Logic and Philosophy of Computation for AI

Curtain raiser: A puzzle and many conundrums

The following puzzle was published by Hans Freudenthal in 1969 and popularized by Martin Gardner. It illustrates quite sophisticated reasoning in epistemic logic using background knowledge of arithmetic.

x and y are positive integers such that $1 < x < y$ such that $x + y < 100$. Ms P and Mr S are mathematicians (and perfect logicians) such that S is told the value of $x + y$ and P is told $x \cdot y$ (the product of x and y). That S is told the sum is known to P, that P is informed of the product is known to S, and so on. Then the following conversation takes place:

- Ms P: *I do not know what x and y are.*
- Mr S: *I knew that you didn't.*
- Ms P: *Now I know x and y.*
- Mr S: *So do I.*

We are told to find x and y.

Do enjoy trying to solve the puzzle. Clearly the first statement of P tells us that both x and y cannot be primes. If they were, given the product, she would have uniquely factorized them and would know x and y. So the "information content" of the first statement is that one of them is a composite number.

But I am not presenting this puzzle to ask you to solve it. Here are some questions for you to think about:

1. P and S are assumed to be *perfect logicians*: what does this mean? Do perfect logicians exist? What are the implications of making such an assumption?

2. Can P and S be computer programs? Do you think it is possible, in principle, for two programs to have such a conversation?

3. The statements refer to each person's knowledge and also mutual knowledge. There is so much certainty in a statement like *I knew that you didn't.* Is this justifiable? Could computer programs reason about the knowledge of other programs in a similar manner?

4. To have a conversation, we need to have some intentions and beliefs, but also linguistic capacities and cognitive capabilities: to decide what to say, to listen and make sense of what the other says, and relate to what was said before, remove irrelevant inferences, and so on. Can a computer converse with a human being in this manner? Can two computers converse with each other in this manner?

5. Clearly a lot of thought is going on between statements in the puzzle above. But how do we certify that someone is thinking? Can you observe your own thought? If yes, how, and what is this process of observation? Can computer programs think (in this sense)? If yes, can we observe their thinking? Can we certify that they are thinking?

6. The puzzle has a natural sequencing, since changing the order of utterances makes them nonsensical. This suggests a flow of time inherent in such inference (in the sense that inference X cannot be carried out before inference Y). Is this merely incidental, or is it essential? Does this impose any limitations on computing agents?

Please think about these questions. These are the kind of issues we wish to discuss in Part III of our course on **Logic and AI**. When we attribute intelligent behaviour to computing agents, many philosophical questions arise on the nature of algorithms, programs and computers.

If you would like to write some answers to these questions, I would be very happy to read them: please send your response to jam@imsc.res,in before November 10, 2020 (so that I can discuss them during the first lecture on November 12, 2020).