

The moral roots of conceptual confusion in Artificial Intelligence research¹

Niklas Toivakainen
University of Helsinki

Introduction

I gather that it would not be an overstatement to claim that the field of Artificial Intelligence (AI) research is perceived by many to be one of the most fascinating, inspiring, hopeful, but also one of the most worrisome and dangerous advancements of modern civilisation. Not only does AI research and related fields such as e.g. neuroscience promise to replace human labor, to make it more efficient, to integrate robotics into social realitiesⁱ and to enhance human capabilities. To many, AI represents or incarnates an important element of a new philosophy of mind, contributing to a *revolution in our understanding of humans and life in general*, which is usually integrated with a vision of a new era of human and super human intelligence. With such grandiose hopes invested in a project it is but surprising that the same elements that invoke hope and enthusiasm in some, generate anxiety and disquietude in othersⁱⁱ.

While I will have things to say about features of these visions and already existing technologies and institutions, the main ambition of this paper is to discuss what I understand to be a pervasive moral dimension in AI research. To make my position clear from the start, I do not mean to say that I will discuss AI from a moral *perspective*, as if it could be discussed from other perspectives *detached from* morals. I admit that thinking about morals in terms of “a perspective” is natural *if* one thinks of morality as corresponding to a theory about a separable

¹ This paper was published in *Philosophy and Computers*, Newsletter of American Philosophical Association, Vol. 14 (2) 2015. The published version is freely available online at <http://c.ymcdn.com/sites/www.apaonline.org/resource/collection/EADE8D52-8D02-4136-9A2A-729368501E43/ComputersV14n2.pdf>.

and distinct dimension or aspect of human life, and that there are other dimensions or aspects, say, scientific reasoning for instance, which are essentially *amoral* or “neutral” with respect to morality. Granting that it is a common trait of modern analytical philosophy and scientific thinking to precisely presuppose such a separation between fact and morality (or “value” as it is usually perceived), I am quite aware that moral considerations enters into the discussion of AI (as is the case for all modern techno-science) as a distinct and separate consideration. Nevertheless I will not be concerned here with a critique of moral evaluations relevant for AI research — as for instance an ethics committee would be — but rather with *radicalizing* the relationship between morality and techno-scienceⁱⁱⁱ. My main claim in this paper will be that the project of AI — as the project of any human endeavor — *is itself* inextricably a moral matter. Much of what I will be doing here is to try and articulate how this claim makes itself seen on many different levels in AI research. This is what I mean by saying that I will discuss the moral dimensions of AI.

AI and techno-scientific understanding of nature

The term “Artificial Intelligence” invites three basic philosophical, i.e. conceptual challenges: what *is* (the meaning of) “Artificial” and “Intelligence” and what is the idea of these two coupled together. For instance, if one takes anything “artificial” to be categorically (conceptually, metaphysically?) distinct from anything “genuine”, “real” or “natural” — which it conceptually seems to suggest — and if we think it sufficient (for a given purpose) that “intelligence” be understood as a computational/mechanical process of some sort, then any chess playing computer program, not to speak of the new master in Jeopardy IBM’s “Watson”^{iv}, would be perceived as real and successful tokens of AI (with good future prospects for advancement) and would not invoke any philosophical concerns in us. But as can be observed when looking at the diverse field of AI research, there are many who do not think that chess playing computers or Jeopardy master Watson display “intelligence” in any “real” sense; that

“intelligence” is not simply a matter of computing power. Rather, they seem to think that there is much more to the concept of “intelligence” and how it relates to the concept of (an actual human) *life* than machines like Watson encompass or display. In other words, the dissatisfaction with what is perceived as a limited or narrow conception of intelligence invites the need for philosophical reflection as to what “intelligence” really means. Later coming back to the concept of “intelligence”, let us though begin by considering the role the term “artificial” plays in this debate and the philosophical and ideological weight it carries with itself.

Suppose we were of the opinion that Watson’s alleged “intelligence”, or any other so called “Good Old Fashion Artificial Intelligence”^v, does not satisfy essential features of intelligence of the “sort” human intelligence builds on and that “more” is needed, say a body, autonomy, moral agency etc. We might think all of this and still think that AI systems *can never become*, out of conceptual necessity, anything more than technological devices or systems, albeit very sophisticated and human or animal like: there will always, so to speak, be an essential difference between a simulation and a real or natural phenomena — this is what the term “artificial” conceptually suggests. But as we are all aware, this standpoint is not shared by all and especially not within the field of AI research and much of “naturalistic philosophy of mind”, as the advocates of what is usually termed “strong AI” hold that AI systems can indeed become “real” or “genuine” “autonomous” “intelligent” and even “conscious” beings^{vi}.

That people can entertain such visions and theories about AI systems one day becoming genuinely autonomous conscious intelligent beings without feeling that they are committing elementary conceptual mistakes, derives from the somewhat dominant conception of the nature of concepts such as “artificiality”, “life” and the “natural/genuine” deep at the heart of the modern techno-scientific self-understanding or worldview. As most of us are aware, modern science developed into its paradigmatic form during the 17th century, reflecting a sort of culmination point of huge social, religious and political changes. Seen from the perspective of scientific theory and method, the

founders and visionaries of modern science turned against the ancient Greek and medieval scholastic “contemplative” natural philosophy, devising new methods and practices which built on (very) different ideologies and aspirations.

It would take not one but many volumes to clarify all the different (trans)formative forces that led up to the birth of the new methods and cosmology of modern techno-science, and many good books have been written on the subject^{vii}. Nevertheless, I shall shortly try to summarize what seems to me – with regards to the topic of this paper – to be some of the decisive differences between modern science and its ancient and medieval predecessors. We begin by noting that in the Aristotelian and scholastic natural philosophy, knowing what a thing *is*, was (also and essentially) to know its *telos* or purpose as it was revealed through the Aristotelian four different causal forces, and especially the notion of “final cause”^{viii}. Further, within this cosmological framework “Nature” or “the natural” stood for that which creates itself or that which is *essential* — and so that which is created by human hands is of a completely different order. Thirdly, both Plato and Aristotle had placed the purely theoretical or formal arts or knowledge *hierarchically* above “practical” knowledge or know-how (arguably reflecting the political and ideological power structures of the ancient Greek society). On the other hand, in the paradigm of modern science, knowing what a thing *is*, is to know how that thing functions, how it is “constructed”, how it can be controlled and manipulated etc. Similarly, in the modern era the concept of “Nature” or “the natural” loses its position as that which is essential and instead becomes more and more perceived as the raw material for man’s industriousness. So in contrast to the Platonic and Aristotelian glorification of the purely theoretical or formal arts/knowledge the 17th century philosophers drew on a new vision “of the importance of uniting *theoria* with *paraxis*, a vision that grants new prominence to human agency and labor”^{ix}. In other words, the modern natural philosopher and scientist sought for a knowledge that would enable him to *dominate* natural phenomena.

This was the cornerstone of Francis Bacon’s scientific revolution. For Bacon, as for his followers —arguably the whole project of modern techno-science — the

duty of human power was to manipulate, change and refine corporeal bodies, thus conceptualizing “knowledge” as the capacity to understand *how* this is done^x. Hence Bacon’s famous term “*ipsa scientia potestas est*” or “knowledge itself is power”. This same idea can also be found at the heart of the scientific self-understanding of the father of modern philosophy and modern dualism (which also sets the basis for much of the philosophy and theory of AI), namely in Descartes articulations. In explaining the virtues of the new era of natural philosophy and its methods he proclaimed that they will “render ourselves the masters and possessors of nature”^{xi}.

Now the main point of this short and crude survey is to try and highlight that had not the modern scientific paradigm been built on a unity between *theoria* and *praxis* and the ideas of the duty of man to dominate over nature, we would not have read Bacon proclaiming that the *artificial does not differ from the natural either in form or in essence, but only in the efficient*^{xii}. For, as in the new Baconian model, when nature loses (ideologically) its position as *essential* and becomes predominantly the raw material for man’s industriousness, nature (and thus life) itself becomes *nothing apart from how man knows it or will someday be able to know it* – and here “knowledge” is conceptualized as that which gives *power* over phenomena. And even more to the point, had such decisive changes not happened we would not be having philosophical discussion about AI in its “strong” sense — i.e. in the sense that the “artificial” can gain the same ontological status as the “real” or “natural”: when such a conceptual change has been made, when the universe is perceived as essentially in no way different than an artifact or technological device, when the cosmos is perceived to essentially be captured through techno-scientific knowledge, then the idea of an AI system as a genuinely autonomous conscious intelligent being becomes a thought to entertain.

As I have pointed out, this modern and Baconian idea is echoed in thinkers all the way from Descartes — whom perceived all bodily functions as essentially mechanical and subject to technological manipulation/control^{xiii} — to modern “naturalist functionalists” (obviously denying Descartes substance dualism) who

advocate AI in its “strong” sense and suggest that life and humans are “*made of* mindless robots [cells] and nothing else, no non-physical, nonrobotic ingredients at all”^{xiv}. Claiming such an essential unity between nature and artifact obviously goes, so to speak, both ways: machines and artifacts are essentially no different than nature or life, but the main argument and emphasis is really that nature and life are essentially no different from artifacts. In other words, I would claim, what is expressed here — in the modern techno-scientific understanding of phenomena — is the idea that it is the artificial (i.e. human power) that is the primary or the essential. I will characterize this ideologically based conception as *a technological or techno-scientific understanding of nature, life and being*. Now the claim I will attempt to lay out is that such a technological understanding is, in contrast to how it is usually perceived, not simply a question of neutral objective facts, but rather an understanding or perspective that is highly morally charged. In the last part of the paper I will try to articulate in what sense (or perhaps a particular sense in which) this claim has a direct bearing on our conceptual understanding of AI.

Is techno-scientific understanding amoral?

The reason that I pose the question of techno-science’s relation to morality is that there resides within the self-understanding of modern techno-science, an emphatic separation between fact and value (as it is usually termed). It may be added that modern science is by no means the only institution in our modern culture that upholds such a belief and practice. In addition to the institutional cornerstone of modern secular societies, namely the separation between state and church, the society at large follows a specialization and differentiation of tasks and authorities/hierarchies^{xv}. Techno-science is one, albeit central, of these differentiated institutions. Now despite the fact that modern techno-science builds strongly on a kind of unity between theory and practice — the truth of a scientific theory is shown by the power of manipulation it produces — it simultaneously developed, due to diverse reasons, a self-image of political and

value (moral) neutrality: a science for the sake of science itself^{xvi}. This meant that while the measure of knowledge was directly related to utility, power of manipulation and control^{xvii} it was thought that this knowledge could be attained most efficiently and purely when potentially corrupt *individual* interests of utility or other values were left outside the methods, theories and practices of science^{xviii}. This principle gives modern science its specific, specialized and differentiated function in modern society: as the producer of “objective” techno-scientific knowledge.

One of the main reasons for calling scientific knowledge “neutral” seems to be founded on an urge to as much as possible detach it from the “use” this knowledge is put to: it can be “misused”, but this is not to be blamed on the institution of science, for it (ideally) deals purely with objective facts. The real problem, one often hears, is the politico-economic power structures that pervert scientific knowledge in pursuit of corrupted ends. This is why we need political regulation, for we know that scientific knowledge has high potency for power and thus destruction or domination. This is why we need ethics committees and ethical regulations, because science itself is unable to ethically determine its moral status and regulate its domain of action: it only deals itself with supposedly amoral objective facts.

I am of course not indicating that *scientists* are morally indifferent to the work they do. I am simply aiming at pointing out that as a scientist in the modern world, one’s personality as a scientist (dealing with scientific facts) is differentiated from one’s moral self-understanding in any other sense than the alleged idea that science has an inherent value in itself. Obviously any scientist might bring their moral selves with them to work and into the laboratories, so the split does not have to occur on this level. Instead the split finds itself at the core of the idea of the “neutral and objective” facts of science. So when a scientist discovers the mechanisms of, say, a hydrogen bomb, the mechanism or the “fact of nature” is itself perceived as amoral — it is what it is neutrally and objectively: the objective fact is neither good nor evil, for such properties do not exist in a disenchanting, devalored and rationally understood nature; nature

follows natural (amoral) laws that are subject to contingent manipulation and utilization^{xix}.

One problem with such a stance relates to what I will call “the hypocrisy of modern science”. On a more fundamental level I would challenge the very idea that scientific knowledge of objective facts of nature/reality is itself “neutral” with respect to morals. Now to begin outlining what I mean with the “hypocrisy of modern science” let me start by noting that the dawn of modern science carries with itself a new, perhaps unprecedented, democratic principle of open accessibility^{xx}. In addition to the Cartesian idea that “Good sense or Reason, is by nature equal in all men”^{xxi}, one might say that the democratic principle was engraved in the method itself, for it was the right *methods* of modern science, not aristocratic or elite minds, that were to produce true knowledge “as if by machinery”^{xxii}. Hence the new ideology and its methods — both Bacon’s and Descartes’ — were to put men on “an equal footing”^{xxiii}. Although the democratization of knowledge was part of the ideology of Bacon, Descartes and the founders of The Royal Society, the concrete reality was and is a completely different story. Just as a short telling example, the Royal Society, founded in 1660, did not have a single female member before 1945. Nor has access to the scientific community ever been detached from individual’s social backgrounds and positions (class), economic possibilities etc., not to speak of cultural and racial factors. There is also the issue of how modern science is connected to forms of both economic and ecological exploitation: modern science with its experimental basis is and has always been highly dependent on large investments and growing capital, capital which at least historically and in contemporary socio-economic realities builds on exploitation of both human as well as natural resources^{xxiv}. Nevertheless, one might argue, such prejudices are more or less part of an unfortunate history and today we are closer to the true democratic ideals of science which have always been there: so we can still hold on to a separation between fact and morals.

All the same, there is another form of hypocrisy that finds itself deep in the roots of modern science and alive and well, if not even strengthened, even today. As

both Bacon and Descartes clearly noted, the new methods of modern science were to make men “masters and possessors of nature”^{xxv}. But the new methods of science would not come only to serve man in his domination over nature, for the power that this new knowledge gave also served man in his domination over man^{xxvi}. As one may quite easily observe when looking at the interconnectedness of the foundations of modern science with political and economic interests of the newly formed nation states of Europe and the Americas, it becomes clear that the history of modern techno-science runs in line with modern military industry and technologies of domination^{xxvii}. For example, Galileo *also* used his own calculations of falling objects in order to calculate ammunition projectile trajectories while Descartes’ analytical geometry very quickly became utilized for improvements of ballistics^{xxviii}. And in contrast to the democratic spirit of modern science — which perhaps can be said to have made some “progress” — the interconnectedness of techno-science and military and weapons research and development (R&D) (and other forms of exploitation/destruction) is still very tight. That is to say, while it is certainly true that modern techno-science is not in any sense original in its partnership and interdependence with military and weapons R&D, it nevertheless, in its conceptual and methodological strive to gain power over phenomena, has created unprecedented means of destruction, domination and oppression — and we must not forget, means of construction and perhaps even liberation. In other words, modern techno-science has not *exclusively* built on or led to dreams of liberation and diminishment of suffering (as it quite often rhetorically promises) but, as one might put it, the complete opposite.

In 1975, the Stockholm International Peace Research Institute’s annual books record that around 400 000 scientists, engineers and technicians (roughly half of the entire world’s scientific man power at that time!) were committed to and engaged with weapons research^{xxix}. At least since the Second World War up until, say, the late 1980’s, military technology R&D relied mostly on direct funding by the state, as state policy (at least in the US) was dominated by what is usually called “spin-off” thinking. The term “spin-off” refers to the idea and belief that

through heavy funding of military R&D also the civilian and commercial sectors will benefit and develop. So, as it was perceived, as military R&D yielded new high-tech devices and related knowledge, some of this knowledge and innovations would then “flow downstream” and finds its place in the civilian commercial markets (in appropriate form). This was, arguably, one of the main “legitimatising” reasons for the heavy numbers of scientists working directly for military R&D.

But this relationship has changed now (if it ever really was an accurate description). For instance, in 1960 the US Department of Defence funded a third of all Scientific R&D in the Western world, whereas in 1992 it funded only a seventh of it^{xxx}. Today this figure is even lower due to a change in the way military R&D relates to civil commercial markets. Whereas up until the 1980’s military R&D was dominated by “spin-off” thinking, today it is possible to distinguish at least up to eight different ways in which military R&D is connected to and interdependent with civil commercial markets, spanning from traditional “spin-off” to its opposite “spin-in”^{xxxix}. The modern computer and supercomputer, for example, are tokens of traditional spin-off and “*Defence procurement pull and commercial learning*” and the basic science that grew to become what we today know as the Internet, stems from “*Shared infrastructure for defence programs and emerging commercial industry*”^{xxxix}. The case of *Deep Brain Stimulation* (DBS), which is used to treat symptoms related to Parkinson’s disease and people suffering from *essential tremor*^{xxxix}, and which falls under the category of “Brain Machine Interfaces” and has its relevance for AI research, will serve as another telling example of the complex and interconnected web of techno-science and the military industrial complex. Developed within the civilian sector, DBS and related knowledge and technology are perceived to be of high importance to military R&D. An official NATO report document from 2009 makes the following observation: “From a military perspective, knowledge [neuroscientific knowledge] development should focus on three transitions: 1) from clinical and patient applications to applications for healthy users, 2) from lab (or controlled) environments to the field, and 3) from *fundamental*

knowledge to operational applications^{xxxiv}. I emphasized the third transitional phase suggested by the document in order to highlight just how fundamental and to the point Bacon's claim that "knowledge itself is power" is and what the unity between theory and practice means in the modern scientific framework: Techno-scientific knowledge of the kind derived for example from neuroscientific and cognitive science research, not only lends itself, but *co-creates* the interdependence between basic scientific research and the military industrial complex, and finds itself everywhere in between "spin-off" and "spin-in" utilisation.

Until today, the majority of applied neuroscience research is aimed at assisting people who suffer from a physical, perceptual or cognitive challenge and not at performance enhancement for healthy users. This situation opens up opportunities for *spin-off and spin-in* between advanced (military) Human System Interaction knowledge and the accomplishments in neurotechnology for patients^{xxxv}.

We should be reminded here that the military-industrial complex is just *one* frontier that displays the interconnectedness of scientific "fundamental knowledge" and end specific utilization ("the means constraint the ends"^{xxxvi}). Just quickly adding to this, we might just as well think of the interconnectedness of basic scientific knowledge in agricultural research and the food markets^{xxxvii} or scientific research of the human and other genomes and for example the drug industry. But I take the case of military R&D to suffice for the point I am making.

Now despite the historical and ongoing (and increasing) connection between modern science and military R&D and other exploitative forces, I am aware of the fact that this connection *can* be perceived to be contingent rather than essential — this is why I called the above a discussion of the "hypocrisy" of modern science. In other words, one may claim that on an essential and conceptual level we might still hang on to the idea of science and its "fundamental knowledge" as "neutral" — although I find it somewhat worrisome that due to reasons described above, alarm bells aren't going off more than they are. Part of the difficulty with coming to grips with the neutrality status of modern science is that

the issue is connected on two different levels. On the one hand, the neutrality of science has been integrated into its methods and to its whole *ethos* when modern science struggled to gain freedom from church and state control since the 17th century^{xxxviii}. Related to this urge to form an institution free from the grips of religious and political power structures and domination, neutrality with respect to value has become an important criterion of “objectivity”: if the methods of science are free from the distorting, corrupting and vulnerable values of individual humans, then only can it be guided in a pure form by the objective stance of rational reason. But, one might ask, is it really so that if science was not value free, and more importantly, if it was essentially morally charged by nature, it would be deprived of its “objectivity”?

To me it seems that “objectivity” is not at all dependent on value neutrality in any absolute sense, or rather, not dependent on being amoral. Of course this does not mean that certain values, perceived by individuals owing up to, say, certain social norms and conventions, might not distort the scientific search for “objectivity”, not to speak of objectivity in other forms of knowing and understanding. Obviously it might do so. The point is rather that “neutrality” and “objectivity” is not the same thing.

Neutrality refers to whether a science takes a stand; objectivity, to whether a science merits certain claims to reliability. The two need not have anything to do with each other. Certain sciences may be completely “objective” — that is, valid — and yet designed to serve a certain political interest...the fact that their knowledge is goal-orientated does not mean it doesn't work.^{xxxix}

Proctor's point is to my mind quite correct and his characterization of scientific “objectivity” as validity that “works” — something that enables one to manipulate and control phenomena — is of course in perfect agreement with Bacon's definition of scientific knowledge^{xl}. The main lesson here, as far as I can see it, is that in an abstract and detached sense it might seem as if scientific “objectivity” really could be politically and morally neutral (in its essence). Nevertheless, and this is my claim, the conceptual confusion arises when we imagine that “objectivity” can in an *absolute* sense be “neutral” and amoral. Surely any given human practice can be neutral and autonomous relative to specific issues/things, e.g. neutral to or autonomous with respect to prevailing

political ideologies, by which we would mean that one strives for a form of knowledge that does not fall victim to the prejudices of a specific ideology. This should nevertheless not lead us into thinking that we can detach “objectivity” from “knowledge” or “knowing” — as if we could understand what “objectivity” *is* independently of what “knowing” something *is*. In this more pervasive sense objectivity is always dependent, as one might put it, on knowing, while knowing itself is always a mode of life and reflects, what might be called, a moral-existential stance or attitude towards life. The mere fact that we choose to call something “knowledge” draws upon certain values, and more essentially, on a dynamics of aspirations that reflect our stance towards our lives, towards other human beings, other forms of life and “the world”. But the recognition that we have come to call some specific stance towards life and the world “knowledge” also includes the questions “Why do we know what we know and why we don’t know what we don’t know? What *should* we know and what shouldn’t we know? How might we know differently?”^{xii}. By this I mean to say that such questions, moral by nature, are included in the question of “why has *this* gained the status of knowledge?” and “why have we given *this* form of knowledge such a position in our lives”? So the moral question we should ask ourselves is what is the moral dynamics that has lead guiding concepts such as “domination”, “power”, “control”, “artificial”, “mechanization” etc. to become constitutional for (modern scientific) “knowledge”.

Obviously I am aware that many philosophers and theorists would object to the way I seem to be implying that moral understanding is prior to scientific or theoretical understanding, and not, as I gather many would claim, that all moral reasoning is itself a form of proto-theoretical rationalization. My claim is obviously in a sense the opposite, for I am suggesting that in order to understand what modern science and its rationale *is*, we need to understand what lies, so to speak, behind the will to project a techno-scientific perspective on phenomena, on “intelligence”, “life”, the “universe” and “being”. In other words, this is not a question that can be answered by means of modern scientific inquiry for it is this very perspective or attitude we are trying to clarify. So, despite the fact that

theories of the hydrogen bomb lead to successful applications, and can in this sense be said to be “objective”, I am claiming that this objectivity is not and cannot be detached from the political and moral dimensions of a the will to build a hydrogen bomb: from a will to power. Rather, it seems to me that the “objectivity” of the facts of the hydrogen bomb are reflections or manifestations of will for such a bomb (power), for knowledge of the “facts” of, say, a hydrogen bomb shows itself as meaningful, as something worth our attention, *only insofar as we are driven or aspire to search for such a knowledge/power*. In other words, my point is that it is not a coincidence or a contingent fact that modern techno-science has devised means of for instance mass-destruction. As Michel Henry has put it

Their [the institution of techno-science] “application” is not the contingent and possible result of a prior theoretical content; it is already an “application”, an instrumental device, a technology. Besides, no authority (*instance*) exists that *would be different from this device* and from the scientific knowledge materializing in it that would decide whether or not it should be “realized”.^{xiii}

Artificial Intelligence or Artificial Life?

My initial claim has been that *if* there is to be any serious discussion about AI in any other sense than what technical improvements can be made in creating an “artificial” “intelligence” — and thus holding a conceptual distinction between real/natural and artificial — then intelligence or “the mind” must be understood *as technological*. The discussion that followed was meant to suggest that (i) the (modern) scientific worldview *is* a technological (or techno-scientific) understanding of the world, life and of being and (ii) that such an understanding is founded on an interest for utility, control, manipulation and domination — for power — and finally that (iii) modern science is fundamentally and essentially morally charged, and strongly so with the moral questions of power, control and domination.

Looking at the diversity of theories and philosophies of AI one will quite quickly come to realize that AI research is always an interplay between on the one hand a technological demand/challenge and aspiration, and on the other hand a

conceptual challenge of clarifying the meaning of “intelligence”. As the first wave of AI research, or “Good Old Fashion Artificial Intelligence” (GOFAI)^{xliii} built on the idea that high-level symbol manipulation alone could account for intelligence and since the Turing machine is a universal symbol manipulator, it was quite “natural” to think that such a machine could one day become genuinely “intelligent”. Today the field of AI is much more diverse in its thinking and theorizing about “Intelligence” and as far as I can see the reason for this is that people have felt dissatisfaction not only with the kind of “intelligence” the “top-down” systems of GOFAI are able to simulate, but more so because people are suspicious with how “intelligence” is conceptualized under the banner of GOFAI. Today there is talk about how cognition and “the mind” is essentially grounded in the body and in action^{xliiv}, thus making “robotics” (the body of the AI system) an essential part of AI systems. We also hear about “situated cognition”, distributed or de-centralized cognition and “the extended mind”^{xliiv}. Instead of top-down GOFAI many are advocating bottom-up “developmental” approaches^{xlivi}.

[L]arge parts of the cognitive science community realise that “true intelligence in natural and (possibly) artificial systems presupposes three crucial properties:

1. The embodiment of the system;
2. Its situatedness in a physical and social environment;
3. A prolonged epigenetic developmental process through which increasingly more complex cognitive structures emerge in the system as a result of interactions with the physical and social environment.”^{xliiv}

My understanding of the situation is that the new emerging theories and practices are an outcome of a felt need to conceptualize “intelligence” or cognition in a manner that more and more resembles how (true and paradigmatic) cognition and intelligence are intertwined with the *life* of an actual (human/living) being. That is to say, there seems to be a need to understand intelligence and cognition as more and more integrated with both embodied and social life itself — and not only understand cognition as an isolated function of symbol-manipulation *alá* GOFAI . To my mind this invites the question as to what extent “intelligence” can at all be separated from the concept of “life”. Or put another way: how “deep” into life must we go to find the foundations of intelligence?

In order to try and clarify what I am aiming for with this question, let us connect the concept of “intelligence” with that of “language”. Clearly there might be a specific moment in a child’s life when a parent (or some other person) distinctly hears the child utter its “first word” — a sound that is recognizable as a specific word and used in a way that clearly indicates some degree of understanding of how the word can be used in a certain context. But of course this “first word” is not a miracle in the sense that before the utterance, the child was completely deprived of language or that it now suddenly “has” language: it is rather a kind of culmination point. Now the question we might ask ourselves is whether there is any (developmental) part of a child’s life — up until the point of the “first word” and beyond — that we could so to speak *skip* without the child losing its ability to utter its “first word” and to develop its ability to use language? I do not think that this is an empirical question. For what we would then have to assume in such a case is that the “first word” is not a culmination of all the interaction and learning that the child had gone through prior to the utterance and this would mean that we could for instance imagine a child that either came into the world already equipped with a “developed” capacity to use language or that we could imagine a child just skipping over a few months (I mean “metaphysically” skipping over them: going straight from, say 1 month old to 5 months old). — But, we might note, in imagining this, we make use of the idea “already equipped with a developed capacity to use language” which all the same builds on the idea that the development and training usually needed is somehow now miraculously endowed within this child. We may compare these thought-experiments with the real case of a newborn child who immediately after birth crawls to his/her mother’s breast, who stops screaming when embraced etc. Is this kind of, what one might call, *sympathetic responsiveness* not constitutive of intelligence and language: if this responsiveness was not there from the start — as constitutive of life itself — how could it ever be established? And could we imagine such an event without the prenatal life in the womb of the mother, all the internal and external stimuli, interaction and communication that the fetus experiences during pregnancy. And what about the pre-fetal stages and conception itself, can these be left out from the development of language and intelligence?

My point here is of course that from a certain perspective, we cannot separate intelligence (or language) from life itself. I say “a certain perspective” because everything depends on what our question or interest is. But by the looks of it, there seems to be a need within the field of AI research to get, so to speak, to the bottom of things, to a conception of intelligence that incorporates intelligence as it is fleshed out in human and other forms of life in its totality — to make the artificial genuine. And if this is the aim, then my claim would be that “intelligence” and “life” cannot be separated and that AI research must try to figure out how to artificialize not only “intelligence” but also “life”. In other words, any idea of strong AI must understand life or being, not only intelligence or “the mind”, technologically, for if it is not itself technological, then how could it be made so.

In the beginning of this chapter I said that AI research is always an interplay between technological aspirations and conceptual enquiry. Now I will add to this that AI is first and foremost driven by a technological aspiration and that the conceptual enquiry (clarification of what concepts like “life” and “intelligence” means or *is*) is only a means to fulfill this end. That is to say, *the technological aspiration shapes the nature of the conceptual investigation: it has predefined the nature of the end result*. What makes the ultimate technological fulfillment of strong AI different from its sibling genetic engineering, is that whereas the latter must in its pursuit to control and dominate the genetic foundations of life always take for granted life itself — it must rely on *re-production* of life: it can only dominate a *given* life — the former aspires in its domination to be an original creator or producer of “intelligence”, and as I would claim, of “life”.

The moral dynamics of the concern for mechanization of intelligence and life

I have gone through some effort in order to make the claim that AI — in its strong sense — presupposes a technological understanding of life and phenomena in general. Further I have tried to make the case that modern science is strongly driven by a technological perspective — a perspective of

knowledge to gain power over phenomena — and that it makes scant sense to detach morals (in an absolute sense) from such a perspective. Finally, I have suggested that the pursuit of AI is determined to be a pursuit to construct an artificial model/simulation of intelligent *life* itself, since to the extent we hope to “construct” intelligence as it is fleshed out in human and other forms of life, it cannot really be detached from the whole process or development of life. What I have *not* said — and I have tried to make this clear — is that I think that modern science or a technological understanding of phenomena and life is invalid or “wrong”, *if* our criterion is, as it seems to be, *utility* or a form of verification that is built on control over phenomena. We are all witnessing how well “it works” and left to its own logic, so to speak, modern science will develop indefinitely — we do not know the limits (if there is such) to human power.

In this final part I want to try and illustrate how what I have been trying to say makes itself shown in the idea of strong AI. My main argument will be that while I believe that the idea of strong AI is more or less implicitly built into the modern techno-scientific paradigm (and is thus a logical unfolding of this paradigm), the rationale behind it is more ancient and in fact reflects a deep moral concern one might say belongs to a constitutive characteristic of the human being. Earlier I wrote that a strong strand within the modern techno-scientific idea builds on a notion that machines and artifacts are no different than nature or life, but that the main argument and emphasis is really that nature and life are essentially no different from artifacts — that it is the artificial, human power, which is taken as primary or essential. Following this suggestion, my concern will now be: what is the dynamics behind the claim that human beings or life itself is formal (since any given AI system would be a formal system) and what kind of understanding or conception of human beings does it build on and what does it overlook, deny and perhaps even repress.

There are obviously logical and historical reasons why drawing analogies between humans and machines is, not only easy (in certain respects), but also tells us something true. Namely, machines have more or less exclusively been created to *simulate* human or animal “behavior” in order to support, enhance,

intensify and replace human labor^{xlviii} and capability^{xlix} and occasionally for the purpose of entertainment. And since this is so, it is only logical that machines have had to build on some analogies to human physiology and cognitive capability. Nevertheless, there is another part to the story — one might call it the other side of the coin — of mechanization that I want to introduce with the help of a quote from Lewis Mumford.

Descartes, in analyzing the physiology of the human body, remarks that its functioning apart from the guidance of the will does not “appear at all strange to those who are acquainted with the variety of movements performed by the different automata, or moving machines fabricated by human industry...Such persons will look upon this body as a *machine* made by the hand of God”. But the opposite process was also true: the *mechanization of human habits prepared the way for mechanical imitations*.¹

It is important to note that Mumford’s point is not to claim any logical priority to the mechanization of human habits over theoretical mechanization of bodies and natural phenomena, but rather to make both a historical observation as well as to highlight a conceptual point about “mechanization” and its relations to human social discipline, regimentation and control^{li}. Building on what I have said earlier, I will take Mumford’s point to support my claim that to both theoretically as well as practically, to mechanize phenomena is always (also) to *force or condition* it into a specific form, to formalize phenomena in a specific way. As Bacon explained the relation between natural phenomena and scientific inquiry, nature reveals her secrets “under constraint and vexed”. Although it is clear that Bacon (as does his contemporary followers) thought that such a method would reveal the “true” nature of phenomena, we should note, or I would claim, that it was and still is *the method itself* which was/is the primary or essential guiding force and thus nature or phenomena had/has to be *forced into a shape convenient to the demands and standards of experiment*^{lii} — this is why we speak of “controlled research environment”. Similarly my claim will be that to theoretically as well as practically — in other words ideologically— mechanize/formalize (human) life, (human) behavior, (human) intelligence, (human) relationships, *is itself to force or condition, so to speak, human nature into a specific form: formalize in a specific way with specific underlying purposes*. Now, as my claim has been, these underlying purposes are essentially something

that must be understood in moral-existential terms — they are the “rationale” behind the scientific attitude to the world and not themselves “scientific objects”. To this I now add that the underlying purposes cannot be detached from *what* (the meaning of) phenomena are *transformed into* under the scientific and mechanizing methods — and this obviously invites the question whether any instance is a development, a re-definition or a confusion, distortion or perversion of our understanding.

Obviously this is a huge issue and one I cannot hope to argue for to the extent that a good case could be made for the understanding that I am advocating. Nevertheless I shall attempt, by way of examples, to bring out a tentative outlining of how this dynamics makes itself shown in human relationships and interaction and how it relates to the idea of strong AI.

Some readers might at first be perplexed as to the character of the examples I intend to use and perhaps think them naïve and irrelevant. Nevertheless, I hope that by the end of the paper, the choice of the examples will be more clear and seen to have substantial bearing on the issue at hand. It might be added that the examples are designed to conceptually elaborate the issue brought up in Mumford’s quote above and to shed light on the dynamics of the idea that human intelligence and life are essentially mechanical or formal.

Think of a cocktail party at, say, the president’s residence. Such an event would be what we would call “formal” and the reason for this is that the expectations on each person’s behavior are quite strict: well organized and controlled, highly determined (although obviously not in any “absolute sense), predictable etc. One is for instance expected not to drink too many cocktails, not to express one’s emotions or desires on the dance floor or otherwise too much, not to be impolite or too frank in one’s conversations and so on: the appropriate and expected behavior follows formal rules. But note, exactly because this is the case, so is its opposite. That is to say, because “appropriate” behavior is grammatically tied to formal rules/expectations, so would also “inappropriate” behavior be: to each

appropriate response and act there are various ways of breaking them, ways which are derived from the “appropriate” ones and become “inappropriate” *from the perspective of the “appropriate”*. So for instance, if I were to drink too many cocktails or suddenly start dancing passionately with someone’s wife or husband, these behaviors would be “inappropriate” exactly because there are “appropriate” ones that they go against. The same goes for anything we would call “informal” since the whole concept of “informal” grammatically presupposes its opposite, i.e. “formal”, meaning that we can be “informal” only in relation to what is “formal”, or rather *seen from the perspective of “formal”*. One could for instance say that at some time during the evening the atmosphere at the party became more informal. One might say that both “formal” and “informal” are part of the same language game. In other words one might think of a cocktail party as a social machine or mechanism into which each participant enters and must use their rational ability to “play” along with the determined or expected rules in relation to their own motivations, and goals, their own fears of social pressure etc.

We all know of course that the formal as well as any informal part of a cocktail party (or any other social institution) is a means to discipline, regulate, control, regiment, effectuate, make efficient, polite, tolerable etc. the way in which human relations are fleshed out: to have formal rules — and all the social conditioning that goes into making humans “obey” these customs — is a way to for instance moderate any political or ideological differences that people might have, to avoid or control embarrassing and painful encounters between people and emotional, passionate and spontaneous reactions and communication etc. In other words a cocktail party is to *force or condition* human nature into a specific formalized form: it is to mechanize human nature and her interpersonal relationships^{liii}. The point to be made here is that understanding the role that formalizing in this sense has, has to include a moral investigation into why human relations create difficulties that need to be managed at all, and what are the moral reactions that motivate to the kinds of formalizations that are exercised.

To make my point a bit more visible, think of a dinner invitation. To begin with we might imagine that the invitation comes with the words “informal dress”, which indicates that the receiver might have had reason to expect that the dress code *could have been* formal: indicating that there is an underlying “formal” pressure in the relationship/invitation. In fact, having “informal dress code” written on an invitation is already a formal feature of the apparently formal invitation. Just the same, the invitation might all together lack any references to formalities and dress codes, which might mean either of three things. (i) It might be that the receiver will automatically understand that this will be a formal dinner with some specific dress code (for the invitation itself is formal), or (ii) it might mean that they will understand — due to the context of the invitation — that it will be an informal dinner, but that they might have had reason still to expect that such invitations usually implied some form of formality (a pressure to understand the relationship as formal). Needless to say though, both of these play on the idea of a “code” that is either expected or not expected. (iii) The third possibility — which is in a sense radical, although a commonly known phenomenon — is simply that the whole idea/concept of formalities/informalities *does not present itself*. That is to say, the invitation itself is neither formal nor informal. If my friend, with whom I have an open and loving relationship, invites me over for dinner, it would be very *odd and indicative of a certain moral tension in our relationship or lack of understanding* if I were to ask him if I should dress formally or informally^{liv}: our relationship is in this sense and to this extent *a-formal*. And, one might say, it will stay *a-formal to the extent no conflict or difficulty arises between us*, potentially leading us to adopt a code of formality *in order to manage, avoid, control etc. the difficulty that has come between us*. There is, so to speak, nothing formal/mechanical *as such* about the relationship or “behavior” and if an urge to formalize comes from either inside or outside it *transforms* the relationship or way of relating to it: it now becomes formalized/mechanized, i.e. it has now been contextualized with a demand for control, regimentation, discipline, politeness, moderation etc. What I take this to be pointing at is that (i) if a relationship does not pose a relational and moral difficulty there will be no need, urge or reason to formalize or mechanize the

relationship. This means that the way we relate to each other in such cases is not determined by social collective identities or roles — at least not dominantly — but is rather characterized by an openness towards each other. (ii) This indicates that mechanization or codification of human relationships and behavior is a *reaction* to certain phenomena over which one places a certain demand of regulation, control etc.

So a cocktail party attendee does not obviously have to understand his or her relationship to other attendees in terms of formal/informal, although the social expectations and pressures might do so. If an attendee meets a fellow attendee openly, kindly and lovingly, as opposed to “politely” (“politely” being a formal way of relating to another, hence part of a “mechanism”), then there is *no* mechanism or determined cause or course of action to specify. Rather such an encounter is characterized by an *openness* (and to which extent it is *open* depends on the persons in the encounter), in which persons encounter each other, at least relatively, independent of what their social collective identities prescribe to them, so to speak, as an *I* to a *you*. In such an openness, as far as it is understood *in* this openness, there is no technological knowledge to be attained for whereas technological understanding always includes a demand over (to control and dominate) phenomena, in an (morally) open relationship or encounter “we do not find the attitude to make something yield to our will”^{iv}. This does not mean of course that we cannot *impose* a mechanical/technological perspective over phenomena, and in this case on human relationships, and that this wouldn’t give us *scientifically useful* information. The point is that *if* this is done, then it must exactly be understood as imposing a certain perspective, seeks to determine means of domination, regulation, control: power. So in this respect it is definitely correct to say that scientifically valid knowledge reveals itself *only* through the methods of science. But, this in itself does not say more than that by using scientific methods *such and such* can be attained: i.e. power over phenomena cannot be attained through moral understanding or insight.

I am by no means trying to undermine how much of our (social) lives follow formal codes and how much of society and human behavior functions

mechanically in one sense or another. It is certainly true that what holds for a cocktail party holds also for many other social phenomena and institutions. And it is also true that any given social or interpersonal encounter carries with itself a load of different formal aspects (e.g. what clothes one wears has always a social stamp on it). In fact, one might say that the formal aspect of human life is deeply rooted in language itself^{lvi}. Nevertheless, the crucial point is that any formal features — which clothes one wears, what social situation or institution one finds oneself in — do not dominate or control the human encounter as far as individuals are able to stay in the openness that invites itself^{lvii}. Another way of putting it is that it is not the clothes one wears or the party one attends that by itself is “formal”. Rather the “formal” makes itself known only as a response to the, quite often, unbearable openness: driven by a desire to control, regiment etc. the moral, and I would add, constitutive bond that makes itself know in encounters between people and even between humans and other life-forms: *the formal is a morally dynamic response to the a-formal openness.*

To summarize, my point is (i) that a technological perspective (i.e. strong AI^{lviii}) is, so to speak, *grammatically* bound to what I have now called formal or mechanical aspirations towards life and interpersonal relationships. (ii) What I have called the a-formal openness cannot, so to speak, itself be made formal/mechanical, but can obviously be mechanized in the sense that the openness can be constrained and controlled. (iii) An AI system can, within the bounds of technological knowledge and resources, be created and developed to function in any given social context in ways that resemble (up to perfection) human behavior *as it is fleshed out in formal terms*. But perceiving such social behavior, i.e. formal relationships, as essential and sufficient for what it is to be a person who has a moral relation to other persons and life in general, is to overlook, deny, suppress or repress what bearing others have on us and we on them.

A final example is probably in order, although I am quite aware that much of what I have been saying about the a-formal openness of our relationships to others will remain obscure and ambiguous — also, I must agree, partly because

articulating clearly the meaning of this is still outside the reach of my (moral) capability. In her anthropological studies of the effects of new technologies on our social realities and our self-conceptions, Sherry Turkle gives an striking story that illustrates something essential about what I have been trying to say. During a study-visit to Japan in the early 1990's she came across a surprising phenomena that she, rightly I would claim, connects directly with the growing positive attitude towards the introduction of sociable robots into our societies. Facing the disintegration of the traditional life-styles with large families at the core, Japan's young generation had started facing questions as to what to do with their elderly parents and how to relate to them. This situation led to a perhaps surprising (and disturbing) solution/innovation: instead of visiting their parents (as they might have lived far away and time was scarce) some started sending actors to replace them.

The actors would visit and play their [the children's] parts. Some of the elderly parents had dementia and might not have known the difference. Most fascinating were reports about the parents who knew that they were being visited by actors. They took the actor's visits as a sign of respect, enjoyed the company, and played the game. When I expressed surprise at how satisfying this seemed for all concerned, I was told that in Japan *being elderly is a role, just as being a child is a role*. Parental visits are, in large part, the *acting out of scripts*. The Japanese valued the *predictable* visits and the well-trained courteous actors. But when I heard of it, I thought, "If you are willing to send in an actor, why not send in a robot?"^{lix}

And of course, a robot would, at least in a certain sense, do just as well. In fact, we are not that far from this already as the elderly-care institution is more and more starting to replace humans with machines and elaborating visions of future mechanization (and not only in Japan!) — as is for instance also the parenting institution! It might be said that Turkle's example, as it is in a sense driven to a quite explicit extreme, shows how interpersonal relationships, when dominated by formal codes and roles hides or masks, shuts out, suppresses or even represses the a-formal open encounter between individuals. As Turkle's report illustrates, what an actor, or robot for that matter, can do is to *play the role* of the child — and here "child" as well as "parent" are formal categories. What the actor (as an actor) cannot do is *to be another person* who responds to *you* and gives expression to, say the fear of losing *you*. The actor (as an actor) might surely take on the role of someone responding/relating to *someone*, but that

means that the actor would derive such feelings from, say, his/her own life and express them to you as another co-player/actor in the script that is being played. In other words the actor (as an actor) would not relate to *you* as *themselves*. If the actor, on the other hand, would respond to *you* as *themselves*, he or she would not anymore be (in the role of) an actor, but would have to set this aside. My claim is that a robot (AI system) could not do this, that is, to set aside the part of acting upon formal scripts. What it can do is to be (play the role of) “a child” or a “parent”, to the extent that these categories designate formal roles, but it could not be a being that is composed, so to speak, of the interplay or dynamics between the formal and the a-formal openness. And even though my or your culture might not understand parental relations as formally as the Japanese in Turkle’s report, it is undeniable that parent-child relationships (due to moral conflicts and social pressure — just look at any psychoanalytical analysis!) take on a formal character — so there is no need to think that this is only a “Japanese phenomena”. One could, or rather should say: it is a constant moral challenge and self-investigation to clarify how much of our relationship to others (e.g. to one’s parents or children) is determined or formed by the formal categories of e.g. “parent”, “child” etc. as they are understood in terms of collective normativity, and to what extent one is open to the other as an *I* to a *you*. To put it once more, the idea of strong AI is, as one might put it, the flip side of the idea that one’s relationships to for instance one’s parents was and is only a matter of “a child” relating to “parents”, i.e. relating to each other exclusively *via* collective social identities.

I am of course aware that anyone who will be advocating for strong AI will simply conclude that what I have called the a-formal openness of human relationship to others and to life is something that must be “naturalized”, “disenchanted” and shown to finally be formal/mechanical in its essence. To this I cannot here say anything more. The only thing that I can rely on is that the reader acknowledges the morally charged dimensions I have tried to articulate, which makes the simple point that understanding what it means to place a technological and mechanical perspective on phenomena always concerns a moral question as to

what the demand for mechanization is *a reaction to* and what it strives for. And obviously, my point has been that any AI system will be a formal system and is conceptually/grammatically bound to a technological perspective and aspiration, which indicates, not that this sets some “metaphysical” obstacles for the creation of “strong AI”^{ix}, but rather that there is inherent confusion in such a fantasy in that it fails to acknowledge that it *is* a technological demand that is *placed on phenomena or life*^{xi}.

Concluding remarks

I realize that it might not be fully clear to the reader how or in what sense this has bearing on the question of AI, and especially on “strong AI”. To make it as straight forwards as possible, the central claim I am advocating for is that technological or mechanical artifacts, including AI systems, all stem from a, what I have called a “formal” (encompassing the “informal”) perspective on phenomena. And as this perspective is one that, as one might put it, *contextualizes* phenomena with a demand for control, discipline regimentation, management etc. and hence *transforms* it — it becomes an artifact of our demand. So my claim is that the idea of strong AI is characterized by a *conceptual confusion*. In a certain sense one might understand my claim to be that strong AI is a logical/conceptual impossibility. And in a certain sense this would be a fair characterization, for what I am claiming is that AI is conceptually bound to what I called the “formal”, and thus always in interplay with what I have called the a-formal aspect of life. So the claim is *not* that we for instance lack a cognitive ability or epistemic “perspective” on reality that makes the task of strong AI impossible. The claim is that there *is no thought to be thought which would be such that it satisfied what want, urge for or are tempted to fantasize* — or then we are just thinking of AI systems as always technological simulations of an non-technological nature. In this sense the idea of strong AI is simply nonsense. But in contrast to some philosophers coming from the Wittgenstein-influenced school of philosophy of language, I do not want to claim that the idea

of “strong AI” is nonsense because it is in conflict with some alleged “rules” of language or goes against the established conventions of meaningful language use^{lxii}. Rather, the “nonsense” (which is to my mind also a potentially misleading way of phrasing it) is a form of confusion arising out of a temptation or urge to avoid acknowledging the *moral dynamics* of the “formal” and “a-formal”, of the openness inherent in our relationship to other and to life. It is a conceptual confusion but it is moral by nature, which means that the confusion is *not* simply an intellectual mistake or shortcoming, but must be understood through a framework of moral dynamics.

Endnotes

ⁱ See Turkel (2011)

ⁱⁱ See for instance Kurzweil (2005) and Malone (2012)

ⁱⁱⁱ In this article I use the term “techno-science” to characterize the dominant self-understanding of modern science as such. In other words I am claiming, for reasons which will become clear — although not argued for sufficiently — that modern science is predominantly a techno-science. I am quite sympathetic with Michel Henry’s characterization, that when science isolates itself from life as it is lived out in its sensible and interpersonal nature — as modern science has done — it becomes a techno-science. As Henry puts it, science alone is technology. See Henry (2012). For more on the issue see for instance Ellul (1990), Mumford (2010) von Wright (1986)

^{iv} See <http://www-03.ibm.com/innovation/us/watson/>

^v For a short discussion of the term Good Old Fashion Artificial Intelligence see p. 12-13 below

^{vi} Dennett (1991, 2006) & Haugeland (1986)

^{vii} See for instance Mumford (2010), Proctor (1991), Taylor (2007)

^{viii} In the Aristotelian system natural phenomena had four “causal” forces: substance, formal, moving and final cause Proctor (1991): 41. Of these causes, the moving or “efficient cause” was the only one which remained as part of the modern experimental scientific investigation of natural phenomena Bacon (2012): II, 9, pp. 70.

^{ix} Proctor (1991): 6

^x Bacon (2012): 1. 124 pp. 60, Lång (1986): 96

^{xi} Descartes (1967): *Discourse on Method part VI*, pp. 119

^{xii} Proctor (1991): 22

^{xiii} See for instance Descartes’ *Discourse on Method* and *Passions of the Soul*, Descartes (1967). In Addition to Descartes we might note that Thomas Hobbes, in addition to Descartes technological conception of the human body, gave a technological account of the human soul as well, holding that cognition is essentially a computational process Hobbes (2010): 27-28, See also Haugeland (1986): 22

^{xiv} Dennett (2006): 3. See also Dennett (1991), Haugeland (1986)

^{xv} Habermas (1984 & 1987), Taylor (2007)

^{xvi} c.f. Henry (2012) chapter 3 “Science alone: Technology”

^{xvii} As Bacon put it: truth and utility are the same thing Bacon (2012): I, 124 pp. 60

^{xviii} Proctor (1991): 31-32

^{xix} One of the main ideological components of modern secularized techno-science has been to devise theories and models of explanation that devalorized the world or nature itself. Morals are a human and social “construct”. See Proctor (1991) and Taylor (2007).

-
- ^{xx} von Wright (1989): 53, Robinson (1997)
- ^{xxi} Descartes (1967): *Discourse on the Method part I*, pp. 81
- ^{xxii} Bacon (2012): Preface, pp.7
- ^{xxiii} Proctor (1991): 26-27
- ^{xxiv} Pereira (2006), Mumford (2010)
- ^{xxv} Descartes (1967): *Discourse on the Method part VI*. pp. 119
- ^{xxvi} c.f. Bacon (2012): 1.129, pp. 62-63. Let me just note here that I am certainly not implying that it is only modern science that serves and has served the cause of domination. This is obviously not the case. My main claim is that, in contrast to at least ancient and medieval science, modern science builds both conceptually as well as methodologically on a notion of power. The consequence of this is and has been the creation of unprecedented means of domination (both in form of destruction and oppression as well as in construction and liberation!)
- ^{xxvii} Mumford (2010), von Wright (1986), Taylor (2007), Mendelssohn (1976)
- ^{xxviii} Mendelssohn (1976): 77 & 207
- ^{xxix} Uberoi (2002): 90
- ^{xxx} Alic *et.al.* (1992): 5
- ^{xxxi} *Reverse spin-off or "spin-in"*. Technology developed in the civil and commercial sector flows upstream, so to speak, into military uses. See *ibid*: 64-75
- ^{xxxii} *ibid*: 65-66 & 69
- ^{xxxiii} See <http://www.parkinson.org/Parkinson-s-Disease/Treatment/Surgical-Treatment-Options/Deep-Brain-Stimulation>.
- ^{xxxiv} van Erp *et.al.* (2009): 11-12, emphasis added
- ^{xxxv} *ibid*: 11
- ^{xxxvi} Proctor (1991): 3
- ^{xxxvii} For an interesting read on the effects of the inter-connectedness between scientific research and industrial agro-business in India see Kothari & Shrivastava (2012)
- ^{xxxviii} Taylor (2008), Proctor (1991)
- ^{xxxix} Proctor (1991): 10
- ^{xl} Another example closer to the field of AI research would be Daniel Dennett's claim that the theoretical basis and methodological tools used by him and his fellow champions of cognitive neuroscience and AI research are well justified because of the techno-scientific utility they produce. See Dennett (2007): 87
- ^{xli} Proctor (1991): 13
- ^{xlii} Henry (2012): 54 (emphasis added)
- ^{xliii} Or Top-down AI which is usually referred to as "Good Old Fashion Artificial Intelligence" (GOFAI). See Haugeland (1986)
- ^{xliv} Barsalou (2008)
- ^{xlv} Clark (2001& 2008), Wilson (2002)
- ^{xlvi} Oudeyer *et. al.* (2007)
- ^{xlvii} Guerin (2008): 3
- ^{xlviii} A telling example is of course the word "robot" which comes from the Check "*robota*" meaning "forced labor"
- ^{xliv} AI seen purely as a form of technology, without any philosophical or metaphysical aspirations, falls under at least three different categories: (i) compensatory, (ii) enhancing and (iii) therapeutic. For more on the issue see Toivakainen (2014) (forthcoming) and Lin *et.al.* (2012)
- ^l Mumford (2010): 41, emphasis added
- ^{li} Sherry Turkle gives contemporary examples of this logic that Mumford is highlighting. Based on her fieldwork as an anthropologist she has noted that sociable robots become either possible or even welcomed replacements for humans when the context of human relationships into which the robots are designed enter is mechanized and regimented sufficiently. For example, when a nurse's job has become sufficiently mechanized/formal (due to resource constraints) the idea of a robot replacing the nurse enters the picture. See Turkle (2012): 107
- ^{lii} In the same spirit, The Royal Society also claimed that the scientist must subdue nature and bring *her* under full submission and control von Wright (1986): 65
- ^{liii} For an interesting discussion of the conceptual and historical relationship between mechanization and regimentation, discipline and control of human habits see Mumford (2010)

^{liv} Obviously I am thinking here of a situation in which my friend has not let me know that the dinner will somehow be exceptional, with perhaps an “important” guest joining us.

^{lv} Nykänen (1995): 130

^{lvi} C.f. Wittgenstein (1953): § 111

^{lvii} For more on this issue see Backström (2007)

^{lviii} Let me note here that the so called “weak AI” is not free from conceptual confusion either. Essentially a product of modern techno-science it must also deal with the conceptual issue of how to relate questions of moral self-understanding with the idea of “knowledge as power” and “neutral objectivity”

^{lix} Turkle (2011):74, emphasis added

^{lx} My point is for instance not to make any claims about the existence or non-existence of “*qualia*” in humans or AI systems for that matter. As far as I can see, the whole discussion about *qualia* is founded on confusion about the relationship between the so called “inner” and “outer”. Obviously I will not be able to give my claim any bearing, but the point is just to encourage the reader to try and see how the question of strong AI does not need any discussion about *qualia*.

^{lxi} I just want to make a quick note here as to the development within AI research that envisions a merging of humans and technology. In other words cyborgs. See Kurzweil (2005) and www.kevinwarrick.com. If strong AI is to make any sense, then this is what it might mean, namely that humans transform themselves to become “artificial” as far as possible (and we do not know the limits here!). Two central points to this: (i) A cyborg will, just as genetic manipulation, always have to presuppose the givenness of life. (ii) Cyborgs are an excellent example of human social and bodily life becoming (ideally fully?) technological. The reason why the case of cyborgs is so interesting is that, as far as I can see, it really captures what strong AI is all about: to not only imagine ourselves, but also to transform ourselves into technological beings.

^{lxii} C.f. Hacker (1990), Kenny (2006)

Bibliography

Alic, John A. *et al.* (1992) *Beyon Spinoff*, Harvard Business School Press,

Backström, Joel (2007) *The fear of openness*. Åbo University Press, Åbo

Bacon, Francis. (2012) *Novum Organum*, Bottom of the Hill Publishing, Memphis

Barsalou, Lawrence. L. (2008) *Grounded Cognition*. In *Annu. Rev. Psychol.* 2008. 59:617–45

Clark, Andy (2001) *Reasons, Robots and the Extended Mind (Rationality for the New Millenium)*, *In Mind and Language*, 16:2: 2001 p.121-145

——— (2008) *Supersizing The Mind*, Oxford University Press, New York

Dennett, Daniel. (1991) *Consciousness Explained*, Little, Brown and Company, Boston,

——— (2006) *Sweet Dreams*, MIT Press, Cambridge, Massachusetts,

Descartes, Rene (1967) *The Philosophical Works of Descartes*. (Fourth edition) transl. & ed. by Elizabeth S. Haldane and G. R. T. Ross. Cambridge University Press, New York

Ellul, Jacques (1990) *The Technological Bluff*, Transl. Bromiley, W. Geoffery, William B. Eerdmans Publishing Company, Grand Rapids, Michigan.

van Erp, Jan B.F. *et al.* (2009) *Brain Performance Enhancement for Military Operations*, TNO: Netherlands Organisation for Applied Scientific Research <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA567925>, [retrieved September 10th 2013]

-
- Habermas, Jürgen. (1984) *The theory of communicative action. Vol. 1, Reason and the rationalization of society*, Heineman, London.
- (1987) *The theory of communicative action. Vol. 2, Lifeworld and system: a critique of functionalist reason*, Beacon Press, Boston
- Hacker, P.M.S. (1990) *Wittgenstein: Meaning and Mind, Volume 3 of an Analytical Commentary on the Philosophical Investigations*. Blackwell, Oxford
- Haugeland, John. (1986) *Artificial Intelligence: The Very Idea*, The MIT Press, Massachusetts
- Henry, Michel (2012) *Barbarism*, translated by Davidson, Scott. Continuum, Chennai, India
- Hobbes, Thomas (2010) *Leviathan*, ed. Ian Shapiro, Yale University Press, New Haven
- Kenny, Anthony (2006) *Wittgenstein* (revised edition), Blackwell Publishing, Oxford, UK
- Kothari, Ashish & Shrivastava, Aseem (2012) *Churning the Earth*. Viking, Published by the Penguin Group, New Delhi, India
- Kurzweil, Ray. (2005) *The Singularity is Near: when humans transcend biology*. Viking, New York
- Lin, Patrick et.al. (2012) *Robot Ethics*. The MIT Press, Cambridge, Massachusetts
- Lång, Fredrik. (1986) *Det Industrialiserade*, Helsingin Yliopistopaino, Helsinki
- Malone, Matthew. (2012) *How Artificial Intelligence will shape our lives*.
<http://www.smartplanet.com/blog/pure-genius/how-artificial-intelligence-will-shape-our-lives/8376>, [retrieved October 2013].
- Mendelssohn, Kurt (1976) *Science and Western Domination*, London Thames & Hudson, London
- Mumford, Lewis (2010) *Technics & Civilization*, fourth edition with a new foreword by Langdon Winner, University of Chicago Press, Chicago
- Nykänen, Hannes. (1995) *Music and the Frailness of Wonder*, In: Ralf Eriksson and Markus Jäntti (eds.), *Economic Value and Ways of Life*, , Avebury, England
- Oudeyer, Pierre-Yves et. al (2007) *Intrinsic motivation systems for Autonomous Mental Development*. In IEEE Transactions on Evolutionary Computation, 11(2), pp. 265—286
- Pereira, Winin (2006) *From Western Science to Liberation Technology* (fourth edition), Earth Books, Kolkata, India
- Proctor, Robert. (1991) *Value Free Science?*, Harvard University Press, Cambridge, Massachusetts
- Robinson, Guy (1997) *Philosophy and Mystification*. Routledge, London
- Taylor, Charles. (2007) *A Secular Age*, The Belknap Press of Harvard University Press, Cambridge
- Toivakainen, Niklas (2014), *Man and his Invention*, In Njohja, No. 3 (forthcoming), Pristina.
- Turkle, Sherry (2011), *Alone Together*, Basic Books, New York
- Wilson, Margaret (2002) *Six Views of Embodied Cognition*, In *Psychonomic Bulletin & Review* 2002, 9 (4), 625-636

Wittgenstein, Ludwig (1953) *Philosophical Investigations*. 3rd. Edition. Translated by G.E. M. Anscombe, Prentice Hall, New Jersey.

von Wright, G. H. (1986) *Vetenskapen och Förnuftet, Vetenskapen och Förnuftet*, Mån-pocket, Stockholm

Uberoi, J.P.S. (2002) *The European Modernity*, Oxford University Press, New Delhi,

van der Zant, Tijn. *et.al.* (2013) *Generative Artificial Intelligence*, In: Müller, Vincent (ed.), *Philosophy and Theory of Artificial Intelligence*, Springer-Verlag Berlin Heidelberg