Name: Patrick Ehrenbrink
Group: Achim Stefan

"*Semantic Engines: An Introduction to Mind Design*" – John Haugeland

<u>Task:</u> Why is the idea that minds are digital computers so appealing and what seem to be its most serious limitations?

<u>Introduction:</u> In his text *"Semantic Engines : An Introduction to Mind Design", written in 1981,* John Haugeland compares the human brain with digital computers. At that time, the existing models (*Dualism, Homunculus and Behaviorism*) which were used to figure out how the brain works were not able to satisfy him, nore seem they fit in the relatively new field of cognitive science. I will exlpain the reasons further, because I think it is necessary for pointing out the innovations of his digital computer theory.

He explains the basic idea of cognitive Science with the attempt to understand intelligent beeings which are semantic engines or automatic formal systems and should be copyable by a computer with an appropriate program.

<u>Conventional theories:</u> Haugeland displeases the **Dualism** because of its lack of compatibility of mind and matter. It fails to explain how this two majour parts could possibly interact and therefore, it is not sufficient for people who work in the field of cognitive science and try to develop intelligent systems or at least systems which appear intelligent.

The **Homunculus-Theory** is also not convincing for him because it assumes that all thought are directed by another intelligence. Obviously this theory can not lead to a philosopic progress. It raises the question it trys to explain. The result would be an infinite loop because the intelligence which leads the thoughts has to have another homunculus for itself.

At least the **Behaviourism**: This theory tries to determine and explain the behaviour of an intelligent system exclusive by external observation of the object. Haugeland thinks that this theory is practically too hard to proof.

<u>Humans are digital computers:</u> As an alternative theory for the three ones mentioned above, Haugeland initiates his theory which supposes every intelligent beeing as a digital computer. I think the base of this idea is that he thinks of an intelligent beeing just as a digital computer and an automatic formal system.

He mentions that no philosopher has a problem to accept a chess playing computer with a program and a processor which follows entirely the laws of physics, but many philosophers do not accept the theory that a chess playing human beeing could be function likewise this chess computer. Whithout violating the laws of physics.

Indeed, the model of the digital computer seems very pleasant looking back to the three theories explained before. I am also very delighted because this theory legetimizes the research

of artificial intelligence and cognitive science in general in a very strong way.

<u>I got an X:</u> There is also a point which tarnishes the vision of an artificial intelligence which might co-exist with us in the future. Haugeland names this point "X".

"X" is in fact not a single point but three properties of intelligent beeings which no computer is able to have.

- **Consciousness** is one of these properties. No matter how fast computers will be in the future, they will not be able realy to understand things which are going on. And since not even humans seem to have an idea of what consciousness exactly is, it seems highly unlikely that we will be able to generate a software which will be able to imitate this.
- **Original intentionality** is also hard to develop because computers do not simply have or generate intentions, we give it to them. And because of this, their intentionality is at best artificial and appended.
- The last "X" is **caring**. Computers do not care about things because they do not have deeper connections to them nore do they have feelings. I think that apart from our disability to give these feelings to them, we have to ask ourselves if it would be sustainable on an ethical level to do that, because computers which would be able to care about things would also suffer under certain circumstances. For instance if they miss a given goal.

<u>Own opinion and final words:</u> The digital computer model is up to now the one which appeals to me most. As Haugeland explains it does not include the flaws of the theories which have been discussed so far and demonstrates possible parallels among mind and software and brain(also body) and hardware.

Even if I have to agree in the objection that computers will never have certain features of human (or other higher biological) beeings, I am very confident of further developments which will help to **converge** computers to real intelligence.

By the way, it seems not necessary to me to develop "real" artificial intelligence. With regard to ethic scruples which I have already mentioned, it is doubtful if we should build conscious machines. Not only because they may suffer from feelings, but also because they would surely become second grade members within our Society.