Sam S. Rakover

**AI and Consciousness**

### Generally, we do not worry about the fact that we are neither the strongest nor the fastest species in the world. But when it comes to sophisticated computers (robots), our attitude changes completely. All of a sudden we are afraid of these apparatuses and attribute evil intentions to them, as portrayed in certain science fiction movies (e.g., [Terminator 2: Judgment Day](https://www.imdb.com/title/tt0103064/?ref_=ttls_li_tt) starring Arnold Schwarzenegger). Why? Here are the main reasons. First, while we admit that we are not as strong or as fast as other animals, we insist that we are the most intelligent. We believe that what distinguishes humans from other creatures is our cognitive abilities. Second, the computer has many traits that are similar to our cognitive capabilities: it can manipulate physical symbols and we refer to this as information processing. Furthermore, a whole new branch of science, cognitive psychology, has been established on the analogy between the way a computer functions and our mental processes. Given these three points, concerns emerged that the machine would rise up against its maker, that sophisticated robots would defeat man and conquer humanity. To counteract this hypothesis, there were attempts to highlight the differences between the computer and the human being. For example, it was said that computers could not beat humans at chess or become Go champions, nor could they paint, write stories, create art, do scientific research, and so on (e.g., du Sautoy, 2019; Simon, 2022). But as it turns out, AI programs have indeed achieved all of the above. In the light of this, Simon has proposed that AI will become “the scientist of the scientist”. I disagree.

### There is a huge difference between a human being and an AI program (or a highly sophisticated computer, or robot) and it does not depend on the question of what a human being can do that a computer cannot. Rather, the difference is based on a particular quality that a human possesses and a computer lacks – consciousness. Rakover (2021) proposes that consciousness is a necessary condition for understanding. In other words, he suggests that without being in a state of consciousness, one cannot even understand what one is doing. It follows that an AI program (or robot) does not understand its own performance and output, even though a human may evaluate an AI program as highly intelligent and creative. Schematically, the following chain of events describes the relationship between a human and a computer:

### REALITY🡪 H 🡪 Input [Computer]🡪 Output interpreted by H🡪 REALITY

### This scheme proposes that on the basis of a given REALITY, a human being (H) enters information (data) into a computer (invented by humans) which is known as input. The computer transforms the input into output that is interpreted by H and compared to REALITY. Thus, without the involvement of a conscious human being, the above chain of events would be impossible. If this approach is correct, then one should not be worried by an AI program (computer). It is under human control. Nevertheless, one crucial question arises here: Could a highly sophisticated computer develop consciousness? I would give a negative answer. To the best of my knowledge, so far no AI program (computer) has developed consciousness. In what follows I will present two important reasons for this.

### *The Chinese room.* Searle’s thought experiment is very famous, so I will describe it briefly. A very sophisticated computer was able to answer any question in Chinese. Searle, who did not understand Chinese, took it upon himself to perform the same operations as the computer (guided by instructions in English) and like the computer, his answers were judged to be very intelligent. Did Searle understand Chinese? His answer was no. Manipulating the Chinese symbols did not endow Searle with an understanding of Chinese. This thought experiment emphasizes two major points. First, computer processes do not create consciousness. Second, the Chinese room highlights the problem with the Turing test: although one may perceive Searle’s answers as indicating that he understands Chinese, Searle himself says that he does not.

### *The Consciousness-Counter*. I have developed the following thought experiment. A group of scientists propose a new theory and a device called the Consciousness-Counter. On the one hand, this device measures any kind of consciousness in objective units of measurement (OUMs), and on the other hand, it can measure in ‘OUMs’ any physical or chemical property. This device was applied in two cases. First, when Mrs. Smith from New York, a lover of Renaissance art, saw the Mona Lisa, her level of ‘art excitement’ was measured by the Consciousness-Counter and it equaled 187 OUM. Secondly, in an art survey, it was found that an environmental sculpture installed in Paris, made of objects that had been discarded and retrieved from the municipal garbage dump, irradiated exactly 187 OUM. Therefore, it was hypothesized that Mrs. Smith’s impression of the Mona Lisa would equal her impression of the environmental sculpture, 187 OUM precisely. However, when Mrs. Smith was flown to Paris and shown the sculpture, the Consciousness-Counter recorded -273 OMU. She detested the sculpture. How can we explain this surprising result? One possible explanation is that the theory and the Consciousness-Counter depended on a methodology developed for investigating the external, natural world and not the inner world – consciousness.

### If you are convinced that no AI program has ever developed consciousness, then there is nothing to be worried about (Schwarzenegger’s movie is not going to come true). Thus, we only have to learn how to appropriately use AI programs for our benefit.

### References

### du Sautoy, M. (2019). *The creativity code: Art and innovation in the age of AI*.

### Harvard University Press.

### Rakover, S. S. (2021). The two factor theory of understanding (TFTU):

### Consciousness and procedures. *Journal of Mind and Behavior*, 42, 347-370.

### Simon, T. (2022). The scientist of the scientist. *AI & Society.*

### https//doi.org/10.1007/s00146-022-015446