

Reasoning and mental models in ethics

Author: Yanick Farmer

To cite this paper: Farmer, Y. (2026). *Reasoning and mental models in ethics*. Archipel. [Hyperlink](#)

Ethical reasoning

Ethical reasoning uses the same general processes as other forms of reasoning. It is its subject matter that differs in certain respects. Ethical problems revolve around certain fundamental dimensions, such as freedom (or self-determination), the good, concern for others, and justice. Certain theories provide a framework for reasoning about these issues. This is the case with classical theories of normative ethics, such as utilitarianism, deontology, and virtue ethics. Other theories have emerged more recently to expand upon these approaches with regard to specific issues.

Ethical reasoning is partly *descriptive*, since any judgment must be based on an adequate analysis of the facts. However, it is also *evaluative*, insofar as it is fundamentally concerned with determining what is right or wrong in the course of an action. The descriptive aspect relies on mental models similar to those used in other forms of reasoning. The evaluative aspect relies in part on the integration of the brain's affective circuitry and its derivatives—namely, affects, emotions, values, principles, and norms. It can be exercised through fast thinking (e.g., intuition) or through “slow” (more elaborate) reasoning. Part of what makes ethical reasoning unique stems from the fact that human beings live in society, and this situation compels them to adhere to social norms to which they must align their actions. In society, there are various types of norms that define what is permissible and what is not. These norms can be legal, deontological, moral, or, more strictly speaking, ethical. In some works, the reasoning that arises from situating a human subject within a system of norms governing, among other things, obligations, prohibitions, permissions, or duties is referred to as *deontic*. The evaluative aspect of ethical reasoning requires adding annotations to mental models of reasoning concerning abstract concepts associated with values, principles, norms, as well as the rules that guide judgment and action.

Some mental models for ethical reasoning

Given the diversity of ethical issues, it is not possible to provide an exhaustive overview of the mental models that can be used to address them. The few examples that follow focus on fundamental concepts in ethics and are original proposals that examine these concepts from the perspective of their underlying structures. These examples address the concepts of the *good*, the *common good*, *self-determination*, and *justice*.

Mental model of the good

Structurally, the good must be understood in terms of the basic biological mechanisms through which the ethical subject seeks to satisfy fundamental needs that revolve around interactions with the environment. To contribute to the satisfaction of these needs, ethical reasoning proceeds by making choices based on sets of basic data that constitute the architecture of rational choice regarding the good. In ethics, the good results from *behavior* that draws upon sets of possible actions (a) linked to *positive* internal states (s) and rewards (r). Indeed, given a set of states S of the ethical agent in the environment, a set of actions A that the agent can perform, and a set of scalar values (rewards R) that the agent can obtain, optimal subsets of actions (a) that maximize the states (s) associated with a positive reward (r) will determine the behavior (which aims at the good). In other words, reasoning about the good means choosing, from among several possible actions, those that produce the most positive internal state possible based on a *personal vision* of what constitutes a positive reward. Value serves as a *weighting factor* in choosing an action to obtain a reward or simply to act upon the environment. It enables a transition from state s^1 to state s^2 through action a.

Ethical problems or conflicts between two subjects, A and B, often arise during an interaction aimed at reaching an agreement of will or intention. When the behaviors of the two subjects are considered in terms of two completely *disjoint* sets of actions in relation to a set of stimuli (I) originating from the environment, then a divergence occurs and agreement of will proves impossible or problematic. In fact, agreement is possible only if there are actions that satisfy the interests of *both parties*—that is, actions that lie at the intersection of the sets of possible actions for A and B.

From this perspective, the good can be defined as an *explicit or implicit rule for constituting (or demarcating) the optimal subset of actions created by the subset of positive states or rewards*.

The ethical agent establishes a hierarchy of personal values based on their conception of the good. The good thus appears as a *metarule* for ordering value-based choices.

Mental model of the common good

The common good can also be understood as a principle arising from the *combination* of subsets of actions carried out by agents who collaborate in order to act collectively on the environment and thereby receive positive outcomes or rewards from it. As is the case with the good at the individual level, the common good underlies the values and norms that govern groups of individuals. In other words, the rule of the common good is the one according to which *collective* values and norms are prioritized. It is even the foundation upon which the justice system governing a society is organized. This close link between the common good and justice will be explained in more detail below.

Mental model of self-determination

Self-determination is made possible when there is a conjunction between a consciously formulated command (or algorithm) and a result. If we return to the basic structure used earlier, the problem of self-determination translates into a *twofold conjunction*: 1) between the command and the action; 2) between the action and the reward. The first conjunction establishes *agency*, while the second establishes *effectiveness*. Indeed, when an intention is not fulfilled, there is no sense of freedom; and a reward that is not the result of an intentional action can lead to a feeling of alienation. The subject's command directed toward action is organized in various areas of the brain, particularly in the motor cortex. It involves *selecting* the appropriate command from a repertoire of possible responses to stimuli, planning the action, and *executing* it. The neural state corresponding to a sense of freedom and self-determination is complex and depends on the interaction of the areas involved in the generation of voluntary action. A debate persists in neuroscience regarding the nature of voluntary action. Some view it as initiating the action, while others believe it is more of a "veto" exercised by conscious will over a response that is prepared somewhat earlier in the process.

The mental model of self-determination divides the concept into two aspects: agency and effectiveness. Agency is absolutely necessary to characterize an action as self-determined or free. The proposed general model can be supplemented by submodels that address issues related to agency. For example, does an uninformed person possess the necessary repertoire to select an

appropriate response to a given problem? The answer is probably no. Consequently, agency depends on considerations of what makes it possible. Submodels can be developed to further explore the question of agency.

Effectiveness is not absolutely necessary for self-determination, insofar as it depends in part on external contingencies that the individual does not always control. Even if a person is highly motivated and self-determined, it often happens that an action fails and does not result in a reward. This does not make the person any less free. That said, the question of effectiveness remains central to the experience of freedom. The repeated success of an action contributes to the enrichment of the repertoire of responses stored in memory and nurtures the subject's *self-confidence*. A subject paralyzed by fear is not free; but a confident subject is much more so. Submodels addressing the issue of effectiveness can also be developed to contribute to the broader reflection on self-determination and freedom.

Mental model of justice

Etymologically, the concept of justice refers to a sense of *fairness*. However, fairness requires determining what is due, to whom, and in what proportions. To understand the structural nature of justice, we must first identify all of its functions, and then determine what makes their fulfillment possible. The first aspect of the concept of justice involves identifying the nature of the *imbalance* that it must correct or prevent. It is important to understand that justice emerged in human societies to maintain their cohesion and harmony through a balance of “rights.” Groups that lack cohesion—due in particular to internal injustices—are more vulnerable to external threats. The second aspect concerns the *domains of assignment*. What exactly does justice allocate? The third aspect relates to the “fair” *proportion* of what is allocated. What are the criteria for this fair allocation? Finally, the fourth aspect concerns the *characteristics* of an effective “system” for administering justice—that is, a system that inspires trust and satisfaction among the individuals to whom it applies.

Generally, justice serves either to impose “punishments” or to distribute goods in a certain proportion that must be deemed equitable. The imbalances it corrects thus relate either to inequalities among individuals caused by the offense or to those resulting from shortages caused by unmet needs. In the first case, we speak of *punitive* or *retributive* justice, whereas in the second case, we speak of *distributive* justice. Once the nature of the imbalance has been identified, justice

must determine the proportion of what it allocates to redress the injustice. This may take the form of a sanction in the context of punitive justice, or an allocation in the context of distributive justice.

To determine a “fair” proportion, it is necessary to rely on criteria. The criteria that enable this type of judgment take the form of values or norms. Since these values or norms must be *shared* by groups of individuals governed by the justice system, they belong to the sphere of the common good specific to that group. The common good influences collective judgments of justice just as the good influences individual judgments: it provides a set of constructs (values, norms) that serve as the ultimate criteria (metarules) for judgment. Justice cannot be realized without a vision of the common good. Because it is rooted in the affectivity encompassed by values and norms, the common good “precedes” justice in the sense that it is its primary condition. Even when the formal structure of a justice system is rigorous and highly developed, the judgments it produces are inextricably linked to the shared emotional fabric of a society. What is shocking, what is revolting, and what provokes outrage all reflect the collective emotional fabric that underpins a society’s vision of the common good.

Defining the characteristics of an effective system falls within the realm of procedural justice. Empirical research has shown that certain characteristics *foster trust* in the justice system. Six of these are particularly important: 1) uniformity; 2) impartiality; 3) quality of information; 4) reviewability; 5) representativeness; 6) values of the common good. Uniformity means that a rule of law must be applied equally to everyone. “Double standards” must be avoided. Impartiality implies that the body evaluating and adjudicating a situation to achieve justice must be neutral and unbiased. The quality of information refers to the evidence that must be gathered to adjudicate a situation. For the judgment to be fair, it is essential that the information gathered be accurate, complete, and transparent. Reviewability paves the way for a possible review of a decision that might be perceived as unfair. Representativeness is the opportunity for a person who is being judged to be heard and even to seek someone to defend their rights (a lawyer or a mediator, for example). The values of the common good are those that a society deems necessary to preserve in order to ensure its cohesion.

Ethical reasoning using the mental model of justice proceeds as follows:

Cause of the imbalance : wrongdoings, needs?



Domain of assignment : punitive or distributive justice



Proportionality (punishment, goods)



Values or norms of the common good (criteria for assessing the proper proportion)

+

Procedural justice (criteria for the acceptability of justice procedures)

=

Overall goal : harmony through a balance of rights

Submodels can be designed to address various issues related to justice. In criminal justice, some models focus on the nature of the crime. Others focus on punishment and rehabilitation. In political philosophy, John Rawls's model focuses primarily on ways to allocate the fair share, particularly using conceptual tools such as the veil of ignorance or the maximin. Amartya Sen addresses the same concern through the concept of capability. During the COVID-19 pandemic, several scientific publications on distributive justice (regarding medical resources) examined the values or norms of the common good that authorities could rely on to ration the use of resources (clinical prognosis, age, social utility, intergenerational equity, etc.).