Towards Robots for Social Engagement

Elisa Cucco, Michael Fisher, Louise Dennis, Clare Dixon, Matt Webster, Bastian Broecker, Richard Williams, Joe Collenette, Katie Atkinson, Karl Tuyls

Department of Computer Science, University of Liverpool, UK

Abstract

In this paper we consider the aspects that ensure successful interaction between social robots and people. As such robots are increasingly autonomous, it is crucial that the user can trust their behaviour, and that their decisions are taken within social and ethical requirements. It is important to specify what actions are expected from the robot, verify that the autonomous robot actually achieve these, and validate that the requirements are exactly what the user wants. To this purpose, our activities have been focused on formal verification of autonomous robotics systems, investigating both reliability and robot ethics and deployment of social robots in both constrained and public environments.

1 Introduction

Social robots are designed to interact with people in a natural, interpersonal manner, often to achieve positive outcomes across applications such as education, health, quality of life, entertainment, communication, and tasks requiring collaborative teamwork. The long-term goal of creating social robots that are competent and capable partners for people is quite challenging. They will need to be able to communicate naturally with people using both verbal and non verbal signals, in order to engage them not only on a cognitive level, but on an emotional level as well, to provide effective social and task-based support to the users. For this reason their main characteristic is a range of social-cognitive skills to understand human behaviour, and to be intuitively understood by people.

Considering their increasing involvement in social-care and education applications, there is also a growing research emphasis in cognitive Human Robot Interaction on identifying the mental models people use to make sense of emerging robotic technologies and investigating people's reactions to the appearance and behaviours of robots.

As those robots are becoming increasingly autonomous and they are directly interacting with humans it is vital that the user can be assured that those robots are safe, reliable and ethical in order to trust them. Thus, a big concern is not only that ethical and reliable behaviours are met, but also that they can be verified [Dennis *et al.*, 2016a; Charisi *et al.*, 2017].

In this paper we focus on the issues, such as safety, cognitive interaction, and trustworthiness, related to the increasingly common situation in which humans and autonomous robots share an environment. We give an overview of our activities related to this problem and in particular report on a practical human-robot engagement in which we have been involved.

2 Social Robots

People are more engaged while interacting with robots that are able to communicate naturally and have some social skills, but it is crucial that they also feel safe.

2.1 Human Robot Interaction

Recent advances in physical human-robot interaction have shown the potential and feasibility of robot systems for active and safe workspace sharing and collaboration with humans. This trend has been supported by recent progress in both robotic hardware and software technology that allow a safer human-robot interaction. Thus, by considering the physical contact of the human and the robot in the design phase, possible injuries due to unintentional contacts can be considerably mitigated.

These robot systems include applications such as coworkers (i.e., cooperative material-handling), but also service robots and assistive devices for physically challenged people. Therefore all of them share the common requirement of safe and close physical interaction between human and robot.

While encompassing safety issues based on biomechanical human injury analysis as well as of human movements, human-friendly hardware design and control strategies, learning and cognitive key components have to be developed, in order to enable the robot to predict human motions in real time in an unstructured dynamic environment. Apart from developing the capabilities for interactive autonomy, human safety and physical interaction have to be embedded at the cognitive decisional level as well; thus the robot will be enabled to react or physically interact with humans in a safe and autonomous way. Furthermore, self-explaining interaction and communication frameworks need to be developed to enhance the system usability and interpretability for humans, for example, to communicate whether a situation is safe or dangerous not only with verbal, but also non-verbal communication cues, such as gestures and emotional feedbacks. The key distinctive aspect of human-robot interaction is then the intrinsic dual aspect of physical and cognitive interaction.

Physical Human-Robot Interaction. Most work in pHRI (physical Human-Robot Interaction) can be classified across three main categories of interaction: *supportive, collaborative* and *cooperative*. The distinction is marked by the increasing frequency and necessity of physical contact with the robot and level of proximity of the user [Siciliano and Khatib, 2007]. Supportive interactions occur when the robot is not the main performer of the task, but instead provides the human with tools and information to optimize the human's task performance or objectives, for example museum tour guide robots, shopping assistant robot and home-care robots. In this context pHRI typically concerns safety, that is preventing and mitigating the effect of unexpected collisions and performing appropriate proxemic behaviour. To support safety as well as the physical interactions, well-structured robot communication is needed. In collaborative interactions both the human and the robot work on the same task, each separately completing the part of the task best suited to their abilities. In this scenario, the human completes a task requiring human dexterity, while the robot completes the part of the task not well suited to direct human involvement, i.e., repetitive tasks, high force applications, chemical deposition or precision placement. Finally cooperative interactions refer to the extension of cooperative manipulation to include force interaction with humans. The human and the robot work in direct physical contact, or indirect contact through a common object, with cooperative and continuous shared control of the task.

The main solution to make robots physically safer is to pursue a mechanical design that reduces the robot link inertia and weight by using lightweight and highly integrated mechatronics designs. Low inertia and high compliance have become the most desirable features(i.e., the DLR LWR-III [Hirzinger *et al.*, 2001]. However, very compliant transmissions may ensure safe interaction but may be inefficient in transferring energy from actuators to the links for their fast motion. Thus, other approaches to gain performance for guaranteed safety are the intrinsecally elastic robots (VIA- Variable Impedance Actuator method [Tonietti *et al.*, 2005] allows the passive compliance of transmission to vary during the execution of tasks, and the SEA-Series Elastic Actuator method [Pratt and Williamson, 1995] consists in locating the largest actuator at the base of the robot and connecting it through a spring, thus achieving low overall impedance, while small motors collocated at the joints provides high-performance motion).

Haptic sensors are capable of measuring contact and detecting collision, while they are also able to read and display emotion sensed by physical interaction, and can improve also the involvement of the human. Indeed, in human development, touch plays a crucial role in developing cognitive, social and emotional skills, as well as establishing and maintaining attachment and social relationships. Recently, more and more social robots are being equipped with tactile skin, thus allowing the robot to react according to the person touching the robot, or recognize social and affective communicative intent in how a human touch the robot.

Cognitive Human-Robot Interaction. A key challenge in robotics is to design robotic systems with the cognitive capabilities necessary to support human-robot interaction. These systems will need to have appropriate representation of the world, the capabilities, expectations and actions of the human and how their own actions might affect the world, their task, and their human partners. Core research activities in this area include the development of representations and actions that allow robots to participate in joint activities with people, a deeper understanding of human expectations and cognitive responses to robot actions and models of joint activity for human-robot interaction [Siciliano and Khatib, 2007].

More specifically research activities in this area include:

- human models of interaction — building an understanding of how people perceive robots and interpret their actions and behaviours, and how these perceptions and interpretations change across contexts and user groups;
- robot models of interaction — the development of models that enable robots to map aspects of the interaction into the physical world and develop cognitive capabilities through interaction with the social and physical environment; and
- models of HRI — creating models and mechanism that guide human-robot communication and collaboration, action planning, and model learning.

Research in cognitive human-robot interaction examines how people, including children and older adults, react to their interactions with social robots. Some approach robots in a scientific-explanatory mode, interpreting a robot's action in an emotionally detached and mechanistic manner, others invest in the interactions emotionally and treat the robots as they were living beings, such as babies or pets [Turkle *et al.*, 2004]. Anthropomorphism, or the attribution of human characteristics to non-human behaviour is an other interesting aspect in HRI research. In [Kiesler *et al.*, 2008] it is shown that people anthropomorphize a physically embodied robot more readily that an on-screen agent, and people behave in a more engaged and socially appropriate manner while interacting with the co-present robot. People also anthropomorphize robots they interact with directly more than they do with robots in general, and with robots that follow social conventions (e.g., polite robots) more than those that do not [Fussell *et al.*, 2008]. Moreover users with low emotional stability prefer mechanical-looking robots to human-like ones [Syrdal *et al.*, 2007]. As might be expected a robot's human-like appearance can have a positive effect on people's propensity towards it but also a too high level of human-likeness may place the robot in an *uncanny valley* [Mori, 1970], which refers to a dip in a hypothetical graph of the relationship between a robot's human-likeness and the human's response, suggesting that a robot that looks like a human, coupled with some remaining non-human qualities, makes users uncomfortable.

2.2 Social Robots Interaction

The way a person interacts with a social robot is quite different from interacting with an autonomous robot. Modern autonomous robots are viewed as tools that humans use to perform hazardous tasks in remote environments. However, social robots are designed to engage people in an interpersonal manner in order to achieve positive outcomes in domains such as education, therapy, or health, or task-related work in areas such as coordinated teamwork for manufacturing, search and rescue, domestic chores and more. The development of socially intelligent and socially skilful robots drives research to develop autonomous robots that are natural and intuitive for the user to interact with, communicate with, collaborate with, and teach new capabilities. Dautenhahn's work is among the most consistent concerned with thinking about robots with interpersonal social intelligence where relationships between specific individuals are important [Dautenhahn, 1995; 1997].

Social robots are designed to interact with people in human-centric terms and to operate in human environments alongside people. Their main characteristic is that they engage people, communicating and coordinating their behaviour with humans through verbal, non verbal or affective modalities. Anthropomorphic design principles, spanning from the physical appearance of robots, to how they move and behave, and how they interact with people, are often employed to facilitate interaction and acceptance. For social robots to close the communication loop and coordinate their behaviour with humans, they must also be able to perceive, interpret, and respond appropriately to verbal and non verbal cues from humans.

Depending on different application scenarios, increasing social skills are needed: robots that need to collaborate with humans simply to achieve, or help in a task, do not need to be particular social. On the other hand, robots that serve as companions in the home for the elderly or assist people with disabilities need to possess more social skills, which will make them more acceptable for humans. Without these skills, such robots might not be used and thus fail in their role as an assistant [Dautenhahn, 2007].

To participate in emotion-based interaction, robots must be able to recognise and interpret affective signals from humans, they must possess their internal models of emotions and they must be able to communicate this affective state to others. In particular, social robots need the ability to recognize, understand and predict human behaviour in terms of the underlying mental states such as beliefs, intents, desires, feelings, etc. For instance social robots will need to be aware of people's goals and intentions so that they can appropriately adjust their behaviour to help the human. Furthermore, the behaviour of social robots will need to adhere to people's expectations. They will also need to be able to flexibly draw their attention to what the user is interested in, so that their behaviour and information can be more useful [Siciliano and Khatib, 2007].

Social robots will need to be deeply aware of the user's emotions, feelings and attitudes to be able to prioritize what is the most important thing to do. In general, emotional displays can inform the interpretations about an individual's internal states (agreement or disagreement about a belief, valuing a particular outcome) and therefore help to predict future actions. An increasing number of socio-emotional robots have been designed to realize such functions to facilitate human-robot interactions. Some of these robots have been designed with emotional responses or emotional inspired decision making systems in order to entertain, i.e., AIBO [Fujita, 2004] or Pepper robots. In this way robots handle better human emotional states, and also motivate people toward more effective interactions, which is particular useful in domains such as education, or therapeutic system.

3 Human-Robot Engagement

For autonomous systems and social robots to be allowed to share their environment with people, they need to be safe and have to behaves within ethical acceptable limits. One vital aspect to human-robot interaction is *trust*. Indeed, no one will use a robot, or even share the environment with it, if they cannot trust its behaviour. In addition, since autonomous robots need to make decisions, it is crucial to have some ethical principles the robot will use to make such decisions, especially when they concern human safety.

3.1 Trust

For the users of a social robot one of the main concerns is that the robot they are interacting with is safe and behaves ethically. *Trust* is the key issue and in order to trust the AI system, the user needs to be informed of all the robot's capabilities. The appearance of trustworthiness might also be an issue, in particular in assisted living technologies. Some concern have been raised related to the impact that such robots can have on elderly [Sharkey and Sharkey, 2012] or children [Matthias, 2011].

Trust also plays a role in choosing an ethical theory to implement in the autonomous robot, even if they are very different. Indeed, trust is a social construct concerned with how the behaviour of the robot appears to the human.

For this reason trustworthiness is considered mainly subjective: a lot of items can change the user's level of trust of a robot, and among them the relationship between trust and harm [Salem *et al.*, 2015]. The concept of trust also involves the robot's reliability and predictability. However, while machine's errors could have an impact on the trust [Salem *et al.*, 2015], also errors occasionally performed by a humanoid robot can increases its perceived human-likeness, and thus, likeability. On the other hand, the nature of the task requested by the robot can affect the users willingness to follow the instructions. People involved in the regulation of the autonomous systems and their integration in the society also need confidence in the system. Finally, developers and engineers need to have confidence in their prototypes as well, and also have the possibility to highlight if there are issues and where they are. Another key requirement for trust is also *transparency*: the human will trust the social robot more likely if he can have some understanding of the robot's action and the reasons for its choices [Charisi *et al.*, 2017].

3.2 Robot Ethics

The main concern of robot ethics is to guarantee that autonomous systems will exhibit an ethically acceptable behaviour in all situations in which they interact with human beings. In particular, robot ethics is an applied ethical field whose objectives is to develop scientific/cultural/technical tools that can be shared by different social groups and beliefs. These tools aim to promote and encourage the development of robotics for the advancement of human society and individuals, and to help to prevent its misuse against humankind [Siciliano and Khatib, 2007].

The responsibility for improper or illegal behaviour of the robot can be attributed to the owners, designers, and/or builders of the machines. The question becomes increasingly difficult as the robot become more autonomous and capable of modifying its behaviour through learning and experience, since obviously the behaviour will be no longer based entirely on their original design.

Most of the ethical requirements that the robot has to follow are set by regulatory or standard bodies. In addition, the manufacturers might have built-in more specific ethical codes without contradicting those prescribed by the regulators. Finally the users could decide to add ethical preferences, to make sure that the robot's actions are personally acceptable [Charisi *et al.*, 2017]. Moreover the choices of criteria for a robot to be considered ethical involve the whole of society, therefore *transparency* is of utmost importance.

Finally, while the first concern is to develop robots that behaves ethically in society, it is important also to concern about how the autonomous robot can protect itself against misuse (e.g., taking advantage of the capabilities of the robot to commit criminal acts). Such misuse can be achieved by hacking an existing system or developing an unethical one.

For instance, sophisticated humanoids raise a number of ethical issues, including the following:

- loss of privacy for the human inhabitants, e.g., if the robots are permitted free access to all rooms in a home or if the robot's computer is accessed by hackers;
- ability of the robots to recognize commands that may lead to unethical behaviour;
- rights and responsibilities of the robots, e.g., should they be treated with respect as if they were human;
- emotional relationships, e.g., how a robot should relate to human anger, can a robot be punished for misbehaviour (and if so, how);
- how should a robot react to multiple instructions from different humans.

From the social and ethical standpoint, the assistive robots bear the most sensitive safety and ethical problems (e.g., patients may become emotionally attached to the robots, so that any attempt to withdraw them may cause distress; the robots will not be able to respond to the patient's anger and frustration, such as when a patient is refusing to take medication; a robot may be called by more than one patient and not being able to prioritize the request).

4 Our activities

We have focused on the verification of ethical behaviour in autonomous systems, and trustworthiness of social-care robots. Recently our interest in human-robot engagement has been increasing and we also have been involved in a practical case study in cooperation with Tate Modern museum.

4.1 Verification of robot ethics

In our society people can trust the decisions of professionals because they are subject to regulations and certification. With autonomous systems, with no human directly in control, ensuring that the system actually matches the required criteria is more difficult. In order to be confident with the robot's behaviour it is crucial to *specify* what actions to expect from the system in particular scenarios, *verify* that the system actually achieves this, and *validate* that the requirements are what the user want [Charisi *et al.*, 2017]. Typically those requirements can be technical, legal or ethical (e.g., never choose to do something dangerous for the user). In particular, is essential that the ethical requirements are certified by a regulator body.

Thus the aim of verification is to ensure that our system meets its requirements. Formal Verification also carries out a comprehensive mathematical analysis of the system to prove whether it corresponds to these formal requirements. By using tools, such as *model checking*, we can prove whether a particular property, that is an expression of the requirements, holds for the model of the system. In this way, the requirements are checked against all possible executions of the system. Verification via model checking is widely used for the analysis of safety and reliability of robotic systems [Dennis *et al.*, 2016b]. We have also recently used formal verification to address ethical issues for autonomous systems [Dennis *et al.*, 2016a; 2015], focusing on the possibility to verify formally whether an autonomous system will behave ethically, given a particular ethical setting.

In work such as [Arkin, 2007; Woodman *et al.*, 2012] the ability of the agent of being also an *ethical governor* has been introduced and verification has been explored in [Dennis *et al.*, 2015]. Such agent will choose the most ethical plan available, allowing unethical choices to occur only when it does not have a more ethical choice.

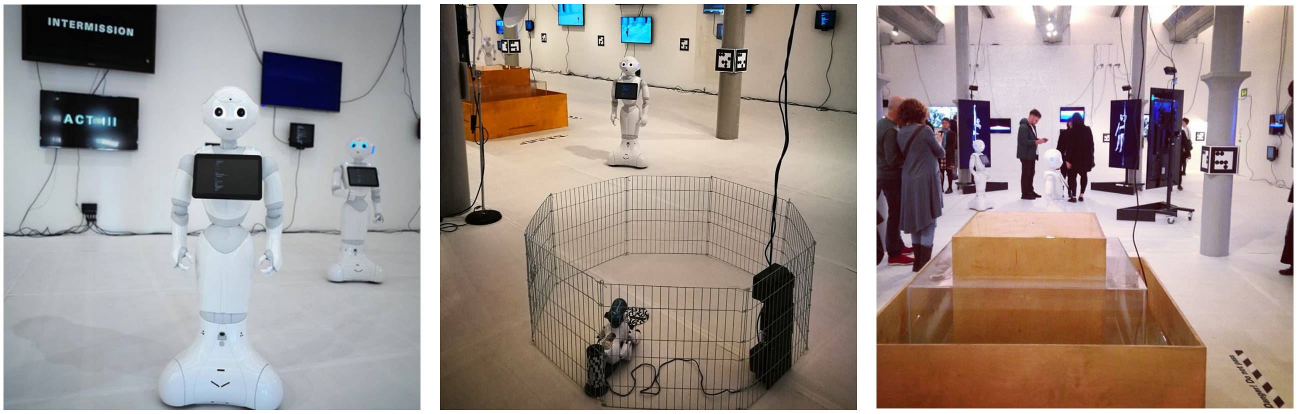
We also have conducted formal verification of an autonomous personal care robot, Care-O-bot, [Dixon *et al.*, 2014; Webster *et al.*, 2015], that is able to autonomously assist a person living within the house. We modelled the robot of Care-O-bot and its environment using Brahms, an high-level multi-agent language. Formal verification was then carried out by translating this to the input language of an existing model checker.

4.2 Practical engagement

For our social experiments in interaction between human and robots we started recently to use Pepper Robots, a humanoid robot developed by Aldebaran and Softbank Robotics (see Figure 1).

Pepper robots. Pepper is a human-shaped robot, designed mostly to be a companion robot. It is the first humanoid robot capable of recognising the principal human emotions, adapting his behaviour to the mood of his interlocutor, and also learning the user's preferences in order to improve the social interaction.

It can observe human expression by its camera system and identify human voice via its speech recognition system. They respectively enable it to function in a complex environments and to identify movements, and to detect where sounds are



(a) (b) (c)

Figure 1: Exhibition at Tate Liverpool. **(a)** Pepper robots. **(b)** Pepper robot and Sony Aibo. **(c)** People moving around robots.

coming from and locate the user's position, while also allowing the robot to identify the emotions transmitted by the user's voice.

Its emotion recognition function render the robot flexible in coping with the situation and interact better and in a more social acceptable way with humans. The constant dialogue between perception, adaptation, learning and choice is the result of what is known as the emotion engine. Furthermore, its anti-collision system (e.g., lasers, infra-red, sonar sensors), enable Pepper to detect both people and obstacles, and therefore to reduce the risk of unexpected collisions.

Human-robot engagement example. As an example of human-robot engagement we report on our involvement, as programming team, with an exhibition at the Tate Modern art gallery in Liverpool[1]. The artist, Cecile B. Evans[2] is interested in the increasing influence that new technologies have on the way we feel, and the way we relate to each other. She created a play where the performance is outsourced to two humanoid robots (Pepper) and a robot dog (Sony Aibo), who collaborate with a group of human users appearing on screens (see Figure 1).

In staging this collaboration between humans and robots, Evans hints at the possibility of the technological singularity — the hypothesis that at some point in the near future, artificial intelligence will surpass human intelligence [Bostrom, 2014]. But the work departs from the conventional narrative of "killer robots" and instead imagines a future scenario in which robots and humans will collaborate, working together to fight against external forces. Together, the users and robots navigate a series of events that they learn about through the screens that uncover aspects of the complex relationship between humans and machines.

Also, while the exhibition was running, we were able to collect feedback from visitors: they were mainly feeling comfortable moving around the robots, amazed at how the robots could move naturally, and interested at the idea of robots helping people in a dangerous situation.

[1] https:/news.liverpool.ac.uk/2016/10/21/robotics-experts-support-new-tate-liverpool-art-installation/

[2] http://cecilebevans.com/

5 Future work

A significant challenge in using social robots, especially in domestic and social-care environments, is ensuring that the interaction with the human is safe, that the user can trust the robots, and therefore that we can verify and validate that all the ethical requirements are met. We are already working on research fields such as verification and validation, dependability and trustworthiness.

In the near future we are planning to support further this research by utilising a social robot laboratory to investigating the operation of autonomous robotic systems in different physical and virtual environments. In particular the facility will improve our research on how humans and robots interact with each other in a domestic environment (social-care or domestic-assistant scenarios). Another future development would be more focused on the trustworthiness. More in particular, how the trust of the user change if the robot exhibit faulty behaviour, especially in a domestic environment (ongoing work with Kerstin Dautenhahn).

6 Conclusions

The future of autonomous robotic systems and their proper integration within our society depends on many different aspects. It is clearly relevant how people perceive the robots and interpret their behaviour. For this reason social robots are provided with increasing social skills.

For autonomous robots to be allowed to share the environment with people they need to be safe and their behaviour has to follow some ethical requirements. Therefore it is important to collect certifications about what to expect from a robot's behaviour, and verify that all these requirements are met. With the increase of autonomy in robotics it is also crucial that the user can trust the robot's behaviour. Indeed, people will never use a social robot, or even share a domestic area with it, if they are not confident that it is behaving safely and that its decisions comply with ethical and social limits.

In order to overcome these issues we have investigated the possibility to use formal verification to guarantee that the autonomous robot is behaving within technical (i.e., safe interaction) and ethical requirements.

Acknowledgements. Work described in this paper continues to significantly benefit from collaboration three leading experts: Kerstin Dautenhahn (*social robotics researcher*); Cecile B. Evans (*artist*); and Alan Winfield (*robot ethics researcher*). We thank them all.

References

[Arkin, 2007] R. C. Arkin. Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture. Technical Report GIT-GVU-07-11, Georgia Tech, 2007.

[Bostrom, 2014] N. Bostrom. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.

[Charisi *et al.*, 2017] V. Charisi, L. Dennis, M. Fisher, R. Lieck, A. Matthias, M. Slavkovik, J. Sombetzki, A. F. T. Winfield, and R. Yampolskiy. Towards Moral Autonomous Systems. *ArXiv e-prints*, March 2017.

[Dautenhahn, 1995] K. Dautenhahn. Getting to Know Each Other: artificial social intelligence for autonomous robots. *Robotics and Autonomous Systems 16:333–356*, 1995.

[Dautenhahn, 1997] K. Dautenhahn. I Could be You: The Phenomenological Dimension of Social Understanding. *Cybernetics and Systems 28(5):417–453*, 1997.

[Dautenhahn, 2007] K. Dautenhahn. Socially intelligent robots: dimensions of human-robot interaction. *Philosophical Transactions B: Biological Sciences 362(1480):679–704*, 2007.

[Dennis *et al.*, 2015] L. A. Dennis, M. Fisher, and A. F. T. Winfield. Towards Verifiably Ethical Robot Behaviour. *CoRR*, abs/1504.03592, 2015.

[Dennis *et al.*, 2016a] L. A. Dennis, M. Fisher, M. Slavkovik, and M. P. Webster. Formal Verification of Ethical Choices in Autonomous Systems. *Robotics and Autonomous Systems 77:1 – 14*, 2016.

[Dennis *et al.*, 2016b] L. A. Dennis, M. Fisher, N. K. Lincoln, A. Lisitsa, and S. M. Veres. Practical Verification of Decision-Making in Agent-Based Autonomous Systems. *Automated Software Engineering 23(3):305–359*, 2016.

[Dixon *et al.*, 2014] C. Dixon, M. Webster, J. Saunders, M. Fisher, and K. Dautenhahn. "The Fridge Door is Open"–Temporal Verification of a Robotic Assistant's Behaviours. In *Proc. Advances in Autonomous Robotics Systems: 15th Annual Conference, TAROS 2014, pp.97–108* Birmingham, UK, September 1-3, 2014.

[Fujita, 2004] M. Fujita. On Activating Human Communications with Pet-type Robot Aibo. *Proceedings of the IEEE 92(11):1804–1813*, 2004.

[Fussell *et al.*, 2008] S. R. Fussell, S. Kiesler, L. D. Setlock, and V. Yew. How People Anthropomorphize Robots. In *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction*, pages 145–152, 2008.

[Hirzinger *et al.*, 2001] G. Hirzinger, A. Albu-Schffer, M. Höhnle, I. Schäfer, and N. Sporer. On a New Generation of Torque Controlled Light-weight Robots. In *Proc. ICRA, pp3356–3363*. IEEE, 2001.

[Kiesler *et al.*, 2008] S. Kiesler, A. Powers, S. R. Fussell, and C. Torrey. Anthropomorphic Interactions with a Robot and Robot-like Agent. In *Social Cognition 2008*, volume 26, pages 169–181, 2008.

[Matthias, 2011] A. Matthias. Robot Lies in Health Care: when is deception morally permissible? *Kennedy Institute of Ethics Journal 25(2):279–301*, 2011.

[Mori, 1970] M. Mori. Bukimi no tani [the uncanny valley]. *Energy 7(4):33–35*, 1970.

[Pratt and Williamson, 1995] G. A. Pratt and M. M. Williamson. Series Elastic Actuators. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems. Human Robot Interaction and Cooperative Robots*, volume 1, pages 399–406 vol.1, 1995.

[Salem *et al.*, 2015] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn. Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust. In *Proc. ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp141–148*. ACM, 2015.

[Sharkey and Sharkey, 2012] A. Sharkey and N. Sharkey. Granny and the Robots: Ethical issues in robot care for the elderly. *Ethics and Inf. Technol. 14(1):27–40*, 2012.

[Siciliano and Khatib, 2007] B. Siciliano and O. Khatib. *Springer Handbook of Robotics*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.

[Syrdal *et al.*, 2007] D-S. Syrdal, K. Dautenhahn, S. N. Woods, M. L. Walters, and K-L. Koay. Looking Good? Appearance Preferences and Robot Personality Inferences at Zero Acquaintance. In *Proc. AAAI Spring Symposium Multidisciplinary Collaboration for Socially Assistive Robotics, Technical Report SS-07-07, Stanford, California, USA*, pages 86–92, 2007.

[Tonietti *et al.*, 2005] G. Tonietti, R. Schiavi, and A. Bicchi. Design and Control of a Variable Stiffness Actuator for Safe and Fast Physical Human/Robot Interaction. In *Proc. IEEE International Conference on Robotics and Automation (ICRA), pp526–531*. IEEE, April 2005.

[Turkle *et al.*, 2004] S. Turkle, C. Breazeal, O. Daste', and B. Scassellati. Encounters with Kismet and Cog: Children respond to relational artifacts. In *Proc. IEEE-RAS/RSJ Int. Conf. Humanoid Robots*, pages 1–20, 2004.

[Webster *et al.*, 2015] M. Webster, C. Dixon, M. Fisher, M. Salem, J. Saunders, K. L. Koay, K. Dautenhahn, and J. Saez-Pons. Toward Reliable Autonomous Robotic Assistants Through Formal Verification: A Case Study. *IEEE Trans. Human-Machine Systems, 46(2):186-196*, 2016.

[Woodman *et al.*, 2012] R. Woodman, A. F. T. Winfield, C. Harper, and M. Fraser. Building Safer Robots: Safety Driven Control. *Int. Journal of Robotic Research*, 31(13):1603–1626, 2012.