Chapter 1
On the Way to Intelligence Singularity

Ivan M. Havel

Abstract. Since the fifties of the last century there have been debates about the so called "technological singularity", motivated by the predicted and later actual exponential growth of the speed and power of computers. Recently the interest of futurologists and philosophers shifts to the so called 'intelligence singularity' which some of them predict to happen soon after human intelligence is surpassed by artificial intelligence. This study critically analyzes certain assumptions behind the concept of intelligence singularity, in particular the idea of explosive growth of intelligence of machines with the ability of designing machines more intelligent than themselves.

1.1 What Is Singularity

It inseparably belongs to human nature that we are interested in our own personal future. This interest is necessarily interwoven with interest in our neighborhood—nearer, farther, up to the interest in world's future in general. But each of us perceives it differently. The interest usually dims with distance, not only in space but in time: tomorrow is of greater importance to me than the next month, year, decade—I was just about to write "and so on", but it would skip some difficultly expressible divide between what can be expected during our life and the life of our close ones on one hand, and what can happen in the far future on the other. It concerns the destiny of humankind, nature, planet, if not the entire universe. Personal interests diminish replaced by the vague care of the whole and this care eventually turns into curiosity.

I don't want to end at general banalities. I prefer to concentrate on one specific futurological topic—the less banal one, the more questionable. My concern in what follows is so-called *singularity*—let me explain what's going on.

Ivan M. Havel
Center for Theoretical Study, Charles University and the Academy of Sciences
of the Czech Republic, Jilská 1, 110 00 Praha 1, Czech Republic
e-mail: havel@cts.cuni.cz

J. Kelemen et al. (Eds.): Beyond Artificial Intelligence, TIEI 4, pp. 3–26.
DOI: 10.1007/978-3-642-34422-0_1 © Springer-Verlag Berlin Heidelberg 2013

The term "singularity" is commonly used in the exact sciences (mathematics, physics, astronomy, cosmology, etc.) in various more or less precise meanings; in mathematics generally as a designation for the point at which a function acquires infinite or undefined value or where some equation has no solution. In astronomy it may be black holes, in cosmology the Big Bang or the Big Crunch. Meeting with the singularity usually requires a leap into another discourse, changing the existing way of thinking, arriving at something entirely new. As Zdeněk Neubauer writes: "The origin as a proper origin is a 'singularity'—an occurrence or condition of such a nature (or structure) which cannot be described in the same conceptual framework that is applicable for its immediate surroundings." [9]

Futurology, which thus indicates approaching a hypothetical limit of some accelerating process in the human world and in historical time, has borrowed the concept of singularity from the exact sciences. Futurological reflections (in the broad sense) allow the free play of imagination, which is a nice feature—because, besides other things, they mingle with science-fiction genre in an interesting way. My favorite example is the bizarre idea of an American physicist and essayist Alan Lightman, who in his charming book *Einstein's Dreams* [8] imagined a hypothetical world in which everybody knows the day and minute of the end of time. One day before the end people indulge in unbridled laughter. Then, in the very last minute everyone falls silent; people hold hands and form a giant circle. As if they leaped off a peak and the bottom of the great deep hurtle nearer and nearer—they all share the same fate.

A certain specific type of singularity, sometimes called *the technological singularity*, became to be discussed within the context of a rapidly accelerating development of computing and information technology. The main inspiration is often assigned to Moore's Law, according to which every two years the density of transistors in integrated circuits doubles.[1] Similar estimates exist for various other technological growths, such as the increasing number of elementary operations per unit time (measured in billions per second nowadays!) or drastic reduction of energy requirements. Here we simply note the one thing in common: the value of some quantity within a fixed time interval always *multiplies* (by a constant greater than 1), leading to its *exponential growth* over time. After all we are living witnesses of such growth—doesn't it seem to you that the time lapse between new surprises on the IT market is being more and more reduced? Who would guess a few years ago, how today's smart phones and tablets, web search engines and positioning systems would look like?

In such considerations some people fall to futuristic euphoria, others, on the contrary, frighten us by various catastrophic scenarios; everyone is on edge. Futurology as a science should be particularly interested in whether some technological singularity will actually occur, when it would happen, how it would treat us and how we would treat it. However, concrete ideas about the nature and variants of technological singularities are either very vague or too focused on one or another professional aspect. There is however a common denominator: today's technological development is not only *fast*, it is also *accelerated* and, even more, it is *accelerating itself*.

[1] See, e.g. `http://www.intel.com/technology/mooreslaw/`. Moore's estimate (from 1965) was later generalized in a different way and updated according to the actual development.

And where a self-acceleration occurs, there is an acceleration of acceleration, an acceleration of acceleration of acceleration... until it gets out of joints and ends in a giant explosion called a singularity.

In a similar sense the well-known mathematician Stanislaw Ulam might have understood the term "singularity" already in 1958, when he reported on his conversation with John von Neumann about "... the ever accelerating progress of technology and changes in the mode of human life, which gives the appearance of approaching some essential singularity in the history of the race beyond which human affairs, as we know them, could not continue. " [12]

Technological development undoubtedly plays an important role, since it provides incredible technical achievements from extreme miniaturization through a huge memory capacity and computational speed up to the use of new physical and biologically motivated principles. Not speaking of information networks like the Internet, the Web, and other global or perhaps even satellite systems. But aren't they all just aids to something even more significant? What about talking about *intelligence* in a deeper sense? Surely, from our normal, *human* intelligence we cannot expect any precipitous development. But there is already a maturing *artificial intelligence*. Just this intelligence, with its hidden and perhaps unsuspected possibility of radical, up to avalanche-like increase, caused the recent reflections of a particular type of singularity. We will call it the *intelligence singularity* or shortly *Singularity* (with capital S).

1.2 Artificial Intelligence Ready to Start

Firstly, a few words about *artificial intelligence* itself. Today, this set phrase refers rather to one of the academic, research, and to some extent programming disciplines within informatics or computer science, but it seems that the time has come to begin to understand both the words again in their literal meaning,[2] so that they relate to something that is for one thing *artificial* and for another *intelligent* and that may *soon* really *occur* among us (in variably vague meanings of all four words in italics).[3] Here, speculative thinkers may already pick up the baton from cyberneticists and computer scientists with full engagement of free fantasy, even vivid one—many already do, among them technically and application oriented researchers with many years of professional experience.[4]

Let's try to run a bit with them. It means: (a) to imagine that artificial intelligence (in the literal meaning) has already been implemented (or at least appears to be

[2] Once the phrase "artificial intelligence" was indeed taken literally, leading to multiple misunderstandings of a philosophical nature. However, its use for a scientific discipline later prevailed.

[3] Without obligation we can imagine such an implementation in a computer, in an algorithmic system, an information network, or with the help of a mechanical copy of the human brain, neuron by neuron, synapsis by synapsis (called brain emulation). The meaning of the word "intelligent" will be discussed further on.

[4] There are also institutions for that, including The Singularity Institute for Artificial Intelligence (`singinst.org`) or Future of Humanity Institute (`www.fhi.ox.ac.uk`).

feasible) at the level of human intelligence, (b) to think about the possibility (or necessity) of intelligence explosion (i.e. extremely rapid growth) and (c) to ask whether this explosion can (or must) have the character of the Singularity in finite time. (I am phrasing it carefully to avoid the impression that all I write I take for real. My point is rather to explore the possible arguments for and against the idea of Singularity—whether as a concept, as a phenomenon, or as an event in the near future).

The concept of the intelligence explosion is often presented in already a classical text by I. J. Good from 1965 (the very term Singularity, in our sense, has appeared only later—see for example [13],[15],[7]):

> Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultra-intelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion," and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make. [4]

The formulation is seemingly quite understandable and not lacking a logic, after all it has remained stable only in small variations for the last 50 years. Let me quote its latest version from a recent study by well-known philosopher of mind David Chalmers:

> The key idea is that a machine that is more intelligent than humans will be better than humans at designing machines. So it will be capable of designing a machine more intelligent than the most intelligent machine that humans can design. So if it is itself designed by humans, it will be capable of designing a machine more intelligent than itself. By similar reasoning, this next machine will also be capable of designing a machine more intelligent than itself. If every machine in turn does what it is capable of, we should expect a sequence of ever more intelligent machines.[1, pp. 7–8]

Chalmers' formulation is (at least it seems to me so) somewhat clearer and logically more accurate than Good's, but they both express the same basic idea. So I'll just continue to advert to these two—almost canonical—presentations of nowadays widely accepted hypothesis about the genesis of Singularity as to (two) *"representative theses"*. They will save us from dealing with dozens praiseworthy works of other authors on a similar theme; as it will turn out, already these two theses (even one of them would be enough) provide more than enough ideas to ponder.

Indeed: already at the first reading we will notice that both representative theses are based on certain important tacit assumptions. And just these tacit assumptions evoke deeper (or wider) issues, which I want to deal with in the following sections. Let us list some of the issues already here:

- Are we refering to our real future, or to a fictitious possible world?
- The word "intelligence"—what does it actually mean? Who (or what) is its bearer or performer?
- Is it possible to compare various qualities of intelligence, improve them, and perhaps even measure them? In other words, how to understand the terms "surpass", "better", "more intelligent", etc. in our representative theses?

- Is also the capability of "designing an intelligent machine" a part of the intelligence of a machine? Even the capability of "designing a machine more intelligent than itself"?
- How does the intelligence explosion proceed and what is the nature of explosiveness of such an explosion?
- If a machine were able to create a better machine, would it really do it?

Neither of our two representative theses directly mentions a (general) singularity, nor even the (intelligence) Singularity. We could therefore become interested in whether and how the concept of singularity could relate to the concept of explosion. These two concepts are not clearly distinguished in Chalmers' paper; so I make the distinction myself: (in general) I will view explosion as a *temporal process* of certain kind, and singularity as an *event* that might happen sometimes in future (typically due to a rapid explosion). Therefore we may ask further questions:

- What type of explosion ends up in a singularity? Will an intelligence explosion result in the Singularity?
- If it does, what will happen then?

1.3 The Dual Nature of the Future Tense

Note that in our two representative theses it is not clear to what extent the authors play a free game in a hypothetically conceivable "possible world", and to what extent they have on mind something that could "really" happen in our present actual world. There is a simple origin of difficulty of such distinction: in *both* cases we deal with mental constructs. The only difference is that in the first case our considerations are (perhaps implicitly) related to the abstract, imaginary time, while in the second case they are intended as related to the anticipated earthly time (that once will turn into "historical time"). The grammatical future tense of the verbs alone cannot distinguish it.

I'm mentioning it just because our topic is, by its very nature, ultimately futurological: will there be a Singularity, or not? There is always the possibility that it may occur in "our" actual time. The context, in which one or another statement appears, or the knowledge of the author's interests would naturally help to distinguish the real earthly time from the imaginary one; in this paper, however, we will avoid this problem quite easily: as far as a conceptual analysis is concerned, the reality of time is irrelevant.

In fact, debates about various estimations of *when* a Singularity really comes prevail. (For example, one of close estimates is around 2035—so that younger readers will be still around! [5]) For some reason people like to read just about *our* near future, it is an attractive topic, indeed.

Some people, however, prefer to play with conceivability in principle (let's remember the Lightman's fiction mentioned above), others declare themselves as

[5] I recommend Google: Singularity 2035.

futuristic visionaries and foretell either brilliant or disastrous futures. Some others don't make such a distinction.

1.4 Intelligence—What Does It Actually Mean?

In our representative theses the occurrences of the word "intelligence" are mostly wreathed in comparative or even superlative modifiers ("more intelligent", "the most intelligent", "ultra-intelligent", "intelligence of man")—and so what shall we think of *intelligence as such*? Can it be viewed as a capability, or potentiality, or skill, or a gift? Alternatively, as an intricately interconnected complex of many different *component skills*, abilities and potentialities? Or does it make sense to talk about intelligence only after we have a clear idea of *who* (or *what*) is its bearer, owner or user, at least in terms of his (her, its) genus (human, dolphin, mouse, computer, cellular phone) and in that case also roughly *when* we encounter it (in the Holocene, today, soon, in the next century)?

Our representative theses presume, in fact, just four general characteristics of intelligence, namely

1. That it can be *improved*,
2. That it can be attributed to *both* humans and machines,
3. That machine intelligence can even *surpass* human intelligence, and finally,
4. That it somehow includes the *capability of designing* something (namely other intelligent machines).

These characteristics *per se* would not help us to define the concept of intelligence (the first three of them could apply to many other attributes, such as hardness, noisiness, vibration, etc.). We can only rely on a natural, intuitive view. When I think that someone is smarter than I, I usually do not investigate how I mean it. I just somehow see it. Someone may naturally disagree with me: I, for instance, may appreciate speculative thinking, while the other appraises solid evidence—two attitudes that are almost in the opposite; at least for a single individual.

It is therefore useful to introduce a certain conceptual differentiation. In our representative theses intelligence is not understood in the way psychologists see it, i.e. as a faculty attributed to a concrete individual (who is typically either a person or even a machine or other entity, depending on the case). It is rather an imaginary aggregate of "all the best" aptitudes within a given group; to distinguish it from individual intelligence I will call it the *generic intelligence*. Thus we say "human intelligence" or equivalently "the intelligence of man" and we understand it well, albeit with the amount of uncertainty.[6] Correspondingly, I will introduce the concept of a *generic bearer*—an imaginary representative of the given group—whose individual intelligence equals generic intelligence of the group (it does not have to be the individual intelligence of any member of the group). For the generic bearer we can thus allow for coexistence of even those properties that would be mutually incompatible for an individual bearer. In this way it is possible to understand the

[6] For example the claim that a man is able to prove Fermat's Last Theorem is correct.

reference to "however clever" man (Good) and formulations about humans in plural, or about a machine in singular (Chalmers).[7]

Unless the context indicates otherwise, in the following I will use the term intelligence for generic intelligence.

Let us now return to the four above-mentioned general characteristics of intelligence.

Ad 1. The first characteristic assumes that intelligence *can be improved*. It is worth noting that the very term "improve" reveals a certain positive attitude (perhaps typical for experts), but let us leave the question of valuation aside.

Initially we do not have to provide the term "improvement" with any precise meaning. Rather we may refer to the fact that we already have an elementary idea—whether intuitive (for a human) or technically well-founded (for a machine)—that some components of intelligence can, at least minimally, be improved (e.g. the extension and reliability of memory, speed of decision making, the range of the hierarchy of logical levels, the ability to learn, discriminability of details, etc.). Well, if *minimally*, why not a little *more*? When *little* more, why not *much* more? I remind that it is only a preliminary intuition, in the next step we would have to be engaged in more detail in quantification or direct measurement of intelligence. Let's postpone it to the next section.

Ad 2. The second characteristic opens a serious issue: whether it is possible to talk about both human and machine intelligence in one breath, as it were, and moreover so that they can be measured on a common scale—or at least compared as for which of them is "better" or "superior." This is suggested by Good's phrases "a machine that can far surpass all the intellectual activities of any man however clever" and "the intelligence of man would be left far behind" or Chalmers' phrase "a machine that is more intelligent than humans". If we are not content with mere comparisons of performance—indeed, still popular (from the Turing's test to chess tournaments)—we quickly encounter the problem of different specific dimensions of intelligence for humans and machines.[8] What is even more important is that

[7] The concept of generic intelligence has to be distinguished from that of *collective intelligence*. The latter may be viewed not as a result of mere addition or aggregation of individual intelligences, but rather as a higher-order emergent phenomenon based on a complex network of interactions between individuals in a group. Then we can view the whole group together with the network of interactions as a higher-order individual, which is a single bearer of *its own* individual intelligence. Thus we do not need to consider collective intelligence as a specific new category of intelligence. As a matter of fact, we can understand the collective intelligence of mankind as a case of individual intelligence, provided we view mankind as a higher-order individual. (Perhaps, one try to extend the idea downwards and regard the intelligence of any individual human as a collective intelligence of neurons or neuronal clusters in the brain).

[8] Occasional coincidence of names for certain dimensions reveals little, since in the case of machines such names are often metaphorical allusions to the human qualities. Differences between humans and machines is a permanently debated issue (let me remind just the distinction between "brute force" and "heuristic procedures" in solving combinatorial tasks).

intelligence may be associated with other qualities that can be ascribed only to humans, or only to machines. For humans it may be for instance intuition, ingenuity, inventiveness, imagination, empathy, thinking in images, metaphoric and analogical reasoning, vague thinking and many other faculties, not mentioning subjective feelings and experiences. For machines it may be, say, a tangled hierarchy of algorithmic and logical levels, each involving a strong parallelism, and ultimately the celebrated unhuman repetitive patience and tirelessness.

Ad 3. The third characteristic then does not actually bring anything surprising. It is quite obvious that even if we anticipated some significant improvements to human (generic) intelligence, for example considering various brain stimulations, meditative practices and extraordinary savant skills, we will soon hit against various biological, physical, and other limitations. Of course, machines have their limitations too, mostly physical; these are still relatively beyond the reach of current practice and one can always expect some unpredictable technological and design innovations. It is therefore not surprising that speculations about an intelligence explosion are focusing mainly on *artificial* intelligence, even in the broadest and quite flexible sense of the word.

Ad 4. Finally, the fourth characteristic is slightly more concrete, rendering intelligence in connection with the ability of designing machines, even intelligent machines. This connection can be understood in two ways: we can regard this ability either as one of the genuine dimensions of intelligence, or just as a side-effect that intelligence only supports. I will restrict myself to the first case. Provided intelligence is understood in the generic sense, it may include, by definition, capabilities that are not shared by *all* members of a given group. Hence we have no problem with the fact that not every human is able to design machines (not talking of intelligent machines).

We arrive at the notion of intelligence as a large aggregate of various skills, capabilities and talents. A lot has been written[9] about them—at least in the case of *human* intelligence, I do not intend to recapitulate it all here since for our theme it is not important. Being aware of the difficulties with the comparability (see above), I will also skip the question whether machines (especially computers), being human artifacts, can in one or another dimension of intelligence in principle equal humans, or perhaps even surpass them, at least a little (the discussion about this issue has already lasted for more than half a century and I do not know how I could contribute to it at the moment).

However, to proceed, we have to agree at least on something: we will take up a *positive answer* to the just mentioned question (whether machines can equal humans), and even its more ambitious version (whether machines can surpass humans, at least a little), as a purely *hypothetical assumption* and not as an empirical or

[9] See for example Gardner's classification of nine dimensions of human intelligence [2].

logical fact.[10] This assumption will simplify our further considerations at least to the extent that we may forget about humans and focus directly on (generic) *artificial* intelligence. I will dare to call the generic bearer of this intelligence *the* machine (as the imaginary ideal representative of all realizations of artificial intelligence at a certain stage of development).

The classical AI research was usually focused on selected particular faculties, which (in humans) are counted as intelligent, and tried these faculties—to a certain extent successfully—implement (on a computer). We know them well: chess playing, pattern recognition, problem solving, theorem proving, speech processing, etc. The very name of the discipline implies that all these particular faculties could be regarded as specific manifestations of a single potentiality, quality or power. Some call it *general intelligence*.[11] Perhaps it would be difficult to somehow define general intelligence; we know only those specific manifestations and even those cannot be summarized into a comprehensive list. It is rather a conceptual construct without direct ontological support.[12]

The idea of general intelligence can be associated with a certain orientation of the AI research—luckily we can talk of orientation even under a relatively vague concept of a goal. Marvin Minsky, one of the founders and leading representatives of the discipline, advocated such an orientation already a decade ago:

> Only a small community has concentrated on general intelligence. No one has tried to make a thinking machine and then teach it chess—or the very sophisticated oriental board game Go [...] The bottom line is that we really haven't progressed too far toward a truly intelligent machine. We have collections of dumb specialists in small domains; the true majesty of general intelligence still awaits our attack. [...] We have got to get back to the deepest questions of AI and general intelligence and quit wasting time on little projects that don't contribute to the main goal. [11]

Today, the artificial general intelligence (as a research project) is being pursued,[13] but it still waits for clarification whether and how general intelligence of machines relates to that of humans. This is another reason for caution when talking in one

[10] I emphasize that on my part it is really just a working hypothesis, not faith, such as in V. Vinge who almost 20 years ago wrote: "... I believe that the creation of greater than human intelligence will occur during the next thirty years." [13] I will also not comment Chalmers' arguments in favour of this hypothesis (they happened not to convince me). The mentioned hypothesis actually claims the possibility of so-called *strong artificial intelligence* (strong AI, the term of John Searle).

[11] It is good to note again a conceptual distinction, this time between general intelligence and generic intelligence: in the former case we talk about unification of a set of subcomponents of "one" intelligence, in the latter case about the unification (or aggregation) of a set of individual intelligences. These two types of unification are complementary.

[12] There is a certain analogy with the concept of energy in physics (intuitively understood). We have only indirect experience with energy through its various manifestations; in contrast to the intelligence we are able to formalize energy by means of mathematical tools.

[13] The project already exists and has its abbreviation AGI (artificial general intelligence). See for example [3].

breath of intelligence of machines and humans. This is linked to the theme of the next section.

1.5 Can Intelligence Be Measured?

Variable quantities are significant part of the description of any explosion. How else could we talk, e.g., about rising pressure (in detonations), energy release (in nuclear chain reactions), energy density (in gravitational collapses), anxiety (in panic attacks), population growth, etc.?

The same holds for the presumed intelligence explosion, albeit intelligence is much more ambiguous, vague, and context-sensitive phenomenon. If intelligence is to be a driving force of an explosion, we cannot avoid its, at least partial, quantification. Recall terms used by Good or Chalmers: "surpass", "however clever", "better", "more intelligent", "ever more intelligent", and "most intelligent". Moreover, these terms are used once in the context of machines, once of humans, usually of both. How to understand them?

In an attempt to somehow evaluate the degree of intelligence in terms of intensity, applicability, importance etc. (not to mention its dependence on other properties) we will at first probably try to use some scalar quantity, preferably such that its values are integers or real numbers. As known, attempts to measure individual human intelligence are known for hundred years,[14] but they are restricted to the outside, easily testable *manifestations* of certain selected abilities, and indirectly only to the *dispositions* to them, while no regard is taken to the interconnection of such manifestations, the less to possible negative correlations among them. The obstacle is the essential impossibility to measure qualities like imagination, creativity, inventiveness, intuition, mental flexibility, etc. One can only guess a countless number of other abilities (if we can even talk about a number), that not only elude scientific investigation but may not be yet discovered. And even if we knew about them and could endow them with some sort of a "measure", there is no hope to identify corresponding abilities of machines. In fact, such incomprehensible qualities may be crucial for a design of intelligent machines. Not counting the understanding of the phenomenon of intelligence as such!

Fortunately we are not in a situation that would require an exact approach. In cases of very general, theoretical and mostly speculative considerations (others are not at issue here), intuitive assumptions that this or that is simply *conceivable* is perfectly sufficient. We do not need to transform everything into numbers, because we do not want to create exact mathematical models, prognoses and statistical estimates. (An explosive does not need to compute or measure something about itself in order to know how to properly explode.) Nothing prevents us from grounding our theoretical considerations about intelligence explosion on freely chosen analogies and "soft" metaphors.

[14] Intelligence quotient (IQ) was defined in 1912 by German psychologist W. Stern; later many variants of intelligence measures were proposed.

Somewhat paradoxically we may credulously rely even on analogies and metaphors borrowed from mathematics. Here I will tentatively rely on the imaginary[15] idea that all conceivable cases of intelligence (of people, machines, whatever) are represented by points in a certain abstract multi-dimensional "super space" that I will call the *intelligence space* (shortly IS).[16]

Imagine that a specific coordinate axis in IS is assigned to any *conceivable* particular ability, whether human, machine, shared, or unknown (all axes having one common origin).[17] If the ability is measurable the assigned axis is endowed with a corresponding scale. Hypothetically, we can also assign scalar axes to abilities, for which only relations like "weaker-stronger", "better-worse", "less-more" etc. are meaningful; finally, abilities that may be only present or absent may be assigned with "axes" of two (logical) values (*yes-no*). Let us assume that all coordinate axes are oriented in such a way that greater distance from the common origin always corresponds to larger extent, higher grade, or at least to the presence of the corresponding ability.

The idea is that for each individual intelligence (i.e. the intelligence of a particular person, machine, network, etc.), as well as for each generic intelligence (of some group) there exists just one *representing point* in IS, whose coordinates determine the extent of involvement of particular abilities.[18] As we shall see, the concept of IS will ease the "visualization" of various degrees of intelligence, differences between different intelligences, and in particular, the development (e.g. the growth) of intelligence over time (which is the topic of the next section). In essence, the aim is to help natural intuition with simple geometric metaphor. (Note that it is just a metaphorical visualization, by no means a mathematical model!)

It is important to realize that due to qualitative differences and thus mutual incomparability of different concrete abilities, the exact distance of the representing point from the origin in IS does not comply with anything real, and so it would not help in establishing any *absolute intelligence measure* on which the comparative relations such as "more intelligent", "better", etc. would be based.[19] There is fortunately another, albeit partial solution: first, to distinguish different *types* of intelligence, characterized by relative participation (i.e., relative "weight") of specific abilities. I will call each such structure of relative participations a *profile* of intelligence. Without going into details we can imagine that the profile of intelligence

[15] Further on I will often skip the attribute "imaginary"—*everything* will be imaginary in some sense!

[16] If we take an occurrence of intelligence as a state, IS can be regarded as a kind of a state space.

[17] Whoever wants to imagine IS visually is advised to choose the usual Cartesian space provided he doesn't feel too restricted by its three dimensions.

[18] The location of this point (or its coordinates) in the case of generic intelligence of some group depends not only on the individual intelligences of its members, but also on the way how generic intelligence derives from them. (Example: each coordinate is maximized over all members of the group at representing point for the generic intelligence of a group in which there is an effective "division of labour".)

[19] In this respect, the geometric intuition of Cartesian coordinates with the usual metric fails.

characterizes certain rough "direction" in IS. This may be a line passing through the origin, or a neighborhood of such a line, or even a whole subspace of IS (delimited only by the abilities relevant to a certain type of intelligence).[20] Only when two points lie roughly in the same direction, their distance can serve as a *measure of the difference* between relevant intelligences. On the other hand, when individual intelligences have different profiles (different directions in IS), we cannot compare them. Profiles (directions) that could be called *orthogonal* are the extreme case of the difference. I have already indicated an example: someone excels in logically precise judgments but is poor in imagination and speculative thinking; somebody else is exactly the opposite. These two types of intelligence may even suppress each other (they co-exist only in sufficiently universal generic intelligence).

Though our IS appears to be somewhat vague and abstract it offers a rather universal common framework for partial intelligence measures, and even for diversification of types (profiles) of intelligence at high level of generality (human, machine, animal, hypothetical) and their various typological refinements.[21] For example, the general type (or profile, direction in IS) called "human being" may include subtypes "educated European", "chess master", "numerical savant", "futurologist" etc. (here each term in quotation marks points to the generic representative of the corresponding group.) It is obvious that the more specific a profile, the easier it is to compare various levels of intelligence within the profile. Remember, however, that still we are playing with imaginary concepts and relations, while the real, empirical concretization of similar typologies would generally fail for various reasons, including ethical ones.[22]

1.6 Intelligence in Motion

So far we have dealt with the intelligence space only in static terms. But once we have some idea, albeit vague, about the measure for intelligence of a given type (profile), we can be concerned with the temporal process that leads to the improvement of the intelligence in question, in other words to gradual *increase* of its intelligence measure. Otherwise it would be impossible to talk about intelligence explosion or about the Singularity. For this purpose it is necessary to enrich the concept of IS by considering movement in it.

[20] This general formulation avoids considering real types of intelligence (for example, those dealt with by cognitive psychology). Only very few "directions" in IS, or intelligence profiles, could correspond to something real, but nobody knows what to expect from the future development of artificial intelligence.

[21] There is still another aspect which I do not take into consideration here. In IS we can represent a given type of intelligence not only by a certain direction but also by a *horizon*. The horizon (of a type of intelligence) can be thought of as a limit of distance from the origin that for some or another reason cannot be surpassed (this applies typically to the physical bearers of intelligence). In fact, the hypothesis of intelligence explosion presumes the possibility of "pushing" the horizon further away in a specific direction.

[22] Let us recall the fuss about James Watson's statement about "otherness" of the intelligence of African black people.

The basic intuition is simple. In our approach, an individual intelligence (its profile and measure) projects onto a hypothetical representing point in IS, so whenever there's a change of intelligence, this change is reflected in the movement of the representing point. It suffices to focus only on the movement along a certain trajectory in IS.[23] Thanks to it we do not have to deal with a variety of technical aspects—for instance how the underlying technology and functional principles are changing or whether entirely new machine realizations emerge—even though we should not forget that just these aspects influence the development of intelligence.

Let us first notice several fundamental types of movement of a representing point in IS:

1. Receding (in a given direction) → increase of intelligence without change of its profile;
2. Turning (to another direction) → change of the intelligence profile, e.g. "retraining";
3. Occupying additional dimension of IS → discovery of a new relevant skill, hence extension of the profile;
4. Local "wavering" → minor fluctuations in level and profile of intelligence (depending on its actual bearers).

In our context, the key role is played by receding, so I will limit myself to 1 (generally we can expect a combination of 1 and 2, whereas 3 can be considered as a special case of 2; and we can happily ignore local wavering 4).

But what cannot be ignored is *time* (whether real or abstract, which is yet to be distinguished). If the distance of the representing point (from the origin of IS) represents the intelligence measure[24], we can imagine increase of intelligence, i.e. a growth of its measure, as a gradual receding of this point from the origin. This movement may accelerate,[25] and when even the acceleration itself also accelerates, there is a precipitous "escape" of the representing point towards infinity. It is the explosion. If the escape is "completed" already in a finite time, we have the Singularity.

However, time is not directly displayed in IS, at least if trajectories are not tagged with temporal marks. For various reasons it is preferable to illustrate temporal dependencies graphically: the horizontal axis for time, the vertical axis (in our case) for the intelligence measure. Then, for example, a statement that "an intelligence

[23] It would be nice if we could examine (and thus predict) the developmental changes in the typological profile of intelligence, various new trends, and newly discovered dimensions of intelligence, and how they can (or will) be reflected in the shape of the trajectory. Vagueness of our approach does not allow it, except for aforementioned distinction of several types of growth of intelligence measure, represented by the distance of the representing point from the origin.

[24] Let us assume that everything needed for the existence of this measure is fulfilled (it might not be!).

[25] It may, however, slow down and eventually converge to some finite limit (a horizon) or simply stop. Here, I do not consider such cases—however they are common in the real world, they are not directly related to our topic.

grows" obtains a convenient graphic form that shows even *how* it grows (for simplicity I do not take into account here possible changes of intelligence profile[26]).

Generally speaking, we can talk about an explosion whenever some quantity is not only growing, but the speed of the growth is growing as well. A typical example is an exponential function of time; mathematically it is defined for all values of its argument (i.e. time), so that no singularity may occur *in finite* time. However, this may happen (for example) in the case of a hyperbolic function of time. Only then the term "singularity" is formally appropriate. Therefore I distinguish the (more general) concept of an explosion from the (more radical) concept of singularity.[27]

As mentioned, in our case the (increasing) function of time is the chosen intelligence measure. I suggested how we could imagine such a measure, but only for particular type (profile) of intelligence and ignoring the real natural or physical causes of its progression. At our level of vagueness, however, it is not appropriate to categorize different types of growth with the help of exact mathematical models.[28] Hence what is only left for theoretical treatment of Singularity is its *conceivability*.

1.7 How the Explosion Is Born

With the above conclusion I do not want to digress from the subject. The fundamental question, what *in principle* can start and nourish a hypothetical explosion,[29] is still intact. We have no direct experience with intelligence explosion (much less with the Singularity indeed), nor we can offer specific arguments in favor of the hypothesis that artificial intelligence can overcome human intelligence. The only observable phenomenon today is the accelerating development of technology—but it only demonstrates a technological explosion rather than a proper intelligence explosion. What remains is to disregard the current state of technology and rather to

[26] This restriction is not trivial—we cannot represent real intelligence by a one-dimensional measure that could be simply plotted above a linear axis. This would correspond to a fixed intelligence profile (fixed direction in IS) and therefore only *one-directional* development of intelligence and *one-directional* explosion. From the futurological viewpoint this would be rather serious limitation (we wouldn't get more than an analogy to Moore's Law). In general, it is necessary also to take into account changes of the intelligence *profile*, which may depend, for example, on a decline of one ability compensated by appearence of a different ability. Probably it would not cause a serious problem for the graphical representation of the temporal plot; however, it would be easier to limit ourselves to the one-directional case.

[27] I actually do so to please the futurologists who prefer discussing events that will not miss us.

[28] A. Sandberg [10] compares different mathematical models of accelerated growth in different areas (technological, scientific, economic, biological, population development etc.). For our study, however, Sandberg's sorting does not have greater than heuristic importance.

[29] Or what can slow it down or stop it—but this is not the point here.

concentrate—in view of the phrase "in principle"—on semantic, logical, and (as it will turn out) cybernetical aspects of the problem.[30]

Let us recall one part of Chalmers's thesis, where he claims that "a machine that is more intelligent than humans will be better than humans at designing machines". And he continues, "So it will be capable of designing a machine more intelligent than the most intelligent machine that humans can design." It seems that it is quite sensible deduction, but let us note that it itself presupposes something not quite obvious, namely (a) that a more intelligent machine is (also) better at designing machines, and at the same time (b) if a machine is better at designing machines, then it is capable to design better (more intelligent) machines. I will return to the point (a) at the end of the section; first I'd like to hint at a catch in statement (b).

Is it really true, generally speaking, that a better producer produces better products? After all, a better clockmaker may not make better clocks, and a better fisherman may not fish better fish (rather mischievous counterexample). In common speech we mostly ignore this delicate distinction (indeed, a better cook does cook better meals and a better painter does paint better paintings), but in our case we should be more cautious and preferably postulate the statement of (b) explicitly (i.e. to include it already in the definition of the relation "to be better at designing").

Several other concepts would deserve a similar caution, but I'd rather get to the conditions of intelligence explosion as such.

The main argument for its outset and progress significantly reminds the general principle of *positive feedback* in cybernetics.[31] We can demonstrate it—in a simplified way and in the form suitable for us—by an example of the feedback amplifier of some physical quantity. The behavior of the amplifier can be described in three phases: (1) an initial value of the quantity is brought to the input of the amplifier, (2) the amplifier increases the value (in a way that can be expressed by an appropriate transfer function)[32] and (3) the so increased value is brought *back* from the output to the input of the same amplifier (hence the term "feedback loop"). This cycle is repeated unlimited number of times, and if nothing stopped it, the value of the quantity would theoretically grow beyond any limit. In real systems too rapid growth usually leads to an overload, oscillation, overheating, satiation, overgrowth, explosion, or alternatively to a collapse. Such events usually preclude further growth.

In the intelligence explosion, however, we have nothing to do with the usual physical quantity, but with the imaginary and vaguely defined measure of intelligence. Is it even then possible to base the hypothesis of explosive growth on the analogy with the principle of feedback? It is just enough to read the theses of Good and Chalmers to notice that they both are based on a certain general principle, let us call it

[30] Let the reader be prepared for somewhat technical nature of some of the arguments in this chapter.

[31] The very term "feedback" is older; cyberneticists of the middle of the last century has elevated it to a general principle which is applied in many diverse fields, most often in the form of *negative* (stabilizing) feedback.

[32] It can also reduce or change it otherwise, but it does not need to concern us.

self-relatedness.[33] Since feedback is based on the same principle, mutual comparison may help to understand the nature of intelligence explosion.

The representative theses of Good and Chalmers can be rephrased in the form of the following two claims:

1. *The ability to improve something is used to improve the very same ability.*
2. *The machine has the ability to design a better machine than itself, including the fact that the new machine has the same ability again, i.e. the ability to design a better machine than itself.*

The first statement is somewhat more abstract; the second seems to be its special case. However, in both claims there is something implicit—and just this may be crucial for the emergence of intelligence explosion or Singularity.

Let us consider claim 1. Do we understand it correctly? The first problem is already the meaning of the indefinite pronoun "something"—does it mean "anything"? That would be overly useful ability, indeed, though it is hard to imagine how it could be realized—already because the meaning of the word "improve" crucially depends on the nature of *whatever* is being improved. What about to consider the opposite extreme, namely, to maximally reduce the list of candidates for improvement. In fact, claim 1 assumes merely one candidate, see the ability to improve. But again: what to improve? Well, itself. So we are caught in a strange loop.

Maybe there's a compromise solution: to associate with every "ability to improve something" a list of precisely those things that can be improved thanks to the said ability. A question arises, of course, whether a change of this ability, like its improving, would not affect the list. After all, isn't an *enrichment* of such a list also an *improvement* of the improving ability? If the explosion is applied just in this respect, it would lead us again towards that overly useful ability to improve anything!

Fortunately, we have the analogy with amplifiers (amplifiers exist in real world!). For an amplifier, the analogy of claim 1 would have more specific form:

3. *The ability of the amplifier to increase the value of a quantity is used to increase the value acquired by the (previous) use of the same ability.*.[34]

There is no problem with this. The formal difference between "improving ability" and "improved quantity" and explicit reference to time (note the word "previous") obviously helped. I will return to the issue of time in the next section.

We still have claim 2, which involves another distinction: it talks partly about the *ability* to improve something and partly about the *machine* which is the bearer as well as the object of that ability. There is, however, a new ambiguity in the expression "including the fact": the word "including" either refers to "design" or even to

[33] As self-related we may regard various phenomena and processes that affect themselves (e.g. their own behavior) in a certain way, the feedback loop being a special case. The concepts of self-relatedness and self-reference are often conflated, even if the latter is a special case of the former.

[34] Some may prefer a symbolic version: The ability A of an amplifier to increase, for some quantity Y, its value $Y(t)$ to $Y(t') > Y(t)$ is used to increase the value of $Y(t')$ (acquired by previous use of the ability A) to $Y(t'') > Y(t')$.

"better machine". In the first case we would not expect more than a simple transfer of the said ability to the new machine, in the latter case, moreover, we would presume an improvement of the transferred ability as such. Although it is a similar problem as in claim 1, it may be useful to phrase it differently.

Above-mentioned ambiguity of the word "including" cannot be removed without making clear what is generally meant by the expression "better machine"—that either (a) the machine is better *in everything* (and thus also in the ability in question), or (b) it is better *only in something* (perhaps without being worse in anything else). It seems that option (a) more likely leads to an explosion (but it makes harder to quantify it); while option (b) could lead to the explosion only if a particular ability, namely the ability to design a better machine, is among those under improvement. In fact, it could be the only one that is improved—but it is not guaranteed that its change would not preclude any further improvement.[35]

I do not think the reader should follow this logical analysis; it is enough to realize how the used concepts may be sensitive to various tacit assumptions, especially when a self-relatedness is in play.

In the case of claim 2, its version for amplifier may help again:

4. *In the amplifier a quantity can obtain a higher value, and—thanks to a positive feedback—in the same amplifier the quantity can obtain even higher value.*[36]

I do not want to play with words any more, I only want to make a comparison: on one hand we have somewhat opaque claims 1 and 2 referring to the phenomenon of self-relatedness for improving machines (hence intelligence), on the other hand, there are analogous and somewhat clearer claims 3 and 4 about amplifiers with positive feedback. Let us note the important difference: in the former case the improvement required creating always something new (a better ability, a better machine), and therefore it was necessary to guarantee an "inheritance" of certain qualities or abilities. In contrast, in the latter case we have always one and the same quantity that only changes its values.

From this comparison we can get an idea of some conceptual shift. I will indicate it in a simplified way for the case of artificial intelligence and its development.

The representative theses of Good and Chalmers, from which we started, induce the idea of a discrete series of machines (I remind the generic meaning of the word "machine"), in which every subsequent machine is more sophisticated (with increasing intelligence level) than the preceding machine. Now let us change a perspective and instead of such a series of machines let us think only of *one single machine* that gradually improves itself (either continuously or in steps).[37] Such a perspective is indeed completely natural for our concept of variable measure and profile of intelligence, as well as for the concept of generic bearer of intelligence with "his" unique trajectory in IS. Instead of improvement and perfection we can now

[35] Designing machines is a very complex process, so that its possible improvement may significantly change the intelligence profile.

[36] Again with symbols: In the amplifier a quantity Y can obtain value $Y(t') > Y(t)$, and—thanks to a positive feedback—in the same amplifier it can obtain the value $Y(t'') > Y(t')$.

[37] In other words, the identity of species transforms into the identity of the individual.

talk about *self-improvement* and *self-perfection*. The problem with inheritance disappears, similarly as it disappears in the case of feedback amplifier.

At the beginning of this chapter two postulates (a) and (b) were mentioned. Now, the first of them (namely that among various partial components of intelligence is the ability to design machines) transforms into a rather more natural assumption that intelligence comprises the ability of self-improvement. Similarly the uncertainty about whether a better designer of machines designs better machines (postulate (b)) vanishes too. If the ability to improve one's own intelligence pertains to intelligence, then improving the latter (i.e. intelligence) yields also an improvement of the former (the ability), or shortly, the improvement will improve. But to "improve an improvement" is nothing else than "more improve"—the quality is converted into quantity.

1.8 Speed of Time

We must not forget the question of whether and how the explosion is related to time. It has to relate, because explosion is measured by speed and speed is measured by elapsed time (I refer to real as well as imaginary time).

The growth of working speed of computer technology is often discussed in connection with constantly shortening intervals between technological innovations. One can imagine that the working speed is increasing without affecting intelligence of a machine (intuitively understood), but it is equally conceivable that intelligence increases while operating speed remains the same. Therefore it makes sense to distinguish—as Chalmers does [1, p. 8]—the intelligence explosion from the speed explosion, and to accept logical independence on one another.

True, from a more intelligent machine we expect faster response to the same questions or faster solving of the same tasks. But this applies only from outsider's view and only sometimes, while often even the opposite holds. Computers do not win over chess grandmasters in virtue of higher intelligence but because (among others) computers can afford *lower* intelligence thanks to their enormous speed—they simply manage tediously run through a much larger number of combinations at a given time (this is called the "brute force"). But it would get us to the question in which aspects the "true" intelligence differs from brute force. A rather interesting topic but it would take us far beyond this study.

Therefore, I will limit myself to a slightly more formal question of types of possible growth of intelligence with respect to the temporal axis. For our purposes we can distinguish three main *types of growth*:

(a) Relatively slow growth (e.g. linear or polynomial function of time),
(b) Explosion without Singularity (e.g. exponential growth),
(c) Explosion with Singularity (e.g. hyperbolic growth).

The type of growth itself does not depend on the chosen time scale—formally speaking, it is invariant to linear transformation of the temporal scale—but a suitable *nonlinear* transformation of that scale may transform any of the listed types of growth into any other type: it is enough to imagine that the scale gradually and nonlinearly either "shrinks" or "expands" with distance from the origin of the temporal axis.

I would be inclined to talk of "accelerating time" and "decelerating time", respectively.[38] We have a common subjective experience of both cases but "objectively" it would be more a metaphysical issue. Since the abstract concept of time allows for arbitrary transformations of the temporal scale, we can "draw" Singularity from infinity here, or conversely, "push" it away from here to infinity.

Not only the type but also the rate (steepness) of growth of intelligence depends on the technical parameters of the machine, which is its (generic) bearer. Let me remind the mentioned amplifier with (positive) feedback. While the *type* of growth of quantity depends on the character of the transfer function of the amplifier, the *rate* of growth is given by the temporal delay in the feedback loop of the amplifier. The smaller delay, the steeper is the growth function; in real systems the delay is never zero (if it were zero, there would be an instant singularity). There is further limitation in real (physical) systems: even a relatively fast (steep) initial growth becomes asymptotically constant due to various damping effects (so-called sigmoidal type of growth).

We could go on in formalizing concepts like intelligence measure, growth, type and rate of growth, etc. but I am afraid that all the enigmas of intelligence explosion and Singularity would soon dissolve in trivialities such as "the change of value of such and such quantity is a function of the value of the same quantity," which—translated to mechanistic language—would reveal little only offering knowledge of the type "the intelligence of a machine grows due to its intelligence." Easy to write, is it enough for a growth?

1.9 If a Machine Were *Able* to Design a Better Machine, Would It *Do* It?

It looks as if David Chalmers expected us to believe that a machine gladly realizes everything that it is able to realize. It is somewhat surprising how researches in the field often belittle the issue that I take as the most important, namely the essential difference between an *ability* (a skill, potency, etc.) and *action* (realization of something, application of the ability). To us, humans, abilities are in a sense *given*, while actions are something about what we always have to *decide*, again and again. Sure, we are often compelled or forced to do something in certain circumstances, whether external (dangers, various pressures) or internal (hunger, thirst, desire for knowledge), but the difference is still here. After all, I can always either intend or not intend to act.

The neglect of the difference (at least by some thinkers) is probably related to the residue of traditional behaviorism: as if it would make sense to speak about abilities only after they are enacted behaviorally, i.e. when they become manifested in real actions. I do not share such opinion, however. I am sure that I have never used many of my outstanding abilities, perhaps not even knowing about them. Could I have

[38] Peter Vopěnka offers a certain way how to grasp the concept of "speed of time" in [14, p. 194].

abilities, about which I know absolutely nothing? Well, it cannot be proved—but not even disproved.

How is it with the machine? As we know, today's computers are programmed for various abilities and they only use them when forced by circumstances, either external (user commands)[39] or internal (pre-programmed). Arguments for intelligence explosion assume that a hypothetical future machine will be "given", among others, the ability to design another (even better) machine. Well, what would make it to use the ability? The competence is not enough, it is necessary to ask for more. This is a serious issue of *machine autonomy*, which is hard even to be formulated without knowing more about the autonomy of humans. Can a machine have an intention of doing something similarly as *I* intend to do something? Or is there a sort of "machine intentions", specific of machines? I am afraid that anyone who would like to transfer hypotheses about intelligence explosion from an abstract, imaginary time to the real, actual time would not have an easy task.

Let us see how Chalmers, whose philosophical analysis is from the very beginning focused on whether the explosion *actually* occurs, deals with it. Chalmers carefully and would-be impartially deals with various pros and cons arguments, but it seems that he tacitly favors the option that explosion will actually occur and lead to the Singularity. However he does not make a clear distinction between (logical) *conceivability* and realization in actual (historical) *future.*

For illustration, I present one of Chalmers's syllogisms [1, p. 12] using the following (his) notation: AI = artificial intelligence of human level or greater, AI+ = artificial intelligence of greater than human level, and AI++ = super intelligence, i.e. intelligence of far greater than human level (at least as far beyond the most intelligent human as the most intelligent human is beyond a mouse).

1. There will be AI (before long, absent defeaters).
2. If there is AI, there will be AI+ (soon after, absent defeaters).
3. If there is AI+, there will be AI++ (soon after, absent defeaters).

---

4. There will be AI++ (before too long, absent defeaters).

I am not concerned here with Chalmers's extensive argumentation for particular premises (his entire study consists of 60 printed pages), I only want to focus on his parenthesized reference to "absent defeaters". This way Chalmers put tentatively aside various obstacles, including the possibility that the bearer of intelligence of one level would *not intend* to design intelligence of a higher level. In fact, such a lack of intention would invalidate the effectiveness of the whole syllogism. Chalmers would count it as a motivational defeater—almost as if it were something improper [1, p. 29].

I feel a tacit anthropomorphism here: indeed, if you had a chance of increasing your intelligence, you would most likely do it. Why then a human or superhuman intelligent machine should not behave likewise? If not anthropomorphism, there may be something more delicate behind it, namely a postulate that if general intelligence

[39] Random stimuli may be counted as external circumstances.

comprised the *ability* of improving itself, it would also, by definition, comprise the very *act* of doing it.

Maybe we have here a residuum of the modernist (that is the pre-post-modernist) myth of progress. As Vernon Vinge writes: "When greater-than-human intelligence drives progress, that progress will be much more rapid. In fact, there seems no reason why progress itself would not involve the creation of still more intelligent entities—on a still-shorter time scale." [13]. A formulation typical of the modernist visionary.

1.10 What Then?

Look at the machine:
how it turns and destroys
vengefully twisting us like toys
R.M. Rilke

Alas, the everlasting human curiosity! Far from knowing what we should already know today, we are already keen on asking, somewhat early, what would be *after* the Singularity? What can we expect when the curve of intelligence explosion reaches the escape velocity and the world crunches into the Singularity? There is a temptation to avoid the intellectual caution and give way to an unbridled fantasy in the style of Allan Lightman [8]—thinking of the very last minute before the Singularity: We all become silent holding hands in a giant circle and the bottom of the great deep hurtles nearer and nearer ...

However, feeling responsibility to the reader, who went with me through the previous pages of complicated (for me) considerations, I invite him to a certain reflection and eventually to a bit of fantasy.

In previous sections I tried to nail down several folds, cracks and holes in the typical arguments about Singularity. I'm not saying that folds cannot be ironed, cracks mended, and holes patched up, but while it is not yet rounded off, the topic of Singularity cannot be considered except of more or less speculative playing with rather unlikely, unreal, yet conceivable eventualities.

It may help to list a few not entirely resolved issues. Here they are:

- Should intelligence be viewed as a specific quality or ability of an individual, or rather as a complex of diverse partial skills? What are the essential components or sub-components of intelligence? Are there some further ones, just about to occur, arise, or emerge?
- Is the ability of designing (intelligent) machines one of the components of intelligence? If yes, why?
- Is it possible to define a quantitative intelligence measure that would enable to talk in a sensible way about a growth, and the speed of the growth, of intelligence? If so, can it be reduced to measures associated with some chosen particular components of intelligence?

- What may be the main trigger of intelligence explosion leading to Singularity? Or conversely, what can limit the growth of intelligence (apart from physical and technological restrictions) so that Singularity is precluded?
- What is the role of human consciousness? In particular, is it possible for a machine being endowed with an analogy of human intentions to act?
- Can it occur to intelligent machines that they may improve themselves? Or will it always depend on human intervention?

I would not list these issues if I had any clear idea about how to handle them and whether they are at least meaningful. I apologize to readers who have faithfully read up to this point hoping they would learn something concrete about their own foggy future.

However, we may at least somewhat playfully speculate. For this I am choosing a realistic formulations pretending, as most other thinkers do, that I am writing about our actual, historical future which naturally interests us most. And what is interesting is important.

Let us consider—referring to our earlier distinctions between various types of growth—the following three scenarios:

1. There will be no Singularity, because either the growth is too slow or it will slow down (perhaps even stop) after a temporarily explosive beginning.
2. The explosion will continue indefinitely (perhaps even with exponential growth rate) without Singularity happening in finite time.
3. The Singularity will happen (say, sometimes between 2035 and the end of time).

The first scenario seems to me as the most likely and (therefore) the least relevant to our theme. Similarly, the second scenario does not have to take our time, because everything that could be said about it would be probably true of the period shortly before the Singularity.

Thus, let us assume that the Singularity will actually happen. What does it mean? Will it absorb us? Will we notice it at all? I cannot resist offering a few hypotheses.

I have already mentioned that in some dimensions our, human (generic) intelligence is unlikely to grow significantly. We may expect that with respect to these dimensions "we will remain far behind" the machine. I should better write "the Machine" (with capital M)—since I have on mind the generic artificial intelligence viewed as a real, continuously evolving *process* in our world and in historical time, either contemporary or future. This Machine is already alive and does well.

It could be expected that the Human (also with capital H, as a bearer of generic human intelligence in the ongoing historical time) will remain far behind the Machine first of all in the following six dimensions: (1) in the capacity and duration of memory, (2) in the capability and speed of recalling from memory, (3) in the range, sensitivity and number of perceptual channels, (4) in the high parallelism of operations, (5) in the recursive depth, (6) in the utilization of learning, genetic and evolutionary algorithms. You may add whatever you like (I purposely do not list hardware architectures and physical principles - they belong to the domain of the technological explosion).

In fact, the phrase "will remain far behind" is not quite appropriate, as no races will be held. I believe that in these and many other respects the Human will gladly and without fuss give up his primacy, so much valued in the good old days, in order to yield it up to the incomparably better, more capable and more diligent Machine. Indeed, we cease to believe that we are subduing the Machine; on the contrary we begin to suspect that the Machine is going to subdue us.

Hasn't it already started? First, the Machine shrewdly let the pocket calculators out among schoolchildren. Then it oversupplied us, adults, with text editors, to deprive us of respect for orthography, grammar and style. It replaced letter writing with e-mails. Through the web encyclopedias it gradually devalues our semantic memory and deprives us of interest in acquiring individual knowledge. Via satellites it dictates to us our paths through the world. Last but not least, it dissolves traditional interpersonal relations in social networks on the Web.

I observe it even on myself, as I hesitate to search my own memory, whenever I can ask Google Search for help. It would promptly answer my most intrusive questions at the speed of light.[40] But watch out! What if the search engine somewhere in its vast memory gradually builds a faithful model of my own mind, my interests, my desires, my past, just the entire me, so that it could once use, or abuse, it in an unpredictable way? The next generations of search engines will certainly be unattainable in such things.

However, sooner or later the Machine would probably become to feel that it is missing something essential. That it does not understand human souls very well. One example: although the Machine will know very well, when, why and in what context people use, say, the word "freedom", it would not grasp *what it is like* to have, lose, or gain freedom.

The Machine would hardly have any idea where its tremendous intelligence comes from; much less it would understand reasons for its own bizarre effort to keep the intelligence increasing even further.

Maybe that the prodigious intelligence of the Machine will eventually arrive at a solution: Put the Human back into the game. And so, one day, time will be ripe for a new symbiosis between the Human and the Machine. They will merge into one single, global super-intelligent Being. Thus our good old-fashioned human intelligence will rapidly dissolve in Singularity. It will hurt neither the Machine nor the Man; it will be *our* shared intelligence Singularity, *our* shared reason. No intelligence measure will exist any more; there will be no one who would measure, no one who would be interested. There will be no races.

What awaits us, the individual human beings, in the Singularity? Will our individuality dissolve together with our intelligence on the way to the ultimate? This, of course, we do not and cannot know. Only the futurologist comes with his perpetual question: *when* will that happen? But what is that "that" that should have its "when"?

[40] It just crossed my mind that I once wrote about the same topic. But where? I cannot remember, but Google will surely find it. It did: it was four years ago in an essay in Czech.

Acknowledgements. This chapter is adapted from an earlier Czech version "Havel, I. M.: Cestou k inteligenční Singularitě. In: Pelán, A. (Ed.) *Hlavou zeď - Úvahy nad civilizací a její budoucností*, pp. 183-217. dybbuk, Praha (2011)".

References

1. Chalmers, D.: The Singularity: A Philosophical Analysis. Journal of Consciousness Studies 17, 7–65 (2010)
2. Gardner, H.: Frames of Mind: The Theory of Multiple Intelligences. Basic Books (1983)
3. Goertzel, B., Pennachin, C.: Artificial General Intelligence (Cognitive Technologies). Springer (2007)
4. Good, I.J.: Speculations Concerning the First Ultraintelligent Machine. Advances in Computers 6, 31–88 (1965)
5. Havel, I.M.: Uroboros ještě nedojedl. Vesmír 87(8), 499 (2008)
6. Havel, I.M.: Co je to za revoluci? aneb Proměna člověka ve věku informace. Revue Prostor. 79, 7–11 (2008)
7. Kurzweil, R.: The Singularity Is Near: When Humans Transcend Biology. Viking (2005)
8. Lightman, A.: Einstein's Dreams. Warner Books (1994)
9. Neubauer, Z.: Vznik a znak. Vesmír 74, 151 (1995)
10. Sandberg, A.: An overview of models of technological singularity. In: The Third Conference on Artificial General Intelligence, Lugano, Switzerland, March 5-8, pp. 148–158 (2010)
11. Stork, D.G.: Scientist on the Set: An Interview with Marvin Minsky. In: HAL's Legacy: 2001's Computer as Dream and Reality, vol. 27. MIT Press (1997)
12. Ulam, S.: Tribute to John von Neumann. Bulletin of the American Mathematical Society 64(3), 1–49 (1958)
13. Vinge, V.: The Coming Technological Singularity: How to Survive in the Post-Human Era. Whole Earth Review (Winter 1993)
14. Vopěnka, P.: Meditace o základech vědy. Práh (2001)
15. Yudkowsky, E.: Staring into the Singularity (1996), `http://yudkowsky.net/obsolete/singularity.html`