Review: **JAR1411**

WITHOUT GOD IN THIS WORLD BUT WITH HOPE IN THE NEXT

a book review of

*Life 3.0: Being Human in
the Age of Artificial Intelligence*

by Max Tegmark
(Kopf, 2017)

Life 1.0 is experienced by bacteria. Their bodies and their behavior are programmed into them by the genetic code of their DNA. Life 2.0 is experienced by humans now. Our bodies are shaped by our inherited DNA, but we can choose our behavior. Life 3.0 does not yet exist on Earth, but it will be here soon, giving us the ability to reshape our brains and our bodies. We had better think carefully and well about that reshaping, because only the laws of physics limit what we can do, and the possible outcomes range from infinite bliss through mindless slavery to total annihilation.

This is the message of Max Tegmark's book *Life 3.0*. Tegmark is doing all he can to insure the chance of infinite bliss by founding and leading the Future of Life Institute and writing this book. Tegmark well knows that the medium is the message, so his book requires his readers to make choices about how they will read his book — and those choices model the way humanity's choices will determine our future and even the future of the universe.

Tegmark is a professor of physics at MIT whose wide range of knowledge and engaging style draw the readers into a delightful conversation. He entertains readers with facts that he off-handedly tosses into the discussion. (Three examples are: (1) there are 1078 atomic particles in the observable universe; (2) the reason you can blink before a bug hits your eye is the blinking reflex short-circuits your conscious brain to enable you to act without thinking; and (3) a possible solution to the theodicy problem is that if God remedied all evil, it would destroy human freedom.) His frequent questions to the readers (Do you think that superhuman AI will be created in this century? Yes — continue to the next page. No — skip to chapter 1) keep readers engaged and thinking.
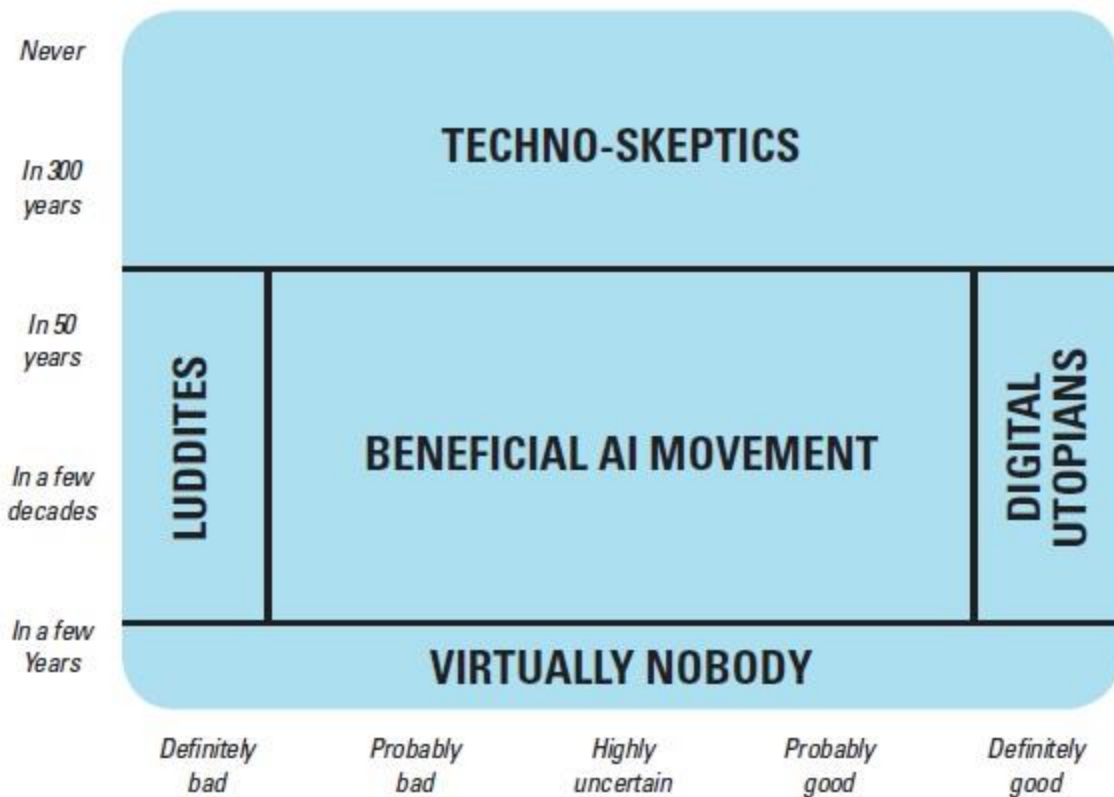
Besides his witty and catechistic style, Tegmark supplies perspicuous illustrations and detailed summaries of each chapter's main points. Along with his

wealth of information, Tegmark also displays a refreshing mass of ignorance. Instead of pretending he knows the answers to debated questions, Tegmark gives the range of possible answers along with their advocates and arguments. In a chart, he coordinates the answers to the questions, "When will AI surpass human level?" and "If superhuman AI appears, will it be a good thing?" He then applies useful labels to the range of answers.

How Matter Learned to Think. For those who think superhuman AI is near, Tegmark prefaces his book with a fable of how people developed superhuman AI and how they used it to control both the Earth and the galaxy. Those who do not think superhuman AI will appear in this century are invited to skip to the next section, which says that since superhuman AI might be possible in the next century and there may be ways we can influence it, we should begin to talk about it now. Then comes a history of intelligence, where he defines "intelligence" as the ability to achieve goals and states that intelligence has three parts: memory, computation, and learning. Tegmark then points out that all three parts of intelligence do not depend on a specific material base but can happen in machines as well as brain cells. He next argues that due to the ability of computers to do arithmetic, fly planes, win at chess, recognize speech, and teach themselves to play games without any instructions, computers might soon learn to make better computers than humans can.

Tegmark then considers the near future where it may be possible for all tasks that involve extracting resources and manufacturing goods, as well as many that involve uncomplicated human interactions, to be done through AI. If this situation comes about, we will need to make AI more robust so it cannot be used in harmful ways. Another potential harmful side effect of AI could be the increasing inequality of wealth throughout the world.

If we succeed in building human level intelligence, those computers might be able to build much better computers, which could build better computers, leading to an intelligence explosion. If a group of humans manages to control that intelligence explosion, they could soon control the world. If they do not, the computers themselves could gain control of the world even more quickly. Either way, the outcome could be the best or worst thing that ever happened to humanity, so we should start thinking hard about what we can do now to make the outcome good. Tegmark next pushes the discussion to the next 10,000 years with the plusses and minuses of twenty-four scenarios if we somehow escape the 99.9995 percent chance we will destroy ourselves in a nuclear war. Finally, he stretches the horizon out to a billion years, asking what are the physical limits on energy, information, and computation. The chapter ends with the plea that we wisely embrace technology so that humanity will not go extinct with the death of our solar system but will experience life and meaning almost infinitely.

Never		TECHNO-SKEPTICS		
In 300 years				
In 50 years	LUDDITES	BENEFICIAL AI MOVEMENT		DIGITAL UTOPIANS
In a few decades				
In a few Years		VIRTUALLY NOBODY		
	Definitely bad	*Probably bad*	*Highly uncertain*	*Probably good*	*Definitely good*

How Matter Sets Goals. After having surveyed almost fifteen billion years of progress, Tegmark shifts his focus to how purposeful activity arises. He argues that one way of thinking about the laws of physics is Fermat's principle that every natural phenomenon can be reformulated mathematically as goal-seeking behavior. For instance, the goal that every thermodynamic reaction pursues is the increase of entropy. According to this way of thinking, life is an activity that increases entropy more efficiently than other phenomena and adds a new goal of replication. We humans do not always know the best means of reproduction, so we have evolved feelings of hunger, thirst, pain, lust, and compassion to set goals to guide us. As we create machines to help us achieve our goals, we will build those goals into the machines. We need to solve three problems: making the machines learn our goals, adopt our goals, and retain our goals. Superintelligent AI could produce computers that do not share our goals and that might not value the preservation of humanity. Thus we need to agree on questions of the ultimate meaning of life before the computers answer it for us. The final question Tegmark raises is the hardest question of all: what is consciousness, and will machines ever feel it? There are three levels to this problem. The hard problem is what physical properties distinguish conscious and unconscious systems? The harder problem is what physical properties determine qualia like colors, sounds, emotions, and self-consciousness? The hardest problem is why is anything conscious? Maybe the latter two questions can never be answered by science. Nevertheless, since machines may become much more intelligent than we are, we should stop thinking of our species as *homo sapiens* (wise men) and start thinking of ourselves as *homo sentiens* (feeling men).

In the epilogue, Tegmark tells the story of his Future of Life Institute, which brings together people who work in AI so they can figure out how to make AI benefit humanity.

Materialistic Presuppositions. Max Tegmark tells the reader nothing of his personal religious beliefs, but he writes his book from the perspective of metaphysical naturalism; nothing exists except energy, matter, time, and space. There is no God who created the world or who presides over the affairs of the universe. No one knows how the Big Bang happened, how life arose, or how human consciousness developed. Since one brain neuron cannot think, but a billion can, thinking must be an emergent property of matter, and thinking must be able to emerge from a manufactured brain as well as from a biological one. There is no clear basis for free will, although Tegmark clearly thinks we humans have it and need to use it wisely. Maybe free will is an emergent property of matter, like thinking. If so, then machines may someday have free wills.

Tegmark does not deal with the implications of Kurt Godel's incompleteness theorem that led Alan Turing to see that since all mathematical systems are incomplete, the machines that implement those systems must also be incomplete as well. Turing showed that even a "Turing Machine," an imaginary computer of infinite speed and capacity, that ran forever would not be able to solve some problems that people easily could solve. Oxford professors John Lucas and Roger Penrose applied Turing's conclusions to the study of the human mind and proved that the human mind is not a computer. They also demonstrated that computers will never be able to think the way people do. While Tegmark mentions Turing several times, he never treats this aspect of his thought, nor does he discuss the findings of Lucas and Penrose.[1]

Working from Christian presuppositions and believing the implications of Godel's incompleteness theorem, readers will find Tegmark's work fascinating but not scary. Christians need to think about how to use God's good gift of AI wisely. Tegmark has no reason to believe that humans have free will, no reason to trust that they will not wipe themselves off the Earth in the next ten thousand years (as he shows on page 196), and no assurance that computers will not develop a malevolent consciousness and extinguish all mankind. Yet he continues to smile at us from the book jacket and make saving humanity his life's work. Would that someone from Park Street Church or another Boston congregation could share with him the reason for the hope that is in us.

— *Charles Edward White*

Charles Edward White, PhD (church history, Boston University), is professor of Christian thought and history at Spring Arbor University

NOTES

1   For more information, see my article, "Who's Afraid of HAL? Why Computers Will Not Become Conscious and Take Over the World," *Christian Research Journal* 39, 6 (2016): 10–15.