Counterfactual Reasoning[*]

Diderik Batens
Centre for Logic and Philosophy of Science
Ghent University, Belgium
diderik.batens@ugent.be

Logically exact definitions of a consequence relation that characterize counterfactuals have been offered more than forty years ago, for example by Nicholas Rescher. Such consequence relations aim at delineating the consequence set of a specific counterfactual hypothesis with respect to a given body of knowledge, which comprises both factual and nomological statements. Articulating consequence relations of the kind should be clearly distinguished from spelling out the logic of the counterfactual implication (as done, for example, by David Lewis).

Needless to say, such consequence relations are defeasible. Moreover, their definitions hide several difficulties. One of them concerns (what Rescher calls) the involved modal categories. A different and harder difficulty concerns the computational aspects: by which reasoning forms are we able to find out whether something is a consequence of a counterfactual hypothesis (in a given setting)? This problem is harder because (i) all realistic examples concern at least predicative languages, and (ii) the defined consequence relation even lacks a positive test (is not semi-recursive).

It will be shown that the first difficulty can be solved by employing a slightly enriched (but still very simple) language, whereas the second is resolved by articulating the adaptive logic that characterizes the consequence relation. The latter move does not resolve the involved computational complexity—no admissible move can do so—but it enables one to understand the way in which humans are nevertheless able to arrive at sensible decisions on the matter.

The proposed solutions will be outlined step by step. Moreover, it will be shown that there is ample room for variation without changing the general framework. Some related topics will be discussed if time permits.