The Moral Principle of Action–Motivation

Mark Tracy

March 6, 2026

1 Introduction of the Principle

Immanuel Kant's Categorical Imperative is to act according to that maxim which you can simultaneously will to be a universal law. However, I think that there are no maxims regarding actions alone that may be coherently willed universally, because one can easily contrive a situation where any action is permissible to prevent a greater evil. Instead, there are combinations of actions and sets of motivations that may coherently be willed to be prohibited or permissible. I propose the following augmentation of Kant's formalization:

Act according to the tuple of action and motivation set which you can simultaneously will to be universally permissible.

2 Explication of Terms

Let us formalize the terms used in this principle. We consider an agent deliberating over actions.

- Let A be the set of possible actions available to the agent.
- Let B be the set of equivalence classes of statements of beliefs of the agent (modulo synonymous phrasing). We denote statements using double quotation marks, e.g., "This is a statement."
- Let C be the set of *relevant anticipated states of affairs*, i.e. anticipated states of affairs that the agent believes to be made more likely by one possible action than by another, i.e.

$$c \in C \quad \Longleftrightarrow \quad \exists a, a' \in A, \exists b \in B : \quad \text{``}P(c|a) > P(c|a')\text{''} \in b.$$

  Note that we assume that the agent considers a prior probability distribution over anticipated states of affairs that may be conditioned on actions, and note that the statement, "$P(c|a) > P(c|a')$," reflects the agent's belief. This set therefore captures the states of affairs that are at issue to the agent in the present decision.

- Let $d : A \to \{0,1\}$ be a one-hot indicator function that signals the action decided upon by the agent, i.e. $d(a) = 1$ if the agent decides to take action a, and $d(a) = 0$ otherwise.
- Let $e : A \to \mathcal{P}(C)$, where $\mathcal{P}$ denotes the power set, associate an action a with the subset of anticipated states of affairs that are relevant with respect to a, i.e.

$$e(a) = \{\, c \in C \mid \exists b \in B, \exists a' \in A : \text{``}P(c|a) \neq P(c|a')\text{''} \in b \,\}.$$

- Let the *motivation set* M_a of an action a be defined as the family of minimal subsets of $e(a)$ that, if the agent believed them irrelevant, then action a would surely not be chosen, i.e.:

$$\begin{aligned} M_a \;=\; & \{m \subseteq e(a) \mid \exists b \in B : \text{``}e(a) \cap m = \emptyset\text{''} \in b \implies d(a) = 0 \text{ and} \\ & \emptyset \neq m' \subset m \implies m' \notin M_a\} \end{aligned}$$

The first condition above, $\exists b \in B :$ "$e(a) \cap m = \emptyset$" $\in b \implies d(a) = 0$, means that a set m of states of affairs is in the motivation set of the agent with respect to action a if it is the case that if the agent were to believe (mistakenly or hypothetically) that the anticipated states of affairs represented by m were all not relevant with respect to a, then the agent would surely not take action a. The second condition, $\emptyset \neq m' \subset m \implies m' \notin M_a$, stipulates that no nonempty proper subset of such a set m is also contained in the motivation set. This enforces minimality of the included sets.

Illustration (conjunctive): Suppose Carl is choosing between staying at their current job or leaving it to find another. Then $A = \{\text{stay}, \text{change}\}$. Suppose that changing jobs offers a better salary and a shorter commute but requires more hours. Suppose further that if, hypothetically, a better salary and a shorter commute were *both* irrelevant to the decision (because, for example, neither action would change either of these figures), then Carl would surely not change jobs; but if either remains relevant, then he would be willing to change jobs. Then:

$$\{\{\text{better salary}, \text{shorter commute}\}\} \subseteq M_{\text{change}}$$

Illustration (disjunctive): Now suppose that if *either* a better salary or a shorter commute were not relevant (because, for example, neither action would change the figure), then Carl would surely not change jobs. Then:

$$\{\{\text{better salary}\}, \{\text{shorter commute}\}\} \subseteq M_{\text{change}}$$

Note that without the second condition specified above (enforcing minimality of elements of M_a), we would also have $\{\text{better salary}, \text{shorter commute}\} \in M_{\text{change}}$. The condition avoids the redundancy of such supersets, which is in general combinatorially explosive.

- For a given decision-making event, and for the action a for which $d(a) = 1$, the corresponding (a, M_a) represents the tuple of action and motivation set, as mentioned in the Rule.

3 Advantages of this Formulation

This Rule allows one to judge the morality of an action *both* by the nature of the action itself *and* by what consequences the decision-making agent believes it is making more or less likely by taking the action. In particular, it stipulates that the tuple of action and motivation set must be universally permissible.

4 Examples of Universalizable Maxims

- Do not lie for the purpose of attaining material personal benefit.
- Do not commit violence for the purpose of attaining material personal benefit.
- Seek out perspectives different from your own for the purpose of better understanding the consequences of your decisions.