A nominal survey paper on robots, autonomy, and AI

Harald Maartmann-Moe

In this paper, I will briefly present a few articles in the field of robots, autonomy, and AI.

Cummings (2004) discusses the significance of automation bias and proposes a classification system for levels of automation in intelligent time-critical decision support systems. The proposal is motivated by the increasing application area and development of intelligent support systems. With a precise language, which these levels of automation contribute to, the ethical concerns of AI in time-critical decision support systems can be discussed with more clarity.

Cummings (2004) states that "higher levels of automation" should be avoided in "critical environments" as neither the systems or the automation can be perfectly reliable. She also condenses the advantageous area of application of AI decision making by underlining the tasks that humans "are better at" (See table 1).

Table 1. Strengths of Humans and Computers in Decision Making

Humans are better at:	Computers are better at:
Perceiving patterns	Responding quickly to control tasks
Improvising and using flexible procedures	Repetitive and routine tasks
Recalling relevant facts at the appropriate time	Reasoning deductively
Reasoning inductively	Handling many complex tasks simultaneously
Exercising judgment	

(Cummings, 2004)

The levels of automation range from full human control to a fully automated black-box system with no power distributed to humans.

Table 2. Levels of Automation

Automation Level	Automation Description
1	The computer offers no assistance: human must take all decision and actions.
2	The computer offers a complete set of decision/action alternatives, or
3	narrows the selection down to a few, or
4	suggests one alternative, and
5	executes that suggestion if the human approves, or
6	allows the human a restricted time to veto before automatic execution, or
7	executes automatically, then necessarily informs humans, and
8	informs the human only if asked, or
9	informs the human only if it, the computer, decides to.
10	The computer decides everything and acts autonomously, ignoring the human.

(Cummings, 2004)

The benefits of intelligent automation in decision support systems are grand in the application areas described in Table 1. The benefits are further emphasized by drawing attention to heuristics and biases in human thought. However, there are dangers to automation as well. System *Brittleness* is brought up by Cummings (2004) as one of these weaknesses. Which describes the inability to account for all relevant parameters which could result erroneous or misleading suggestions. There are even pitfalls when automation merely supports human decision making: Automation bias concerns the tendency of humans to refrain from looking for contradictory information when presented with an automated suggestion (Cummings, 2004). Cummings (2004) gives an example study from aviation where "[...] it was recommended that unless decision aids are perfectly reliable, status displays should be used instead of command displays." As automation bias caused an error rate of 41% in certain circumstances where the error rate previously had been 3%.

In sum, there are biases to decision systems no matter the distribution of control and responsibility between humans and AI. Fortunately, these biases can be curbed through a better understanding of them in the said systems' design. Cummings (2004) draws attention to an example preliminary study where the decision support system includes a display of a reliability trend, which reduced the effects of automation bias.

Forlizzi (2007) enters the context of homes in an ethnographic study of robotic products. The study provides insight concerning the home as a context and the accompanying complex challenges. The paper extends the research on how adding qualities of autonomy and intelligence affect responses to a product. This study can be described in four stages: (1) getting to know participants and activities of daily living through conversational interviews, (2) a visual story diary on current cleaning events, (3) the family received a robotic or non-robotic cleaning product and once again –after some time-- participated in a visual story diary, (4) finally follow-up interviews were conducted supplied with diary entries. The use of theory was diverse: Forlizzi addressed the states of the HRI and HCI communities in describing the motivation and relevance of the study, a product ecology and the term *social products* was introduced. The non-robotic counterpart in the study did not have the same ability to enact long-term change. The robotic product had a greater impact on the ecology. Forlizzi (2007) argues that intelligent and autonomous products' "fundamental changes in the structure and infrastructure of the home" to fit in. The case findings support this statement, as the case demonstrated lasting change to "[...] who cleaned and how they cleaned." (Forlizzi, 2007). The case is further utilized to emphasize the importance of aesthetics of products to be used in the home. The case is also used as an example in the presentation of Forlizzi's (2007) five dimensions of social products: function, aesthetics, symbolism, emotion, and social attribution- Forlizzi (2007) offer a design implication to achiever less stigmatizing and more adoptable systems: include simple social attributes in their design.

Forlizzi and DiSalvo (2006) present lessons from two ethnographic case studies of an autonomous mobile robot. The studies entailed introducing cleaning technology to families followed by semi-structured interviews. The aim of the study was to “provide a grounded understanding of how design can influence human-robot interaction in the home.” An autonomous, mobile robot was utilized. They present three main lessons: "first, the way the technology is introduced is critical; second, the use of the technology becomes social; and third, that ideally, homes and domestic service robots must adapt to each other." This paper aims to better our understanding of the home and the "[...] complexities that will be encountered when deploying robotics into the domestic environment." The home has been a new and growing context for HRI, hence the new challenges and complexities introduced along with the home as a context should be thoroughly investigated.

Coeckelbergh (2010) undertakes four objections to the introduction of AI assistive technologies in health care practices. Coeckelbergh distinguishes *replacement* from *assistance* in this debate, claiming the latter is less controversial, therefore focusing on the former. He argues that higher standards are required of technology when considering its introduction, compared to the standards existing solutions hold. The first objection Coeckelbergh undertakes is dubbed 'deep care' and regards the view that "[...] the technology does not really care about the patient.". This objection is related to the argument concerning higher standards, as Coeckelbergh points out that not all current practices "really care", yet technology has to fulfill this demand in order to be introduced. Coeckelbergh also argues that deep care can be emotionally problematic for the caretaker and that there might not be time for deep care in current practices. The second objection is dubbed 'good care' and concerns the quality of care. The objection is centered around the human emotional and social needs. Coeckelbergh considers Sparrow and Sparrows (2006) view that only other humans can meet these social and emotional needs. The third objection dubbed 'private care' is met by the counterarguments that privacy is not a new issue, and it can be accounted for in design, use, and regulation. Coeckelbergh also argues that privacy should be balanced against other principles, and existing practices are not optimal in the area of privacy. The fourth objection is dubbed 'real care'. The objection is based on the notion that the technology does not solve issues, it merely fools people by simulating the feelings corresponding to those of that of a better situation. People might feel healthier, but they are not. Coeckelbergh describes a 'Care Experience Machine' thought experiment to discuss what an optimal solution might be. To face these challenges with introducing AI assistive technology Coeckelbergh proposes a modified Capabilities approach with emphasis on the inherent social dimension of care. This approach is given with the encouragement to balance it against other principles, as to avoid paternalism among other pitfalls.

Coeckelbergh mentions the work of Linda and Robert sparrow during the discussion of the objection dubbed 'good care'.

> *Robert and Linda Sparrow, for instance, have argued against the replacement of human nurses by robots in elderly care for the reason that robots are incapable of meeting the social and emotional needs of elderly persons, which can only be done by means of contact with humans (Sparrow and Sparrow 2006). Their main concern is, rightly so, the quality of care.*
>
> (Coeckelbergh, 2010)

The view of robots as incapable of meeting the social and emotional needs of elderly persons possibly opposes elements of Forlizzi (2007) concept of social products. The term social products (Forlizzi, 2007) also concerns technology that we create social relationships with -- we create social relationships with the technology itself. Forlizzi's (2007) paper does not describe the nature or limit of the social relationships created with technology. Therefore, the term may both be compatible or incompatible with Sparrow and Sparrows (2006) view of robots.

As Coeckelbergh (2010) does not provide counterarguments to a view of robots as incapable of meeting the social and emotional needs of elderly persons, Coeckelbergh might also have objections to the notion of social products as Forlizzi defines the term (2007).

Mutlu and Forlizzi (2008) examine the issues around robotic technology in organizational contexts. They found little work on this subject, and therefore characterize their work as *nascent theory*. Their study is qualitative utilizing ethnographic data collection and grounded theory analysis. They used natural environment open-ended interviews and participant- and fly-on-the-wall-observations. Through grounded theory analysis they produce a diagram of their findings summarized as follows:

> *Patient profile and the kind of healthcare service provided cause differences in units' workflow, goals, social/emotional context, and use of their physical environment. (1) When staff interruptibility is low, interruptions by the robot are perceived as worsening the workflow. (2) A misalignment between the goals of the unit and the benefits provided by the robot might cause people to reject the use of the robot. (3) Intimate relationships between caretakers and patients cause a lower tolerance for interruptions. (4) In high traffic and/or cluttered hallways, the robot is perceived as taking precedence over people.*
>
> (Mutlu & Forlizzi, 2008)

Mutlu and Forlizzi (2008) stress the importance of "[…] the complex social dynamics […]" in the context of organizations. They conclude that "When technologies such as service robots are adopted by organizations, they have an impact on social dynamics and work practices of many groups. […] we showed that aspects of workflow, and social/emotional, political, and environmental context influenced how workers at a hospital used, perceived, and interacted with the robot." In their ethnographic study, they saw dramatic differences in the relationships to-, and approval of- a robotic system. These differences came as a result of differences in organizational factors (Mutlu & Forlizzi, 2008). The robotic system was met with approval in the context of the post-partum unit. People even formed affable relationships with the robots in the post-partum unit, yet the same robotic system was met with disapproval in the context of the medical unit.

«[…] nurses at the medical units, who had low tolerance for interruptibility, found the robot to be a nuisance […]»

(Mutlu & Forlizzi, 2008)

The social aspect of tolerance for interruptions was critical in the approval and adoption of the robotic system.

Mutlu and Forlizzi also promote the use of participatory design methods, and stress aspects of similar social exchange theory "When the cost of using the robot outweighs the benefits provided by its adoption, people are less willing to use the robot.» (Mutlu & Forlizzi, 2008).

Mutlu and Forlizzi propose designing for social relationships as affable social relationships were visible in the context where the robotic system succeeded. This proposal arrives sensibly, however, one might question whether an affable social relationship and the systems' success has some degree of reverse causality: Was it the affable social relationship that caused the robotic system to succeed, or was the affable social relationship formed due to the robotic systems' success? In the latter case, designing for (affable) social relationships could be the equivalent of *designing for one of the consequences of success*, as opposed to the general goal in any design: *designing for success*. This objection depends upon the definition of success as an affable social relationship might be considered success – or overlap with success --

in some cases. If this is a given There might be bi-directionality with (affable) social relationships and a robotic systems' success.

Susi and Ziemke (2005) address the confusion related to the term affordance. They describe how the term has been used to describe different phenomena, and thus contributed to misunderstandings and inaccuracy. To clarify the affordance confusion Susi and Ziemke (2005) undertake terms, views, and concepts introduced by Jakob von Uexküll (1864-1944), David Kirsh (1989-), Martin Heidegger (1889-1976), and James J. Gibson (1904-1979) as they have all typically fall under the term "affordance".

Susi and Ziemke (2005) detail the concept of Functional tone from von Uexküll's work. "the receptor image of rods and wholes [a ladder] had been supplemented by the effector image of his own action; through this, it had acquired a new meaning. The new meaning manifested itself as a new attribute, as a functional or effector tone" (von Uexküll as cited in Susi and Ziemke, 2005). Condensed by Susi and Ziemke (2005) "functional tone concerns the way people ascribe certain meanings to objects from the subject's point of view». Functional tone is constructed by the subject in interacting with the object and shaped by the subject's mood. Hence, objects may have different functional tones with different moods, and with different subjects. Functional tone is a property of the subject, not the object.

Susi and Ziemke (2005) discuss a second term often confused within the umbrella term "affordance", the concept of equipment developed by Martin Heidegger. Condensed by Susi and Ziemke (2005) "equipment relates objects to a whole chain of other objects and the context of an activity, which gives the object its meaning in the first place". Susi and Ziemke find similarities between Heidegger's and von Uexküll's concepts as neither supports the Cartesian subject/object distinction (Susi & Ziemke, 2005).

Heidegger emphasizes context in describing equipment, summarized by Susi & Ziemke (2005) "A piece of equipment has to fit into the context of an activity, since it is only within a meaningful context an object is what it is.». "for Heidegger, there is an interdependent relation between subject and object, and they cannot be considered as separate entities.» (Susi & Ziemke, 2005).

Gibson's concept of affordance is also described by Susi and Ziemke (2005) "affordances concern possible actions in an environment (which can be made salient or, in case of unwished actions, may be hidden)». Gibson shares Heidegger's view of secondhand knowledge's effect on object use and perception and his opposition to the traditional psychological dichotomization between mind and body and the existence of intermediary processes in perception (Susi & Ziemke, 2005). However, Gibson locates affordance in the physical environment, to be perceived --through light-- by subjects rather than perceiving the qualities of the objects (Susi & Ziemke, 2005). Gibson also describes tools as objects that extend our capacities for perceiving and acting.

Susi and Ziemke (2005) also present the concept of entry point from Kirsh's work. They describe entry points as something that "…invite us to do something…" Susi and Ziemke (2005) and condense the concept to "entry points describe different characteristics of objects, from a cognitive point of view, which affect the way people react to them.» to highlight differences in the discussed terms. Entry points can both be "located" in the object, or in the subject. Susi and Ziemke (2005) touch on the importance of understanding the agent-environment relationship in "the design of artificial subjects like robots, and their possibility to perceive and actively adapt the world according to their needs.» Susi and Ziemke (2005) main contribution in this paper was clarity in the fuzzy area of "affordance".

Ziemke (2016) critically examines the concept of embodied cognition. He has found that robots are typically considered embodied in the HRI field. Yet, *embodiment* is rarely defined. Ziemke (2016) therefore questions “What is a body?” and “What is embodied?”. Robots are typically described as embodied as “[…] they are physical and interact with their environment through sensors and actuators.”. This attribution is based on a specific understanding of bodies, and embodiment as being grounded in sensorimotor interaction. However, there are alternative understandings such as those presented by Johnson (2007) that would disqualify many robots accredited “embodied”. Ziemke (2016) discusses a view of embodiment “grounding sensorimotor interaction in bodily regulation”, and “sensorimotor interaction with the environment is itself deeply rooted in the underlying biological mechanisms and more specifically layered/nested networks of bodily self-regulation mechanisms.” homeostasis is of necessity. If the view of embodiment rescinding robots of this an embodied status is correct, it would imply that embodied AI cannot develop through robots —or models of robots— lacking layered/nested networks of bodily self-regulation mechanisms. The main contributions of the paper are this implication, along with the healthy questioning of fundamental assumptions in embodied AI, and a discussion of intentionality, robotic functionalism, autonomy, and central works in cognitive psychology and embodied AI.

Sharkey and Ziemke (2001) distinguish between two ends in a view of embodiment as a spectrum: phenomenological and mechanistic.

> *“Phenomenal embodiment refers to embodiment of a mental or subjective world. It has its roots in von Uexku ̈ll’s idea of bringing together an organism’s perceptual and motor worlds in its Umwelt (subjec- tive or phenomenal world).»*

(Sharkey & Ziemke, 2001)

> *In the mechanistic view, cognition is embodied in the control architecture of a sensing and acting machine. There is nothing else.*

(Sharkey & Ziemke, 2001)

Strong intelligence requires the system itself to understand what it is processing, while weak AI serves more as a tool (Sarle as cited by (Sharkey & Ziemke, 2001)).

Von Uexküll (as cited Sharkey and Ziemke 2001) distinguishes between robot bodies, and living bodies through their origin. A robot body is constructed by assembling independently produced parts, while a living body is grown outwards from single cells.

Bibliography

Coeckelbergh, M. (2010). Health Care, Capabilities, and AI Assistive Technologies. *Ethical Theory and Moral Practice*, *13*(2), 181–190. https://doi.org/10.1007/s10677-009-9186-2

Cummings, M. (2004). Automation Bias in Intelligent Time Critical Decision Support Systems. American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2004-6313

Forlizzi, J. (Ed.). (2007). How Robotic Products Become Social Products: An Ethnographic Study of Cleaning in the Home. In *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction* (pp. 129–136). New York, NY, USA: ACM. https://doi.org/10.1145/1228716.1228734

Forlizzi, J., & DiSalvo, C. (2006). Service Robots in the Domestic Environment: A Study of the Roomba Vacuum in the Home. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction* (pp. 258–265). New York, NY, USA: ACM. https://doi.org/10.1145/1121241.1121286

Mutlu, B., & Forlizzi, J. (2008). Robots in Organizations: The Role of Workflow, Social, and Environmental Factors in Human-robot Interaction. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (pp. 287–294). New York, NY, USA: ACM. https://doi.org/10.1145/1349822.1349860

Sharkey, N. E., & Ziemke, T. (2001). Mechanistic versus phenomenal embodiment: Can robot embodiment lead to strong AI? *Cognitive Systems Research*, *2*(4), 251–262. https://doi.org/10.1016/S1389-0417(01)00036-5

Sparrow, R., & Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds and Machines*, *16*(2), 141–161. https://doi.org/10.1007/s11023-006-9030-6

Susi, T., & Ziemke, T. (2005). On the subject of objects: Four views on object perception and tool use. *TripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, *3*(2), 6–19.

Ziemke, T. (2016). The body of knowledge: On the role of the living body in grounding embodied cognition. *Biosystems*. https://doi.org/10.1016/j.biosystems.2016.08.005