

SEMANTICS, GÖDEL'S THEOREM, AND NEURAL NETWORKS

Yu-Houng Houng

ABSTRACT

John Searle's Chinese Room Argument and Gödel's Incompleteness Theorem has been taken as fatal attacks to the traditional digital computer models of the mind. The main thrust of Searle's argument is that a programmed computer is purely a syntactic machine which is not sufficient for modeling the semantic aspect of the mind. Gödel's theorem has been taken, on the other hand, to show that a programmed computer is necessarily inferior to human mind for it in principle cannot "see" the truth of its Gödel sentence while human beings can.

Searle further proposes that an extended version of his argument, i.e., Chinese Gym Argument, shows the similar consequence for neural network models. Roger Penrose rejuvenates the argument from Gödel's theorem and intends to show that human mind is non-algorithmic, hence no Turing Machine approach can succeed in modeling the mind.

In this paper, I discuss the relations between neural network models and Searle's argument and Penrose's argument. I contend that Searle might be right in refuting digital computer models, but he totally misunderstands the nature of neural network models. His Chinese Gym Argument cannot be applied to neural networks. With regard to Penrose's argument, I argue that Gödel's theorem does not have the philosophical implications with which he think can defeat Turing Machine models, not to mention neural network models. Although I don't agree with Penrose with respect to the above issue, he nevertheless is correct in pointing out that human mind might be non-algorithmic, and hence no algorithmic models can suffice in modeling the mind.