

Will Hominoids or Androids Destroy the Earth? —A Review of How to Create a Mind by Ray Kurzweil (2012)

Michael Starks

#### ABSTRACT

Some years ago I reached the point where I can usually tell from the title of a book, or at least from the chapter titles, what kinds of philosophical mistakes will be made and how frequently. In the case of nominally scientific works these may be largely restricted to certain chapters which wax philosophical or try to draw general conclusions about the meaning or long term significance of the work. Normally however the scientific matters of fact are generously interlarded with philosophical gibberish as to what these facts mean. The clear distinctions which Wittgenstein described some 80 years ago between scientific matters and their descriptions by various language games are rarely taken into consideration, and so one is alternately wowed by the science and dismayed by its incoherent analysis. So it is with this volume.

If one is to create a mind more or less like ours, one needs to have a logical structure for rationality and an understanding of the two systems of thought (dual process theory). If one is to philosophize about this, one needs to understand the distinction between scientific issues of fact and the philosophical issue of how language works in the context at issue, and of how to avoid the pitfalls of reductionism and scientism, but Kurzweil, like nearly all students of behavior, is largely clueless. He, is enchanted by models, theories, and concepts, and the urge to explain, while Wittgenstein showed us that we only need to describe, and that theories, concepts etc., are just ways of using language (language games) which have value only insofar as they have a clear test (clear truthmakers, or as John Searle (AI's most famous critic) likes to say, clear Conditions of Satisfaction (COS)). I have attempted to provide a start on this in my recent writings, such as *The Logical Structure of Consciousness* (behavior, personality, rationality, higher order thought, intentionality) (2016) and *The Logical Structure of Philosophy, Psychology, Mind and Language as Revealed in the Writings of Ludwig Wittgenstein and John Searle* (2016). Those interested in all my writings in their most recent versions may consult my e-book *Philosophy, Human Nature and the Collapse of Civilization - Articles and Reviews 2006-2016* 662p (2016). I will give a very brief presentation of this framework since I have described it in great detail in many recent papers and several books, available on this site and others.

Also, as usual in 'factual' accounts of AI/robotics, he gives no time to the very real threats to our privacy, safety and even survival from the increasing 'androidizing' of society which is prominent in other authors (Bostrum, Hawking etc.) and frequent in scifi and films, so I make a few comments on the quite possibly suicidal utopian delusions of 'nice' androids, humanoids, democracy, diversity, and genetic engineering.

I take it for granted that technical advances in electronics, robotics and AI will occur, resulting in profound changes in society. However, I think the changes coming from genetic engineering are at least as great and potentially far greater, as they will enable us to utterly change who we are. And it will be feasible to make supersmart/super strong servants by modifying our genes or those of other monkeys. As with other technology, any country that resists will be left behind. But will it be socially and

economically feasible to implement biobots or superhumans on a massive scale? And even if so, it does not seem remotely possible, economically or socially to prevent the collapse of industrial civilization.

So, ignoring the philosophical mistakes in this volume as irrelevant, and directing our attention only to the science, what we have here is another suicidal utopian delusion rooted in a failure to grasp basic biology, psychology and human ecology, the same delusions that are destroying America and the world. I see a remote possibility the world can be saved, but not by AI/robotics, CRISPR, nor by democracy and equality.

Those who wish to read all my articles please consult the ebook *Philosophy, Human Nature and the Collapse of Civilization -- Articles and Reviews 2006-2016* by Michael Starks 3<sup>rd</sup> Ed 675p (2017)

Some years ago I reached the point where I can usually tell from the title of a book, or at least from the chapter titles, what kinds of philosophical mistakes will be made and how frequently. In the case of nominally scientific works these may be largely restricted to certain chapters which wax philosophical or try to draw general conclusions about the meaning or long term significance of the work. Normally however the scientific matters of fact are generously interlarded with philosophical gibberish as to what these facts mean. The clear distinctions which Wittgenstein described some 80 years ago between scientific matters and their descriptions by various language games are rarely taken into consideration, and so one is alternately wowed by the science and dismayed by its incoherent analysis. So it is with this volume.

If one is to create a mind more or less like ours, one needs to have a logical structure for rationality and an understanding of the two systems of thought (dual process theory). If one is to philosophize about this, one needs to understand the distinction between scientific issues of fact and the philosophical issue of how language works in the context at issue, and of how to avoid the pitfalls of reductionism and scientism, but Kurzweil, like nearly all students of behavior, is largely clueless. He, is enchanted by models, theories, and concepts, and the urge to explain, while Wittgenstein showed us that we only need to describe, and that theories, concepts etc., are just ways of using language (language games) which have value only insofar as they have a clear test (clear truthmakers, or as John Searle (AI's most famous critic) likes to say, clear Conditions of Satisfaction (COS)). I have attempted to provide a start on this in my recent writings, such as *The Logical Structure of Consciousness (behavior, personality, rationality, higher order thought, intentionality)* (2016) and *The Logical Structure of Philosophy, Psychology, Mind and Language as Revealed in the Writings of Ludwig Wittgenstein and John Searle* (2016). Those interested in all my writings in their most recent versions may consult my e-book *Philosophy, Human Nature and the Collapse of Civilization - Articles and Reviews 2006-2016* 2nd Ed 675p (2017). I will give a very brief presentation of this framework since I have described it in great detail in many recent papers and several books, available on this site and others.

Actually, “reduction” is a complex language game or group of games (uses of words with various meanings or COS) so its use varies greatly depending on context and often it’s not clear what it means. Likewise with modeling or simulating or reproducing or equivalent to or the same as etc. Likewise with the claims here and everywhere that “computation” of biological or mental processes is not done as it would take too long but not computable or calculable means many things or nothing at all depending on context and this is usually just totally ignored.

Chapter 9 is the typical nightmare one expects. Minsky’s first quote “Minds are simply what brains do” is a truism in that in some games one can e.g., say ‘my brain is tired’ etc. but like most he has no grasp at all of the line between scientific questions and those about how the language games are to be played (how we can use language intelligibly). Descriptions of behavior are not the same as descriptions of brain processes. This ‘reductionism’ is a hopelessly bankrupt view of life, -- it just does not work, i.e., is not coherent, and this has been explained at length, first by Wittgenstein and subsequently by Searle, Hacker and many others. For one thing, there are various levels of description (physics, chemistry, biochemistry, genetics, neurophysiology, brain, thought/behavior) and the concepts (language games) useful and intelligible (having clear meaning or COS) at one level work differently at another. Also one ‘mental state’, ‘disposition’ or ‘thought’ or ‘action’, can be described in first person or third person by many statements and vice versa, one statement may describe many different ‘mental states’, ‘dispositions’, ‘thoughts’ or ‘actions’ depending intricately on context, so the match between behavior and language is hugely underdetermined even for ‘simple’ acts or sentences. and as these become more complex there is a combinatorial explosion.

There is no clear meaning to describing my desire to see the sun set at the lower levels and their never will be. They are different levels of description, different concepts (different language games) and one cannot even make sense of reducing one to the other, of behavior into neurophysiology into biochemistry into genetics into chemistry into physics into math or computation and like most scientists Kurzweil’s handwaving and claims that it’s not done because its inconvenient or impractical totally fails to see that the real issue is that ‘reduction’ has no clear meaning (COS), or rather many meanings depending acutely on context, and in no case can we give a coherent account that eliminates any level.

Nevertheless, the rotting corpse of reductionism floats to the surface frequently (e.g., p37 and the Minsky quote on p199) and we are told that chemistry “reduces” to physics and that thermodynamics is a separate science because the equations become “unwieldy”, but another way to say this is that reduction is incoherent, the language games (concepts) of one level just do not apply (make sense) at higher and lower levels of description, and it is not that our science or our language is inadequate. I have discussed this in my other articles and it is well known in the philosophy of science, but it is likely never going to penetrate into “hard science”.

The psychology of higher order thought is not describable by causes, but by reasons, and one cannot make psychology disappear into physiology nor physiology into biochemistry nor it into physics etc. They are just different and indispensable levels of description. Wittgenstein famously described it 80 years ago in the Blue Book.

“Our craving for generality has [as one] source ... our preoccupation with the method of science. I mean the method of reducing the explanation of natural phenomena to the smallest possible number of primitive natural laws; and, in mathematics, of unifying the treatment of different topics by using a generalization. Philosophers constantly see the method of science before their eyes, and are irresistibly tempted to ask and answer in the way science does. This tendency is the real source of metaphysics, and leads the philosopher into complete darkness. I want to say here that it can never be our job to reduce anything to anything, or to explain anything. Philosophy really is “purely descriptive.”

Like nearly all ‘hard’ scientists and even sadly ‘soft’ ones as well, he has no grasp at all of how language works, e.g., of how ‘thinking’ and other psychological verbs work, so misuses them constantly throughout his writings (e.g., see his comments on Searle on p170). I won’t go into an explanation here as I have written extensively on this (see my recent ebook *Philosophy, Human Nature and the Collapse of Civilization -- Articles and Reviews 2006-2017* by Michael Starks 3rd Ed. 675p (2017)). So, like most scientists, and even most philosophers, he plays one language game (uses the words with one meaning or Condition of Satisfaction) but mixes it up with other quite different meanings, all the while insisting that his game is the only one that can be played (has any ‘real’ sense). Like most, he also is not clear on the distinction between scientific issues of fact and the issues of how language can be used intelligibly. Also he does not have a clear grasp of the distinction between the two systems of thought, the automaticities of nonlinguistic system S1 and the conscious deliberations of linguistic system S2 but I have described this extensively in my writings and will not do so here.

Another thing that Kurzweil never mentions is the obvious fact that there will be severe and probably frequently fatal conflicts with our robots. Just think about the continual daily problems we have living with other humans, about the number of assaults, abuses and murders every day. Why should these be any less with androids--and then who takes the blame? There would not seem to be any reason at all why androids should be less in conflict with each other, and with us, than other humans are already. Asimov’s law of robotics –do not harm humans, is a fantasy that is unattainable in practice for androids just as it is for us. I admit (as Searle has many times) that we are ‘androids’ too, though designed by natural selection, not having ‘intelligence’ from one viewpoint, but having limitless ‘intelligence’ from another.

What is to stop androids having all the mental ailments we have—neuroses, psychoses, sociopathies, egomania, greed, selfish desire to produce endless copies of one’s own ‘genome’ (electrome, digitome, silicome?), racism (programism?), something equivalent to drug addiction, homicidal and suicidal tendencies? Of course humans will try to exclude bad behavior from the programs but this will have to be after the fact, i.e., when it’s already dispersed, and as they will be self programming and updating, any badness that confers a survival advantage will spread rapidly. This is of course just the android equivalent of humanoid evolution by natural selection (inclusive fitness).

John Searle killed the idea of strong AI with the Chinese room and other descriptions of the incoherence of various language games (as Wittgenstein had done superbly long before there were computers, though few have noticed). He is regarded by some as the nemesis of AI, but in fact he has just kept it on track and has no antipathy to it at all. Searle has said repeatedly that of course machines can think and

feel, for we are such machines! Made of proteins etc., and not metal, but machines in a very fundamental sense nevertheless. And machines that took about 4 billion years of experimentation in a lab the size of the earth with trillions or trillions of machines being created and only a tiny number of the most successful surviving. The efforts of AI seem or at least robotics, seem trivial by comparison. And as he notes it is possible that much or all of our psychology may be unique to fleshy beings, just as much of AI may be to solid state androids. How much might be true overlap and how much vague simulation is impossible to say.

Darwinian selection or survival of the fittest as it applies to machines, is a major issue that is never addressed by Kurzweil, nor most others, but is the subject of a whole book by philosopher-scientist Nik Bostrum and of repeated warnings by black hole physicist and world's longest surviving ALS sufferer Stephen Hawking. Natural selection is mostly equivalent to inclusive fitness or favoritism towards close relatives (kin selection). And there is no countervailing 'group selection' for 'niceness' (see my review of Wilson's *The Social Conquest of Earth*(2012)). Yes we do not have DNA and genes in robots (yet), but in what is perhaps philosopher Daniel Dennett's most (only?) substantive contribution to philosophy, it is useful to regard inclusive fitness as the 'universal acid' which eats through all fantasies about evolution, nature and society. So, any self-replicating android or program that has even the slightest advantage over others will automatically eliminate them and humans and all other lifeforms, protein or metal, that are competitors for resources, or just for 'amusement' as humans do with other animals. Exactly what will prevent programs from evolving selfishness and replacing all other competing machines/programs or biological life forms? If one takes the 'singularity' seriously, then why not take this just a seriously? I commented on this a long ago and of course it is a staple of science fiction. So AI is just the next stage of natural selection with humans speeding it up in certain directions until they are replaced by their creations, just as the advantages in our 'program' resulted in the extinction of all other hominoid subspecies.

As usual in 'factual' accounts of AI/robotics, Kurzweil gives no time to the very real threats to our privacy, safety and even survival from the increasing 'androidizing' of society which are prominent in other nonfiction authors (Bostrum, Hawking etc.) and frequent in scifi and films. It requires little imagination to see this book as just another suicidal utopian delusion concentrating on the 'nice' aspects of androids, humanoids, democracy, computers, technology, ethnic diversity, and genetic engineering. It is however thanks to these that the last vestiges of our stability/privacy/security/prosperity/tranquility/sanity are rapidly disappearing. Also, drones and autonomous vehicles are rapidly increasing in capabilities and dropping in cost, so it will not be long before enhanced AI versions are used for crime, surveillance and espionage by all levels of government, terrorists, thieves, stalkers, kidnappers and murderers. Given your photo, fingerprints, name, workplace, address, mobile phone #, emails and chats, all increasingly easy to get, solar powered or self-charging drones, microbots, and vehicles will be able carry out almost any kind of crime. Intelligent viruses will continue to invade your phone, pc, tablet, refrigerator, car, TV, music player, health monitors, androids and security systems to steal your data, monitor your activities, follow you, and if desired, extort, kidnap or kill you. Its crystal clear that if the positives will happen then the negatives will also. This dark side of AI/Robotics/The internet of things goes unmentioned in this book, and this is the norm.

Though the idea of robots taking over has been in scifi for many years, I first started to think seriously about it when I read about nanobots in Drexler's Engines of Creation in 1993. And many have worried about the 'grey goo' problem—i.e., of nanobots replicating until they smother everything else.

Another singularity that Kurzweil and most in AI do not mention is the possibility that genetic engineering will soon lead to DNA displacing silicon as the medium for advanced intelligence. CRISPR and other techniques will let us change genes at will, adding whole new genes/chromosomes in months or even hours, with superfast development of organisms or brains in vats without bothersome bodies to encumber them. Even now, without genetic engineering, there are precocious geniuses mastering quantum mechanics in their early teens or taking the cube of a 10 digit number in their head. And the programming of genes might be done by the same computers and programs being used for AI.

Anyone who takes AI seriously also might find of interest my article on David Wolpert's work on the ultimate law in Turing Machine Theory which suggests some remarkable facets of and limits to computation and 'intelligence'. I wrote it because his work has somehow escaped the attention of the entire scientific community. It is readily available on the net and in my book: Wolpert, Godel, Chaitin and Wittgenstein on impossibility, incompleteness, the liar paradox, theism, the limits of computation, a nonquantum mechanical uncertainty principle and the universe as computer—the ultimate theorem in Turing Machine Theory (2015).

To his credit, Kurzweil makes an effort to understand Wittgenstein (p220 etc.), but (like 50 million other academics) has only a superficial grasp of what he did. Before computers existed Wittgenstein discussed in depth the basic issues of what computation was and what makes humans distinct from machines, but his writings on this are unknown to most. Gefwert is one of the few to analyze them in detail, but his work has been largely ignored.

On p222 Kurzweil comments that it is 'foolish' to deny the 'physical world' (an intricate language game), but it is rather that one cannot give any sense to such a denial, as it presupposes the intelligibility (reality) of what it denies. This is the ever-present issue of how we make sense of (are certain about) anything, which brings us back to Wittgenstein's famous work 'On Certainty' (see my various reviews of his books) and the notion of the 'true only' proposition. Like all discussions of behavior, Kurzweil's needs a logical structure for rationality (intentionality) and (what is equivalent) a thorough understanding of how language works, but it is almost totally absent. As much of my book deals with these issues I won't go into them here except to provide the summary table of intentionality.

After half a century in oblivion, the nature of consciousness is now the hottest topic in the behavioral sciences and philosophy. Beginning with the pioneering work of Ludwig Wittgenstein in the 1930's (the Blue and Brown Books) to 1951, and from the 50's to the present by his successors Searle, Moyal-Sharrock, Read, Hacker, Stern, Horwich, Winch, Finkelstein etc., I have created the following table as an heuristic for furthering this study. The rows show various aspects or ways of studying and the columns show the involuntary processes and voluntary behaviors comprising the two systems (dual processes) of the Logical Structure of Consciousness (LSC), which can also be regarded as the Logical Structure of Rationality (LSR-Searle), of behavior (LSB), of personality (LSP), of Mind (LSM), of language (LSL), of reality (LSOR), of Intentionality (LSI) -the classical philosophical term, the Descriptive Psychology of

Consciousness (DPC) , the Descriptive Psychology of Thought (DPT) –or better, the Language of the Descriptive Psychology of Thought (LDPT), terms introduced here and in my other very recent writings.

The ideas for this table originated in the work by Wittgenstein, a much simpler table by Searle, and correlates with extensive tables and graphs in the three recent books on Human Nature by P.M.S Hacker. The last 9 rows come principally from decision research by Johnathan St. B.T. Evans and colleagues as revised by myself.

**System 1 is involuntary, reflexive or automated “Rules” R1 while Thinking (Cognition) has no gaps and is voluntary or deliberative “Rules” R2 and Willing (Volition) has 3 gaps (see Searle)**





FROM DECISION RESEARCH								
Subliminal Effects	No	Yes/No	Yes	Yes	No	No	No	Yes/No
Associative/Rule	RB	A/RB	A	A	A/RB	RB	RB	RB
Context Dependent/Abstrac	A	CD/A	CD	CD	CD/A	A	CD/A	CD/A
Serial/Parallel	S	S/P	P	P	S/P	S	S	S
Heuristic/Analytic	A	H/A	H	H	H/A	A	A	A
Needs Working Memory	Yes	No	No	No	No	Yes	Yes	Yes
General Intelligence	Yes	No	No	No	Yes/No	Yes	Yes	Yes
Cognitive Loading Inhibits	Yes	Yes/No	No	No	Yes	Yes	Yes	Yes
Arousal Facilitates or Inhibits	I	F/I	F	F	I	I	I	I

Public Conditions of Satisfaction of S2 are often referred to by Searle and others as COS, Representations, truthmakers or meanings (or COS2 by myself), while the automatic results of S1 are designated as presentations by others ( or COS1 by myself).

\* Aka Inclinations, Capabilities, Preferences, Representations, possible actions etc.

\*\* Searle's Prior Intentions

\*\*\* Searle's Intention In Action

\*\*\*\* Searle's Direction of Fit

\*\*\*\*\* Searle's Direction of Causation

\*\*\*\*\* (Mental State instantiates--Causes or Fulfills Itself). Searle formerly called this causally self-referential.

\*\*\*\*\* Tversky/Kahneman/Frederick/Evans/Stanovich defined cognitive systems.

\*\*\*\*\* Here and Now or There and Then

It is of interest to compare this with the various tables and charts in Peter Hacker's recent 3 volumes on Human Nature. One should always keep in mind Wittgenstein's discovery that after we have *described*

the possible uses (meanings, truthmakers, Conditions of Satisfaction) of language in a particular context, we have exhausted its interest, and attempts at *explanation* (i.e., philosophy) only get us further away from the truth. He showed us that there is only one philosophical problem—the use of sentences (language games) in an inappropriate context, and hence only one solution— showing the correct context.

On p 278 he comments on our improving life and references ‘Abundance’ by his colleague Diaminidis – another utopian fantasy, and mentions Pinker’s recent work “The Better Angels of Our Nature: Why Violence Has Declined”, but fails to note that these improvements are only temporary, and are bought at the cost of destroying our descendant’s futures. As I have reviewed Pinker’s book and commented in detail on the coming collapse of America and the world in my articles, I will not repeat it here.

Every day we lose about 200 million tons of topsoil into the sea (ca. 12kg/person/day) and about 20,000 hectares of agricultural land becomes salinified and useless. Fresh water is disappearing in many areas. And every day the mothers of the 3<sup>rd</sup> world (the 1<sup>st</sup> world now decreasing daily) ‘bless’ us with another 300,000 or so babies, leading to a net increase of about 200,000—another Las Vegas every 10 days, another Los Angeles every month. About 4 billion more by 2100, most in Africa, most of the rest in Asia. The famously tolerant Muslims will likely rise from about 1/5<sup>th</sup> to about 1/3 of the earth and control numerous H bombs. Thanks to the social delusions of the few hundred politicians who control it, America’s love affair with ‘diversity’ and ‘democracy’ will guarantee its transformation into a 3<sup>rd</sup> world hellhole and the famously benevolent Chinese will take center stage. Sea level is projected to rise at least one to three meters by 2100 and some projections are ten times higher. There is no doubt at all that it will eventually rise much higher and cover much of the world’s prime cropland and most heavily populated areas. It’s also clear that the oil and natural gas and good quality easy to get coal will be gone, much of the earth stripped of topsoil, all the forests gone, and fishing dramatically reduced. I would like to see a plausible account of how androids/AI will fix this. Even if theoretically possible, at what cost in money and pollution and social distress to created and maintain them? The second law of thermodynamics and the rest of physics, chemistry and economics works for androids as well as hominoids. And who is going to force the world to cooperate when its obvious life is a zero sum game in which your gain is my loss? There is no free lunch. Even if robots could do all human tasks right now it would not save the world from constant international conflicts, starvation, disease, crime, violence and war. When they cannot be made to cooperate in this limited time of abundance (bought by raping the earth) it is hopelessly naïve to suppose that they will do it when anarchy is sweeping over the planet.

I take it for granted that technical advances in electronics, robotics and AI will occur, resulting in profound changes in society. However, I think the changes coming from genetic engineering are at least as great and potentially far greater, as they will enable us to utterly change who we are. And it will be feasible to make supersmart/super strong servants by modifying our genes or those of other monkeys. As with other technology, any country that resists will be left behind. But will it be socially and economically feasible to implement biobots or superhumans on a massive scale? And even if so, it does not seem remotely possible, economically or socially to prevent the collapse of industrial civilization.

So, ignoring the philosophical mistakes in this volume as irrelevant, and directing our attention only to the science, what we have here is another suicidal utopian delusion rooted in a failure to grasp basic biology, psychology and human ecology, the same delusions that are destroying America and the world. I see a remote possibility the world can be saved, but not by AI/robotics,CRISPR, nor by democracy and equality.