A scalable approach to causal responsibility

Sjur Dyrkolbotn

When allocating responsibility for events involving many agents, we require a theory of causality that is scalable. Existing counterfactual theories fail in this regard, since the number of different counterfactual scenarios we must consider tends to grow exponentially in the number of potential causes in the model of the event in question. Importantly, this explosion in complexity will take place even if our knowledge of the phenomenon – as encoded by the model – rules out most of the counterfactual scenarios we are forced to consider. In this talk, I sketch a theory of causality that can be used to address this problem. It is based on the idea that we can identify causes by applying the but-for test at counterfactual states that are regarded by the model as possible states of the world. The challenge that arises is how to recognise cases of pre-emption without considering further interventions that violate the constraints set by the model. I propose a fixed-point construction to solve this problem and argue that it is captures the phenomenon of pre-emption as well as existing definitions, while avoiding conceptual anomalies associated with quantifying over impossible states.