Artificial Intelligence: Philosophical and Epistemological Perspectives

Pierre Livet and Franck Varenne

Abstract Research in artificial intelligence (AI) has led to revise the challenges of the AI initial programme as well as to keep us alert to peculiarities and limitations of human cognition. Both are linked, as a careful further reading of the Turing's test makes it clear from Searle's Chinese room apologue and from Deyfus's suggestions, and in both cases, ideal had to be turned into operating mode. In order to rise these more pragmatic challenges AI does not hesitate to link together operations of various levels and functionalities, more specific or more general. The challenges are not met by an operating formal system which should have from the outset all the learning skills, but -for instance in simulation- by the dynamics of a succession of solutions open to adjustments as well as to reflexive repeats.

1 Introduction

The question "Can machines think?" as Alan Turing asked in the first sentence of his famous paper Turing (1950) became a plausible question with the appearance of new computing machines in the forties of the last century. It is linked with other questions, about mathematical creativity: "Can computers discover interesting mathematical theorems?", or about decision: "Can computers find better solutions to collective decision problems?" (see chapter "Collective Decision Making" of Volume 1) "Artificial Intelligence" has been coined in 1956 in order to cover these questions (see chapter "Elements for a History of Artificial Intelligence" of Volume 1). This slogan was deliberately provocative and raised passionate debates about the previously

P. Livet (✉)
Aix-Marseille Université, Marseille, France
e-mail: pierre.livet@univ-amu.fr

F. Varenne
University of Rouen, Rouen, France
e-mail: franck.varenne@univ-rouen.fr

P. Marquis et al. (eds.), *A Guided Tour of Artificial Intelligence Research*,
https://doi.org/10.1007/978-3-030-06170-8_13

420055_1_En_13_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:18/6/2019 Pages: 19 Layout: T1-Standard

assumed monopoly of living bodies and particularly humans on these cognitive abilities. This debate smoulders under the embers and is reactivated from time to time, but its main arguments have been already expressed long ago. Computers are better than humans for operating on multiple and discrete formalized data (including games like checkers, chess and now Go) but their capacity to deal with the relevance problem and to be creative (in the – rather fuzzy human sense) are still disputable. The intensity of the debate has decreased, but other interesting philosophical and epistemological questions have been raised in a less passionate way (Ganascia 1990). Whether real progress has been made on these new questions is still a matter of debate, but the progress of AI leads us to change the way we ask these questions and at least the way we could conceive the answers. These questions are strongly related to each other. They are even entangled and, as a consequence, difficult to expose separately. Maybe showing how the old debates can be reconsidered from the present perspectives could be useful for making the shift of perspective easier to appreciate.

2 Three Classical Debates: Turing's Test, Searle's Chinese Chamber, Dreyfus' Phenomenological Arguments

2.1 Turing's Test

In his paper in Mind, Turing sets out his test in a relatively complicated way, but this complexity is meaningful. A human people H1 is supposed to dialog with two entities. The first one is a human being (H2 – in the initial game, it was a women), the second one a machine that is supposed to try to imitate the human being H2. If H1 cannot distinguish between the machine and the human being H2, the machine wins the game. Why not to compare directly H1 with the machine? The reason is that the properties that we want to test in this game are not the intrinsic properties of H1 and the machine - there are obvious differences. What we want to test instead, is to what extent the observable speech behaviour of the machine is similar to the observable speech behaviour of H2, at least at the best level of approximation that is available for H1. Nowadays, for usual conversations, computers can pass the test - except for tricky contextual dependencies and creative metaphors. Do not forget that this success is partly due to the limitations of the discrimination capacities of human agent and human observer in a usual interaction. The peculiarities of the Turing's test are also useful in order to avoid a difficulty of the simpler version (interaction reduced to H1 and the machine): H1 could believe he interacts with a human being because he over-interprets the sentences of the machine. For example, people interacting with the program ELIZA Weizenbaum (1966, 1983) found it "human": ELIZA was programmed for building sentences, using some elements of the questions asked by people and answering by asking other questions related to the psychoanalytic stereotypes triggered by words belonging to the human sentences. People were over-interpreting the personal relevance of these stereotypes. Maybe the

comparison between the sentences of ELIZA and the ones of a real psychoanalyst would have allowed them to make a difference? A human being could invent creative and relevant new associations that other human beings could regard as relevant - even if they do not understand their precise meaning.

One can make the hypothesis that the best AI machine could not produce such new associations. But such hypothesis cannot be proved, as there are no strict criteria for the relevance of associations that are creative but not precisely understood. Turing's test does not require such impossible proof, but its result is necessarily vague, since it depends on the limitations and approximations of human cognition.

2.2 Searle's Chinese Room

Searle claims that the story of the Chinese room (Searle 1980) – shows that AI fails to pass the Turing's test – in its simple but stronger version: Indiscernibility between the intrinsic properties of H1 and the machine. Of course, this stronger and simple version (SSTT) is not Turing's version – a test of the indiscernibility of the behavioural properties of H2 and the machine. Does the Chinese room story show that AI fails to pass the real or true Turing's test (TTT)? (Saygin et al. 2000) Searle is in a room. Someone passes him through a hole an ideogram Chinese text. Searle does not know Chinese, but has at his disposal a handbook that gives him, for any combination of ideograms, another combination (the handbook is the equivalent of a program). He identifies one combination in the Chinese text, writes the correspondent combination of the handbook and passes it back through the hole. As the handbook is well made, the sequence of combinations happens to be a meaningful conversation in Chinese. Nevertheless Searle cannot pretend to understand Chinese.

Searle claims that the difference between a human being and a program is that the operations of the program are only syntactic while the human being's cognitive processes require semantics. Let us now apply TTT. Do the human being's discriminative abilities make him able to distinguish between two behaviours, if the first one is the result of syntactic operations and the second one the result of syntactic plus semantic processes? In the Chinese room, Searle believes that the ideograms are related together by the syntactic correspondences specified in the handbook. Maybe they have a semantic counterpart, but only for the Chinese people outside the room. Or, to be more precise, the semantics of the handbook is a purely formal one: the formal correspondence between two sequences of ideograms. This formal correspondence is very poor relatively to the real semantics that relates written symbols to a lot of things and activities in the world. Therefore we have shown that Searle can distinguish two kinds of semantics, one limited to the transcription operations between ideograms inside the room, and the other that relates ideograms to things inside and outside the room. Searle seems to believe that this result shows that these syntactic operations are the intrinsic properties of the AI machine, while the human abilities to establish semantic relations are intrinsic human properties. This conclusion would require the stronger test, which requires comparing the intrinsic properties of the

420055_1_En_13_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:18/6/2019 Pages: 19 Layout: T1-Standard

machine to the intrinsic properties of Searle. But semantic relations are not purely intrinsic they need external relata. And, according to Searle, the internal basis of mind is a biological one let say, the dynamics of our neural system. And they may be a similarity between the operations of a program and the dynamics of neural connections- namely, that syntactic operations preserve semantic relations in each case. Of course, focusing on the syntactic operations that preserve semantic relations seems to beg the question. The stronger version of Turing's test cannot be conclusive because its conditions cannot be satisfied. Let us come back to the TTT. We would have to build two Chinese chambers, one for Searle, the other for the AI machine. Another human operator would have to distinguish the AI machine and Searle's behaviours. As by construction there are no differences between the operations of the machine and Searle's processes, the AI machine would pass the test.

Why does Searle conclude in favour of the opposite proposition? Because he believes that a semantic relation involves all the possible relations that a human being can have with his world. In this case he can complain that in the Chinese room, the relations to the outside world are limited to the transcription operations, while the Chinese people have many more relations at their disposal. If Searle were not accustomed to these richer relations, he would not complain to be limited in the Chinese room. Let us now consider the AI machine. In the TTT without Chinese room, the machine is related to the external world by some devices, at least the ones that make it able to interact with H1 and H2. So the "real" TTT would imply to compare Searle in the real world with the AI machine in the real world. And the old problem of TTT would reoccur: Are we able to detect salient and stable differences between Searle's interactions with the real external world and the interactions of the AI machine with the same external world? Are we able to give to AI machines the tools for developing semantic relations that are admissible by human beings? This question has no definite answer because we, as human beings, do not know the limits of the realm of such relations.

Let us shift to what seems a less difficult question: When our AI machines present limitations because the syntax we gave them cannot preserve our semantic relations, will we be able to solve this problem? Will we have the syntactic abilities to solve it? In the formal domain, in which differences are strictly defined, one could interpret Gödel's theorem as showing that this is not always possible. But in the pragmatic domain of ordinary life, differences are not so strict. Vagueness of human semantics could lead us to consider too easily some differences between men and AI machines as admissible while regarding also too easily other differences as not admissible.

2.3 AI and the Phenomenological Approach

Dreyfus and Dreyfus (1988) and Dreyfus (1992) have tried to show that as AI uses computation on symbols governed by formal rules, it cannot be akin to human intelligence. Their starting point was an analysis of the human experience in the first person of shifting from one level of competence to a higher one, from the status of

novice to the ones of advanced beginner, of competent, of proficient and of expert. The novice tries to apply rules, but as he goes further he accumulates experiences in different contexts and become able to select the relevant elements in each context (Collins 1996). At the end of this contextual learning process, he has no longer to search for the right decision, he just sees it. What he has to do appears to him obvious, but this does not imply that he is able to explain how he manages to know that it is the good decision. This learning process cannot be the result of external programming, but only of a kind of progressive incorporation of diverse kinds of sensitivity to little clues that get their meaning only from the contextual situation. This learning story leads Hubert Dreyfus to believe that without a biological body human intelligence cannot be approached. But the story seems only to show that without integrating multiple experiences and being able to evolve autonomously in interaction with its environment, no system could pass the Simple version of the Turing test (STT).

Let us assume that STT implies that H1 detects the difference between the human experience in the first person and the experience of a machine. As it seems impossible to know by experience what could be the experience in the first person of an AI machine, the phenomenological perspective seems to make impossible to give a meaning to the STT. Nevertheless, we can apply the TTT. For example, we may give the role of H1 to a novice, the role of H2 to an expert and examine whether the novice detects any difference between the machine and the expert. We have to experiment every possible combination, with expert in the role of H1, the novice in the role of H2, etc. We could rank the results of TTT's on a scale. For example, it seems easier for AI machine to pass the test when H1 is a novice and H2 is an expert, because the novice understands the machine as badly as the expert. Some situations have a flavour of paradox. AI machines might pass the test when H2 is a novice and H1 is proficient or expert - if the operations of the machines are clumsy and if the limits of their program are similar to the limits of the heuristics of novices. They pass the test when H1 and H2 are novices, if the possible differences of the limits of the machine with the limits of the novice are difficult to detect for a novice. A competent person can be unable to detect differences between competent H2 and a high level machine, because, again, she could have similar difficulties to understand both the expert and the machine. The most unfavourable situation for the machine to pass the test is that H1 and H2 are proficient or expert. But in the domain of games that can be formalized, as chess and GO (see chapter "Artificial Intelligence for Games" of Volume 2), AI machines pass the test even with experts. The conclusions of these different TTT's seems to be that the limitations of AI cannot be defined separately neither from the limitations of human abilities nor from the fact that we build AI machines as complement of our own limitations. Nevertheless there are limitations for which we have difficulties to define what the useful complements would be, because we do not know our own modus operandi. For example, we do not know precisely how a human being can become an expert, how he can integrate his experiences and become able to extract from them methods of evaluating and choosing in various situations the relevant way for him to define the problem and to find a solution. Would we have the opportunity to know better our integration processes, maybe it would be difficult to optimize them. Let us suppose that we use some kind of "deep learning". Deep

420055_1_En_13_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:18/6/2019 Pages: 19 Layout: T1-Standard

learning programs accelerate a lot the treatment of information and take into account many more combinations (see chapter "Designing Algorithms for Machine Learning and Data Mining" of Volume 1). But they are sensitive to the configuration of the available information, and this configuration can induce biases. For example, racist sentences are frequent on Internet, and finding ways of discarding them is a difficult problem for human, and also for machines. The conclusion of this section would be that we have to acknowledge that in these disputes around the test, we cannot separate the question of the human limitations and the one of the AI limitations. Given this conclusion, we prefer to evaluate the results of AI by comparing them with the challenges that AI researchers have themselves defined and by analysing the conceptual evolution of their research program.

3 Initial Challenges of AI Program of Research

AI program of research, at its beginning, could be defined as computo-representationalist and internalist. Intelligence was supposed to handle problems by building representations of their elements. Representations were expressed in formal symbols. Intelligent inferences and reasoning were operated by computational devices. The context, the environment had also to be represented and these representations also were internal to the computational system. AI was supposed to take up different challenges. We will formulate them, not necessarily in their initial terms, but in terms that are inspired by the evolution of AI:

1. To solve complex problems, without previously knowing a procedure for demonstrating their solutions – or at least to give a satisfying approximation of a solution. Procedures that could be logically validated are preferred (see chapter "Automated Deduction" of Volume 2). This computational challenge has been partly taken up.
2. To build computational systems that can learn, can find generalizations of their procedures and extend them to other domains than the one of their learning example basis (see chapter "Statistical Computational Learning" of Volume 1 and chapters "Automated Deduction", "Belief Graphical Models for Uncertainty Representation and Reasoning", "Planning in Artificial Intelligence" of Volume 2). This challenge is also partly taken up, even if it is always problematic to avoid generalizations that do not take into account contextual differences.
3. To simulate creativity in more informal domains (to prove interesting theorems, find new solutions and methods). This challenge can be seen as a combination of the first and the second challenges. The difficulty lies in the fact that combinations that bring forth new results have to be selected in order to extract only the relevant ones.
4. To give a formal account of usual human reasoning. Simply building inferential systems is not sufficient here, because the relevance of usual reasoning is sensitive to contexts and the relevant definition of context cannot always been determined in advance. The development of inferences can lead to revise it (see chapters

420055_1_En_13_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:18/6/2019 Pages: 19 Layout: T1-Standard

"Argumentation and Inconsistency-tolerant Reasoning" of Volume 1 and chapter "Planning in Artificial Intelligence" of Volume 2).

5. To understand human languages and converse with people in a relevant way. Problems here are not only linguistic, but also pragmatic and contextual (see chapter "Artificial Intelligence and Natural Language" of this volume).
6. To decide and monitor relevant actions. This challenge implies to solve the frame problem (i.e.: While developing an action, one has to take into account only the changes of the situation that are relevant for achieving the task); the qualification problem (when one has to determine what are the properties and qualities of the situation of the action, one has to select only the qualifications that are relevant for the task), and the problem of ramification (when examining the tree of possible alternatives, one has to avoid exploring the branches that are not relevant for the action) (see chapter "Reasoning about Action and Change" of Volume 1 and chapter "Planning in Artificial Intelligence" of Volume 2). Here again the problems are not only syntactic and semantic, but imply pragmatic relevance and its extensions to elements that were at first external to the representations of the system (see chapters "Artificial Intelligence and Pattern Recognition, Vision, Learning" and "Robotics and Artificial Intelligence" of this volume).

Other more ambitious challenges of the computationalist program of cognitive sciences were philosophical challenges like giving a computational account of intentionality, qualia, and consciousness (Garbay and Kayser 2011; Kayser et al. 2004; Penrose 1989; Pitrat 1995). In order to take up these challenges, research on programs has to try to overcome the problems of relevance and of interactions that depend on context and environment. The recurrence of these problems has made difficult to maintain a pure internal perspective and led to revise the problematic.

4 How Evolution of AI Shifts Epistemological Perspectives on Intelligence

At the beginning of AI, the usual analysis of the sequence of an intelligent behaviour was roughly the following: Perceiving a situation, extracting the more meaningful features, retrieving in memory the knowledge useful for inferring the consequences of these features and combining different known data until finding a combination (for example an action plan) that satisfies expected requirements (see chapters "Artificial Intelligence and Pattern Recognition, Vision, Learning" and "Robotics and Artificial Intelligence" of this volume). Evolution of AI led us to notice that this representation of intelligence omitted essential elements. Perception cannot be reduced to the neutral collection of a lot of data coming from the external world and triggering a deductive process. It has to be conceived as continuous interaction with environment, an interaction oriented towards goals. These goals can be modified because of the evolution of the external world, but also when one updates the action because of the already acquired results. In the same way, combining different pieces of knowledge

420055_1_En_13_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:18/6/2019 Pages: 19 Layout: T1-Standard

that could be useful is not a simple task, because their kinds could be very different. In addition one has to allow some local combinations to imply contradictions while avoiding that these contradictions paralyze the system (see chapter "Argumentation and Inconsistency-tolerant Reasoning" of Volume 1). An individual (and a fortiori a collective of individuals) has to be able to use information that she does not completely control, or that refers to actions and processes that are still in progress. One needs to know how to integrate partial, uncertain and vague data on the action during the action itself (see chapters "Knowledge Representation: Modalities, Conditionals, and Nonmonotonic Reasoning", "Representations of Uncertainty in Artificial Intelligence: Probability and Possibility", and 'Case-Based Reasoning, Analogical Reasoning, and Interpolation'" of Volume 1). This continuous process of updating and revising was not central in the initial agenda of AI.

New questions arise. Is the heterogeneity of the components needed for an intelligent behaviour irreducible? Is it impossible to conceive a formal language that could express every component of an intelligent knowledge (universal expressiveness)? What would be the relation between this formal language and the natural languages? Is a specific logic required for each of the different modes of combination that are specific to each kind of components, or does a universal logic exist that rules any kind of combination? How to deal with vagueness?

With regard to the problem of expressiveness, AI brings a new perspective, the philosophical implications of which have not still be explored (see chapters "Validation and Explanation" and "Knowledge Engineering" of Volume 1). Human intelligence does not require finding a formal expression for anything in order to compute everything. Intelligence has a pragmatic dimension, related to the urgency of decision. If a fire flares up, one has to decide immediately whether there are ways of extinguishing it or whether it is better to get out. AI has made possible a finer evaluation of the trade-off between expressiveness and efficiency of the systems of representation: If the language of description is "too much" expressive, the decision system will be "too" slow. One of the results of AI is to make us aware of a lot of similar trade-offs in intelligent behaviours. These trade-offs are related to the different variants of the frame-problem: How to define what we are entitled to neglect? How fine the grain of description has to be depends of the variety of the rhythms of the task. The grain of the definition of the sub-tasks has to be variable, each level of granularity (see chapter "Representations of Uncertainty in Artificial Intelligence: Probability and Possibility" of Volume 1) - implying a more or less fine grain of the language of description – implying a different speed of inference that has to fit the required speed of decision. When the goals of the system change, new elements have to be expressed in a finer grain while previous details can now be neglected. To some extent, computer programs that use this category of self-observation modules called "activity tracking" and "activity awareness" for their algorithms can help to increase this optimization at a runtime (see chapter "Reasoning with Ontologies" of Volume 1).

These remarks lead to focus on the meta-cognitive dimension of the intelligent behaviour (Proust 2013). This dimension cannot be reduced to "reflection" but consists also in finding efficient ways of combining memory accessibility and infor-

420055_1_En_13_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:18/6/2019 Pages: 19 Layout: T1-Standard

mation about the states of present processing. In order to make the right decision, knowing how long it would be to make an in-depth reasoning, or what is the ratio between the amount of knowledge that has been already mobilized and the amount of information that could be processed with a bigger effort and in a longer time are meta-information that matter. It becomes clear that we need a measure of the distance between the present state of reasoning or the process of decision and the solution that we are looking for. If this distance increases or does not decrease, a meta-level examination of the cognitive or decisional strategy and possibly a change of strategy have to be triggered. When the question about what is the more efficient level of granularity becomes relevant for the evaluation of an intelligent behaviour, reasoning can no longer be reduced to its version in the framework of classical logic. In this framework, logical reasoning is submitted to the constraint of saving truth: When it is applied to true premises, it has to give true conclusions. But when premises may have different granularity, they are not required to be true in an absolute sense. What matters is not to get absolutely true conclusions, but conclusions that are still relevant if considered at the degree of granularity that is relevant for the task. This evolution leads to build non-classical logics (AI has been very productive in this domain), or to put in the background the logical concerns and to work on ways of processing uncertain and vague propositions. One could believe that the normative role of logic is challenged, but it is not the case, as logic is still needed in order to ensure that programs really do what we want they to do (see chapters "Automated Deduction", "Logic Programming" of Volume 2, and chapter "Theoretical Computer Science: Computational Complexity" of this volume) (Sallantin and Szczeciniarz 2007).

Relations between rationality and strict logical validity become less close when AI has to deal with the problems of the computation time and of computational complexity. Cook's theorem (1971) reminds us that satisfiability in propositional logic is NP-complete. Instead of focus on logical truth and its formal versions, a more pragmatic orientation can be taken. Reasoning has to be coherent, but also functional. The coherence of information, as long as the information does not disturb the efficiency of the task in progress, may be not continuously checked, but incoherence of information that would cause troubles and decrease in efficiency has to be corrected. The way of checking and validating the coherence in a reasoning or decision task may be specific to the type of the task. Generalizing one kind of validity to a larger context may require redefining this validity. Complex context dependency, including dependency of contexts on other contexts in accordance with a given architecture, may require distinguishing and articulating different levels of validity. Generalization may imply to take more risks and to pass from truth and validity to the less demanding notion of normality.

Taking into account the dependency of a task on its context, or criteria of what is desirable for a task, as well as the uncertainty of the data and information that are at one's disposal, leads to consider "normal" inferences. Normal inferences are valid in normal circumstances, but if the context changes, they may be defeated (see chapter "Main Issues in Belief Revision, Belief Merging and Information Fusion" of Volume 1). The guiding principles of intelligent behaviour are not restricted to normality (see chapter "Norms and Deontic Logic" of Volume 1) and normativity (see chapters

"Knowledge Representation: Modalities, Conditionals, and Nonmonotonic Reasoning" and "Representations of Uncertainty in Artificial Intelligence: Probability and Possibility" of Volume 1), but these notions are related to ways of hierarchizing the different possible states (see chapter "Artificial Intelligence and High-Level Cognition" of this volume). Defining a hierarchy is usually the privilege either of a culture or a subjectivity. Intelligence depends also on cultural and even subjective factors. In this perspective, our well-founded inferences allow us to anticipate what can be "normally" expected. In order to decide an action, one has to compare what will normally follow this action with what will normally follow if we choose another action or do nothing. Usually, we make a distinction between "normal" denoting "things that happen frequently", and "normal" in the sense of "normative". We are in a first step more sensitive to information that is "abnormal" in comparison of what we expect and want, and in a second step we try to explain the reason why our expectation has been defeated by finding some anomaly in the environment. When we have identified an anomaly, we try to infer, by abduction, what plausible process could have produced it. AI has contributed by developing non-monotonic logics, and different systems of probabilistic revision, and has tried to give an operational content to the notion of norm and, correlatively, to the notion of exception by playing with the double meaning of "normal". It is possible to relate these two meanings by conceiving the "normative normal", this expression denoting not what is necessarily frequent but what would become so if things evolve the right way.

This notion of normal relation is close to the Humean notion of cause (Hume 1987) as implying the regularity of the relations between antecedent and consequent. There are no claims in AI about the metaphysical question of the existence of causal laws, but AI requires giving an operational content to the notion of causality (see chapter "A Glance at Causality Theories for Artificial Intelligence" in Volume 1 and chapter "Main Issues in Belief Revision, Belief Merging and Information Fusion" in this volume). For example, an intelligent system has to diagnose breakdowns, and to try to find out what are the causes of breakdowns. A robot has to plan its actions and to know what effects its actions will cause. AI seems to favour the "interventionist" conception of causality: A is regarded as a cause of B in context C if A is an exogenous intervention and if, in context C, B is known as a normal consequence of this intervention, while if A would have been absent, B would not be regarded as a normal consequence. Von Wright (1971) was one of the firsts to propose this conception, and Bayesian networks are one of its formal expressions (see chapter "Belief Graphical Models for Uncertainty Representation and Reasoning" of Volume 2). Philosophers suspected this conception of being too anthropomorphic, but this objection has less weight since AI has operationalized this conception.

AI has shed a new light on another topic of philosophical considerations, the status of language and meaning (see chapter "Artificial Intelligence and Natural Language" of this volume). In its beginnings, AI was mainly focused on the first two members of the tripartition between syntax, semantics and pragmatics. The problem of the dependence of meaning on the context was still let aside. But the acceptability of a sentence cannot be ensured without taking into account the context, and this notion of context implies a mixture of semantics and pragmatics. More recent

researches in the domain of computational linguistic are no longer restricted to the articulation of syntax and semantics. They exploit statistical correlations in huge corpus of linguistic data and associate them to different uses of language in different social and pragmatic situations. They analyse also the evolution of the network of words and of co-occurrences of words, as well as the dynamics of conversations. They relate more directly the regularities of co-occurrences of linguistic symbols and their pragmatic contexts of use. The tools developed in AI in order to deal with the problem of normality have influenced numerous works on language, including the use of analogies. Relations between syntax and semantics can be considered under a new light, and even the approach of pragmatics first developed in the philosophical current of analysis of "ordinary language", which was still focussed on a "grammar" of uses, and not on an evolutionary network of statistical correlations, could be renewed by these new perspectives.

The development of AI leads also to consider the pragmatic conditions of AI itself: The pragmatics of the interactions between programs and users (human beings are not the only category of users) (see chapter "Negotiation and Persuasion Among Agents" of Volume 1). The dynamics of these interactions has to be examined, simulated and formalized. One can refer to Sallantin's proposal (Sallantin and Szczeciniarz 1999): Thinking of a proof not only as a formal deduction, but as giving to the person that understand it new abilities to make inferences, and ways of overcoming cognitive obstacles. This implies to take into account together the dynamics of a proof and its capacities of interaction - or even "transaction" - with its potential public. In a different perspective, Jean-Yves Girard (2001) remains in the core of the fundaments of logic - the benchmark for the validity of programs – but nevertheless regard proofs as interaction - interactions with possible counter-proofs.

These epistemological shifts lead AI researchers to raise new problems and put out new challenges: Finding representation formats and processes for manipulating these formats that make possible to change the operational mode in accordance with contexts and problems. This implies to have at one's disposal a sufficiently diverse and rich variety of normalities; ensuring the possibility to set new normalities for new contexts. Operational devices have then to be possibly reusable and also adjustable and revisable. We could name this challenge the problem of a "dynamical capacity": Operations that include the possibility of dynamically modifying their functionality (see chapters "Reasoning about Action and Change" and "Formalization of Cognitive-Agent Systems, Trust, and Emotions" of Volume1, and chapter "Planning in Artificial Intelligence" of Volume 2. See also Livet (2002a, b)

Solving the problem of the dynamical capacity would also solve the problem of frame, of qualification, of ramification. Their solutions need to reintegrate in monitoring and planning of action the observed deviations of the course of action (deviations with respect to the initial anticipations about what plans of action were relevant for the goals of the action). In order not to have to compute every possible state, these deviations have not been taken into account, but some of them need to be considered. These unanticipated or neglected elements are the source of difficulties in the three forms of the frame problem.

420055_1_En_13_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:18/6/2019 Pages: 19 Layout: T1-Standard

The reader has perhaps noticed that the philosophical challenges (intentionality, qualia, consciousness) are no longer the main concerns of AI. One could think that this loss of interest is the result of the revision of the pretensions of the initial program of AI. But there is a more subtle reason. It could be an effect of a revision of the relations between the properly philosophical horizon of these claims and challenges, on the one hand, and the new horizons and challenges of AI, on the other hand.

5 The Right Place of Philosophical Challenges

Among the many challenges AI was said to take up, philosophical ones seemed to be the most difficult. Maybe this was an illusion. It may have come from the fact that philosophical requirements concerning intentionality, qualia, and consciousness had been somehow idealized and defined in a too speculative and operational way. Philosophers, when pointing out the properties that have to be satisfied, define them in their strongest form and choose their more normative and ideal definition. The properties defined in such a way could be called "horizon-properties". Philosophers do not even indicate how to find operational processes that satisfy these properties. Let us consider again things in the TTT spirit. We would have to compare to the horizon properties of the philosophers their AI-corresponding horizon-properties: The properties that the actual accomplishments of AI would be supposed to satisfy in their idealized development - a kind of horizon of the operational. For example the Universal Turing machine could be regarded as an idealized development (that can be carried out ideally, but not in a realistic time) of real Turing machines with their effective program. We could also demand, symmetrically, from philosophers that they give us operational versions of the philosophical horizon-properties, but this would require that philosophy of mind and its collaboration with cognitive psychology and neurosciences would be more advanced. In these symmetrical perspectives, it would be possible to show that, if we suppose that AI is able to give at least partial solutions to the problem of "dynamical capacity", which is an horizon-property for AI, we can believe that it will take up the philosophical challenges. Finding a solution to this problem requires that the initially built up representation structure still offer the possibility of changes (this is the dynamical part) and that the possibility of these changes is at least partially present as dispositions (this is for capacity) of the operations associated to this structure (partially, because the modifications of these operations, while belonging to their dispositions, have also to be triggered by their interaction with a new context). The functioning of an operational dynamic capacity would imply:

1. A computo-representational functioning with its regularities.
2. The insertion of its operations in a new environment, insertion that triggers variations.

420055_1_En_13_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:18/6/2019 Pages: 19 Layout: T1-Standard

3. The capacity of the computo-representational system of re-formatting itself and inducing re-categorizations - if by "re-categorizations" we mean recalibration of these operations in accordance with the normalities of the new context.

Intentionality in the philosophical sense (capacity of making reference to external as well as internal referents as considered under a specific aspect) requires the satisfaction of these three conditions, but they are sufficient for intentionality. The reason is that insertion in an environment makes possible to have relations to referents, which are grasped through representations under a given format - an ersatz of the notion of aspect– defined by the computo-representational system (see chapter "Validation and Explanation" of Volume 1). These representations do not by themselves give us access to the exogenous referents, but they can be modified in accordance with changes of the referent. If we do not take a static perspective, but a dynamical one (see chapters "Main Issues in Belief Revision, Belief Merging and Information Fusion", "Case-Based Reasoning, Analogical Reasoning, and Interpolation", "Argumentation and Inconsistency-Tolerant Reasoning" and "Reasoning about Action and Change" of Volume 1), the evolution of the referent is strongly related to the evolution of the mode of representation (the format), and this relation is a way to satisfy the philosophical criterion for intentionality. Of course we have no guarantee that this always happens, this is only a possibility, but an operational version of philosophical intentionality cannot ensure a better guarantee.

The same three conditions are sufficient for the production of "qualia" (qualitative experiences in the first person, "what it is like" to smell the odour of a rose). What is needed is that the effective functioning of the interactive operations with a specific environment brings forth particularizations of the canonical format of representation. In AI, these particularizations can be the results of various constraints: constraints inherent to the physical implementation of computational operations, or specifications of the modalities of data capture, as well as specific versions of a program, or particular parameterizations, or specific effects of some coordination between different kinds of programs, etc. If these particularizations can trigger some evolution of the operations of categorization, it is possible to regard them (in a dynamical, not in a static perspective) as both the operational and the AI horizon-version of the philosophical qualia: phenomenal contents that modify the content of basic categorizations by particularizing them and even giving them a singular content.

Consciousness presupposes qualia and their capacity to particularize categorizations. To be conscious is to be able to integrate information that is not already categorized, or is particularized, together with information already categorized and more generic. We can be convinced of that proposition if we consider the phenomenology of our conscious states: The conscious representation of a situation involves usual categorizations together with related elements that may not be categorized yet, but can be used as a pool for possible evolutions of categories. Our conscious integration has to save both the peculiarities of information in its phenomenality and the capacity of generalization or systematicity of the information.

The paradox of the Chinese room is overcome when our three conditions are satisfied. A computer program that could re-categorize most of the variants that

420055_1_En_13_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:18/6/2019 Pages: 19 Layout: T1-Standard

the Chinese external observer could introduce in the sentences that express Chinese questions, so as to be able to give answers to these questions, would surely be regarded as "understanding" Chinese. Nevertheless, there would be no guarantee that for each variation of context the program could give a relevant answer. But human beings are no more ensured to give optimally relevant answers when their environment is turned upside down.

If AI researchers would imitate philosophers and their preference for idealized horizon-properties, and raise the degree of idealization of the operational capacities of AI up to the level of the horizon-property that we have called "dynamic operational capacity", it would be difficult for philosophers to distinguish the AI version of this horizon-property and the philosophical version of "dynamical operational capacity" that they have to propose if they consider an operational version of their idealized properties. Idealized operational version and operational version of idealized properties would be very similar.

6 Attempts to Take Up New Challenges

We have shown that AI could take up the challenge of dynamical capacity (relevance, flexibility with respect to changes of context, capacity of defining new normalities), but it remains to show that this is operationally plausible.

As we are tempted to relate dynamical capacity and evolution, approaches that have a more interactionist and evolutionary flavour (as connectionism or revision of rules or genetic algorithms) might have seemed more promising.

Let us regard genetic algorithms (see chapter "Meta-Heuristics and Artificial Intelligence" of Volume 2) as a way to deal with the problem of dynamical capacity. At the beginning, we have sequences of coded symbols (structures) that have functional capacities; they are submitted to the process of variations and selection of a simulated evolution, and the results of these structural variations are selected in accordance with the new desired functional capacities. It seems to give a good example of a dynamical capacity. But in this process, the relation between computational operations and functional normalities that was ensured in the first phase by the initial coding has been lost, and nothing guarantees that one will found clear relations between the new functionalities associated with the new coding and the initial ones. We may also wonder whether the introduction of an evolutionary process and the fact that these evolutions are not deterministically programmed give more chances to take into account the context, or to correctly simulate the re-constitution of a cognitive and pragmatic capacity that a change of context can bring about in human intelligence.

People often too rapidly admit that it is enough for a computational system to be evolutional and interactive (including for robots interaction with the real world) in order for this evolution and interactivity to correctly simulate the evolution and interactivity of human intelligence. But one evolutional interaction can be very different from another one.

420055_1_En_13_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:18/6/2019 Pages: 19 Layout: T1-Standard

This possible dissimilarity is accepted in AI. It is admitted that the operational solutions to the problem of dynamical capacity does not require that a unique program is sufficient for taking up the challenge. The problem may imply a combination of different functionalities and forms of representation that either co-exist (without conflict) or are activated sequentially. Computers combine in an articulated way different levels and kinds of functionalities the structure and processes of which can be very different (electronic level, compilation levels, computational level, semantic level, and may be normalities level).

Most often, simulations on computers replace deductive processes by a series of computation on symbols that do not necessarily emulate any logical inference. To this extent, simulation can be seen as more generic as many cognitivist or even connectionist use of computers in AI. More importantly, a simulation presents at least two different phases: An operative one and on observational one. During the operative phase, a simulation gives rise to computations between discrete symbols. In the observational phase, the computer simulation triggers a series of observational measurements or reuse procedures that are performed on the symbolic collective patterns or clusters arising from the operative phase. These two phases can be performed by the computer program and be run simultaneously, i.e. at runtime. As a consequence, computer simulations can take advantage, first, of the duality of their intrinsic phases and, second, of the superposition of different (at least two) symbolic levels that, accordingly, has to be recognized. Of course, the level of computation needs a substrate of electronic elements the compatibility of which with operations on symbols is ensured. But, this implies different movements back and forth between the levels and between different kinds of operations. The elements that interact in a given inner level of the computational system can be regarded as full symbols in the sense that they are referring to certain things or symbols in the target system. But they could also be regarded as sub-symbolic (not strictly decomposable – and composition is not possible for every element, while some allowed compositions can be used as emerging symbols at the upper level) relatively to the symbolic functioning of an upper symbolic level in the denotational hierarchy: Hence, the connectionist notion of sub-symbolhood can be multiplexed, contextualized and, as such, generalized. One can observe, for instance, that complex computational models involve entangled submodels, some of which are only partially formalized, others ones that use different formalisms. This entanglement at runtime demands prior contextualized sub-symbolizations. The plenary use of computer simulation involves all these processes. This is the case for standard systems of multi-modelling as DEVS (Zeigler 1976) (their supporters try nowadays to simulate a universal modelling system), or for some complex computational models (for instance coupled agent-based/equation-based models) in empirical sciences like physics, biology and the social sciences. What is proper to the forms that are the results of a specific effective mode of functioning is taken without modification in its specificity (Varenne calls that an "iconic" mode 2009, 2018). Conversely what is symbolic from the beginning (and because of that, multi-realizable) is discretized in order to get a operational specific (iconic) aspect at a lower level and, because of that, the capacity to interact with the iconic aspects of other heterogeneous components of the system of models.

We have said that simulation implies the possibility of going back and forth between the iconic level of specific functioning and the level of results endowed with symbolic genericity. These back and forth moves make then possible to give generic capacities to iconic specific functioning and conversely to recharge symbolic capacities with iconic specificities. This transformation of specific into generic, and conversely, offers a possibility of simulating the simulation, as it makes possible the auto-observation of simulations that are internal to the system. Sensitivity to context will be made manifest by a decreasing of genericity, which will lead to looking for particularity, and, if things go well, to a converse movement towards a readjusted genericity, resulting in adaptation to new normalities. Of course one has to allow in the program the possibility of modifying the basic functioning as a function of the problems that appear at the generic level, in order to implement the chosen heuristics or models of symbolic cognition. The computer simulation of human practices of simulation (a well-known ability of human practical and theoretical cognition) could be regarded as one of these steps of increasing particularity that are intertwined with procedures of re-symbolization in form recognition, re-categorization, frugal heuristics, and so on and so forth.

Let us suppose that one respects the constraint of ensuring a sufficient level of compatibility and co-functionality (as in DEVS). In addition, let us assume that the computational system has implementations of each sub-symbolic level that save the iconicity in such a way that the form of the computation results is preserved and recognizable at the other levels and, at the end, recognizable by observers external to the system otherwise simulation is only a computational trick that allows to solve a computation problem. One is then able to use all available means in order to simulate the role of context (recharging particularity, then recharging genericity). One can use sub-symbolic functioning and operations on iconic symbols. One can ensure that fine syntactic differences (differences of implementation) trigger operations that have the result of modifying categories while saving their main functioning. This modification may result from a cumulative process that reaches a threshold, or it may result from emergence (in the weak sense of the term, as it is only a result from the computational although inescapable properties of the step-by-step operations). One could design procedures of revising categories by merging different systems and hierarchies of categorization in such a simulation. It may be also possible to automatize and modularize devices for building correspondences between the different ontologies that can be extracted ex post from various re-categorizations (see chapter “Collective Decision Making” of Volume 1) (Livet et al. 2010).

At this state of the art of AI, the conditions that have to be satisfied in order to deal with the problem of dynamical capacity require computational means in order to favour the back and forth move between at least two levels (the specific sub-symbolic realizations and the new symbolic combinations that they may induce), and the development of a three phases dynamics: Functioning, perturbation, re-adjustment or revision.

420055_1_En_13_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:18/6/2019 Pages: 19 Layout: T1-Standard

7 The Laboratory of Agent-Based-Simulation

We can find examples of this use of multi-levels relations and plurality of phases in Agent-Base-Models (ABM) via their simulations of the interactions of multiple agents (see chapter "Negotiation and Persuasion Among Agents" of Volume 1) (Livet 2007). The agents are automata or well-defined pieces of programs. They interact in accordance with their own rules. Their rules are selectively triggered as reaction to their environment (time, spaces), to their neighbours, to the messages they receive and to their inner representations and goals. The collective result of their interactions can be interpreted as an emergence of collective forms. These forms have not been defined from the beginning, neither in the model nor in the program. These forms change the environment, the evolution of which can make new structures emerge, and other ones be perturbed and disappear. Rules and parameters are adjusted either in order to obtain a structure that is similar to the one that the model is supposed to study, or to trigger transitions from one structure to another, if one wants to study the evolution of a system. Jacques Ferber has suggested that it would be useful to inscribe the different phases of the development of an agent-based-model in different frames and to articulate these frames or "quadrants" in accordance with the distinction between individual and collective. The main steps in this process are the following:

1. Defining the functions that are internal to each individual agent (Individual internal quadrant).
2. Defining its interactions with environment - that requires transducing the effects of the functions of the individual agent into external observables (individual external quadrant).
3. Defining the observable effects of the collective aggregation of the behaviours of individual agents (collective-external quadrant).
4. If agents are cognitive ones, defining how they can use the observation of the collective behaviours in order to modify their own functioning (collective-internal quadrant).

One can apply these frames to the relations between local computational functioning and their aggregate effects under a common format.

The process that goes from 1 to 2, then to 4, and comes back to 1 is symmetrical with the process between iconic and symbolic genericity that we have previously considered. In this previous process, recharging particularity induced a symbolic re-categorization; in this new process, collective compatibility induces individual re-categorization. The two perspectives are different ways to deal with the problem of the context. Most computer scientists do not try to solve this problem by defining sub-contexts as sub-categories inside a general frame of formalization (a global perspective that we regard as static, because the higher level categories are not changed). They give a dynamical solution to the problem: The categories that belong only to a single step are not solutions to the problem. Only the dynamic of transformations from one structure of categorizations to another can give solutions - solutions that are only temporary and open to revisions.

Other proposals have been suggested in order to solve the problem of dynamical capacity. "Embodied" AI and cognition – grounded in the reality of the perceived environment and constrained by the conditions of action in real situation, which implies to solve problems of trade-off between taking time for exploring the situation and time of reaction to the situation - can be regarded as an attempt to solve it. In the embodied cognition perspective, interactions with environment trigger recharging particularity. Nevertheless, solutions to the problem of the contextual relevance have still to be elaborated. Context consists in the combination of the present task and the state and changes of the environment, including the changes induces by action in the environment as well as the changes that they induce in the agenda of the system. The constraints implied by the need to achieve the task in limited time and in a real environment may lead to simplify the problem, but these simplifications could be peculiar to local environments, and we would want them to be more general. Here again, AI seems to have to give up a static and generic perspective for a dynamical perspective that combines particularity and genericity, maybe by introducing revisability.

8 Conclusion

Although it seems to raise ultimate and somewhat "eternal" philosophical questions, the AI research program has known many significant changes in the last decades. In connection with these changes, the conception of human intelligence that the first AI was supposed to compete with has been modified as well. We were tempted to attribute to human intelligence the capacity of simultaneously activating different virtues, in a static perspective. But recent parallel or convergent developments of AI and of research programs in cognitive psychology and complex models simulations suggest that human intelligence cannot activate (and even possess) all these virtues at the same time. It probably has to let aside some virtues in order to activate others, and conversely. Activating these different virtues can only be done dynamically. Intelligence uses bootstrapping and recursivity, or, less formally, one uses one's know-how in order to improve one's action. But intelligence has also to use what stands up to it and raises problems, as it makes possible to detect on which point the intelligent behaviour was faulty. Intelligence implies modes of reacting to what the previous representations and operations could not control, and this requires to combine its initial capacities and incentives to variations suggested by new situations. The new project of AI is to analyse these intelligent dynamics by simulating them. AI and human intelligence appear to have more limited ambitions but more evolutional capacities.

420055_1_En_13_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:18/6/2019 Pages: 19 Layout: T1-Standard

References

Collins HM (1996) Embedded or embodied? a review of Hubert Dreyfus' what computers still can't do. Artif Intell 80(1–2):99–117

Cook SA (1971) The complexity of theorem-proving procedures. In: Proceedings of the third annual ACM symposium on theory of computing, pp 151–158

Dreyfus HL (1992) What computers still can't do - a critique of artificial reason. MIT Press, Cambridge

Dreyfus HL, Dreyfus SE (1988) Mind over machine - the power of human intuition and expertise in the era of the computer. Free Press, New York City

Ganascia J (1990) L'âme-machine: les enjeux de l'intelligence artificielle. Science ouverte, Éditions du Seuil

Garbay C, Kayser D (2011) Informatique et sciences cognitives: influences ou confluence? http://hal.archives-ouvertes.fr/hal-00713699

Girard JY (2001) Locus solum: from the rules of logic to the logic of rules. Math Struct Comput Sci 11(3):301–506

Hume D (1978) A Treatise of Human Nature. Oxford Clarendon Press, book I, Part III, section XIV, an edition with analytical index by Selby-Bigge, second edition with text revised by Nidditch

Kayser D, Magda FA, Stefan-Gheorghe P (2004) Intelligence artificielle et agents intelligents. Printech

Livet P (2002a) Emotions et rationalité morale. Presses Universitaires de France

Livet P (2002b) Révision des Croyances. Hermès

Livet P (2007) Towards an epistemology of multi-agent simulation in social sciences. The Bardwell Press, Cumnor, pp 169–194

Livet P, Mller JP, Phan D, Sanders L (2010) Ontology, a mediator for agent-based modeling in social science. J Artif Soc Soc Simul 13(1):3

Penrose R (1989) The emperor's new mind: concerning computers, minds, and the laws of physics. Oxford University Press, Oxford

Pitrat J (1995) De la machine à l'intelligence. Hermès

Proust J (2013) The philosophy of meta-cognition, mental agency and self-awareness. Oxford University Press, Oxford

Sallantin J, Szczeciniarz J (1999) Le concept de preuve à la lumière de l'intelligence artificielle. Nouvelle Encyclopédie Diderot, Presses Universitaires de France, http://books.google.fr/books?id=48PgPAAACAAJ

Sallantin J, Szczeciniarz JJ (2007) Il concetto di prova alla luce dell'intelligencia artificiale (post-face)

Saygin AP, Cicekli I, Akman V (2000) Turing test: 50 years later. Minds Mach 10(4):463–518. https://doi.org/10.1023/A:1011288000451

Searle JR (1980) Minds, brains and programs. Behav Brain Sci 3:417–457

Selinger EM, Crease RP (2002) Dreyfus on expertise: the limits of phenomenological analysis. Cont Philos Rev 35(3):245–279–279. https://doi.org/10.1023/A:1022643708111

Turing AM (1950) Computing machinery and intelligence. Mind 49:433–460

Varenne F (2009) Qu'est-ce que l'informatique ? Chemins Philosophiques, J. Vrin. http://books.google.fr/books?id=KTRPPgAACAAJ

Varenne F (2018) From models to simulations. Routledge. https://www.routledge.com/From-Models-to-Simulations/Varenne/p/book/9781138065215

Weizenbaum J (1983) Eliza - a computer program for the study of natural language communication between man and machine (reprint). Commun ACM 26(1):23–28

Weizenbaum J (1966) Eliza - a computer program for the study of natural language communication between man and machine. Commun ACM 9(1):36–45

Wright G (1971) Explanation and understanding. Cornell classics in philosophy. Cornell University Press, Ithaca

Zeigler B (1976) Theory of modeling and simulation, 1st edn. Wiley Interscience, New York