Draft of the work published on the volume *Singularity Hypotheses* edited by A.H Eden, J.H Moor, J.H. Søraker and E. Steinhart, Springer, 2013.

The slowdown hypothesis

Alessio Plebe
Department of Cognitive Science
v. Concezione 8, 98121 Messina, Italy
alessio.plebe@unime.it

Pietro Perconti
Department of Cognitive Science
v. Concezione 8, 98121 Messina, Italy
perconti@unime.it

Abstract

The so-called *singularity hypothesis* embraces the most ambitious goal of Artificial Intelligence: the possibility of constructing human-like intelligent systems. The intriguing addition is that once this goal is achieved, it would not be too difficult to surpass human intelligence. While we believe that none of the philosophical objections against strong AI are really compelling, we are skeptical about a singularity scenario associated with the achievement of human-like systems. Several reflections on the recent history of neuroscience and AI, in fact, seem to suggest that the trend is going in the opposite direction.

1 Introduction

The so-called *singularity hypothesis* embraces the most ambitious goal of Artificial Intelligence: the possibility of constructing human-like intelligent systems. The intriguing addition is that once this goal is achieved, it would not be too difficult to surpass human intelligence. A system more clever than humans should also be better at designing new systems as well, leading to a recursive loop towards ultraintelligent systems (Good 1965), with an acceleration reminiscent of mathematical *singularities* (Vinge 1993). According to David Chalmers, the singularity hypothesis is to be taken very seriously. If "there is a singularity, it will be one of the most important events in the history of the planet. An intelligence explosion has enormous potential benefits: a cure for all known diseases, an end to poverty, extraordinary scientific advances, and much more. It also has enormous potential dangers: an end to the human race, an arms race of warring machines, the power to destroy the planet" (Chalmers 2010, p.9).

Back when AI suffered from a significant lack of results with respect to the claims put forth by some of its most fervid enthusiasts, and faced strong philosophical criticism (Searle 1980; Dreyfus and Dreyfus 1986), skepticism about the possibility of it achieving its main goal spread, leading to a loss of interest in the singularity hypothesis as well. Our opinion is that, despite the limited success of AI, progress in the understanding of the human mind, coming especially from modern neuroscience, leaves open the possibility of designing intelligent machines. We also believe that none of the philosophical objections against strong AI are really all that compelling.

This, however, is not our main point. What we will address instead, is the issue of a singularity scenario associated with the achievement of human-like systems. With this respect, our view is skeptical. Reflection on the recent history of neuroscience and AI suggests to us instead, that trends are going in the opposite direction. We will analyze a number of cases, with a common rate pattern of discovery: important achievements in simulating aspects of human behavior become on one hand, examples of progress, and on the other, a point of slowdown, by revealing how

complex the overall functions are of which, they are just a component. There is no knockdown argument for posing that the slowdown effect is intrinsic to the development of intelligent artificial systems, but so far, there is good empirical evidence for it. Furthermore, the same pattern seems to characterize the recent inquiry concerning the core notion of intelligence.

We will present two lines of reasoning in this paper. First, we will provide a simple formalization of the slowdown hypothesis in mathematical terms, showing in an abstract way what the causes of the slowdown are, and their effects on the evolution of AI research. The aim of the formalization is not to propose a mathematical model of some kind of automatism inscribed in the logic of scientific discovery, but simply to show in formal terms how a given field of research (in this case, AI) could end up in a slow down progress (under the circumstances we will discuss in what follows). We will then move inside various domains of AI, observing how the history of their scientific development provides support for our hypothesis. One of these will be the field of artificial vision, where the long history of research and the rich body of evidence obtained make it a significant case in point.

Furthermore, we will discuss how the recent inquiry concerning the core notion of intelligence seems to show a similar pattern, with a series of new and far reaching fields of research that have grown around the initial one, such as the role played by consciousness in the social nature of intelligence. On the whole, we will argue that the slowdown effect is due both to reasons that are internal to the logic of scientific discovery, and to the changes in the expectations held in regard to a much idealized subject of inquiry: "intelligence".

2 Formalization of the slowing down

In this section we will try to give a mathematical formalization of the reasons why the research enterprise for an artificial intelligence is characterized by the slowdown effect. Let us call Δ the normalized distance between the performance of an artificial system and that of one held as point of reference, assuming that $\Delta = 0$ means equally valid performances. In very general terms Δ can be the sum of distances over a set $\mathcal{P}$ of simple elementary processes p, producing some measurable performance b_p, assumed to be 1 when fully intelligent, and 0 when absolutely dull.

$$\Delta = \sum_{p \in \mathcal{P}} 1 - b_p \tag{1}$$

With a leap of faith in progress, we can imagine that the performance b_p will become better and better as long as research efforts accumulate in time, for all possible single processes p, therefore $b_p(t)$ is a function of time t, continuously increasing towards 1.

The core of our reasoning is a very common phenomena found in the research of any process involved in human intelligence. It is the fact that a higher and a more detailed knowledge of a cerebral process often spawns a new field of investigation, that is discovered to be a necessary component of the overall intelligent behavior. Process spawning can arise for different reasons. For example, a deeper investigation might reveal that a process, previously thought to be atomic, is in fact the result of two almost independent subprocesses, each deserving its own research specificity, or, while searching for a known brain process, a new and different one, that was previously unknown, is discovered. Still yet, a known process that had little empirical evidence and that could not be reproduced artificially, might begin to become more clear thanks to some new scientific discovery, leading to a new research direction.

A very crude simplification is to take the same trend of performance in time for all processes and assume that some of them, after a certain research time T, will spawn a new process.

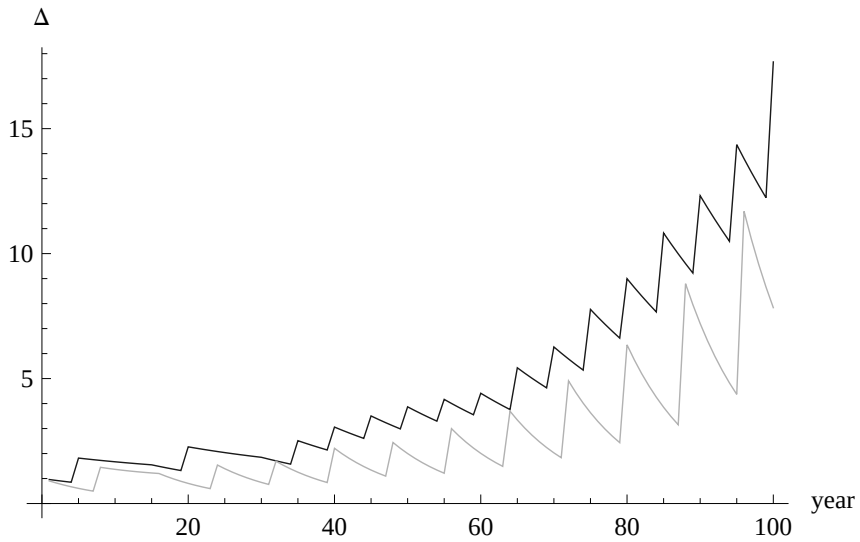

Figure 1: Examples of Δ trends according to equation (2). Parameters for the dark gray plot are $\alpha = 25, T = 5, \phi = .6$, for the light gray curve $\alpha = 10, T = 8, \phi = .7$.

We can rewrite equation (1) in this way:

$$\Delta(t) = \sum_{i=0}^{\left\lceil (\phi 2)^{\left\lfloor \frac{t}{T} \right\rfloor} \right\rceil} e^{\frac{t - \left\lfloor \log_{\phi 2} i \right\rfloor}{\alpha}} \tag{2}$$

where α is the rate of improvement in performance, and T is the amount of research time after which a current process may spawn a new one, and ϕ is the fraction of processes that actually do spawn after T elapsed. Operators $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ are respectively the floor and ceiling functions. In Fig. 1 two examples are shown, with different parameter values. For the dark plot it is assumed that a new process will come about every five years, while for the light plot this event happens every eight years. It is immediately apparent how the spawning phenomena prevents Δ from converging to 0, the level of perfect intelligent behavior, on the contrary it diverges in time towards increasingly worsening values. It may seem paradoxical in Fig. 1 that at the very beginning of research Δ would be smaller than after a century, during which research has expanded in many directions. In fact, it is to be expected, if you take into account that in equation (1) Δ is an estimate of the intelligence level reached by an artificial system with respect to the set of processes $\mathcal{P}$ only. It is not an absolute measure. In principle equation (1) could give an estimate of the absolute general intelligence, using a theoretical $\widetilde{\mathcal{P}}$, the set of all possible processes necessary for a general intelligent system, including many processes for which no research has yet begun. For all those unexplored $\tilde{p}$, it holds $b_{\tilde{p}}(t) = 0$ all the time. Equation (3) can be rewritten as:

$$\Delta(t) = \left|\widetilde{\mathcal{P}}\right| - \left\lceil (\phi 2)^{\left\lfloor \frac{t}{T} \right\rfloor} \right\rceil + \sum_{i=0}^{\left\lceil (\phi 2)^{\left\lfloor \frac{t}{T} \right\rfloor} \right\rceil} e^{\frac{t - \left\lfloor \log_{\phi 2} i \right\rfloor}{\alpha}} \tag{3}$$

In practice, however, there is no way of knowing $\widetilde{\mathcal{P}}$ in advance, due to the fact that precisely knowing all the processes contributing to a general intelligent agent would mean knowing almost everything about intelligence. What happens instead is that every time a new component is discovered to play an important role in intelligence, almost immediately or shortly afterward, a new

research effort for simulating this process artificially, begins. Let us take the example of consciousness: the focus on this problem in philosophy and the awareness of its crucial influence in how the mind works, has triggered a growing amount of research on machine consciousness.

There are clearly many factors neglected in the simple formulation of equation (2) but it reflects real research trends in artificial intelligence. Some will not influence overall trends in a significant way . For example, the birth of new processes would not be synchronous, but each parent process may spawn a new one at a different time. The effect, however, would be just that of having a less regular curve derived by equation (2), but if T is the average time of spawning of all the processes, the long term trends will be similar. Some of the neglected factors would indeed make the forecast of (2) worse, and some better, for the sake of fairness we will discuss the inclusion of a few terms only from among those that warrant more optimism.

A reasonable argument would be that it is unrealistic to believe that all processes p simulating intelligence equally contribute in the summation of equation (2). One may argue that the first studied processes are more important, and as long as research continues, and new fields are spawned, the new fields are components that are gradually less and less crucial to the overall goal of reaching intelligent behavior. Along the same line, one may challenge as unrealistic the expectation that the spawning process will continue forever, and at the same rate as when the AI enterprise began.

We can take into account these two factors, with a formulation that is slightly more complex than that in (2), as follows:

$$\phi(t) = \phi_\infty + (1-\phi_\infty)\, e^{\frac{-t}{\gamma}} \tag{4}$$

$$\Delta(t) = \sum_{i=0}^{\left\lceil (2\phi(t))^{\left\lfloor \frac{t}{T} \right\rfloor} \right\rceil} i^{-\beta} e^{\frac{t-\left\lfloor \log_{2\phi(t)} i \right\rfloor}{\alpha}} \tag{5}$$

where β is the decay in importance of the processes added late on the overall performance of the system, and γ is the decay of the number of processes that during their advanced research stage may spawn a new field of research.

In Fig. 2 two examples of the evolution in time of Δ with the new formulation are shown. As previously for the dark plot, it is assumed that a new process will come about every five years, while for the light plot this event happens every eight years. Contrary to the plots in Fig. 1, in this cases Δ does not move towards worse values, it remains around the value of 1 during the 100 years of the simulation, but again the spawning of new processes hampers the continuous decrease towards the optimal value of 0.

A possible objection to the formalization here presented, might be that in principle Δ can only reach asymptotically its best value 0, even in absence of process multiplication, while the singularity hypothesis postulates that human intelligence cannot only be approximated, but even surpassed. For this reason we defined the value 1 of the measured performance b_p of a process p as the reference best value, without a strict reference to human intelligence, therefore, it can be held to be more than the average human performance. The argument that the fundamental speed up in the singularity hypothesis is based on the ability of the intelligent system to design and implement a newly created system automatically, will be discussed in §6.

In concluding our sketch of a formal justification of the slowdown effect, we would like to mention certain aspects, missing in equations (2) and (5), that will make the development of a fully intelligent system even slower. All processes p are treated as independent, each with its own continuous progress in time, by proceeding in this way we neglected the problem of the interactions between the many processes involved in general intelligent behavior. More realistically, a new process often requires its own research and development, but it also requires the effort to understand and simulate the interfaces between this new process, and at the very least, the one that exists with its parent, not mentioning those between the many other related processes involved.

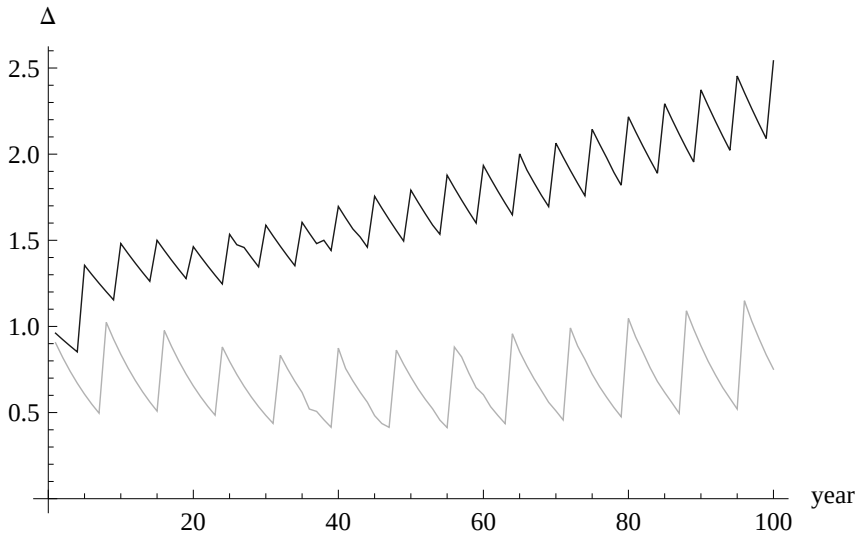

Figure 2: Examples of Δ trends according to equation (5). Parameters for the dark gray plot are $\alpha = 25, T = 5, \beta = .9, \phi_\infty = .7, \gamma = 10$, for the light gray curve $\alpha = 10, T = 8, \beta = .8, \phi_\infty = .7, \gamma = 100$.

Often, understanding the interactions between processes is more demanding than simulating the processes itself. Moreover, it seems that the singularity hypothesis requires an exponential growth of computational and design capacities. We argue, on the contrary, that even the case of an indefinite linear growth is questionable, and that this sounds as an *a fortiori* argument supporting the slowdown hypothesis.

There are also cases in which a new field of investigation may reveal novel solutions for many other ongoing research investigations, a paramount example is the first connectionist approach to modeling neural networks. On the other hand, not all newly initiated research directions turn out to be fruitful. The history of AI, as any other scientific domain, is full of new attempts that initially seemed promising, and later revealed themselves to be wrong or useless. An example is research of the so called $2\frac{1}{2}$*-dimensional sketch* in vision. As a side effect of a successful line of research that affects many others, such as that of connectionism, mentioned above, several older processes may die, substituted by others based on the new paradigm. Summing up, the interactions between processes would certainly make the evolution of Δ more complex than the abstract formulations here suggested, in ways likely to enhance the slowdown effect even more.

3 Scientific idealization: the "zooming in" and "zooming out" effect

The suspicion that the maturation of brain function simulations is characterized by the slowdown effect emerged before attempting to formalize its mechanism, from the observation of a typical pattern in the social history of science, as shown by both the typical pathway of scientific idealization and the recent history of several scientific domains.

Idealization plays an important role in scientific inquiry. In a sense, every scientific enterprise is based on a sort of scientific idealization, that is, "the intentional introduction of distortion into scientific theories" (Weisberg 2007, p.639). Let's take into consideration the well-known case of Galileo's use of the inclined plane to study the force of gravity. According to Aristotle, a continually acting force would be necessary to keep a body moving horizontally at a uniform velocity. Galileo believed that if there was be no air resistance and no friction, and if a perfectly round and smooth ball was rolled along a perfectly flat inclined endless plane, the speed of the body would accelerate in a predictable way. Of course, such a scenario does not exist in the real world. In fact,

it is an idealized state of affairs. In order to arrive at knowledge about gravitational acceleration from the observation of a falling body, Galileo supposed that the inclined plane was an idealized frictionless object. The aim of this theoretical move is to make the problem computationally tractable. Galilean idealization is a computational advantage in elaborating theories with strong predictive power. It is even possible to compute the gravitational acceleration taking into account the influence of the friction, but it is a known fact that Galileo's theoretical move to imagine a frictionless plane allowed significant achievements in the study of classical mechanics. Galilean idealization chooses only certain traits among the plethora of features a given phenomenon is endowed with. It is a deliberative act by the scientist, which is eventually justified by the research program he adopts. In other words, the price the scientist has to pay in order to explain a given phenomenon, is to leave aside or deliberately ignore some of its (not crucial, one has to hope) features.

The gap between the real and the ideal object is a matter of how much the scientist's attitude is flexible in regards to the scientific idealization. In the history of science the process of idealization has changed continuously in its scope. It is like a "zooming in" and "zooming out" effect which depends on how much of the features of the object are neglected or not. The rationale of this "zooming in" and "zooming out" effect consists in allowing the scientific enterprise to respond to social influences in a flexible way. If there is a significant social pressure to include a given trait into the scientific explanation, the community of researchers can modify the "zooming in" of the idealization and incorporate that feature.

Besides this mechanism of regulation of the relationship between science and society, the "zooming in" and "zooming out" effect also depends on the internal dynamics of the logic of scientific discovery. This is exactly what happened to the concept of intelligence from the time of Alan Turing's pioneering studies. The quest for Artificial Intelligence is an attempt to produce a human-like creature from a rather restricted idea of what intelligence actually is. In this perspective, in fact, intelligence is conceived as a computational feature of a disembodied mind. All that really matters according to this perspective is the sensibility to a set of formal characteristics and a good information processing device. The computational nature of this conception of intelligence leads AI scholars to believe that they are dealing with cumulative progress. The singularity hypothesis is based on the assumption that AI findings are cumulative. This expectation, however, depends on the stability of the idealization "zooming". If the amount of features we are interested in grows in a remarkable way, the cumulative effect vanishes. In fact, the growth of knowledge involves increases in a horizontal direction rather than in a vertical cumulative one. As long as scientists devoted their efforts to abilities such as arithmetic computations, the cumulative effect was remarkable. Calculating machines have long excelled their masters, and this has happened without the help of the machines' capacity to design other machines. While these results are increasingly promising, scientific idealization of intelligence has deeply changed. In the last decades, the studies on intelligence have become increasingly more focused on many of the aspects of the phenomenon ignored in the past. Intelligence is now considered as a multifaceted cognitive process, with a proliferation of proposals of new kinds of intelligence, from emotional to musical, from spatial to social.

Howard Gardner (2006) famously argued there are many kinds of intelligence, including numerical intelligence, language mastering, body-kinesthetic, memory, and of spatial perception. On the whole, intelligence now appears to be a more ecological capacity than it did in the past: it is deeply influenced by motor schemata and is constituted by subsymbolic (and perhaps encapsulated) processes. In this way, however, the number of aspects one has to take into account become increasingly more significant, and the pathway of scientific discovery heads more towards a slowdown rather than towards a singularity effect.

4 The case of artificial vision

Let consider now the case of artificial vision, the research aimed at designing artificial systems capable of a visual perception comparable with humans or other animals. This is an interesting benchmark, because its history dates back almost as long as AI, and because it has been and currently is, the most understood cognitive function in the brain.

In Fig. 3 is a sketch of process spawning, the phenomenon at the basis of our equation (5), in the case of vision, showing how the overall domain tends to branch into many autonomous fields of research. There are clearly many different criteria for splitting the domain of artificial vision into singular processes, we have used the principle of only including simulations that target visual behavior clearly identified in visual neuroscience, simulated in a way that adheres to the knowledge of the equivalent brain process. It is far from being exhaustive or objective, even inside the mentioned principle, in that the choice of spawning a new research field and citing a specific work as the beginning of that field, is largely subjective.

The point is not in the details of which processes are included or not, but rather in the fact that as long as artificial vision progresses, more and more new areas are discovered that are important components to be addressed in order to achieve an *intelligent* enough vision.

The scenario of natural language understanding also appears to be well described by our equation (5), with many new independent research fields that have started as this discipline has progressed, with one main difference from vision. Our current understanding of how our brain processes language is far from being as complete as that of the knowledge we have on the processing stages in vision. The precise brain areas involved in language and the characterization of the computational functions of those areas remain obscure. It is therefore, impossible to sketch a tree of stable processes spawning new ones where language is concerned. Lacking relationships with analogous brain processes, it is not possible today to know which of the many ongoing research fields in understanding natural language will continue to progress and to spawn new components, or that instead will lead to a dead end.

The history of efforts in simulating aspects of intelligence offers another recurrent clue related to the slowdown effect, not directly captured by the formalism given in §2. In a number of cases a common pattern of discovery can be detected: an important achievement in simulating aspects of human behavior become on one hand, an example of progress, on the other, gives the illusion of easy progress, while its follow-up reveals how complex the overall functions are of which, it is just a component.

While the discovery of receptive cells in cortical area V1 (Hubel and Wiesel 1959) was a major breakthrough in the understanding of the visual system, that gave confidence in believing that this achievement would be a first step towards artificial vision comparable to that of humans, today it is clear that the computation done by V1 is but a small fraction, and the simplest, of that involved in the whole vision process. Not only, in V1, there is an overlap of processes that are much more complex than just the selectivity to orientation and ocularity.

A puzzle in the early era of neural computation was the simulation of language, requiring syntactic processing. Jerry Elman (1990) made another breakthrough with his recurrent network, that exhibited syntactic and semantic abilities. It was a toy-model, with a vocabulary of just a few words, however, it was then presumed that it would open the road to fast progress in simulating language. In the twenty years that have followed, no other model has achieved results that are comparable to Elman's. Minor improvements were gained at the price of much more complex systems (Miikkulainen 1993).

The biggest success in mathematical modeling of brain functions has been the H-H model of neural polarization (Hodgkin and Huxley 1952). Decades later a powerful simulator became available, based on the core equations of the HH model (Wilson and Bower 1989). Oddly enough, no mathematical model of similar importance for the brain has been developed since. Today, mathematical models are lacking or are oversimplified and limited for the most important phenomena at

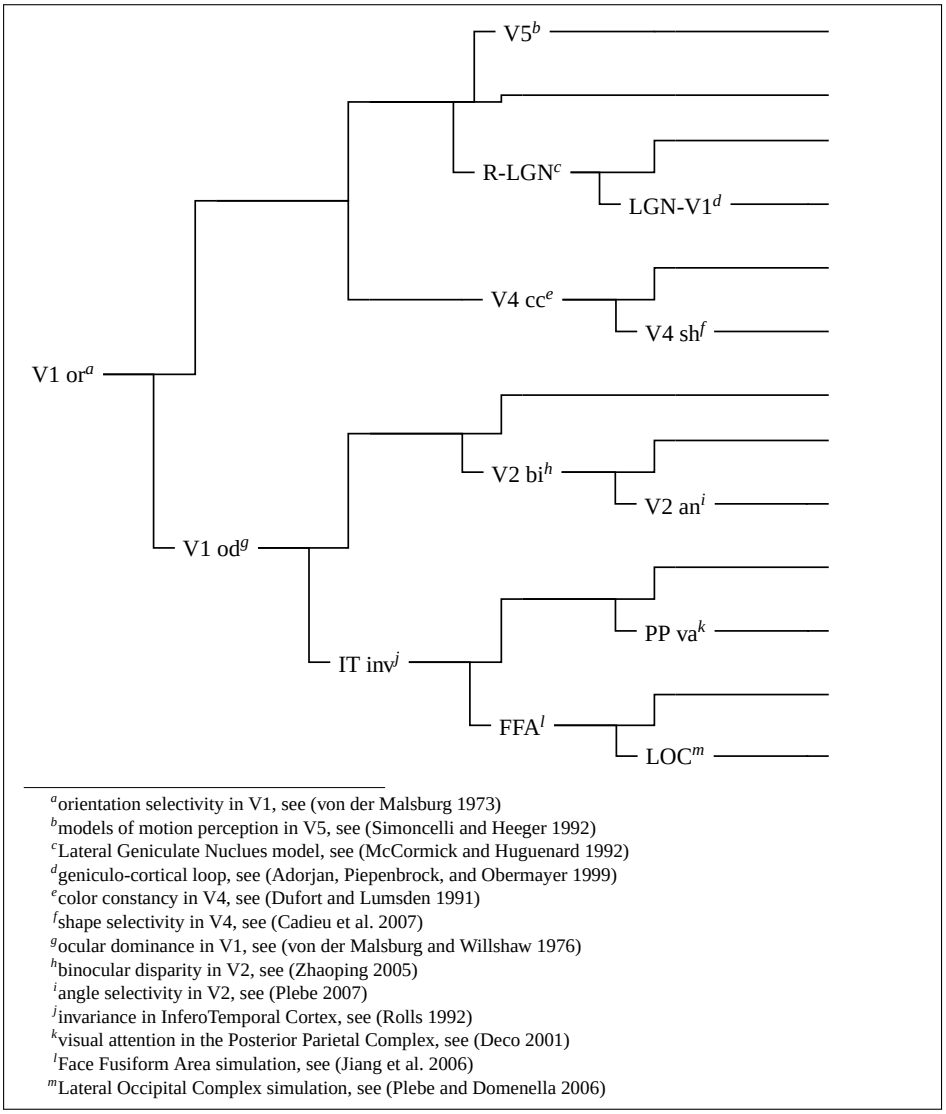

Figure 3: Sketch of research field branching in the case of vision.

a cellular level, such as synaptic transmission (Rudolph and Desrexhe 2005), synaptic facilitation and depression (Dittman, Kreitzer, and Regehr 2000) or dendritic growth and axon arborization (van Pelt et al. 2010), and the reliability of numerical simulations of biological neurons, in itself, has been questioned (Rudolph and Desrexhe 2007).

5 From "intelligence" to the (rest of the) mind

The variety of things we call "intelligence" and that we would like to give an account for within a scientific framework, is much greater now than in the past. We now know much more about intelligence, but this knowledge does not accumulate as before. On the contrary, it is lost in many different pathways, offering a more detailed understanding of "intelligence", but a less cumulative one than in the past. The point, however, is that since the very beginning of the adventure of Artificial Intelligence, the attempt to simulate intelligence has actually involved the whole mind. The privilege granted to the property of intelligence was based on the belief that it is the crucial property of the mind. Understanding intelligence is understanding the mystery of the mind. In reality, the true aim of Artificial Intelligence has never been intelligence, but the mind as a whole. The challenge was to produce a mind like that of humans, and do it artificially. To be like that of human beings, however, intelligence must be endowed with many skills that traditionally were not even associated with it, including the ability to experience emotions appropriate to circumstances and to reason metaphorically. This latter feature, for example, opened the way for attempts to develop computational models of metaphor. The road has proven difficult and at the same time fascinating because it could also shed some light on the metaphorical nature of thought and language. This, in turn, has drawn attention to the role that metaphors and frames have in many aspects of social life, as in the case of political discourse.

Consider the case of moral reasoning. It is a crucial kind of intelligence, if we want to create something that resembles human performance in a significant way. For this reason, researchers, including Wendell Wallach and Colin Allen, have devoted their efforts to the investigation of machine ethics. Is it possible to design machines endowed with ethical principles? That is, that have the capacity to reflect upon different alternatives or to compute a procedure for discovering a way to resolve the ethical dilemmas machines might encounter (Wallach and Allen 2008; Anderson and Anderson 2011)? In this perspective, developing an ethic for machines appears as an interdisciplinary endeavor. This circumstance, however, calls into question many other issues, including the normativity of moral judgments and the sensitivity to the social context of such judgments. The general point is that intelligence has a social nature and when observed in its human form, requires some kind of consciousness. It seems that the kind of intelligence typical of the singularity hypothesis is inspired by methodological solipsism, a widespread tenet in classical cognitive science. According to this way of thinking, the mind is considered as something that pertains to a given individual and consists of a set of skills that can be gradually amplified so as to exceed those of humans. But, if we accept the major claims of current cognitive science, we are also driven to consider the mind as a social and ecological thing. Externalists argue that to specify the content of many mental states one must take into consideration its reference rather than the manner in which, it is given to the mind (Menary 2010). According to theorists of embodied cognition, mental contents are determined by the way the body acts in the environment (Shapiro 2011). Moreover, the success of the account of social cognition, with its idea of the core ability to interpret behavior as a consequence of the mental states of its performer, has finally shown the limits of solipsism (Tomasello 2009). In sum, intelligence is no longer conceived as a mere individual property. In this way, the ghost of normativity and its intractable character appear on the horizon of experimental science.

Overall, if we observe the manner in which the science of mind has concretely evolved in recent decades, it seems that we are faced with a systematic tendency to investigate new areas of

research originating from previous ones. In this way, as soon as results are achieved in a certain field, new questions arise and new areas of investigation open up.

6 Machines designing machines

There is a peculiar aspect of intelligence, which is crucial in the singularity hypothesis, and that is worth underlining. The key idea is that a machine that is more intelligent than humans will be better than humans at designing machines. Designing machines or developing algorithms, is a very special ability only a select few of humans have, and are not included in the common meaning of general intelligence. One may argue that it is possible to escape the slowing down implicit in equation (5), and move towards singularity, simply because it is not interested in general intelligence, but rather in the very specialized aspect of designing algorithms. How many examples of this aspect of intelligence can be found in artificial systems, even at very early stages? Not one. As far as we know, there is no example of any artificial system able to design new algorithms. The computer science domain dubbed "automatic programming" of "generative programming" or "automatic code generation" have nothing to do with designing algorithms. They are just tools that help human programmers write their code, at a higher abstraction level, for example , or by using templates and prototypes. So far, no sign of the aspect of intelligence that is most crucial to the singularity hypothesis, has been seen yet.

It is interesting to analyze under the perspective of aspects of intelligence, some of the most exciting achievements of in AI. Playing chess was one of the first challenges undertaken by AI (Shannon 1950), and was the one that Hubert Dreyfus (1972) bet computers would never even come close to being able to do as well as human beings. The historic victory of IBM's Deep Blue over world chess champion Gary Kasparov, therefore looked like a momentous one. Despite the positive reaction and renewed enthusiasm for the perspective of a strong AI, Deep Blue's victory involves a very marginal aspect of general intelligence. Unlike previous chess computer programs, Deep Blue's architecture was entirely focused on highly efficient database mining (Campbell, Hoane, and hsiung Hsuc 2002). It worked on a database of about 700,000 grandmaster games, and at each move the current position was searched over the entire database for the closest one, at the speed of 200 million positions per second. A similar approach allowed a more recent success to Watson, another IBM supercomputer that won the American quiz show *Jeopardy!*. This system includes a structured and complete version of Wikipedia, that can be searched at a speed of about 500 GB per second. In this case the search is based on a sophisticated analysis carefully specialized for the type of clues used in the *Jeopardy!* challenge (Ferrucci et al. 2010). The 30-clue session is organized into six categories, that range from broad subject headings like "The European Union" to less informative puns like "One buck or less", to specific like "Cambridge". Wide samples of *Jeopardy!* questions were analyzed to classify the so-called LAT (*Lexical Answer Type*), which is a word in the clue that indicates the type of the answer, independent of assigning semantics to that word. For example in the chess category clue "Invented in the 1500s to speed up the game, this maneuver involves two pieces of the same color" the LAT is in the string "this maneuver", and the answer is "castling". Related to LAT is the focus of the clue, that is the part of the question that, if replaced by the answer, makes the question a meaningful stand-alone statement. For example in the clue category "Cambridge": "In 1546 this king founded Trinity College, the largest of Cambridge's colleges" the focus is "this king" and the answer is "Henry VIII". In the Watson system the result of the question analysis is used to interrogate its huge knowledge base adopting a variety of search techniques, generating many candidate answers that are further filtered and ranked. The system is impressively complex and efficient, however it is clearly highly tailored to the *Jeopardy!* quiz interaction, and would be almost unable to maintain a simple, ordinary conversation.

We do not claim that knowledge base mining is not intelligent, on the contrary an aspect of

human intelligence is certainly the ability to retrieve from long term memory what is relevant to the current stream of thought. However, this aspect becomes rather shallow and limited when the whole system is highly specialized and with a single focus, like browsing chess grandmaster games only, or answering clues in the *Jeopardy!* format only.

There is another reason why AI champions, such as the two cases here analyzed, are unlikely to be steps on the way to a general intelligence machine. These systems are not only highly customized to fulfill their goals, they also lack any reference to how the brain works when pursuing the same goals. The extreme search of performance obliged the designers to abandon any attempt to implement processes with biological plausibility. As a result, the solutions cannot be counted as processes to sum with others in the direction of a general intelligence, they cannot be affected by the continuous progress of neuroscience, and cannot spawn new processes for new components of intelligence. IBM itself, seems well aware of this, and its current most promising line of research in AI is in a completely opposite direction to that of Deep Blue and Watson: that of mimicking in detail the functions of biological neurons in computer chips, for building the future *Cognitive computing* (Modha et al. 2011). The $21 million DARPA funded project SyNAPSE (*Systems of Neuromorphic Adaptive Plastic Scalable Electronics*) is an ambitious program to engender a revolutionary neuromorphic chip comprising one million neurons and 10 billion synapses per square centimeter. Several tech commentators have suggested this may be the beginning of singularity[1].

We are convinced that this is one of the main roads to intelligent machines. However, the fastest hardware for simulating neural circuits would be useless, if we do not yet have clear ideas of what those circuits should be computing. Given how limited our knowledge still is on the computations done in our brain, to process language, for example, or sustain consciousness, we suspect that the path will be slow and problematic. Chalmers is right when he claims that if there is a singularity, it will be a turning-point in human history. However, if the slowdown hypothesis has some basis, then perhaps we should also worry about its possible chilling effect on the course of research on the mind. In fact, if it were to confirm the trend of research on the mind to slow its progress due to the persistent tendency of making the explanation of a given aspect of the mind depend on understanding other mental phenomena, then we risk appreciating the magnificent complexity of the human mind, but without knowing how to cope with it.

References

Adorjan, Peter, Christian Piepenbrock, and Klaus Obermayer. 1999. "Contrast adaptation and infomax in visual cortical neurons." *Reviews in the Neurosciences* 10:181–200.

Anderson, Michael, and Susan Leigh Anderson, eds. 2011. *Machine Ethics*. Cambridge (UK): Cambridge University Press.

Cadieu, Charles, Minjoon Kouh, Anitha Pasupathy, Charles E. Connor, Maximilian Riesenhuber, and Tomaso Poggio. 2007. "A Model of V4 Shape Selectivity and Invariance." *Journal of Neurophysiology* 98:1733–1750.

Campbell, Murray, A.Joseph Hoane, and Feng hsiung Hsuc. 2002. "Deep Blue." *Artificial Intelligence* 134:57–83.

Chalmers, David. 2010. "The Singularity: A Philosophical Analysis." *Journal of Consciousness Studies* 17:7–65.

Deco, Gustav. 2001. "Biased competition mechanisms for visual attention in a multimodular neurodynamical system." In *Emergent neural computational architectures based on neuroscience: towards neuroscience-inspired computing*, edited by Stefan Wermter, Jim Austin, and David Willshaw, 114–126. Berlin: Springer-Verlag.

Dittman, Jeremy S., Anatol C. Kreitzer, and Wade G. Regehr. 2000. "Interplay between Facilitation, Depression, and Residual Calcium at Three Presynaptic Terminals." *Journal of Neuroscience* 20:1374–1385.

[1] `http://www.techjournalsouth.com/2011/08/ibms-brain-like-cognitive-chips-can-learn-video/`

Dreyfus, Hubert. 1972. *What Computers Can't Do: A Critique of Artificial Reason*. New York: Harper and Row Pub. Inc.

Dreyfus, Hubert L., and Stuart E. Dreyfus. 1986. *Mind Over Machine: The Power of Human Intuition and the Expertise in the Era of the Computer*. New York: The Free Press.

Dufort, P.A., and C.J. Lumsden. 1991. "Color categorization and color constancy in a neural network model of V4." *Biological Cybernetics* 65:293–303.

Elman, Jeffrey L. 1990. "Finding structure in time." *Cognitive Science* 14:179–221.

Ferrucci, David, Eric Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya A. Kalyanpur, Adam Lally, William Murdock, Eric Nyberg, John Prager, Nico Schlaefer, and Chris Welty. 2010. "Building Watson: An Overview of the DeepQA Project." *The AI magazine* 31:59–79.

Gardner, Howard. 2006. *Multiple Intelligences: New Horizons*. New York: Basic Books.

Good, Irving John. 1965. "Speculations Concerning the First Ultraintelligent Machine." In *Advances in Computers*, edited by Franz L. Alt and Morris Rubinoff, Volume 6, 31–88. New York: Academic Press.

Hodgkin, Alan Lloyd, and Andrew Fielding Huxley. 1952. "A quantitative description of ion currents and its applications to conduction and excitation in nerve membranes." *Journal of Physiology* 117:500–544.

Hubel, David, and Torsten Wiesel. 1959. "Receptive fields of single neurones in the cat's striate cortex." *Journal of Physiology* 148:574–591.

Jiang, Xiong, Ezra Rosen, Thomas Zeffiro, John VanMeter, Volker Blanz, and Maximilian Riesenhuber. 2006. "Evaluation of a Shape-Based Model of Human Face Discrimination Using fMRI and Behavioral Techniques." *Neuron* 50:159–172.

McCormick, David A., and John R. Huguenard. 1992. "A Model of the Electrophysiological Properties of Thalamocortical Relay Neurons." *Journal of Neurophysiology* 68:1384–1400.

Menary, Richard, ed. 2010. *The Extended Mind*. Cambridge (MA): MIT Press.

Miikkulainen, R. 1993. *Subsymbolic Natural Language Processing: and Integrated Model of Scripts, Lexicon and Memory*. Cambridge (MA): MIT Press.

Modha, Dharmendra S., Rajagopal Ananthanarayanan, Steven K. Esser, Anthony Ndirango, Anthony J. Sherbondy, and Raghavendra Singh. 2011. "Cognitive Computing." *Communications of the Association for Computing Machinery* 54:62–71.

Plebe, Alessio. 2007. "A Model of Angle Selectivity Development in Visual Area V2." *Neurocomputing* 70:2060–2066.

Plebe, Alessio, and Rosaria Grazia Domenella. 2006. "Early Development of Visual Recognition." *BioSystems* 86:63–74.

Rolls, Edmund. 1992. "Neurophysiological Mechanisms Underlying Face Processing within and beyond the Temporal Cortical Visual Areas." *Philosophical transactions of the Royal Society B* 335:11–21.

Rudolph, Michael, and Alain Desrexhe. 2005. "An Extended Analytic Expression for the Membrane Potential Distribution of Conductance-Based Synaptic Noise." *Neural Computation* 17:2301–2315.

Rudolph, Michelle, and Alain Desrexhe. 2007. "An Extended Analytic Expression for the Membrane Potential Distribution of Conductance-Based Synaptic Noise." *Neural Computation* 17:2301–2315.

Searle, John R. 1980. "Mind, Brain and Programs." *Behavioral and Brain Science* 3:417–424.

Shannon, Claude. 1950. "Programming a computer for playing chess." *Philosophical Magazine* 41:256–275.

Shapiro, Larry. 2011. *Embodied Cognition*. London: Routledge.

Simoncelli, Eero P., and David J. Heeger. 1992. "A Computational Model for Perception of Two-dimensional Pattern Velocities." *Investigative Opthalmology and Visual Science Supplement* 33:1142.

Tomasello, Michael. 2009. *Why We Cooperate*. Cambridge (MA): MIT Press.

van Pelt, Jaap, Andrew Carnell, Sander de Ridder, Huibert D. Mansvelder, and Arjen van Ooyen. 2010. "An algorithm for finding candidate synaptic sites in computer generated networks of neurons with realistic morphologies." *Frontiers in Computational Neuroscience* 4:1–17.

Vinge, Vernor. 1993. “The Coming Technological Singularity: How to Survive in the Post-human Era.” *Proc. Vision 21: Interdisciplinary Science and Engineering in the Era of Cyberspace*. Lewis Research Center: NASA, 11–22.

von der Malsburg, Christoph. 1973. “Self-organization of orientation sensitive cells in the striate cortex.” *Kybernetic* 14:85–100.

von der Malsburg, Christoph, and David J. Willshaw. 1976. “A mechanism for producing continuous neural mappings: ocularity dominance stripes and ordered retino-tectal projections.” *Experimental Brain Research* 1:463–469.

Wallach, Wendell, and Colin Allen. 2008. *Moral Machines: Teaching Robots Right from Wrong*. Oxford (UK): Oxford University Press.

Weisberg, Michael. 2007. “Three kinds of idealization.” *The Journal of Philosophy* 12:639–661.

Wilson, Matthew A., and James M. Bower. 1989. “The Simulation of Large-Scale Neural Networks.” In *Methods in Neuronal Modeling*, edited by Christof Koch and Idan Segev, 291–333. Cambridge (MA): MIT Press.

Zhaoping, Li. 2005. “Border Ownership from Intracortical Interactions in Visual Area V2.” *Neuron* 47:143–153.