ON THE POSSIBLE OUTPUTS OF HUMAN BRAINS AND INTELLIGENT MACHINES

Andrew C. Papanicolaou

AI machines (e.g. robots) can become conscious

- **Robots could become autonomous and even revolt against humanity**

Two categories of phenomena: **Behaviors--Experiences**

Features of behaviors

Electrochemical signals among brain cells belong to the category of behaviors as well.

Conscious experiences: the second category of phenomena

Why are experiences categorically different from behaviors (including brain signals)

Varieties of experience: sensations, feelings, sentiments, thoughts, aspirations, intensions….

Why present-day AI systems are not conscious

Could future AI system that pass the Turing test be conscious?

Objections to that claim: the "Chinese room" experiment

The "Chinese experiment" proved that passing the Turing test does not imply that the agent that passes it is conscious

The experiment did NOT prove that no robot can be conscious

Leibniz: no complex pattern of signals, human or artificial, suffices for consciousness

The scientific "emergence" hypothesis

A third experiment

The proof of the proposition "no robot can be conscious"

The only kind of consciousness we know emerges from patterns of biological signals

Objections to the proof and why they are baseless

On the possibility of revolt of even unconscious robots