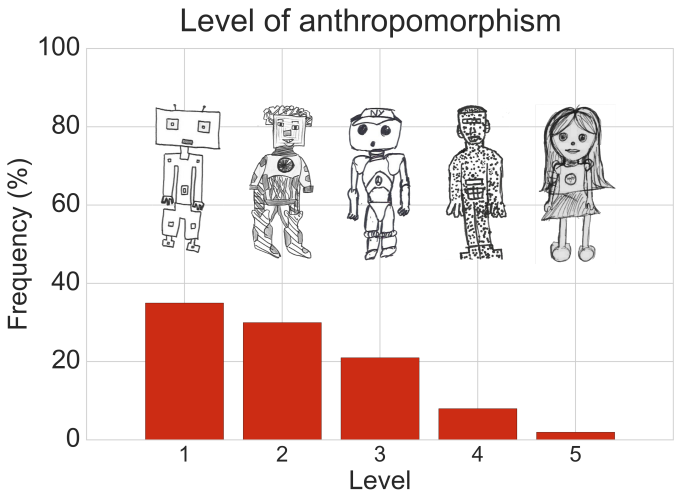


Figure 7.4: Frequency of anthropomorphism shown in the drawings (only for the robots inside of the "anthropomorphic" type). An example of each level is shown above each bar.

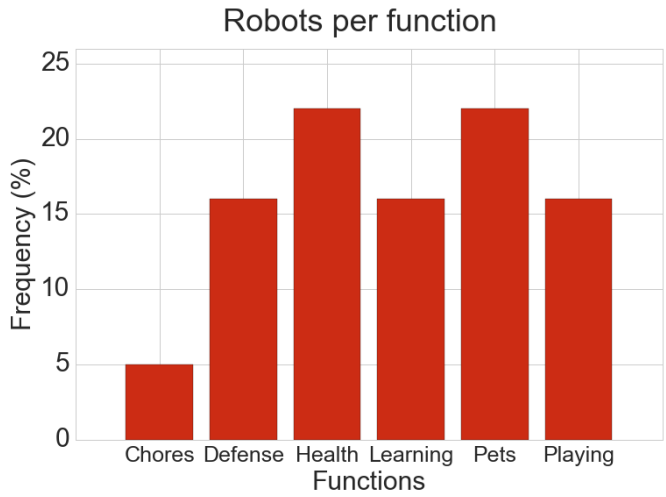


Figure 7.5: Frequency of envisioned robot functionality as extracted by children's design.

female ones, as shown in figure 7.6.

Regarding the depicted functionality, we can see differences depending on

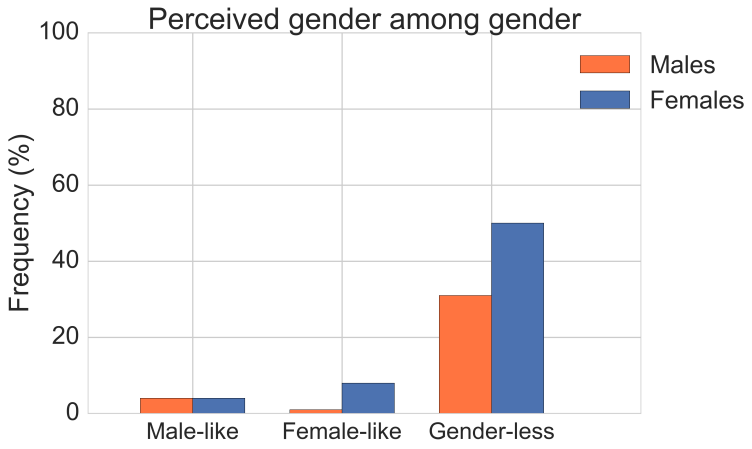


Figure 7.6: Frequency of robot gender as extracted from children's drawings.

gender (Figure 7.7). In the case of the chores- or pets-related robots, the frequency of these functions in the drawn robots is equally divided between genders (2% for each gender in chores-related robots and 11% in the learning-related ones). The main difference comes from the defence-related robots, all of them drawn by boys (16% of the total amount of drawings), which also explains the fact that in the other functionalities (health, learning, and playing) the frequency of robots drawn by girls is higher. This is most evident in the learning-related ones, where a 2% of the drawings were produced by boys, and a 13%, by girls.

Differences in size, body features, and facial expressions

Children tended to draw genderless robots compared to male or female ones, while there was no interaction between gender and functionality (Figure 7.6). In terms of body features, all robots were drawn with eyes and almost all had a mouth and hands (Figure 7.8).

In terms of facial expressions, 48 children drew a robot with a happy face whereas 74 children drew a robot with neutral facial expression. In total, 30 children drew themselves with the robot. All drawn children with the robot displayed a happy facial expression while the frequency of drawing

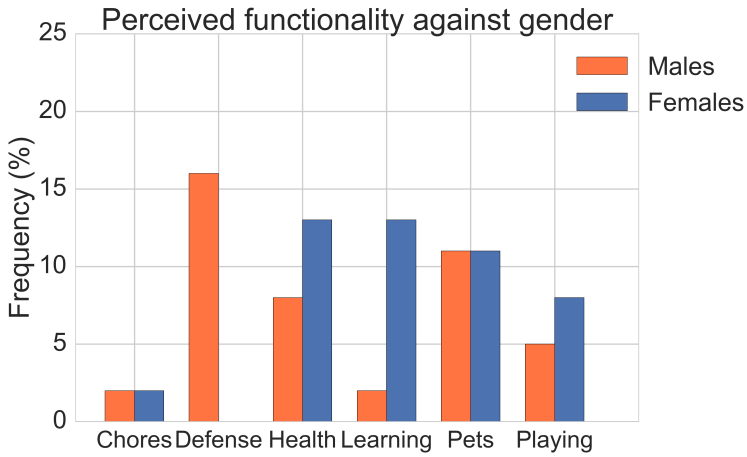


Figure 7.7: Fig. 7. Frequency of robot functionality by gender as extracted from children's drawing'.

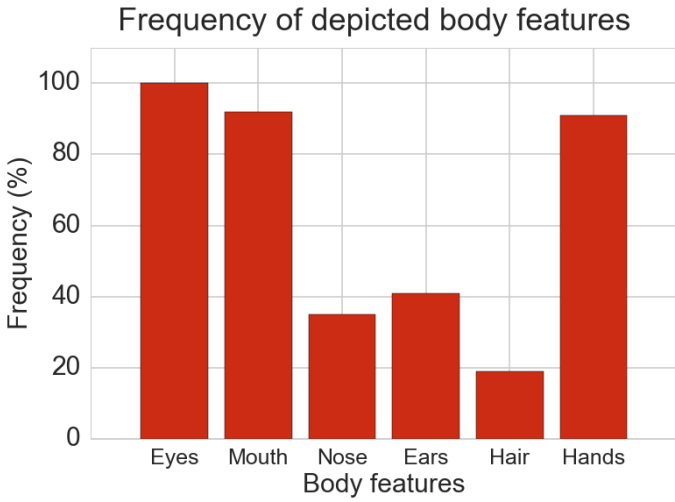


Figure 7.8: Frequency of body features present in the drawings.

the child larger (n = 10), smaller (n = 10) or equal (n = 10) to the robot was evenly distributed.

7.2.7 Discussion and conclusion

Robots will soon become an almost ubiquitous part of our daily lives Gates (2007). Therefore, we investigated the characteristics children expect from robotic companions in educational settings and how they envision them in terms of design and functionality. To do so, a sample of 142 children between nine and ten years old interacted with three different robots whose morphology ranged from non-anthropomorphic (CodiBot) to anthropomorphic. Here, we varied the level of anthropomorphism, as we presented two anthropomorphic robots: the Nao and the Zeno, with the latter being classified as highly expressive and with a human-like face.

Children were asked to rate each robot in preference and evaluate the interaction. Additionally, we asked them to draw a robot of their preference and we analysed their drawings. From this sample, 34 of them interacted with the Zeno robot in a one-to-one interaction focused on physical exercise. Meanwhile, the children that did not interact with the robot were drawing their robots or watching the interaction. At the end, children answered the questionnaires. Our results put in evidence that children preferred humanoid robots that resemble machines than humans in terms of morphology. In terms of gender, most of them envisioned a genderless robot, similar to what has been observed in Bumby and Dautenhahn (1999).

We observed several similarities between drawings within the different groups, which suggests that children did affect each other during the drawing activity. Indeed, group members are likely to imitate the behaviour of other members of the group (nesdale2001social) and mutually influence their artwork (boyatzis2000naturalistic). It is possible that children's designs may have been influenced by the media (bushman2006short) or their previous interaction with the three robots, as we observed several similarities with the Nao robot.

Contrarily to what we could expect, only the Nao was depicted in the drawings although all children interacted the same amount of time with each robot. Additionally, two children per group interacted with the Zeno robot

performing the healthy living task, however, none of these children drew a robot that resembled the Zeno. Thus, any resemblance with the Nao robot cannot be explained by the exposure time with the robot. These resemblances are consistent with children's preferences since the Nao was rated first in liking and in accordance with earlier work that suggests that bodily features should not be identical to humans Woods (2006); Bumby and Dautenhahn (1999) but instead have some human-like characteristics.

As a limitation, we must say that not all the children interacted with the three robots in the same order, as they were divided into smaller groups (between two and three people) that rotated turns and were also able to move freely among them. Thus, we cannot provide results regarding the effect of interaction order on their expectations from robotic companions.

Another conclusion that can be extracted from the drawings is the heterogeneity of expectations children have from robotic companions. The robot's expected functionality is not always constrained to one specific field: children see robots as multipurpose tools, mainly related to educational and domestic purposes (drawing "a" of figure 9 represents an example). Additionally, children's image of robots as defence-related agents (e.g. soldiers, policemen, etc) cannot be ignored; they are possibly influenced by cinema culture, as suggested in Benítez Sandoval and Penaloza (2012). A representation of each type of functionality can be found in Figure 7.9.

Consistently with Flannery and Watson (1995), we found gender differences in the kind of scenes sketched by children: boys produced more defence-related robots and drawings including aggressiveness situations; girls depicted more details in terms of clothing. Moreover, girls used a larger part of the page, as already observed in Iijima et al. (2001).

As previously stated, we highlight the importance of inviting children to co-design robots to properly assess their expectations and needs. Moreover, although studies like the current one provide insights about the expected morphology and functionality of robots for children, we should not forget that other aspects have to be considered. When designing educational

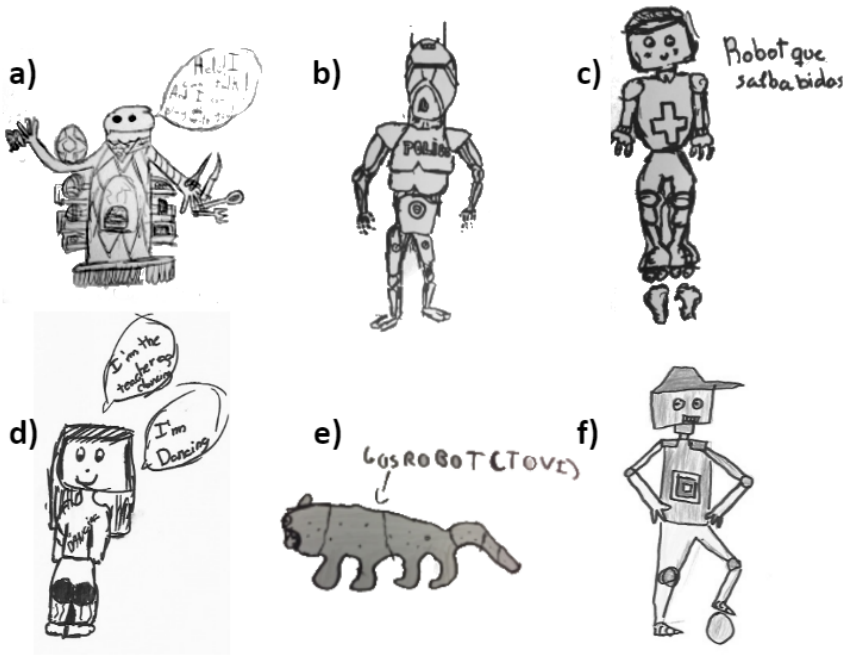


Figure 7.9: Drawings depicting the six types of functions defined: a) Chores (an example of multipurpose one, as it also relates to playing), b) Defence, c) Health, d) Learning, e) Pets and f) Playing.

robots for children, we also have to consider the goal these children would like to achieve with them.This is an important aspect because there can also be possibilities of misuse, such as expecting the robot to do their homework, instead of helping them with it (Figure 7.10).

This work mainly focused on the collaborative design of robots with children. A way to systematically explore collaborative design would be to ask children to draw their robot of preference without previously allowing them to interact with it. Currently, robots meant to be used by children are designed by adults, neglecting children's perceptions and attitudes towards robots. The active participation of children in the design of smart technology is advocated by Druin (1999) as they are likely to provide valuable feedback to the design process that better addresses their interests and

Figure 7.10: Drawing of an educational robot saying "Hello, I am the machine to do homework."

needs.

Extracting constructive information can be done with a variety of methods, ranging from writing, interviews and drawing Hourcade (2008). Additionally, children can be presented with various robotic platforms whose morphology gradually varies from mechanical to anthropomorphic ones, as in our case the "step" from machine-like (CodiBot) to human-like (Nao, Zeno) was great both in terms of functionality and morphology. Nonetheless, the current study provides valuable insights on robot design that is created for children by children.

The present study primarily addressed the design of robotic applications in terms of morphology and functionality. The examination of the attribution of emotional states, mental capabilities, perceived personality and interaction styles of robotic platforms goes beyond the scope of this study, however, such issues need to be addressed in future work. Finally, given the fact that the role assumed by the robot affects how users perceive it Blancas et al. (2015), a systematic approach is needed to ensure the robot's role meets children's expectations.

The fact that learning-related robots (those depicted as teaching or were reported in writing as "robots to learn" or "robots to do homework") were not the most frequently depicted in the drawings should not be a constraint for the use of educational robots. Instead, it should be seen as a demonstration of the heterogeneity of the functionality that robots can have for children. The most popular functionality of robots was either related to health or pets. One could take advantage of their popularity and design educational robots to scaffold children's learning process in subjects related to them, like biology or chemistry.

The three main body features present in the drawings are eyes (depicted in all of them), followed by mouth and hands, which relate to the expected anthropomorphism of the robots. The result from this study is then a prototype of a robot with anthropomorphic (but machine-like) characteristics that does not resemble any specific gender. From a technical perspective, the focus of the design should be centred in its eyes, mouth, and hands and from a functional perspective, it seems that multiple functionalities are preferred as opposed to a single one.

7.3 The effects of gaze in an educational scenario

To test the design principles and the ability of our system to provide the robot with the necessary means of social capabilities, we devised a dyadic teaching scenario. Robotic instructors can provide educational content in a more social manner than other artefacts or devices. Synthetic agents have already been introduced in schools as teacher assistants Chang et al. (2010) or peers Kanda et al. (2007).

As we already demonstrated, tiny behavioural cues may affect the image the robot projects and are therefore of high importance in the context of teaching. Gaze, especially of both the teacher and the student affects knowledge transfer and learning rate Phelps et al. (2006). By using gaze, emotions and body language, robots may be able to play on the empathic lever to increase knowledge transfer in the benefit for the student.

The first question raised during the development of a Synthetic Tutor Assistant (STA) was whether it could act as an effective peer for the learner, both regarding social interaction and learning. Thus, we examined the social competence of the robot by decomposing it to two key factors that we manipulate. Additionally, by examining the performance of the participants, we can assess the interaction between the components of social competence to the task competence. Hence, the focus of this study was to investigate how the modulation of the behavioural parameters of an agent, i.e., gaze model and emotion expression can affect the acquisition of knowledge of a particular topic and the subjective experience of the learner.

Our hypothesis was that eye contact would strengthen the interaction between the student and the robot while facial expressions would act as a reinforcement of the participant's actions (the robot displayed a happy face when the human partner answered correctly and a sad face when the answer was incorrect) and could be considered as a reward. Furthermore, gaze can be decomposed into communication signal (eye contact) and action support (e.g. look at the object of selection). Thus, we wanted to evaluate the effect of each separate gaze model on the interaction. To test our hypothesis, we

Figure 7.11: Experimental setup of the robot interacting with a human using the Reactable for the educational game scenario. In the image, you can see the participant holding an object used to select an item from the Reactable (round table with projected images of countries and capitals). The human partner was facing the iCub. The projected items were mirrored, so each side has the same objects.

used the H5W_Alpha in the role of a robotic tutor. We based the robot's behaviour on the DAC architecture and the proposed behavioural modulation system. The scenario devised was a pairing game where the task was to correctly match an item to its corresponding category, using the Reactable to project the digital objects.

7.3.1 Methods and setup

The educational task consisted of the H5W_Alpha (the iCub) controlled by the DAC architecture, the Reactable and the Kinect sensor to track the human partner. An example of the setup is displayed in Figure 7.11. The system was designed to run autonomously in each trial, using the allostatic control as the main component to guide the learner during the task. In total, our system operated for approximately 24 hours.

The pairing game and in general the use of technology-enhanced environments was grounded on the premises of constructivism Papert (1980), an educational model that emphasises the collaboration and feedback of two or more peers who learn together. In this study, the robot behaved similarly to a constructivist tutor. The STA did not provide information or feedback directly. Instead, it would provide feedback like *"well done!"* or *"it is ok, you will do better next time!"* only after the participant had provided an answer. Additionally, it helped students understand the rules of the game by commenting on the invalid actions of the partner. For example, if the player would choose two items or two categories instead of one of each, the robot would say: *"first choose an item and then the category it belongs to"*.

The pairing game

The devised educational task was a turn-taking pairing game, where participants had to match objects appearing on the Reactable (for more information about the Reactable, please see section 5.2.2) to their corresponding categories. The selection of virtual objects was achieved using an object (paddle) or a cursor (fingertip). At the beginning of each session, the robot verbally introduced the game, initiated the interaction and assumed the first turn.

The task comprised three levels of increased difficulty. The gradual increase of difficulty allowed to scaffold the task, and consequently, the improvement of the learning process Azevedo and Hadwin (2005). For each level, four objects and four categories appeared. Both the learner and the robot had the same objects mirrored on each side. Upon correct choice, the object (but not the category) would disappear from both the learner and the STA. Validation of the four associations was required before proceeding to the next level. The players received visual feedback from the Reactable regarding their correct (green blink) or incorrect (red blink) matches.

We tested the pairing game with both children and adults and adjusted the content of the game to their estimated knowledge. Thus for children, the game was about recycling, where players had to match different types

of waste (like glass or plastic bottle) to the corresponding recycling bin. For adults, the game was about geography and more specifically matching a country with its capital. Since almost all participants were native Europeans, the difficulty was defined by choosing foreign, non-European countries. With children, the game, the robot's utterances and the questions were in Spanish, whereas with adults in English. An example of the interaction is found in Figure 7.12.

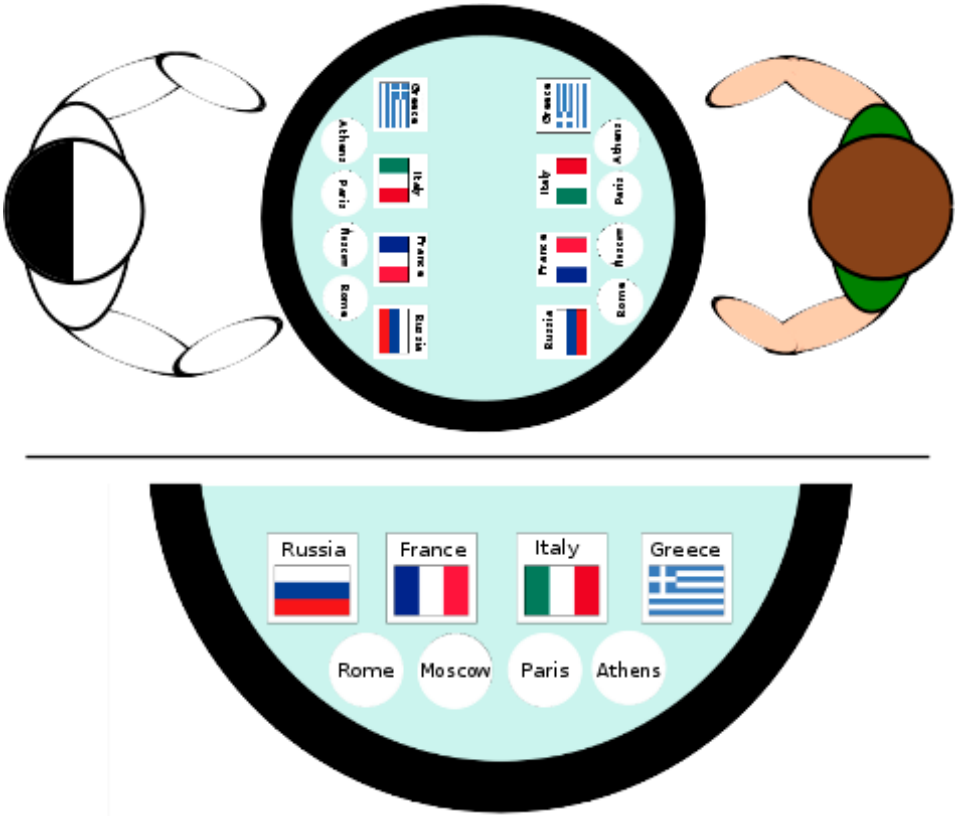


Figure 7.12: Example of the pairing game setup (geography). The robot (left) and the participant (right) had a mirrored screen. The round objects on the bottom represented the capitals and the square objects on the top the countries with their flags. For each correct association, only the capital item disappeared, and only the remaining non-associated items were displayed. For recycling, the setup was the same only the images of the categories (recycling bins) and items (various kinds of waste) changed.

Condition	Embodiment	Action supporting gaze	Eye contact	Facial expression
THI	No	No	No	No
NoR	Yes	No	No	Neutral
ToR	Yes	Yes	No	Neutral
T&HoR	Yes	Yes	Yes	Varying
HHI	Yes	Yes	Yes	Varying

Table 7.1: Behavioural parameters used for each of the five conditions, ranging from the simplest one (THI) to the more complex one.

Experimental conditions

To evaluate the effect of social cues to knowledge acquisition and subjective experience and assess our architecture, we devised five experimental conditions (see Table 7.1). We defined each condition by summing up complementary behavioural elements (i.e. gaze behaviour and facial expressions), creating a scale from the most artificial to the most natural interaction.

The simplest system allowed us to investigate the role of the embodiment during the interaction. It composed of the Reactable and a speaker uttering the same speech synthesis as the robot **(HTI: Human-Table Interaction)**.

In the **NoR (Non-Oriented Robot)** condition, the robot's gaze was fixed to a point at the centre of the table. The head of the agent compensated for the torso during pointing gestures so that the fixation point remained constant. This condition ensured that no accidental eye contact between the robot and the learner occurred. The robot's facial expressions were set to neutral.

The **Task-Oriented Robot (ToR)** condition liberated the robot from the fixation point of the previous condition and allowed for action-sustained gaze: the robot looked at the pointing location. Additionally, it looked at the selected items of the participant. Here, all possible gaze targets were located at the table, making any eye contact with the student very unlikely.

Similarly to the NoR condition, the agent's expressions were neutral.

In the **Task and Human-Oriented Robot (T&HoR)**, we enhanced the previous gaze model by adding fixation point at the partner's face. The synthetic agent looked at the partner when speaking to promote eye contact. Additionally, its facial expressions changed according to the actions of the student (happy for correct choices and sad for incorrect). The robot's behaviour regarding gameplay, verbal interaction and reaction to the participant's actions remained the same.

Finally, as a control condition to acquire the baseline of the interaction, we added the **HHI (Human-Human interaction)** condition, where two naive humans played the game. We tested all conditions with adults, whereas only the **NoR, T&HoR** and **HHI** with children.

7.3.2 Procedure and measurements

The procedure was as follows: the experimenter welcomed the participants and introduced the task to them. After carefully reading the task's instructions and pose any questions, participants filled the pre-knowledge questionnaire, followed by the interaction with the STA. Consequently, participants filled a post-knowledge and subjective experience questionnaire and were debriefed by the experimenter. A specialist on child interaction (monitor) was always present during the pre- and post-questionnaires, to provide clarifications and read out loud the questions. The interaction with the robot lasted approximately 12 minutes and the entire procedure about 30 minutes.

The pre- and post-knowledge questionnaires served to measure the effects of the intervention on knowledge acquisition. They consisted of Single Answer Multiple Choice questions where they had to match an item to its category. The items and categories of the questionnaire were relevant to the task (geography for adults and recycling for children). The questions for recycling came from the website "Residu on vas"[6], property of the Catalan

[6] http://www.residuonvas.cat

Wastes Agency (12 items in total). Adults had to match capitals with their countries (12 items) and flags with the name of their country (12 items).

We hypothesised that the most elaborate interaction scenario with the robot (T&HoR) could result in the highest improvement during post assessment among the robot conditions.

The Subjective Experience Questionnaire measured the effect of the STA's social behaviour and the robot's psychological believability. The Basic Empathy Scale Jolliffe and Farrington (2006) measured empathy; the Godspeed Bartneck et al. (2009) evaluated the synthetic agent regarding anthropomorphism, animacy, likeability and perceived intelligence. Finally, the Tripod Survey Ferguson (2008) assessed student's engagement and perception of their teachers. We hypothesised that the more complex the behaviour of the robot, the higher it would score on the subjective experience. Finally, we recorded the log files from the game and behavioural data for all participants.

We recruited 74 adults (age M = 25.18, SD = 7.55; 24 females) from the Pompeu Fabra University campus. We randomly distributed participants among conditions (THI = 13, NoR = 15, ToR = 15, T&HoR = 15, HHI = 16), and none of them reported familiarity with robots or a significant knowledge of programming or computer science. 34 children participated in the recycling game (age M = 9.81, SD = 1.23; 11 female) and were randomly assigned to one of the three experimental conditions (NoR = 12, T&HoR = 14, HHI = 8).

7.3.3 Results

Knowledge acquisition

First, we report a significant knowledge improvement in adults in all conditions: THI ($t(13) = 7.\ 697$, $p < 0.001$); NoR ($t(14) = 2.170$, $p = 0.048$); ToR ($t(14) = 3.112$, $p = 0.008$); T&HoR ($t(16) = 3.174$, $p = 0.006$) and HHI ($t(13) = 3.454$ $p = 0.004$). Despite the fact that all conditions indicated an increase of the knowledge between pre and post tests, a Kruskal-Wallis test

showed that there was no statistically significant difference in the improvement in the different conditions $x^2(4) = 2.709$, $p = 0.608$.

In contrast, children showed a trend in improvement in all three conditions; however, there were no significant differences between the pre and post questionnaires. From these results, it appeared that the behavioural cues exhibited by the robot did not affect the knowledge acquisition of the participants. We were expecting higher knowledge transfer in the more complex and social conditions for both adults and children. Given the fact that we did not observe any significant differences in the recycling game, we hypothesise that the associations taught were too simple. Additionally, the significant improvements in all conditions in adults could be attributed to the visual feedback received by the Reactable. In both cases, we need to revise the game's difficulty, to make sure we provide a more challenging task.

Subjective experience

Regarding the subjective experience, there was no statistically significant difference among conditions for the Empathy and Tripod parts. These results indicate that the manipulation of social cues did not affect how humans perceived the robot as a tutor and it did not elicit different empathic responses. However, we found statistically significant differences between groups for the Godspeed part as determined by one-way ANOVA (F(4,35) = 4.981, p = 0.003). As expected, in the Godspeed questionnaire humans scored higher (HHI, .06 $\pm$ 0.87) than the robot in two conditions (NoR, 2.84 $\pm$ 0.72, $p = 0.003$; ToR, 3.19 $\pm$ 0.46, $p = 0.044$, but surprisingly not the T&HoR) and the table (THI,3.02 $\pm$ 0.56, $p = 0.031$) (Bonferroni post-hoc test). Results suggested that the synthetic agent scored significantly lower than a human in all conditions but the one where its behaviour was as close as possible to that of a human: action-sustained gaze (look at where one points), eye contact and facial expressions as a feedback to humans' actions. No statistical differences in the other two tests were found.

We found no statistically significant differences in any of the subjective

experience questionnaires from children. We suspected that such results may be because these questionnaires were originally designed for adults and not children, despite the presence of a monitor who was there to assist children in the comprehension of the questions.

Behavioural data

Regarding the behavioural data, the most relevant results were related to gaze, as other behaviours (such as speech, waving, etc.) were not present enough times to analyse them systematically. Looking at the other player was classified as either looking at the face of the robot or the human (in the robot and human conditions) and looking at the speakers on top of the table at the THI condition. In adults, a Kruskal-Wallis test showed that there was a highly statistically significant difference in the time spent looking at the other player between the different conditions (x2(4) = 15.911, $p = 0.003$). A Mann-Whitney Test showed significant differences between the THI (2.72 ± 5.53) and the NoR (16.37 ± 21.17) conditions ($p = 0.026$); the THI (2.72 ± 5.53) and the ToR (7.80 ± 7.76) conditions ($p = 0.029$); the THI (2.72 ± 5.53) and the T&HoR (19.87 ± 12.01) conditions ($p < 0.001$); the ToR (7.80 ± 7.76) and the T&HoR (19.87 ± 12.01) conditions ($p = 0.028$); and the T&HoR (19.87 ± 12.01) and the HHI (3.66 ± 4.13) conditions ($p = 0.002$) (See figure 7.13).

We expected that the more humanlike the behaviour of the STA, the more people would look at it. Surprisingly, results showed that humans spent little time looking at each other, possibly because they were focused on the task and exchanged very few utterances between them. In contrast, the significant difference in the time spent looking at the robot in the most social (T&HoR) and the HHI condition could be explained by the robot's behaviour. The robot looks at the human and comments on the player's actions, whereas in the case of people, they didn't.

We also found a statistically significant difference between conditions for the mean gaze duration in children one-way ANOVA (F(2,26) = 8.287, $p = .0021$)). A Bonferroni post-hoc test revealed that the time spent looking

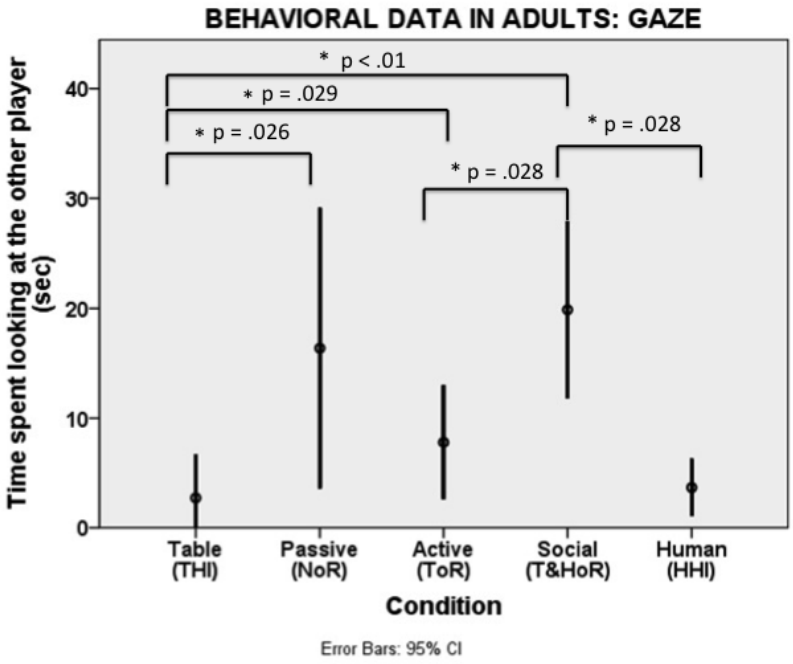


Figure 7.13: Time spent looking at the other player (in seconds) in adults among conditions. Asterisks "*" depict significance.

at the other player (in seconds) was significantly lower in the NoR (14.70 $\pm$ 8.81", $p = 0.012$) and the HHI conditions (11.74 $\pm$ 8.02", $p = 0.003$) compared to the T&HoR condition (30.97 $\pm$ 15.16")(figure 7.14).

Our expectation regarding the difference between the NoR and T&HoR conditions was correctly met: people looked more at the agent who looked back at them. However, we were not expecting a difference between T&HoR and HHI condition. We believe that the reason why the difference in mean gaze duration occurred was that humans remained focused on the game and were mainly looking at the table instead of looking at the other player. Furthermore, there were much less verbal interactions between them.

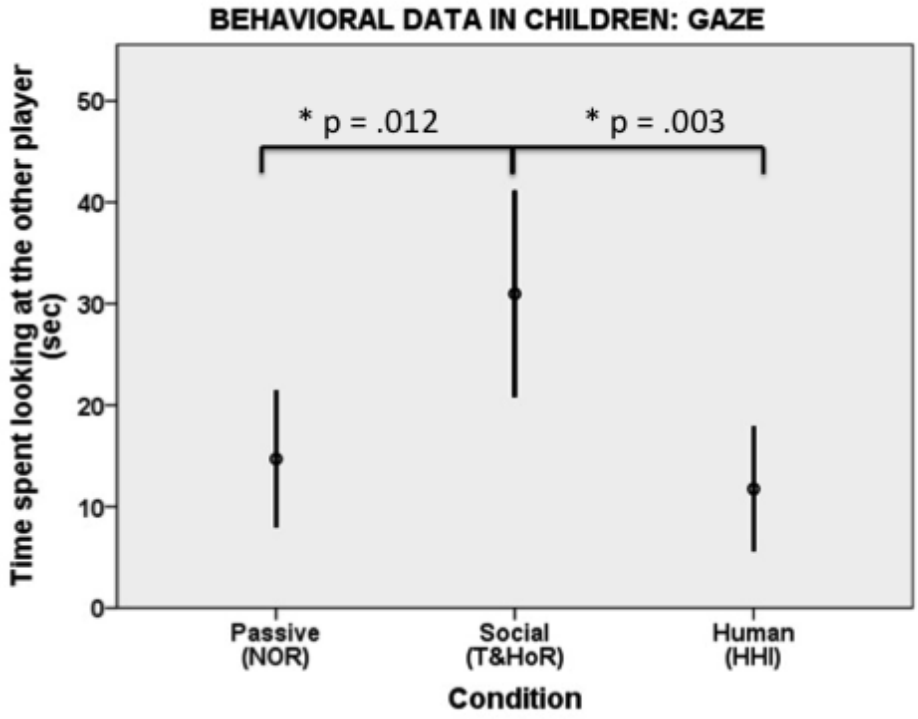


Figure 7.14: Time spent looking at the other player (in seconds) in children among conditions. Asterisks "*" depict significance.

7.3.4 Discussion and conclusions

The goal of this study was to evaluate components that were previously found to be affecting psychological plausibility in the social competence criterion in the task competence criterion of our taxonomy. Hence, we investigate different models of gaze (both as means of a communication channel as a way to support action) and how nonverbal channels of the robot affect knowledge transfer in the context of an educational task. To do so, we devised a pairing game of increased difficulty; the goal of the game was to match each item with the category it belongs. For adults, the game was about geography and players had to match a capital with its country. Children had to pair various types of waste to their containers (recycle bins).

Unfortunately, our results did not support the hypothesis that the behavioural cues of the robot would enhance knowledge acquisition, showing no interaction between the components of social competence and task competence. We could not support the tutoring success of the robot, as adults showed significant differences in all conditions and children no significant differences, although a trend was observed. The results were not sufficient to draw any concrete conclusions about knowledge transfer and we cannot conclusively explain why in all conditions adults showed an improvement.

One explanation could be that the visual feedback from the Reactable was enough to account for knowledge acquisition. Additionally, the results in knowledge acquisition in the HHI condition could not be comparable with the other conditions, as human players did not provide the same feedback (in knowledge or time) as the Reactable or the robot. The reason we investigated the HHI condition was to establish a baseline that would allow us to compare the interaction with the robot and the Reactable to the interaction with a human. However, to understand the interplay of social cues to knowledge acquisition, we need to revise both the pairing task and our control conditions. It is possible that the task, though the difficulty increased with each trial, was still relatively easy. We need to ensure a more challenging task to be able to extract better insights.

Regarding the perception of the robot in adults, humans scored significantly higher in the Godspeed questionnaire compared to the HHI and THI, NoR, ToR but not the T&HoR condition. Such results were expected as a human would score higher than a machine: we would expect humans to be more psychologically plausible. Surprisingly this did not happen in the most social case, where the robot exhibited two gaze models (eye contact and action-sustained) and facial expressions. It seemed that indeed the robot's complex behaviour positively affected humans' perception. However, we did not find similar results in children. A possible explanation could be that a different evaluation method was needed for children, as certain concepts introduced by the questionnaires may not be fully understood by such a young age. In our next experiments, we will be using simpler and more

visual questionnaires that are more appropriate for this age.

What is interesting is the fact that though not significant, in the NoR (the robot was looking at a fixed point on the table) condition, the robot's evaluation was lower than the Reactable in adults. Here, we can see the interplay of social competence and morphology: it seems that although the embodiment is important Dautenhahn (2007), it may have a negative effect if it is not accompanied by social cues. This effect would be associated with the mismatch between perception and expectation. Humans may have built a model of cognitive abilities based on the robot's physical shape, that was not met by the synthetic agent's behaviour. Players may have felt that they were not being addressed by the agent. This mechanism was supported by Ham et al. (2011), where the incorporation of gestures resulted in more positive evaluations only when eye contact was present.

Our behavioural data indicated that adults seemed to look more at the robot in the most complex social condition compared to the other ones. Surprisingly, in the HHI condition, players did not appear to look at each other much. Similar results were observed with children. The fact that humans did not speak much when playing the game could explain the lack of eye contact. What distinguished the T&HoR from the other conditions was the fact that the robot displayed eye contact and facial expressions. Such complex social cues may be more salient and attract the attention of the participant and may contribute to the social competence of the robot. Just the spoken utterances (what humans did not have) could not explain this difference, as the robot was speaking in the exact same way in all conditions. This behaviour is important for the development of social and educational robots, as gaze following directs attention to areas of high information value and accelerates social, causal and cultural learning Meltzoff et al. (2010). Indeed, such cues positively impact human-robot task performance with respect to understandability Breazeal et al. (2005).

7.4 The validation of H5W_STA on an educational scenario

In this section, we evaluate the H5W_STA's task competence in an educational scenario, where the robot teaches a physics task to children in the ecologically valid environment of a school. As a robot, we chose the Nao instead of the iCub because children seemed to prefer it from other humanoid robots and also, as the Nao being smaller than the iCub, it was safer to install in the school premises.

7.4.1 Science–based educational scenario

We designed an interaction scenario in real–life settings based on an inquiry–based learning task. Typically, inquiry–based learning tasks involve active exploration of the world by asking questions, making discoveries and testing hypotheses. The proposed scenario aimed at teaching children about physics concepts based on the Piagetian balance-beam experiments. Exploiting a formal teaching scenario allowed us first to perform a separate validation study to evaluate the minimum set of tools needed to efficiently and effectively teach children the physics task. To do so, we utilised three different mediums for content presentation in the context of the Balance Beam scenario: a physical scale, a virtual scale, and an augmented reality application coupled with the Smart Balance Beam (SBB). Consequently, we utilised the most appropriate tool with the synthetic agent, where we varied the robot's behavioural strategies when providing help to the student.

The balance beam task

The usage of the balance beam task in the present work constitutes a simple inquiry learning task where children's performance can be fully described in terms of the application of a hierarchy of rules of increasing complexity that can be operationally controlled.

The *balance beam problem* was first described by Inhelder and Piaget to characterise and explain children's stages of cognitive development Inhelder

and Piaget (1958). Following the Piagetian work, Siegler Siegler (1976); Siegler et al. (1981) developed a methodology which allowed him to classify children's cognitive developmental stages on the base of four rules of increasing complexity that children of different ages would apply while solving the balance beam task.

Briefly, in the balance beam scenario, different numbers of weights are placed at varying distances from the fulcrum on the equally spaced pegs positioned on both arms of the scale. Children explore the physics of the balance problem using tangible materials and are guided by an artificial agent (e.g., a robot or the content presentation apparatus) that serves as the physical manifestation of the STA. Students are then asked to predict the behaviour of the beam given the configuration provided: if it will stay in equilibrium, tip to the left or tip to the right. To succeed in this task, they have to identify the relevant physical concepts (i.e., weight and distance) and understand the underlying multiplicative relation between the two variables (i.e., the "torque rule"). The goal of the interaction is that the learner acquires knowledge about balance and momentum by going through a series of puzzle tasks with the balance beam. The artificial agent is there to encourage the students, to help them get through the different tasks and to provide feedback; thus, learning improves by continuously monitoring the learner's progress.

The puzzles we provided have four levels of increased difficulty, matching Siegler's rules (Figure 7.15):

- Level I: different weights are placed at the same distance from the centre of the balance.
- Level II: same weights are placed at various distances from the centre of the balance.
- Level III: different weights are placed at different distances from the centre of the balance.

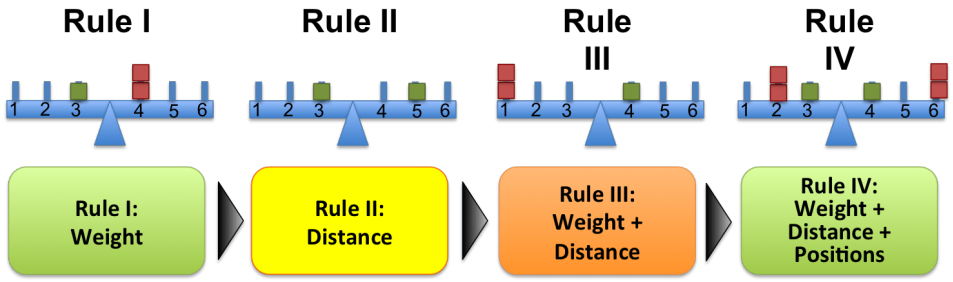


Figure 7.15: Schematic illustration of the four rules assessed by Siegler Siegler (1976). At each developmental stage one or both dimensions (i.e., weight and distance) are considered. Here we consider two weights: green and red (red is twice as heavy as the green). For example, Rule I exclusively considers the weight, whereas Rule III considers both weights and distance from the fulcrum.

- Level IV: follows the principles of Level III, however now the number of weights at each side varies.

For the balance exercises, we have devised two types of tasks. In "Task 1", participants are given a predefined configuration of weights and their distances and are asked to predict the behaviour of the scale (i.e. "tip left", "tip right", "stay in balance"). In contrast, "Task 2" provides the users with the desired outcome and they have to place the weights in the appropriate configuration. Here, difficulty levels are represented as "rules" or "constraints". For example, in Level I, users can place one weight at each side, however the distance must be the same. In Level II, they can put the weights at a distance of their choice, however each side must have the same weight. Similarly, in Level III, distance did not matter, however each side had to have a different weight, and in Level IV, they had to use two weights per side. In our tasks, we have implemented two kinds of weights: red and yellow (weights twice as much as the red).

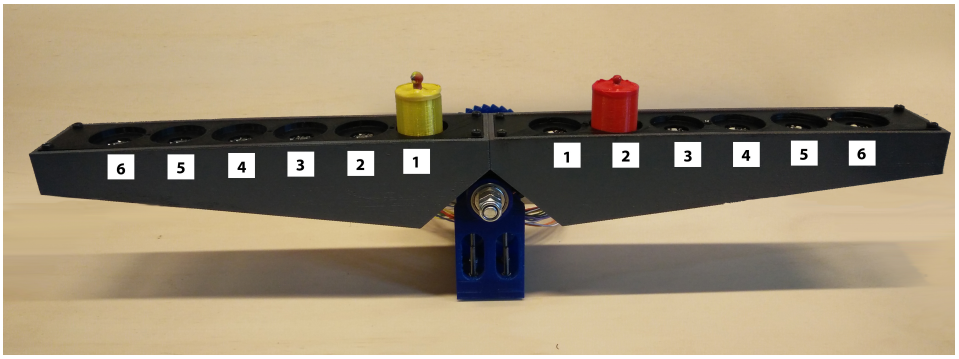


Figure 7.16: Picture of the physical balance beam with a yellow weight placed in position number one on the left side of the fulcrum and a red weight placed on position number two on the right side. Given the fact that the yellow weight is twice as heavy as the red, the scale is in equilibrium.

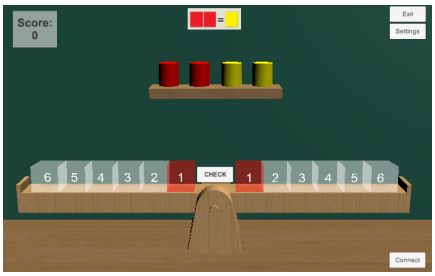
7.4.2 Validation of the content presentation tools

To present the content of the Balance Beam we have used three different tools. A physical scale, the EASELscope and the Smart Balance Beam (SBB). The physical scale consists of six equally spaced placeholders per side on top of which object weights can be placed 7.16. All weights have the same size, and the difference in weight lies in the number of magnets they contain. To visually discriminate between them, weights are also represented by colours (i.e. red, yellow and green). Both the scale and weight are 3D-printed (MakerBot Europe GmbH & Co. KG, Germany).

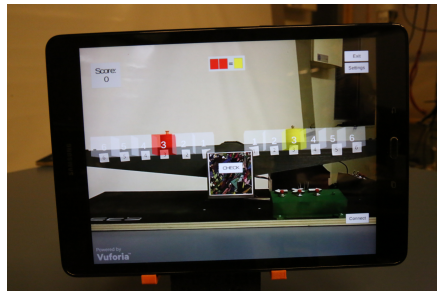
The EASELscope

The EASELscope is a handheld device (tablet) that employs 3D multimodal content management and offers a Virtual and Augmented Reality interface. The Graphical User Interface (GUI) of the application is realised using the platform independent Unity 3D engine[7]. We used the Vuforia Qualcomm[8] augmented reality library to display 3D objects with frame and image markers independent of the position of the objects in the environment.

[7] https://unity3d.com/
[8] https://www.vuforia.com/



(a) VR application.



(b) AR application.

Figure 7.17: The Virtual and Augmented Reality applications. Example of the vR when an exercise is generated. The user has to place the appropriate weight (in this case red) to the indicated positions (position "1" in the left and right side of the fulcrum). Example of the AR application where the physical balance (SBB) is superimposed with the content generated by the tablet.

As a mediating device for the balance beam task, the EASELscope is responsible for presenting a predefined configuration of weights for "Task 1" where users provide their answer by choosing the desired outcome (button). For "Task 2", the apparatus provides the desired result and users must place the appropriate weights and their distances from the fulcrum on the scale.

The main difference between the Virtual (VR) and Augmented Reality (AR) lies on the typology of the task used: in the VR case, a virtual beam is graphically presented on the screen (Figure 7.17a). Once an exercise is generated, the user has to place the correct objects on the virtual scale. In the AR case, the EASELscope superimposes the task configuration on top of the SBB, and the user has to manually place the physical weights to the required positions (Figure 7.23b). In both cases, only when the setup matches the one of the exercise, the user can provide an answer about the outcome of the balance beam. Each exercise is defined by a difficulty level which depends on the number of variables that are manipulated. User replies are given via the EASELscope by pressing the corresponding buttons ("left", "balance", "right").

The Smart Balance Beam

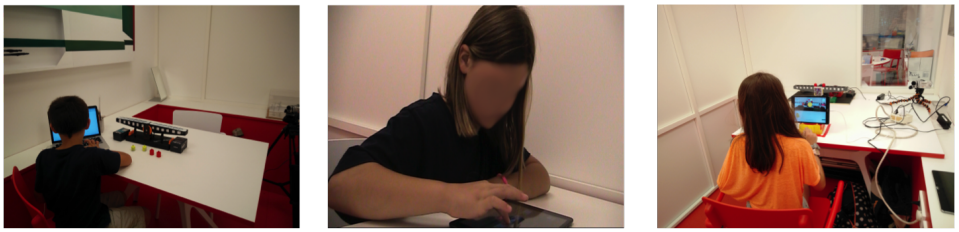
The Smart Balance Beam (SBB) is a motorised beam consisting of a servomotor to animate the beam, an Arduino micro-controller and weight objects. Both the SBB and the physical balance share the same construction (3D-printed), shape and weights. Each placeholder consists of a LED and a colour sensor to inform the STA the exact location of each object. The full set of sensors of the SBB allows for the acquisition of information about the user's actions (e.g. weights selected, the position of weights, reaction times, etc.) that can be used to further analyse the interaction throughout each experimental session.

The SBB can operate as a standalone device or it can be interfaced with the EASELscope tablet. This versatility allows for testing different configurations depending on the aim of the experiments. In the standalone mode, the Arduino micro-controller calculates the product of the weights and their distances from the fulcrum for each side and animates the scale accordingly. When interfaced with the EASELscope (Augmented Reality), all computations are performed externally. Information about the weights and their distances is sent to the Exercise Generator (a module of the STA that generates the exercises) which calculates the outcome of the scale and sends the corresponding command to the Arduino which in turn performs the animation. To overcome the registration problem (objects in the real and the virtual worlds need to be properly aligned to each other to maintain the illusion), an image marker is placed on the scale.

Methods

In total, 76 children (39 females) from two elementary schools in Barcelona took part in the study (age: 9-10 years old). They were all Catalan and Spanish native speakers. All experiments were conducted in the facilities of each school.

Children were randomly assigned to one of the three conditions 7.18:



(a) (b) (c)

Figure 7.18: Example of the experimental setup for the physical balance (a), the Virtual Reality (b) and the Augmented Reality (c) conditions.

- **Physical Balance (PB)**: participants interacted with the physical non-motorised balance. In this task, subjects first had to place the weights on the balance beam, report their answer and their confidence level and then remove the supports from the balance to observe the outcome. (28 participants)
- **Virtual Reality (VR)**: participants interacted with the VR version of the EASELscope. Once an exercise was given, they had first to place the objects to the corresponding positions, provide an answer, report their confidence level and then observe the outcome (animation of the beam). (29 participants)
- **Augmented Reality (AR)**: the setup comprised of the SBB and the EASELscope. Participants used the EASELscope to superimpose the provided information. Similarly to all conditions, participants had to first place the objects correctly to the scale, provide an answer and confidence level (via the EASELscope) and then observe where the scale would fall (animation of the SBB). (19 participants)

Experimental setup and protocol

To evaluate the three different educational tools we conducted an experiment at two schools in Barcelona. Children were first introduced to the scope of the experiment and were given an example of

Data collection

Upon introduction, children filled a questionnaire that included: demographics (gender, age), if they owned a tablet or a smartphone (access to smart technology), presentation of their weekly activities and a set of exercises of the Balance Beam.

To assess how children spend their time, we presented a set of activities in the form of playing cards (such as: doing sports, playing video games, dancing etc.) (see Figure 7.19). We asked children to select the five activities they perform the most and order them by the time they spend.

The pre-assessment questionnaire served as a way to assess their initial knowledge about the balance beam physics task. This questionnaire consisted of 8 predefined configurations of "Task 1" (2 per level) where children had to decide where the scale would fall. Additionally, we provided them with four questions of "Task 2" where the outcome of the balance was given and they had to decide where to place the weights. In two exercises, one weight was already placed on the scale and in the remaining two, no weights were placed on the scale. For all exercises, children had to provide their confidence level (10 item scale where 0 corresponds to not confident at all and 10 to completely confident).

The post-assessment questionnaire consisted of a set of exercises similar to those of the pre-assessment phase. Again, students had to report their confidence level.

This conclusive phase allowed us to estimate possible effects on performance improvements (compared to the pre-assessment phase). Following the methodological approaches suggested in Charisi et al. (2016), we evaluated the task itself and the interaction between the child and the tablet using the Fun Toolkit Read and MacFarlane (2006) questionnaire where we asked children if they a) would do it again, b) would recommend it to a friend and c) found the task difficult. The usage of the Fun Toolkit ensured that the questions directed to children did not have a too high complexity.

Figure 7.19: A sample of the playing cards used to assess the weekly activities performed by the kids.

Additionally, we extracted log files from the interaction regarding the exercise provided as well as reported confidence; error rates, performance, reaction time, possible mistakes in reproducing a given configuration were acquired by the log files obtained from both the EASELscope as well as the EASEL modules responsible for the interaction. Finally, we obtained behavioural data from the interaction using video recordings.

Results

Regarding the evaluation of the task and the interaction with the content delivery tool, children overall reported that they liked the activity and would do it again. No statistical differences were found among conditions $\chi(2) = 1.4$, $p = 0.5$. Means and standard deviations for the different conditions are PB (4.83 ± 0.38), VR (4.5 ± 0.78) and AR (4.8 ± 0.54). Similarly, all children reported that they would totally recommend the activity to their friends and no statistical differences were found among conditions. No statistical differences were found among conditions $\chi(2) = 0.94$, $p = 0.62$. Means and standard deviations for the different conditions are PB (4.9 ± 0.34), VR (4.6 ± 0.78) and AR (4.9 ± 0.31). Finally, regarding the difficulty of the task, children overall evaluated the task with average difficulty, however, no differences were found among conditions $\chi(2) = 1.1$, $p = 0.58$. Means and standard deviations for the different conditions are PB (2.1 ± 0.78), VR (2.38 ± 1.1) and AR (2.37 ± 1.01). Comments received regarding the task included: "We almost never work with technology and it has been very fun", ""I have learned a lot about robotics", "It has been fun, I would do it again", "It has been very interesting" and [...]"It has all been very but very fun, I'm sure that my classmates will like it a lot, thanks to both of you and to all the other boys and girls that work for this thing of doing robots".

Regarding the performance, the participants of the physical balance (PB) performed significantly higher compared to the Virtual Reality (VR) ($p = 0.004$) and Augmented Reality (AR) ($p = 0.005$). In total, participants in all conditions performed better in the post-questionnaire compared to the pre-questionnaire, with significant differences in the physical (PB) ($p = 0.029$) and virtual (VR) ($p = 0.035$) reality.

Finally, if we split participants in low confidence and high confidence, we observe that participants that reported low confidence significantly improved in the post test, whereas no significant difference was found in the improvement between pre- and post-tests in the high confidence (Figure 7.20).

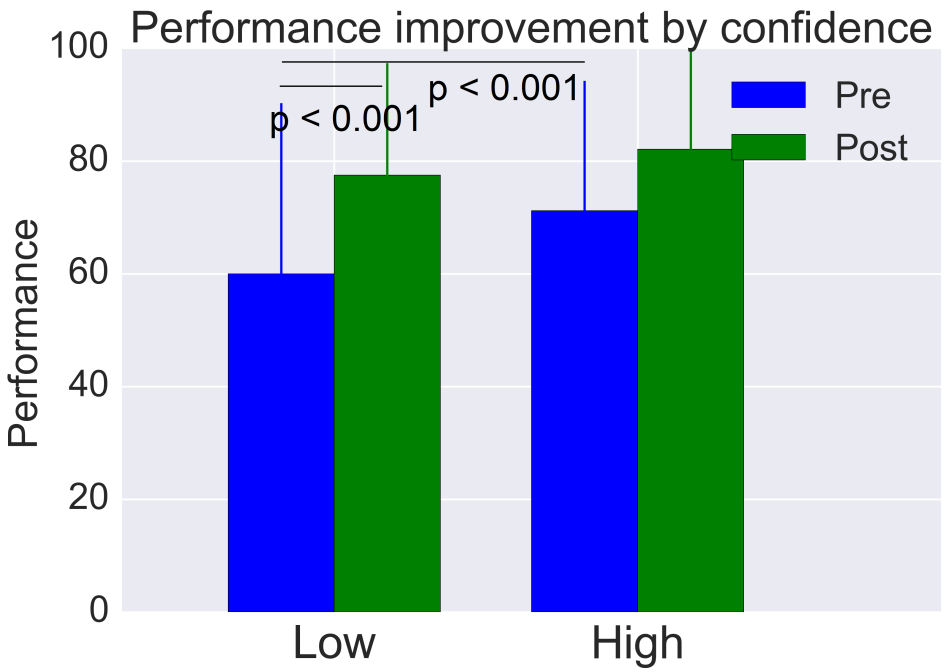


Figure 7.20: Differences in performance between participants who overall had low confidence and high confidence.

Additionally, we found gender differences between the reported confidence as males reported significantly higher confidence compared to females (see Figure 7.21).

7.5 Conclusions and discussion

Overall, we found a ceiling effect in the performance in all conditions. This means that contrary to literature, students of 9 years old are already able to solve all levels of the Balance Beam task. Additionally, regarding the interaction with the content presentation tools, we found that children enjoyed the task, they would do it again and overall evaluated it with average difficulty. The fact that participants performed significantly better in the PB compared to the AR and VR conditions could be due to the assignment of participants to conditions. In future experiments, participants will be ran-

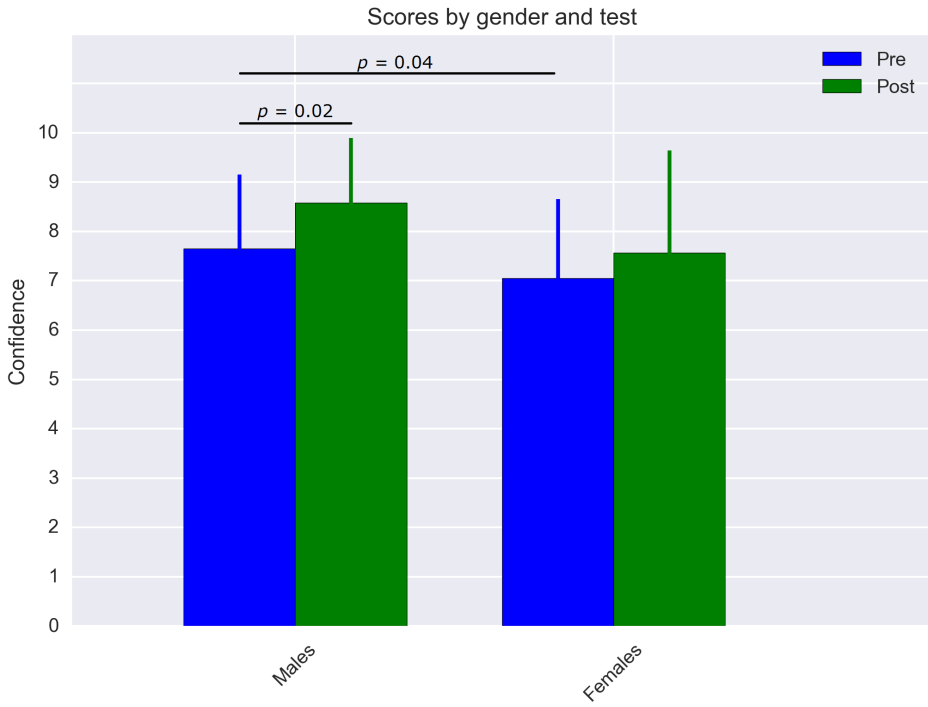


Figure 7.21: Differences in confidence between males and females.

domly assigned to the various experimental conditions in a way that will try to ensure that no baseline biases occur (for example based on their performance on the pre-questionnaire). Additionally, the simplicity of the tool might allow children to better focus on the task, whereas in the AR or VR conditions include the usa of technology that can be considered distractive.

7.5.1 Validation of the robot's strategies. Does distraction help?

When developing an artificial tutoring system, it is important to pay attention to the content provided (for example the difficulty of the task). However, equally important is knowing what the system should do when the learner finds himself in a situation where he needs help. For example, the kind of feedback the learner receives during a task plays a significant

role in the enhancement of learning and the improvement of student achievement. By feedback, we mean information provided by an agent regarding the performance or understanding of an individual.

There are numerous way to provide informative feedback, as feedback may be **relevant to the task** (correct or incorrect answer), the **processing of the task** (making sure the learner understands how to complete successfully the task), **self-regulation** (encouraging and informing students how to better continue with the task) or may be **directed to the "self"** (like praising one's intelligence or responses) Hattie and Timperley (2007). "Doing a great job" or "You will do it better next time" are good for promoting motivation, however, they provide little information on what students are doing wrong Hattie and Timperley (2007). For feedback to be effective, it needs to reduce the discrepancy between the learner's current understanding and what is desired.

Additionally, knowing when to provide feedback is pivotal. In fMRI studies, it is suggested that the striatum supports learning when the feedback is immediate, whereas the hippocampus supports learning when the feedback is delayed Foerde and Shohamy (2011); Shohamy et al. (2004). The results of the study suggest that the contribution of a variety of neural systems in learning is modulated by the time feedback occurs. Although there is evidence that delayed feedback may benefit knowledge transfer, learners seem to prefer immediate feedback Mullet et al. (2014). The positive effects of immediate feedback were highlighted in a meta-analysis of 53 studies Kulik and Kulik (1988). For example, immediate feedback enhances information acquisition, retention and increases the probability of answering correctly in the future a question that was previously answered incorrectly Epstein et al. (2001). Especially for tasks that are demanding, timely feedback is key, as there may be interferences of conflicting information (or alternative rules) held in working memory during the delay Opitz et al. (2011).

The aim of this study is to explore how performance is affected by the various strategies of the Synthetic Tutor Assistant to provide help to the learner

in a dyadic interaction. Additionally, we wanted to investigate how the level of perceived confidence estimated at the beginning of the experimental scenario affects learning and performance in the task. We decompose the help strategies in two main categories: *feedback* and *distractions*. Although various feedback mechanisms have been explored in learning environments, the role of distraction has not been sufficiently examined.

Distraction is the process of shifting one's attention or focus to events or stimuli that block or diminish the acquisition of desired information. Distractions can be internal or external and can be relevant or irrelevant to a specific task. The negative effects of distracting tasks have been studied in a variety of situations, like the impairment of performance while driving Horberry et al. (2006) or performing laparoscopic surgical tasks by increasing the time needed for completion of the task Goodell et al. (2006). Nonetheless, it seems that distraction may affect performance but not learning Eysenck and Thompson (1966). Nonetheless, distraction can be considered as a coping mechanism and therefore beneficial. For example, chimpanzees use self-distraction to accumulate higher rewards Evans and Beran (2007). Distraction is also used as a mechanism that reduces pain experiences or increases pain tolerance Aarts and Dijksterhuis (2000).

Several studies discriminate the positive or negative effects of distraction based on its relevance to a specific task. It seems that congruent (or relevant) distraction facilitates performance, increases response times, reduces forgetting Weeks and Hasher (2014) and opposite effects are observed when distraction is incongruent. Similarly, learned motor skills learned were remembered when distractions were relevant Song and Bédard (2015). Interestingly, in the same study, recall effects were higher when participants were distracted in both learning and recall phases than only in the learning phase.

Here we report a 6 weeks-long in-school study that evaluates the effects of help provided to the learner by an artificial tutor while performing a scientific inquiry-learning task. This study is a follow-up from Experiment

1 where we evaluated the tutoring tools that the synthetic tutor is using in the current study.

Methods and setup

Given the outcome of the previous study, we selected the Virtual Reality application as the most optimal apparatus. The motivation behind this approach is that we needed to gain information regarding the actions of the learner when performing the task, to provide relevant and informative feedback. From the three suggested tools, the Augmented Reality (AR) tool that comprised of the EASELscope and the Smart Balance Beam showed the lowest performance, with no significant differences between the pre- and post- assessment phases. Self-report and video recordings analysis suggested that the poorer performance observed in the AR condition could be explained by the ease of getting distracted by the platform itself, an aspect that was not observed in the other two conditions. Between the physical and Virtual Reality (VR) conditions, we chose the AR, since it provides valuable information regarding the correct or incorrect placement of the weights, the user's answer and confidence level that the lack of sensory elements of the physical balance beam does not provide. As a robot for the interaction, we chose the humanoid Nao, as it was assessed with the highest score in terms of preference (see section 7.2.3).

The observation of a ceiling effect in performance in the previous task led us to target a younger group of children, aged 8-9 years old. All students were Catalan or Spanish native speakers and consequently, experiments were conducted in Spanish. To ensure a sufficiently challenging task and to get the baseline on when to provide help we conducted two pilot experiments. The outcomes of the pilots are discussed in the following section and are used as part of the main experiment.

Pilot study 1

The purpose of the first pilot study was two-fold. On the one hand, it served as an evaluation of the difficulty of the task and on the other hand,

it allowed us to extract a series of parametric information (e.g. average response time, number of errors etc.) to fine-tune the variables studied in the main experiment. For this pilot, we used 7 children.

Regarding the difficulty of the task, we first examined "Task 1" of the balance beam where we added two extra difficulty levels and another weight (green weight, three times heavier than the red). The experiment composed of two conditions: self-requested and system-provided help. In the self-requested condition, children could request help from the robot by pressing a button on the EASELscope. Before providing help, the robot asked the user "What is it that you do not understand?". On one hand, by not providing direct help, we allowed children to verbalise and reason about their doubts and lack of knowledge regarding the task. On the other hand, this verbalisation provided us insights concerning the task itself (e.g. if the child did not understand the instructions). Allowing the learners to request feedback, gave them the possibility to regulate their own learning processes.

In the system-provided help, the robot would provide help in two different cases: if the student took too long to provide an answer and if the student remained on the same difficulty level for a certain number of trials. These parameters were approximated as an average from the outcome of experiment 1.

The results of this study indicated that children were still able to solve the task with little effort. In the case of self-requested help, only 2 children asked for help and the reason they asked for help was that they did not correctly understand the task.

Pilot study 2

Given that children still showed high performance and did not ask for help, a second pilot study was necessary. To increase the difficulty of the task, we employed "Task 2" of the balance beam problem. Children were given the outcome of the balance (fall to the left, right or in balance) and they had to place the weights following Siegler's four rules (difficulty levels). The

reason we chose the second task of the balance problem is that in the pre- and post-assessment questionnaires, children encountered more difficulties in providing the correct answer.

Similarly to the first pilot study, we had two conditions: self-requested and system-provided help. The help that was provided by the robot was following the experimental setup of our main study. For this pilot, we used 12 children.

As expected, we observed a drop in performance, as "Task 2" was considered more difficult, however not all children asked for help. Nonetheless, we got sufficient data to parametrise the main experiment.

Main study

Given the outcomes of the two pilot studies, we used "Task 2" exercises. Given the fact that children did not often ask for help, we eliminated the self-request feedback condition and kept the system-provided help. As mentioned previously, the agent provided help in two situations: if the student took too long to place the weights and therefore provide an answer, or if the student remained blocked in the same level for several trials.

The time after which the robot provided help was extracted from the average time children needed to complete each trial and was approximated to 35 seconds. Similarly, if the student was at the same level, the robot would provide help after the seventh trial for every four trials.

The robot provided two kinds of help: hints or distractions. Hints were further subdivided to "Open" and "Closed". Open hints are relevant to the task in general (e.g. "Remember that the yellow weight is two times heavier than the red weight"). Closed hints were specific to the difficulty of each level (e.g. "Remember that if the distance is the same, what is important is the weight" for level I, "Remember that if the weights are the same, what is important is the distance" for level II etc.).

We divided distractions in two subcategories: "Trivia" and "Jokes". Trivia

refers to known facts such as "Did you know that the male seahorse is the one that gets pregnant?". Jokes are funny stories like "What did the traffic-light say to the other? Do not look at me, I am changing!". The provided jokes were appropriate for the age group of the children. In fact, several of the jokes told by the robot were provided by participants in the pilot tests. On average, we made sure that the robot's utterances last approximately the same to avoid any biases.

The robot in all conditions positively encouraged the learner in correct and incorrect answers (e.g. "Well done!" or "Excellent" for correct answers and "It is ok, you will do better next time!" for incorrect answers). The difficulty of each exercise was adapted by the Allostatic Controller module based on a simple Learner's model that computed online the performance of the user. If the performance was above 75%, the proposed puzzle's difficulty would increase. In contrast, if the student would make several mistakes, the level would decrease. For each exercise, children had to report their confidence level before viewing the answer on the apparatus. The maximum number of exercises was capped to 24. However, from the 16th trial, the synthetic tutor would ask the student if he wanted to continue and the child provided the answer via the EASELscope.

To avoid novelty effect, we introduced the robot to the children during class where we provided information regarding the interaction with the robot. Additionally, we asked them not to share any information regarding the nature of the interaction with the rest of their classmates that had not yet taken part in the experiment to avoid any biases. The experimental sessions consisted of five main phases: an introduction, a pre-assessment questionnaire, the intervention phase, a post-assessment questionnaire and a semi-structured interview.

The introduction served as a way to explain to the child the nature of the task and provide an example of the interaction with the tablet. Upon the completion of the second phase, the experimenter guided the child to the room where the Nao was located and briefly introduced the robot, the

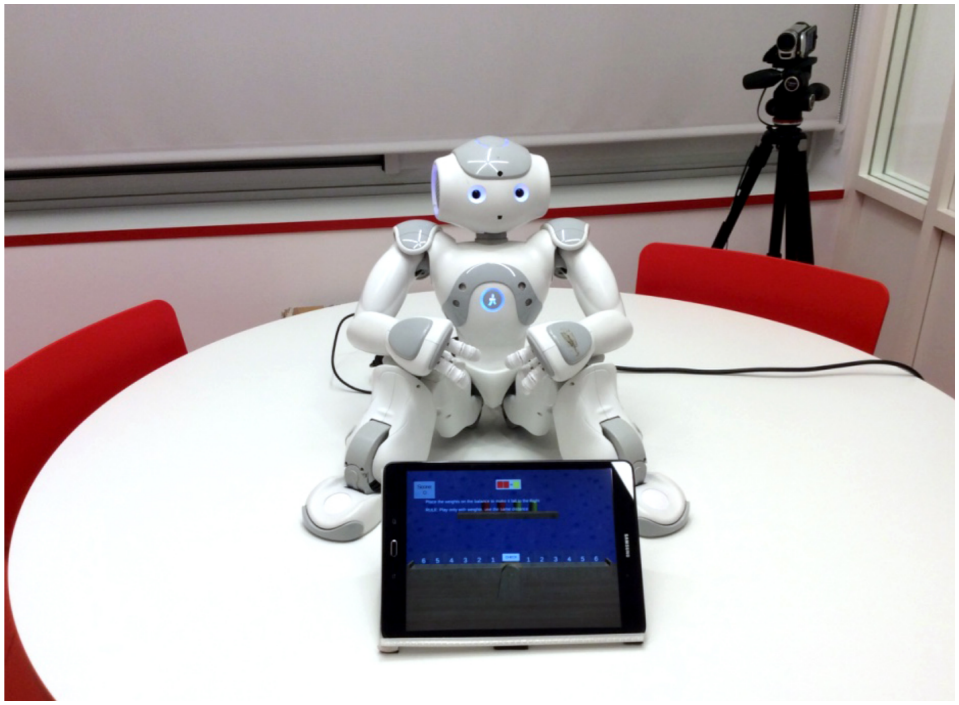


Figure 7.22: The experimental setup. The child sits in front of the robot and interacts with both the robot and the EASELscope. The EASELscope is used to present the different exercises and get the answer from the child. During the interaction, the synthetic tutor (Nao) looks at the child and provides feedback according to the child's actions.

EASELscope and the task (see figure 7.22).

The interaction began with the robot introducing itself, the task and the kind of help it would provide. When the robot finished the introduction, the first exercise would appear on the tablet. Throughout the whole experiment, the robot would maintain eye-contact with the child while occasionally looking at the tablet. Though the robot's facial expressions are limited to the colouring of the robot's eyes, we accompanied the robot's shy or happy speech with the appropriate animations that are already installed in the robot.

Finally, when the intervention phase terminated, the experimenter guided

the participants to the first room where they completed the post-assessment questionnaire and the semi-structured interview.

We examined the following hypotheses:

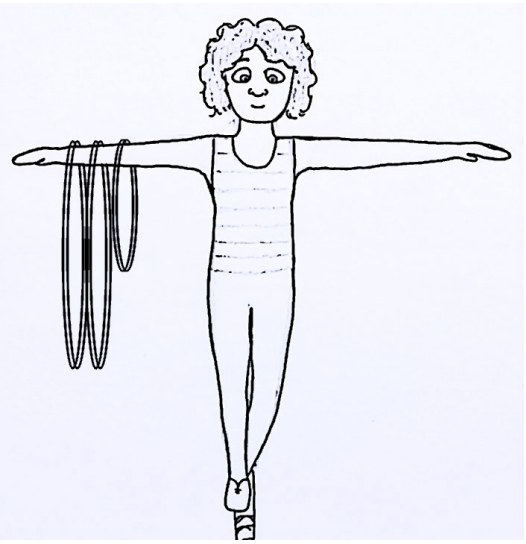
- participants in the closed hint condition would perform better than those of the open hint, given the fact that the feedback they receive is not general; in contrast, it is relevant to each level
- overconfident participants would persist with the task without considering performance and/or feedback
- underconfident participants would give up faster due to their expectancy of low success
- children would prefer the jokes than the trivia feedback because it is inherently more fun

The study was approved by the Ethical Committee of Universitat Pompeu Fabra. Before the study, the parents of the children signed an informed consent letter that allowed them to participate.

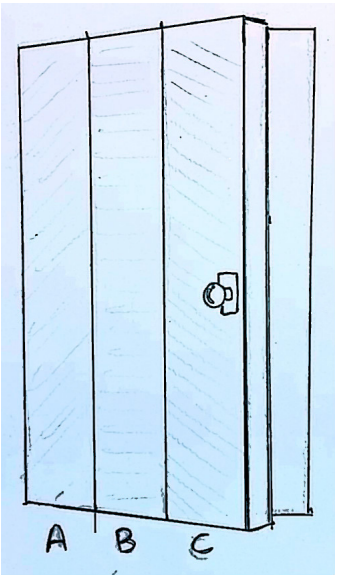
The data collection followed the principles of experiment 1. However, to evaluate the difficulty of each exercise, we asked children (in both pre- and post-assessment questionnaires) to evaluate the difficulty of the exercise (5 point scale)

Additionally, to assess if the kids could generalise the balance beam principles to a different problem, we provided them with a set of three questions (see Figure 7.23). Again, students had to report their confidence level and evaluate the difficulty of the exercise.

Finally, we added a semi-structured interview (see appendix A.1) to gain better insight regarding children's views of the robot and the interaction and we obtained behavioural data from the interaction and the semi-structured interview using video recordings.



(a) Balancing acrobat: if one large circle weights twice as much as a small circle, how many circles does the acrobat need to put to his empty hand to maintain equilibrium?



(b) Closing a door: which part of a door (a), (b) or (c) will you use the least force to close it?

Figure 7.23: Examples of questions aimed to assess children's ability to generalise the balance beam principles.

Results and conclusion

We deployed the system for 18 days in the two schools in Barcelona. Overall, the system has been running for 108 hours and was very robust. In only three interactions the robot crushed and had to be restarted. Nonetheless, we could resume the session without a problem as the module responsible for the interaction allowed to restart a session from a specific level/trial or moment without losing any data. During the interview, we asked participants to report any perceived errors the robot might have made. None of the reported the robot making any mistakes.

We found no significant differences in performance between conditions as shown in Figure 7.24 though a trend can be observed.

Interestingly, we found significant differences in confidence between "open"

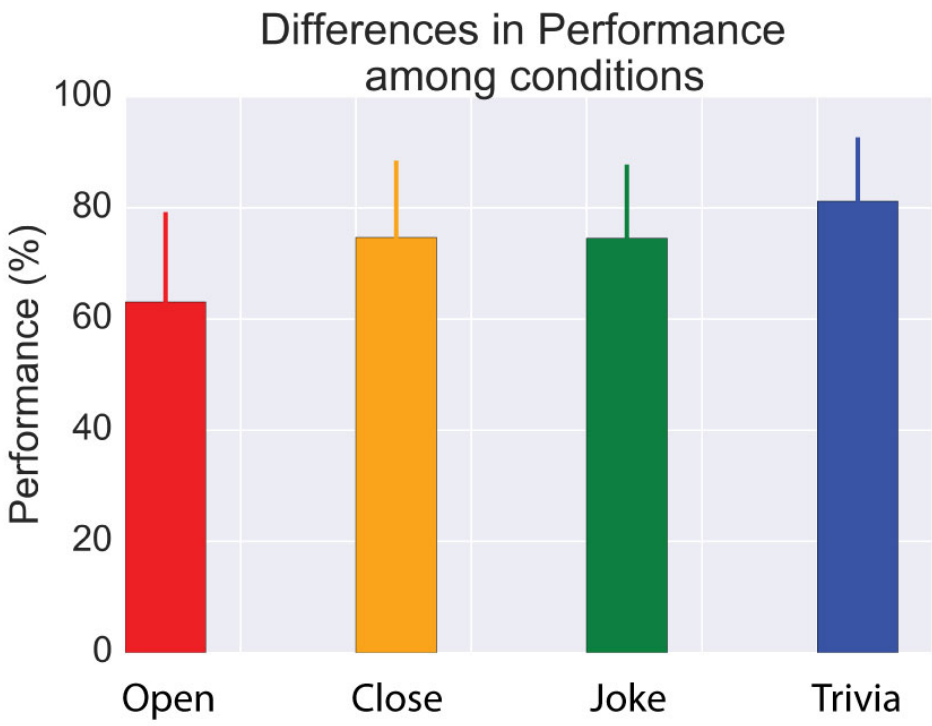


Figure 7.24: Overall performance among conditions. No significant differences in performance were found.

and "close" conditions ($p = 0.05$) and "open" and "trivia" conditions ($p = 0.01$) (results corrected for multi-comparisons). The difference in confidence between "open" and "close" conditions may lie to the nature of the help the robot provides. In the "open" condition, the help is very general to the task; in contrast, in the "closed" condition, the help is relevant to the level. Students might have felt more confident in answering a specific question when they consider help to be more relevant than when not. To explain the significant difference in the level of confidence between the "open" and "trivia" conditions, extracting information from the semi-structured interviews may shed some light. When we asked children if the robot helped them, and in what way, many of them that were in the "joke" or "trivia" reported that the robot had helped them. Children reported jokes as help-

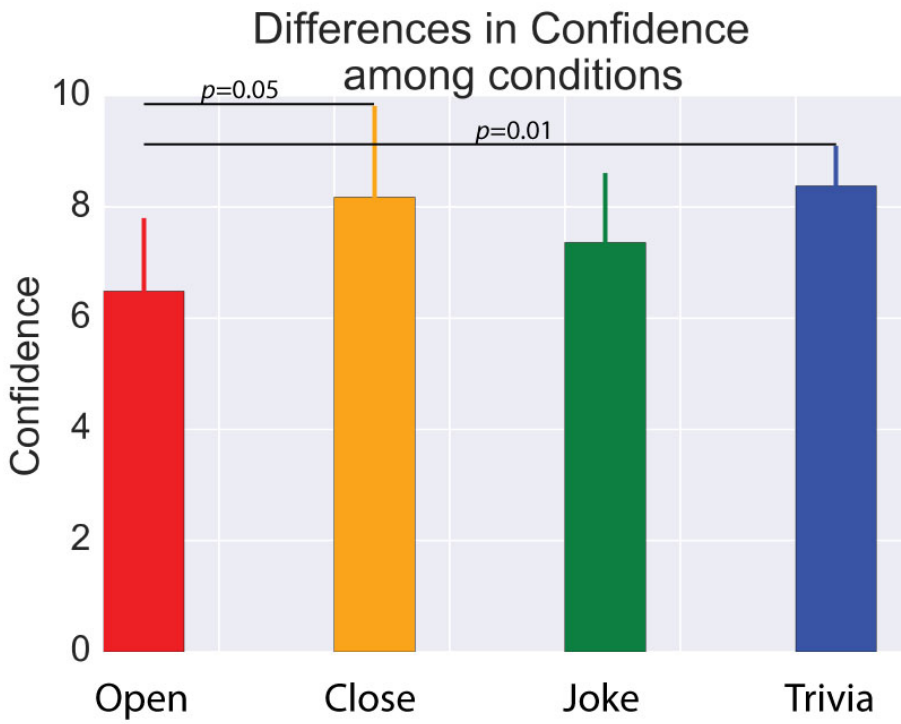


Figure 7.25: Overall difference of confidence among conditions. We found significant differences in confidence between "open" and "close" conditions and "open" and "trivia" conditions.

ful because they were funny and acted as a distraction or even alleviated possible stress. Trivia knowledge was also helpful because the robot made them feel they already knew more than before.

In general, when children were asked if the robot helped them, most of them said yes. When we asked in what way, most replied that the robot was helpful because it was reminding them that they did not place correctly the objects, or overall, because it was giving them instructions or explanations regarding the task.

Regarding the evaluation of the overall task, 70% of the children reported that they would definitely repeat the activity and only 2 of them said they probably would not. Similarly, 76,1% of the children said they would defi-

nitely recommend this activity to their friends/family. Finally, the majority of the children did not find the task difficult. When children were asked to write any comments or if they would change anything in the activity (a non-mandatory procedure) some wrote: "Many thanks to the Nao that helps everyone", " I would not change a thing" or "no it is very entertaining". Additionally, when asked could the robot see what you were doing, most of the children replied yes. When we asked them to reason about it, their replies varied but could be classified in three main responses: "because it has eyes", "because it was responding to what I was doing" or "because it could look at me".

Overall, this experiment allowed us to validate the behaviour of the robot in three out of the four criterions of our proposed taxonomy. The analysis of the data is ongoing and not all aspects of the experiment have been analysed, however, in the presented preliminary results, we can observe that children overall enjoyed the experiment and the robot. Based on children's answers in the semi-structured interview, in terms of task competence, the robot appeared helpful to almost all students; additionally, children believed that the robot was aware of what the student was doing and they all claimed they enjoyed the interaction. Finally, when we asked children if they learned something new about the task, they all reported that they learned how to use the scale and how to make it in balance. Some also reported learning jokes or trivia, however, further quantification of the replies is needed to be able to draw any conclusions.

In terms of social competence, when we asked participants if they liked the robot and if they found it friendly, they all replied yes and they found it very funny and kind. When we asked them if the robot found them friendly, children either reported that they did not know or that the robot indeed found them friendly, because it was responding to them in a nice way and encouraging them. In terms of autonomy, the whole system was operated continuously for many hours and only three times we had to restart the application, and there, mainly because the motors of the robot were over-heated and the robot stopped responding. In addition, during the interview,

none of the children reported the robot doing any mistakes or even restarting the interaction. Hence, we validated the prototypic H5W_STA that is controlled by the DAC architecture as explained in sections 5 coupled with an action-sustained gaze model and emotion expression. Children enjoyed the interaction with the robot and our prototype was partially validated as a psychologically plausible agent, however, further analysis of the interaction, the behavioural data and the exercises are needed to extract more concrete information.

Chapter 8

Conclusion

With the continuous advancement of technology, the introduction of robots in our daily lives is only a matter of time. Hence, a fundamental question that all roboticists will have to answer is: "How can we create robots that are successful in interaction and are accepted by people?". We argue that acceptance is important because it defines whether humans will use the robot and interact with it. The challenge of answering this fundamental question lies in the definition of success in interaction and the measurement of acceptance. We propose that a key determinant is the psychological plausibility of the robot. More specifically, we view humans as active modellers of the world or hypothesis-driven systems that use internal models of the world, make predictions and learn from their mistakes. Hence, violations of a user's predictions may lead to the rejection of an observation and consequently its plausibility.

We argue that for a robot to be accepted by people, it needs to be believable, or what we would call psychologically plausible, and its behaviour should fall within human expectation or prediction. We propose to decompose plausibility into discrete parts that we can understand and test empirically. By understanding the interactions of the robot's components and how they affect plausibility, we can define a mini psychological engine (or a set of features if you will), that will allow us to construct robots that

individuals can perceive as believable agents. To do so, we offer four benchmarks for consideration: *morphology*, *autonomy*, *social competence* and *task competence*.

Before we go into further details regarding the proposed taxonomy, we take a step back and examine early attempts to create convincing machines. More specifically, in chapter 2, we focused on archetypes that are characterised by efforts made to understand biological organisms and imitate their physical appearance, functionality and complex life-like behaviours. Although these machines might not be considered autonomous (despite the fact that they work on their own) or social, their creators took special care for their morphology and task competence. For example, to create a machine modelled after the duck, Vaucanson ensured that his creation not only resembled one externally but also imitated key actions of the living animal, like flapping its winds, eating, drinking etc. Vaucanson also tried to reproduce the internal biological processes of the duck, as shortly after it eats, it defecates.

We drew attention to machines that morphologically and behaviourally resembled humans. In contrast to Vaucanson's duck, these automata intended to depict external human activity rather than reproduce any biological process. Although their functionality varied from playing a musical instrument to writing or drawing, their creators faced the same challenge: they needed to be believable and convincing. To achieve plausibility, their appearance resembled a human, and their task competence resulted from mechanical ingenuity and extensive studies of human anatomy. For example, the "Flute Player" was equipped with a set of movements that were necessary for a person to play the flute: it moved its fingers and varied the position of its lips while modifying the air exertion through its mouth to play a higher or lower pitch. At the same time, the automata makers of that period implemented a set of behaviours that usually accompanied the tasks their automata performed. The most common feature in the Jaquet-Droz creations was the gaze behaviour: all automata looked at where they acted. Interestingly, the draughtsman occasionally stopped, raised its head presumably to examine his work better and then resumed its drawing. The writer would dip its

quill to the inkwell and then shake it lightly to remove any excess. Finally, the musician displayed breathing patterns and torso movements similar to those of a real organist. These anthropomorphic behaviours closely resembled those of humans (breathing, shake the ink off the quill, examine one's work). What is more, the gaze behaviour may have also provided a false impression of "awareness" of their own actions and the surrounding environment or even possibly the attribution of Theory of Mind. Taken together, this set of behaviours intended to make machines plausible, as they would fall within human expectations.

However, the automata of that period were purely reactive and could not perceive the world and act upon it. If we want to create robots that socially interact with humans, they are required to understand their environment and respond accordingly; hence, sensing is essential. A key challenge for robotics is to produce convincing behaviours with response to the environment. Walter's and Braitenberg's approach was to link perception and actuation to exhibit plausible behaviours. Despite their simplistic design and control, their creations shared a common hypothesis: the production of complex and convincing behaviour lied in the interplay between perception and action, which in turn are responsible for the competence they show for a predefined task.

The more sophisticated a robot becomes, the more challenging its control is, and different approaches than directly connecting a sensor with an actuator are required. The use of a layered controlled architecture to control such complex robots is lately gaining ground. *Shakey* was the first robot controlled by a layered architecture and became the basis of contemporary robots. The approaches of automata makers and early robots provide modern roboticists with a set of tools, guidelines and behavioural features that can be used to create psychologically plausible agents. Although there are still no direct guidelines regarding the design of social robots, nor commonly accepted validation methods, the existing strategies in developing robots that interact with humans are divided into two broad categories: morphology and behaviour.

Research indicates that the morphology and behavioural features of the robot affect the way people perceive it. Nonetheless, their interactions are not yet clear. Following the example of automata makers, we presented in chapter 3 the typical approach which is to design human-like robots or robots that allow humans to anthropomorphise them, however, we also discussed the implications of non-anthropomorphic design. We argue that what is more important than machine-like or human-like design is the transparency of the robot, that is, allowing people to read its communication channels correctly. Additionally, we presented the uncanny valley, a phenomenon that links the spectrum of machine-like to human-like robot appearance with feelings of familiarity or affinity. According to the uncanny valley, the more human-like the robot is, the higher it will score in affinity. Nonetheless, close resemblance to humans causes negative feelings (or a distinctive drop), but if the appearance of the robot becomes indistinguishable from that of humans, the relation becomes positive again. We argue that we need to redefine the uncanny valley, as there are many cases that the relationship between human-likeness and affinity does not follow the suggested curve. The perceptual tensions might be caused by inconsistencies of the robot's features. Hence other factors (like task or social competence) might explain this phenomenon. Finally, we propose that the suggested measurement of eeriness or familiarity does not explain the perceptual tensions. Instead, a better measurement would be the psychological plausibility or believability that accounts for acceptance of a robot's behaviour or design.

To understand the behavioural traits that allow humans to perceive robots as believable agents, we examined existing approaches that are typically employed in HRI scenarios in chapter 3. The most common approach is to implement anthropomorphic components to robots since humans intuitively apply the same social rules to machines as when they interact with users. The behavioural components we have identified are the expression and perception of emotions, communication via dialogue and the exhibition of motivated behaviour. Indeed, a key component of the generation of behaviour is motivation. We presented the various theoretical approaches to

motivational systems and what most of them seem to have in common is the ability of the organism to maintain a "steady state" or what is called *homeostasis*. The application of homeostatic systems on drives allows the organism to select actions that will enable it to satisfy its needs. However, one of the most fundamental challenges is the orchestration of multiple drives (or homeostatic systems), especially in the cases of conflict. We argue that a solution to this problem is coupling the homeostatic system with an allostatic one. We propose that endowing robots with motivational systems allow for the generation of plausible and robust behaviours that facilitate adaptation and allow the robot to stay focused on its task.

Another critical component for the production of behaviour as well as the social interaction is the implementation of an emotional system. Researchers propose that the functional role of emotions is not limited to communication (utilitarian). Instead, emotions are also involved in the organisation of behaviour (epistemic). We propose that emotions are closely linked to notions of motivation and self-regulation as serving a vital role in the regulation of control systems underlying behaviour and communication. Finally, we examined the role of empathy in HRI scenarios. Empathy allows organisms to predict and understand behaviours and is linked to prosocial behaviours. Research suggests that humans are able to empathise with robots; however, there seem to be a variety of factors that affect the elicitation of empathic responses, such as the appearance, behaviour and personality of a robot. In most cases, empathic responses are linked with the perception of intelligence or the attribution of Theory of Mind: the more plausible a behaviour is, the easier it is for people to accept the robot and empathise with it. We, therefore, suggest that the elicitation of empathic responses can be used as a measurement to assess the psychological validity of a robot.

The research in the field of automata and modern robots led us to identify some parameters that may play a fundamental role in the psychological plausibility of the robot. More specifically, we saw that action-sustained gaze was a feature that was commonly used in the automata of that period, while almost all automata exhibited micro-behaviours that accompanied

the task they were performing. The reason behind this is simple: humans would expect that micro-behaviour to occur and by implementing it, the behaviour of the automata would fall within humans' expectations. Additionally, other features, like emotional or motivational systems not only contribute to the robust behaviour of the robot, but they also seem to affect how humans perceive them. Given the vast possibility of features that one could explore, we provide an alternative method. In chapter 4, we generalise the relevant features in four broad categories or criteria that contribute to the psychological plausibility. These categories aim at covering the basic aspects of robots: how they look (morphology), how well they execute the task they were created to perform (task competence), how well they comply with the social norms of their assumed role (social competence), and how robust is their behaviour (autonomy). In many cases, a robot might fulfil some of the criteria and fail in others; additionally, robotic designers may focus on one approach and neglect the rest. To create a truly believable and psychologically plausible robot, all four criteria should be taken into consideration.

So how can one know what to do to create a believable agent? The way our taxonomy is formulated allows us to further decompose each of these criterions to testable components, study their interactions and examine how they affect human perception. For example, in morphology, the key criterion is to ensure that the robot's design serves its purpose. If the purpose of the robot is to guide visitors in a museum, it would be preferable for the robot to move in space instead of being static. If part of its task is to communicate with human partners emotionally, its design should support some communication channels that allow people to read their internal states.

Another essential criterion is the transparency of its communication channels. Take the case of a highly realistic humanoid robot and a robot that looks like an animal. Each of these robots has their own way of expressing their internal states. The highly realistic humanoid robot will use facial expressions to communicate its emotions or perhaps even body postures, prosodic features or gestures. The robotic animal will use lights, sounds,

perhaps even movement. One might think that the humanoid robot would be favoured in interaction and communication. In the previous chapters, we examined the role of human-like characteristics in communication and how they may facilitate the interaction, given the fact that similar to human communication channels might be easier read and understood.

However, the exact opposite may occur. Humans may prefer to interact with a robotic animal and not the realistic humanoid robot. Users would not expect an animal to use lights to communicate, and they would expect a human to use facial expressions, so why would they prefer or accept the animal and not the humanoid robot? The answer lies in expectations and the way the emotional expression is delivered. For example, expressing emotional states with light is quite straightforward: internal LEDs will emit light at a particular frequency and pattern. In contrast, the expression of emotion in a realistic humanoid face is a much more complicated procedure: the corresponding motors should synchronise to produce the expression and the time of execution is key. One may argue: this is the definition of the uncanny valley because the highly sophisticated face of the robot may fall within the slope of creepiness. Indeed preference might be biased toward the animal robot, but not because of the appearance of the robot resembles that of humans and falls within the uncanny valley's curve, but because its behaviour might not meet the expectations of the observer. Most humans have little problems in recognising a face, that is because during their lifetime, they have been trained by seeing thousands of them. People know how long it may take to smile, how a smile looks like, so if a robot's face highly resembles that of a human, they would expect similar behaviours. According to Moore (2012), these perceptual tensions may be due to anomalous movement of the motors or similar inconsistencies. Based on the uncanny valley theory, the robotic animal would not have scored high either, given its non-human appearance.

The robot's expressions need to be transparent and understandable. The realism of the face is not important if it fails to communicate its internal states correctly. For example, the humanoid's smile might be mistaken

with disgust or fear expression with surprise. What we therefore suggest is, that for the robot to be plausible, its morphology and design should fall within human expectations and correctly communicate the intended emotional cues.

Regarding autonomy, one may ask "What makes a robot autonomous?". We view autonomy in terms of performance and how humans perceive it. Regarding performance, we suggest that a robot is autonomous when it operates robustly without the need of human intervention. From a psychological perspective, people perceive an entity as autonomous when they infer motivated, intentional or purposeful actions or when they attribute Theory of Mind to it. Hence, to be plausible, we propose that the robot operates in an autonomous and robust way and that humans perceive it as an entity who is guided by internal motivations and performs actions that serve to accomplish its goals.

We define task competence as the ability to perform a task successfully. For example, a robot that acts as a tour guide in a museum is called to interact with humans on a frequent basis. Visitors might be asking the robot for directions or information regarding a particular display, or the robot could randomly approach people and inform them of the newest installation. To access its competency, we can decompose the task into smaller functions: "Does the robot correctly identify and track humans?", "Is it able to recognise the visitors' requests correctly?", "Does it provide the appropriate information?", "Can it recover from error?" and so on. If the robot satisfies the derived criteria and meets humans' expectation about the specific task, we can claim its competency. Finally, social competence measures the robot's social success. There are a number of factors that may affect the social competence of a robot like the expression of emotional states or personality, the use of gestures or other paralinguistic cues, the gaze model, the display of motivated behaviour, to name a few.

The proposed taxonomy allows us to identify, understand and evaluate components that may affect the perceived plausibility of the robot. Here, we

focused on the gaze model and the expression of emotional states, however the same approach can be applied and extended to other factors. We believe that our taxonomy can act as a guideline for the creation of robots that are accepted by humans and an evaluation method for HRI studies.

A core challenge in the creation of plausible agents lies in implementation, as all benchmarks need a control system that generates the appropriate behaviours. In chapter 5 we presented the Distributed Adaptive Control (DAC) architecture as the control system for our robot. The key ingredients of our proposed system are actions, goals and drives: drives are the robot's intrinsic needs and define its goals and actions are generated to satisfy those goals. This engine allows for the generation of autonomous and deeply parametrisable behaviours by solving five essential problems (H5W: Why, What, Where, When and How) in the domains of perception, self-representation and the action. We claim that such an integrated architecture (i.e. encompassing all sensorimotor aspects as well as cognitive processes) is a necessary condition for generating plausible reactions and adaptive behaviours of robots in complex, dynamic and uncontrolled social contexts. In parallel, the motivational and emotional system of the DAC architecture allows users to view the robot as an autonomous agent, guided by its own needs.

Finally using DAC we can achieve fast implementation, prototyping and platform abstraction in any interaction scenario. Most of the exchange of information within the architecture's modules is high level; hence small changes in the Somatic layer (like how information is received and how an action is produced) solves the problem of embodiment in any robot design. The configurability of the modules requires little extra coding and the modularity of the architecture modularity of the architecture allows us to turn on and off various modules. It is this versatility of the DAC architecture that allowed us to implement the presented scenarios.

To assess the proposed architecture, in chapter 6 we presented the behaviour of the robot during an interaction. We demonstrated that the robot's in-

terplay between the emotional and motivational system lead to robust interactions while the robot's actions aimed at satisfying its needs. To assess the transparency of its expressivity, we conducted a pilot study where we systematically varied the facial components of the robot (like eye aperture or mouth configuration) and asked participants to evaluate the expression regarding valence and arousal. The acquired results indicated a relationship between mouth configuration and valence and eye aperture and arousal. Additionally, we used the most salient combinations of eye, mouth and eyebrow configurations in our experiments.

As mentioned in chapter 3, a lot of factors can affect the way users perceive the robot. To evaluate whether complex social behaviours affect user perception, in chapter 6 we presented a study where we manipulated the following components: touch, speech, gaze model, facial expressions, interpersonal distance and proactive behaviour. We defined complexity as the number of employed components, and we evaluated psychological plausibility in terms of anthropomorphism, likeability, perceived intelligence and animacy. Results showed that interactions with the robot displaying complex social behaviours scored higher in psychological plausibility, suggesting that the examined components play a major role in the interaction and perception of the robot.

We propose that another way to assess social competence is by measuring the empathic responses of the user. Empathy is the ability to understand the other's emotional state, and in most cases, empathic responses are linked to the attribution of Theory of Mind to the object of empathy. Attributing Theory of Mind to someone or something allows us to understand, predict and explain its behaviours and act accordingly. If an individual shows empathic responses towards an agent, it might mean that the individual has attributed Theory of Mind to it. We assume that the attribution of Theory of Mind or the elicitation of empathic responses contributes to the psychological plausibility to the agent. Hence, we use empathic reactions as a measurement taken from the observer to evaluate the plausibility of the robot. Two factors that affect empathic responses are the facial expressions

and gaze model, so we manipulated them to assess how they affect the empathiser. Our results suggested that a major contributor to the elicitation of empathic responses is not the behaviour of the robot, but the sensitivity of the empathiser, highlighting the importance of individual differences in any interaction. Hence, if one wants to model acceptance, taking into consideration the individual differences of participants alongside the behavioural components of the robot is essential.

Finally, in chapter 7 we evaluated the task competence of the robot, focusing on the tutoring domain. The reason we chose tutoring is that the effects of the robots task competence can be measured by the performance of the user during the task and any possible increase in the knowledge acquisition after the task. At the same time, the social components of the robot affect the interaction, so we can use this task to evaluate the robot's social competence. The target group for this set of experiments is children, hence we first explored children's preferences regarding the design of the robot. Our results suggested that children do not prefer highly realistic robots, but instead, more machine-like ones while they conceptualised robots whose task is multipurpose. The two robots we have employed for our studies, namely the Nao and the iCub robot are well-aligned with children's preferences.

We first evaluated the interaction between the social components of the robot and its task competency, by modulating the facial expressions and gaze model of the robot using both children and adult participants. Our results in adults showed a significant improvement in knowledge in all scenarios and no significant improvements in children, meaning that with the way our experiment was configured, we could not draw conclusions regarding the task competence of the robot. However, they highlight the role of the gaze model in plausibility, as the robot scored higher in the Godspeed questionnaire.

Our final tutoring scenario is the Piagetian Balance beam, in which the robot acts as a peer tutor, guiding students through the process of learning physics. We first validated the individual features of the non-anthropomorphic

content presentation tool and the results obtained were used in the main study where we explored the robot's task competence by varying the various types of feedback given. Results indicated that children did enjoy the interaction and found the provided feedback helpful, contributing to both the task and social competence of the robot.

To conclude, we have contributed to the domain of Human-Robot Interaction in a variety of ways. The most significant contribution of this thesis lies in the proposed taxonomy for the systematic evaluation of the psychological plausibility of the robot. To our knowledge, there are no similar approaches that offer systematic validation and provide guidelines for the development of social robots. A second contribution is the identification and validation of essential components for interaction like the expression of emotions, motivated behaviour as well as gaze. Similarly, we highlighted another important factor that affects the interaction: the individual differences between participants, something that has been neglected by many studies in the field. Nevertheless, knowing how to validate the robot's components is not sufficient to create a psychologically plausible robot. The production of convincing behaviours is equally important and results from the interplay between sensation and actuation. Consequently, the robot's control architecture plays a definitive role.

Another contribution is the introduction of the DAC biologically grounded layered architecture as the control system that guides the robot's behaviour. The studies presented and the H5WRobot prototype are the first attempts to extend DAC's implementation to embodied social agents for dyadic interactions. DAC's layered control structures facilitate the evaluation of our taxonomy. They allow for perception and action and include an emotional and motivational system, components that are crucial for the autonomous behaviour of the robot. Finally, we argued that the uncanny valley needs to be redefined, as its existing formalisation provides little empirical validation and inadequate information that can act as a guideline for future robotic applications. Alternatively, we offered our proposed taxonomy as a method to explore the uncanny valley.

We do not claim to have found the solution to every problem in the domain of HRI. We are also well aware that this is the beginning of the work and further studies are needed to systematically assess the role of motivation and emotion on the emergence of believable behaviour. Empiric validation of social components can provide valuable insights to the creation of a plausible agent and can help us refine our taxonomy and the robot's psychological engine. Nevertheless, we firmly believe that this work can set the ground for further investigation with the ultimate goal to create an agent that humans can accept. These social and psychological aspects create interesting challenges to the field of artificial intelligence, as robots that interact with humans will be required, among other things, to learn from their environment, understand complex social interactions and reason about human intentions and actions. Psychologists and social scientists may also benefit, as robots can become the testbed for psychological, emotional and communication models, aimed at helping us better understand ourselves, and in turn, improve the interaction.

The exploration of the social and cognitive components of robots contributes to their believability and at the same time, creates interesting ethical challenges. One of the eminent issues is safety, as robots will be required to operate in close proximity with their users. This raises the concern of privacy, both in terms of handling the data collected and user's sense Feil-Seifer et al. (2007). Additionally, the robot's behaviour should comply with law and ethics and in parallel, its social impacts need to be examined. Humans can treat robots as companions and this may have a crucial influence in their relationship, hence, research needs to explore this area for potential psychological harm from attachment Lin et al. (2011). These issues need to be addressed while the development of social robots advances. For this reason, the field of social robotics should be the outcome of close collaboration between psychologists, social scientists, neuroscientists and engineers; only then can we get closer to fully functional social and plausible machines.

Appendix A

Appendix

A.1 Semi-structured interviews

The purpose of the semi-structured interview was to get a better understanding of both the success of the interaction and what children thought of the robot. The questions asked were the following:

- How did it go?
- Did you enjoy the task?
- What did you like best?
- Was there also something that you didn't like?
- Can you tell me something about the robot?
- Did the robot help you? Why?
- Did you like the way the robot help you?
- If you go back to your classroom what would you tell other children about the robot?
- Could the robot see what you were doing? Why do you think so?

- Could the robot hear what you were saying? Why do you think so?
- Did you help the robot or did the robot help you?
- Do you think the robot is smart?
- Do you think the robot is friendly? Why?
- Do you think the robot finds you friendly?
- How old do you think the robot is? Gender? Role?
- Did the robot make any mistakes?
- Was the task easy or difficult?
- Did you know how to make the scale in balance?
- Did you learn something new about the task? What did you learn?

A.2 DAC architecture implementation

The setup and the architecture of H5W_Alpha has been extensively presented in Lallée et al. (2015).

Our setup employs a variety of sensors that are either internal to the robot or external (e.g. Kinect, Reactable). Given that each sensor has its own reference system, it is important to coordinate and calibrate the sensors with each other. Within the DAC architecture, we represent multiple reference frames by mapping them onto a single ego-centric frame of reference that the robot will use as its main standpoint. This approach facilitates sensor and robotic platform abstraction, and in the case of H5W_Alpha, allows the gaze behaviour to easily switch between agents in the environment (Kinect) and virtual objects (Reactable).

To deal with the "anchoring problem" Coradeschi and Saffiotti (2003) (link the incoming sensorimotor data to symbolic representations), we formalised the information exchange between modules around the H5W problem (How, When, Why, What, Where, Who). By exchanging information in the form of an *Entity* or *Relation*, we achieve a coherent view of the world by integrating all the sensorimotor and semantic representations.

The details of solving the issues of sensorimotor abstraction, knowledge representation, population and retrieval can be found in Lallée et al. (2015).

The developed modules for the H5W_Alpha have later been extended to the so called H5W_STA architecture and the main module that has been kept almost the same is the `Objects Properties Collector` module that acts as the working memory of the architecture and is the interface for the communication of information between the rest of the modules. Although the logic of the architecture remains the same, the modules have been extended in a way that allows for fast prototyping, as most of the variables that define a module's functionality can be changed through a configuration file. This way we ensure that no further compilation is needed to change key parameters of the system.

Each implemented model in the DAC H5W_STA architecture can be mapped to one or more of the core components of the DAC model that they embody. The developed models are schematically illustrated in Figure A.1 and have been described in Vouloutsi et al. (2016).

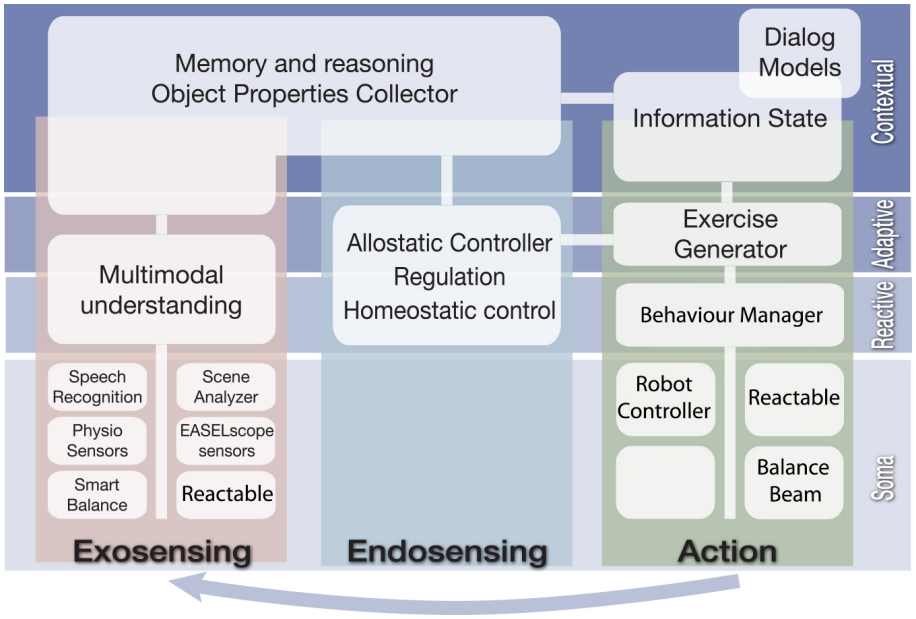


Figure A.1: Overview of the current implementation, where each module is mapped to the core components of the DAC architecture. Image adapted by Vouloutsi et al. (2016).

The integrated H5W_STA architecture is based on the EASEL european project and is a practical incarnation of the conceptual architecture of the Distributed Adaptive Control (DAC) theory of the design principles underlying perception, cognition and action. Each implemented module in the EASEL architecture can be mapped to one or more of the core components of the DAC model that they embody and the specific modules are schematically illustrated in Figure 5.1.

The Speech recognition (ASR module), the SceneAnalyzer, the PhysioReader and the EASELScope sensors embed the exosensing component of the Soma layer through which the states of the world are acquired and the in-

ternal drives are established. More precisely, the ASR module is based on the open-source Kaldi speech recognition toolkit Povey et al. (2011) with an EASEL specific vocabulary, language model and recognition grammar.

The SceneAnalyzer builds upon several other libraries to deliver integrated recognition of multimodal features of the users and their behaviour Zaraki et al. (2014, 2013). The physiological signal acquisition module uses non-obtrusive and robust methods for obtaining information about the user's physiological state: by integrating sensors in the robot or in the EASELscope tablet, information can be unobtrusively obtained without sensors worn or strapped to the body of user. The EASELScope sensors allow detection of the current state of the balance beam allowing the EASEL system to respond to the actions of the user with respect to the learning materials.

In the reactive layer, ASAPRealizer module Reidsma and van Welbergen (2013); Van Welbergen et al. (2014) is responsible for the choreography of the behaviour (verbal and non-verbal) of the STA using the generic robot-independent Behaviour Markup Language (BML). We also use an easily configurable XML binding between the BML and the motion primitives of each robotic platform (that serves as the physical instantiation of the STA). Such approach abstracts away from specific motor control by exposing more general behaviour specifications to the dialog manager and provides generalisation across embodiments. The ASAPRealizer maps to the behaviour component of the DAC architecture by directly controlling the actuators of the somatic layer.

The Allostatic Control (AC) module currently implemented in the EASEL architecture embraces both the Reactive and the Adaptive layers of DAC. An homeostatic controller continuously classifies the current state of each drive by sending fast requests for corrective actions to keep drives within optimal boundaries. The allostatic controller maintains consistency between drives in an adaptive way by assigning priorities to the different drives and making the appropriate corrections to maintain coherence (e.g., by adapting the difficulty of the task to the learner's behaviour). The learning algorithms

of the allostatic controller Vouloutsi et al. (2013a); Lallée et al. (2014) allow the STA to adapt its drives and homeostatic boundaries to a specific student's behaviour and skills. Successful interactions (i.e., contextual cues and actions) are then stored as memory segments in the Object Properties Collector (OPC) module to build a model of the user.

The Exercise Generator contains the collection of learning exercises with all their different properties and difficulty levels. It selects the appropriate exercise given the current state of the tutoring model, the student model, and the output of the AC. This information is shared with the Flipper Dialog Manager to allow the robotic assistant to discuss with the child the progress of the exercise.

The Object Properties Collector (OPC) embodies the Contextual Layer's memory components of DAC. At this stage of development, the OPC implements the memory for events that can be stored and distributed to the other STA's modules as well as its short-term memory component. At each instant of the interaction, the OPC can temporarily retain ongoing perceptions, actions and values (i.e., outcomes of the current interaction) as segments of memory (relations). These relations allow the definition of rules for specific interactions that can be further stored as long-term memories if a high-level goal is successfully achieved. Between the OPC and the sensors lies the multimodal understanding module, a light and simple interpreter of speech, emotions and speaker's probabilities, which simplifies the requirements for the Flipper Dialog Manager's scripts. Flipper offers flexible dialog specification via information state and rule-based templates to trigger information state changes as well as behaviour requests ter Maat and Heylen (2011).

The robots (Zeno (Hanson Robotics, Hong Kong) and FACE Lazzeri et al. (2013b)), the virtual robot avatars and the EASEL-scope hidden state visualiser correspond to the DAC effectors and represent the main interface of the STA with the world (Figure 7.17a). The EASELscope offers an augmented reality (AR) interface that allows the learner to interact with the

task materials. It can be used to present extra information to the child about the learning content. This allows the system to vary between different ways of scaffolding the learning of the user. For instance, using the EASELscope we can present "hidden information" about the balance beam, such as the weights of the pots, or the forces acting on the arms of the beam; both are types of scaffolds in learning that would not be possible without the scope (Figure 7.17a).

Each module within the framework has been integrated in a cohesive setup and the configuration options, models, behaviour repertoires and dialog scripts will allow us to validate the EASEL system through specific experiments with child-robot interaction in the proposed learning scenarios.

The way the architecture is organised gives us three key advantages: scalability, configurability and abstraction. This allows us to easily add sensory components with negligible changes to the main core of the system: it is sufficient to add the input to the multimodal understanding module that, in turn, will store the new information with an appropriate format in the OPC module. Furthermore, all modules are fully configurable: for instance, we can add new behaviours (ASAPRealizer), drives (Allostatic Control) as well as dialogues (Flipper) in an easy way with the usage of configuration files such as XML scripts. Thus, any additional implementation for the needs of the EASEL architecture (in terms of scenarios or sensory inputs) can be done in a flexible way. Finally, the proposed architecture permits abstraction from the physical manifestation of the STA, in a way that using the same scenario, we can choose the robotic platform (or even avatar) with small changes in the main core of the system.

At this stage, we are now ready to start validating our educational architecture and focus on concrete long-term studies on human-robot symbiotic interactions in learning tasks.

Bibliography

Each reference indicates the pages where it appears.

H. Aarts and A. Dijksterhuis. Habits as knowledge structures: automaticity in goal-directed behavior. *Journal of personality and social psychology*, 78(1):53–63, 2000. 187

L. Y. Abramson, M. E. Seligman, and J. D. Teasdale. Learned helplessness in humans: Critique and reformulation. *Journal of abnormal psychology*, 87(1):49–74, 1978. 141

H. Admoni and B. Scassellati. Social eye gaze in human-robot interaction: A review. *Journal of Human-Robot Interaction*, 6(1):25–63, 2017. 42

M. Alemi, A. Meghdari, A. Ghanbarzadeh, L. J. Moghadam, and A. Ghanbarzadeh. Impact of a social humanoid robot as a therapy assistant in children cancer treatment. In *International Conference on Social Robotics*, pages 11–22. Springer, 2014. 27

J. R. Anderson. *How can the human mind occur in the physical universe?* Oxford University Press, 2007. 78

J. R. Anderson, M. Matessa, and C. Lebiere. Act-r: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, 12(4):439–462, 1997. 78

A. Anning and K. Ring. *Making sense of children's drawings*. McGraw-Hill Education (UK), 2004. 145

M. A. Arbib and J.-M. Fellous. Emotions: from brain to robot. *Trends in cognitive sciences*, 8(12):554–561, 2004. 51

M. Argyle and J. Dean. Eye-contact, distance and affiliation. *Sociometry*, 28(3):289–304, 1965. 70

R. C. Arkin. *Behavior-based robotics*. MIT press, 1998. 19, 20

R. C. Arkin, M. Fujita, T. Takagi, and R. Hasegawa. An ethological and emotional basis for human–robot interaction. *Robotics and Autonomous Systems*, 42(3):191–201, 2003. 46, 100

R. Azevedo and A. F. Hadwin. Scaffolding self-regulated learning and metacognition–implications for the design of computer-based scaffolds. *Instructional Science*, 33(5):367–379, 2005. 163

S. Badia, A. Valjamae, F. Manzi, U. Bernardet, A. Mura, J. Manzolli, and P. Verschure. The effects of explicit and implicit interaction on user experiences in a mixed reality installation: The synthetic oracle. *Presence*, 18(4):277–285, 2009. 76

W. A. Bainbridge, J. Hart, E. S. Kim, and B. Scassellati. The effect of presence on human-robot interaction. In *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, pages 701–706. IEEE, 2008. 30

Y. Bar-Cohen and D. Hanson. *The coming robot revolution: Expectations and fears about emerging intelligent, humanlike machines*. Springer Science & Business Media, 2009. 23

S. Baron-Cohen. *Mindblindness: An essay on autism and theory of mind*. MIT press, 1997. 58

R. B. Barr and J. Tagg. From teaching to learning—a new paradigm for undergraduate education. *Change: The magazine of higher learning*, 27 (6):12–26, 1995. 139

I. B.-A. Bartal, J. Decety, and P. Mason. Empathy and pro-social behavior in rats. *Science*, 334(6061):1427–1430, 2011. 52

C. Bartneck and J. Forlizzi. A design-centred framework for social human-robot interaction. In *Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on*, pages 591–594. IEEE, 2004. 30

C. Bartneck, C. Rosalia, R. Menges, and I. Deckers. Robot abuse - a limitation of the media equation. In *Proceedings of the interact 2005 workshop on agent abuse, Rome*, 2005. 53, 133

C. Bartneck, M. Van Der Hoek, O. Mubin, and A. Al Mahmud. Daisy, daisy, give me your answer do!: switching off a robot. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 217–222. ACM, 2007a. 53, 134

C. Bartneck, M. Verbunt, O. Mubin, and A. Al Mahmud. To kill a mockingbird robot. In *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*, pages 81–87. IEEE, 2007b. 53, 60, 120, 134

C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1 (1):71–81, 2009. 107, 114, 128, 167

A. Batliner, M. Blomberg, S. D'Arcy, D. Elenius, D. Giuliani, M. Gerosa, C. Hacker, M. Russell, S. Steidl, and M. Wong. The pf_star children's speech corpus. In *Ninth European Conference on Speech Communication and Technology*, 2005. 148

D. S. Bauser, P. Thoma, and B. Suchan. Turn to me: electrophysiological correlates of frontal vs. averted view face and body processing are associated with trait empathy. *Frontiers in integrative neuroscience*, 6, 2012. 60

P. Baxter, T. Belpaeme, L. Canamero, P. Cosi, Y. Demiris, V. Enescu, A. Hiolle, I. Kruijff-Korbayova, R. Looije, M. Nalin, et al. Long-term human-robot interaction with young users. In *IEEE/ACM Human-Robot Interaction 2011 Conference (Robots with Children Workshop)*, 2011. 27

R. D. Beer, H. J. Chiel, and R. F. Drushel. Using autonomous robotics to teach science and engineering. *Communications of the ACM*, 42(6):

85–92, 1999. 137

G. Bekey and J. Yuh. The status of robotics. *IEEE Robotics & Automation Magazine*, 15(1):80–86, 2008. 23

E. Benítez Sandoval and C. Penaloza. Children's knowledge and expectations about robots: A survey for future user-centered design of social robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 107–108. ACM, 2012. 157

F. B. V. Benitti. Exploring the educational potential of robotics in schools: A systematic review. *Computers & Education*, 58(3):978–988, 2012. 28

M. Bennewitz, F. Faber, D. Joho, M. Schreiber, and S. Behnke. Towards a humanoid museum guide robot that interacts with multiple persons. In *Humanoid Robots, 2005 5th IEEE-RAS International Conference on*, pages 418–423. IEEE, 2005. 27

U. Bernardet and P. F. Verschure. iqr: A tool for the construction of multi-level simulations of brain and behaviour. *Neuroinformatics*, 8(2):113–134, 2010. 91

A. Betella and P. F. Verschure. The affective slider: a digital self-assessment scale for the measurement of human emotions. *PloS one*, 11(2):1–11, 2016. 101

A. Betella, M. Inderbitzin, U. Bernardet, and P. F. Verschure. Non-anthropomorphic expression of affective states through parametrized abstract motifs. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 435–441. IEEE, 2013. 63

M. Bisani and H. Ney. Joint-sequence models for grapheme-to-phoneme conversion. *Speech communication*, 50(5):434–451, 2008. 149

M. Blancas, V. Vouloutsi, K. Grechuta, and P. F. Verschure. Effects of the robot's role on human-robot interaction in an educational scenario. In *Conference on Biomimetic and Biohybrid Systems*, pages 391–402. Springer, 2015. 138, 159

M. Botvinick, A. P. Jha, L. M. Bylsma, S. A. Fabian, P. E. Solomon, and

K. M. Prkachin. Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *Neuroimage*, 25(1):312–319, 2005. 61

J.-D. Boucher, J. Ventre-Dominey, P. F. Dominey, S. Fagel, and G. Bailly. Facilitative effects of communicative gaze and speech in human-robot cooperation. In *Proceedings of the 3rd international workshop on Affective interaction in natural environments*, pages 71–74. ACM, 2010. 144

J.-D. Boucher, U. Pattacini, A. Lelong, G. Bailly, F. Elisei, S. Fagel, P. F. Dominey, and J. Ventre-Dominey. I reach faster when i see you look: gaze effects in human–human and human–robot face-to-face cooperation. *Frontiers in neurorobotics*, 6(3):1–11, 2012. 42, 62, 70

R. E. Boyatzis. *The competent manager: A model for effective performance*. John Wiley & Sons, 1982. 61

E. A. Boyle, A. H. Anderson, and A. Newlands. The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language and speech*, 37(1):1–20, 1994. 30

M. M. Bradley and P. J. Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994. 101

M. Brady. The advent of intelligent robots. In R. Langton Gregory and P. Marstrand, editors, *Creative Intelligences*, pages 88–110. Intellect Books, Norwood, New Jersey, 1987. 19

V. Braitenberg. *Vehicles: Experiments in synthetic psychology*. MIT press, 1986. 15, 17

C. Breazeal. Regulation and entrainment in human-robot interaction. *The International Journal of Robotics Research*, 21(10-11):883–902, 2002. 28

C. Breazeal. Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, 59(1):119–155, 2003a. 45, 70

C. Breazeal. Toward sociable robots. *Robotics and Autonomous Systems*, 42(3):167–175, 2003b. 26, 28, 41, 114

C. Breazeal and R. Brooks. Robot emotion: A functional perspective. *Who*

needs emotions? The brain meets the robot, pages 271–310, 2005. 45, 51

C. Breazeal, A. Brooks, J. Gray, G. Hoffman, C. Kidd, H. Lee, J. Lieberman, A. Lockerd, and D. Chilongo. Tutelage and collaboration for humanoid robots. *International Journal of Humanoid Robotics*, 1(02):315–348, 2004. 32

C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman, and M. Berlin. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 708–713. IEEE, 2005. 42, 71, 173

C. L. Breazeal. *Designing sociable robots*. MIT press, 2004. 28, 30, 31, 42

G. Briggs and M. Scheutz. Investigating the effects of robotic displays of protest and distress. *Social Robotics*, pages 238–247, 2012. 53, 60

R. A. Brooks. Intelligence without representation. *Artificial intelligence*, 47 (1-3):139–159, 1991a. 20

R. A. Brooks. Intelligence without reason. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, IJCAI'91, pages 569–595. Morgan Kaufmann, 1991b. 20

C. Brown. Gaze controls with interactions and decays. *IEEE Transactions on Systems, Man and Cybernetics,*, 20(2):518–527, 1990. 70

K. Bumby and K. Dautenhahn. Investigating children's attitudes towards robots: A case study. In *Proc. CT99, The Third International Cognitive Technology Conference*, pages 391–410, 1999. 156, 157

J. Burgner-Kahrs, D. C. Rucker, and H. Choset. Continuum robots for medical applications: A survey. *IEEE Transactions on Robotics*, 31(6): 1261–1280, 2015. 23

M. Cabanac. What is emotion? *Behavioural processes*, 60(2):69–83, 2002. 49

J.-J. Cabibihan, H. Javed, M. Ang, and S. M. Aljunied. 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, 2013.

27

S. Calinon, D. Bruno, M. S. Malekzadeh, T. Nanayakkara, and D. G. Caldwell. Human–robot skills transfer interfaces for a flexible surgical robot. *Computer methods and programs in biomedicine*, 116(2):81–96, 2014. 24

D. Cameron, S. Fernando, E. Collins, A. Millings, R. Moore, A. Sharkey, V. Evers, and T. Prescott. Presence of life-like robot expressions influences children's enjoyment of human-robot interactions in the field. In *Proceedings of the AISB Convention 2015*. The Society for the Study of Artificial Intelligence and Simulation of Behaviour, 2015a. 42

D. Cameron, S. Fernando, A. Millings, R. Moore, A. Sharkey, and T. Prescott. Children's age influences their perceptions of a humanoid robot as being like a person or machine. In *Conference on Biomimetic and Biohybrid Systems*, pages 348–353. Springer, 2015b. 36

D. Cameron, S. Fernando, A. Millings, M. Szollosy, E. Collins, R. Moore, A. Sharkey, and T. Prescott. Designing robot personalities for human-robot symbiotic interaction in an educational context. In *Conference on Biomimetic and Biohybrid Systems*, pages 413–417. Springer, 2016. 28, 42

D. Cañamero. Modeling motivations and emotions as a basis for intelligent behavior. In *Proceedings of the first international conference on Autonomous agents*, pages 148–155. ACM, 1997. 45, 76

L. Cañamero. Bridging the gap between hri and neuroscience in emotion research: robot as models, 2014. 51

L. Cañamero and P. Gaussier. Emotion understanding: robots as tools and models. *Emotional development: Recent research advances*, pages 235–258, 2005. 51

W. B. Cannon. The wisdom of the body. *The American Journal of the Medical Sciences*, 184(6):864, 1932. 44, 80

H.-L. Cao, P. G. Esteban, A. De Beir, R. Simut, G. Van de Perre, D. Lefeber, and B. Vanderborght. Robee: A homeostatic-based social behavior controller for robots in human-robot interaction experiments. In *Robotics and*

Biomimetics (ROBIO), 2014 IEEE International Conference on, pages 516–521. IEEE, 2014. 45

J. Cassell, T. Bickmore, H. Vilhjálmsson, and H. Yan. More than just a pretty face: affordances of embodiment. In *Proceedings of the 5th international conference on Intelligent user interfaces*, pages 52–59. ACM, 2000. 30

Á. Castro-González, M. Malfaz, and M. A. Salichs. An autonomous social robot in fear. *IEEE Transactions on Autonomous Mental Development*, 5(2):135–151, 2013. 45

D. W. Chambers. Stereotypic images of the scientist: The draw-a-scientist test. *Science education*, 67(2):255–265, 1983. 145

C.-W. Chang, J.-H. Lee, P.-Y. Chao, C.-Y. Wang, and G.-D. Chen. Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *Educational Technology & Society*, 13(2):13–24, 2010. 161

V. Charisi, D. Davison, F. Wijnen, J. Van Der Meij, D. Reidsma, T. Prescott, W. Van Joolingen, and V. Evers. Towards a child-robot symbiotic co-development: A theoretical approach. In *AISB Convention 2015*. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB), 2015. 139

V. Charisi, D. Davison, D. Reidsma, and V. Evers. Evaluation methods for user-centered child-robot interaction. In *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pages 545–550. IEEE, 2016. 181

C.-C. Cheng, K.-H. Huang, and S.-M. Huang. Exploring young children's images on robots. *Advances in Mechanical Engineering*, 9(4):1–7, 2017. 36

S. Chennu, V. Noreika, D. Gueorguiev, A. Blenkmann, S. Kochen, A. Ibáñez, A. M. Owen, and T. A. Bekinschtein. Expectation and attention in hierarchical auditory prediction. *Journal of Neuroscience*, 33 (27):11194–11205, 2013. 57

H. Christensen, J. Barker, N. Ma, and P. D. Green. The chime corpus: a resource and a challenge for computational hearing in multisource environments. In *Eleventh Annual Conference of the International Speech Communication Association*, 2010. 148

A. Clark. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204, 2013. 56

A. Clark. *Surfing uncertainty: Prediction, action, and the embodied mind.* Oxford University Press, 2015. 56

E. C. Collins, T. J. Prescott, and B. Mitchinson. Saying it with light: A pilot study of affective communication using the miro robot. In *Conference on Biomimetic and Biohybrid Systems*, pages 243–255. Springer, 2015. 32, 42, 63

G. Colombetti. From affect programs to dynamical discrete emotions. *Philosophical Psychology*, 22(4):407–425, 2009. 49

L. Cominelli, D. Mazzei, M. Pieroni, A. Zaraki, R. Garofalo, and D. De Rossi. Damasio's somatic marker for social robotics: Preliminary implementation and test. In *Conference on Biomimetic and Biohybrid Systems*, pages 316–328. Springer, 2015. 51

D. J. Cook and S. K. Das. How smart are our environments? an updated look at the state of the art. *Pervasive and mobile computing*, 3(2):53–73, 2007. 76

S. Coradeschi and A. Saffiotti. An introduction to the anchoring problem. *Robotics and Autonomous Systems*, 43(2):85–96, 2003. 215

S. Coradeschi, H. Ishiguro, M. Asada, S. C. Shapiro, M. Thielscher, C. Breazeal, M. J. Mataric, and H. Ishida. Human-inspired robots. *IEEE Intelligent Systems*, 21(4):74–85, 2006. 31

D. G. Cowan, E. J. Vanman, and M. Nielsen. Motivated empathy: The mechanics of the empathic gaze. *Cognition and Emotion*, 28(8):1522–1530, 2014. 60

A. D. Craig. How do you feel now? the anterior insula and human awareness.

Nature reviews neuroscience, 10(1):59–70, 2009. 49

H. Cramer, A. Amin, V. Evers, and N. Kemper. Touched by robots: effects of physical contact and proactiveness. *arXiv. org e-Print archive*, pages 1–11, 2009a. 83

H. Cramer, N. Kemper, A. Amin, B. Wielinga, and V. Evers. Give me a hug?: the effects of touch and autonomy on people's responses to embodied social agents. *Computer Animation and Virtual Worlds*, 20(2-3): 437–445, 2009b. 83

H. Cramer, J. Goddijn, B. Wielinga, and V. Evers. Effects of (in)accurate empathy and situational valence on attitudes towards robots. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 141–142, March 2010a. doi: 10.1109/HRI.2010.5453224. 53

H. Cramer, J. Goddijn, B. Wielinga, and V. Evers. Effects of (in) accurate empathy and situational valence on attitudes towards robots. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pages 141–142. IEEE, 2010b. 60

T. Dalgleish. The emotional brain. *Nature Reviews Neuroscience*, 5(7): 583–589, 2004. 51

F. D'Ambrosio, M. Olivier, D. Didon, and C. Besche. The basic empathy scale: A french validation of a measure of empathy in youth. *Personality and Individual Differences*, 46(2):160–165, 2009. 52, 129

K. Darling, P. Nandy, and C. Breazeal. Empathic concern and the effect of stories in human-robot interaction. In *Robot and Human Interactive Communication (RO-MAN), 2015 24th IEEE International Symposium on*, pages 770–775. IEEE, 2015. 53, 60

K. Dautenhahn. Socially intelligent robots: dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):679–704, 2007. 26, 28, 173

K. Dautenhahn and T. Christaller. *Remembering, rehearsal and empathy-towards a social and embodied cognitive psychology for artifacts*. GMD-Forschungszentrum Informationstechnik, 1995. 20

K. Dautenhahn and I. Werry. Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition*, 12(1): 1–35, 2004. 27

K. Dautenhahn, B. Ogden, and T. Quick. From embodied to socially embedded agents–implications for interaction-aware robots. *Cognitive Systems Research*, 3(3):397–428, 2002. 20

K. Dautenhahn, S. Woods, C. Kaouri, M. L. Walters, K. L. Koay, and I. Werry. What is a robot companion-friend, assistant or butler? In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 1192–1197. IEEE, 2005. 41

F. D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–340, 1989. 55

M. M. De Graaf and S. B. Allouch. Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems*, 61(12): 1476–1486, 2013. 56

F. de Vignemont and T. Singer. The empathic brain: how, when and why? *Trends in Cognitive Sciences*, 10(10):435 – 441, 2006. ISSN 1364-6613. doi: http://dx.doi.org/10.1016/j.tics.2006.08.008. URL http://www.sciencedirect.com/science/article/pii/S1364661306002154. 52

F. B. de Waal. Do animals feel empathy? *Scientific American Mind*, 18(6): 28–35, 2007. 52

F. B. de Waal. The antiquity of empathy. *Science*, 336(6083):874–876, 2012. 52

C. F. DiSalvo, F. Gemperle, J. Forlizzi, and S. Kiesler. All robots are not created equal: the design and perception of humanoid robot heads. In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 321–326. ACM, 2002. 35

C. Distante, G. Indiveri, and G. Reina. An application of mobile robotics for olfactory monitoring of hazardous industrial sites. *Industrial Robot: An International Journal*, 36(1):51–59, 2009. 23

A. Druin. Cooperative inquiry: developing new technologies for children with children. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 592–599. ACM, 1999. 145, 158

A. Duff and P. F. Verschure. Unifying perceptual and behavioral learning with a correlative subspace learning rule. *Neurocomputing*, 73(10):1818–1830, 2010. 75

B. R. Duffy. Anthropomorphism and the social robot. *Robotics and autonomous systems*, 42(3):177–190, 2003. 34, 35, 41

B. R. Duffy, C. Rooney, G. M. O'Hare, and R. O'Donoghue. What is a social robot? In *10th Irish Conference on Artificial Intelligence & Cognitive Science, University College Cork, Ireland, 1-3 September, 1999*, 1999. 30

M. Dunbabin, J. M. Roberts, K. Usher, and P. Corke. A new robot for environmental monitoring on the great barrier reef. In *Proceedings of the 2004 Australasian Conference on Robotics & Automation*. Australian Robotics & Automation Association, 2004. 23

A. H. Dyson. Appreciate the drawing and dictating of young children. *Young Children*, 43(3):25–32, 1988. 145

N. Eisenberg and P. Miller. The relation of empathy to prosocial and related behaviors. *Psychological Bulletin*, 101(1):91–119, 1987. 52

M. Ekman, P. Kok, and F. P. de Lange. Time-compressed preplay of anticipated events in human primary visual cortex. *Nature Communications*, 8, 2017. 57

P. Ekman. An argument for basic emotions. *Cognition & Emotion*, 6(3-4): 169–200, 1992. 49, 92

P. Ekman and W. V. Friesen. Facial action coding system. 1977. 103

K. Eng, D. Klein, A. Babler, U. Bernardet, M. Blanchard, M. Costa, T. Delbrück, R. J. Douglas, K. Hepp, J. Manzolli, et al. Design for a brain revisited: the neuromorphic design and functionality of the interactive space 'Ada'. *Reviews in the Neurosciences*, 14(1-2):145–180, 2003. 39, 63, 76, 80

K. Eng, R. J. Douglas, and P. F. Verschure. An interactive space that

learns to influence human behavior. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 35(1):66–77, 2005. 46, 63, 91, 145

H. G. Engen and T. Singer. Empathy circuits. *Current opinion in neurobiology*, 23(2):275–282, 2013. 52

M. L. Epstein, B. B. Epstein, and G. M. Brosvic. Immediate feedback during academic testing. *Psychological reports*, 88(3):889–894, 2001. 186

V. Estivill-Castro. Robotic activities that engage year 6 students into stem: Visual descriptions of behaviour. In *ICERI2016 Proceedings*, 9th annual International Conference of Education, Research and Innovation, pages 3299–3307. IATED, 14-16 November, 2016 2016. ISBN 978-84-617-5895-1. doi: 10.21125/iceri.2016.1771. URL http://dx.doi.org/10.21125/iceri.2016.1771. 137

T. A. Evans and M. J. Beran. Chimpanzees use self-distraction to cope with impulsivity. *Biology Letters*, 3(6):599–602, 2007. ISSN 1744-9561. doi: 10.1098/rsbl.2007.0399. URL http://rsbl.royalsocietypublishing.org/content/3/6/599. 187

H. J. Eysenck and W. Thompson. The effects of distraction on pursuit rotor learning, performance and reminiscence. *British Journal of Psychology*, 57(1-2):99–106, 1966. ISSN 2044-8295. doi: 10.1111/j.2044-8295.1966.tb01009.x. URL http://dx.doi.org/10.1111/j.2044-8295.1966.tb01009.x. 187

D. Feil-Seifer, K. Skinner, and M. J. Matarić. Benchmarks for evaluating socially assistive robotics. *Interaction Studies*, 8(3):423–439, 2007. 58, 211

R. Felder and R. Brent. Understanding student differences. *Journal of engineering education*, 94(1):57–72, 2005. 140

J.-M. Fellous. From human emotions to robot emotions. *Architectures for Modeling Emotion: Cross-Disciplinary Foundations, American Association for Artificial Intelligence*, pages 39–46, 2004. 47, 51

R. Ferguson. The tripod project framework. *The Tripod Project*, 2008. 167

Y. Fernaeus, M. Håkansson, M. Jacobsson, and S. Ljungblad. How do you play with a robotic toy animal?: a long-term study of pleo. In *Proceedings of the 9th international Conference on interaction Design and Children*, pages 39–48. ACM, 2010. 39

S. Fernando, R. K. Moore, D. Cameron, E. C. Collins, A. Millings, A. J. Sharkey, and T. J. Prescott. Automatic recognition of child speech for robotic applications in noisy environments. *arXiv preprint arXiv:1611.02695*, 2016. 28, 149

M. S. Fibla, U. Bernardet, and P. F. Verschure. Allostatic control for robot behaviour regulation: An extension to path planning. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 1935–1942. IEEE, 2010. 46, 77

K. A. Flannery and M. W. Watson. Sex differences and gender-role differences in children's drawings. *Studies in Art Education*, 36(2):114–122, 1995. 157

K. Foerde and D. Shohamy. Feedback timing modulates brain systems for learning in humans. *Journal of Neuroscience*, 31(37):13157–13167, 2011. 186

T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3):143–166, 2003. 26, 30, 31, 33, 35, 41, 138, 143, 150, 152

D. François, S. Powell, and K. Dautenhahn. A long-term study of children with autism playing with a robotic pet: Taking inspirations from non-directive play therapy to encourage children's proactivity and initiative-taking. *Interaction Studies*, 10(3):324–373, 2009. 27

W. Freeman. Wg walter: The living brain. In *Brain Theory*, pages 237–238. Springer, 1986. 15

N. H. Frijda. *The emotions*. Cambridge University Press, 1986. 47

N. H. Frijda. The laws of emotion. *American psychologist*, 43(5):349, 1988. 48

N. H. Frijda and B. Mesquita. The social roles and functions of emotions.

1994. 47

A. Frischen, A. P. Bayliss, and S. P. Tipper. Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin*, 133(4):694–724, 2007. 70

K. Friston. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138, 2010. 56

C. Frith. Role of facial expressions in social interactions. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 364(1535):3453–3458, dec 2009. 143

C. Frith and U. Frith. Theory of mind. *Current Biology*, 15(17):R644–R645, 2005. 58

C. D. Frith and U. Frith. Interacting minds–a biological basis. *Science*, 286 (5445):1692–1695, 1999. 58

D. M. Fryer and J. C. Marshall. The motives of jacques de vaucanson. *Technology and Culture*, 20(2):257–269, 1979. 12

M. Fujita. Aibo: Toward the era of digital creatures. *The International Journal of Robotics Research*, 20(10):781–794, 2001. 24

M. Fujita and K. Kageyama. An open architecture for robot entertainment. In *Proceedings of the first international conference on Autonomous agents*, pages 435–442. ACM, 1997. 24

H. L. Gallagher and C. D. Frith. Functional imaging of "theory of mind". *Trends in cognitive sciences*, 7(2):77–83, 2003. 58, 63

S. Gallagher. *How the body shapes the mind*. Cambridge Univ Press, 2005. 57

B. Gates. A robot in every home. *Scientific American*, 296(1):58–65, 2007. 156

G. Geiger, N. Alber, S. Jordà, and M. Alonso. The reactable: A collaborative musical instrument for playing and understanding music. *Her&Mus. Heritage & Museography*, 2(2):36–43, 2010. 79, 84

Y. Geng, D. Xia, and B. Qin. The basic empathy scale: A chinese validation of a measure of empathy in adolescents. *Child Psychiatry & Human*

Development, 43(4):499–510, 2012. 129

E. v. Glasersfeld. A constructivist approach to teaching. *Constructivism in education*, pages 3–15, 1995. 139

R. Gockley, J. Forlizzi, and R. Simmons. Interactions with a moody robot. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 186–193. ACM, 2006a. 27

R. Gockley, R. Simmons, and J. Forlizzi. Modeling affect in socially interactive robots. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 558–563. IEEE, 2006b. 27

J. Goetz, S. Kiesler, and A. Powers. Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on*, pages 55–60. Ieee, 2003. 33

E. Goffman. *Behavior in public places*. Simon and Schuster, 2008. 34

A. I. Goldman et al. Theory of mind. *The Oxford handbook of philosophy of cognitive science*, pages 402–424, 2012. 58

K. H. Goodell, C. G. Cao, and S. D. Schwaitzberg. Effects of cognitive distraction on performance of laparoscopic surgical tasks. *Journal of Laparoendoscopic & Advanced Surgical Techniques*, 16(2):94–98, 2006. 187

M. A. Goodrich and A. C. Schultz. Human-robot interaction: a survey. *Foundations and trends in human-computer interaction*, 1(3):203–275, 2007. 41

G. Gordon, S. Spaulding, J. K. Westlund, J. J. Lee, L. Plummer, M. Martinez, M. Das, and C. Breazeal. Affective personalization of a social robot tutor for children's second language skills. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 3951–3957. AAAI Press, 2016. 28, 138

K. Goris, J. Saldien, B. Vanderborght, and D. Lefeber. Probo, an intelligent huggable robot for hri studies with children. In *Human-Robot Interaction*. InTech, 2010. 32

K. Goris, J. Saldien, B. Vanderborght, and D. Lefeber. How to achieve the huggable behavior of the social robot probo? a reflection on the actuators. *Mechatronics*, 21(3):490–500, 2011. 32

M. S. Gou, V. Vouloutsi, K. Grechuta, S. Lallée, and P. F. Verschure. Empathy in humanoid robots. In *Biomimetic and Biohybrid Systems*, pages 423–426. Springer, 2014. 53, 122

D. Gouaillier, V. Hugel, P. Blazevic, C. Kilner, J. Monceaux, P. Lafourcade, B. Marnier, J. Serre, and B. Maisonnier. The nao humanoid: a combination of performance and affordability. *CoRR abs/0807.3223*, 2008. 31

B. Graf, U. Reiser, M. Hägele, K. Mauz, and P. Klein. Robotic home assistant care-o-bot® 3-product vision and innovation platform. In *Advanced Robotics and its Social Impacts (ARSO), 2009 IEEE Workshop on*, pages 139–144. IEEE, 2009. 28

S. Graham and B. Weiner. Theories and principles of motivation. *Handbook of educational psychology*, 4:63–84, 1996. 43

F. Grenier and A. Lüthi. Mouse brains wired for empathy? *Nature neuroscience*, 13(4):406–408, 2010. 52

A. Haber and C. Sammut. A cognitive architecture for autonomous robots. *Advances in Cognitive Systems*, 2:257–276, 2013. 18

L. Hall. Inflicting pain on synthetic characters: Moral concerns and empathic interaction. *Proceedings of the Joint Symposium on Virtual Social Agents*, pages 144–149, 2005. 134

J. Ham, R. Bokhorst, R. Cuijpers, D. van der Pol, and J.-J. Cabibihan. Making robots persuasive: the influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power. In *Social Robotics*, pages 71–83. Springer, 2011. 173

J.-H. Han, M.-H. Jo, V. Jones, and J.-H. Jo. Comparative study on the educational use of home robots for children. *Journal of Information Processing Systems*, 4(4):159–168, 2008. 138

D. Hanson. Exploring the aesthetic range for humanoid robots. In *Proceedings of the ICCS/CogSci-2006 long symposium: Toward social mech-*

anisms of android science, pages 39–42, 2006. 39

D. Hanson, S. Baurmann, T. Riccio, R. Margolin, T. Dockins, M. Tavares, and K. Carpenter. Zeno: A cognitive character. In *Ai magazine, and special proc. of AAAI national conference, Chicago*, pages 9–11, 2009. 31

N. A. Harrison, T. Singer, P. Rotshtein, R. J. Dolan, and H. D. Critchley. Pupillary contagion: central mechanisms engaged in sadness processing. *Social cognitive and affective neuroscience*, 1(1):5–17, 2006. 61

J. Hattie and H. Timperley. The power of feedback. *Review of educational research*, 77(1):81–112, 2007. 186

M. Heerink, B. Krose, V. Evers, B. Wielinga, et al. The influence of social presence on acceptance of a companion robot by older people. *Journal of Physical Agents*, 2(2):33–40, 2008. 56

F. Heider. Social perception and phenomenal causality. *Psychological review*, 51(6):358, 1944. 57

F. Heider and M. Simmel. An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2):243–259, 1944. 40, 57

A. Hennenlotter, U. Schroeder, P. Erhard, F. Castrop, B. Haslinger, D. Stoecker, K. W. Lange, and A. O. Ceballos-Baumann. A common neural basis for receptive and expressive communication of pleasant facial affect. *Neuroimage*, 26(2):581–591, 2005. 61

K. Highfield, J. Mulligan, and J. Hedberg. Early mathematics learning through exploration with programmable toys. In *Proceedings of the Joint Meeting of PME 32 and PME-NA*, pages 169–176. Citeseer, 2008. 138

P. J. Hinds, T. L. Roberts, and H. Jones. Whose job is it anyway? a study of human-robot interaction in a collaborative task. *Human-Computer Interaction*, 19(1):151–181, 2004. 35

M. L. Hoffman. *Empathy and moral development: Implications for caring and justice*. Cambridge University Press, 2001. 51

T. Horberry, J. Anderson, M. A. Regan, T. J. Triggs, and J. Brown. Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis & Pre-*

vention, 38(1):185–191, 2006. 187

J. P. Hourcade. Interaction design and children. *Foundations and Trends in Human-Computer Interaction*, 1(4):277–392, 2008. 159

W. Huitt. Motivation to learn: An overview. *Educational psychology interactive*, 12, 2001. 43

C. Hull. *Principles of behavior*. Appleton-century-crofts, 1943. 44, 80

D. Hume. *The natural history of religion*. Stanford University Press, 1957. 33

IFR. World robotics: Service robots 2016, 2016. 24, 25, 143

M. Iijima, O. Arisaka, F. Minamoto, and Y. Arai. Sex differences in children's free drawings: a study on girls with congenital adrenal hyperplasia. *Hormones and Behavior*, 40(2):99–104, 2001. 157

M. P. Inderbitzin, A. Betella, A. Lanatá, E. P. Scilingo, U. Bernardet, and P. F. Verschure. The social perceptual salience effect. *Journal of experimental psychology: human perception and performance*, 39(1):62–74, 2013. 144

B. Inhelder and J. Piaget. *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures*. Basic Books, 1958. 174

iRobot. Annual report, 2015. 25

S. Jeong, D. E. Logan, M. S. Goodwin, S. Graca, B. O'Connell, H. Goodenough, L. Anderson, N. Stenquist, K. Fitzpatrick, M. Zisook, et al. A social robot to mitigate stress, anxiety, and pain in hospital pediatric care. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, pages 103–104. ACM, 2015. 27

W. Johal, G. Calvary, and S. Pesty. Non-verbal signals in hri: Interference in human perception. In *International Conference on Social Robotics*, pages 275–284. Springer, 2015. 34

D. Jolliffe and D. P. Farrington. Development and validation of the basic empathy scale. *Journal of adolescence*, 29(4):589–611, 2006. 128, 167

M. Joosse and V. Evers. A guide robot at the airport: First impressions. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 149–150. ACM, 2017. 27

M. Joosse, M. Lohse, J. G. Pérez, and V. Evers. What you do is who you are: The role of task context in perceived social robot personality. In *Robotics and automation (ICRA), 2013 IEEE international conference on*, pages 2134–2139. IEEE, 2013. 62

S. Jordà. On stage: the reactable and other musical tangibles go real. *International Journal of Arts and Technology*, 1(3):268–287, 2008. 84

C. F. Julià, D. Gallardo, and S. Jordà. Mtcf: A framework for designing and coding musical tabletop applications directly in pure data. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, volume 20011, pages 457–460, 2011. 85

H. K. Kabali, M. M. Irigoyen, R. Nunez-Davis, J. G. Budacki, S. H. Mohanty, K. P. Leister, and R. L. Bonner. Exposure and use of mobile media devices by young children. *Pediatrics*, 136(6):1044–1050, 2015. 137

M. Kabátová and J. Pekárová. Lessons learnt with lego mindstorms: from beginner to teaching robotics. *Robotics in education*, 10:51–56, 2010. 27

P. H. Kahn, H. Ishiguro, B. Friedman, and T. Kanda. What is a human?-toward psychological benchmarks in the field of human-robot interaction. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 364–371. IEEE, 2006. 58

P. H. Kahn, N. G. Freier, T. Kanda, H. Ishiguro, J. H. Ruckert, R. L. Severson, and S. K. Kane. Design patterns for sociality in human-robot interaction. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 97–104. ACM, 2008. 58

P. H. Kahn, B. T. Gill, A. L. Reichert, T. Kanda, H. Ishiguro, and J. H. Ruckert. Validating interaction patterns in hri. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pages 183–184. IEEE, 2010. 58

P. H. Kahn Jr, J. H. Ruckert, T. Kanda, H. Ishiguro, A. Reichert, H. Gary, and S. Shen. Psychological intimacy with robots?: using interaction patterns to uncover depth of relation. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, pages 123–124. IEEE Press, 2010. 58

P. H. Kahn Jr, T. Kanda, H. Ishiguro, N. G. Freier, R. L. Severson, B. T. Gill, J. H. Ruckert, and S. Shen. "robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental psychology*, 48(2):303–314, 2012. 53

T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro. Interactive robots as social partners and peer tutors for children: A field trial. *Human-computer interaction*, 19(1):61–84, 2004a. 28, 137, 138, 143, 144

T. Kanda, H. Ishiguro, M. Imai, and T. Ono. Development and evaluation of interactive humanoid robots. *Proceedings of the IEEE*, 92(11):1839–1850, 2004b. 41

T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro. Friendly social robot that understands human's friendly relationships. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 3, pages 2215–2222. IEEE, 2004c. 28

T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro. A two-month field trial in an elementary school for long-term human–robot interaction. *Robotics, IEEE Transactions on*, 23(5):962–971, 2007. 28, 138, 161

T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita. A communication robot in a shopping mall. *Robotics, IEEE Transactions on*, 26 (5):897–913, 2010. 27

J. Kätsyri, K. Förger, M. Mäkäräinen, and T. Takala. A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in psychology*, 6:1–16, 2015. 38

D. Keltner and P. Ekman. Emotion: an overview. *Encyclopedia of psychology*, 3:162–167, 2000. 143

D. Keltner and J. J. Gross. Functional accounts of emotions. *Cognition & Emotion*, 13(5):467–480, 1999. 47

A. Kendon and M. Cook. The consistency of gaze patterns in social interaction. *British Journal of Psychology*, 60(4):481–494, 1969. 133

J. Kennedy, P. Baxter, and T. Belpaeme. Can less be more? the impact of robot social behaviour on human learning. In *Proceedings of the 4th International Symposium on New Frontiers in HRI at AISB*, 2015a. 42

J. Kennedy, P. Baxter, and T. Belpaeme. The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 67–74. ACM, 2015b. 28, 138

J. Kennedy, P. Baxter, E. Senft, and T. Belpaeme. Social robot tutoring for child second language learning. In *Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on*, pages 231–238. IEEE, 2016. 62

J. Kennedy, P. Baxter, and Belpaeme. The impact of robot tutor nonverbal social behavior on child learning. *Frontiers in ICT*, 4:1–16, 2017. 42, 62

C. D. Kidd and C. Breazeal. Effect of a robot on user perceptions. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 4, pages 3559–3564. IEEE, 2004. 30

C. D. Kidd and C. Breazeal. Robots at home: Understanding long-term human-robot interaction. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pages 3230–3235. IEEE, 2008. 27, 41, 42

C. D. Kidd, W. Taggart, and S. Turkle. A sociable robot to encourage social interaction among the elderly. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 3972–3976. IEEE, 2006. 27

E. H. Kim, S. S. Kwak, and Y. K. Kwak. Can robotic emotional expressions induce a human to empathize with a robot? In *RO-MAN 2009 - The*

18th IEEE International Symposium on Robot and Human Interactive Communication, pages 358–362, Sept 2009a. doi: 10.1109/ROMAN.2009.5326282. 53

E. H. Kim, S. S. Kwak, and Y. K. Kwak. Can robotic emotional expressions induce a human to empathize with a robot? In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pages 358–362. IEEE, 2009b. 60

R. Kirby, J. Forlizzi, and R. Simmons. Affective social robots. *Robotics and Autonomous Systems*, 58(3):322–332, 2010. 27

H. Klinck, K. Stelzer, K. Jafarmadar, and D. K. Mellinger. Aas endurance: An autonomous acoustic sailboat for marine mammal research. In *Proceedings of the International Robotic Sailing Conference (IRSC)*, 2009. 23

M. Knapp, J. Hall, and T. Horgan. *Nonverbal communication in human interaction*. Cengage Learning, 2013. 34, 70

T. Koda. Agents with faces: A study on the effects of personification of software agents. master's thesis, 1996. 120

S. Krach, F. Hegel, B. Wrede, G. Sagerer, F. Binkofski, and T. Kircher. Can machines think? interaction and perspective taking with robots investigated via fmri. *PloS one*, 3(7):1–11, 2008. 62

J. Kruger and D. Dunning. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, 77(6):1121–1134, 1999. 140

J. A. Kulik and C.-L. C. Kulik. Timing of feedback and verbal learning. *Review of educational research*, 58(1):79–97, 1988. 186

I.-H. Kuo, C. Jayawardena, E. Broadbent, R. Stafford, and B. MacDonald. Hri evaluation of a healthcare service robot. *Social robotics*, pages 178–187, 2012. 58

S. S. Kwak, Y. Kim, E. Kim, C. Shin, and K. Cho. What makes people empathize with an emotional robot?: The impact of agency and physical embodiment on human empathy for a robot. In *RO-MAN, 2013 IEEE*,

pages 180–185. IEEE, 2013. 53

J. Laird. *The Soar cognitive architecture*. MIT Press, 2012. 78

S. Lallee, C. Madden, M. Hoen, and P. F. Dominey. Linking language with embodied and teleological representations of action for humanoid cognition. *Frontiers in neurorobotics*, 4:1–12, 2010. 76

S. Lallée, U. Pattacini, J.-D. Boucher, S. Lemaignan, A. Lenz, C. Melhuish, L. Natale, S. Skachek, K. Hamann, J. Steinwender, et al. Towards a platform-independent cooperative human-robot interaction system: Ii. perception, execution and imitation of goal directed actions. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 2895–2902. IEEE, 2011. 76

S. Lallée, K. Hamann, J. Steinwender, F. Warneken, U. Martienz, H. Barron-Gonzales, U. Pattacini, I. Gori, M. Petit, G. Metta, et al. Cooperative human robot interaction systems: Iv. communication of shared plans with naïve humans using gaze and speech. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 129–136. IEEE, 2013. 42, 70

S. Lallée, V. Vouloutsi, S. Wierenga, U. Pattacini, and P. Verschure. Efaa: a companion emerges from integrating a layered cognitive architecture. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 105–105. ACM, 2014. 46, 78, 113, 218

S. Lallée, V. Vouloutsi, M. Blancas Munoz, K. Grechuta, J.-Y. Puigbo Llobet, M. Sarda, and P. F. Verschure. Towards the synthetic self: making others perceive me as an other. *Journal of Behavioral Robotics*, 6(1): 136–164, 7 2015. 60, 78, 79, 138, 144, 215

P. J. Lang, M. M. Bradley, and B. N. Cuthbert. International affective picture system (iaps): Technical manual and affective ratings. *Gainesville, FL: The Center for Research in Psychophysiology, University of Florida*, 2, 1999. 103

N. Lazzeri, D. Mazzei, A. Zaraki, and D. De Rossi. Towards a believable social robot. In *Conference on Biomimetic and Biohybrid Systems*, pages 393–395. Springer, 2013a. 31

N. Lazzeri, D. Mazzei, A. Zaraki, and D. De Rossi. Towards a believable social robot. In *Conference on Biomimetic and Biohybrid Systems*, pages 393–395. Springer, 2013b. 218

J. LeDoux. Rethinking the emotional brain. *Neuron*, 73(4):653–676, 2012. 47, 49, 50

J. E. LeDoux. Emotion circuits in the brain. *Annual review of neuroscience*, 23(1):155–184, 2000. 49

I. Leite, C. Martinho, A. Pereira, and A. Paiva. icat: an affective game buddy based on anticipatory mechanisms. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 3*, pages 1229–1232. International Foundation for Autonomous Agents and Multiagent Systems, 2008. 138

I. Leite, G. Castellano, and S. Mascarenhas. Modelling Empathy in Social Robotic Companions. *Proceedings of the 19th International Conference on Advances in User Modeling*, pages 135–147, 2012. URL http://gaips.inesc-id.pt/gaips/component/gaips/publications/showPublicationPdf?pid=201{&}format=raw. 52

I. Leite, C. Martinho, and A. Paiva. Social robots for long-term interaction: a survey. *International Journal of Social Robotics*, 5(2):291–308, 2013. 41

A. M. Leslie. Pretense and representation: The origins of" theory of mind.". *Psychological review*, 94(4):412, 1987. 63

R. W. Levenson. The intrapersonal functions of emotion. *Cognition & Emotion*, 13(5):481–504, 1999. 47

D. Lewis and J. Greene. *Your child's drawings... their hidden meaning*. Hutchinson, 1983. 145

M. Lewis and L. Cañamero. An affective autonomous robot toddler to support the development of self-efficacy in diabetic children. In *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*, pages 359–364. IEEE, 2014. 27

D. Leyzberg, S. Spaulding, M. Toneva, and B. Scassellati. The physical presence of a robot tutor increases cognitive learning gains. *Proceedings*

of the Annual Meeting of the Cognitive Science Society, 34, 2012. 138

D. Leyzberg, S. Spaulding, and B. Scassellati. Personalizing robot tutors to individuals' learning differences. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, HRI '14, pages 423–430, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2658-2. doi: 10.1145/2559636.2559671. 138

P. Lin, K. Abney, and G. A. Bekey. *Robot ethics: the ethical and social implications of robotics*. MIT press, 2011. 211

D. Lindsley. Emotion. In S. S. Stevens, editor, *Handbook of experimental psychology*, pages 473–516. Wiley, New York, 1951. 43

C. E. Looser and T. Wheatley. The tipping point of animacy how, when, and where we perceive life in a face. *Psychological science*, 21(12):1854–1862, 2010. 35, 63

G. Maffei, D. Santos-Pata, E. Marcos, M. Sánchez-Fibla, and P. F. Verschure. An embodied biologically constrained model of foraging: from classical and operant conditioning to adaptive real-world behavior in dac-x. *Neural Networks*, 72:88–108, 2015. 72, 77

E. Marcos, P. Pani, E. Brunamonti, G. Deco, S. Ferraina, and P. Verschure. Neural variability in premotor cortex is modulated by trial history and predicts behavioral performance. *Neuron*, 78(2):249–255, 2013. 141

A. H. Maslow. A theory of human motivation. *Published in*, 1943. 44, 80

M. B. Mathur and D. B. Reichling. Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition*, 146:22–32, 2016. 35, 37

R. Matthews, N. J. McDonald, P. Hervieux, P. J. Turner, and M. A. Steindorf. A wearable physiological sensor suite for unobtrusive monitoring of physiological and cognitive state. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 5276–5281. IEEE, 2007. 76

D. Mazzei, L. Cominelli, N. Lazzeri, A. Zaraki, and D. De Rossi. I-clips brain : A hybrid cognitive system for social robots. In *Conference on*

Biomimetic and Biohybrid Systems, pages 213–224. Springer, 2014. 31

A. Melonio and R. Gennari. Co-design with children: the state of the art. Technical report, Technical report, KRDB Research Centre Technical Report, 2012. 145

A. N. Meltzoff, R. Brooks, A. P. Shon, and R. P. Rao. "social" robots are psychological agents for infants: A test of gaze following. *Neural Networks*, 23(8):966–972, 2010. 173

G. Metta, G. Sandini, D. Vernon, L. Natale, and F. Nori. The icub humanoid robot: an open platform for research in embodied cognition. In *Proceedings of the 8th workshop on performance metrics for intelligent systems*, pages 50–56. ACM, 2008. 83

G. Metta, L. Natale, F. Nori, G. Sandini, D. Vernon, L. Fadiga, C. Von Hofsten, K. Rosander, M. Lopes, J. Santos-Victor, et al. The¡ i¿ icub¡/i¿ humanoid robot: An open-systems platform for research in cognitive development. *Neural Networks*, 23(8):1125–1134, 2010. 31, 79

A. Michotte. *The perception of causality.* Basic Books, 1963. 57

S. Milgram. Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, 67(4):371–378, 1963. 122, 123, 132

S. Milgram. Some conditions of obedience and disobedience to authority. *Human relations*, 18(1):57–76, 1965. 87, 123

S. Milgram and E. Van den Haag. Obedience to authority, 1978. 87, 122, 123

C. Misselhorn. Empathy with inanimate objects and the uncanny valley. *Minds and Machines*, 19(3):345–359, 2009. 35, 52

B. Mitchinson and T. J. Prescott. Miro: A robot "mammal" with a biomimetic brain-based control system. In *Conference on Biomimetic and Biohybrid Systems*, pages 179–191. Springer, 2016. 32

F. Mondada, M. Bonani, X. Raemy, J. Pugh, C. Cianci, A. Klaptocz, J.-c. Zufferey, D. Floreano, and A. Martinoli. The e-puck , a robot designed for education in engineering. *Robotics*, 1:59–65, 2006. 137

R. K. Moore. A bayesian explanation of the "uncanny valley" effect and

related psychological phenomena. *Scientific reports*, 2:864–869, 2012. 39, 205

M. E. Moran. The da vinci robot. *Journal of endourology*, 20(12):986–990, 2006. 11

M. E. Moran. Jacques de vaucanson: The father of simulation. *Journal of endourology*, 21(7):679–683, 2007. 12

M. Mori, K. F. MacDorman, and N. Kageki. The uncanny valley from the field. *IEEE Robotics & Automation Magazine*, 19(2):98–100, 2012. 37, 38

W. Moyle, C. Jones, B. Sung, M. Bramble, S. O'Dwyer, M. Blumenstein, and V. Estivill-Castro. What effect does an animal robot called cuddler have on the engagement and emotional response of older people with dementia? a pilot feasibility study. *International Journal of Social Robotics*, 8(1):145–156, 2016. 62

O. Mubin, C. J. Stevens, S. Shahid, A. Al Mahmud, and J.-J. Dong. A review of the applicability of robots in education. *Journal of Technology in Education and Learning*, 1:209–215, 2013. 137, 143

H. G. Mullet, A. C. Butler, B. Verdin, R. von Borries, and E. J. Marsh. Delaying feedback promotes transfer of knowledge despite student preferences to receive feedback immediately. *Journal of Applied Research in Memory and Cognition*, 3(3):222 – 229, 2014. ISSN 2211-3681. doi: http://dx.doi.org/10.1016/j.jarmac.2014.05.001. URL http://www.sciencedirect.com/science/article/pii/S2211368114000448. Cognition and Education. 186

B. Mutlu and J. Forlizzi. Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction. In *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on*, pages 287–294. IEEE, 2008. 27

B. Mutlu, J. Forlizzi, and J. Hodgins. A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *Humanoid Robots, 2006 6th IEEE-RAS International Conference on*, pages 518–523. IEEE, 2006. 70

C. Nass, Y. Moon, B. Fogg, B. Reeves, and C. Dryer. Can computer per-

sonalities be human personalities? In *Conference companion on Human factors in computing systems*, pages 228–229. ACM, 1995. 41

N. J. Nilsson. Shakey the robot. Technical report, DTIC Document, 1984. 17

S. Nishio, H. Ishiguro, and N. Hagita. *Geminoid: Teleoperated android of an existing person*. INTECH Open Access Publisher Vienna, 2007. 31

T. Nomura. Robots and gender. *Gender and the Genome*, 1(1):18–25, 2017. 36

I. Nourbakhsh, C. Kunz, and T. Willeke. The mobot museum robot installations: A five year experiment. In *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volume 4, pages 3636–3641. IEEE, 2003. 27

I. R. Nourbakhsh. *Robot futures*. MIT Press, 2013. 19

J.-H. Oh, D. Hanson, W.-S. Kim, Y. Han, J.-Y. Kim, and I.-W. Park. Design of android type humanoid robot albert hubo. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 1428–1433. IEEE, 2006. 31

T. Ono, M. Imai, and R. Nakatsu. Reading a robot's mind: a model of utterance understanding based on the theory of mind mechanism. *Advanced Robotics*, 14(4):311–326, 2000. 71

T. Ono, M. Imai, and H. Ishiguro. A model of embodied communications with gestures between human and robots. In *Proceedings of the Cognitive Science Society*, volume 23, 2001. 144

B. Opitz, N. K. Ferdinand, and A. Mecklinger. Timing matters: the impact of immediate and delayed feedback on artificial language learning. *Frontiers in human neuroscience*, 5(8):1–9, 2011. 186

J. Otterbacher and M. Talias. S/he's too warm/agentic! the influence of gender on uncanny reactions to robots. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 214–223. ACM, 2017. 36

J. Panksepp. Empathy and the laws of affect. *Science*, 334(6061):1358–1359,

2011. 52

J. Panksepp and L. Biven. The archaeology of mind, 2011. 49

S. Papert. *Mindstorms: Children, computers, and powerful ideas*. Basic Books, Inc., 1980. 163

S. Papert and I. Harel. Situating constructionism. *Constructionism*, 36(2): 1–11, 1991. 139

D. Parisi and G. Petrosino. Robots that have emotions. *Adaptive Behavior*, 18(6):453–469, 2010. 51

B. Parkinson. Emotions are social. *British journal of psychology*, 87(4): 663–683, 1996. 47

M. A. Peters and D. Araya. Transforming american education: Learning powered by technology. *E–Learning and Digital Media*, 114(2):102–105, 2011. 137

R. Pfeifer and C. Scheier. *Understanding intelligence*. MIT press, 2001. 20

F. G. Phelps, G. Doherty-Sneddon, and H. Warnock. Helping children think: Gaze aversion and teaching. *British Journal of Developmental Psychology*, 24(3):577–588, 2006. 161

J. Piaget. *The moral judgement of the child*. Simon and Schuster, 1997. 64

J. Piaget and M. T. Cook. *The origins of intelligence in children*. WW Norton & Co, 1952. 141

J. Piaget and B. Inhelder. *The psychology of the child*. Basic Books, 1950. 139

G. Pointeau, M. Petit, and P. F. Dominey. Successive developmental levels of autobiographical memory for learning through social interaction. *IEEE Transactions on Autonomous Mental Development*, 6(3):200–212, 2014. 100

M. E. Pollack, L. Brown, D. Colbry, C. Orosz, B. Peintner, S. Ramakrishnan, S. Engberg, J. T. Matthews, J. Dunbar-Jacob, C. E. McCarthy, et al. Pearl: A mobile robotic assistant for the elderly. In *AAAI workshop on automation as eldercare*, volume 2002, pages 85–91, 2002. 27

D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel,

M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, et al. The kaldi speech recognition toolkit. In *IEEE 2011 workshop on automatic speech recognition and understanding*, number EPFL-CONF-192584. IEEE Signal Processing Society, 2011. 217

A. Powers, S. Kiesler, S. Fussell, and C. Torrey. Comparing a computer agent with a humanoid robot. In *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*, pages 145–152. IEEE, 2007. 30

R. S. Prawat. Constructivisms, modern and postmodern. *Educational psychologist*, 31(3-4):215–225, 1996. 139

D. Premack and A. J. Premack. Origins of human social competence. In M. S. Gazzaniga, editor, *The cognitive neurosciences*, pages 205–218. The MIT Press, 1995. 57

D. Premack and G. Woodruff. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, 1(04):515–526, 1978. 58, 63

T. J. Prescott, T. Epton, V. Evers, K. McKee, M. Hawley, T. Webb, D. Benyon, S. Conran, and R. Strand. Robot companions for citizens: Roadmapping the potential for future robots in empowering older people. *Bridging Research into Ageing and ICT Development. Prague, Czech Republic: Institute for Informatics and Digital Innovation*, 2012. 28

T. J. Prescott, N. Lepora, and P. F. Vershure. A future of living machines? international trends and prospects in biomimetic and biohybrid systems. In *Proc. SPIE 9055, Bioinspiration, Biomimetics, and Bioreplication*, pages 905502–905502, 2014. 72, 78

A. Ramachandran and B. Scassellati. Adapting difficulty levels in personalized robot-child tutoring interactions. In *Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014. 138

J. C. Read and S. MacFarlane. Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In *Proceedings of the 2006 conference on Interaction design and children*, pages 81–88. ACM, 2006. 181

B. Reeves and C. Nass. How people treat computers, television, and new media like real people and places. *CSLI Publications and Cambridge*, 1996. 41, 42, 70

D. Reidsma and H. van Welbergen. AsapRealizer in practice–a modular and extensible architecture for a BML Realizer. *Entertainment computing*, 4 (3):157–169, 2013. 217

D. Reidsma, V. Charisi, D. Davison, F. Wijnen, J. van der Meij, V. Evers, D. Cameron, S. Fernando, R. Moore, T. Prescott, et al. The easel project: Towards educational human-robot symbiotic interaction. In *Conference on Biomimetic and Biohybrid Systems*, pages 297–306. Springer, 2016. 28

C. Rennó-Costa, J. E. Lisman, and P. F. Verschure. A signature of attractor dynamics in the ca3 region of the hippocampus. *PLoS computational biology*, 10(5):1–15, 2014. 141

L. D. Riek, T.-C. Rabinowitch, B. Chakrabarti, and P. Robinson. How anthropomorphism affects empathy toward robots. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 245–246. ACM, 2009. 35, 53

L. D. Riek, T.-C. Rabinowitch, P. Bremner, A. G. Pipe, M. Fraser, and P. Robinson. Cooperative gestures: Effective signaling for humanoid robots. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pages 61–68. IEEE, 2010. 35

J. Riskin. The defecating duck, or, the ambiguous origins of artificial life. *Critical Inquiry*, 29(4):599–633, 2003a. 11, 12, 13, 14

J. Riskin. Eighteenth-century wetware. *Representations*, 83(1):97–125, 2003b. 13, 14

B. Robins, K. Dautenhahn, R. Te Boerkhorst, and A. Billard. Robots as assistive technology-does appearance matter? In *Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on*, pages 277–282. IEEE, 2004. 35

B. Robins, K. Dautenhahn, R. Te Boekhorst, and A. Billard. Robotic assistants in therapy and education of children with autism: Can a small

humanoid robot help encourage social interaction skills? *Universal Access in the Information Society*, 4(2):105–120, 2005. 27

B. Robins, K. Dautenhahn, and J. Dubowski. Does appearance matter in the interaction of children with autism with a humanoid robot? *Interaction studies*, 7(3):509–542, 2006. 36

T. Robinson, J. Fransen, D. Pye, J. Foote, and S. Renals. Wsjcamo: a british english speech corpus for large vocabulary continuous speech recognition. In *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, volume 1, pages 81–84. IEEE, 1995. 148

E. T. Rolls. On the brain and emotion. *Behavioral and brain sciences*, 23 (02):219–228, 2000. 47, 49

C. Rosalia, R. Menges, I. Deckers, and C. Bartneck. Cruelty towards robots. In *Robot Workshop-Designing Robot Applications for Everyday Use, Göteborg*, 2005. 53, 133

A. M. Rosenthal-von der Pütten, N. C. Krämer, L. Hoffmann, S. Sobieraj, and S. C. Eimler. An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics*, 5(1):17–34, 2013a. 134

A. M. Rosenthal-von der Pütten, F. P. Schulte, S. C. Eimler, L. Hoffmann, S. Sobieraj, S. Maderwald, N. C. Krämer, and M. Brand. Neural correlates of empathy towards robots. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 215–216. IEEE Press, 2013b. 52

A. M. Rosenthal-Von Der Pütten, F. P. Schulte, S. C. Eimler, S. Sobieraj, L. Hoffmann, S. Maderwald, M. Brand, and N. C. Krämer. Investigations on empathy towards humans and robots using fmri. *Computers in Human Behavior*, 33:201–212, 2014. 60

M. E. Rosheim. In the footsteps of leonardo˜ articulated anthropomorphic robot\. *IEEE Robotics & Automation Magazine*, 4(2):12–14, 1997. 11

K. Rymarczyk, Ł. Żurawski, K. Jankowiak-Siuda, and I. Szatkowska. Emotional empathy and facial mimicry for static and dynamic facial expres-

sions of fear and disgust. *Frontiers in psychology*, 7, 2016. 60

A. M. Sabelli, T. Kanda, and N. Hagita. A conversational robot in an elderly care center: an ethnographic study. In *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on*, pages 37–44. IEEE, 2011. 27, 41

M. Saerbeck, T. Schut, C. Bartneck, and M. D. Janse. Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1613–1622. ACM, 2010. 41, 137, 138, 143, 144

Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and K. Fujimura. The intelligent asimo: System overview and integration. In *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on*, volume 3, pages 2478–2483. IEEE, 2002. 31

M. Salem, A. Weiss, P. Baxter, and K. Dautenhahn, editors. *Robots guiding small groups: the effect of appearance change on the user experience*, Canterbury, UK, 2015. University of Kent. 27

M. Sanchez-Fibla, U. Bernardet, E. Wasserman, T. Pelc, M. Mintz, J. C. Jackson, C. Lansink, C. Pennartz, and P. F. Verschure. Allostatic control for robot behavior regulation: a comparative rodent-robot study. *Advances in Complex Systems*, 13(03):377–403, 2010. 46, 74, 77

M. A. Sánchez-Montañés, P. Konig, and P. F. Verschure. Learning sensory maps with real-world stimuli in real time using a biophysically realistic learning rule. *IEEE Transactions on Neural Networks*, 13(3):619–632, 2002. 50

B. Scassellati. Using social robots to study abnormal social development. In *In Proceedings of the Fifth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, pages 11–14, 2005. 41

B. Scassellati, H. Admoni, and M. Matarić. Robots for use in autism research. *Annual review of biomedical engineering*, 14:275–294, 2012. 27

K. R. Scherer. Neuroscience projections to current debates in emotion psy-

chology. *Cognition & Emotion*, 7(1):1–41, 1993. 49

B. J. Scholl. Objects and attention: The state of the art. *Cognition*, 80(1): 1–46, 2001. 40, 57

M. Schulte-Rüther, H. J. Markowitsch, G. R. Fink, and M. Piefke. Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: a functional magnetic resonance imaging approach to empathy. *Journal of cognitive neuroscience*, 19(8):1354–1372, 2007. 60

M. E. Seligman. Learned helplessness. *Annual review of medicine*, 23(1): 407–412, 1972. 141

J. P. Seward. Drive, incentive, and reinforcement. *Psychological review*, 63 (3):195, 1956. 80

J. Seyama and R. S. Nagayama. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and virtual environments*, 16(4):337–351, 2007. 37

T. Shibata, T. Mitsui, K. Wada, A. Touda, T. Kumasaka, K. Tagami, and K. Tanie. Mental commit robot and its application to therapy of children. In *Advanced Intelligent Mechatronics, 2001. Proceedings. 2001 IEEE/ASME International Conference on*, volume 2, pages 1053–1058. IEEE, 2001. 32

D.-H. Shin and H. Choo. Modeling the acceptance of socially interactive robotics: Social presence in human–robot interaction. *Interaction Studies*, 12(3):430–460, 2011. 55

N. Shin and S. Kim. Learning about, from, and with robots: Students' perspectives. In *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, pages 1040–1045. IEEE, 2007. 138

M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita. Interactive humanoid robots for a science museum. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 305–312. ACM, 2006. 27

M. Shiomi, D. Sakamoto, T. Kanda, C. T. Ishi, H. Ishiguro, and N. Hagita.

A semi-autonomous communication robot—a field trial at a train station. In *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on*, pages 303–310. IEEE, 2008. 27

D. Shohamy, C. Myers, S. Grossman, J. Sage, M. Gluck, and R. Poldrack. Cortico-striatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. *Brain*, 127(4):851–859, 2004. 186

B. Siciliano and O. Khatib. *Springer handbook of robotics*. Springer, 2016. 24

R. S. Siegler. Three aspects of cognitive development. *Cognitive psychology*, 8(4):481–520, 1976. 175, 176

R. S. Siegler, S. Strauss, and I. Levin. Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development*, pages 1–84, 1981. 175

T. Skybo, N. Ryan-Wenger, and Y.-h. Su. Human figure drawings as a measure of children's emotional status: Critical review for practice. *Journal of pediatric nursing*, 22(1):15–28, 2007. 145, 152

M. Slater, A. Antley, A. Davison, D. Swapp, C. Guger, C. Barker, N. Pistrang, and M. V. Sanchez-Vives. A virtual reprise of the Stanley Milgram obedience experiments. *PLoS ONE*, 1(1):135–147, 2006. ISSN 19326203. doi: 10.1371/journal.pone.0000039. 53

J.-H. Song and P. Bédard. Paradoxical benefits of dual-task contexts for visuomotor memory. *Psychological science*, 26(2):148–158, 2015. 187

A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich. Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 33–40. ACM, 2006. 58

P. Sterling and J. Eyer. Allostasis: a new paradigm to explain arousal pathology. In S. Fisher and J. Reason, editors, *Handbook of life stress, cognition and health*, pages 629–649. John Wiley & Sons, 1988. 45

W. D. Stiehl, J. Lieberman, C. Breazeal, L. Basel, L. Lalla, and M. Wolf.

Design of a therapeutic robotic companion for relational, affective touch. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 408–415. IEEE, 2005. 27

A. Stoytchev and R. C. Arkin. Incorporating motivation in a hybrid robot architecture. *Journal of Advanced Computational Intelligence Vol*, 8(3), 2004. 45, 46

Y. Sun and S. S. Sundar. Psychological importance of human agency: How self-assembly affects user experience of robots. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, pages 189–196. IEEE Press, 2016. 58

J. Sung, H. I. Christensen, and R. E. Grinter. Robots in the wild: understanding long-term use. In *Human-Robot Interaction (HRI), 2009 4th ACM/IEEE International Conference on*, pages 45–52. IEEE, 2009. 25

J.-Y. Sung, L. Guo, R. E. Grinter, and H. I. Christensen. "my roomba is rambo": intimate home appliances. In *International Conference on Ubiquitous Computing*, pages 145–162. Springer, 2007. 39

Y. Suzuki, L. Galli, A. Ikeda, S. Itakura, and M. Kitazaki. Measuring empathy for human and robot hand pain using electroencephalography. *Scientific reports*, 5:1–9, 2015. 60

K. Swan, M. v. Hooft, A. Kratcoski, and D. Unger. Uses and effects of mobile computing devices in k–8 classrooms. *Journal of Research on Technology in Education*, 38(1):99–112, 2005. 137

K. Taber. Constructivism as educational theory: Contingency in learning, and optimally guided instruction. *Educational theory*, pages 39–61, 2011. 139

T. Tamura, S. Yonemitsu, A. Itoh, D. Oikawa, A. Kawakami, Y. Higashi, T. Fujimooto, and K. Nakajima. Is an entertainment robot useful in the care of elderly people with severe dementia? *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 59(1):M83–M85, 2004. 27

F. Tanaka, A. Cicourel, and J. R. Movellan. Socialization between toddlers

and robots at an early childhood education center. *Proceedings of the National Academy of Sciences of the United States of America*, 104(46): 17954–17958, 2007. ISSN 0027-8424. doi: 10.1073/pnas.0707769104. 137

A. Tapus, C. Ţăpuş, and M. J. Matarić. User-robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intelligent Service Robotics*, 1(2):169–183, 2008. 27

A. Tapus, C. Tapus, and M. J. Mataric. The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia. In *Rehabilitation Robotics, 2009. ICORR 2009. IEEE International Conference on*, pages 924–929. IEEE, 2009. 27

M. Tenorth, A. C. Perzylo, R. Lafrenz, and M. Beetz. The roboearth language: Representing and exchanging knowledge about actions, objects, and environments. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 1284–1289. IEEE, 2012. 76

C. L. Teo, E. Burdet, and H. Lim. A robotic teacher of chinese handwriting. In *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2002. HAPTICS 2002. Proceedings. 10th Symposium on*, pages 335–341. IEEE, 2002. 28

M. ter Maat and D. Heylen. Flipper: An information state component for spoken dialogue systems. In *Intelligent Virtual Agents*, pages 470–472. Springer, 2011. 218

F. Thomas, O. Johnston, and F. Thomas. *The illusion of life: Disney animation*. Hyperion New York, 1995. 39

S. Thrun. Toward a framework for human-robot interaction. *Human–Computer Interaction*, 19(1-2):9–24, 2004. 24, 41

S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A. B. Cremers, F. Dellaert, D. Fox, D. Haehnel, C. Rosenberg, N. Roy, et al. Probabilistic algorithms and the interactive museum tour-guide robot minerva. *The International Journal of Robotics Research*, 19(11):972–999, 2000. 26

R. Triebel, K. Arras, R. Alami, L. Beyer, S. Breuers, R. Chatila, M. Chetouani, D. Cremers, V. Evers, M. Fiore, et al. Spencer: A socially

aware service robot for passenger guidance and help in busy airports. In *Field and Service Robotics*, pages 607–622. Springer, 2016. 27

M. Trincavelli, M. Reggente, S. Coradeschi, A. Loutfi, H. Ishida, and A. J. Lilienthal. Towards environmental monitoring with mobile robots. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pages 2210–2215. IEEE, 2008. 23

D. B. D. K. M. G. K. O. D. W. M. S. G. G. S. T. J. J. M. G. M. N. M. S. P. N. E. Tucker Balch, Jay Summet and A. Gavin. Designing personal robots for education: Hardware, software, and curriculum. 2008. 138

S. G. Tzafestas. *Sociorobot World: A Guided Tour for All*, volume 1048. Springer, 2015. 31, 32

A. van Breemen, X. Yan, and B. Meerbeek. icat: an animated user-interface robot with personality. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 143–144. ACM, 2005. 32

H. Van Welbergen, R. Yaghoubzadeh, and S. Kopp. AsapRealizer 2.0: the next steps in fluent behavior realization for ECAs. In *Intelligent Virtual Agents*, pages 449–462. Springer, 2014. 217

V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis. User acceptance of information technology: Toward a unified view. *MIS quarterly*, pages 425–478, 2003. 55

P. Verschure. Connectionist explanation: Taking positions in the mind-brain dilemma. *Neural networks and a new artificial intelligence*, pages 133–188, 1997. 77

P. F. Verschure. Real-world behavior as a constraint on the cognitive architecture: Comparing act-r and dac in the newell test. *Behavioral and Brain Sciences*, 26(05):624–626, 2003. 78

P. F. Verschure. Distributed adaptive control: A theory of the mind, brain, body nexus. *Biologically Inspired Cognitive Architectures*, 1:55–72, 2012. 4, 43, 47, 49, 51, 56, 65, 71, 72, 73, 78, 91, 112, 140

P. F. Verschure. Formal minds and biological brains ii: from the mirage

of intelligence to a science and engineering of consciousness. *IEEE Intelligent Systems, Trends & Controversies, "The Convergence of Machine and Biological Intelligence"*, pages 7–10, 2014. 72, 78

P. F. Verschure. Synthetic consciousness: the distributed adaptive control perspective. *Phil. Trans. R. Soc. B*, 371(1701):20150448, 2016. 56

P. F. Verschure and R. Pfeifer. Environment interaction: a case study in autonomous systems. In *From Animals to Animats 2: Proceedings of the Second International Conference on Simulation of Adaptive Behavior*, volume 2, pages 210–217. MIT Press, 1993. 75

P. F. Verschure, B. J. Kröse, and R. Pfeifer. Distributed adaptive control: The self-organization of structured behavior. *Robotics and Autonomous Systems*, 9(3):181–196, 1992. 45, 112

P. F. Verschure, T. Voegtlin, and R. J. Douglas. Environmentally mediated synergy between perception and behaviour in mobile robots. *Nature*, 425 (6958):620–624, 2003. 4, 65, 72, 140

P. F. Verschure, C. M. Pennartz, and G. Pezzulo. The why, what, where, when and how of goal-directed choice: neuronal and computational principles. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655):1–14, 2014. 72

V. Vouloutsi, S. Lallée, and P. F. Verschure. Modulating behaviors using allostatic control. In *Biomimetic and Biohybrid Systems*, pages 287–298. Springer, 2013a. 45, 46, 78, 110, 113, 218

V. Vouloutsi, L. L. LopezSerrano, Z. Mathews, A. E. Chimeno, A. Ziyatdinov, A. P. i Lluna, S. B. i Badia, and P. F. J. Verschure. The synthetic moth: A neuromorphic approach toward artificial olfaction in robots. In K. Persaud, S. Marco, and G.-G. A, editors, *Neuromorphic Olfaction*. CRC Press/Taylor & Francis, Boca Raton (FL), 2013b. 23

V. Vouloutsi, K. Grechuta, S. Lallée, and P. F. Verschure. The influence of behavioral complexity on robot perception. In *Biomimetic and Biohybrid Systems*, pages 332–343. Springer, 2014. 42, 116

V. Vouloutsi, M. B. Munoz, K. Grechuta, S. Lallee, A. Duff, J.-Y. P. Llobet,

and P. F. Verschure. A new biomimetic approach towards educational robotics: the distributed adaptive control of a synthetic tutor assistant. *New Frontiers in Human-Robot Interaction*, page 22, 2015. 138

V. Vouloutsi, M. Blancas, R. Zucca, P. Omedas, D. Reidsma, D. Davison, V. Charisi, F. Wijnen, J. van der Meij, V. Evers, et al. Towards a synthetic tutor assistant: the easel project and its architecture. In *Conference on Biomimetic and Biohybrid Systems*, pages 353–364. Springer, 2016. 78, 79, 216

L. S. Vygotsky. *Mind in society: The development of higher psychological processes*. Harvard university press, 1980. 139, 141

K. Wada and T. Shibata. Living with seal robots in a care house-evaluations of social and physiological influences. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 4940–4945. IEEE, 2006a. 41, 42

K. Wada and T. Shibata. Robot therapy in a care house-results of case studies. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 581–586. IEEE, 2006b. 27

K. Wada and T. Shibata. Living with seal robots?its sociopsychological and physiological influences on the elderly at a care house. *Robotics, IEEE Transactions on*, 23(5):972–980, 2007a. 27

K. Wada and T. Shibata. Social effects of robot therapy in a care house-change of social network of the residents for two months. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 1250–1255. IEEE, 2007b. 27

B. J. Wadsworth. *Piaget's theory of cognitive and affective development: Foundations of constructivism* . Longman Publishing, 1996. 141

D. D. Wagner, W. M. Kelley, and T. F. Heatherton. Individual differences in the spontaneous recruitment of brain regions supporting mental state understanding when viewing natural social scenes. *Cerebral Cortex*, 21 (12):2788–2796, 2011. 34, 62

J. Wainer, D. J. Feil-Seifer, D. A. Shell, and M. J. Mataric. The role of physical embodiment in human-robot interaction. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 117–122. IEEE, 2006. 30

W. G. Walter. An imitation of life. *Scientific American*, 182(5):42–45, 1950. 15

W. G. Walter. A machine that learns. *Scientific American*, 185(2):60–63, 1951. 15

M. L. Walters, D. S. Syrdal, K. Dautenhahn, R. Te Boekhorst, and K. L. Koay. Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Autonomous Robots*, 24(2):159–178, 2008. 35, 37

M. L. Walters, K. L. Koay, D. S. Syrdal, K. Dautenhahn, and R. Te Boekhorst. Preferences and perceptions of robot appearance and embodiment in human-robot interaction trials. *Procs of New Frontiers in Human-Robot Interaction*, pages 136–143, 2009. 35

K. C. Wassermann, K. Eng, P. F. Verschure, and J. Manzolli. Live soundscape composition based on synthetic emotions. *IEEE MultiMedia*, 10 (4):82–90, 2003. 46

J. C. Weeks and L. Hasher. The disruptive–and beneficial–effects of distraction on older adults' cognitive performance. *Frontiers in psychology*, 5:1–6, 2014. 187

F. Wijnen, V. Charisi, D. P. Davison, J. Meij, D. Reidsma, and V. Evers. Inquiry learning with a social robot: can you explain that to me? 2015. 137

T. Wisspeintner, T. Van Der Zan, L. Iocchi, and S. Schiffer. Robocup@ home: Results in benchmarking domestic service robots. In *Robot Soccer World Cup*, pages 390–401. Springer, 2009. 58

S. Woods. Exploring the design space of robots: Children's perspectives. *Interacting with Computers*, 18(6):1390–1418, 2006. 39, 157

Y.-H. Wu, C. Fassert, and A.-S. Rigaud. Designing robots for the elderly:

Appearance issue and beyond. *Archives of Gerontology and Geriatrics*, 54(1):121–126, 2012. 35, 36

P. R. Wurman, R. D'Andrea, and M. Mountz. Coordinating hundreds of cooperative, autonomous vehicles in warehouses. *AI magazine*, 29(1): 9–20, 2008. 23

C. Zaga, M. Lohse, K. P. Truong, and V. Evers. The effect of a robot's social character on children's task engagement: Peer versus tutor. In *International Conference on Social Robotics*, pages 704–713. Springer, 2015. 138

A. Zaraki, D. Mazzei, N. Lazzeri, M. Pieroni, and D. De Rossi. Preliminary implementation of context-aware attention system for humanoid robots. In *Biomimetic and Biohybrid Systems*, pages 457–459. Springer, 2013. 217

A. Zaraki, D. Mazzei, M. Giuliani, and D. De Rossi. Designing and evaluating a social gaze-control system for a humanoid robot. *IEEE Transactions on Human-Machine Systems*, 44(2):157–168, 2014. 147, 149, 217

J. Złotowski, H. Sumioka, S. Nishio, D. F. Glas, C. Bartneck, and H. Ishiguro. Appearance of a robot affects the impact of its behaviour on perceived trustworthiness and empathy. *Paladyn, Journal of Behavioral Robotics*, 7(1):55–66, 2016. 35