STEPHEN HAWKING AND MACHINE INTELLIGENCE[1]

Enrico Beltramini

ABSTRACT: This note addresses the thought of Stephen Hawking on machine intelligence in form of a preliminary reflection. The note focuses on the gap between the possibility of an artificialized science, which was considered by Hawking, and that of an artificialized humanity, a possibility that Hawking did not consider. In this context, the artificialization of the human is led not by human scientists and technologists, but by artificialized science.

KEYWORDS: Stephen Hawking; Artificial Intelligence; Disenchantment

FROM THE UNIVERSE TO THE HUMAN

Cosmologists and theoretical physicists have shown a growing interest in both the development of formal sciences (i.e., logic, mathematics, statistics, theoretical computer science, robotics, information theory) and the future of humankind. A case in point is Max Tegmark (1967-), theoretical physicist at MIT, famous for his previous book on the mathematical character of the universe (Tegmark 2014), who argues that the only possibility of guaranteeing the survival of the human species is space colonization. This space colonization, in turn, will be ignited by machine superintelligence (i.e., machine intelligence that surpasses human intelligence). Thus, the current reality of an organic *Homo Sapiens* resident on this planet will be replaced by a universe filled with machine superintelligence with a human face. Tegmark's message is clear: not only might the current technological discussion of Artificial Intelligence (AI) have profound impact on the trajectory of life for countless millennia but also far beyond our own planet (Tegmark 2017). At this point, his reflection has left behind familiar debates about the

[1] Acknowledgement: I would like to thank the anonymous referee for the comments and suggestions on an earlier draft of this paper.

job market, privacy, and weapons of mass destruction to venture into realms that hitherto were associated with philosophy.

Another case in point is Stephen Hawking (1942-2018). He made several remarks about AI. In general, he was not a big advocate of an unconstrained evolution of AI as an independent intelligence. An independent (or autonomous) machine intelligence is able to continue to upgrade itself and therefore to advance technologically at an incomprehensible rate, *independently from humans*. Hawking thought that this autonomous machine intelligence would pursue independent, i.e., independent from humans, goals.[2] His tendency was not toward a malicious AI, rather an indifferent AI, an AI indifferent to human concerns. He synthesized his view in a 2015 question-and-answer session on Reddit:

> The real risk with AI isn't malice but competence. A superintelligent AI will be extremely good at accomplishing its goals, and if those goals aren't aligned with ours, we're in trouble. You're probably not an evil ant-hater who steps on ants out of malice, but if you're in charge of a hydroelectric green energy project and there's an anthill in the region to be flooded, too bad for the ants. Let's not place humanity in the position of those ants (Hawking 2015).

Hawking encouraged scientists and technologies to create AI and to make sure that its use was beneficial to humanity. He never put it this way, but it can be inferred that he was for the use of a non-autonomous intelligence. As people mourn his loss, they should remember that his legacy goes far beyond science. He also had a remarkable legacy as a social theorist, including his nuanced perspectives on what specific dangers AI could bring to humanity.

Hawking was primarily one of the greatest theoretical scientists of the past century. His work on black holes and relativity, i.e., discovering that black holes evaporate and helping found the modern quest for quantum gravity, contributed to him being regarded as one of the most brilliant theoretical physicists since Albert Einstein. Hawking was a scientist, and he considered science, particularly physics, the preeminent form of knowledge. He famously argued that fundamental questions about the nature of the universe can only be resolved by science, not philosophy. In the first chapter, accurately titled 'The Mystery of Being,' of a recent book co-authored with Leonard Mlodinow, Hawking explained that people have always asked existential questions such as, "How can we understand the world in which we find ourselves?" "How does the universe behave?" "What is the nature of reality?" "Where did all this

[2] In this article I equate 'autonomous intelligence' to Artificial Intelligence, while I do not necessary always equate Artificial Intelligence to 'autonomous intelligence.'

come from?" "Did the universe need a creator?" Hawking then argued:

> Traditionally these are questions for philosophy, but philosophy is dead. Philosophy has not kept up with modern developments in science, particularly physics. Scientists have become the bearers of the torch of discovery in our quest for knowledge (Hawking and Mlodinow 2010, p. 13).

Whether Hawking is correct on this point, and whether or not philosophy is still a viable contributor to human understanding of the world, is beyond the scope of this article. This article instead discusses the basic assumption behind Hawking's remark on philosophy, that is, that people wonder about the universe and their place in it. It seems to me that the direction of people's concern has changed: it is no longer the world, or reality, but the human itself has become the focus of concern. The gesture that, broadly speaking, exemplifies contemporary sensibility operates against the ontological and scientific type of metaphysics inherited from Aristotle and his medieval commentators. It is not the cosmos out there, rather the cosmos inward, that seem to spring up out of current Zeitgeist. It is not the grandiose stars, galaxies, and black holes that matter, but the interiority of human mind, intelligence, and life. It is not the space, but the existence, that matters. Not surprisingly in the last decades, the direction of the scientific enterprise has changed: it is no longer the object, but the subject itself that is the focus of inquiry. So, the revolutionary potential of technology and the science have shifted their focus, and all expect that technology and science will radically remake those traditional domains that fall within the boundaries of the human.

DISENCHANT AND RECONSTRUCT THE HUMAN

'Disenchantment' is a word made famous by sociologist Max Weber (Weber 1976). It means that reality is what is left when the magical, spiritual, even religious meaning is stripped out of nature (Taylor 2007). Disenchantment is the process that leads to a notion of nature without a supernatural component. This process has been primarily a scientific enterprise: humans have developed science and through scientific inquiry they have disenchanted nature. When Hawking noted that science has replaced philosophy because "fundamental questions about the nature of the universe could not be resolved without hard data such as that currently being derived from the Large Hadron Collider and space research," he is implicitly assuming that the nature of the universe is made in such a way that hard data are useful to answer fundamental question about it (Hawking 2011). If the nature of the universe was, to say, consciousness, hard data would be useless. But it is this specific understanding of the fabric of the universe -- a universe that is disenchanted, nothing more than inert matter -- that is the result of the scientific treatment of nature.

Now that the direction of the scientific enterprise has changed, and the human itself is the focus of inquiry, it can be expected that the next form of remaking, of course, will probably be the disenchantment of the human. The process will probably lead to a disenchanted human. A disenchanted human is what is left when the entire humanistic heritage is dismissed as 'non-scientific,' nothing more than a myth. In the process of disenchantment, the human is remade. The human is remade in the sense that the humanistic heritage, i.e., the traditional ontological claim of a human exceptionalism, is dismissed. With 'human exceptionalism' it is meant that humans are a species set apart; it is the belief that human beings hold a special status in nature. Thus, the termination of humanism, or human exceptionalism, ultimately makes humanness identical to nature. The scientific inquiry of the human is, after all, an attempt to reduce the human to nature. In the process of disenchantment, it has been said, the human is remade. Science makes myth out of traditional ontological claims of the universe and relegates them to the history of ideas. All things being equal, it is fair to suppose that science will also make myth out of traditional ontological claims regarding the human.[3]

Once the initial form of remaking, i.e., disenchantment, will be complete, a more invasive form of remaking can be expected. This form of remaking of the human by technology and science is the creation of the new human. This new human will be the result of the manipulation of nature by biotechnological and genetic tools. The new human will also be the result of advanced studies on the nature of mind, and the possibility of some form of digital immortality. Digital immortality (or 'virtual immortality') is the hypothetical concept of storing or transferring a person's personality in more durable media, i.e., a computer, and allowing it to communicate with people in the future. The only real question is one of how radically the human will be remade. Here, everyone differs, and in quite predictable ways. No matter what position people take, however, it is clear that the new human will set humanity once and forever apart from nature. If this new human is special (or exceptional), it will be not special because of some intrinsic quality, eventually developed by natural selection or imprinted at the beginning of time by a supernatural being. The new human will be special because of science and technology.

[3] Declaring that traditional ontological claims regarding the human will not suffer the fate of other traditional ontological claims more generally amounts to declaring that some kind of transcendental argument operates in the background. Or, maybe better, that a meta-philosophical condition exists that makes the human special. So, it is either that a meta-philosophical condition limits and corrupts scientific inquiry or an unrestrained and pure scientific inquiry displaces the ontological claim of a human exceptionalism.

DENATURALIZED SCIENCE

Thus, the remaking of the human by science and technology will equate to a dual movement: a naturalization of the human, humanness as nature, and an artificialization of the human, the human like a machine. More importantly, this dual movement will be governed by an increasingly denaturalization (artificialization) of science. Philosopher Luciano Floridi argues that

> the increasing and profound technologisation of science is creating a tension between what we try to explain, namely all sorts of realities, and how we explain it, through the highly artificial constructs and devices that frame and support our investigations. Naturalistic explanations are increasingly dependent on non-natural means to reach such explanations (Floridi 2017, p. 272).

He considers unfortunate this incongruency between means and ends. His suggestion to fix the incongruency is to equate explanation to construction: his point is that naturalistic explanations are ultimately constructions, that is, non-natural. Thus, no naturalization of the human is at work, because "the natural is in itself artefactual (a semantic construction)" (Floridi 2017, p. 269). But what if it does? What if the naturalization of the human is at work? What if the naturalization of the human through the artificial is not, as Floridi argues, a predicament, eventually an embarrassment, but rather part of the dual movement that is remaking the human? In this case, the naturalization of the human, and its twin movement, the artificialization of the human, are both led not simply by human scientists and technologists, but also by artificialized science and technology, including AI. The scientists who "have become the bearers of the torch of discovery in our quest for knowledge" as Hawking claimed and have discovered new theories that "lead us to a new and very different picture of the universe and our place in it," are not necessary humans and they do not lead us to a new picture of the universe (Hawking 2011). They may be pieces of an increasingly powerful AI which lead humans to a new picture of themselves as well as to a very different existence. To put it differently, humans become increasingly dependent of AI. As a matter of fact, an increasingly dependent humanity must already deal with an increasingly autonomous intelligence.

The remaking of the human through an increasingly artificialized science and technology brings me back to Stephen Hawking and his remark on AI. He claimed that humans must be careful when it comes to AI: they must make sure that AI is not only useful but also beneficial. But who are the humans Hawking is talking about? In his book *The Grand Design: New Answers to the Ultimate Questions of Life*, he portraited humans as "a curious species." Then he continued:

> We wonder, we seek answers. Living in this vast world that is by turns kind and cruel, and gazing at the immense heavens above, people have always asked a multitude of questions" (Hawking and Mlodinow 2010, p. 13).

But this idea of humans, so far from evolution and nature, so deliciously humanist in its character, is also far from the human in becoming, the human that science and technology are remaking. Because it is not only that humans are making science and technology, but also that science and technology are remaking the human. More precisely, an increasingly artificialized science and technology are remaking the human. Hawking is correct to warn humanity about the risk of AI, but AI is already remaking humanity, and it's likely the entire conversation needs to be rethought.

CONCLUSION

The idea that humanity is transforming itself through the development of technologies is not new. Marx argued something like this in the '1844 Manuscripts', that humans form themselves through their transformation of nature and can only truly know nature insofar as they have produced it (Marx 1956). Those who have examined the technologizing of the world, like Marshall McLuhan, Walter Ong, Hobart and Schiffman, and Susan Greenfield, that is, those who have studied the profound effects on humans of different technologies of communication (from the first writing to electronic media), also made this claim and saw more threatening aspects to it. Their assumption, however, was that humans are intelligent, while technology is not. Enter AI. AI appears to be remaking humanity in such a way that not only is life even more disenchanted than it was with reductionist science, but people are being dumbed down and humanity is increasingly blind to how it is destroying the conditions for its existence, transformed or not. Other scholars would like to offer more clarification of the argument about the direction civilization now appears to be taking, and what is wrong with this direction.

Floridi's claim that through the development of knowledge we are creating nature rather than merely explaining it still operates in the domain of human exceptionalism. It can rather be claimed that, in creating nature, humans are subordinating themselves to machine intelligence which is transforming them. If in his effort to acknowledge human creativity Floridi is correct in his arguments, and the implications of his arguments are those suggested in this article, a terrifying trend is at work in civilization. With that said, Hawking might not have understood the extent of the threat of AI to humanity, and philosophy still has something of interest to contribute.

ebeltramini@ndnu.edu
Notre Dame de Namur University

BIBLIOGRAPHY

Luciano Floridi (2017), "A Plea for Non-naturalism as Constructionism," *Minds & Machines* 27:269–285.

Stephen Hawking (2015), Question and Answer Section, The New Reddit Journal of Science, available at https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/ (accessed March 17, 2018).

Stephen Hawking (2011), Speech at Google Zeitgeist conference in England, in Matt Warman, "Stephen Hawking tells Google 'philosophy is dead," *The Telegraph*, May 17, available at http://www.telegraph.co.uk/technology/google/8520033/Stephen-Hawking-tells-Googlephilosophy-is-dead.html (accessed March 17, 2018).

Stephen Hawking and Leonard Mlodinow (2010), *The Grand Design: New Answers to the Ultimate Questions of Life*. London, Bantam Press.

Karl Marx (1956). Economic and Philosophic Manuscripts of 1844, translated by Martin Milligan from the German text, revised by Dirk J. Struik, contained in Marx/Engels, Gesamtausgabe, Abt. 1, Bd. 3. London, Prometheus Books.

Charles Taylor (2007), *A Secular Age*. Cambridge, Mass., Harvard University Press.

Max Tegmark, *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality*. New York: Alfred Knopf, 2014.

Max Tegmark, *Life 3.0. Being Human in the Age of Artificial Intelligence*. New York: Alfred Knopf, 2017.

Max Weber (1976), *The Protestant Ethic and the Spirit of Capitalism*. London: George Allen & Unwin, 2nd ed.