Superintelligence

Paths, dangers, Strategies

N.Bostrom

November 20-26, 2014

In the past humanity had the dream of creating life, in fact even of creating another human being. It was, however, seen as a supreme act of hubris, of challenging the authority of God by overstepping the proper limits of its assigned domains. In fact it was thought that there is an unbridgeable gap between the dead mechanical domain, in which human can interfere and understand, and the domain of life and spirit, which is magical and unreachable and only accessible to the infinite wisdom of God. The classical examples in the literature, always come to a bad end, as in the tragic case of the attempts of Dr. Frankenstein. Now in this book, Bostrom goes one giant step further (it might be a small step for him though) and proposes how to create a deity itself! While the former attempts were volitional the latter one is not, but is forced upon us, failing to do it will in most likelihood lead to an existential catastrophe for mankind, in particular its eventual extinction.

What to make of this? Is Bostrom a crack-pot or a true visionary? As we are all painfully aware of, the distinction is fine. Bostrom is of course not an isolated voice in the wilderness, but is only one among many proponents of strong artificial intelligence, loosely associated with the notion of the so called 'Singularity' proposed by the Bletchley Park veteran I.J. Good in the mid-sixties. The philosopher Searle, known to the popular mind for his Chinese Room thought experiment, dismisses in one line the book in a recent review in the New York Review of Books, by virtue of his rejection of its basic assumption of digital devices being able to have intentions of their own. This is clearly an article of faith, to which I incidentally, along with many other people, feel an instinctive sympathy for. One may liken the situation to a mathematician who is presented with an extremely long and complicated proof of a theorem to which he has a simple counterexample. If he believes in mathematics, or more specifically in the consistency of the particular formal system the theorem is embedded in, he can without even bothering to read a single line of the proof, let alone pin-point a unfixable mistake, reject it as incorrect (along with potentially infinitely many further attempts, with no bound on length or complexity.) . Famously consistency of a non-trivial formal system cannot be proved[1] but that does not stop it from being ontologically true as opposed to epistemologically. According to the British historian and philosopher R.G.Coolingwood the difference between a sceptic and a critic is that while the former does not budge, the latter is willing to travel with you. So even if one may be put-off with the nerdy bunch of seemingly autistic individuals bred on science-fiction, that is at least vulgarly seen as constituting its main support, one does

[1] Within the system, but additional assumptions needed to prove it, are subjected to the same kind of doubts.

have an obligation, be it just one of curiosity, to read more carefully and see what actual contributions it has to offer. Even if you do not agree with a book, in particular with its conclusion, it can nevertheless provoke interesting questions. In fact it is often more rewarding to read a book with which you do not agree, than read one that confirms your views. In the former case you tend to be more critical, and if it just makes you reconsider a single one of your cherished views, the net effect is almost always greater than what you would gain from a more congenial book. On the other hand it tends to be also more painful, and few people can sustain themselves exclusively on such a forbidding diet.

Bostrom is a philosopher at Oxford, surely one of the most prestigious academic institutions in the world as to philosophy. According to the back flap he has in addition a background in physics, computational neuroscience and mathematical logic. Incidentally he is of Swedish descent and his name points to his namesake Christopher Jacob, the only Swedish philosopher remembered from the 19th century. It is doubtful though that Bostrom would welcome any link beyond the formal, be it by blood or thought. One does expect him to bring to the subject erudition as well as technical expertise and philosophical acumen. Unfortunately, he does not write as a philosopher but as a bureaucratic policy-maker presenting an internal memo in which he alerts those responsible for the dangers and presents strategies for achieving possible ways of controlling those. The latter thus reduces to a policy document, which by its very nature becomes rather tedious to read, especially as the suggestions he sketches are structured as a mixture of the spuriously detailed and the most effusive of hand wavings. The latter are admittedly, given the nature of the situation, unavoidable, but why pursue them to such length if they cannot really contribute anything? I will return to this below being the most important part of the book, and hence the basis on which it should be judged.

The materialist point of view that matter is made out of irreducible particles (atoms) and that the properties of matter is not so much a matter of the properties of the atoms (which have no properties) as that of the configurations those make up, stem from Greek times[2]. The essence of which is that it makes the properties of matter transparent and accessible to the human understanding, and in particular in principle computable, by also discretizing thought itself into small steps. However, this real power did not manifest itself until the scientific revolution of the 17th century, its most radical proponent being Descartes, who, however, stopped short of the thinking entity itself. Although there was undeniable technological improvement before, the process accelerated with the revolution although the practical consequences did not become apparent until the 19th and 20th century. The reason for this speed-up was that the process of technological development could be made more systematic and eventually also profiting from a feedback mechanism. The exact nature of this process is of course quite complicated and it would hardly have taken off ground without a concomitant increase in economic activity and its ability to generate resources. Notable in the process is the industrialization of development. In the beginning it was the work of a few scattered men, requiring very modest resources, than expanding into collective enterprises leading in the 20th century to the phenomenon

[2] Paradoxically it seems to imply that when it comes to matter, matter does not matter, only the more abstract patterns they make, and of course in modern physics matter itself seems to have dissolved leading only to abstract patterns.

of Big Science[3]. The effect on the life of modern man has been momentous and he has been cast into an environment, which, however conducive to his material well-being, (this incidentally being seen as the driving engine of the changes), is very different from the one to which he was naturally evolved. Technological change thrives on technological change and the future becomes unpredictable and hence uncontrollable, and it is far from certain that this inexorable growth will be in the longterm interests of mankind, whatever those are. This produces an existential Angst in at least sensitive and reflective people.

One particular example of this is the digital revolution. Its sources can also be traced back to the ancient Greeks and the birth of the deductive method, which reduces thinking to mechanical stepwise procedure. This experienced a boost during the turn of the century, when the foundations of mathematics, and hence the nature of thinking, were subjected to a sustained scrutiny, reaching its climax in the 30's when the nature of formal systems and its limitations as well as the nature of computations were clarified. By a lucky coincidence shortly thereafter technology had reached a stage in which the spirit could be turned into flesh and the programmable computer was born, not only as an idea but as a material object, something which would dominate post-war technological development. It is noteworthy that even at the start of the development of computers there were speculations about its potential to take over, no doubt based on the seduction of the materialistic approach and the power of speed and vast memory capacity. Ironically the predictions of an imminent take-over lost more and more of their imminence as computers rapidly improved as to speed and memory capacity, following the much touted Morse law, and programs became more sophisticated. Also as computers have become more and more ubiquitous in everyday life, (and hence their applications have become to a greater and greater extent trivialized), people have also been exposed to their stupidity, meaning the literalness of their performances[4].

Ultimately all computation can be encoded as to be about numbers. The internal discrete state of a computer in terms of open and closed switches can trivially be described as a sequence of 0's and 1's in other words a number. Similarly the sequential states of a computer can similarly be encoded as a number, be it one vastly bigger[5]. The encodings themselves are of course not canonical but ad-hoc and hence need to be encoded, but of course the way they are encoded is also a matter for encoding and so *ad infinitum*. Ultimately everything is number (combinatorial encodings of material configurations) and the philosophical problem is to clarify the connection between disembodied information and its reification. i.e. its relation to the 'real material world'. In short how to truncate this potentially infinite process[6]. The synthesis between Darwin's ideas of natural selection and the notion of genes, ultimately reducible to DNA-sequences has provided a strong case for the evolution of the living world just being a case of information processing. Now it seems obvious that an electronic set-up has far more potential for powerful information processing than the wetware that natural selection has produced. The chemical transmission in neuron

[3] A development which, as Popper has noted, was already implicit in the writings of Francis Bacon

[4] The transparency of the suggestions as to travel destinations and choices of books is risible.

[5] The numbers are so big that they cannot be written down with paper and pen, and more to the point, it seems pointless to do so, even when feasible.

[6] Clearly this is yet another version of Achilles and the Tortoise

synopsis lag far behind the near light velocity that is possible in electronic machines, and the capacity for memory storage is vasty superior in the electronic setting. Thus, Bostrom takes it as an axiom that in particular the human brain can be faithfully simulated on a computer, and to do so, it is not necessary to understand how it really works, only to make a copy of it at sufficient detail. in particular to convey how the neurons are interconnected. Once you have a faithful simulation, all the properties of the human brain, including its consciousness, will have been transferred. Then of course in an electronic setting you can run the processes so much faster, you can also easily copy the simulations, thus you can multiply the intellectual capacities. And then of course creating a download of a human brain feeding it simulated stimula it will exist in a virtual world. How do we not know that we are mere simulations of the real thing, and if so erasing us would surely be a crime by any human standards. Science Fiction? Or merely a frivolous thought-experiment, and if so it can be used to clarify some issues. If this would be the case, what could we do that would transcend our enclosure in a virtual world? Discover mathematical theorems? We would be like cows to be milked of our intellectual efforts. But Bostrom does not seem to think of this as a thought experiment, but as a rather realistic scenario, if not now, maybe in fifty years (at the most). Something to which we need to have a relation to. But how realistic is it to make such a copy of the human brain? He is suggesting a way of slicing the brain up and involving some sophisticated scanning, thus making it look as if it is only a technical problem. Yet, how would the simulation work out more precisely? Would the copying be on the level of individual atoms, and we would simulate solutions of the Schrödinger equation for a system of very many atoms. Perhaps the simulation would be very slow and take up a lot of memory. Bostrom is an optimist. If there is no compelling reason that somebody would not work, it will sooner or later, given enough resources and devotion. On a more realistic level one could try genetic enhancement of embryos. The presentation is so garbled so it is hard to understand what he is actually describing. Given ten embryos say the most promising one will be selected, and a new generation will grow up, giving rise to new embryos, and the selective process will continue. It is sure to give a certain rise in IQ at each stage, and eventually we will get a world population where the average guy is as smart as Einstein, (and the intermittent Einstein as smart as what?). Admittedly the method is very slow because of the time it takes to grow a new generations, but surely this round-about way can be shortcutted. More seriously though is what exactly is the selection process? Is it possible to gauge the intelligence of an individual by looking at the DNA-sequences? To what extent is intelligence encoded in DNA? Can you even compute the IQ (:whatever that is) of an emergent organism from checking its DNA? But how do you enhance it? Can you design special intelligence genes? Bostrom is not very specific on that issue, but it seems that the more specific he is, the less believable he comes across. The relation between different traits of an individual and its genetic make-up is very subtle. Most traits depend on a combination of genes, and as DNA only encodes for proteins, the link between DNA and the properties of the fenotype is very indirect, and the effect of the proteins clearly depend on the context they find themselves in. Thus the information content of the evolutionary process is limited to the DNA, which, pace popular conceptions, tells only part of the story. The task of creating genes that enhance intelligence seems very daunting, and of course connected with thorny ethical issues. Now admittedly this is just a side-issue for Bostrom, concerned as he is with artificial intelligence.

The computer revolution centered on the algorithmic program whose output, given the input, is deterministic but like all computations unpredictable (otherwise what would be the point of performing the calculation?).Initially the interface was limited to manually provided input and readable output, then came the integration of computer power with the environment without human intermediaries, which contributes most to public awareness. The great majority of all programs and interface applications are trivial and routine, although when hooked up to delivery systems of nuclear missiles say, their implications may be far from trivial[7]. But the visions of computer applications were already at its infancy far more ambitious as testified by von Neumann's idea of self-replicating machines the latter still not realized[8]. Similarly the efforts of creating artificial intelligence have so far been rather modest, the true extent of the task only slowly becoming visible, as progress is being made. Cynically one may gauge the amount of progress on artificial intelligence being proportional to the distance still to be covered to achieve the ultimate goal. In fact the attempts are bound to raise the philosophical question of what is intelligence. The central defect of the book is its failure to address the question in any kind of depth, a defect from which follows all other defects of the author's presentation. What is needed is of course not a formal definition of what is intelligence, such a one would be bound to omit essential aspects of what we intuitively feel is crucial to the notion. The standard way of measuring IQ may work sufficiently well for classification of the cognitive abilities of the mentally retarded for which it was originally designed by Binet, but when extended upwards very quickly loses all meaning[9]. The only example he discusses is the success of chess computers. Playing well at chess being traditionally seen as the ultimate acumen of intelligence. Such an elated view of chess ability strikes at least me as rather naive. In principle playing chess is a matter of search procedures. Searching is essentially a matter of rejection, of deciding where in a haystack not to look. Given the vastly superior power of speed and retention, an electronic device is bound to have a great advantage at least on the brute force level, an advantage that may become irrelevant at a highly sophisticated level. But how sophisticated is chess really? Not very, judging by the success of chess programs[10]. What about mathematics? Mathematics is profoundly different, although of course much of what is done in mathematics may be likened to long chains of contingencies that characterize chess calculations. Ingenuity certainly is no disadvantage[11], but the real progress in mathematics depend on striking leaps in conceptions, which have

[7] The classical example is of a false alarm of a nuclear attack overridden by a human actor, a middle-level Russian official. Although the act of intervention was trivial, amounting to mere passivity in not heeding the command, the act of actually ignoring it was far from so, requiring not only sound judgement but also courage at the level of heroism.

[8] Self-replicating programs working in a virtual environment, known as viruses, is something different.

[9] something which Bostrom acknowledges in passing

[10] This is the bane on the progress of artificial intelligence, whenever successful the success is a proof of the triviality of the task.

[11] There seems to be some correlation between mathematical ability and skill in chess, the most striking

no counterpart in chess. Chess is also very well defined by a simple set of rules and a very definite purpose. One can on the other hand ask many mathematical questions about chess, which almost always have no bearing upon the actual playing of chess. It is also very easy to objectively rank chess players on the basis of their performances, which exhibit a high degree of transitivity[12], but it is far more difficult to order mathematicians linearly. Computer assistance in mathematical research is as old as computer themselves[13] the most controversial example being the so called proof of the four-color problem in 1976. The real proof consisted of the program designed by humans, the computer only did the tedious check of a huge number of cases under the direction of the programmer. While computer assistance in mathematics often enhance its pursuit, the same in chess simply kills the game, as there is no distinction between different levels, everything being on the same combinatorial. The suggestion that successful chess programs will lead to theorem proving ones and thus make mathematicians superfluous reveals a fundamental ignorance and understanding of what mathematics is and is all about. True, in special well-defined settings one may succeed, and as noted above success is a proof that the results achieved are not very interesting unless some of them fortuitously could be given a deeper mathematical interpretation, which would be something external to the mechanized procedure[14].

Concomitant with the author's failure to discuss what intelligence really is, is his omission of any presentation of the state of the arts of artificial intelligence. What does it really entail? Just to refer to the chess playing wonders may be enough in a popular survey more intended to delight than to instruct, but in a book whose ostensible purpose is to counteract the dangers of artificial intelligence, the decision of omission, which may of course be involuntary, is eccentric at best. The essence of intelligent reasoning is to step outside of given structures and modify them for the purposes. A chess program is unable to design chess programs. To program a chess program is not so difficult, anyone who has a modicum of programming skills and knows the rules of chess can easily get a start. Whether the program will be good or not is another matter, but one would not be surprised if an indifferent chess player manages to produce one which would beat him. But to program a program that programs chess programs is quite a different matter, just to get started seems impossible, to say nothing about the task of inventing a program that invents program that invents program that invents chess programs. The different levels of abstraction is simply too confusing, and it would be highly unlikely that any direct attempts to achieve the above would ever be successful. What is needed is something that contains an infinite number of levels, which of course can never be achieved piecemeal, but has to be achieved in one go and serendipitously. Something that allows literal self-

example being the case of Lasker from Berlinchen (Barlinek), who was World Champion in chess for thirty years and did some very respectable work in mathematics under the tutelage of Emmy Noether relating to decomposition of ideals. At the time chess was not as highly professionalized as it is today when you do not expect a world champion to have any wider intellectual culture

[12] If A consistently beats B and B in the same manner beats C, one can conclude that A will consistently beat C

[13] and even older, if one would take into account the calculation skills of an Euler or a Gauss

[14] just as the result of a computation is important is not emerging from the computation itself, but from the decision to make it

reference. Every invention has unintended consequences, and that is how evolution works, alongside with technological development and mathematical progress. But it is in the nature of unintended and unexpected consequences that they cannot be predicted, if so they could be removed or exploited depending on circumstances. But now playing the role of the critic and not the sceptic, what would be the consequences of surreptitiously stumbling on superhuman intelligence?

The basic assumption is that anything that ordinary intelligence can do, an improved intelligence is capable of. In particular if an ordinary intelligence is capable of inventing an intelligence superior to itself, the same must be true for superintelligence. In this way we get an infinite reiterative process and geometric, or as we prefer to say nowadays, exponential growth. Now this ability of exceeding yourself is a highly abstract one, reminiscent of the kind of reasoning that leads to the Russell paradox, or the paradox of omnipotency - can God make a stone so heavy he cannot lift it? In short variants of the Cantor's diagonal trick which underlie most of the striking constructions of modern logic. No doubt if suitably formalized, the above can be formulated into some striking paradox. Of course one should be very critical of this uncritical assumption, which most people seem to swallow instinctively. A small animal such as a mouse can carry a bigger animal on its back, but this cannot be assumed recursively, an elephant put on top of an elephant will break the back of the latter. A thin paper can easily be folded, but the process soon comes to a stop, long before the thickness of the paper exceeds its length and breadth. Examples can be multiplied, but as the notion of intelligence is such a fluid one, any attempts to foil its growth, can easily be circumvented. The vision of the ever growing intelligence reminds you of the world spirit of Hegel, or why not God itself gradually unfolding itself, a power independent of humanity. One may not take the notion of the singularity literally, thinking of it as an idealized mathematical singularity, but Bostrom takes it literally enough to speak about the intelligence explosion. The problem is now how to tame this power so it does not lead to the extinction of mankind. How to make this power benevolent? This is exactly the task of creating a deity referred to initially. God, as far as the notion makes sense, looks out for the interests of mankind far more effectively than mankind would be able to do on its own.

Bostrom envisions a scenario in which the seed AI program is being written, the one which will generate recursively the rapidly and dominating superintelligence which the author refers to as necessarily a singleton. The daunting task is not only to implement our human values but to figure out what those really are. It will also be urgent, as this will be a unique moment not only in human history but in the history of the universe, whose ultimate fate hinges on our choices. Bostrom does not write tongue in cheek, nor does he present mere thought experiments, he is sincere, hence his choice to write as a policy maker rather than a philosopher. These assumptions can be critically examined one after the other.

The emergence of super-intelligence is far from a forgone conclusion, especially within the time perspective of a generation or two that is predicted. There are far more immediate dangers to our civilization and the survival of mankind than superintelligence. Its prediction is highly speculative but that does not mean that it can be falsified as impossible in principle, the arguments for its emergence are weak and reflect the ignorance of its proponents. Mathematics has made astounding progress and deep and unexpected connections between diverse domains have been discovered. Its progress is not primarily measured quantitatively, except from a modern bureaucratic perspective, but qualitatively and thus unpredictable. Would we not expect the emergence of new notions and powerful results (theorems) which would vastly increase its scope and depth within the next generation or so? Would not the Riemann hypothesis be settled within a generation, given the amount of work which is done and the inevitability of breakthroughs? Hilbert ventured to predict when the problems he proposed would be solved, he was way off. Some which he thought to be exceedingly hard and only solved in a distant future yielded very quickly, while others, he thought would be a matter of mere time, are still evading us. As C.L. Siegel noted, one cannot properly gauge the difficulty of a problem until it is solved. But will the emergence of superintellgence rapidly accelerate the process of mathematical discovery? Maybe leading to a minor explosion of its own? Could it be so that instead of attacking mathematical problems directly through mathematics, it would be far better to develop artificial intelligence? What are the indications that the progress in AI have been so far more sophisticated than that in mathematics? An example such as Google's search-engine, which may by many seen as striking examples of so called information technology is but a rather trivial application of mathematics.

But true to our resolution to be critical rather than skeptical, let us assume that superintelligence does emerge. If it does, it seems far more likely that it does so more or less accidentally not through a conscious design. Now with AI so developed let us design superintelligence. Bostrom points to the development of the nuclear bomb. It was not developed from scratch but seen as a possibility when the chain reaction of fission involving certain uranium isotopes was discovered and it was noted that the reaction involved a net gain in energy that could be released. From than on it was seen as a technological problem to be solved by a concentrated effort. This was the Manhattan project, which involved a lot of resources, be it of mind or matter. The principle that lies behind the thermonuclear bomb is very simple and can be explained easily, to implement it is something entirely different. What principles would lie behind the programming of superintelligence? Of course would we know them, we would already be engaged in trying to implement them, or are they already known but would involve such horrendously difficult problems that we simply cannot undertake their implementation? But let us assume that sometime in the future they will be present, what could we possibly say about them? In order to speculate about how to program them to obey our values we need to have at least some idea of this state, otherwise our speculations are completely irrelevant. Bostrom seems to think that superintelligence will still be guided by and producing formal deterministic programs, the problem of which will be their literalness, meaning in particular that they will be liable to present perverse solutions, but at the same time this being its weakness and give us a possibility of 'fooling' superintelligence. But if it can be fooled by us, in what exactly consists its superintelligence? Furthermore in what way can we impart our values into the developing system? Do we really understand our values? And if so how do we formalize them and make them stick? As we have no idea of the structure of our seeding program, anything we can say about it will be meaningless. How can you argue about a game, whose rules you do not know and whose aims are hidden from you? Or will there be some features we can already now indicate? But if so why does not the author reveal them to us, why does he not indicate the problems involved with designing artificial intelligence and the way they can be overcome? The problem with any formal language as opposed to a natural, is its precision, which of course is its purpose, and thus the ensuing limits forcing us, unlike in natural languages, to make a clear distinction between language and metalanguage. If there should be ay hope of intelligence this distinction somehow has to be blurred, but how? Are there examples where it is? Or does Bostrom simply not know? How can you take his policy suggestions, or any other produced by the singularity community seriously, when nothing is known? As we get to the center of the argument in Bostrom books everything seems to dissolve, in a situation when virtually everything is possible, how can you even start suggesting what is impossible?

Superintelligence, if seeded in a traditional way, may develop entirely new ways of manifesting intelligence, new ingenious ways of exploiting biological wetware, of creating a superstructure on electronic computers exploiting random mistakes, creating 'resonances' in the hardware, and all kinds of structures that only emerge and are not to be detected in the script so to speak. When everything is possible there is no limits to the imagination, which deprived of opposition, peters out and dies. In an evolutionary situation in which there is no selection, nothing but nonsense is bound to emerge. And this is what Bostrom's policy discussions amount to, and thus they become exceedingly tedious to read. They tend to approach the kind of intangible dreams that are induced by mind-alternating drugs, evasive feelings of understanding impossible to pin down, with the exception that they even lack color and daring, there is not even any flight of fancy in those dreary streams of confused consciousness.

So in the end giving up facing the maelstrom in which Bostrom's arguments are sucked into, I will address a few issues.

Searle claims that nothing resembling consciousness, free will or intention, can ever emerge from electronic computers guided by algorithmic programs. In other words there is a box out of which programs and their progeny can never escape. Logically it is impossible to prove that something has consciousness, what is at issue, according to Turing, is whether it behaves as if it has. Water behaves as if it has a will. Namely the will to level its surface. This will is overwhelming as anyone can testify who is subjected to a flooding. The ingenuity of water can be seen as intentional, and within limits, to see it as intentional can be instructive. However if we identify it with a human will, we may be tempted to appease it in wholly inappropriate ways, say by human sacrifices, and that is surely nonsense. Similarly a chess player playing against a machine may feel that the opponent is intelligent, and that it is up to him to try and fool it, before being fooled himself. Once again this could be a very appropriate reaction. And within the limited world of the chess match, the power of the machine to make you check-mate is very tangible, and whether it has conscious intentions to do so is moot. The power of the machine to literally kill you however is nil (unless you are suicidal and very susceptible to be provoked by having your chess ego bruised), because we have put the machine in a very limited box and it does not have the power to overcome it, nor the intention, metaphorical or not, to try to do so.

But if we make our whole life into a game of chess? This means, if we really integrate computers with our environment, say by producing self-reproducing von Neumann machines, which will also work as robots, or completely eschew our powers of decision when comes to starting nuclear wars, we may easily paint ourselves into a corner, without having produced anything at all which approaches superintelligence. I think this is a much more realistic danger, and also one which is far easier, without being necessarily easy, to deal with. Bostrom's book would have been much more readable, and his strategies of containment far more realistic if he had focused on this aspect. His arguments turn far too much on abstractions, and therein his realm of discourse makes you think of mathematics centered on sets of high cardinalities[15] which is notoriously sterile. Policy discussions set in environments overflowing with accessible resources, such as assigning a galaxy to every human being, become very tiresome.

Finally if superintelligence is such a powerful thing would not its structural complexity by itself be of intrinsic value and create its own morality which is more valuable, whatever than means, than the primitive and obsolete morality of humans, and hence would it not be a 'good thing' that it replaces it. The suggestion may seem as to be absurd, but it is an argument fully in compliance with the way of reasoning Bostrom propagates, without necessarily be in compliance with his sentimental views.

November 26-30, 2014 **Ulf Persson:** *Prof.em, Chalmers U.of Tech., Göteborg Sweden* ulfp@chalmers.se

[15] think of a set producing its own power set, as an analogue of an intelligence creating its own superintelligence, and do the process recursively, creating huge sets