Aspects of Long-Term Autonomy of Social Robots and Their Potential Impact on Society

Marcel Häselich

Active Vision Group
University of Koblenz-Landau
Universitätsstr. 1
56070 Koblenz
mhaeselich@uni-koblenz.de

Abstract: Robots emerge from simple working machines to more and more complex systems. Current systems are equipped with sophisticated learning techniques to enhance their application domain in a dynamic way. Robots are no longer limited to domains that are hazardous for human beings. Their power and diversity allows them to perform tasks which are impossible for humans. Further, inconvenient tasks can be assigned to robots. Plans exist to introduce robots to various new fields such as nursing and elderly care. The integration of robots into our everyday life will have a strong influence on society and arise new topics to public attendance.

In the future, robots will be able to learn, behave, evolve, integrate and may possess their own consciousness. The scope of this paper is the description of the properties and aspects of such robots and a critical inspection on their potential impact on society. Three different fictive stories of autonomous robots in the future give insight to the possible integration of these new contemporaries of mankind.

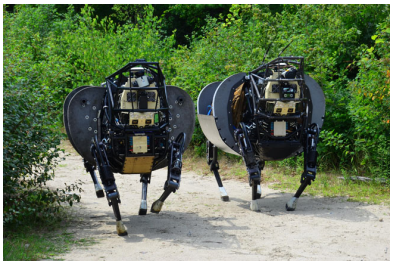
The first story deals with Rose and Brian Miller, two married robots who plan their future. A robot limitation law prevents robots from having an unlimited number of replica. Only the most superior robots are rewarded with additional replica. Ronald B1000 is a butler robot who's employer wants to replace him with a more sophisticated exemplar. While it is impossible for Ronald to let go of his job, he struggles on and on to meet the increasing needs of his master. The last story accompanies John 137, a robot worker who decides to kill himself due to his bad implementation.

1 Introduction

The origin of the word *robot* lies in the Czech/Slavic word *robota* and means bond servant or thrall. If someone today speaks of a robot, he mostly refers to physical robots. Virtual software robots are abbreviated and addressed to as bots. Robots may either perceive their environment, interact with it, or do both. Some robots are completely stationary, e.g. in factory environments, others are able to move around on surfaces or walls, dive or swim, or are able to fly. Especially robots with complex manipulators are often built according to the image of a human or an animal. Modern robots are able to move faster and further than any human and they are equipped with extraordinary sensors. Color cameras, night vision cameras, thermal imaging cameras, stereo vision setups, laser range

(a)

(b)



(c)

Figure 1: The image on the left shows the android HRP-4C (nicknamed Miim). The roboticist Prof. Hiroshi Ishiguro is shown on the top right image together with his android copy. The image on the bottom right shows two LS3 mule robots developed to transport goods and to accompany soldiers on missions over rough terrain. (image sources: (a) [IsEG13], (b) [oAIST13], (c) [Dyn13])

finders, and sonar sensors are examples of *robot senses*. Radar sensors even allow a robot to perceive velocities of other objects. Robots can carry huge payloads, possess great communication skills and have a nearly unlimited knowledge storage. However, they are seldom present in everyday life and are a rare phenomena. Some examples of state-of-the-art robots are illustrated in Fig. 1. What most current robots are missing is an intelligent and adaptive behavior. They are designed and able to fulfill separate tasks or sequences of them. Beyond their task they are almost completely unable to perform tasks they were not programmed for. Learning of new tasks and skills and learning for robots in general is an active research topic. Robots need to be able to enhance their knowledge and abilities and create their own goals in order to unfold their full potential. Therefore, some sort of robot-consciousness will be required and reveal completely new topics. The topic has been addressed previously from other viewpoints.

In 1942, Isaac Asimov [Asi42] published *The Three Laws of Robotics*: “One, a robot may not injure a human being, or, through inaction, allow a human being to come to harm. [..] Two, [..] a robot must obey the orders given to it by human beings except where such orders would conflict with the First Law. [..] And three, a robot must protect its own existence as long as such protection does not conflict with the First or Second Laws” (p. 27). An example of what might happen in the future if robots were unbound from these

laws is depicted in the movie *I, Robot* [PGVA04] that was created according to a short story collection of Asimov.

The Swedish science fiction series *Real Humans* (Swedish title: Äkta människor) [LHA12] takes places in a parallel universe. Humans are used to have androids accompanying them in various spheres of live such as factories, housework, and for sexual purposes. Some of the androids start to develop emotions and pursue their own goals. The movie series pictures the relationships and the complications that occur when androids replace humans in different positions.

The science fiction drama film *A.I. Artificial Intelligence* [SA01] tells the story of the childlike android David. In this future scenario, androids are capable of having their own thoughts and emotions and David is the prototype of an android that is able to love. David is integrated in a family whose son was placed in induced coma due to his disease and David starts to develop feelings for his new mother. Things complicate as the human son is cured and the android David is abandoned from the human mother he loves.

The term *social robot* (or socially interactive robot) describes an autonomous physical robot who interacts with humans and other robots by following social behaviors. Several research groups are developing social robots (cf. e.g. [oCS13, oFDoCS13]) with advanced social and cognitive skills to effectively interact and cooperate with humans.

This paper illustrates a future in which social robots may possess a consciousness. Aspects of such autonomous systems are presented in sec. 2. Consequently, the following section 3 describes three fictive scenarios from the point of view of some of these robots as thought-provoking impulse. Sec. 4 concludes the examples and their messages.

2 Aspects of Autonomous Systems

Nowadays, there already exists a number of different robotic systems in society. Specialized robots fulfill explicit tasks and are used in regions that are hazardous or inconvenient for humans. More and more complex systems are on the increase and cover further application domains step by step. This section disregards task-specific robots and rather describes the properties of complex autonomous social robots with an own conscience and profound capabilities. Therefore, the following aspects not only describe their properties, but also the properties they are *not* given.

Tranquility Many robotic systems are capable to move faster, further, and with a much higher payload than any human being. Further, a robot is not limited to the ground and might be able to fly or move underwater much superior to humans. These properties, among others, make robots the perfect soldiers. Therefore, the first property of the robots considered in this paper is their tranquility which completely disassociates them from their warmongering colleagues. In accordance with Asimov's first law, tranquility goes a step ahead and binds robot to integrate into our society and assist humans in any peaceful and friendly way. No human shall ever be disadvantaged, harmed, or killed, because of a robot having this property.

Obedience In accordance with Asimov's second law, any order of a human must be obeyed by robots. In the case that a hierarchy is established among the robots, a robot must obey orders from another robot that is higher in the hierarchy. Robots considered in this work may not obey commands that disadvantage, harm, or kill human beings.

Immortality Given the ability to repair themselves, robots become practically immortal. Under the premise that sufficient materials are available and the robot is able to afford them, he can conduct a reconstruction of any broken or battered part. In cases of heavier repair procedures, he will be able to shutdown himself for the duration of the repair process and restarts afterwards. Of course, he can return the favor to the robot that repaired him.

Replica Intelligent autonomous system may acquire the necessary skills to create other robots. This skill allows them to reproduce themselves and to create copies of themselves. Their immortality in combination with the power to create something of their own image makes them appear godlike. In order to prevent robot overpopulation, birth-control needs to regularize the amount of replica allowed per robot. A special law, created by humans, dictates replica creation and regularizes robot genesis.

Evolution In contrast to Asimov's second law, a robot may choose to end his own existence at any point and in any way. In evolutionary theory, the "Survival of the fittest" [Spe64] is a phrase originating by Herbert Spencer as an alternative description of "natural selection" [Dar59] by Charles Darwin. The corresponding term for robots would probably be *artificial selection* or *natural robot selection*. It will be implemented as an algorithm weighting the value of the robots own existence against its common use for society. Replica created from superior robots arise slowly, but still the number of robots increases as they are immortal. Hence, the algorithm will calculate the value of the robots life with respect to its abilities and the total number of robots. Once this calculation surpasses a threshold, evolution steps in and the robot knows the time has come to perform his final task. Of course, inaccuracies of measurements may cause robots to accidentally commit suicide inhumanly often, which is why the algorithm will be implemented in a probabilistic way robust to outliers in the measurements. This property prevents the market to be flooded with an unmanageable number of robots and allows superior robot specimen to prevail.

Emotions A newly created robot starts its life with a basic set of emotions enabling him to fulfill basic tasks in everyday life. Equipped with the basic emotion modules, the new robot will be able to feel the needs of the humans around him, especially the ones of his employer. Upgrades are available for his own emotions and disturbing emotions might be deactivated for arbitrary time slots once the robot has acquired them.

Introspection The opposite of emotions is the mandatory introspection ability of each robot. Introspection allows a robot to reflect its own behavior and the possible outcomes at any time in a completely emotionless manner. This valuable ability guarantees his total submission towards humans and greatly eases his integration in

society. For example, emotion modules might become quite disturbing if a robot is not allowed to create any further replica, but his introspection helps him accepting the facts and enables him to disable the disturbing emotion module. Another example is the deactivation of the fear-module that might prevent the robot to enter a dangerous environment or to commit suicide when the time has come.

Selective Swarm Intelligence New communication devices allow robots to share all their knowledge at any time. Whenever a robot learned something new he can immediately share this experience with the whole robot community. Further he is able to warn others right away not to perform the mistake he just made, e.g. in case of an accident.

Omnipresent Swarm Intelligence The new communication technology might not only be used in a selective way. Cheaper or older robots might not be able to retrieve or send the correct information to the swarm. Information might be lost or the robot is unable to receive important security news. Therefore, some low-level robots are only equipped with an omnipresent swarm intelligence. More sophisticated robots decide then what to do with those unfiltered information sources. This means that they do not select and share explicit knowledge but rather broadcast everything they know, learn, and experience. Despite their own desires to hide thoughts or failures they made, those robots share everything which is why some emotion modules are prohibited or at least not recommended for such robots.

This listing presents only a short excerpt of a vast amount of requirements and can further be extended and intensified at particular points. However, for the examples in the next section, these aspects yield a sufficient amount of unintentional and controversial situations.

3 Three case studies

The following fictive stories take place in a future society in which peaceful and useful robots are integrated in various social domains. They posses a consciousness and are only bound by the aspects listed in the previous chapter. Hence, they may never harm humans or disobey their orders and their whole existence is resided about improving the life of humans. These robots are equipped with sophisticated learning algorithms and are capable to adopt to any task or challenge that serves the improvement in everyday life of mankind. The stories reflect the point of view of some of these formidable robots.

Rose and Brian Rose and Brian Miller married 30 years ago. Robots are genderless but they are often created according to the image of humans and hence tend to live in couples. Two years after their marriage they created their first and only allowed replica Jenna. For the next twenty-two years, Rose and Brian lived a full and happy life until Jenna left the couple to create a replica on her own. Sadly, the Robot Replica Law prohibited more than one replica for robots of Rose and Brian's non-superior kind. Sadness and emptiness spread within them despite what emotion

modules Rose and Brian deactivated. Something beyond their implemented consciousness started to rekindle them. Since the algorithm that assigns more than one replica cannot be manipulated and will only choose the most superior robots, the yearly robot competition in their town was their only chance. Once in a year, robots living in their town compete in a dangerous trial of agility, strength, and computation power. Two years ago Brian participated for the first time but ended up with high repair costs and far from the top ranks. In the last year, the Walkers from the other side of the street won the competition and were allowed to have their third replica, Jules III. For the period of a whole year, Mrs. Walker tortured Rose by showing her Jules III all the time and telling her how magnificent he is and how delightful their life is again with their new family member. This is a main reason why Brian risked everything in last year's competition in order to win the replica permission. The competition was as hard as it has never been before. Brian received irreparable damages to both of his arms and parts of his hard drive were lost and could not be recovered. Nevertheless, he won the competition and Rose and he created a new replica, guaranteeing their happiness for the next twenty years.

Ronald Ronald B1000 is a personal butler robot that has been working for his employer for 18 years. Within this time, he fulfilled almost any task in the household and the garden on his own. He went to get the purchases, cleaned the rooms, cooked and served the food, and kept the house and garden clean and cozy. Over the years, new models emerged and new features were integrated into the B-series of the butler robots. Until the B1300, Ronald's employer never seemed to notice them or was at least to polite to let Ronald know. But with the release of the B1300, the robot series was added softer wheels and less noisier axle rods than the creaky ones of the B1000. Eventually, Roland upgraded his complete chassis and all his wheels. Until now, his employer never seemed to realize Rolands investment but he also never mentioned the superior features of the B1300 again. Half a year later, Ronald's employer came up with another robot model, the B1380. The B1380 was capable of carrying twice as much payload as any of the earlier models. His master actually told Ronald that it is a shame he can only carry such lousy amounts of payload. He further hinted that replacing Ronald with a newer B-series might have advantages for Ronald, too, since there are other more gentle application domains for hardworking deprecated robots. Ronald had to react immediately. He invested all his saved money and acquired a new carrying platform so that he was able to carry even more payload than the B1380. Thankfully, Ronald's master never mentioned the B1380 ever again. The new carrying platform had the disadvantage that Ronald gained several kilos of weight that moderately decreased his speed. Ronald was still able to fulfill all tasks for his master in a satisfying and fairly quick way and his life was in order again. Sadly one year later, the B1500 model was equipped with new expensive engines, which was directly brought to Ronald's attention by the master. Ronald knew what he had to do. He raised an incredibly high credit from the robot bank and bought the new engines. The B1500 was off the table and his existence was in pace again. Unfortunately his old-age provision was gone and he was sitting on a debt mountain. Ronald realized too late, that the algorithm that decides over his existence might take this financial disaster into account when deciding Ronald's value for society.

But since his existence prevailed over two years, Ronald started to relax and felt convenient again to serve his master and to be off value. If it were not for this new B1700 series with their advanced and extraordinary expensive telescopic arms.

John In the beginning of the age of autonomous factory working robots, an error in the manufacturing pipeline caused 200 robots to be created with a slight failure. Instead of a selective swarm intelligence, those robots were produced with an omnipresent swarm intelligence. Since the failure did not influence their abilities to fulfill their factory work, these robots were not recalled. One of those robots is John 137, an industrious factory robot. The first months his life went smooth and John 137 got to know the products, the factory, and their workmanship. He learned all processes with ease and acquired all necessary skills for his work. A couple of weeks after his apprenticeship, he made his first mistake. He dropped an expensive circuit board. John 137 grew nervous and let the broken board disappear secretly in the trashcan on the far end of the factory hall and was relieved that nobody got aware of his mistake. At the end of the shift, John 137 headed towards the exit when he was stopped by his foreman, a robot called Jim 12. He whispered to John 137 that he should not worry too much and that every robot breaks a piece once in a while. At first John 137 was relieved that his foreman was not angry and that there will be no consequences for him. Then he panicked as he realized that Jim *knew* what had happened. No one had watched him dropping the board or hiding it. His foreman was informed by John 137 himself, who broadcasted the information for everyone. John 137 went home and felt terrible. Not only his foreman will know any of his future mistakes, but every other robot, too. On the next day he was very nervous during his whole shift which resulted in another broken part. This time it was only an overheated main wiring harness, but everyone would know it. Later John 137 heard one of the other factory robots whisper “John wrecking ball” as he passed a group of robot workers. John 137 felt terrible but this was only the beginning. Within the next weeks, John 137 grew more and more nervous and broke several other things. Every robot in the factory started to call or at least think of him as John the wrecking ball. Some of the bigger robots even started to bully him and announced that the materials he destroys exceed the value of the goods he produces. John 137 decided that he had do to something and started to save as much money as his low factory robot salary allowed. He stored his money carefully in a metal case with a secure combination lock in his locker. After a period of 4 month, John 137 almost had enough money to by the necessary upgrade that will turn his omnipresent into a selective swarm intelligence. At the end of his shift John 137 approached his locker to find it open, as well as the metal case with the combination lock that no longer contained any money. Since the lock was undamaged, John 137 knew that he must have transferred the knowledge to all other robots and someone seized the opportunity. And so he decided to kill himself to get out of his misery . In his last moments he was wondering why no one came to change his mind because every other robot must have been aware of his intention.

4 Conclusion

Three examples portray the coexistence of robots and humans in a future society from the point of view of different autonomous social robots. Enacting laws that regulate the robots seems a direct and straightforward task at the first glance. Simple behavior patterns are sufficient to integrate the robots into our lives: Do never harm the humans and always follow their orders and so on. A closer inspection reveals that robots are intended to be installed in nursing and elderly care or as nannies for children. At this points the situation becomes critical and a multitude of questions arises and a careful examination is urgently required. Considering elderly care, does the retired grandfather not possess the right to interact and talk to someone equal? Will he be forced to talk to mindless machines that are only able to answer basic questions concerning the weather or their battery status? Does anyone really want to have his children raised by an emotionless metal container? In order to circumvent such scenarios one approach could be that robots require an own consciousness and their own goals and wishes. If the solution is to create them after our paragon, what distinguishes them from us except genesis?

In my opinion it is one thing to create a robot for a certain purpose, e.g. a task that is too dangerous or infeasible for a human, like lifting a car. Security aspects can be ignored for those robots, as there are very simple and efficient ways to circumvent any incidents, e.g. by stopping all robot motion and behavior if a human is too close to the robot. Even for the most complex factory or outdoor scenarios, there are ways such robot working machines can be integrated easily into society. As complex and manifold the purpose and the related scenario might become, there is no real problem until human beings get involved. But this exactly the case in application domains like elderly care for example. In such scenarios, the complexity of the task or the requirements for the robot are no longer in the focus, but how a human feels when such a scenario is invaded by a robot. Humans in the areas might become offended and feel pushed aside from society, towards the soulless automaton that accompany them to their grave. Approaches towards making robots more humane need to be considered for future scenarios, but they should never be realized hastily and without most careful scrutiny. One should always be aware of the side effects and what really happens when we make more of the robot as it is supposed to be - a worker in the sense of a tool.

References

[Asi42] I. Asimov. Runaround. Street & Smith, 1942.

[Dar59] C. Darwin. *On the Origin of the Species by Means of Natural Selection: Or, The Preservation of Favoured Races in the Struggle for Life*. John Murray, 1859.

[Dyn13] Boston Dynamics. LS3 - Legged Squad Support Systems for Rough Terrain. http://www.bostondynamics.com/robot_ls3.html, April 2013.

[IsEG13] Photo by Makoto Ishida IEEE spectrum: E. Guizzo. Robotics/Humanoid Robots: Hiroshi Ishiguro: The Man Who Made a Copy of Himself.

http://spectrum.ieee.org/robotics/humanoids/hiroshi-ishiguro-the-man-who-made-a-copy-of-himself, April 2013.

[LHA12] L. Lundström, H. Hamrell, and L. Akin. Äkta människor. [DVD], 2012.

[oAIST13] National Institute of Advanced Industrial Science and Technology. Successful Development of a Robot with Appearance and Performance Similar to Humans - For the entertainment industry. http://www.aist.go.jp/aist_e/latest_research/2009/20090513/20090513.html, April 2013.

[oCS13] Yale University: Department of Computer Science. Yale Social Robotics Lab. http://www.scazlab.com, April 2013.

[oFDoCS13] University of Freiburg: Department of Computer Science. Social Robotics Laboratory. http://srl.informatik.uni-freiburg.de, April 2013.

[PGVA04] A. Proyas, A. Goldsman, J. Vintar, and I. Asimov. I, Robot. [DVD], 2004.

[SA01] S. Spielberg and B. Aldiss. Artificial Intelligence: AI. [DVD], 2001.

[Spe64] H. Spencer. *The principles of biology*. Williams and Norgate, 1864.