Scientia*et*Fides 10(1)/2022 ISSN 2300-7648 (print) / ISSN 2353-5636 (online)

Received: April 10, 2020. Accepted: August 20, 2020 DOI: http://dx.doi.org/10.12775/SetF.2022.008

From Absolute Mind to Zombie: Is Artificial Intelligence Possible?*

MORITZ ERNST MARIA BILAGHER
UNESCO Internal Oversight Service (IOS) Evaluation Office, Paris
m.bilagher@gmail.com
ORCID: 0000-0001-9368-4517

Abstract. The dream of achieving artificial intelligence (AI) and, in particular, artificial consciousness ('strong AI'), is reflected in mythologies and popular culture as utopia and dystopia. This article discusses its conceptual possibility. It first relates the desire to realise strong AI to a self-perception of humanity as opposed to nature, metaphorically represented as gods or God. The realisation of strong AI is perceived as an ultimate victory on nature or God because it represents the crown of creation or evolution: conscious intelligence. The paper proceeds to summarise two debates relevant to AI: one educational and one technological. The technological debate, almost invariably presupposing a materialist framework, is related to the mind–body problem of philosophy; the educational one to understanding the concept of intelligence. By proposing a definition of intelligence linked to an idealist conception of reality, postulating mind as participation in Absolute Mind, I attempt a convergence of these debates, rejecting the possibility of strong AI.

Keywords: computing, religion, consciousness, mythology, intelligence.

* The ideas and opinions expressed in this article are those of the author and do not necessarily represent the view of UNESCO.

Introduction: An enduring fantasy

One of the most enduring fantasies of the human species is the ability to reproduce its own consciousness. With this I refer not to the ability to procreate but the ability to manufacture life: to 'play God'. In terms of mythology to this effect, European culture has the homunculus, which has roots in alchemy and was seemingly first mentioned by Paracelsus (1537); in the same period, slightly later, a Jewish tradition described the creation of a golem; and, in modern times, Mary Shelley wrote the story of Frankenstein (1818). Myths and legends aside, the spectacular growth of technological expertise since the 19th century, and since World War II in particular, has engendered hopes, but also fears, that this fantasy may, one day, materialise. Developments in biotechnology, robotics and computer sciences seem particularly promising and, interestingly, around 10 years ago, one author (Halal 2009) thought that we should approach great achievements in these fields around exactly the present time:

> At about 2020, the very time when the planet is likely to teeter between calamity and salvation, our forecasts suggest that routine human thought should be automated by far more sophisticated IT networks, a second generation of more powerful computers, smart robots that think and talk, and artificial intelligence that approaches human skills.

Although one may question whether this prediction has indeed come about, it is certain that great technological progress has been achieved since 2009. However, while advances in the mentioned fields take hold, one may ask: why would we want to (re-) create conscious life? It seems plausible that the dream of the re-creation of self-consciousness stems from the awareness that life is finite, to which, as Dobzhansky has ventured to assert (1969), much of philosophy is the attempt at an answer. As the mind is considered the crown on creation, or evolution, the recreation of this is of particular interest. The quest for conscious life is, thus, a quest for immortality.

In this essay, I will explore the conceptual possibility that the pursuit of (re-) creating the human mind, Artificial Intelligence (AI), can be realised. I will do so by first relating the quest for AI to a fundamentally religious dimension, which sees human beings searching for the 'tree of life' mentioned in the Biblical book of Genesis, symbolising immortality. I will then proceed to summarise the current state of the debate around AI, asserting that there are two strands relevant to this, rather than only one: the technological debate, but also the educational and psychological one, on the concept of intelligence. I argue that, whereas these debates could inform one another, there seems to be little interaction between them. I will state that to place the discussion on AI on firm ground, a solid definition of intelligence is needed, which is currently lacking. I will propose such a definition and, finally, explain how my understanding of intelligence as rooted in a divine, Absolute Mind, precludes the realisation of what is known as 'strong AI', i.e. intelligence that has human properties, such as consciousness.

1. Man, nature and technology

It is not surprising that the common fantasy of striving towards religious dimensions, such as eternity and immortality, finds its most ardent hopes in the growth of technological knowledge. Technology has, after all, realised many things that seemed unattainable, akin to the miracle in religion: the abilities to fly (aviation), to capture visual material (photography) and communicate with people at a distance (telecommunication) all previously belonged to the realm of magic. This has placed technology, and science, with which it is allied (Kwa 2005), in a polemical relation with religion.[1] Claims that events in the Bible presented literal facts, as in Ussher's calculation of the date of Creation, which subsequent study made unlikely, led to an antagonism that contends that if one is right (say, religion), the other side must be wrong (science). One of the best exam-

[1] Although, according to Brooke (2006, 293), the view that science and religion have "existed in more or less perpetual warfare … is problematic because of the level of generality assumed."

ples of this may be Darwin's theory of evolution, which directly opposes a literal interpretation of the Story of Creation. It is no coincidence that biology is the science of life, and thus probably also not a coincidence that some of the most vocal opponents of religion, among which Dawkins, Wilson and Harris are, or were, biologists.

If we wish to understand the deeper reason between this antagonism and, in connection with this, the reason that technology is at times perceived as the strongest stake towards immortality, we have to review the relation we as human beings have with our environment: nature. In this respect, the 16th and 17th century Dominican friar and Italian philosopher Campanella's identification of nature with God (Manuel and Manuel 1982)[2] is enlightening; not as an ontological statement, but as an anthropological one. By this, I mean that God Himself is not by necessity nature, but that what or who is commonly understood to be God is curiously similar to the whole of events, mechanisms and states of affairs that we refer to as nature. This seems most directly clear in the deities of worldwide religions: in the old Greek pantheon, Neptune is the sea, Zeus lightning, Gaia the earth; among Norse-Saxon gods, Thor or Thuner is the god of thunder; and of Japan's Shinto deities, Amaterasu represented the sun, Tsukiyomi the moon and Susano the sea. Even in Judaism, Spinoza asserted, wind (*ruagh*, commonly translated as 'spirit') is thought to come from Yahweh (1670/1862).

But the above is also true implicitly: for example, we notice historically that the weaker nature becomes, the weaker religion becomes, too. The more nature has been understood by scientists, in astronomy (Copernicus, Galileo), geology (Lyell, Hutton) and biology (Darwin); and the more technology has allowed humanity to control the forces of nature, the less the need has been felt for proxies such as ceremony or ritual to communicate with, appease or even control the collection of forces that are indicated with these names (nature and God). Nowadays, nature has become so weak that it needs an ecological movement to protect it;[3] almost like religion.

---

2 Campanella declared, according to 'terror-struck witnesses': "There is no God. There is only nature which we call God." Over 50 years later, Spinoza would formulate similar thoughts.

3 Although, by way of climate change, it may still turn on us – as a result of precisely the type of hubris that will be discussed in a following paragraph.

2. Hubris: Climbing Mount Olympus

The subjugation of 'external nature' through rationalisation, symbolised by its transformation from earth (the objective terrain) to land (a means of economic production), described by Marx (1909), may not have an obvious parallel in 'internal nature'. In his autobiography, Darwin compared the need for religion with the fear and hatred of a monkey toward a snake (1887/1958).[4] The myth of Prometheus, in Greek mythology, where this hero is punished for the crime of stealing fire from the gods may reflect such a fear or lack of rationalisation. The fire stands for control over nature (MacGregor 2008, 6), which technology is understood to represent; the gods of the Greek pantheon for nature itself. To convey this idea, that human beings would want to 'play God', the Greeks had the concept of *hubris*, which is clearest in Plato's myth of humanity's rebellion against the gods on Mount Olympus. After this rebellion failed, they were cut in half, each eternally yearning for the other half (380 BCE/1995a).

The morale of stories such as these is clear: entering the domain of the gods is associated with great dangers. To the human mind, or at least to that of the Greeks, it is for the gods only to control nature, which they represented, and if a human being would try to do the same, s/he would be punished. What is interesting here is that, according to Kwa, with the configuration of the Greek pantheon, the necessary preconditions for experimental science and technological development were given: the Greek gods are surprisingly human-like. In fact, according to St. Augustine (415 AD/2007), Tullius complained about Homer that he had wished the Greek poet had transferred divine characteristics on human beings rather than portray the gods as human beings. As a consequence, even though human beings often have to pay for their hubris, sometimes they can outwit the gods – which might not be possible in Judaism, for example, where Yahweh is omniscient and omnipotent. Given God's identity with nature, understanding how nature works is the first step towards control over it.

[4] This passage was deleted from the first edition of this autobiography on the request of Darwin's wife Emma, but later restored by his granddaughter.

From a perspective of technology – and of experimental science – nature has to be subjugated; in the Jewish faith, it might have been very difficult to imagine so..

3. Utopias and dystopias

Where Greek mythology emphasises the possibilities of technology (if also its dangers), and therefore, in embryonic form, the potentiality of AI, the Judaic myth of the fall from Paradise symbolises the birth of free will, which seems relevant to any intelligence. To illustrate this, we have to affirm that while, in line with our contemporary scientific understanding of our ancient history (evolution theory), Paradise never existed, the myth refers to something that has. I contend that the myth of Paradise is the product of a confluence of two things: present human consciousness, which we associate with conscious perception and rational thought and a 'pre-human memory' of when we were not yet human. The latter element of this equation is based on Freud's idea that individuals traverse the entire development of the species in condensed fashion (1926/1991) and on Jung's concept of the collective unconscious. That is to say, it seems plausible that our minds, whatever they are, contain coded memories not only of our own, personal experiences, but also of our ancestors. Genes are sometimes mentioned as a concrete example of a 'memory of the species': after all, they remember how to build new members.

Following evolution theory, we do not only have human ancestors, but also pre-human or animal ones whose experiences are likely to have been completely different from our own. We can hypothesise that they lived in a state of innocence. They were subject to instinct and thus, with the absence of choice between good and evil, the choice which Dostoyevsky's great inquisitor described as a horrible burden (1880/1958), probably did not yet know what we call guilt. With the transfer from the animal to a human state, our ancestors acquired free will, and the human being became what Heinisch (1960) called a *Doppelwesen* (ambiguous being), with its intrinsic uncertainties, indetermination and doubts.

Paradise is the conscious memory of an (non-conscious) innocence that has gone. With the fall from Paradise, humanity acquired the knowledge of good and evil, which describes the rational, conscious and intelligent human mind, and this was therefore not really a fall, but an ascent.[5] However, the pursuit of knowledge, or philosophy, required more than only the *ability* to know or reason; it also required a *reason* to trigger its existence. I consequently argue that the other crucial requirement for thinking is evil or suffering; the notion that there is an imperfection to be mended (and this is why the fall is understood as a fall, not an ascent). Only the presence of evil can lead to the fundamental question: why? And, with that, to the utopian tradition of Plato's Republic, More's Utopia and Marx' socialist society (all attempts to re-establish original perfection). However, in contrast to Paradise, or its Christian mirror image, the Kingdom of Heaven, these are utopias that can be brought about by human hands, which Mumford called utopias of reconstruction (1922/1959) rather than utopias of escape, which can only be achieved through divine intervention. This line of demarcation runs parallel to the technological principle: the notion that humanity can control the natural order rather than be controlled by the gods of nature.

It should now be mentioned that in Paradise there was not only a tree of knowledge; there was also a tree of life. This is significant in that, when the first humans fall from grace in this myth, it is announced that they will only be redeemed when this other tree is found. In the Christian Bible, it is also mentioned in the Book of Revelation (22:2), where it is said to be found in the utopia of New Jerusalem. The historical developments we are now witnessing may be those which perceive technology, and AI in particular, as a tree of life.

---

[5] While for several theologians, such as Augustine, Thomas Aquinas and even Tillich the Fall was certainly a fall, Fromm (1984) seems to have shared the view that the Fall may have been an ascent: "Acting against the command of authority, committing a sin, is in its positive human aspect the first act of freedom, that is, the first human act. […] The act of disobedience as an act of freedom is the beginning of reason."

4. Mind as product

If we indeed conceive of the history of humanity as the story of our relation with the forces of nature, of rationality and randomness, then what becomes almost instantly clear is that what distinguishes these two is mind. The human mind is placed against animal instinct; or the directed intelligence of humanity against the blind force of nature. This distinction is clear in Kant's description of nature as everything that operates under laws (1788/1996) as opposed to the human mind, which, paradoxically, attains freedom in its submission to the ethical law, that is, the knowledge that one thing is right and another wrong, because of its ability to make a wrong and therefore also a right choice. It is this freedom of choice that subjects human to the ethical law and, with that, makes human holy and an "end in itself" (id., 210). The distinction inherent in this, between nature and humanity, is sharply drawn in Bacon's metaphor for the scientific experiment, of nature as a slave that can be put to the rack in order to extract her secrets (Kwa 1991).

The combination of these two states of affairs, that humanity uses technology to impose its will on nature, on the one hand, and the definition of everything without mind as nature on the other, has made a new possibility visible: the use of technology to create a mind, or its product(s): intelligence and consciousness. This is a dream of AI, which is mainly pursued in Computer Sciences. This is not to say that no other disciplines are involved in this enterprise: in the quest for AI, there has been a realisation that it may be the human set-up that conditions our minds; that consciousness is related to the fact that we have bodies. For this reason, over 10 years ago, a robotics research team at the University of Essex explored the 'Zombie enigma': "if you build a machine that is functionally identical to a human … would it be aware, in the way that we associate with being alive?" (Akass 2007, 20). In a 2012 article, the project leaders admitted that "the goal of consciousness was not reached within the project" (Diamond et al.). This was not the only claim to machine consciousness (Bishop 2018): "For example, in 2002, Kevin Warwick announced his 'Cybernetic learning robots' to be 'as conscious as a slug' (Warwick 2002)."

However, to do justice to the Janus-like nature of AI (i.e. combining the imposition of human will on nature, through technology, with a technological mind opposed to a natural one), the discussion of this phenomenon may need to consider two debates, which are quite distinct. It would first be necessary to understand what intelligence is. This question is central to the educational sciences. Education addresses the subjects of teaching and learning, including what is to be learnt and how to know whether progress has been made towards learning goals (educational assessment). Intelligence tends to be regarded as one of several factors explaining learning achievement. While the subject of intelligence has traditionally been controversial in the social and behavioural sciences, only once this is addressed can one debate whether a machine could acquire such intelligence. This has, until now, mainly been a technological debate.

5. Artificial intelligence – the state of affairs

It is commonly understood that the debate on artificial intelligence via technology began after the Second World War (Warnick 2004). One of its main points of reference is the Turing Test, devised by the known mathematician. This test postulates that if there is a person in one room and a computer in another and an observer, upon conversing with both but without seeing either, cannot decide who the person is and who the computer, the computer must be understood to be intelligent (Turing 1950). As one may expect, this argument has been strongly criticised, for example by Searle (1980), who countered the Turing test with his Chinese room argument. This gives the example that if an English-speaking person in a room is given sets of cards with Chinese symbols (one with questions, the other with answers) and English instructions for correlating these, when posed questions in Chinese it would seem as if s/he understood Chinese. In reality, s/he is only following instructions and relating pre-set answers to a number of questions. The Turing test would therefore fail as a sufficient criterion for AI, as its conditions could be met while there is no real intelligence, like the person in Searle's example does not really speak Chinese.

To clarify what precisely is missing in the Turing Test criterion, Searle makes a distinction between 'strong AI' and 'narrow AI' (1980), which has since dominated the debate on the subject (see e.g. Warnick 2004) and is also reflected in Chalmers' 'easy' and 'hard' questions around the nature of consciousness (1995). The difference between the two is that narrow, or weak, AI refers to the extent to which computers can perform tasks for which human beings need some intelligence, such as playing chess (and doing it well, as Deep Blue's 1997 defeat of Kasparov suggests), performing calculations or retrieving information. To Searle, it is uncontroversial that this has been achieved and, as inhabitants of the 21st century, we can witness its ubiquity in smartphone-embedded speech assistants, in tailored advertising based on 'cookies' and face recognition.

Strong AI, on the other hand, refers to the extent to which a machine-based entity has an understanding of the procedures it is executing. A computer can be programmed to perform operations, and some of these may seem similar to operations of the human mind, but from the point of view of strong AI, this is not intelligent behaviour as the computer is not aware of this. That is, the computer does not have a phenomenal internal world or 'qualia'. These problems in recreating consciousness arise because it is not known how physiological processes are related to internal representations or subjective experience, which is known as the 'binding problem'.

6. Intelligence and consciousness

Because of the problems encountered in AI in the late 1980s, including the binding problem, this research agenda is thought to have experienced a crisis. This was characterised by the realisation that we cannot essentially understand intelligence apart from consciousness, which led to the emergence of Consciousness Studies in the 1990s, mainly under the leadership of David Chalmers. AI is still one of the most important subjects of Consciousness Studies, but its main question has since arguably been rephrased from 'can a machine exhibit intelligent behaviour' to 'can a machine be intelligent?' In this, it would be hard not to see intelligence as awareness.

The dream of AI and the initial optimism in the discipline (Searle 1991) reveal an underlying notion of what intelligence and the mind are: the mind in AI is, essentially, the brain; and intelligence or conscious perception (often used interchangeably in this debate[6]) the product of this mind. Or, as Dembski noted (2001, in Warnick 2004, 165): "whereas the goal of neuroscience is to reduce intelligent agency to neurophysiology, the goal of AI is to reduce it to computation". The almost inevitable consequence of this line of thought is the idea that the brain is a sophisticated machine and the rhetoric that compares mind with hardware and thoughts with software (Searle 1991), frequent. Even though such rhetoric suggests that we have an understanding of what AI is or how it might work, Chalmers objected that, for all the sophistication involved in responses rooted in neuroscience, they do not respond to essential questions that AI raises. Saying that awareness consists of "35–75 hertz neural oscillations in the cerebral cortex" (Chalmers 1995, 204) do not explain how they instil perception as: "The really hard problem of consciousness is the problem of experience. When we think and perceive, there is a whir of information-processing, but there is also a subjective aspect" (id., 201). The description, in summary, is not an explanation.

The question implicit in Chalmers hard problem refers to a very old problem in philosophy: the mind-body problem (Searle 1991). This is the question how universals (abstract and mind-related entities) relate to particulars (concrete and possibly bodily instances), without which, as Rorty said, we would not have had our 2,500 years' history of philosophy (1980). In this case, the question is how something mechanical and technological can be connected with subjective mental impressions and awareness and, thus, consciousness and, ultimately, intelligence. Therefore by trying to explain consciousness, a pursuit that arose to an extent from the AI-programme, in turn engendered by the successes of technology in the second half of last century, Chalmers *cum suis* returned, almost circularly, to the great questions of philosophy. Maybe this is not surpris-

[6] In itself, this equation may not be without merit, but for Schopenhauer (1819/1996), for example, they are different and referred to as '*Verstand*' (perception) and '*Vernunft*' (intelligence).

ing as even one of most ardent advocates of strong AI, Kurzweil, argued that to understand consciousness "there is a critical role for philosophy, which we sometimes call religion" (2001, in Warnick 2004).

In the meantime, even to date, there seems to be a consensus that we have not come closer to strong AI. As Reggia (2013) concluded in an extensive review: "no existing approach to artificial consciousness has presented a compelling demonstration of phenomenal machine consciousness, or even clear evidence that artificial phenomenal consciousness will eventually be possible." Graziano and Kastner (2011) argued that while all the approaches to explaining consciousness that they discussed recognise that "the content of consciousness includes a great complexity of interlinked information … none … explain how it is that we become aware of that information." Finally, Fjelland (2020) noted that "although development of artificial intelligence for specific purposes (ANI) has been impressive, we have not come much closer to developing artificial general intelligence (AGI)."

7. The concept of intelligence

If the question of intelligence leads to the question of consciousness, and that to the mind-body problem, then it seems crucial to ask: what is intelligence? Only once we will have established this, after all, can we devise criteria for machines to test whether they are, in fact, intelligent. In this regard it is surprising that the advocates of Consciousness Studies have not turned to the educational sciences, in spite of regular references to learning in the AI discourse, for example in relation to DeepMind, which was made to 'learn' how to diagnose kidney disease, which has largely been considered to have failed (Marcus 2019):

> [It] has been putting most of its eggs in one basket, a technique known as deep reinforcement learning. … In some ways, deep reinforcement learning is a kind of turbocharged memorization; systems that use it are capable of awesome feats, but they have only a shallow understanding of what they are doing. As a consequence, current systems lack flexibility, and thus are unable to compensate if the world changes, sometimes even in tiny ways.

One of the main sources for understanding intelligence in education derives from a taxonomy of educational objectives by Bloom et al. Rather than giving a definition of intelligence, it identified a set of learning tasks that range from simple, lower order tasks (memorisation and reproduction) via intermediate level tasks (application) to complex tasks, such as analysis. This betrays a hierarchical structure as the complex tasks are considered more difficult (1971) and thus, by implication, more indicative of intelligence. Bloom's taxonomy, as this framework came to be known, identified three domains of learning: the cognitive, affective and psychomotor skills, related to the coordination of bodily movements.

This system, like any other, has not proved beyond criticism. Howard Gardner argued that the notion of intelligence in education was too narrowly defined, as the stress was often on cognitive intelligence. This was certainly true of Bloom's taxonomy: the cognitive domain was by far the most elaborate, and the third volume of the taxonomy, related to psychomotor skills, was never published. Gardner stated that there are many more ways in which people can generally be considered to be 'capable' and 'intelligent' (1993). He identified multiple intelligences, including the rhythmical-musical, visio-spatial and interpersonal. The idea that intelligence is not a monolithic concept was taken further by others, to the extent that socio-emotional learning (SEL) is now widely considered an important agenda. It can also be interpreted as a response to the allegedly narrow definition of IQ as a measure of intelligence. Already in 1996, the American Psychological Association (APA) conceded that "standardized tests do not sample all forms of intelligence," such as "creativity, wisdom" and "practical sense."

The almost intrinsic controversy around the subject of intelligence was illustrated, some 10 years ago, by the scandal around revelations that Watson, co-discoverer of the human DNA-structure (the 'double helix'), had said that intelligence varies with ethnic background (Hunt-Grubbe 2007). Controversially, similar statements had been made earlier in a declaration of a conglomerate of intelligence researchers in the Wall Street Journal (Arvey et al. 1994). What statements such as these lacked, is a clear idea of what intelligence is. While the signatories of the declaration, and Watson,

relied on IQ as measure of intelligence, they also admitted that the concept of intelligence stands for "a very general mental capability" that "involves the ability to reason, plan, solve problems, think abstractly", and so on, for a full paragraph. The multiplicity of things that intelligence stands for in this definition does not support its credibility. It seems plausible, as the group further states, that intelligence "reflects a broader and deeper capability for comprehending our surroundings" but, if nothing else, this approximation of a definition can hardly be a criterion that can help us establish the possibility of strong AI.

Encouraged by the re-emergence of philosophical questions in the thinking around AI, for a definitive answer to the question what generic intelligence[7] is, I believe we have to turn to Plato. Plato said, in the Republic, that the characteristic of a real philosopher is that s/he has the ability to see unity in manifoldness (370 BCE/1995b). This sentence sums up the essence of intelligence: the ability to discover patterns, see trends, establish links. It rests, essentially, on the operation of *analysis* (disassembly of things, objects or concepts in component parts). This, along with *synthesis* (assembly of things; constructing objects or concepts from their component parts) is probably by no coincidence one of the operations Bloom et al. identified as higher order thinking skills. As secular as this may seem, for Plato, this idea was firmly religious. This can be recognised in that what led Plato to his claim is that there exist two realities: one of sensory perception, or empirical reality, which is illusory, deceptive and manifold; and a real world that is conceptual, abstract, hidden from sensory perception and united. According to Plato, what distinguishes the philosopher is the ability to recognise this second world (unity) in the first one (sensory perception). In a world of events that can seem chaotic, the philosopher sees regularities, laws and meaning.

[7] I do not intend here to invalidate Gardner's idea that there are multiple intelligences, and so this idea of generic intelligence overarches, but does not replace other 'intelligences'.

8. The fundamental problem

However, there is more to Plato's idea. As said, for Plato, the conceptual world is the real world. This is the world that intelligent minds, or philosophers, have access to. By saying this, Plato indicated that the empirical, physical world is only a product (or reflection) of this, as he illustrated with his allegory of the cave. Therefore, what is real to Plato is what is mental, the products of abstraction, not what is tangible, physical and material. Moreover, for Plato, the tangible derives from the abstract. Without the concept of a cat,[8] one could say there would not be actual cats as there would not be objects that could be recognised as cats. Thus, the material world emanates from the conceptual, abstract and immaterial world rather than vice versa. This reflects an essentially idealist position and, in that sense, Plato's ontology varies intrinsically from contemporary philosophical discourse, adopted in AI, which takes materialism (realism) to be the valid position. And it is on this, in my view, mistaken, assumption that the debate around artificial intelligence hinges. The idea that AI is possible is based on the notion that materialism is valid or, as Aydede and Güzeldere (2000, 264) stated: "for the project of AI to have any hopes of accomplishing its grand goal, it has to rely on an entirely materialist framework."

Reggia (2013) argued that artificial consciousness investigators tend to dismiss the idealist perspective "because it leads to solipsism (the belief that only one's own mind is certain to exist), and because viewing the physical world as essentially 'a dream' makes the scientific study of idealism very problematic." One example of such an investigator is Argonov (2014), who argued that for the mind–body problem, philosophy has "two basic alternatives: … 'materialism' … and '(substance) dualism'." The possibility that idealism might be a correct ontology is simply not considered. There exists, however, no philosophical evidence for the assumption that idealism should lead to solipsism. For example, psychoanalysis provides a framework where the mind is collective rather than individual.

[8] Bertrand Russell illustrated Plato's theory with the example of a cat in his History of Western philosophy (1945/2007).

The hopes of AI are thus based on the following image: if it were possible to create networks like that of the brain, replacing the biological elements with inorganic ones, it might be possible to recreate the buzz or algorithm or oscillations that we associate with consciousness (Warnick 2004; Searle 1991). However, even if this were a valid idea, what the exponents of AI still fail to do is explain how this would work – again, a description is not an explanation (even the occasional statement that the existence of the brain is evidence for the possibility of strong AI is incorrect, as it is not artificial). My contention is that such an explanation is intrinsically impossible, because the materialist position is incorrect. Some exponents of Consciousness Studies and AI seem to think in the 'wrong direction': they axiomatically take concrete reality as their starting point and, as a consequence, abstract reality, as its corollary. In reality, though, without awareness of that material world, the material world itself would not exist. It is only via awareness that sensory data are perceived, as Jung observed, and it are sensory data via which images and general impressions of the world are constructed (1982). Trying to reconstruct an abstract world from concrete objects is an inherent impossibility because it is an inversion of the order of things.

9. The thinking of machines

It may be useful to emphasise that the position outlined above indicates that Plato's notion, that the world of unity precedes that of sensory perception, is true in more than only a metaphorical sense. If we agree that reality starts to take shape through the lens of concepts by which we order our sensory impressions, and thus what we see in them is greater, the more we have a grasp of such concepts, we can say that intelligence creates reality. By way of allegory, we can illustrate the difference from computer thinking implicit in this, as follows: images in computing are normally built up of pixels. If a computer 'sees' an image, it starts to read the picture as it were from the 'bottom up', starting with single pixels and inferring from these the existence of an image construed by a multitude of pixels. Human intelligence works exactly the other way around: when

shown an image, the human mind will first see the whole image, as noted in Gestalt psychology, and only then decompose the image in its constituent parts (analyse it). The direction of thinking of machines and human beings is therefore inherently at variance. Computers take bits as their starting points, and think 'upwards' through bytes and kilobytes and so their perception is intrinsically synthetic. Human minds, on the other hand, are intrinsically analytical. The big picture (or concept) comes first, and the constituent parts of it (instances, e.g. objects) later.

Given this state of affairs, it seems inherently impossible to create computers[9] that have the same thinking direction as human beings, and thus to recreate intelligence artificially (AI). Engineers have created a range of computing devices, with great variance in size, function and mobility and the resources – software – they have created have certainly opened up new worlds. But what these really tell the story of is the power of the human imagination, intelligence and creativity. Surely, many applications have more than only automated existing processes (the Internet is probably the clearest example of a conceptual innovation), as they have created the necessary preconditions for a new imagination to emerge. The absence of limits to possibilities in software engineering, brought about by an entirely new reality, the virtual one, has invoked ideas that might never have seen the light of day if computers had never existed. But the final dream, that of recreating creativity itself, must by necessity remain elusive.

Yet, in a way, this quest is reminiscent of an earlier search for the tree of life: Jung noted that the medieval alchemists, trying to discover the philosopher's stone that would turn metal into gold, did not find this actual stone, but did in fact find it after all: through the learning process involved in the trial and error of their alchemical science, they themselves became wise. They, in a way, created themselves through their learning;

[9] At the time of writing this essay, new computer concepts are being developed, such as quantum-, chemical and even biological computers. I can not exclude that they will have different operating principles from present-day computers, but the question is to what extent they would then still be computers. I therefore stick with the definition by Searle (1991, 30) who emphasised that "[i]t is essential to our conception of a digital computer that its operations can be specified purely formally."

similarly, the search for AI can be perceived as a way of self-creation; the quest for AI as an alibi for the creation of the scientists' own intelligences.

Conclusion: The design of the mind

Although we have indicated that the human, intelligent mind is by definition different from that of a computer, which is due to the different directions of thinking, we have not yet indicated why this is so. And here we arrive at the heart of this argument. If we accept the thesis that intelligence equates fundamentally to the ability to see unity (patterns) in manifoldness (instances), and is therefore fundamentally analytical, then we assume that its nature is fundamentally deductive. This induction–deduction tension runs in part parallel to the concrete–abstract dichotomy; the human mind is deductive in that its basic direction is not from matter to theories to explain that matter. It already has some basic patterns ingrained, which Kant called *a priori* knowledge (1781/1990, 2). For example, we do not find concepts such as association, taxonomy or difference in the natural world. These originate in the human mind. When we perceive the natural world, we do so with our minds, and therefore perceive it from a position in which these ordering principles are already inherent – we go from the abstract to the concrete and specific.

Furthermore, in contradiction with the conception that idealism should lead to solipsism, the concepts with which we approach the natural world refer to a realm that is not individual but shared. We refer to these concepts as universals, not only because they are ideal and therefore unchangeable (as opposed to objects in material reality, which are subject to empirical laws of change), and may even be applicable independently of time and place, but also because they are universally human. Whereas interpretation, opinion, values and so on can differ by culture and even by individual, universal concepts do not change from person to person. So here we encounter a paradox: if we know that universals come from an ideal world, and if we assume that they are non-specific, then it seems as if they come from a realm we associate with religious dimensions: a timeless (eternal) and placeless (omnipresent) world; a Mind that

is the pre-human archetype of all minds. If this vocabulary is reminiscent of psychoanalysis, then that is not a coincidence: the concept of mind in the discourse of computation, and AI in particular, seems strikingly similar to that attributed to the philosophers by Freud. Freud alleged that for philosophers, the psyche coincides with the contents of consciousness, while psychoanalysis argues that it is basically unconscious (1991).

If it were true that we derived our concepts from a pre-empirical space that is in fact not a space, and if we could identify this with Plato's unity that lies behind the empirical world of (manifold) sensory perception, then it seems plausible that humanity's intelligence depends on the extent to which s/he participates in this Mind. That is to say, our intelligence, which is the ability to see unity in manifoldness, derives from the fact that, behind the scenes of conscious perception, we are, as it were, part of that unity. As conscious beings, naturally, we experience ourselves as individuals; however, through our minds, in an ideal world, we participate in a unity that, again through our minds, permeates physical reality. From this perspective, the etymology of the word religion is unsurprising: its root is *religere*, which in Latin means to bond, to 're-unite'. By engaging in religion, as individuals in a world of sensory perception, we re-establish our link with a reality behind this one and, as a consequence, with one another.

The direction of thinking of human beings, therefore, goes by definition from that which is beyond our individuality, this Mind or Absolute Mind as Von Baader calls it (1851; in Jung, 1950), to the world. It starts with concepts, and applies these to specific observations, which it then perceives, classifies and evaluates. It goes from 'up' (the heavens that represent metaphysics) to 'down' (the earth that symbolises the world of the tangible). Machines, which are built from elements of that physical world, can by definition only think in the other direction, hopefully applying some ideas from elementary building blocks. It is because they are not part of a Mind that they can, by definition, not become intelligent. That is to say, their processing power can enormously increase, according to Moore's Law, and this can certainly enhance their synthetic capacity, but they cannot operate analytically: they can only exist in time, not in timelessness.

This leaves us with the question of whether this is a problem. The answer is, probably, no. The human imagination is quite powerful as it can imagine something inside a machine, even if it does not exist; but to encounter intelligence we will still have to look at one another; and fortunately so, I might add.

Note

The original version of this article won the 2008 GD & MC Harris Prize for best essay on computing and religion, Kellogg College, University of Oxford. It has since been updated to reflect a current state of affairs.

References

Akass, C. 2007. "The Zombie enigma." *PC World*, November: 20–21.

American Psychological Association. 1996. *APA task force examines the knowns and unknowns of intelligence*. Retrieved 28 July 2020 from: http://web.archive.org/web/20041130090438/http://www.apa.org/releases/intell.html

Argonov, V. 2014. "Experimental methods for unraveling the Mind–Body Problem: The phenomenal judgment approach." *Journal of Mind and Behavior* 35, No. 1 and 2: 51–70.

Arvey, R.D. et al. 1994. "Mainstream science on intelligence." *Wall Street Journal* 13 December: A 18.

Augustine, A. 2007. *De stad van God* (The city of God). Transl. G. Wijdeveld. Amsterdam: Ambo.

Aydede, M. & Güzeldere, G. 2000. "Consciousness, intentionality and intelligence: some foundational issues for artificial intelligence." *Journal of Experimental & Theoretical Artificial Intelligence* 12: 263–277.

Baader, F. von. 1851. *Fermenta cognitionis*. Leipzig: Publisher unknown.

Bishop, J.M. 2018. "Is anyone home? A way to find out if AI has become self-aware." *Frontiers in Robotics and AI* 5, No. 17: pp. 1–2.

Bloom, B. et al. 1971. *Taxonomy of educational objectives*. Ann Arbor: Edward Bros.

Brooke, J.H. 2006. "Contributions from the history of science and religion." In *Oxford handbook of science and religion*. Oxford: Oxford University Press.

Chalmers, D. 1995. "Facing up to the problem of consciousness." *Journal of Consciousness Studies* 2, No. 3: 200–219.

Darwin, C. 1958. *The autobiography of Charles Darwin 1809–1882*. [With original omissions restored. Edited with Appendix and Notes by his grand-daughter Nora Ballow.]. London: Collins.

Dembski, W.A. 2002. "Kurzweil's impoverished spirituality." In *Are we spiritual machines: Ray Kurzweil vs. the critics of strong AI*, edited by J.W. Richards. Seattle: Discovery Institute Press.

Diamond, A., Knight, R., Devereux, D. & Holland, O. (2012). "Anthropomimetic Robots: Concept, Construction and Modelling." *International Journal of Advanced Robotic Systems* 9: 1–14.

Dobzhansky, T. 1969. *The biology of ultimate concern*. London: Rapp & Whiting.

Dostoyevsky, F. 1958. *De gebroeders Karamazow* (The brothers Karamazov). Transl. J. van der Eng. Amsterdam: G.A. van Oorschot.

Fjelland, R. 2020. "Why general artificial intelligence will not be realized." *Humanities and Social Sciences Communications* 7, No. 10: 1–9.

Freud, S. 1991. *Zelfportret*. *De weerstanden tegen de psychoanalyse*. *Het vraagstuk van de lekenanalyse*. (Self-portrait. The refutations of psychoanalysis. The question of lay analysis). Meppel / Amsterdam: Boom.

Fromm, E. 1984. *The fear of freedom*. London: Routledge Classics.

Gardner, H. 1993. *Frames of mind*. London: Fontana.

Graziano, M.S.A. & Kastner, S. 2011. "Human consciousness and its relationship to social neuroscience: A novel hypothesis." *Cognitive Neuroscience* 2, No. 2: 98–113.

Halal, W.E. 2009. "The automation of thought: Information technology holds the key to global maturity." *Technology Analysis & Strategic Management* 21, No. 2: 277–280.

Heinisch, K.J. 1960. *Der Utopische Staat* (The utopian state). Reinbek bei Hamburg: Rowohlt.

Hunt-Grubbe, C. 2007. "The elementary DNA of dear Dr. Watson." *Times Online* 14 October. Retrieved 28 July 2020 from: https://library.wur.nl/WebQuery/file/cogem/cogem_t49203c8e_001.pdf

Jung, C.G. 1982. *Psychologie en religie* (Psychology and religion). Rotterdam: Lemniscaat.

Jung, C.G. 1950. *Gestaltungen des Unbewußten* (Representations of the unconscious). Zürich: Rascher.

Kant, I. 1990. *Critique of pure reason*. Amherst: Prometheus books.

Kant, I. 1996. *Kritik der praktischen Vernunft. Grundlegung zur Metaphysik der Sitten* (Critique of practical reason. Groundwork of the metaphysics of morals). Frankfurt am Main: Suhrkamp.

Kurzweil, R. 2001. *One half of an argument*. Retrieved 28 July 2020 from: http://www.edge.org/3rd_culture/kurzweil/kurzweil_print.html

Kwa, C. 2005. *De ontdekking van het weten* (The discovery of knowledge). Amsterdam: Boom.

Kwa, C. 1991. "Wetten en verhalen (Laws and stories)." *Kennis en Methode* 1: 105–120.

MacGregor, N. 2008. "Foreword." In: *The Greek myths: Origins of the gods*. Supplement to *The Guardian*.

Manuel, F. & Manuel, F. 1982. *Utopian thought in the Western world*. Cambridge, Mass.: Belknap / Harvard University Press.

Marcus, G. 2019. DeepMind's losses and the future of artificial intelligence. *Wired*, 14 August. Retrieved 28 July 2020 from: https://www.wired.com/story/deepminds-losses-future-artificial-intelligence/

Marx, K. 1909. *Capital (Vol. III)*, edited by F. Engels. Chicago: Charles H. Kerr.

Mumford, L. 1959. *The story of utopias*. Gloucester Massachusetts: Peter Smith.

Paracelsus 1537. *De natura rerum*, Book I.

Plato. 1995a. *Sokrates' leven en dood* (Socrates' life and death). Transl. G. Koolschijn. Amsterdam: Athenaeum / Polak & Van Gennep.

Plato. 1995b. *Constitutie* (The Republic). Transl. G. Koolschijn. Amsterdam: Athenaeum – Polak & Van Gennep.

Reggia, J.A. 2013. "The rise of machine consciousness: Studying consciousness with computational models." *Neural Networks* 44: 112–131.

Rorty, R. 1980. *Philosophy and the mirror of nature*. Oxford: Basil Blackwell.

Russell, B. 2007. *The history of Western philosophy*. London: Routledge.

Schopenhauer, A. 1996. *Die Welt als Wille und Vorstellung* (The world as will and representation). Frankfurt am Main / Leipzig: Insel Verlag.

Searle, J.R. 1991. *Mind, brains and science: the 1984 Reith lectures*. London: Penguin.

Searle, J.R. 1980. "Minds, brains, and programs." *Behavioral and Brain Sciences* 3, 3: 417–457.

Shelley, M. 1818. *Frankenstein; or, the Modern Prometheus*. London: Lackington, Hughes, Harding, Mavor & Jones.

Spinoza, B. 1862. *Tractatus theologico-politicus*. London: Trübner & Co.

Turing, A.M. 1950. "Computing machinery and intelligence." *Mind* 59: 433–460.

Warnick, B. 2004. "Rehabilitating AI: argument loci and the case for Artificial Intelligence." *Argumentation* 18: 149–170.

Warwick, K. 2002. "Alien encounters." In *Views into the Chinese Room*, edited by J. Preston and J. M. Bishop. Oxford, UK: Clarendon Press.