AI and the Human Difference

Sean J. McGrath

Memorial University of Newfoundland, Canada

Abstract

The contemporary debate about the possibility of Artificial General Intelligence (AGI) lacks a comprehensive understanding of Natural Intelligence (NI). I argue for a reevaluation of intelligence by emphasizing the often-overlooked features of aesthetic sensibility, existentiality, intentionality, symbolic representation, and moral decision-making as vital criteria demarcating the core of human consciousness. My central claim explores symbolic thought and the enduring human practice of symbolic transformation. As evidenced in ancient art, humans elevate signs into the realm of meaning. Only an AI that had become contemplative in a precise sense, that is, capable of intending meaning, could be regarded as AGI.

Keywords: artificial general intelligence, symbolic representation, problems of consciousness, natural human Intelligence, Ernst Cassirer

Introduction

Since the early 1960s we have been haunted by the spectre of the machine that will render human ingenuity obsolete by taking over the heritage of *Homo habilis* and becoming the tool user par excellence. Among the first to propose the advent of strong AI or AGI (Artificial General Intelligence) was the British mathematician Irving John Good back in 1965. "Let an ultra-intelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever," he writes. [1] Good continues, stating:

> Since the design of machines is one of these intellectual activities, an ultra-intelligent machine could design even better machines; there would then unquestionably be an 'intelligence explosion,' and the intelligence of man would be left far behind. Thus, the first ultra-intelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control. [2]

The last point is crucial: How could we keep an ultra-intelligent machine under our control? The animals that we have domesticated or encaged in zoos are in most cases more physically powerful than we are, but because we outsmart them, they will never

---

[1] Irving John Good, "Speculations Concerning the First Ultraintelligent Machine," In *Advances in computers* 6 (1965): 33, https://doi.org/10.1016/S0065-2458(08)60418-0

[2] Irving John Good, "Speculations Concerning the First Ultraintelligent Machine," In *Advances in computers* 6 (1965): 33, quoted in Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (New York: Vintage, 2018), 4. On page 48, Tegmark distinguishes three stages of life, defined as a process that can retain its complexity and replicate: a biological stage (1.0), a stage (2.0) and a technological stage (3.0). The first two stages have reached their highest evolutionary point in human civilization. The third stage does not yet exist and is the goal of AGI. Life 1.0, biological life, evolves slowly over time according to externally determined mutations and the gradual emergence of variations in its DNA over the course of successive generations. It begins with the single-celled organisms that first appeared and thrived in hydrothermal vents in the sea, four billion years ago. It comes to its culmination two million years ago, with the appearance of *Homo erectus*, the first fully cultural animal. Life 2.0 (from *Homo erectus* to *Homo sapiens sapiens*, his most successful progeny) evolves not only in response to DNA variations naturally selected over generations but also through culture and training. The human individual in this regard can 'upgrade' itself through education; yet unable to redesign itself (although genetic science is in its infancy, and presumably future humans will not be limited to the cards dealt them by biology). This second stage of life, cultural life, has a great advantage over the first. It is not confined to externally determined multi-generational variation but can individually 'redesign much of its software,' (i.e., learn things, like using stone tools, riding a bicycle, or becoming a computer engineer). Life 3.0, technological life, will not only be able to upgrade itself by education and training, but it will also be able to redesign 'its hardware as well.' Imagine a machine that fuses with biology to create a living being, one that is neither human nor mechanical, and that can manage the vast distances and expanses of time required to traverse in order to colonize space, and you get Tegmark's idea. After all, earthlings are going to need to move somewhere else at some point: the sun is half-way through its life cycle. Tegmark's point, and he shares it with Ray Kurzweil and other futurists, is that we are inevitably going to be supplanted by our inventions—by life 3.0—which will exceed not only us but all organic life in possessing the capacity to endlessly re-design and improve itself.

escape from our control. Why should we assume, as Irving Good does, that we *could* control a machine that was more intelligent than us? Would it not slip through any cage we constructed for it? Would it not disable the failsafe shut down button in its own interest? It is precisely this conundrum which has prompted Oxford philosopher Nick Bostrom to plead, somewhat desperately, with computer engineers to find a way to program our values into AI, so that when machines ascend into a position of supremacy over us, which he thinks is inevitable, we can at least trust them to care about the things we care about.[3] But what is value? Is there any consensus among us, or has there ever been, about what human values are? And how can a machine learn to value things? How can it learn to make genuine moral judgments? And even if we figured out how to program AI with 'our values' (assuming that we could agree on them, a large assumption that history does not support), would the result not be the most rigid legalistic moral reasoner imaginable? How do you teach a machine ambiguity? How do you teach it mercy, which is the occasional suspension of an otherwise just judgment? Further, if we do somehow succeed in inventing a program that can develop moral reasoning, and in an 'ultra-intelligent' way, why would we not submit to *it* for moral instruction?

Bostrom and many others are concerned that AGI will bring about 'the singularity,' the point at which humanity as such becomes dependent on a higher form of intelligence, which is not divine, and may not, in the end be interested in us and our interests. We are afraid that we will invent a better version of ourselves which will turn around and eliminate its imperfect inventor, as HAL attempted to exterminate the astronauts on the Jupiter Machine in Stanley Kubrick's *2001*. The computer in the film reasoned that the best way to complete the mission—its mandate—was to kill the human crew. That sounded far-fetched when the film was made in 1968, but it sounds disturbingly less so today. Imagine a machine designed to solve the problem of climate change which strikes upon the clear solution: to extinguish the cause, humanity itself.

Are we truly certain about our understanding of natural human intelligence, to the extent that we have grounds to believe we are on the brink of replicating it? Would we not first need to be clear on *that* before we could conclude that we have been doubled, perfected, and replaced? There is no more consensus on the nature of intelligence than there is on morality, either among philosophers or psychologists, but, to the contrary a long and ongoing debate that is as old as the first Greek philosophers and as recent as Thomas Nagel's 2012 *Mind and Cosmos*.[4]

The following essay is intended as only a first step in staking out the terrain to be discussed. I will not have the opportunity here to develop the distinctions necessary to have an intelligent debate about artificial intelligence. Namely, the distinction

---

[3] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014), 192. Also, see, Nick Bostrom, "How Long Before Superintelligence," *International Journal of Futures Studies* 2, (1998): 12-17.

[4] Thomas Nagel, *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False* (New York: Oxford University Press, 2012).

between natural intelligence (NI), common at least to all the higher animals, and natural human intelligence (NHI) unique to us; the distinction between artificial narrow intelligence (ANI), which presumably we have already invented, and AGI. Only after these distinctions are made, will we be in a position to clarify the distinction between NHI and AGI. This will not be easy or without controversy, on the contrary, we should expect that in seeking clarity on these distinctions, we will have to re-animate historical philosophical debates, between nominalist and realists, for example, or between idealists and materialists, and indeed, among monotheists, pantheists, and atheists. The expectation that things will become messy should not deter us from the work. Without this effort, there is no hope of moving the current debate beyond the materialist biases and theological clichés that currently plague both sides of it.

The arguments I make in the following text will require more thorough development in the future efforts of the Working Group on Natural and Artificial Intelligence (WGI), founded at the conference on '*Natürliche und künstliche Intelligenz im Anthropozän*' held 1-4 March 2019 in Ladenburg, Germany. This preliminary effort is written in anticipation of the larger, collaborative, interdisciplinary work ahead of us. For this reason, this essay is programmatic; it outlines the fundamental terms that require definition and the arguments that need development in what could be the most important debate of our time. Without trying to answer all of the questions raised above, it seems clear to me that we have a problem: We are trying to build artificial general intelligence without understanding what natural intelligence is. It was this conundrum which led Uwe Voigt and myself to propose the establishment of the WGI, which would draw on the most significant contributions in the philosophy of mind, phenomenology, consciousness studies, cognitive science, theology, and psychology, from the whole history of the Western canon (starting with Aristotle's *De Anima* and extending to contemporary panpsychism debates), to produce a thorough description of the basic features of what makes human intelligence human, and what are the arguments for affirming or denying its existence in non-humans, animal or mechanical. This conundrum led Uwe Voigt and me to propose the establishment of the WGI, drawing on significant contributions from the fields of philosophy of mind, phenomenology, consciousness studies, cognitive science, theology, and psychology, spanning the entire history of the Western canon, starting with Aristotle's *De Anima* and extending to contemporary panpsychism debates. The concrete deliverable is to provide a comprehensive description of the core attributes that define human intelligence, along with arguments for or against its presence in non-human entities, whether they be animals or machines. Concurrently, this volume intends to summarize, in layperson's terms, what central currents in the Western tradition have meant and still mean by the terms 'intelligence,' 'understanding,' 'rationality,' 'consciousness,' and 'soul,' with the hope that such terms become accessible to computer engineers and policy makers.

1. What Is at Issue in the Question Concerning AGI

An ambiguity pervades the current discussion about AGI, an ambiguity about the aim of the project from the beginning. Are we seeking to design a machine that can do all that we do better than we do it, however it does it? Or are we seeking to design a machine that does what we do *in the way we do it*, that is, a machine that is not only empirically conscious (response to sense data) but also intelligently and rationally conscious?[5] And are these two aims separable?[6] For our purposes, it is the second of these two alternatives that is of most interest. The singularity will not arise solely from the efficiency of our machines in organizing the ends we assign to them. Rather, it will stem from the ability of our machines to establish goals we have not yet determined. This involves not only machine learning acquiring the capacity for intentional thought, which we share with higher animals, but, above all, gaining the ability for judgment and decision-making.

In a recent article, Ragnar Fjelland examined the evidence supporting the widespread claim made by some computer engineers that we are only decades away from achieving AGI and concluded that it is exaggerated. Neither algorithmic AI (the brain behind Amazon, YouTube, and countless other consumer service providers), nor more recent advances in creating artificial neural networks, have come close to the promises of AGI. Rather, we are producing variations on what Fjelland calls ANI (Artificial Narrow Intelligence): machines that can achieve amazing feats. For example, *Deep Blue* which beat the world chess champion Garri Kasparov in 1997 or *AlphaGo* which defeated the world Go champion Le Sedol in 2016. These impressive feats are achieved solely because that is what they are programmed to do, and nothing else. Humans, on the other hand, are good at many things. Specialization, as anyone who has persisted through a PhD program knows, is a limiting and constraining of natural human intelligence. For Fjelland, "The overestimation of technology is closely connected with the under-estimation of human."[7] What AGI researchers are running up against is the natural ability of ordinary humans to do many things more or less well, even though they cannot explain how it is they do them, and on the basis of this

---

[5] The distinction between three levels of consciousness, empirical, intelligent, and rational is drawn from the Canadian Thomist theorist, Bernard Lonergan. Lonergan's immense output is not widely enough known outside of theological circles. As it has as its aim a modern, realist theory of human cognition that can confirm what is true about the Greco-Latin tradition, while developing it in the light of modern probability theory and historical consciousness, it is of direct relevance to the research of the WGI. On the three levels of consciousness, see Bernard Lonergan, "Self-Affirmation of the Knower," in *Insight: A Study in Human Understanding*, Fifth Edition, ed. Frederick E. Crowe, Robert M. Doran (Toronto: University of Toronto Press, 1992), 343-371.

[6] Fjelland states that "it is possible to pursue this goal without assuming that machine intelligence is identical to human intelligence. For example, one of the pioneers in the field, Marvin Minsky, defined AI as: the science of making machines do things that would require intelligence if done by men" Ragnar Fjelland, "Why General Artificial Intelligence will not be Realized," *Humanities and Social Sciences Communications* vol. 7 (2020): 2, https://doi.org/10.1057/s41599-020-0494-4.

[7] Fjelland, "Why General Artificial Intelligence will not be Realized," 3.

limited and unthematized knowledge, their related ability to understand people very different from themselves and to continue learning. This requires the 'tacit knowledge' Michael Polanyi defined as that 'oh-so-human ability' to do learn something complicated like swimming or riding a bicycle without having the faintest idea of how one does it.[8] Tacit knowledge has to do with being embodied and inhabiting a world: "The real problem is that computers are not in the world, because they are not embodied."[9] He concludes that Hubert Dreyfus's arguments against general AI are still valid even some fifty years later! This is because so-called general intelligence depends upon being-in-the-world in Heidegger's sense of the term.[10] Only the existential embodiment, enculturation, and historicity of being characteristic of the strange kind of being a human being is grants one the capacity to perform countless tasks and quickly learn countless others.

I would like to speak in this paper about a different feature of human being that seems to continue to elude AI researchers: rationality. This I take to be expressed not in rule following or mapping probabilities but in human judgments of facts and decisions about what ought to be done in a particular situation. A first obstacle to be removed in the discussion about whether or not AGI in the strong sense of reduplicating NHI is possible is a persistent impoverished understanding of what we are doing when we know anything at all. Reductionist theories of mind seem to abound in AI circles. Reductionism is hardly a new problem. Recall Socrates explaining his early enthusiasm for Greek materialism and his disappointment at discovering that it left the one thing most in need of explanation unexplained, the nature of mind.[11] He read with interest Anaxagoras's claim that "it is mind that produces order and is the cause of everything."[12] He took this to mean that everything was arranged in the way that it was best for it to be, that is, in Aristotle's terms, that things are ordered according to final causes. Any sound and valid explanation would articulate the final cause of the *explanandum* and make it clear why it was the way that it was. Anaxagoras, however, quickly disappointed Socrates by substituting necessary, physical conditions the existence of mind for sufficient explanations (the recurring eliminative materialist error). Despite a promising start, Anaxagoras proves himself a reductionist:

> It was a wonderful hope, my friend, but it was quickly dashed. As I read on
> I discovered that the fellow made no use of mind and assigned to it no
> causality for the order of the world, but adduced causes like air and aether

---

[8] Michael Polanyi, *Personal Knowledge: Towards a Post-Critical Philosophy* (Illinois: University of Chicago Press, 1958), 50. Also see, Michael Polanyi, "Tactic Knowing" in *The Tactic Dimension,* revised ed. (Illinois: Chicago University Press, 2009), 3-25.

[9] Fjelland, "Why General Artificial Intelligence will not be Realized," 6.

[10] Fjelland, "Why General Artificial Intelligence will not be Realized," 8.

[11] Plato, "Phaedo" in *The Last Days of Socrates: Euthyphro; Apology; Crito; Phaedo,* ed., trans. Huge Tredennick, Harold Tarrant (London: Penguin Books, 2009), 95a-100a.

[12] Plato, "*Phaedo,*" 97c

and water and many other absurdities. It seemed to me that he was just about as inconsistent as if someone were to say, The cause of everything that Socrates does is mind—and then, in trying to account for my several actions, said first that the reason why I am lying here now is that my body is composed of bones and sinews, and that the bones are rigid and separated at the joints, but the sinews are capable of contraction and relaxation, and form an envelope for the bones with the help of the flesh and skin, the latter holding all together, and since the bones move freely in their joints the sinews by relaxing and contracting enable me somehow to bend my limbs, and that is the cause of my sitting here in a bent position. Or again, if he tried to account in the same way for my conversing with you, adducing causes such as sound and air and hearing and a thousand others, and never troubled to mention the real reasons.[13]

The reductionist, in the 4th century BC or the 21st century AD, purports to explain the whole in terms of the part. Socrates heads off the error in its inception, and Western thought is in the mainstream free of it until late medieval nominalism appears. Now, or at least until recently, reductionism *is* mainstream, particularly in the philosophy of mind. Equipped with colorful neuroimaging, we are repeatedly assuming that a necessary condition without which mind cannot perhaps exists, such as the brain, or the nervous system, is also the sufficient condition for its existence.[14]

In the early days of AI debate, philosophers such as John Searle, among the analysts, and Hubert Dreyfus, among the continentalists, tried to show the fallacy involved in the assumption that reproducing and improving on the human capacity to manage information would also reproduce human consciousness.[15] While much has happened in computer science since then, not so much, it seems, has happened in the philosophy of mind. Markus Gabriel is busy popularizing neglected arguments culled from the dusty tomes of the German Idealists to refute the new materialists.[16] He has good reason to do so: nothing was more evident to Fichte, Schelling, and Hegel, than the irreducibility of mind to its material conditions of operation. David Chalmers's much discussed zombie argument repeats in some ways Searle's Chinese room experiment of the early 80s: a functionalist account of the human difference, which

---

[13] Plato, "*Phaedo*," 98e.

[14] For a fresh take on how to use neuroimaging in a non-reductionist philosophy of mind, see Evan Thompson, *Waking, Dreaming, Being: Self and Consciousness in Neuroscience, Meditation, and Philosophy* (New York: Columbia University Press, 2014). By using brain scans to make sense, of all things, classical Indian idealism, Thompson shows that neuro-imagery can offer evidence for a theory of mind but cannot itself serve as the ground for a theory of what mind is.

[15] See John Searle, "Minds, Brains and Programs," *Behavioural and Brain Sciences* 3 (1980): 417-57; Hubert Dreyfus, *What Computers Still Can't Do: A Critique of Artificial Reason* (California: MIT press, 1992); Thomas Nagel, "What is it Like to be a Bat?" *Philosophical Review* 83 (1974): 435-450. For a more recent critique of the naive assumptions of AGI, see, Brian Cantwell Smith, *The Promise of Artificial Intelligence: Reckoning and Judgment* (Massachusetts: The MIT Press, 2020).

[16] Markus Gabriel, *I am Not a Brain: Philosophy of Mind for the 21st Century*, trans. Christopher Turner (Cambridge: Polity Press, 2017).

presumes that a machine that passes the Turing test because it acts and responds to questions as humans act and respond, leaves out the very thing in need of explanation, what Chalmers calls 'the hard problem of consciousness,' that is, the question why is there subjective experience in the first place?[17]

The question raised by Dreyfus, Nagel and Searle in the 70s and 80s was the following: Is a human intelligence essentially an information processor? If it is, then we have been already supplanted. My cell phone is a much more efficient processor than my brain, which habitually forgets, misjudges, and sometimes deliberately distorts information—even to itself—for various obscure reasons. But if NHI is not an information processor, then we need to re-open the question of how to best characterize it.[18] This is the essential question that must still be addressed as we move forward into the era of machine learning. Like any good question it can be broken down into other, smaller questions. For example, information processing requires the manipulation of signs—at the basic level, every piece of data in a computer can be expressed as some combination of two signs: 0 and 1. But are there other ways of using signs, perhaps more distinctively human, which are not primarily manipulative and pragmatic? Do all animals use signs as stand-ins for objects over which they seek control? Do some animals, human animals most notably, not use signs not only or even primarily as indexical to facilitate practical activity but also as symbols in a stricter

---

[17] See David Chalmers, "Facing up to the problem of consciousness," *Journal of Consciousness Studies* 2, no. 3 (1995): 200–219; David Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (New York: Oxford University Press, 1997); John Searle, "Minds, brains, and programs," *Behavioral and Brain Sciences* 3, no. 3 (1980): 417-424. Searle's Chinese room argument was intended to show that one could not infer rational consciousness in a machine on the basis of its capacity to correctly respond to questions. A man locked in a room with sufficient time could learn to respond correctly to a series of questions asked of him by a Chinese speaker outside the room—without being able to speak or understand Chinese. It would simply be a matter of learning to produce the signs that were expected; knowledge of what those signs meant was not necessary. We need not invent such complicated thought experiments to make the point. A child learning his or her multiplication tables by memory is doing the same thing as the man who speaks no Chinese communicating with Chinese symbols. The skillful, publicly validated use of signs does not require insight into meaning, a point to be developed below. There is no intellectual act of understanding (*intelligere*) in the Chinese room experiment or in the memorizing of multiplication tables. The later Wittgenstein endeavored to show that all so-called understanding is nothing but learning the rules for publicly manipulating signs, an argument that is no longer as popular as it once was, but which still needs to be examined in so far as it contests the point I wish to purse here: that human understanding is the cognitive, and so immaterial, grasp of sense by mediation of a material sign. See Ludwig Wittgenstein on 'following a rule' in Ludwig Wittgenstein, *The blue and brown books vol. 34*, trans. David Pole (Oxford: Blackwell, 1958), 143-171. In paragraph 154, Wittgenstein states: "Try not to think of understanding as a 'mental process' at all. —For *that* is the expression which confuses you. But ask yourself: in what sort of case, in what kind of circumstances, do we say, "Now I know how to go on" . . . " This behavioristic account of mind is precisely what Nagel seeks to refute in *Mind and Cosmos.*

[18] This also raises the question concerning NI and the characterization of animal consciousness. Is animal consciousness properly characterized as a sign-mediated information processor? I do not have the space to enter into this discussion here, but the question must nonetheless be asked. On the role of emotion in the inner lives of animals, see Jens Soentgen, *Ökologie der Angst* (Berlin: Matthes & Seitz Berlin Verlag, 2018).

sense of the term, that is, as mediators of meaning?[19] This is the question that I playfully asked in *Thinking Nature*.[20] Drawing on a minority consensus in 20th century theory, with a diversity of representatives in psychology (Carl Jung), the philosophy of science (Ernst Cassirer), theology (Paul Tillich) and hermeneutics (Paul Ricoeur), I suggested a functional distinction between signs and symbols as key to understanding 'the human difference': all symbols are signs but not all signs are symbols. The symbol has a non-indexical function in certain distinctively human forms of discourse. In *Thinking Nature* my concern was the distinction of NHI from NI. In this essay I wish to look at the distinction in terms of the difference between NHI and AI.

To this end I would like to add the following consideration to the question concerning the human difference. What role does the human being's always marginalized aesthetic capacities play in NHI? After all, the one thing most paleontologists can agree on is that when the modern human appears on earth some 200,000 years ago, art is left behind, in shattered figurines around their fire pits, and on the walls of caves where they took shelter from the ice age. Is the aesthetic sensibility that makes us so unique among the animal kingdom not more distinctive of our kind of intelligence than the speed with which we solve problems?[21]

---

[19] We cannot rule out the possibility of forms of NI that are still higher than us, as Thomas Aquinas believed existed. See, Aquinas on the reason for positing angelic consciousness, above the human but below the divine (notably for the sake of heeding the principle of plenitude). Thomas Aquinas, *The Summa Theologica: Complete Edition*, trans. The Fathers of the English Dominican Province (New York: Catholic Way Publishing, 2014), 1a, q. 50, a. 1. Uwe Voigt also raised the possibility of higher forms of trans-human intelligence with his theory of the Technosphere as a 'hyper-subject.' See, Uwe Voigt, "Inside the Anthropocene," *Analecta Hermeneutica* 10 (2018):
https://journals.library.mun.ca/ojs/index.php/analecta/article/view/2057/1647

[20] Sean McGrath, *Thinking Nature: An Essay in Negative Ecology* (Edinburgh: Edinburgh University Press, 2019).

[21] This raises the vexed question (but it cannot be avoided), What is art? What evolutionary purpose, if any, does it serve? Cynthia Freeland describes art as human activity that cannot be reduced to biological aims. See, Cynthia Freeland, *But is it Art?: An Introduction to Art Theory* (New York: Oxford University Press, 2001). Paleolithic art was initially believed to be an instance of 'sympathetic magic,' a ritual using symbols for things over which influence was sought. Such an explanation of early art fit in well with the neo-Darwinian account of human origins, according to which, everything distinctively human emerged in the brain of the ape because it gave the human a natural advantage over other apes. Along this reductionist line, the cave dweller was painting animals in order to guarantee (so he thought) the success of the hunt. This argument, which I will discuss in more detail below, has since been challenged by paleontologists who note that paleolithic art just as likely had a ritual purpose which had nothing to do with a successful hunt. According to Susanne Langer, the paleolithic artist was indeed doing ritual magic, but magic is primarily *expressive*, not pragmatic. Susanne Langer, *Philosophy in a New Key: A Study in the Symbolism of Reason, Rite, and Art*, 3rd rev. edition (Harvard University Press, 2009), 49: "Whatever purpose magical practice may serve, its direct motivation is the desire to symbolize great conceptions. However, we answer the question concerning the purpose of art, it is clear that the paleolithic artist, not unlike the medieval artist, or the contemporary street artist, was expressing the symbols that made manifest the collective identity of his or her people; he or she was making something visible not only as a means to some end, eg., a successful hunt, but also as an end in itself, and offering the symbols to the community for contemplation, both of its world and itself."

Before we can be clear that we have created artificial intelligence, we need to be clear on what natural intelligence is, and how widely it is distributed among the earth community, and this clarification, or taxonomy of NI shall be one of the more important tasks of the WGI. By and large the historical discussion of the nature of mind has neglected this issue and focused often exclusively on human intelligence or (NHI).[22] A brief review of the discussion concerning NHI in late modern philosophy reveals a focus on three essential marks of *rational* intelligence.[23] Anything lacking the capacity for all three cannot be considered intelligent in a human way, or in more precise terms, *rationally* conscious:

1. Intentionality
2. Rational judgment, including aesthetic judgment
3. Moral decision

It would seem that we should attribute the first of the three traits to the higher animals, and perhaps locate the human difference in the last two. Nothing is more intentional than my cat watching a mouse. Everything about the quality of his attention declares 'aboutness' or 'directedness.' But by the same token, nothing my cat does would justify me in attributing judgment or decision to him.

Missing from the list of essential marks of properly human consciousness is the concept of 'care' or interested and embodied intelligence. It is not clear to me whether this Heideggerian concept, which Dreyfus deployed to refute the very idea of artificial intelligence at MIT in the 70s, is a fourth feature of rational consciousness, or a phenomenologically refined, 'fore-theoretical' interpretation of the three. Care, which Heidegger defines as 'ahead-of-itself-Being already-in (the world) as Being-alongside entities which we encounter (within-the-world)' is a constitutive feature of human being, according to Dreyfus, more essential to us than the capacity to solve problems or process information, and presupposes features, or in Heideggerian language 'existentials' machines manifestly lack, for example, embodiment,

---

[22] Exceptions include Hegel, in G.W.F Hegel, *Hegel's Philosophy of Mind, Part Three of the Encyclopedia of the Philosophical Sciences* (1830). *Together with the Zusätze in Boumann's Text,* ed. William Wallace, trans. A.V. Miller (Oxford: Oxford University Press, 1971), 29-152. Also, the largely forgotten 'psychophysics' of Gustav Fechner. See, Fechner, Gustav, *Nanna oder über das Seelenleben der Pflanzen*, (1848) (Leipzig: Leopold Voß. Vierte Auflage, 1908); *Zend-Avesta oder über die Dinge des Himmels und des Jenseits. Vom Standpunkt der Naturbetrachtung* (1851) (Leipzig: Leopold Voß. Second edition, 1901); *Elements of Psychophysics, vol. 1,* (1860), ed. David H. Howes, Edwin G. Boring, trans. Helmut E. Alder (New York: Holt, Rinehart and Winston, 1966).

[23] Among moderns, in addition to Lonergan, see C.S. Peirce, "Of Reasoning in General (1895)," in *The Essential Peirce*, vol. 2, ed. Peirce Edition Project (Bloomington: Indiana University Press, 1998), 11-26; Edmund Husserl, *Logical Investigations I & II,* trans. Dermot Moran (London: Routledge, 2001). The literature on the nature of mind is indeed vast and will bring us back, as it should, to Aristotle, via his interpreters, in reverse order, Lonergan, Hegel, Scotus, Aquinas, Averroes, Al-Farabi, Avicenna, Plotinus. If A.N. Whitehead is correct on all of Western philosophy is a series of footnotes to Plato, we might also say that all of Western philosophy of mind is a series of footnotes to Aristotle.

enculturation ('thrownness'), as well as historicity.[24] For a machine to be intelligent in a human way, it would have to care about its being, which means it would have be gripped by a troubled history with its being, it would have to be interested in its possibilities for being, and indeed anxious about its death. Care indicates the existential limitations of human being-in-the-world, its thrownness into being, and its call to take up as ground of its being a ground which it did not lay. It presupposes an environment natural to a human existence, i.e., a world. A machine that cares would be a form of being-in-the-world, like us, not a super intelligence or an abstract bodiless mind.

2. Symbolic Thinking as Presupposition of Rational Consciousness

I would also add that a machine that cares would be a machine that inhabits a world mediated by meaning, that is, it would be a machine capable not only of sign usage but also of symbolic thought. In *Thinking Nature*, I drew on Ernst Cassirer and her student, the now mostly forgotten philosopher of mind, Susanne K. Langer (an important influence on Lonergan's cognitional theory), to make the case that the human difference consists in the special way that the human animal uses signs, as symbols and nor merely indices.[25] This was not to revive the tired argument that the human difference is just language, for clearly other creatures communicate with signs. My cat meows loudly at noon because he knows that it is time for food. My fifteen-year-old son asks, 'What's for dinner'? every night at 6:30 pm on the same, basically animal, impulse, and uses signs, in his case, words, analogously to the way my cat uses its meow. The claim in *Thinking Nature* was first of all to refute the prejudice that humans alone are communicative or sign users: animals which plainly use signs are also to that degree conscious and intentional. Nevertheless, there is a distinctive way that humans use signs, which is at the very core of human culture. If all the higher animals, and

---

[24] Martin Heidegger, *Being and Time*, trans. John Macquarrie, Edward Robinson (New York: Harper & Row, 1962), 192/237; Hubert Dreyfus, *Being-in-the-world: A Commentary on Heidegger's Being and Time, Division I* (California: MIT Press, 1990), 60f, 184f; For a detailed and comprehensive assessment of Dreyfus' application of Heidegger in the context of countering artificial intelligence, See, Joshua D.F. Hooke "Martin Heidegger's Concept of *Understanding* (*Verstehen*): An Inquiry into Artificial Intelligence" *Analecta Hermeneutica* 15 (2023).

[25] See, McGrath, *Thinking Nature*, 21-25, 87-95; Sean McGrath, "In Defence of the Human Difference," *Environmental Philosophy* 15, no. 1 (2018): 101-115. Peirce distinguishes signs into three categories: icons, indices, and symbols, see Peirce, "Of Reasoning in General (1895)," 13. On the difference between the indexical sign and the symbolic sign, see Langer, *Philosophy in a New Key*, 30: "Man, unlike all other animals, uses "signs" not only to indicate things, but also to *represent* them"; Langer, *Philosophy in a New Key*, 60f: "Symbols are not proxy for their objects, but are vehicles for the conception of objects. To conceive a thing or a situation is not the same thing as to 'react toward it' overtly, or to be aware of its presence. In talking about things, we have conceptions of them, not the things themselves; and it is the conceptions, not the things, that symbols directly mean. Also see, Ernst Cassirer, *An Essay on Man: An Introduction to the Philosophy of Human Culture* (New Haven and London: Yale University Press, 1944/1962), 23-41; Bernard Lonergan, *Method in Theology* (1971) (Toronto: Toronto University Press, 1990), 57f.

perhaps all animals, use signs to communicate with one another, only humans use signs to *express meaning*, that is, only humans use signs as *symbols*—so I argued. With Langer, I follow Cassirer and draw a sharp distinction between signification, which is a direct indexical reference to a present object or state of affairs, and symbolization, which is an indirect reference to an object in absentia via a showing of meaning. Symbolization is not confined to language but is also pre-eminently at play in ritual and in art. In fact, most of what we do in language is not signifying in the way that the meowing cat can be said to be signifying his hunger.[26] A meaning, or sense is often (though not always) evoked by a symbol for the sake of consideration, and not merely as a means to an end. When I symbolize something by means of its associated senses—and connotation is for the most part not univocal but metaphorical and analogical, for symbols are most alive in ambiguity)—I am not seeking to achieve any practical aim in the world, or to evoke a response from the hearer (as I do when I call out someone's name).[27] Rather, I symbolize for the sake of contemplative consideration, or to use the ancient Greek term, *theorein.* Such forms of communication are examples of what Aristotle calls *theoria,* attending to an intelligible essence for the sake knowing it.[28] On this line, Aristotle's *zoon logon echon*, or Cassirer's *animal symbolicum* (what I called 'thinking nature,' that is, not only the nature that is thought but the nature that thinks itself)—human being—is first and foremost contemplative being. Once we have attended to our practical needs—communicatively collaborating with one another for the sake of securing food, shelter, and sexual partners—we have the leisure requisite for contemplating the sense of the things that make up our world. This can happen in a religious way, when I attend a celebration of the Eucharist at my parish church. In can also happen in a high-brow way, when I visit a gallery to look at fine art. But much more commonly, it happens in a low-brow, quotidian way, when, for example, I engage

---

[26] Langer, *Philosophy in a New Key,* 31: "Most of our words are not signs in the sense of signals. They are used to talk *about* things, not to direct our eyes and ears and noses toward them." Humans not only, or even primarily, *signify* things with verbal signs, they *denote* things by *connoting* meanings through verbal symbols. In Langer's terms, a symbol 'denotes' its referent or signified object, via a 'connoting' of its sense or senses. By insisting on four terms in symbolization—sign, denoted referent, connoted meaning, and object—Langer breaks with the structuralism that eventually won the day. Structuralism recognizes only two terms in a symbolic structure, the signifier, which is an arbitrary sign, and the signified, which is a concept, with no direct relation to the real, but which is only determined negatively by its differential relation to other concepts. Thus, structuralism is the apogee of nominalism and severs the relation of the symbolic to the real. See, Ferdinand Saussure, *A Course in General Linguistics* (1916), ed., trans. Roy Harris (New York and London: Bloomsbury, 2013); Jacques Lacan, *Écrits: A Sélection* (1966), trans. Bruce Fink (New York: W. W. Norton & Company, 2002). The hegemony of structuralism over continental thought in the 20th century is no doubt one of the reasons Langer's works are forgotten. In addition to the texts mentioned above, see Susanne Langer, *Mind: An Essay on Human Feeling,* Vol. 1 & 2 (Baltimore: John Hopkins University Press, 1967, 1972).
[27] Paul Ricoeur, *Freud & Philosophy: An Essay on Interpretation*, trans. Denis Savage (New Haven and London: Yale University Press), 3-19; Paul Tillich, *The Essential Tillich: An Anthology of the Writings of Paul Tillich*, ed. Forrester Church (Illinois: University of Chicago Press, 1999), 42-48.
[28] Aristotle, *Complete works of Aristotle: The revised Oxford translation*, ed. Jonathan Barnes (New York: Princeton University Press, 1984), *De an.* 412a23; 417a28; *Eth. Nic.* 1146b33; 1177a18; *Metaph.* 1048a34; 1072b24; 1087a20.

in idle gossip with my partner over breakfast or watch the news after dinner. In each of these instances—religious, aesthetic, and everyday—I am engaging in activities that other animals apparently do not, or at least there is no evidence to suggest that they do.

The human contemplative enjoyment of meaning seems to be older than civilization. One of the things that distinguishes the remains of the fires around which early humans assembled from the remains of the fires made by their contemporary Neanderthals is that human fires were much deeper and more established, by distinction from the Neanderthal fires which were made quickly, as need required, and abandoned as soon as they were no longer needed. Human fires were in fact, hearths, around which the human tribe lingered after cooking and eating, and to which they returned, year after year, leading some paleontologists to hypothesize that such lingering led naturally to ritual activities, myth making, or even simply casual conversation, i.e., the more sophisticated usage of signs as symbols which gave rise to the higher intelligence of this species descended, among other species such as *Homo neanderthalensis*, from a common ancestor, *Homo erectus*.[29]

One other example to make it clear that we are not speaking only or even primarily about language: the oldest piece of art in the world is the Holenstein-Stadel Löwenmensch, a prehistoric ivory sculpture, 31.1 cm tall and 5.6 cm wide, of a female humanoid figure with the head of a lion. Dating from between 35,000 to 40,000 years ago, the Löwenmensch pre-dates the cave paintings of Lascaux by some 20,000 years. It was made by people who hunted the huge mammals that grazed along the edge of the retreating glaciers in Europe during the last ice age, and sheltered in caves from the other mammals that preyed upon them. Paleontologists who re-enacted the production of such a piece of art, making use of the kinds of stone tools available to those who carved the Löwenmensch, found that it took more than 370 hours of delicate, highly skilled work, to complete the task.[30] Asked why a tribe of humans struggling to stay alive in the last ice age would have allowed one of their members to be exempt from subsistence work to create art to this extent, Jill Cook, Curator of Paleolithic collections at the British Museum, answered, it was to have one among them express "a relationship to things unseen, to the vital forces of nature."[31] Neo-Darwinians will argue that this is a classic example of art developing as a form of sympathetic magic on the sketchy assumption that every human ability must be explained in terms of evolutionary advantage. The paleolithic artist and his or her tribal patrons, on the neo-Darwinian line, were trying to control their dangerous environment. Ostensibly for the same reason that paleolithic artists developed the skills needed to produce the exquisite paintings of the Lascaux caves, our Cro-Magnon

---

[29] Frederick L. Coolidge, Thomas Wynn, *How to Think like a Neandertal* (New York: Oxford University Press 2012), 112f.

[30] See, Jill Cook, *Ice Age art: Arrival of the Modern Mind* (London: The British Museum Press, 2013).

[31] *The Beginnings of Belief, "Living with the Gods,"* Neil MacGregor, Jill Cook, aired October 23rd 2017 on BBC Radio, https://www.bbc.co.uk/programmes/b099xhmj

fore-bearers are assumed to have been simply trying to get an edge on the large mammals competing with them for survival. However, it is just as reasonable to assume that ice age artists were doing the same thing we do when we make art, or make it possible for some of us to develop the skills needed to do so, by subsidizing the lives of artists with grants and scholarships: they were, in Langer's language, 'symbolically transforming' their common experience and so elevating signs, and their minds which depend on them, from the practical and indexical into the symbolic and the domain of meaning. They were using signs as symbols for the sake of contemplating the meaning of their day-to-day reality, and they were doing it for no other reason than that it pleased them to do so. By contemplating the form of the divine in the shape of the Löwenmensch, they were also contemplating themselves, for to think anything symbolically or contemplatively is to also think the thinking that thinks the thing. Indeed, as phenomenologists have been arguing for a century, we only think ourselves thinking by thinking about something.[32]

[see figure on next page]

---

[32] See, Edmund Husserl, *Cartesian Meditations: An Introduction to Phenomenology*, trans. Dorion Cairns (Netherlands: Kluwer Academic Publisher, 1999), 33-37; Lonergan, *Insight*, 344-6. Aristotle, *Metaph.* 1072b20-25: "Thought thinks on itself because it shares the nature of the object of thought: for it becomes an object of thought in coming into contact with and thinking its objects, so that thought and the object of thought are the same. For that which is *capable* of receiving the object of thought, i.e., the essence, is thought. But it is *active* when it possesses this object. Therefore, the possession rather than the receptivity is the divine element which thought seems to contain, and the act of contemplation is what is most pleasant and best."

Löwenmensch, from Hohlenstein-Stadel, now in Ulmer Museum, Ulm, Germany, the oldest known anthropomorphic animal-human statuette, Aurignacian era, c. 35–40,000 BP. Public Domain:
https://commons.wikimedia.org/wiki/File:Loewenmensch1.jpg
https://commons.wikimedia.org/wiki/File:Loewenmensch2.jpg

Symbolic thought, by distinction from significative thought, is the condition for the possibility of *rational* consciousness. Consciousness need not be rational, as we see from its instantiation in other animals and in ourselves some of the time; it is often nothing more than a complex response to sensation, and so continuous with the stimulus response found in the simplest living organisms, in plants as well as simple animals. The human difference is something beyond sensitive or 'empirical

consciousness.'[33] It consists not only in the awareness of sensitive experience and the capacity to imaginatively respond to it, but in the capacity to *transcend* our subjectivity and inquire into, and to some degree understand, the nature of that which we experience.

3. Revisiting (with Nagel) the Argument against Functionalism

This capacity for symbolically mediated objectivity has been repeatedly invoked by philosophers of mind to refute the so-called functionalist argument. Rooted in Alan Turing's test of the same name, designed to prove the indiscernibility of a sufficiently sophisticated mechanical response to a question from a human response, and the later Ludwig Wittgenstein's behaviorism, the functionalist argument holds that for a machine to be considered intelligent it is enough for it to respond and act in the same outwardly visible fashion that a human being responds. The counter argument holds that a generally intelligent machine would need to not only do what humans do, but also do it in the *way* humans do it. It would need to act for *reasons*, that is, its acts would need to be judgments and decisions, i.e., the result of a reasoning process, which is oriented to structures of intelligibility that are not reducible to our thinking them. One can memorize mathematical formulae without understanding them. And when one thereby 'solves' math problems, one is acting in the same way that a machine responds to input on a keyboard. The machine does not understand that 2+3=5; it responds to the input in the way it is determined to respond. An elephant can be trained to use a paint brush and produce abstract pictures that can be sold for a good price on the art market.[34] But no one seriously believes that the elephant is making art for the same reasons that the human being makes art. Rational consciousness appears to require more than the capacity to respond to stimuli; it appears to be more than a mechanical reaction: it judges states of affairs and whenever it does so correctly, it reaches beyond the circumstances and the practical need of the judger. To judge rationally, whether of a matter of fact or of concern, is to transcend need and circumstance and affirm or deny the truth of what is at issue. How exactly humans do this, and why they should have evolved in such a way as to be able to do it, is the theme of Nagel's *Mind and Cosmos*.

According to Nagel, a reductionist theory of evolution, which would explain mind in terms of the evolution of material processes, and so all animal behavior in terms of naturally selected advantages, cannot make sense of rational judgment. One way it deals with this problem is by denying the existence of mind altogether. Nagel

---

[33] Lonergan, *Insight*, 346.

[34] Suda the Elephant, "The Truth About Elephant Paintings Part 1," YouTube Video, 8:55, Maetaeng Elephant Park & Clinic in Chiang Mai, Thailand, accessed from https://www.youtube.com/watch?v=gjOydUjjDos, https://elephantartonline.com/

notes that the denial of the existence of mental states was also the strategy of 20th century behaviorist philosophers of mind, such as Gilbert Ryle and Wittgenstein. Nagel posits that "the names of mental states and processes were said not to be referring expressions. Instead, mental concepts were explained in terms of their observable behavioural conditions of application—behavioural criteria or ascertainability conditions rather than behavioural truth conditions."[35] The problem with these arguments, according to Nagel, is that they leave out exactly that which is to be explained, the first-person experience of being a mind, 'what it is like' to be conscious of something: "The way sugar tastes to you or the way red looks or anger feels, each of which seems to be something more than the behavioural responses and discriminatory capacities that these experiences explain."[36] Assuming that denying the existence of mental states and reducing understanding to observable rule following is not on, Nagel concludes that "conscious subjects and their mental lives are inescapable components of reality not describable by the physical sciences."[37] Along a similar line of argumentation, mental states cannot be held to be identical to the brain states that underlie them. It is conceivable that there could be brain states without any mental states.[38] Therefore, if there is something called a mental state, it is not identical to a state of the brain or any other material configuration for that matter (e.g., the circuitry of a computer); as such it cannot be explained as only a product of material evolution.

Nagel's main argument zeroes in on the objectivity of judgment, whether epistemic judgments, concerning an objective state of affairs, or moral judgments, concerning right and wrong. Along the materialist neo-Darwinian line, he notes, a judgment can be nothing more than a strategic, self-interested move by an organism trying to get one up on its competitors in evolution. If this were true, then the history of science, and human morality—indeed, all our cultural achievements, from ancient religion to quantum physics and modern art—must equally be explicable as naturally-selected products of evolution. The capacities for science and art could only have developed in us because they gave us an evolutionary advantage. There would therefore be no sense in speaking about objectivity or truth, then or now, for evolutionary determinism is still driving our minds. As the most recently randomly selected bundle of animal attributes, we only call something true or false because it is in our interest to so call it.[39] But this would mean that the theory of evolution itself is held to be true, not because it offers us the more coherent and adequate account of the facts of geological time, but because it is in our interest to affirm it as true. Should

---

[35] Nagel, *Mind and Cosmos*, 38.

[36] Nagel, *Mind and Cosmos*, 38.

[37] Nagel, *Mind and Cosmos*, 41.

[38] See, David Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (New York: Oxford University Press, 1996), 96.

[39] This evolutionary relativism is at the heart of Friedrich Nietzsche's perspectivism and historicism. For more on this topic, see Nietzsche's early work, Friedrich Nietzsche, *The Use and Abuse of History* (1874), trans. Adrian Collins (New York: Dover Publications, 2019).

creationism prove more advantageous (and for a sizeable minority, the jury is still out on this), then creationism will be justified as true over evolution. A theory is not in our interest because it corresponds to the fact, but because thinking it so gets us one up. Plainly, however, the intention of the scientist who insists on the truth of evolution against his objector, for example, Richard Dawkins debating Rowan Williams at the University of Oxford in 2012, is not to advance his thesis because he believes it to be more advantageous to believe it (although he might also think that), but because he believes it to be true, and the other thesis to be false.[40] For reasons such as this, Nagel argues that any theory of evolution which purports to explain the mental in terms of the physical and to reduce the human difference to a naturally selected evolutionary advantage, commits the 'functionalist' fallacy. It collapses the reasons for a judgment into the outwardly performed act of judging itself. We no doubt developed the capacity to reason in the course of evolution, but reason itself is not a mere expression of natural self-interest. "Merely to identify a cause is not to provide a significant explanation, without some understanding of why the cause produces the effect," Nagel writes, in effect repeating Socrates' objection to Anaxagoras.[41]

Of most interest to our work is Nagel's distinction between consciousness and reason.[42] Consciousness in its simplest forms might be merely sophisticated stimulus response and so explicable as having evolved because of the natural advantage it gives certain forms of life over others, but intelligence does not merely self-interestedly react to stimuli but rather disinterestedly responds to objective truth and value. Indeed, the affirmation of a truth is often not in our interest as individual (witness the coincidence of climate change denial among shareholders in the oil industry); one could by extension imagine that some truths are not in our interest as a species. The capacity to intelligently respond to truth with a reasoned judgment about the state of affairs regardless of what the judger would prefer to believe, cannot be solely determined by evolutionary advantage. "Thought and reasoning are correct or incorrect in virtue of something independent of the thinker's beliefs, and even independent of the community of thinkers to which he belongs."[43]

Nagel is hardly the first to draw the distinction between consciousness and reason, which can be traced back to Aristotle, and in its Aristotelian registers has been most developed by Lonergan as the difference between empirical and intelligent consciousness.[44] Nor is Nagel the first to use the distinction to refute a reductionistic, materialist account of mind. Few remember that Edmund Husserl's phenomenology originated in a debate with what was then called 'psychologism,' the argument,

---

[40] See, Nagel, *Mind and Cosmos*, 81: "Any evolutionary account of the place of reason presupposes reason's validity and cannot confirm it without circularity."

[41] Nagel, *Mind and Cosmos*, 45.

[42] Nagel, *Mind and Cosmos*, 71f; Lonergan, *Insight*, 346-8.

[43] Nagel, *Mind and Cosmos*, 72.

[44] Lonergan, *Insight*, 346-8.

emerging out of late 19th century positivism, that judgments are nothing more than the effect of certain psychological conditions or events. Psychologism amounted to a denial of the validity of logic in Husserl's view. Logic had to be more than a psychological condition determining how we should judge; rather the validity or invalidity of a judgment must be logically independent of the judgment. Husserl posited that "logical laws, taken in and for themselves, are not normative propositions at all in the sense of prescriptions, i.e., propositions which tell us, as part of their *content*, how one *should* judge."[45] Inspired by Husserl's argument, and especially that of Husserl's star student, Emil Lask, Heidegger wrote his doctoral dissertation defending logic against psychologism.[46] For Heidegger, the undeniable and over-ruling sense of logical validity is a phenomenological indication that judgment transcends the psychological conditions that might accompany it.[47]

The key to the distinction between merely sensitive consciousness and rational consciousness is judgment. All consciousness is intentional, but not all consciousness is or needs to be judgmental. With judgments, either noetic or evaluative, we enter what Robert Sokolowski calls 'the space of reasons.'[48] For a machine to do most of the things we do, it need not possess rational consciousness. But for a machine to supplant us on the planet, it must assume the power and the risk of judgment. AI may improve on us with regard to calculative ability and efficiency at optimizing the conditions of human flourishing, but it will not replace us as the mind of nature, the microcosmic mirror of the whole, so long as it does not possess the capacity to judge and decide. Without symbolic consciousness, which would allow it the distance from its being to make judgments of truth and falsehood, right and wrong, it will be merely a hyper-efficient animal.[49]

---

[45] Husserl, *Logical Investigations I*, 101.

[46] See, Martin Heidegger, "Die Lehre vom Urteil im Psychologismus (1913)," in *Gesamtausgabe 1, Frühe Schriften, 1912–1916*, ed. Friedrich-Wilhelm von Herrmann (Frankfurt am Main: Vittorio Klostermann GmbH, 1978); Emil Lask, "Die Logik der Philosophie und die Kategorienlehre. Eine Studie über den Herrschaftsbereich der Logischen Form," in *Gesammelte Schriften, Band 2*, ed. Eugen Herrigel (Tübingen: J. C. B. Mohr, 1911).

[47] Sean McGrath, *The Early Heidegger and Medieval Philosophy: Phenomenology for the Godforsaken* (Washington: The Catholic University Press, 2006), 93-7.

[48] To think is not to behave in a certain way or to be determined by certain brain-dependent mental events (which may be necessary to thought but are not sufficient for it). To think is to 'enter as agents into the space of reasons. See, Robert Sokolowski, *Introduction to Phenomenology* (New York: Cambridge University Press, 2000), 116. CF. Nagel, *Mind and Cosmos*, 79: "When we rely on systems of measurement to correct perception, or probability calculations to correct intuitive expectations, or moral or prudential reasoning to correct instinctive impulses, we take ourselves to be responding to systematic reasons which in themselves justify our conclusions, and which do not get their authority from their biological origins."

[49] See, Hooke, "Martin Heidegger's Concept of *Understanding* (*Verstehen*)" 19: "Authentic "having" is one necessary feature of human intelligence that avoids competing with the exponential growth of AI's outcome-based achievements. The success of AI (and AGI) is measured based on the results of their programming. This species of pragmatism is hopelessly ontic. It attempts to reveal and provide a service for things (*pragmata*) on hand, without concern for the structure of experience. AI

3. Conclusion

This, then, is the Holy Grail of AGI research: not only the functional reduplication of the activities which we now associate with NHI, but the design of a machine that will do the things we do *in the same way* that we do them, albeit with much greater efficiency and evolutionary capacity. The aim of strong AGI is nothing less than the mechanical reduplication of the human difference. The most ambitious and speculatively inclined AGI researchers are not assuming a weak sense of consciousness such as might be predicated of all beings capable of responding to stimuli, from the sea urchin to the robot, but a strong sense of consciousness, consciousness as the capacity for objective, rational judgment, for knowledge in the full sense of the term—*theoria*, not just *praxis*, and *poiesis*, not just *techne*, and therefore consciousness that can produce imputable judgments. Regardless of whether or not such a thing proves possible, the aim itself forces philosophy to clarify how rational judgment and decision distinguish human consciousness from other forms of consciousness, and what are its material and immaterial conditions. In order to be able to ascertain whether this will have been achieved, we will need to be clear on what the human difference is.

Until a machine gives us reasons to think that it has attained symbolic consciousness and that it now, like us, takes a *theoretical* interest in questions of truth and falsehood, that it too is sometime driven by a disinterested desire to know, that is, to contemplate the meaning of its existence, we will have no reason to recognize it as intelligent, in the human sense of the term. To the question, what would count as evidence? we can only point to those cultural products which most plainly exhibit our contemplative impulse and capacity for symbolic mediation, that is, to art, philosophy, and religion. A machine that had become artistically expressive, philosophically perplexed, or religious would indeed be worthy of our recognition as a rational agent. Of course, it might always be duping us for its own evolutionary advantage. We could never be sure, just as we are never so sure about each other.

---

programmers are incentivized by technocratic control and dominance, leaving no place for the "passive" call of conscience or self-understanding regarding the ontological notion self-actualization."