



4

The Problem of Autonomy and Determination in the Digital World

Somewhat unsure, Neo stands in the doorway to the kitchen of the “Oracle”—an older cookie-baking lady, who cheerfully smokes one cigarette after another. She is expected to tell him if he is the “chosen one” or not, that is the one who will destroy the digital program of the “Matrix,” to which all humans are linked to mistaking it for real life. Destroying the Matrix would mean the end of the machines’ reign and the beginning of a real and self-determined life.

“And don’t worry about the vase,” she tells him shortly after he has crossed the threshold.

“What vase?” he asks and turns around, accidentally smashing a vase.

“How did you know?” Neo asks.

The Oracle looks at him, amused. “What really is going to bake your noodle later on is, would you still have broken it if I hadn’t said anything.”

He is not alone with this question. The spectator too wonders about that. Would Neo have broken the vase if the lady hadn’t said anything and he accordingly wouldn’t have turned around in the first place? After all the oracle is itself a program, whose top priority is guarding the Matrix. A strategic communication would thus be more than plausible.

The question that is going to bake our noodle now is if predicting the future is at least in principle possible. From a philosophical point of view, fortune telling has the following problem: If a fortune teller can predict future actions, it must be true that people behave just as Turing machines (machines named after Alan Turing), that is, behaving according to laws of determination where each state determines an exact after-state. If this was the case, then indeed all our behavior would be predictable. However, as our behavior is

dependent on our beliefs, a fortune teller would have to be capable of predicting these too. If this was the case, we could predict the knowledge of future societies—which is first of all incompatible with the idea of genuine progress of knowledge and secondly gives rise to logical problems that Karl Popper already pointed out in *On Clocks and Clouds*.¹ His argument was that if one assumes all future knowledge is predictable, this future knowledge would already need to be part of the knowledge today and would therefore not really exist as future knowledge. A true revolution in knowledge implies that future knowledge is not part of former knowledge. The belief in total determination would then come into logical conflict with this assumption.

In a scene from the second part of *The Matrix* franchise, the Oracle and Neo are sitting on a park bench inside the Matrix. Neo is unsure as to if he can trust the Oracle; after all, she as an AI is part of the system he seeks to destroy. As all AIs in the film, she too doesn't believe that humans possess freedom of will and freedom of action. She presupposes that humans too are guided by algorithms, determining each future state on the basis of the past one. At one point in the conversation, she draws his attention to a couple of birds picking up some grain on the floor.

"We are all here to do what we're all here to do [...] Look, see those birds? At some point a program was written to govern them, a program was written to watch over trees, the wind, sunrise and sunset. There are programs running all over the place."

The oracle is wrong, however. Humans do not behave like birds or software-controlled machines. Humans reflect on what they do. They are able to act according to reasons. This capability to make decisions based on good reasons is the essence of our human freedom and responsibility and distinguishes us from animals. If every action can already be predicted *before* any deliberation (be it only the probability of all possible actions), there wouldn't be any free and responsible agents. To be exact, there wouldn't be any agents at all. Instead of actions, there would only be behavior. The birds in the Matrix don't act; they just behave in a certain way.

It is not the only scene in which the protagonists debate the problem of free will. In a future scene, Neo and Morpheus pay a visit to the so-called "Merovingian," an AI system which is able to write programs itself. There, the problem of free will is brought up again. The scene is set in a luxurious restaurant. Chandeliers are hanging from the ceiling, there is soft music in the background, and beautiful women are sitting at elegant tables. The Merovingian is sitting at a table with his wife, Persephone. Neo and Morpheus who want to

¹ Popper (1996).

use the Merovingian in order to get to the central computer of the Matrix come up to his table to talk to him. But the Merovingian only smiles condescendingly at them. Humans may believe they are free but in his opinion, they are not capable of choosing their own goals. In his logic, there is no freedom of will. Humans, just like machines or animals, do what they have been programmed to do. They act how they are supposed to act. And if they don't, then it is because is an anomaly in their system.

Indeed, many software-controlled systems are run by probabilistic functions. These do not assign one state to a following state but use a probability distribution. Such probabilistic functions make for "self-learning" robots and complex software systems. In the movie *The Matrix*, these self-learning AIs come up in form of Agent Smith for example. He, or let's say this system, has learned how to hack himself into other software systems and multiply himself by that. "Neo delivered me," Agent Smith will say at one point. But the truth is: his freedom is just an illusion. What he stands for is simply a chaotic system, which may act unpredictably but only as a result of randomness and not as the result of deliberation and weighing reasons.

Moving from deterministic machines to probabilistic machines does not cancel out the categorical differences between humans and machines. The alternative is not between determinism and probability but between determination and freedom.

"You see," the Merovingian tells Neo und Morpheus, "there is only one constant, one universalism. It is the only real truth: causality. Action. Reaction. Cause and Effect."

"Everything begins with choice," Morpheus contradicts him.

"No. Wrong. Choice is an illusion," the Merovingian says.

Is the Merovingian right? Are humans—just like natural objects—subject to the principle of causality?

In philosophy, there are three answers to this question: non-compatibilism, compatibilism, and semi-compatibilism.

Non-compatibilists believe that in the world described by natural science, there can be no freedom and no responsibility because determination and freedom are incompatible. Non-compatibilists are generally "naturalists." They believe that scientific laws govern all which happens and that there is consequently no room for freedom of will. Freedom of will is solely a useful illusion, which allows us for example to make children feel responsible for their misbehavior. Threatening someone with sanctions, however, will influence and determine human action. Naturalism as an ideology is very present in Neuroscience. It negates human freedom and responsibility by referring to

our supposedly determined brain system, which is controlled by genetic, epigenetic, as well as sensory stimuli.²

The problem of this position is not only that it goes against the intuition of most people, but also that it is obviously wrong.

The individual development of our character is not only dependent on our environment and on genetics but also on decisions. This coincides with the findings of Aristotle as he has formulated them in the *Nicomachean Ethics*. Aristotle makes it clear that virtues (character features, dispositions, etc.) are not solely based on habit and education but are also an expression of one's decisions (*prohairesis*).

Of course, experience and habit are important for the development of virtues. But humans are capable of changing their beliefs and making conscious decisions, of changing their attitudes in the future (Aristotle speaks of virtue as "hexis," which means that attitude and belief is the result of weighing reasons and finally gaining an opinion, especially after hardship or a time of crisis in one's life. But also emotive attitudes, for example admiring another person relies on the belief³ that this person has achieved something particular or has shown a great deal of helpfulness, etc. We are not just solely "products" of education and socialization but are actively involved in forming our own character.

The question of freedom of will has been debated in philosophy since antiquity. In today's philosophy, the so-called "compatibilism" dominates the discussion—it is the theory that a complete determination is compatible with human freedom of will and of action. Even if most of these compatibilists are naturalists and hang on to the idea that in the end everything is determined by physical processes, they believe that despite that, it is possible to look at humans as free and responsible agents. They believe that in order to be called free it suffices that humans are able to fulfill their wishes, independently of if they are free to choose these wishes or not. By that, freedom is being made possible in a deterministic world. Freedom of action is being defined as freedom to do what I wish to do—independently of how these wishes came about. This compatibilist view is hard to bring into accordance with our self-image as free agents. The feeling of being an author of one's life demands more than just choosing the best means to a given end. Being authors of our lives, we want to determine our own goals.

Genuine authorship is not compatible with the idea that humans are algorithmically controlled in their goals and actions. The capability to deliberate,

² Singer (2001).

³ Nida-Rümelin (2018).

to weigh reasons against each other plays a central role for our self-image according to which it cannot be that our valuations and decisions are already fixed *before* we even begin to weigh reasons for and against it. Weighing reasons should not be regarded as a predetermined process. The result of this process is open and this is what makes out our freedom.

The debate about freedom of will is ultimately a debate about the question if our self-image as free and responsible beings, who let themselves be guided by beliefs and reasons, is an illusion or not. If one breaks away the element of freedom of action, one could neither hold people responsible for their actions nor morally judge, praise, or criticize them. There is a lot at stake here. Nothing less than our lifeform.

Hegel once said that “freedom is the comprehension of necessity.” Indeed if freedom consists of acting according to one’s reasons, accepting necessities becomes a form of freedom. This which is necessary is no longer looked upon as a restriction. However, this should not lead to the idea that every kind of limitations on freedom should be accepted by humans. The prisoner in his cell might find peace of mind if he gave up his wish to leave the prison; he should not however fool himself into believing that his situation doesn’t mean a massive limitation of his possibilities. The same goes for a servile attitude towards authorities solely with the aim of avoiding conflicts. The opportunist who arranges his wishes according to what can be achieved with the least of obstacles to be confronted loses her willpower and ultimately the authorship of her life. In an extreme case, she will degenerate to being simply a function of external circumstances, only functioning according to what is being demanded of her.

At the end of the second part of *The Matrix* franchise, Neo enters a brightly lit room, the central computer of the Matrix. In this room, there are hundreds of television screens. In the right corner, an old man with a white beard is sitting on an office chair. Opposite of him is Neo, in long, black priest-like clothing with a high-fitting collar.

Obviously, this scene is meant to insinuate a meeting between God and his creation. The white-haired man who speaks of himself as the “father” of the Matrix insists—just as the Merovingian has done before him—on the fact that humans, just like machines, are subject to the laws of causality. Then he presents him with the following choice: he can either save the world or his lover. He cannot do both. Neo chooses the latter. Later in the story however Neo will prove him wrong and save both. Obviously, machines just like humans cannot be fortune-tellers.

The Matrix becomes thus a hymn to the human spirit and to humans who possess the freedom of will to act according to their reasons, make their own decisions, and act in and upon the world.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits any noncommercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if you modified the licensed material. You do not have permission under this license to share adapted material derived from this chapter or parts of it.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

