

Ethical Acts in Robotics

Bernhard Irrgang
 Department of Philosophy of Technology
 Institute for Philosophy
 Dresden University of Technology, Germany

[This paper was translated by Arun Kumar Tripathi, Dresden University of Technology, Germany/]

I. Introduction

To consider the question, *can Computers/Robots act morally*, one must first distinguish between the participant perspective (subjective experience: first person perspective) and the observer perspective (objective experience: third person perspective) in phenomenological and hermeneutic terms. From the observer (i.e., third person) perspective, Ricoeur's approach of modelling acts (actions) without the subjective action is possible. Therefore it seems the *Automatic Machines* can act on their own without moral values being assigned to them. Paul Ricoeur thinks about the dialectic of a self-identity (*Ipse-Identity*) and about the same-identity (*Idem-Identity*) (Ricoeur, 1990, p. 14). Based on the philosophy of René Descartes ("I think, therefore, I am") Ricoeur refers to the two aspects of personal identity, where these two aspects refer to "I think - I am". This is all that cannot be doubted, according to Descartes. And, the self-executable thoughts in thinking and execution, is a category of existence and a category of thought to be parallel to each other. Thus, the starting point is the question of execution of Thoughts (*Denken*). According to the failure of a materialistic anthropology we can determine human-beings (Menschen) only by their execution of Thoughts, which is in thereby a execution of "I" <Ich> at the same time. The execution of Thoughts can be referred to an object of studies on the one hand, and on the other hand it can be reflexive. This defines the first tension in the context of the interpretation of execution of Thoughts. Second, it refers to the execution of "I" and it reconstructs the execution of "I think" in the context of a Hermeneutics "I-Self" and "Same" (Ricoeur, 1990, p. 29). According to Ricoeur, selfhood implies otherness to such an extent that selfhood and otherness cannot be separated. The self implies a relation between the same and the other. This dialectic of the self and other contradicts Descartes' cogito ("I think, therefore I am"), which posits a subject in the first person (an "I," or an ego) without reference to an other. Thus, the hermeneutics of the self differs from the philosophy of the cogito.

In the context of philosophy of language, Ricoeur reconstructs the person and the identifying designations (Ricoeur, 1990, p. 39). It concerns the subject of the act of speech (Ricoeur, 1990, p. 60). For robotics, the concept of Ricoeur is interesting and be applicable to an action without an “acting I” (*handelndes Ich*). (Ricoeur, 1990, p. 73). Following Ricoeurian concepts, one must ask the question, whether a robot can act on its own. Or, can the actions of robots be evaluated? In the context of possible behaviours of a robot, probably it does not do to deny that, generally speaking, robots can completely fulfill a pattern of action (scheme of action). In the most general case, an action pattern (action-scheme) is present, if at least the intentionality and the goal of action are preliminarily modelled.

The <Self> and <I> must be differentiated, since the self description is the category of reflection. "The Self (Ipse)" and "The Same (Idem)" is also to differentiate as the actual description for an Identity. Philosophy of subject or philosophy of "cogito = sum" are to substituted/replaced with the hermeneutics of self. In the philosophy of history, the phases of overestimation and underestimation of the subject is followed and studied. In a hermeneutics of the self, the concern is with the determination of self -- i.e., the subject of spoken, action, narration and of the moral attribute. Selfness and subjectivity can be understood as a dimension of human action and the unity of self. Self testifying and attributes are here inter-linked with others. The person is to be understood as a fundamental single thing (individual).

Now, are there human actions in Computer Science or in Robotics? The computer extends the range of human action (the activity of a human being). First, programming is to defined as an action within computer science. Computer science gives special attention to the complexity of problems. This becomes a pure action in the case of manufacturing processes (Kastendiek, 2003, pp. 157-159). Programming can also be understood as a medium to provide knowledge and technology (mediation of knowledge and technology). Programming is a practical activity, which is developed in the computer program (*Informatik*). It is based on the mediation of theoretical and mathematical aspects of the history of computer science, which is in turn based on the practical solution of human and social problems. Thus it is not a scientific method, but it can be defined as an action (Kastendiek 2003, pp. 161-163). Yet there authors, who attribute the character of action to computers themselves.

Some are inclined to attribute functionalistic & operational parameters of responsibility to computers. Computers are made responsible for their decisions and therefore it is necessary to

make them more trustworthy. Computers are neither moral nor social beings, although their common employment can produce extensive social effects. Today, ever more decisions in a society are left to individual information systems, because human-beings are not able to make appropriate and irresponsible decisions within extremely short response times. Therefore, these decisions are generally no longer made by individual human beings. Klaus Haefner realizes the first step in the actual inevitability of the transmission of responsibility attributed to computers, which is an integration of human beings into a very complex overall system. It is increasingly responsible for human beings. Indeed, one must differentiate between the descriptive analysis of causality on the one hand and the normative expression of responsibility on the other hand, in the sense of the someone-real-responsible-person to be made (*jemanden-wirklich-verantwortlich-machen*) (Lenk, 1990, pp.104-106).

William Bechtel and John Snapper are working on this concept. They specify quite reliable conditions for flexible program-controlled decision systems, which they believe permit them to provide legal and moral responsibilities to these systems. It was at first unclear who should bear responsibility for computer errors: operator, owner, manufacturer, programmer, software producer, etc. (Lenk, 1990, p. 107). Who is going to take the responsibility for errors in the system? Snapper and Bechtel use concept of *Responsibility* only in the sense of a descriptive-attribute for defining the problem and giving solutions. The normative expression in the sense of some-responsible-person-in-the-making or someone-to-be-held-responsible is not discussed (Lenk, 1990, p. 111).

This is the underlying mission: In the long run, the new systems, robots or highly networked systems, have the characteristics to be ever able to "act" more autonomously, to be able to unreel without being able to reach "autonomy" in an ethical sense. To untether the behaviour of programs in the long run will also in certain frameworks and certain necessary decisions, could have serious consequences for human beings. Programs could lose their character as tools and instead become autonomously active participants. In the robots existing today, humans must still intervene by controlling, steering, supervising, etc., but the degree of automization is growing. Thus, it raises the responsibility question for the behaviour of robots and AI expert systems, as well as for that of their technical designers and users. Early discussions in this vein were limited to the topic of artificial intelligence (AI), but today's debate has been extended to such topics as data processing and networking. Thus, the computer can still be seen as a tool, but also seen as an autonomous machine. The specific

kind of action determines which characteristics lie in the computer. The scope of action is to be registered regarding the level of action with acting computers, and thus the computer can be regarded as an all-purpose tool or a universal tool (Kastendiek, 2003, pp. 9-11).

In addition, the thought (*Gedanke*) of a machine-similar action should be introduced. First, it is necessary to differentiate between action and behaviour. Similar and repetitive actions can easily be made by machines, whereas humans require a lot of training and substantial efforts of willpower in order to implement actions in this way (consider, for instance, military drill). Humans need a long time for such things (again, consider a machine-similar drill practice). The thoughtless routine must be internalized and constantly repeated. Afterwards, the machine actions and human actions are determined. Nevertheless, actions which do not follow explicit rules may later still be summed-up (*subsummierbar*) under a rule.

Retrospectively formulated rules do not apply, however, to the future. How can one split and divide behaviour? That is one of the central problems of methodical kind in the description of documents and behaviour (Collins, 1991, pp. 30-45).

Thereby, it gives a clear picture of an increasing authority and competence of machines. If humans, for example, feel too tired to implement routine tasks, they could simply delegate the whole process to their machines. What is of concern is the extent to which can delegate their actions to machines. In other words, we cannot delegate actions -- rather we can only the delegate the behavioural co-ordinations of the act. We can delegate the action successfully only up to this degree, which is the part of an action and in sufficient measure are suitable to be reduced to a specific behaviour, that are recordable without the loss of a formula. And so we can say that dancing steps coincide with the steps of many participants, if we describe a physical activity in terms of the production of standardized artifacts or mental activities (Collins, 1991, pp. 65-71).

II. Perspectives of Ethical Acts (Action)

Computers are the part of our culture. We have expectations that those computers think like machines. In addition, we are inclined to attribute intelligence to them. All these things were conceived in recent time. But what is decisive is the machine-machine interface. The Turing test concerns the man-machine interface, takes only 20 seconds, but that is not sufficient time for a genuine evaluation of machine. Machines with an experimenter can be trained (Collins, 1991, pp. 187-194). A programmer can never anticipate what an interrogator may say or ask

(Collins 1991, p. 203). The power of a computer can be only understood in the context of the social groups to which it is assigned, and human beings then compensate for the defects from artifacts. What can cause or teach and carry out an expert system for a social group? Nothing can be completely described; otherwise, machines could take over the competence-oriented jobs of workers? Well, this can apply, but this applies alone to behaviour-specific competence. The only error is to believe that a tool is an actuator (*Akteur*) -- i.e., an action. There is a substantial difference between the way in which humans act and the way in which machines copy a document (action). Intelligent machines are, however, the most useful and interesting tools (Collins, 1991, pp. 215-224).

It is to be considered that robots can have the shape of humans and thus be able to simulate humans; this is hardly the case now, but it will be the case in the near future that their abilities will be similar to humans. They need the correct environmental model, thus they are not environmentally open (friendly) like humans; and they need a constant control and maintenance and they need stable environment conditions, in order to do the correct behaviour for the day. Probably, the instinct-secured behaviour of some animals is more autonomous and more similar to humans than talented robots in the foreseeable future. So far automats with only very limited functionality can be realized, because they lie within a range of pre-rational intelligence, within which certain human competence is to be simulated. The robots of future will be copied still more from their models, humans or animals. And in robots the artificial and biological elements will unite, since computers, especially networked computers, are disembodied Artificial Intelligence. The realization of this intelligence in a body-free Cyberworld (a sort of disembodied world) is a possibility on one side, and the realization in robots is on the other. But we will still have to wait and see which way will be best for AI development.

For artificial intelligence to become true intelligence, it must become artificial soul and not be limited to a pure mind or pure cognition. In addition, it's necessary to understand that feeling and motivation actually contribute to the increase of intelligence to a considerable degree and that they are possible in the computer (Doerner/Spitzer, 2002, p. 24f). Doerner understands the imbedding of the human action regulation, and also the imbedding of the cognitive processes by feelings into a structure about modulation and action tendencies -- which are for its part situation-oriented. For it is important that the brain puts on, sequentially, memory minutes of our activities. But memory brings with it the basis for self

modification. The free Will is the released Will, or the released motive. Man will build artificial souls, because we are making an attempt to copy the psychological processes to give us much greater understanding. Human beings should know substantially better, and that is so far the case (Doerner/Spitzer, 2002, pp. 33-36); whether, however, a human soul could be designed independently of a human body in a form other than an extremely rough model of a soul, I think is very doubtful. Besides, it would have to be clarified as to which psychological characteristics are to be given to such a soul.

The desire to humans can represent a specific perspective for robotics. It is objected that the communicative competence of humans is not in principle technically replaceable. Technology proceeds right from the outset, where it substitutes for human activities, procedurally differently than nature (Decker, 1999, p. 23). Cognitive robotics is for engineers, at the beginning of technical conversion of a new body-soul theory -- functionalism (Decker, 1999, p. 37). But it ignores the role of humans, who can not be simulated. In robotics, the focus is increasingly on functional coupling between nerve cells and technical systems in the form of neuro silicon couplings into cell cultures (Decker, 1999, p. 41). In order to be able to take responsibility, technical systems must be controllable by humans. By a complementary system, organization work quality -- in the double sense of competence-promoting tasks and effective achievement -- contribution becomes possible. Occasionally, one must consult the collectiveness of beneficiaries, in order to compensate the damage to those the technology hurt by malfunctioning (Decker, 1999, p. 65). This refers both to producer and consumer.

Many perturbations in the 20th century were in revolt against the privileged position of certain humans. Then machines appeared as a concurrent for the privileged position of humans. Humans are not any different than machines in this respect. Computers are today, in symbolic algebra, better than humans. Computers with the correct software actually think about facts, make decisions and have goals. The task is to build robots that act as if they seem to have fear, or simulate the fear. We draw a limited comparison with our feelings: in the animal realm we are still safe in our place, and so we confess some feelings to animals (Brooks, 2002, pp. 172-189). It is similar for machines, and the machine operates in the same way as humans. John Searle subordinated human brains, and nerve cells somewhat particularly, to consciousness, but he nowhere give a reference to what specifically makes consciousness. He also does not also say why a system based on silicon cannot have consciousness. Artificial Life (AL) and the artificial evolution create whole new thought categories. Models of neural networks still need to be developed, and the question of

consciousness is a very difficult question indeed. The idea of robot as our slaves is a beginning of morally justified slavery. And so, if robots are to have the status of humans, then this form of slavery must also be eliminated (Brooks, 2002, pp. 192-215).

In 20 years our personal computers will be 1000 times more efficient than today and will then exceed our intellectual capacity. This is the forecast of Rodney Brooks. Perhaps we can build a computer that is more intelligent than we are, and perhaps not. Do robots then represent a danger for humans? Is it in the paradigm of condemnation or release by robots to be arranged? The resemblance of robots to humans could be limited. Some basic assumptions of robotics should be discussed, in order to focus tightly on their limitations: (1) the question of their independent reproduction: Machines that repair themselves and reproduce themselves would nonetheless depend on an infrastructure. Humans are worried about these developments, because the supply of materials and the production and maintenance of computers remains the task of humans. This infrastructure is still much too complicated for robots. Also, in chip production over the next 25 years much human participation will continue to be necessary. (2) robots do not have feelings, but sentimental robots will be designed. (3) the survival instinct of robots; robots that guarantee their own power supply are not conceivable in foreseeable time. For a long time robots will be dependent mainly on electricity, which is operated by humans. (4) loss of control of humans; Such development does not come overnight and not rapidly. We will notice it only if we begin to build machines that we can no longer control.

The three disappointments of human narcissism, Copernicus, Darwin, and Freud, were the milestones on the way to appreciating the finiteness of the human mind and its embodiment (*bodily existence*). AI could be in the next stage, and could teach us to read a new interpretation of the hermeneutics of anthropology. A condition for it, however, is that we learn from the visions we have been given -- to learn, for example, from the discovery of Hans Moravec's illusions. In billions of years of untiring arms races, our genes succeeded finally in putting itself out for rebooting. And so we prepare a post-biological, supernatural future for ourselves. In this future world, the new humans will be torn between an eagerness for future changes and a regret for displacing their own artificial descendants. Today our machines are still simple creatures, which require (like all newborn babies and children) the parental care and welfare service, and are hardly as intelligent as they are thought to be. Since these children of our spirit are not dependent on the failure of the biological evolution, they

will unfold unrestrained. Thus, it comes to an alert redemption between the Intelligences. Since the industrial revolution, it comes down to replacing the human bodily functions by machines. The time is already very close for us to understand that each important physical and mental function of humans will have its artificial counterpart. The primitive evolution began with reproduction of clay/tone crystals. Our genes and body from flesh and blood will soon lose their meaning, and it is not difficult to imagine human thinking, free from the connection to a mortal body. Hans Moravec pleads in such a way for a post-biological world (Moravec, 1990, pp. 9-14).

III. Autonomous Robots

Robots that can act "autonomously" -- or more precisely said, that can behave in many situations somewhat similar to animals -- continue to be technical products. They remain tools, and do not become doer (acting) subjects. How should they become subjects? (How could one attribute subject to the machines? Only by a miracle!) Robot-action belonged to the cases of actions without a doer (acting) subject. Since the industrial revolution, an illustration of technical action patterns in machines are to be seen. Here it develops with the reproduction of the spin procedure in machines and with the linkage of weaving looms provided the first integrated machine illustration of more complex human technical action. A modelling and simulation of human competence that includes the use of application tools is not somewhat new in principle. (Which indeed could be possible in the future by robots.) It is an extension of the basic pattern based on non-technical courses of action such as perception, mobility etc. It can also be used for non-technical purposes. Thus a further mechanization of everyday life will take place. But in principle, this is also not new.

A robot does not mutate to the acting (*handelnden*) subject. The acting subject with a person or a personality is not the product of a genetic code or a computer program, even if many reductionistic science programs believe in this project, and system theory is not the least force that wants us to believe this. Now it is in principle conceivable that robots are further developed than we appreciate, and are able with their own energy balance to be at least partly equipped to act universally and multi-functionally -- also, capable of completing several tasks to be programmed, so that by the interaction of everything important it possibly appears to some humans that an autonomously acting subject might be act here. The question is, howeverm actually open: in what respect will programming be like that something is present such as the consciousness of the internal actions of the machine. We may not subordinate

these simply because the action forms of such a computer are to simulate human behaviour from a third person perspective. The aspect of execution of a human agent or a human action pattern cannot be simulated in a machine, because the execution perspective concerns the first person perspective. To that extent possibly a comparability is here, given in the external aspect of human behaviour and in the robot behaviour by a similarity of courses of action.

Of course we can attribute a status to an animal or a robot exactly in the same manner as humans award a certain status to others in a certain role. But if we award dignity in an almost indiscriminate way, in attributing the moral sense to robots, animals and humans, then the concept of dignity devaluates itself in a dramatic way. It cannot be consulted then around a certain privileged position of humans. However if we do not want to maintain this kind of difference any longer, then in principle the totalitarian regime of torture methods is finally justified. To that extent, it appears necessary not to follow those efforts, to aim at a kind of posthuman condition -- but instead to develop a suitable anthropology that can be established on the basis of an understanding that is not against the modern realizations for animals and computer action patterns, i.e. on the basis of an understanding that recognizes the cognitive structures. I assume that the simulation of cognitive structures does not fit by any means automatically to knowledge, just as few of the simulations from actions to humans. It is morally assignable caused (*zurechenbaren*) action. Without this being strictly provable, it yields pragmatically ground of evidence of why human action and human morality remains embodied and are mediated or remain intact in an embodied mind. To that extent, probably an isolated brain cannot act morally in a technical matrix. Here, however, a line of limitation is present, on which one must contemplate.

A concrete observation: "autonomous" robots carry an actual potential of danger that can arise from them: for example, perhaps household robots do not give proper consideration to babies who are in the way. Also, different accidents, e.g. on collision courses of household robots, are possibly not an acceptable risk. On the other side, it could be the case that these dangers that we usually design and build in human-like household robots are good for babies. There are possibly other special technical solutions for special uses, through which robots not like humans go shopping. On the other side, there are efforts to build a universally acting robot which could become indeed human-like. But why should we actually build such universal robots? At least from today's perspective, it appears meaningful to design robots, computers, or expert systems that can fulfill certain specialized tasks and also can also make

use of them. In such specialized functions, robots are the tools that remain dependent on humans in a certain way, and belong to the area of responsibility of humans, just like traditional technological applications.

Why could it be meaningful to build a universally applicable robot if there are humans who can fulfill such tasks? Perhaps in space, where robots can work better together with other robots, because in space human life is endangered. One will have to contemplate such areas of application, and need a knowledge to work out the hazard potential from limited working robots within the specialized ranges. Afterwards estimations and considerations must be made whether we finally want to design a universally applicable robot and possibly generate a new technology that will dismiss the human control. From an ethical perspective, a technology that is completely out of control and can cause heavy damage, should not be built, because this is the limit of a potential of responsibility of assumption of capacity of each human, which can also be valid for the collective responsibility of humans.

In all other respects, the question still exists: why must artificial humans and robots be absolutely human-like in order to be useful. Perhaps it is the fascinating dream of humans (*Faszinosum Mensch*), which energizes us to convert such Thought experiments into practice, in order to find the technical models of ourself, which will give an understanding of who we finally are. But we can achieve success on this by technical artifacts and technical models. I also assume that such universal and unspecified robots with great flexibility in the behaviour patterns and a due portion at learning competence, as well as a certain humanlikeness, nevertheless do not finally become humans, which are after all equipped with knowledge in a theoretically philosophical sense or with action competence in a moral sense. Even if they are possibly a useful simulation for the everyday life of human beings. The potential of humans is more comprehensive than that of a common-sensical understanding of the technologized everyday life. Also, there is the possibility that if robots' action patterns as quasi autonomous process cycles are programmed to independently learn in a certain way, by no means is it the case that these robots can also reach a consciousness of their actions and execution of their action patterns. We do not subordinate consciousness to a computer machine in order to examine the arithmetic operations contentwise. Also, a text processing system probably does not understand anything about the texts that it processes.

But even if the complexity of programming increases extremely and machines learn, how could it become possible that an arithmetic performance capacity (which obviously produces no consciousness) can then eventually develop consciousness -- or even self-confidence? Perhaps we can once model humans as a programmable blueprint to sketch for a robot. But if we actually let humans themselves be simulated (and not a deterministic animal in a program), why should this simulation be regarded as the proper one? Liberty or freedom can be simulated in a random number generator (*Zufallsgenerator*), but human liberty or freedom can't be simulated. The model or the copy of an object must be from the same material, and have the same form. We nevertheless usually estimate and hold the original high. But as a copied painting and the original painting have something obvious in common, robots -- due to their material difference and organizational difference with humans -- will not be able to reach the identity.

IV. Decision Making

In a Thought experiment (*Gedankenexperiment*) one can equip humans gradually with more *Exoprothesen*. In the same way, after a complete replacement of all organs, men could be said to become Androids (Erlach, 2000, 36)? But here Erlach errs; if humans are supplemented by prostheses, replacement artificial organs, and cosmetic surgery, they do not become artificial humans, Androids or robots. But eventually the overall human system will break down. The Thought experiment leads to an interesting result: that there is theoretically a transition field between humans and robots -- yet what is relevant however is the question of the extent to which this process is technically feasible. The success of this process does not depend only on technological fate, but particularly on the maintenance of a vital function of humans, with the replacement of parts. The Thought experiment is thus limited in an attempt of its execution (*Durchführung*).

Without acting I (*handelndes Ich*), the goal structure of action pattern can also be ethically evaluated. If a goal structure can be reconstructed, an action is present, also if no acting I (*handelndes Ich*) is to be identified I can judge whether a robot carries out a morally positive or not-positive action pattern. However, a non-embodied (*disembodied*) (non-human) action is only in the abstract sense an action, for whose consequences are also attributable to the robot and held responsible -- yet possibly to the programmer who sketched the action pattern and implemented it in robotactio. Action is itself an Interpretation construct (construction of interpretation) and is defined as a write-up phenomenon. Pure physical procedures do not

differentiate the specified actions of events, and so positivists conclude that they also do not provide the moral actions and human liberty (freedom). But our experience reflects otherwise (Irrgang, 2005).

The psychological & thoughtful (*denkerische*) identity, based on the continuous execution of thoughts (*Denken*) (materially spoken, non-stoppable brain processes), constitutes that embodied execution of beginning of the life of the brain up to the death of the brain. In order to be morally responsible, however, an act needs a participant, who is characterized by personality or subjectivity. Robots can proceed according to a certain action pattern (goal structure and consequence of evaluation), which can also be evaluated morally. The robot cannot be pulled to account, since it was not the actual participant, but it was a generator, that is a programmer, of the action patterns. What can now identity and selfness of an embodied execution mean? The crucial and decisive dimension of reply is based on the fact that the wholeness of an execution remains protected. The wholeness of an execution however can be judged in the long run only in a retrospect by its conclusion and it is completely evaluated. This wholeness of an embodied execution begins with humans sometime between conceptual design (*Konzeption*) and birth and ends with the death of the brain. With a robot, the technical execution begins as process without participation of subjectivity or personality by others. The engineer brings a robot into the enterprise and sets him out of operation; in contrast to the embodied execution, the technical enterprise of a robot can also be interrupted.

V. Concluding Remarks

Robots do not become humans, and cloned humans or genetically redesigned humans will remain humans. To that extent, a certain naturalness of human embodiment (*physicality*) is a central issue for establishing an anthropology of *homo faber* in the 21st century. The difference between humans and robots can not be erased. A Cyborg is defined as a cybernetic organism and is ultimately a robot or artificial human. Humans with artificial hearts remain humans, also ones with heart tissue. Humans, even with genetically improved code, cannot mutate and become a robot. Thus, the difference between “natural” and “artificial” is preserved within the area of anthropology despite all sharpening (*Verschleifung*). The Kaleidoskope resolves the clear differences between naturally produced humans and technically designed robots, but it does not remove them completely.

The human beings who are reconstructed almost totally after a fatal accident and equipped with prostheses do not become a Robot: he/she remains, a human being when he/she survives. A human can be supplemented prosthetically, but he/she can't survive his/her brain death, based on our current knowledge. One could use parts of a corpse (*dead bodies*) theoretically for the production of robots. That doing so would let them remind themselves of events from the life of a dead one, is highly improbable. The designer baby likewise becomes no Cyborg in the sense of a robot, even if the most general definition of Cyborgs permits this interpretation. I would thus suggest that we clearly differentiate the concept of Cyborgs from that of Robots. Nevertheless, we can also make the reverse Thought Experiment (*Gedankenexperiment*) and try to build a computer into humans! This will not be technically possible, because probably a brain is not able to control a robot. Of course one can build and transplant some chips, but the limitation between humans and robot perhaps falls into category of fantasy -- not however within the range of what is technically realizable. There are limitations between morally attributable (*zurechenbaren*) acts (first person perspective) and machine regulated, implemented behavioural patterns which are capable of being implemented (third person perspective). It is an amazing fact, which fairy-tales researchers are ready to believe, that this difference is of concern for building artificial humans.

The identification of a robot with a natural-born cyborg has disastrous consequences: Either genetically designed humans as well robots do not have human dignity, or however human rights must be awarded to robots. And all this can only be accomplished due to a concept of weakness (*Begriffsschwäche*) and a concept of reflection. Also the thought that between genetically adapted humans, who are wrongly called Cyborgs and robots, in them some kind of sex or even love could be possible, which is nevertheless strains our fantasy to a great extent. Here, the film "Blade Runner" conveys the correct answers in the form of literary mediation. In addition, technologically reprogrammed humans become designer human beings and humans with tissues technology produced for organs for Man-Machine hybrids or natural born Cyborgs, but they cannot be robots.

The difference between humans and robots will remain at least as long as humans are thought to be nourished by technical (in particular electrical) energy. Extreme forms of man-machine hybrids or human brains in a nutritive solution, i.e. in a technical matrix remain for technical -- like ethical -- reasons ultimately Science Fiction. A brain in a nutritive solution i.e. in any matrix would be actually posthuman. It would be only in the most reduced way

spirit, because this brain was at least bound to a body. This brain is conceivably the most extreme form of a Cyborg, but it does not however become a robot. Probably such a brain also still has memories of the embodiment (*physicality*) and a subjectivity developed in former times. A conception of subjectivity, which may develop a brain in nutritive solution, is to be apostrophized from the perspective of the human brain today living as embodied mind as strange and posthuman. A human brain in a nutritive solution or a matrix is not human, but remains a brain in a nutritive solution. Only if such a brain were able to control a machine the question about the morality of the courses of motion i.e. actions of this hybrid thing arises. It is done in such a way, when whether any replacement (*ersetzbarkeit*) of physical (*corporeal*) functions would be possible with man-machine hybrids by technical surrogation (*Surrogate*). This corresponds to the logic of technical expansion, but probably not the logic of human embodiment (*Bodily existence*) (Irrgang, 2005).

The Science Fiction of the future of humans usually comprises of negative utopias of the posthuman. Posthumanity is growing on the basis of a wrong picture of humans and a wrong view about technology. On the other hand, only philosophical reflection helps one to understand the clarification. The negative utopias proceed from the failure of humans, but the Humanism tradition wins in the technical thinking of strength. Therefore, the discussion over posthumanity is probably not much more than the last rearing up of a declining cartesian dualism. *Homo faber artefactus*, the autonomously acting robot, is not by any means the reality, and there is not even silver-lining on the horizon. But it casts its own shadows (shade) ahead. It is time to speculate about the century, over an ostensibly and supposedly posthuman human-being (*Bodily existence; Menschsein*) in its way of technical existence and its Technosphere about an anthropology of technology in the 21st century, that advances with the technologization of humans and its subfunctions with an unforeseen speed.

References:

- Brooks, R. (2002) *Menschenmaschinen. Wie uns die Zukunftstechnologien neu erschaffen; übersetzt aus dem Amerikanischen von A. Simon*. Frankfurt.
- Collins, Harry M. (1991) *Artificial Experts. Social Knowledge and Intelligent Machines*. Cambridge Mass./London (1990).
- Decker, Michael (1999) (Ed.) *Robotik. Einführung in eine interdisziplinäre Diskussion* Bad Neuenahr Ahrweiler.
- Dörner, Dietrich, Manfred Spitzer (2002) *Vernunft – Gehirn – Computer: Was bleibt vom Menschen?* Bamberg
- Erlach, K. (2000) *Das Technotop. Die technologische Konstruktion der Wirklichkeit* Münster Hamburg London.

Irrgang B. (2005) *Posthumanes Menschsein? Künstliche Intelligenz, Cyberspace, Roboter, Cyborgs und Designer-Menschen - Anthropologie des künstlichen Menschen im 21. Jahrhundert* Steiner Verlag.

Kastendiek, Antonia (2003) *Computer und Ethik statt Computerethik* Münster.

Lenk, Hans (1990) Information zwischen Verantwortung und humaner Eigentätigkeit. Chancen und Probleme der neuen Informationstechnologien für Gesellschaft, Moral und Ausbildung; in: B. Irrgang, J. Klawitter (Hg.): *Künstliche Intelligenz; Stuttgart, 99-115.*

Ricoeur, Paul. (1990) *Soi-meme comme un autre* Paris (Ricoeur, Paul. *Oneself as Another. Translated by Kathleen Blamey. Chicago: The University of Chicago Press, 1992.*).

Author biography

Dr. Bernhard Irrgang (professor for philosophy of technology at Technological University Dresden) teaches courses in the philosophy of technology and bio – and hermeneutic ethics in the Institute for Philosophy at Dresden University of Technology Germany. He has expertise in different fields of philosophy of technology, especially pertaining to the philosophical issues to technology, gene technologies, cultural theory of techniques, information technology, artificial intelligence and expert systems, technological assessments, and medical ethics. Main emphases are on: Philosophical questions and topics concerning technology; Technology transfer as a Cultural transfer; Technological development and early technological cultures; Technoscience research; STS research (Science, Technology and Society studies).

Note

An earlier version of this paper was published in the Conference Proceedings of the Sixth International Conference of Computer Ethics: Philosophical Inquiry (CEPE2005, July 17-19, 2005, Enschede, The Netherlands) eds. Philip Brey, Frances Grodzinsky and Lucas Introna, CEPTES, Enschede, pp. 241-250.

Source: Ubiquity -- Volume 7, Issue 34
(September 5, 2006 - September 11, 2006)

<<http://www.acm.org/ubiquity>>