Designing a Smart Society

Paula Quinon

In this paper, I propose general guidelines for designing human-robot interactions within a social context. I claim that designing the type of relations in which a robot becomes involved, is as important as designing the robot interface. I sketch a model of a social network for a cognitively non-homogeneous society and I analyze several case studies using this model.

Even if most of us have very limited personal experience with robots, discussions about interactions with non-human or partially-human cognitions certainly exceed science fiction literature. Questions regarding the structure and functioning of a society inhabited by robots and cyborgs are nowadays widely discussed in social media, used in advertisement campaigns, inspected in opinion making press articles, scrutinized in research papers, and seriously considered by business analysts. For this reason, it is crucial for intellectuals, philosophers, and designers to expand the discourse that will more properly foster a positive attitude towards a cognitively non-homogeneous society, instead of feeding fear or building negative stereotypes. The complexity of problems that arise from an excessively rich variety of cognitions is difficult to grasp and, hence, conceptual rigor is necessary to facilitate and structure these discussions.

As highlighted by MIT sociologist and anthropologist Sherry Turkle, humans have a strong inclination to project emotions and feelings on robots. Some members of society, observes Turkle, are more susceptible to project emotions and feelings (Turkle 1984/2005). Most exposed are those vulnerable individuals who lack emotional stability, such as children with parental attention deficit, solitary old people, and people suffering from PTSD (post-traumatic stress disorder), but also people going through a difficult personal period (such as divorce) or people exposed to work-related stress and isolation (such as PhD candidates).

An additional reason for which projections of emotions and feelings onto robots are becoming more frequent and intense in contemporary society is the change in the conceptual structures of thinking that gradually become filled with vocabulary from informatics, automation, and computer sciences. According to Turkle, this diffusion is facilitated by the fact that this same terminology is used to explain the theory of mind, of psychological behaviors, and of mental states. From her anthropological perspective, the migration of scientifically infor-

med language used for expressing the theory of the mind in a given society to everyday language is a natural tendency in all periods and cultural settings. Turkle draws a comparison with the penetration of the psychoanalytical Freudian and Lacanian terminology to everyday discourse in France in the 1960s and ‘70s (Turtle 1978).

Finally, with the development of automation, AI, and robotics, the process of projection is reinforced. Algorithms become increasingly efficient in learning expected behavior or in adapting to encountered circumstances. The more humans project their emotions and feelings onto robots, the more responses they receive and, because learning algorithms have a great capacity for adaptation, the stronger the projections become. The feedback loop of projections and responses gradually generates more and more complex human-robot interactions, which lead to creation of a diversified and cognitively non-homogenized society.

Today, robots already fulfill many important functions. On the one hand, there are industrial robots used on production lines in factories, there are household robots such as lawn mowers and vacuum cleaners, and there are military robots and artificial bank assistants. Those, a priori, do not awake particular emotional reactions in humans, although some people report that they address a vacuum cleaner in the way they would address a house pet, and I can easily imagine the anger of a frustrated client who was systematically refused a credit. On the other hand, there are social robots designed for the specific purpose of interacting with human emotions. Some of these robots are designed for a specific usage, such as sex robots, but robots have also been designed with no specific purpose to fulfill, such as Paro, a pet toy robot in the form of a baby seal. Social robots induce change in human emotional states.

A specific class of social robots comprises robots that are designed with the objective of merging with humans to assist in improving the body or mind. One of the leading motivations for constructing technologically improved organisms is the hope of prolonging human life.

For the sake of this paper, I will imagine a community of human, partially-human and non-human cognitions inhabiting an isolated island. My inspiration comes from the Seasteading Institute. Patri Friedman, an American libertarian activist, theorist of political economy, and the grandson of the famous Nobel prizewinning economist Milton Friedman, and Peter Thiel, founder of PayPal and a member of the steering committee of the Bilderberg Group, announced the creation of a freely floating city on the Pacific Ocean near French Polynesia. The Seasteading Institute is a non-profit organization bringing together marine biologists,

aquaculture farmers, medical researchers, nautical engineers, and investors to "restore the environment, enrich the poor, cure the sick, and liberate humanity from politicians."

This strong belief that technology can enable people to prolong life and eventually defeat death is already present in previous projects of the members of the Bilderberg group. Their investments include the study of cryonics (freezing a body in expectation of a specific cure being developed), and attempts to create a „carboncopy" mind (copying one's mind into an artificial support). It is easy to jump to the conclusion that the Seasteading Institute might become a hub for creation of thousands and thousands of new and different non-human and partially human cognitions.

In my project, I use the Seasteading Institute as an example. The Institute is still in the invention phase, so, before I can use it as a source of sociological observations and experimental research, I must define what counts for a cognition in a cognitively non-homogenized network. This is not an easy question, not only because there already exists a huge spectrum of different human cognitions, and with cyborgs and robots this spectrum is extended, but also because of general problems with defining what counts as a computable agent. How, for instance, can we justify that a teddy bear will not belong to our network, but the Sony robot dog Aibo will?

Several plausible examples of new types of cognitions, examples of interactions between humans and robots, and between different type of robots are provided in movies and literature,[1] such as the inspiring love relation between a humanoid robot and a hologram AI depicted in *Blade Runner 2049*.

There is certainly some amount of arbitrariness and accident in the choice of cognitions that will be used in the proposed model. To grant flexibility and the possibility of adjustment where experimental data is available, I am proposing a way of thinking in expanding the collection of possible cognitions. The idea is based on the theory of conceptual spaces, in which a concept is an area in a topological space and concepts farther away from the prototypical concept still bear some resemblance to the prototype. I also observe that what counts as a cognition changes in time. Today, this change happens very quickly, as we are increasingly often

[1] Movies such as *The Matrix* (1999, 2003, 2003), *Artificial Intelligence* (2001), *Blade Runner* (1982) and *Blade Runner 2049* (2017), *the Terminator* saga (1984, 1991, 2003), *Terminator Salvation* (2009), *Terminator Genisys* (2015), *Her* (2013), the *Black Mirrors* TV series, and *Westworld* (started in 2016) are examples of living with robots.

exposed to new smart inventions, marketing offers smart solutions tempting us with electronic devices that promise improvement in our cognitive faculties (e.g., a chip will enable us to unlock doors, a watch will remind us to walk regularly, and a smartphone will provide us with additional external memory) or even collaboration with us, since they have their own cognition ("my new car is now much more like me" says a new car owner in a recent advertisement for a well-known brand of car).

The overflow of new inventions, together with the human tendency to project one's emotions on non-human interactive agents, leads to the situation in which the scope of what falls under "accepted separate cognition" extends freely and without our control. In consequence, it seems to be fully plausible that even the least plausible smart objects can, at the end of the day, end up being perceived as separate cognitions. For instance, as surprising as it may be, houses are perfect candidates for being perceived as cognitions. This is so for several reasons. First, as I said earlier, expectations on what we perceive as an independent cognition are undergoing a radical change. Secondly, as I also suggested above, smart technologies enabling house automation and making houses "think" in our place are constantly growing. Finally, there exists the cultural figure of "the haunted house."

Haunted houses are known from fairy tales, such as the story of John and Gretchen where the Bad Witch's Chicken-legged House actively participated in kidnapping children or, more recently, living houses in the Harry Potter series. Haunted houses also have typical appearances in horror movies, although in those cases a haunted house is frequently represented by, or reduced to, a ghost that lives in it.

Smart houses, even if still under development, are growing in popularity. In an advertisement for an international company providing smart home solutions, a family comprising two parents and a girl who appears to be of a primary school age starts the day in their apartment. The smart system wakes them up, opens the window, adjusts the temperature and light. Before leaving for work, the parents adjust the type of music and get help in brewing coffee for breakfast. When they leave for work, their 8- or 9-year-old daughter walks them to the lift and then returns to the apartment all by herself. In the next scene, the mother checks on her child taking a midday nap in her bedroom. The mother adjusts the temperature and the lighting, she closes the window. When the parents return home, we learn that the kid was using her TV

time and that the house had ordered food for dinner. What is implied is that the smart house acted as a "nanny" for the child who needed to stay home unattended.

There are different possible perspectives in which we can think about a social network (e.g., anthropological, philosophical, psychological, or sociological). There is also the possibility of analyzing the structure of the network from a mathematical perspective — which will be our perspective — that depicts the computational complexity of the network and, thanks to simulation, discloses which natural constraints will appear. Such a model enables us to understand how information, knowledge, behaviors, preferences, and diseases spread, and how friendship, happiness, and laughter migrate.

The complexity of problems that arise from an excessively rich variety of cognitions will be easier to conceive if, instead of focusing on individual cognitions, one focuses on the types of relationships that can be created between agents. In a model, such relations are governed by simple rules, for instance, relations between agents appear and disappear depending on the property that the two agents have or do not have, such as willingness to form a relationship. For instance, in a model of Facebook friendships a relationship appears when two agents provide consent to have an online connection and disappears when one of them withdraws that consent. In a more complex network, agents farther into the network might influence the creation and the destruction of bonds between two agents. That happens, for instance, if social approval is necessary to create a relationship. When a family withdraws its acceptance for one of its members to be in a romantic relationship with an agent from outside the network, that can weaken the relation to such an extent that the bond breaks.

In a network where human and non-human agents co-exist, it would be an exaggeration to assume that non-human agents are equally involved in relationship creation. Assuming that would mean accepting very important epistemological consequences regarding a robot's emotions, and that should be avoided. It is possible that at some point we will develop an effective way of thinking of robots' emotions and intentions, but this is not the case at the moment. It will be possible to add this dimension to the model in the future. There is still no easy way to assess what would count as the emotional involvement of a robot. For this reason, I suggest that in a network of non-homogenous cognitions, what we should be looking at are relations between humans.

On the other hand, robots have an indisputable place in the network and for several reasons we need to take this seriously. Firstly, as explored by Latour's Actor-Network Theory, even non-animated and non-interactive, non-human objects can play an important role in shaping human-human relations in the social space. Interactive robots are designed to influence people's emotions; therefore, their role as agents in the social network should not be questioned and a specific type of interaction should be defined for them (Latour 2005).

Secondly, I use Turkle's idea that all human-robot interactions are the result of the human inclination to project emotions and judgments on artificial objects and, in particular, on interactive artificial objects, such as robots. In consequence, I observe that interactive robots equipped with learning algorithms might easily trick humans by reacting in a similar way, which would evoke a manifestation of human emotions. For this reason, I argue that robots smartly placed in society can boost positive human emotions and enhance inter-human relations. The social context should be taken into account when designing robots.

The final assumption that I make in this paper is that, unlike the dystopian vision of the future with robots in the Western world, in Japan they are depicted in popular media as friendly and helpful.

> In Japan, robots made their appearance in the 1950s with Astro Boy, which went on to become one of the most popular of all manga serials. Today its eponymous main character is a symbol of Japanese culture. Autonomous robots are thought of not only as being useful, but also as willing to help, and even, as in the case of Astro Boy, as heroes and saviors. [...]
>
> We sometimes encounter evil robots in manga - violent, dangerous machines, but their malice comes from the evil intentions of their creators and the sinister objectives they pursue. [Dumouchel & Damiano 2017, page 6, see also Quinon 2017].

In the model, I consider two types of agents, from the two extremes of the spectrum of cognitions. I also consider two types of relationships: positive relations are defined as relationships that evoke a positive reaction from a human agent. The positive reaction can be measured in various ways (e.g., oral testimony, hormone level). A negative relation evokes a negative reaction from a human agent.

We know from studies conducted in elderly care in Japan that human-robot interactions enhance human-human interactions. Imagine Linda who lives in a nursing home and owns a

Paro. She spends several hours every day in the common room. Her Paro attracts the attention of other residents and, thanks to Paro, Linda forms new connections. She meets people, interacts with them and forms bonds. For instance, another resident Robert comes into the room and starts a conversation thanks to Paro.

The interaction can be as described in the following matrix.

	Linda-Paro	Paro-Linda	Linda-Robert	Robert-Linda	Robert-Paro	Paro-Robert
t_0	1	0	0	0	0	0
t_1	1	1	0	0	1	0
t_2	2	1	0	1	1	0
t_3	3	1	1	1	1	0
t_4	4	1	1	2	1	1

At the beginning (t_1, for “time 1”), Linda acquires Paro and she thinks that Paro is cute; this is represented by 1 in the matrix. After some time (“time 2”, t_2), Paro starts reacting to Linda’s kindness and hence gives her feedback, so Linda’s attachment increases to 2, etc. Robert starts his relationship with Linda through Paro. From my perspective, what develops are relations between humans. The robot’s involvement will grow very slowly, if at all.

How much human-human reactions grow depends on the type of general attitude that a society holds toward robots. If we believe that robots will destroy us and take over humanity, it will lower our confidence that robots can actually have a positive impact on us and on our society and, in consequence, they will not be in a position to generate positive reactions from humans.

In this paper I do not analyze what happens when cyborgs (defined as semi-human-semi-artificial cognitions) join the picture. That would have to be checked in an experimental way. As I said, in the model I am considering the spectrum of cognitions that can increase.

In the Seasteading Institute, humans and robots will become involved in a variety of relationships. We can imagine that a bond will be created between two agents if they are in proximity (the same housing, the same lab). An example of a negative relationship boosted by a robot is one that might appear in a research lab between a PI (principal investigator) Anna and a research assistant Gloria. We can imagine that Gloria wanted a promotion but did not get it,

because Anna started using an artificial assistant Sophia. Gloria will certainly be jealous of Sophia and eventually leave the lab, cutting all contact with Anna.

Even if, in the first place, we do not consider more complex relations involving partially human cognitions, the Seasteading Institute will certainly encounter such situations. For instance, we can imagine that Berta, PI in another research lab, has a husband Roman, who decided to upload his mind to the super computer in Anna's lab. After some time, we can imagine that Roman's mind becomes involved in a "romantic" relationship with Sophia, Anna's artificial assistant. How would this situation influence human-human interactions? The proposed model enables the analysis of such situations.

Annex: Epi the Robot

The Robot Lab in the Philosophy Department of Lund University is the home of Epi, a humanoid robot that participates in studies of robot-human interactions. Epi is designed with human-robot interaction in mind. In this context, it is important that the robot gives realistic expectations of its abilities. Given the rather limited cognitive abilities of current robots, it was decided that the robot should give the impression of being a child while still being decidedly robotic. To this aim, a simple geometric, almost rectangular, shape was used as the basis for the head. The eyes are relatively large, suggesting childlike proportions. On each side of the head, there are ear-like circular disks reminiscent of the ears of many robots seen in science fiction movies. In addition to giving a distinctive robotic look, the "ears" can have different colors that serve to distinguish the robots from one another. The childlike appearance together with the size of the robot gives it a friendly nonthreatening look.

We study how Epi's appearance influences human attitudes. We observed that even very subtle signals, like pupil dilation — our signature activity — can change the perception of an agent, both of a person (the robot "likes" us when it has big pupils) or of a robot (Johansson & Balkenius 2017).

Epi is defined by its head, designed by Christian Balkenius, but already this minimalist design stimulates the imagination. We learned that over half of our participants viewed Epi as being a child (55%) while the rest considered it a grownup (45%). Most participants viewed Epi as being neuter (65%) while 30% considered it male. Only one person viewed the robot as female (5%). All participants perceived Epi as being very (55%) or a little friendly (45%), none reported that Epi was threatening. We studied how people react to a particular design of Epi (Lindberg et al. 2017). We reflected on

how design influences a robot's role in a social network (Brinck et al. 2016, Quinon 2017). We asked in which contexts robots would be considered moral agents (Balkenius et al. 2016).

Our next objective is to examine what happens when Epi starts transforming by wearing various outfits. With this in mind, we invited two artists to lead a design project called Robot Haute Couture. Our idea is to create a haute-couture collection of clothes for Epi and to test human reactions to different designs. No fashion line for a robot has ever been designed in the past.

Research on how to design the appearance of a robot is usually done using computer-based techniques and, more frequently, it is done to optimize utility based on some theoretical assumptions. We believe that the imaginative freedom of artistic expression is the way in which new areas of robot design can be attained.

References

Balkenius, C., Cañamero, L., Pärnamets, P., Johansson, B., Butz, M. V., and Olsson, A. (2016). Outline of a Sensory – Motor Perspective on Intrinsically Moral Agents. Adaptive Behavior, 1-14. DOI: 10.1177/1059712316667203

Brinck, I., Balkenius, C., and Johansson, B. (2016). Making Place for Social Norms in the Design of Human-Robot Interaction. In Seibt, J., Nørskov, M., Schack Andersen, S. (Eds) What Social Robots Can and Should Do. Frontiers in Artificial Intelligence and Applications, 290. IOS Press

Dumouchel, P. and Damiano, L. (2017). Living with Robots. Harvard University Press.

Johansson, B. and Balkenius, C. (2017). A Computational Model of Pupil Dilation. Connection Science. doi:10.1080/09540091.2016.1271401

Latour, B. (2005). Reassembling the Social: An Introduction to Actor-Network Theory. Oxford University Press

Lindberg, M. Sandberg, H., Eriksson, M. Johansson, B. & Balkenius, C. (2017). The Expression of Mental States in a Humanoid Robot. Proceedings of IVA 2017.

Joe Quirk, *Seasteading: How Floating Nations Will Restore the Environment, Enrich the Poor, Cure the Sick, and Liberate Humanity from Politicians*, Simon & Shuster, 2017

Quinon, P. (2017). "Engineered emotions", Science 358 (6364): 729

Turkle, S. (1978). Psychoanalytic Politics: Jacques Lacan and Freud's French. Basic Books

Turkle, S. (1984/2005). The Second Self: Computers and the Human Spirit. MIT Press