

Abstract

Deriving the Formula of Humanity

In the *Groundwork of the Metaphysics of Morals*, Section II, Kant tries to establish that if there is a supreme principle of morality, then it is (or is equivalent to) the Formula of Humanity. He offers a “derivation” of this principle. This paper focuses on the derivation’s initial steps.

Granting as the derivation contends that a categorical imperative would require there to be something unconditionally good, why must this unconditionally good thing be humanity, rather than something else? The argument Kant appears to invoke in response to this question, namely an argument by elimination, is very disappointing. But Christine Korsgaard and, later, Allen Wood have claimed that this argument does not represent Kant’s best effort at a derivation. Kant, they contend, actually undertakes a “regressive” argument. I question this claim. I then try to bolster Kant’s argument by elimination with the help of material he presents near the end of *Groundwork* II. Kant points out that rational nature is that which can possess a good will. Since it is, he suggests, rational nature alone can be the unconditionally good “ground” of a categorical imperative. For all other candidates are such that if they had this status, the good will would be “subordinate” to them. But, claims Kant, the good will cannot without contradiction be subordinated to any other object. Suitably interpreted, this reasoning strengthens Kant’s argument by elimination, I contend.