

DO ROBOTS DIE?

Gergely DELI

Professor (University of Győr)

If we would like to reveal the connection between robots and natural law, this is the first question we should ask. But let's take a step back and clarify the basic concepts we will use. The natural law (*lex naturalis*) is the rational creature's participation in the eternal law (*lex aeterna*).¹ For the time being, until they are created by humans, robots are only theoretical beings with general intelligence and physical dimensions. Will robots participate in the eternal, divine law?² Robots, if they will exist at all, will necessarily have to obey the laws of physics. The most important law of physics that seems to govern the universe is the second law of thermodynamics. According to this law entropy (in other words, the disorder of things) increases continuously until it reaches its possible maximum value.³ This means that the development of the cosmos can be described as a gradual growth of disorder.⁴ All things aim at reaching the state of thermodynamic equilibrium, "the state of universal rest and death".⁵ The physical world we know is driven by the entropy; this is the eternal law of the cosmos. The process of the entropy is irreversible and unchanging. Just like the eternal divine law, the *lex aeterna* according to Saint Augustine.⁶ The concepts of entropy and eternal law have interesting common traits. However, the idea that entropy and *lex aeterna* are identical cannot be accepted. According to the natural law the eternal law is connected with eternal life, and not with the heat death of the universe. The question arises as to how life and death in the sense of physics and in the sense of natural law relate to each other.

¹ *Summ. Theol.* II-I, q. 91, 2a.

² *Summ. Theol.* II-I, q. 91, 1a.

³ Max TEGMARK: *Élet 3.0. Embernek lenni a mesterséges intelligencia korában.* [= Life 3.0. Being Human in the Age of Artificial Intelligence] Budapest, HVG Könyvek, 2018. 277.

⁴ Carlo ROVELLI: *Az idő rendje.* [= L'ordine del tempo] Budapest, Park, 2018. 166.

⁵ William THOMSON: On the Age of the Sun's Heat. *Macmillan's Magazine*, vol. 5. (1862) 388–393.

⁶ *De lib. arb.* I, 6; ld. még *Summ. Theol.* II-I, q. 91, 1c.

From the perspective of physics life is a network of processes that increase entropy and catalyze each other's effects.⁷ Death, that consists in the dissolution of the order according to its own rules, is an inherent part of this process. According to Saint Paul death is the wages of sin.⁸ God created man immortal, and allowed him to be in control of his own life or death.⁹ Decay and death are consequences of sin. What is the link between entropy and sin?

In Paradise decay threatened man in two ways, and, accordingly, man could defend himself in two ways against these deadly attacks. The first possible cause of death is the loss of moisture. This explains why Kant abhorred perspiration and the wasting of semen.¹⁰ The loss of moisture is caused by natural heat, that is, according to Thomas Aquinas, the instrument of the soul.¹¹ Man could defend himself against this internal cause of decay by eating.¹² As a physicist would put it: life is a physical process that is supported by the low entropy of food.¹³ The second, external cause of death is when moisture extraneously generated is added to the pre-existing moisture, diminishing the active vitality of the latter.¹⁴ Thomas Aquinas illustrates this idea with two examples. If we add water to wine, first we get a vinous fluid. However, if we add more and more water, in the end the wine becomes watery. In the second case he takes nutrition and the process of aging as an example. Children are able to utilize nutriment in a sufficient quantity that allows them to grow. In adults food is only sufficient to maintain the tissues of the body, while in old age it is not even up to this: shrinkage and in the end death of natural causes follow. In the language of physics: food is transformed into vital energy, but this process entails heat loss, and the lost heat cannot be gained back and reused without expenditure of energy: the entropy does not come back.¹⁵ In Paradise man could ward this danger off by eating the fruit of the tree of life. According to Saint Thomas Aquinas this fruit was able to prevent the first human couple from dying after partaking of extraneous food, that is, they were not exposed to entropy.¹⁶ In Saint Augustine's words: "*Cibus aderat ne esuriret, potus ne sitiret, lignum vitae ne illum senecta dissolveret*"¹⁷ – Man had food to appease his hunger, drink to slake his thirst; and the tree of life to banish the breaking up of old age.

⁷ ROVELLI op. cit. 165.

⁸ Romans 5, 12.

⁹ *Summ. Theol.* I, q. 97, 1a.

¹⁰ Ehregott Andreas WASIANSKI: *Immanuel Kant in seinen letzten Lebensjahren. Ein Beytrag zur Kenntniß seines Charakters und häuslichen Lebens aus dem täglichen Umgange mit ihm.* Königsberg, F. Nicolovius, 1804. 36., 38. For further explanation see Jean-Baptiste BOTUL: *Immanuel Kant szexuális élete.* [= La Vie sexuelle d'Emmanuel Kant] Budapest, Palatinus, 2004. 58–67.

¹¹ *Summ. Theol.* I, q. 97, 4a.

¹² *Summ. Theol.* I, q. 97, 4a.

¹³ ROVELLI op. cit. 165.

¹⁴ *Summ. Theol.* I, q. 97, 4a.

¹⁵ ROVELLI op. cit. 164.

¹⁶ *Summ. Theol.* I, q. 97, 4a.

¹⁷ *De civ. Dei* 14, 26.

In Thomas Aquinas' opinion, since we no longer live in Paradise, death has two causes: an internal, unnatural cause and an external, natural cause. The external cause of death is entropy, and it consists in the natural dissolution or dissipation of the body (*naturalis dissolutio corporis*). The internal cause of death is the loss of moisture due to the heat generated by the functioning of the soul, but it would not be sufficient to cause decay. The soul is immortal, and the way it works in us cannot lead to death. In theory we could recover the lost moisture by food intake. Death and the continuous decrease of energy is due to an unnatural cause: sin. Every sinful act proceeds from inordinate desire for some temporal good.¹⁸ Sin interferes in the order of life, upsets the energetic equilibrium, and on the whole increases entropy.¹⁹

The first from the two causes of death, entropy is a universal physical law, and, as such, it necessarily affects robots, too. Robots cannot evade entropy, because if they will exist, they will be constructed by human beings subject to the power of entropy, using materials that are parts of the created physical world and are also subject to the law of entropy. Robots will necessarily be mortal, just like us. As we saw, the other cause of death is the way our soul functions, or more exactly, the fact that we commit sinful acts. If the soul is separated from God because of the sin, it is not able to stop the corruption of the body. Will robots have souls, can they have souls, and if so, will they sin? To answer this question, we must have a clear idea of the essence of the soul. This examination leads us to another mysterious phenomenon: the phenomenon of time.

Time is nothing else but the distinction between past and future. With our brains full of memories and predictions, we need time to be able to interact with the world. In this way time becomes the source of our identity.²⁰ There are three elements of our identity, or, in the terminology of natural law, of our soul that are relevant to our subject.

The first characteristics worth mentioning is the fact that in his or her inner consciousness or soul each of us adopts a certain view of the world. The world exists for us in the here and now, the way we see it. The way the world reflects in us depends on the correlations that are relevant for our survival.²¹ Gregory of Nyssa defines the relationship of God and the soul in a similar way, breaking with the idea of Neoplatonism on the unity of God and the soul.²² "Like the eye is not able to see itself, but only its image in the mirror, the soul can get to know itself only by beholding the Original Beauty reflected in the mirror and in the figure of her own beauty"²³ – says Gregory of Nyssa, inverting the famous mirror analogy. If it is God that reflects in our soul, we reflect the most important correlation in the universe.

¹⁸ *Summ. Theol.* II-I, q. 77, 4a.

¹⁹ *Summ. Theol.* II-I, q. 83, 1a.

²⁰ ROVELLI op. cit. 190–191.

²¹ ROVELLI op. cit. 175.

²² IVÁNKA, Endre: Nyssai Szent Gergely mint a keresztény misztika úttörője. [= Gregory of Nissa as the Pioneer of the Christian Mystique] In: IVÁNKA, Endre: *Heidegger filozófiája és az ókori metafizika.* [= The Philosophy of Heidegger and the Ancient Metaphysics] Budapest, Paidion, 2004. 33.

²³ PG XLVI 509.

The second point refers to the special way our nervous system works, namely, it dissects the world for us to be able to digest it.²⁴ We don't perceive our existence as a continuum, but a fragmented reality of memories, desires and past experiences. Eventually, that is how God created it from small particles, using fine discretion. Granularity is not only characteristic of our perception of time, but also of the structure of time (the so-called Planck time).²⁵ The idea of granularity of time was shared by Saint Isidore of Seville,²⁶ Venerable Beda²⁷ and Maimonides.²⁸ Time is basically a kind of inner sense, a way to dissect the world. Time helps us to bring order into our inner state, or, in other words, our soul.²⁹

The third building block of our identity is memory.³⁰ To explain the role of memory in our identity, Saint Augustine compares it to listening to a hymn. If we listen to a hymn, the sensation a sound generates in us depends from the previous and the subsequent sounds. Thus music makes sense only in its temporal dimension. But if we live in the present, how can we grasp the meaning of music? According to Saint Augustine this is linked to the base of our self-consciousness, that is, to the fact that we are endowed with a soul, and it implicates that we are able to remember and anticipate things.³¹

Therefore robots will have souls, if they will exist in time. As we could see, it has three important components. On the one hand, robots should be similar to humans in being able to "reflect" the most crucial correlation in the world, God. According to the doctrine of the Catholic faith we are not capable of it through our own efforts, but only through the grace of God. The mere physical characteristics of the neural networks of the brain do not necessarily imply that we are able to recognize God. From the point of view of natural law robots will have souls only if God, by His grace, wants them to have souls.³² But we will talk about this later. On the other hand, robots should perceive the world as a highly fragmented reality. This condition is likely to be met, because the functioning of robots will necessarily be subject to the physical laws of our segmented world. Thirdly, robots should have memory and anticipation. It is not unrealistic, either, since the memory capacity of our computers have already surpassed by far the memory capacity of the human brain. On the other hand, thanks to meta-algorithms, some special artificial intelligences are already endowed with something we, as human observers, would define as preliminary understanding or intuition.

To sum up, robots will have souls, if they will be able to recognize the most crucial correlation in the world, God, among all the important correlations they will reflect. Man does not have a say in God's decision on whether by His grace robots will have

²⁴ ROVELLI op. cit. 174–175.

²⁵ ROVELLI op. cit. 8.

²⁶ *Etymologiae*, 1, 16.

²⁷ *De temp. rat.* Caput III.

²⁸ *Dalālat al-hā'irīn* [= Guide for the Perplexed], 52.

²⁹ ROVELLI op. cit. 185.

³⁰ ROVELLI op. cit. 178.

³¹ ROVELLI op. cit. 183.

³² *PG XLVI* 717.

souls or not. Let's address what we, humans can do. God, the most crucial correlation in the world dwells within us to the extent we prepare a place for Him through our love and purity.³³ The expression "dilation of the heart" (*dilatatio cordis*)³⁴ used by Saint Teresa in *The Interior Castle* should be understood in this sense.

Ultimately, for the souls of the robots the pivotal question is the robots' approach to the crucial correlations of the world. The algorithms the souls of the robots are constructed of function according to the correlations or patterns of the world perceived by the robots. In this aspect robots are similar to us, humans. Both robots and us base knowledge on experience. We obtain knowledge by processing impressions and inputs from the perceptible external, material world. In addition to that, human beings and robots are surrounded by the same world, thus the object we gain knowledge of is also the same. However, it is possible that the conclusions we and the robots will arrive at will not be identical. The fact that we try to acquire knowledge of the same object does not necessarily implicate that the results of this cognitive process will also be the same.

Although every piece of our knowledge is based on experience, not every piece of knowledge derives from experience. Our senses allow us to gain knowledge of separate phenomena, separate features and separate data. The relation between separate data cannot be detected through sensory experience. The arrangement and order of these data, the relations and causal links between them, the categories they belong to come from the conclusions we arrive at on the basis of what we experienced. It is not sure that robots and humans will recognize the same connections between the data obtained. Robots and humans may liberate different forms from the same matter. In other words, although we get the same inputs as robots, because we and robots live in the same world, it is not sure that our conclusions and outputs will be the same, too.

The existing special artificial intelligences have already proved that robots are able to arrive at conclusions that are different from our knowledge. In many cases these conclusions reach or even surpass the level of our knowledge both quantitatively and qualitatively. This is convincingly demonstrated by move 37 of the algorithm AlphaGo in the second game of the Go match against Lee Sedol. This move was a qualitative leap compared to what mankind was able to reach in the three thousand year history of Go.³⁵ Of course, in a context where the goal is clearly defined (like in a Go match), it is easy to evaluate the conclusions of the robots. If the robot's conclusion contributed effectively to reaching the goal, or, in this case, to winning the game, it was an appropriate conclusion. But there are cases when the goal cannot be defined clearly. Salvation can be regarded as the goal of human life, but, as we cannot define it, we are not able to tell whether the conclusion of a robot is salvific or not.

According to Thomas Aquinas man has a faculty called *intellectus agens* that helps him to liberate the form from the matter via abstraction. It is well-known that in Thomas Aquinas' opinion abstraction does not mean the recognition of common

³³ PG XLVI 1269.

³⁴ *Libro de las moradas o castillo interior* 4, 2, 5.

³⁵ Cade METZ: The sadness and beauty of watching Google's AI play Go. *Wired Magazine*, 2016/3. <https://www.wired.com/2016/03/sadness-beauty-watching-googles-ai-play-go/> [2019.06.06.]

characteristics that recur in a certain number of cases, but the direct recognition of substantial characteristics, even if they occur only in a single case.³⁶ In other words, according to Thomas Aquinas there is a human capacity that helps us to recognize the substantial, so-called eidetic correlations of the world.

Currently robots lack this faculty, they are not able to distinguish between important and unimportant patterns that can be found in the world. This can be explained by the fact that the current stage of development of the artificial intelligence concentrates on self-learning algorithms that use the common characteristics that occur in a statistically significant number in the huge mass of separate data. We know for sure that today robots lack the faculty of eidetic recognition, but we are not able to tell whether they will acquire this faculty in the future. On the other hand, there is no exact proof of the existence of this faculty in humans, either. This idea is based on the authority of some of the greatest philosophers and on the belief, that God granted the human brain the ability to recognize Him. From this perspective we cannot tell for sure whether robots will have souls in the future. As we are not able to demonstrate unequivocally that our cognitive ability can recognize God, the most crucial correlation in the world, we will not be able to assess such abilities in robots, either, if they will obtain them in the future. Therefore we should beware of feeling superior to robots.

Whether God exists or not, or whether there are objectively substantial correlations in the world, the emergence of robots will not affect significantly the existing important principles of natural law. It is possible that we will modify or discard a previous principle of natural law because of principles of natural law recognized by robots. However, this process should remain under human control. It is in the nature of things that man cannot use a robot to decide whether a law revealed by another robot is a substantial, that is to say, a divine law. After all, man has to rely on his eidetic faculty granted him by God, if he has such a faculty or if we believe that he has such a faculty. As humans we can only use our human reason to tell whether the potential new laws discovered by robots are eidetic or divine laws.

Let's return to our initial question on the connection between robots and natural law. Robots can help us to get closer to God, but it remains our own responsibility to act.

³⁶ IVÁNKA, Endre: Hogyan oldja meg a tomizmus a kantianizmus ismeretelméleti problémáját. [= How does the Thomism solve the Epistemological Problem of Kantianism] In: IVÁNKA (2004) op. cit. 50.