# THE TRIVIAL NOTIONS SEMINAR

## Sam Marks

will speak on

# Decision theory

**or: what to expect when your expectations depend nontrivially on what you're expecting**

ABSTRACT

A decision algorithm is a procedure that takes in a list of possible actions (and information about the state of the world) and outputs a chosen action. Determining what a decision algorithm will do is easy in many cases, but can become quite complicated when important facts about the current state of the world depend on facts about the decision algorithm. For instance, consider the following algorithm L for playing the Prisoner's Dilemma against an adversary A. L searches for a proof that A will cooperate against L; if it finds a proof, then L cooperates. If it finds no such proof (within a fixed large amount of time), then L defects. Question: what will L do when playing Prisoner's Dilemma against another copy of L? The goal of this talk is to give an introduction to decision theory and discuss some interesting examples of situations like this one, which arise when a decision algorithm needs to have a working model of itself.

## Thursday, October 14, 2021

## at 11:30 am

## Jefferson Tent, North Lawn