What ultimately matters, indeed?

Gordana Dodig-Crnkovic
Mälardalen University, Computer Science Laboratory
721 23 Västerås, Sweden
gordana.dodig-crnkovic@mdh.se

ABSTRACT

Computer science is winning the ground that was the traditional domain of physics. Answer to the question what ultimately matters nowadays belongs more to computer science than to physics. Search for answers to questions about truth, meaning, mind, subjectivity, consciousness etc. lies among others within AI. A brief skeptic argument against Penrose’s neo-vitalism is presented. The next question is if we can claim understanding the phenomena on the ground of their experimental reproducibility?

1 INTRODUCTION

The Paradigm of Science of the 20th-century was physics: relativity, quantum mechanics and finally, chaos. Physics was for a long period of time the ideal of scientific understanding of reality.

Questions in focus of previous century were:

- What is the (physical) Universe (microcosm, macrocosm)
- How is the Universe built up?
- What is the matter/energy, time, space?

Year 2000 we have answers to those questions that seem to fulfil our present needs. At the same time the efforts necessary to further improve knowledge within physics exceed by order of magnitude corresponding efforts needed to improve other basic scientific disciplines of interest. Therefore the paradigm of Science changes successively.

Parallel with growth of the body of physical theory there was the emergence of the "intentional sciences": disciplines that deal with symbols, references and interpretations, such as logic, cognitive science, psychology, neuroscience, parts of biology and computer science.

These new sciences are changing our conceptions of reality, and of relation between science and reality. Truth and meaning have been brought within the scope of science pertaining to the completely new context. Our views of realism and metaphysics are being modified. Scientists are starting to scrutinize fields of norms and values.

Traditionally, it is in philosophy or in the religious domain questions of most general significance have been asked, as e.g.: What ultimately matters?

In the last century expectations on science have grown enormously. Whether it knows it or not, science seems to be entering straight into the classically philosophical and theological territory taking the place of highest authority.

New questions in focus are:

- What is life (it’s laws, mechanisms, limitations)
- What is mind?
- What is meaning? (Issues related to meaning such as truth, proof, symbol manipulation (how symbols acquire meaning), “location” of meaning, form).

Computer science has many interesting new fields where new insights are reached that can contribute clarifying above ideas.

2 COMPUTER SCIENCE AND AI

Questions of relevance for computer science (as e.g. concepts of mind and meaning) have many practical consequences in the field of Artificial Intelligence, AI. Among others there is an interesting current controversy about Machines and Minds, see Ref. [1-4].

The questions simplified are as follows:

- Can machines be intelligent (think)?
- Can machine have self-consciousness?
- Can machine have a soul?

As usually in history of important controversies there are two confronting groups claiming opposite answers to those questions. That debate is in many ways instructive. First of all because it reveals our basic attitude to the question of what ultimately matters. Secondly, and at least equally interesting and illustrative is the argument itself.

There are a number of results of mathematical logic used to show that there are limitations to the powers of (discrete-state) machines.
The best known of these results is Gödel's theorem (1931) which shows that in any sufficiently powerful logical system statements can be formulated which can neither be proved nor disproved within the system, unless the system itself is inconsistent. It is established that there are limitations to the powers of any particular (discrete state) machine due to the Gödel's theorem. Yet it has only been stated without any sort of proof, that no such limitations apply to the human intellect.  

The same question can be put in different forms, the classic one being: Are human/thinking/feeling machines possible? Or, to cite Turing on Arguments from Various Disabilities: Will we ever be able to make machines to:

"be kind, resourceful, beautiful, friendly, have initiative, have a sense of humor, tell right from wrong, make mistakes, fall in love, enjoy strawberries and cream, make some one fall in love with it, learn from experience, use words properly, be the subject of its own thought, have as much diversity of behavior as a man, do something really new."…

The question is: does it necessary need to be one single machine? Do we need humanoid machines after all?

One can as well ask more pragmatic questions, as e.g.: Can a machine be made such as to:

- Pass Turing test
- Create an artifact that can be acknowledged as genuine by experts (compose music, write a sonnet…)
- Prove theorems/ check theorem proves through “mechanization” of reasoning (deduction)  
- Possess best knowledge within certain field and can act like expert system (medical expertise helping to set accurate diagnosis), etc.

There is namely a difference between the ambition of representing the common behavior (including knowledge) of the average person and the attempt to construct the machine able to compete with best of scientists, artists, philosophers etc. within their special fields.

Part of AI research’s objectives is to understand the computational principles underlying intelligence in man and machines and to develop methods for building computer-based systems to solve problems, to communicate with people, and to perceive and interact with the physical world (see Appendix).

3 PENROSE, CONSCIOUSNESS, CONSISTENCY, MINDS AND MACHINES

In his book Shadows of the Mind, Ref. [1], Roger Penrose presents arguments, based on Gödel's theorem, for the conclusion that human thought is uncomputable.

Penrose's First Argument

The best way to address Gödelian arguments against artificial intelligence is to ask: what would we expect, given the truth of Gödel's theorem, if our reasoning powers could be captured by some formal system F. One possibility is that F is essentially unsound, so that Gödel's theorem does not apply. But what if F is sound? Then we would expect that:

- F could not prove its Gödel sentence G(F);
- F could prove the conditional "If F is consistent, then G(F) is true";
- F could not prove that F is consistent.

If our reasoning powers are capturable by some sound formal system F, then we should expect that we will be unable to see that F is consistent. F is likely to be some extremely complex system, perhaps as complex as the human brain itself, and there is no reason to believe that we can determine the consistency of arbitrary formal system.

There does not seem to be anything especially paradoxical about this situation. Many arguments from Gödel's theorem offer us no reason to believe that we can see the truth of our own Gödel sentence, as we may be unable to see the consistency of the associated formal system.

Penrose's Second Argument

Unlike the previous argument, this argument does not depend on the claim that if we were a sound formal system F, we would be able to see that F is sound. Because of this, it is a more novel and interesting argument.

---

1 Ref. [7], A. Turing, "But I do not think this view can be dismissed quite so lightly. We too often give wrong answers to questions ourselves to be justified in being very pleased at such evidence of fallibility on the part of the machines. Further, our superiority can only be felt on such an occasion in relation to the one machine over which we have scored our petty triumph. There would be no question of triumphing simultaneously over all machines. In short, then, there might be men cleverer than any given machine, but then again there might be other machines cleverer again, and so on."

2 Ref. [6]

3 There is a lot of projects of automated deduction which result in the construction of “provers” (to automatically produce proofs) or “checkers” (to automatically check proofs produced by people).
The argument is developed in a roundabout way, but is summarized in the fantasy dialogue with a robot mathematician in 3.23. In a simplified form, the contention goes as follows:

- Assume my reasoning powers are captured by some formal system F ("I am F"). Consider the class of statements I can know to be true, given this assumption.
- Given that I know that I am F, I know that F is sound (as I know that I am sound). Indeed, I know that the larger system F' is sound, where F' is F supplemented by the further assumption "I am F". (Supplementing a sound system with a true statement yields a sound system.)
- So I know that G(F') is true, where this is the Gödel sentence of the system F'.
- But F' could not see that G(F') is true (by Gödel's theorem).
- By assumption, however, I am now effectively equivalent to F'. After all, I am F supplemented by the knowledge that I am F.
- This is a contradiction, so the initial assumption must be false, and F must not have captured my powers of reasoning after all.
- The conclusion generalizes: my reasoning powers cannot be captured by any formal system.

Penrose maintains that the abilities of human mathematicians to discover new mathematical truths cannot be explained on the basis of anything a machine could do. He argues in Ref. [1], p110, that this kind of thinking must be based on "insights that cannot be systematized - and, indeed, must lie outside any algorithmic action!"

So the appealing argument in favor of humans is to postulate some novel "basic" principle by which our minds are animated throughsome Vital Force or Essence we call Mind, or Consciousness, or Soul!

Nevertheless, since Harvey, Darwin and Pasteur, the idea of a Vital Force has nearly vanished from biology. Why is it still so much a part of attempts to explain human mind?

An entire generation of logical philosophers has thus wrongly tried to force their theories of mind to fit the rigid frames of formal logic.

In short, there is no basis for assuming that humans are consistent - not is there any basic obstacle to making machines use inconsistent forms of reasoning!

In any case, we have much the same problem with ourselves. Try asking a friend to describe what having consciousness is like. Most likely you hear only the usual patter about knowing oneself and being aware, of sensing one's place in the universe, and so on. Why is explaining consciousness so dreadfully hard?

I'll argue that this is something of an illusion, because consciousness is actually easier to describe than most other aspects of mind. Indeed, our problem is a far more general one, because our culture has not developed suitable tools for discussing and describing thinking in general. This leads to a kind of irony; it is widely agreed that there are "deep philosophical questions" about subjectivity, consciousness, meaning, etc. But people have even less to say about questions they would consider more simple:

- How do you know how to move your arm?
- How do you choose which words to say?
- How do you recognize what you see?
- How do you locate your memories?
- Why does Seeing feel different from Hearing?
- Why does Red look so different from Green?
- Why are emotions so hard to describe?
- What does "meaning" mean?
- How does reasoning work?
- How do we make generalizations?
- How do we get (produce) new ideas?
- How does Commonsense reasoning work?
- Why do we like pleasure more than pain?
- What are pain and pleasure, anyway?

In short: we humans do not possess much consciousness! That is, we have rather limited natural ability to sense what happens within and outside ourselves.

It seems to me that all of this stands upon a single and simple mistake. It overlooks the possibility of including systems that are mistaken about mathematics to some degree, or systems that can change their minds. By ruling such machines out, you've simply begged the question whether human mathematicians can be kinds of machines - because people do indeed change their minds, and can indeed be mistaken.
4 CONCLUSION

The researchers in Artificial Intelligence have discovered a wide variety of ways to make machine do pattern recognition, learning, problem solving, theorem proving, game-playing, induction and generalization, and language manipulation, to mention only a few. AI is a steadily growing field within computer science. As an illustration see Appendix, AI Subject index.

To be sure, no one of different AI programs seemed much like a mind, because each one was so specialized! But now we're beginning to understand that there may be no need to seek either any single magical "unified theory" or and single and hitherto unknown "fundamental principle". Thinking may instead be the product of many different mechanisms, competing as much as cooperating, and generally unperceived and unsuspected in the ordinary course of our everyday thought.

What has all this to do with consciousness? Well, consider what happened in biology. Before the 19th century there seemed to be no alternative to concept of "Vitality" - that is, the existence of some sort of life-force. There simply seemed no other way to explain all the things that animals do. But then, as scientists did their work, they gradually came to see no need for a "unified theory" of life. Each living thing performed many functions, but is slowly became clear that each of them had a reasonably separate explanation!

The next question is if we can claim understanding the phenomena only on account of their experimental reproducibility? Does the answer to question “how?” automatically means the answer to question “why?”. If we have the machine that can distinguish sweet from bitter, can we say we understand what “sweet” and “bitter” mean?

5 LITERATURE

1. Shadows of the Mind, Roger Penrose
2. The Emperor's New Mind, Roger Penrose
3. The Society of Mind, Marvin Minsky
4. The Mind’s I, Fantasies and Reflections on Self and Soul, D.R.Hofstadter & D.C.Dennett Ed.
5. Gödel, Escher, Bach: An Eternal Golden Braid, Douglas R.Hofstadter
6. Experiments in Musical Intelligence, D Cope, 1996
7. Computing Machinery and Intelligence. A. M. Turing, Mind 59, 1950. (Also reprinted in Ref [4]).