

“Can machines have Free Will? Analysis of the Concept of Free Will in Relation to the Psychophysical Problem” – summary

The goal of my thesis is to analyse the notion of free will from the perspectives of various stances in philosophy of mind. It employs an approach quite different than the more standard one, in which philosophers try to answer the question “do we have free will?” directly. Therefore, the first chapter of the thesis contains an analysis of how they do it with the main focus on the contemporary compatibilism/incompatibilism debate – the question “is free will compatible with determinism?” According to compatibilists there is no contradiction in assuming that free will and determinism can coexist, and, of course, incompatibilists think the opposite is true. I show that in the end both groups of philosophers end up talking about something different – for one of them free will is a thing compatible with determinism and for the other it is something incompatible with determinism and these are not the same thing. So when they answer the question they were supposed to answer – “is free will compatible with determinism?” – each of the groups answers a different one, because the term “free will” means something different to them. Therefore I conclude that it is not a valid approach to this problem and look for another way to grasp it. The idea of the thesis is to analyse the notion of free will in relation to the mind-body problem.

The subsequent three chapters contain discussions about the relation between free will and dualism, materialism and transcendental idealism. The considerations about free will in dualistic stances serve as a background for the final two chapters, which contain the most important conclusions of the thesis. According to dualistic stances there are two separate substances or kinds of properties – mental and physical. They interact with each other, and have some degree of independence, different for each particular dualistic stance. The most popular dualistic stance is Cartesian Dualism, according to which there are two separate substances – mind and body – and they interact with each other through the pineal gland. I argue that in this setting the question of whether human beings have free will is quite easy to answer, the main problem however is shifted to grasping the relation between the mind and the body.

In the third chapter, I discuss the relation between free will and materialism with a focus on whether it is sensible to talk about free will in the context of eliminative materialism. Eliminative materialism is a very controversial stance according to which propositional attitudes do not exist. Propositional attitudes are attitudes of an agent towards a proposition, for example “A believes that the Earth is round.” expresses an attitude of an agent A towards the proposition “the Earth is round.” According to eliminative materialists, thinking in terms of propositional attitudes is a part of “folk psychology” and will be replaced by mature neuroscience. To analyse what it could mean, I go more in depth into the analysis of how artificial neural networks process information. Given that neural networks are commonly employed by naturalistic philosophers as a model of the human

mind, the analysis of the notion of free will in this context should give a better understanding of whether it is a sensible approach to understanding the human mind. The conclusion is that eliminative materialism and other materialistic stances do not give any room for free will and any talk about it on materialistic ground is meaningless.

In the last chapter, devoted to transcendental idealism, I present a solution to the free will problem and Kant's third antinomy of reason that incorporates eliminative materialism on the empirical level. Kant's third antinomy of reason states that pure reason is in an apparent contradiction, by getting to the conclusion that the world is determined on one hand, and that human beings are free on the other. According to Kant it is possible to retain this view by considering a human being from two perspectives – phenomenal and noumenal. As a phenomenon, human beings are part of the nature and therefore fully determined. This is where eliminative materialism helps to explain human behaviour – when human beings are treated as temporal and spacial objects (so, as phenomena), there is no need to ascribe propositional attitudes to them. On the other hand, as noumena, human beings are capable of free decisions and actions which is crucial for morality.

The thesis ends up with a conclusion in which I answer the question stated in the title. Machines cannot have free will and if human beings are, in relevant ways, like machines, they do not have free will either.

Krzysztof Kwenc