UNFOLDING THE INTERACTION BETWEEN HUMANS AND ROBOTS – AND THE POTENTIAL PITFALLS

In this point of view, we will illustrate some of the mechanisms that evoke our tendency to treat robots as humans. To explore this path, we will take a deep dive into the human psyche and uncover why we characterize robot behavior as human behavior. Subsequently we showcase the various implications that these heuristics have in our current interaction with robots, and how we can better manage these in the future.

IT'S A MATCH!

MANAGING EXPECTATIONS IN HUMAN-ROBOT INTERACTION

While robot Sophia[1] is joking around with Jimmy Fallon in The Tonight Show and Google Assistant[2] makes phone calls for you, robots in Japan are taking care of significant problems like labour shortage and elderly loneliness. Robots are and will increasingly be an integrated part of our society, and we will increasingly interact with them. A way to optimize this interaction is to make robots look and act like humans. Such "anthropomorphization" sounds perfectly harmless, but there are two sides to this coin. In this article you will find out why we treat robots as if they are human and more importantly, why managing expectations in human-robot interaction (HRI) is pivotal to its effectiveness.

1. Sophia is a social humanoid robot developed by Hong Kong-based company Hanson Robotics. Fun fact: In October 2017, Sophia became a Saudi-Arabian citizen – the first robot to receive citizenship of any country.
2. Google Duplex, a virtual voice bot showcased in May 2018, claimed to be indistinguishable from a human when making phone calls.

WHY DO WE TREAT ROBOTS LIKE HUMANS?

What is a robot?

Before going into depth on this topic, it is useful to clarify how we understand robots and bots. Robots, on the one hand, share three basic similarities: they all have some kind of mechanical construction designed to achieve a specific task. They usually include electrical components which power and control the machinery, and they contain some level of computer programming code. Examples include everything from industrial robots to social robots like Pepper to NASA's lunar rovers.

Bots, on the other hand, are software applications that run automated tasks, often over the Internet. Bots typically perform tasks that are simple, structurally repetitive and perform it faster than a human would be able to. The mere difference between robots and bots is that robots have a physical manifestation while bots do not – bots are purely digital entities. Like chatbots.

Both concepts stem from the same origin, the word "robota", which is Czech for "forced labor". Whether it is a robot arm that attaches a wheel to your car or a bot that tells you the product you wanted to buy online is not in stock – a human made them do it, or at least set the boundaries for their response.

Today's technological developments enable us to make more use of robots and bots, especially by using them in different and more advanced ways. Frank Wammes, CTO and Innovation Lead at Capgemini explains: "Interaction with robots will be a common thing of the future, as it is actually already a thing of the present. We just don't realize that a lot of activities we do are already robot interaction. Usage will morph into our lives, as did the mobile phone which all of us considers now to be a key component of our daily life."

The intelligence of robots

Both robots and bots can be viewed as Artificial Intelligence (AI), yet not all of them are considered AI. AI is often seen as letting machines mimic "intelligent"thought and behavior. But as machines become increasingly capable, the number of tasks still considered "intelligent" slowly decrease. Tesler, therefore, defined AI as: "whatever machines haven't done yet".

Take for example your dish washer, thermostat, or your computer that can play chess. When introduced, these were considered "smart" solutions. Today, it is normal to have them around, and we do not consider them as smart anymore.

Whether a (ro)bot is actually intelligent has nothing to do with (lack of) its physical appearance. While some bots can be considered intelligent, most of them are simple and straight-forward applications. And while some robots look a lot like human beings, it does not make them more intelligent than the ones that do not.

Some measure of intelligence can exist within a physical appearance (robots) or without physical appearance (bots). How intelligent we perceive a (ro)bot, however, is affected by its physical appearance, due to what we call 'anthropomorphism'.

Considering all things human

Anthropomorphism, stemming from the Greek words "anthropos" (man) and "morphe" (form or structure), can be defined as the tendency to attribute human characteristics to animals and inanimate objects with an aim to make sense of a situation.

Anthropomorphism has been around since classical times. Xenophanes, a Greek philosopher, was the first to coin the term. He satirized traditional religious views of his time by arguing that human projections of gods are visualized as a reflection of their own image: "Ethiopians say that their gods are snub-nosed and black, Thracians that they are pale and red-haired."

Modern psychologists generally characterize anthropomorphism as a cognitive bias. Anthropomorphism is a process by which people use their schemas (a set of mental images of preconceived ideas) about other humans as a basis for supposing the characteristics of non-humans in order to make quick judgements about the environment.

The extent to which we engage in anthropomorphization is dependent on the extent to which the object resembles a human being to us. The more human-like an object is (human movement, eyes, facial expressions, etc.), the more likely we tend to attribute human characteristics to it. For the same reason, we anthropomorphize (ro)bots for the same reason. Robotics expert Volker Darius explains: "We give robots a physical appearance, a gender, a name, to make them tangible for human beings." Not surprisingly, the more a robot resembles a human, the more we tend to anthropomorphize the robot.

Anthropomorphization carries many implications. Although we have the tendency to anthropomorphize, we do not attribute human qualities to all objects we encounter. Entities that possess similar traits to humans, are more likely to be anthropomorphized. Nicholas Epley, a researcher from the University of Chicago, devised a theory to explain the reasons we unconsciously anthropomorphize, categorized in three root-causes:

ELICITED AGENT

We have a far richer knowledge on humans compared to non-humans. For this reason, we are more likely to use anthropomorphic explanations for other entities since we lack adequate mental models for non-humans.

EFFECTANCE MOTIVATION

We have the tendency to anthropomorphize when we are motivated to explain non-human behavior due to our tendency to resolve uncertainty, seek meaning, and feel efficacious.

SOCIALITY MOTIVATION

When social connection on the human level is lacking, we compensate for this by ascribing human attributes to non-human agents.

Epley proposes that by our very nature we try to understand an entity's behavior by ascribing mental states to it. This involves ascribing human characteristics to the entity with little or no reference to its real competences. Obviously, the tendency to anthropomorphize is a source for error and perilous situations.

WHAT HAPPENS WHEN WE TREAT ROBOTS LIKE HUMANS?

What are the consequences of anthropomorphizing robots?

Last week, we established the causes for anthropomorphization. Let's dive into the implications that are involved with our unconscious tendency to anthropomorphize our observations of computer behaviors. This tendency is also known as the ELIZA effect named, after the 1966 chatterbot ELIZA, developed by MIT computer scientist Joseph Weizenbaum. ELIZA's script was meant to parody a "50s psychotherapist by constantly rephrasing the patient's replies as new questions". Though ELIZA never said anything "new", only rephrasing what the human user said, it prompted powerful emotional responses from these users.

As usual, there are two sides to the story. On the one hand, anthropomorphization positively affects interaction between humans and robots. Users experience smoother and more natural communication and collaboration with (ro)bots due to the resemblance to human interaction. On the other hand, anthropomorphization can also evoke unfounded assumptions and misperceptions about the robot, leading to disappointment and potential interaction breakdown. Emotions like unrealistic fears ("Robots will take over the world!"), naivety ("It's just code") or inordinate optimism ('"Robots will soon resolve all human problems, hail singularity!"). Hence, designing robots to resemble humans should be carried out wisely by the builder and should serve a specific purpose.

Should we build human-like robots?

As said, the sense in anthropomorphization lies in the fact that we have a natural tendency to search for familiarity. Anthropomorphization can thus be used to smoothen HRI and integration. "It's your new colleague, give her a name", Volker Darius advises his clients looking into robotics.

However, anthropomorphization does not always make sense. Humanoid shapes are not the pinnacle of most efficient design for any given task. Human-like machines make it easier for humans to interact with them, simply because of our natural tendencies. Pepper, for example, is easy to interact with due to its humanoid shape features. Contrasting, Pepper is dreadful at making coffee.

Being aware of both the enabling and limiting aspects of anthropomorphization, we should consciously assess the appropriateness of creating anthropomorphized machines. There is little sense, for example, to appending human characteristics into a waffle iron. In other words, enhanced anthropomorphic features will induce expectations of intelligent behavior, but a humanoid waffle iron' functionalities are bound to result in disappointment. We argue that the degree of anthropomorphization should be in line with the human expectations of the social interaction.

How can we shield ourselves against the negative effects of anthropomorphization?

Evidence indicates that we are susceptible to telling our darkest secrets to anthropomorphic robots. Their cute faces and child-like voices can be a facade to exploitative code and unintended reactions. 'How can we protect ourselves from these deceptions?' This concern has also been voiced by British scientist Stuart Russel at the time the vice chairman of the World Economic Forum Council on robotics and AI. He called for a ban on "highly human-like humanoid robots". "We are just not equipped in our basic brain wiring to see something that's perfectly humanoid and not treat it as a human being, so in some sense, a humanoid robot is lying to us using the lower levels of our brain we don't get to control", Russel argues. This statement refers to our more unconscious brain processes, over which we have less control.

The perils of anthropomorphization lie in these natural tendencies of human beings. It can blind us to the true potential and limitations of technology. A robot can speak "I am your friend", without harboring any feelings of friendship like a human. It's not that it has feelings of hatred against you, it just is not able to have feelings at all. Potential deception is lying in wait. Awareness of these social dynamics is vital in the interaction with robots. We are urged to be consciously aware of the dynamics of HRI and the potential pitfalls that humans, by nature, are susceptible to. Especially, now with the knowledge of the ELIZA effect, we must be mindful of the imaginable unethical behaviors of robot developers with immoral intentions. Ethical developers should also be considerate of the potential implications. Referring back to the anthropomorphic waffle iron and the dynamic associated with it, there is an equilibrium between social expectations and anthropomorphization that needs to be maintained

How should robot designers deal with these consequences?

When developing a robot that is meant to interact with humans, one should ask: what is the job that the robot should perform? But more importantly, what should it not perform? How do we ensure the robot does what it is expected to do? And – what can we enable with the use of current technologies? These questions do not imply striving for a robot as advanced as possible, but that the humans interacting with the robot know what to expect from the it, and what not.

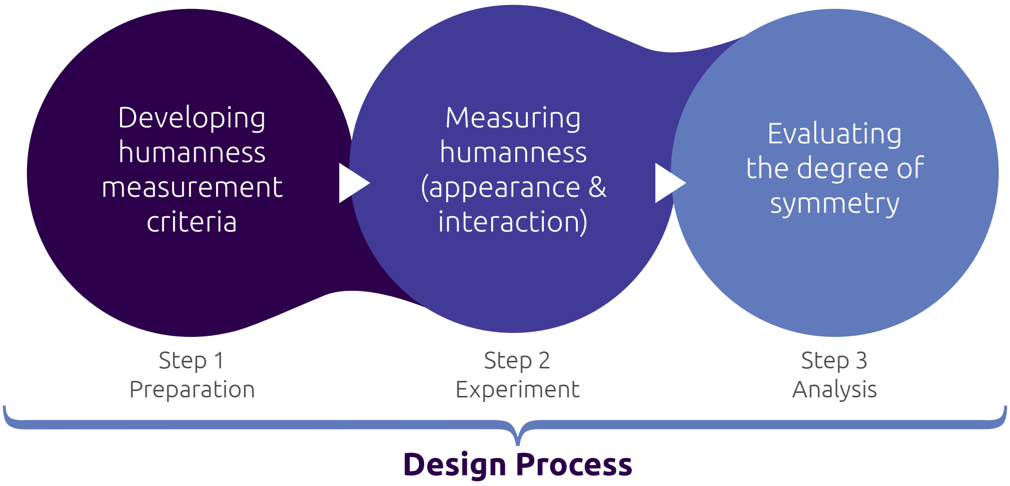
To facilitate an appropriate design for a robot to perform a particular job, the designer should follow a set of questions to guide the development process (see Figure 1). These questions are not dictating the technical design of a robot but approach it from an anthropomorphic point of view. The first and second question entice the developer to consciously think about the desired performance. This should be closely linked to the third question regarding the projected expectations from a user's point of view. These three questions should form the foundation of robot design with a view towards it, as they define the degree of anthropomorphization that is appropriate. Lastly, the designer should think about the enablers to facilitate functionalities and interaction through technology. There is a thin line between finetuning (hinting) the design and overselling (declaring) it. Designers propose that robots should hint at, rather than declare, anthropomorphization. This would help users form closer and smoother connections with robots. The more humanoid and intelligent your robot looks, the higher the expectations that need to be met and the greater the risk of disappointment.

Figure 1: **Guiding Questions**

1. What is the job that the robot should perform?
2. What should it not be doing?
3. How do we make sure that the robot does what it is expected to do?
4. What can we enable with current tech nologies?

Developers can evaluate the usage of anthropomorphic features using three simple steps (see Figure 2): designing measurement criteria, measuring humanness in both appearance and interaction, and evaluating the degree of symmetry in the two values. In the design of the robot the humanlike appearance and interaction – the symmetry - should be in equilibrium.

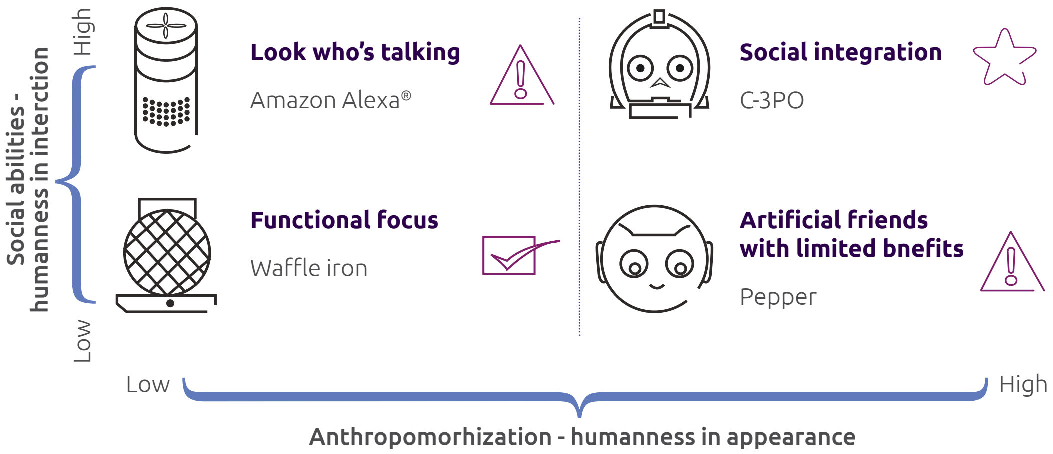
Figure 2: **Design Process**



A disequilibrium, as said before, will lead to disappointment from the user. Interactions between the human and the robot will not play out as intended. Hence, safeguarding symmetry in robot appearances and interaction is a continual challenge for designers and a topic of awareness for users. Examples of alignment and misalignment are bountiful around us.

Take for example C-3PO (see Figure 3) a robot showing many human-like features - highly anthropomorphic. At the same time, his social interaction is also very human-like. You can say that his human appearance and his social abilities are in sync. This is what makes human contact with C-3PO so smooth and may be one of the reasons he had so many owner[3].

On the other hand, when human appearances are lacking while social abilities are highly present, interaction is not as smooth as it could be. If Amazon's Alexa® would be more anthropomorphic users might be more inclined to use its services. But resembling a human is a disadvantage when social abilities are lacking, as can be seen with Pepper®. The android robot looks amazingly cute and human, which fuels our paradigm with expectations of highly social abilities. Unfortunately, this is not the case and people get easily disappointed in the interaction with Pepper®. Finally, the waffle iron is not anthropomorphic and not social, but it makes you excellent waffles on a Sunday morning – exactly what you expect it to do. Form and function should be in balance.



3. https://c3p0fanclub.wordpress.com/2009/10/26/a-chronological-list-of-c-3pos-owners/

MANAGING INTERACTIONS BETWEEN MAN AND MACHINE

How can we manage HRI expectations and responses?

An often-cited thought experiment to envision a dystopian future of robotics is the Paperclip Maximizer. An AI, tasked with generating as many paperclips as possible. Here, anthropomorphization drives us to expect the machine understands a command – producing paperclips – in the same way us humans do. Instead, the AI transforms earth and all humans on it into paperclips, because that was the singular task it was given.

It illustrates that an AI can be a powerful optimizer without sharing any (recognizable) human values. A robot that is not specifically programmed to be benevolent to humans could thus be as dangerous as if it were designed to be malevolent. This is why Isaac Asimov, biochemist and writer of "I, Robot" designed the Three Laws of Robotics (see figure) in 1950. Although meant as guidelines on coding human ethics into robots, the validity and relevance of these laws are under discussion.

Current AI research focuses heavily on machine learning - giving computers the ability to "learn" implicitly from observing data, without explicit programming. These implicit learning techniques (like neural networks) are hard to combine with explicitly coded laws like Asimov's. The thought process of these "black box" models is much less clear than rule-based systems and may be completely alien to us. To some experts[4], this prompts the addition of a fourth law: a robot must be transparent in its reasoning.

Furthermore, the experts argue that another law should be added: a robot should not deceive its user, nor be designed in a deceptive way. The machine's nature should be transparent- not a robot in disguise.

This was a major point of critique on Google's Duplex, the technology behind a new Google assistant feature. It is a fully automated system that mimics a PA by making appointments for you. Its conversation with a hairdresser was extremely hard to mistake from a human's. Google has since issued a statement that Duplex will always identify itself as a bot. Whether this identification clause is a core part of the Duplex system's coding or just a quick fix on top remains to be seen.

Due to the highly rapid development of robotics, technology and AI in our lives, laws of robotics are continuously subject to change. Moreover, our understanding of HRI is still in its infancy. Newly acquired knowledge impacting HRI should be a constant feed for updating the laws of robotics. Hence, while developing robots it is very important to keep in mind what we want to use it for, so that we make sure that we "raise" the robot in the right way - especially when it is no longer only coded explicitly but also "learning" implicitly. Give the machine the right data to "grow up with". Code love, not hate and make sure to communicate effectively with the people worried by or scared of these robots. Expectation management of what robots can and cannot do is always of the essence .

Figure 4: **Laws of Robotics**

Asimov

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Robotics experts

4. A robot must be transparent in its reasoning
5. *A robot should not deceive its user nor be designed in a deceptive way*

4. Principles of robotics - EPSRC website. (2018). Retrieved from https://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/

What should we do first?

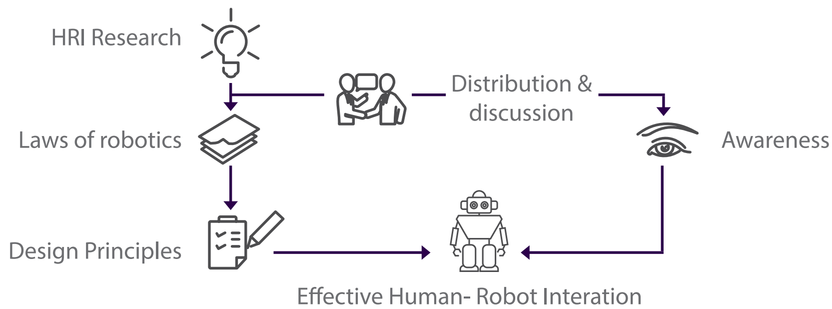
Anthropomorphism can enable efficacy with users, a sense that increases one's apparent competence interacting with robots, with negative and positive implications. To ensure a smooth, effective and ethical interactions between humans and (ro) bots, we need to ensure three information flows are present: (1) increased awareness of users on the effects of anthropomorphization, (2) the enforcement of appropriate design of robots through laws of robotics and design principles and (3) a continuous feed of research in the field of HRI to refine the laws of robotics. Facilitating anthropomorphism may therefore serve as an effective method for improving the usefulness and acceptance of certain technological agents.

Anthropomorphism in robots is a double-edged sword, facilitating smooth interaction with humans but also creating unrealistic expectations due to human mental models that supplant capabilities from humans to social robots. Both users and developers of robots should be aware of the intended interaction when balancing form and function.

In this point of view we illustrated some of the mechanisms underneath our tendency to treat robots like humans. We discussed the mechanisms, their underlying causes, the potential implications thereof and how to manage al of the above in a world where the distinction between man and machine is increasingly hard to define.

For questions or remarks, feel free to contact the authors.

Figure 5: **Human-Robot interaction**



For more information:

Tanja Runhaar
Email: tanja.runhaar@capgemini.com
Tel. +31 6 3441 4373

Maurice Smulders
Email: maurice.smulders@capgemini.com
Tel. + 31 6 1558 6333

About Capgemini Invent

As the digital innovation, consulting and transformation brand of the Capgemini Group, Capgemini Invent helps CxOs envision and build what's next for their organizations. Located in more than 30 offices and 10 creative studios around the world, its 6,000+ strong team combines strategy, technology, data science and creative design with deep industry expertise and insights, to develop new digital solutions and business models of the future.

Capgemini Invent is an integral part of Capgemini, a global leader in consulting, technology services and digital transformation. The Group is at the forefront of innovation to address the entire breadth of clients' opportunities in the evolving world of cloud, digital and platforms. Building on its strong 50-year heritage and deep industry-specific expertise, Capgemini enables organizations to realize their business ambitions through an array of services from strategy to operations. Capgemini is driven by the conviction that the business value of technology comes from and through people. It is a multicultural company of 200,000 team members in over 40 countries. The Group reported 2017 global revenues of EUR 12.8 billion. People matter, results count.

Visit us at
www.capgemini.nl/invent

Capgemini Invent
Postbus 2575 - 3500 GN Utrecht
Tel. + 31 30 689 00 00
www.capgemini.nl/invent

IN/06-015.18