Attributions of Causation and Prevention: A test of Counterfactual and Generative Theories

Clare Walsh, ***University of Plymouth***

&

Steven Sloman, ***Brown University***

How do people understand questions about causation? Counterfactual theories propose that causation can be defined in terms of a counterfactual conditional. In its original formulation, it proposes that "event c caused event e" provided that if c hadn't occurred then e wouldn't have occurred in the closest world to our own in which c occurred. In contrast, generative theories define causation in terms of the process through which the effect occurs. Causation involves a transmission along a causal pathway and may involve the exchange of some conserved physical quantity, such as energy or momentum. These theories can also be used to define prevention. Counterfactual theory may be extended to so that "C prevents E" if it is the case that if C hadn't occurred E would have occurred". Prevention is problematic for generative theories because if A prevents B, then B does not occur and hence there is no continuous process connecting the preventor and effect. A possible fix is to assume that prevention is qualitatively different from causation. Dowe (2000) proposed that A prevented B if there was a causal interaction between A and another process x and if A hadn't occurred then x would have caused B. Hence according to this view, causation and prevention have quite different meanings and the definition of prevention involves a counterfactual.

Our aim is to test whether people's judgments of causation conform to these alternative definitions. In most situations involving causation, counterfactual and generative theories make the same predictions. Usually there is a continuous causal process linking the action and the outcome and a change to the outcome would lead to a change in the effect. One counterexample to this occurs when an outcome is over-determined. In these cases, there is a causal process but a change to the cause may not bring about a change to the effect though counterfactual theories have been modified to cope with this problem. We aim to test an alternative situation where counterfactual and

generative theories make different predictions. We compare cases that involve a complete causal mechanism linking the action to the outcome (i.e., a clear process of generative transmission) to ones where the action interrupts a mechanism and hence there is no causal mechanism linking the action to the outcome. But in both cases the counterfactual alternatives are identical – a change in the action brings about a change in the outcome. Hence, generative and counterfactual theories make different predictions. If causal attribution depends on understanding how the outcome comes about, i.e., on the specific process generating the outcome, people should make different judgments for the two scenarios. In contrast, if causal attribution depends on the simulation of counterfactual alternatives, people should respond in the same way to the two scenarios.

Our results show that people attribute causation more often to an action that is linked by a continuous mechanism to an outcome than to one which removes an obstacle blocking an outcome (Experiment 1). People attribute prevention more often to an event that interrupts a mechanism than to the omission of an event (Experiment 2). Finally, people judge "C prevents E" to mean something different from "C causes not E" (Experiment 3). We discuss the implications for counterfactual and generative theories of causation.