

MIT Press • Open Encyclopedia of Cognitive Science

Morality

H. Clark Barrett¹

¹University of California, Los Angeles

MIT Press

Published on: Jul 01, 2026

DOI: <https://doi.org/10.21428/e2759450.74a15520>

License: [Creative Commons Attribution-NonCommercial 4.0 International License \(CC-BY-NC 4.0\)](#).

Morality concerns what is—or is considered to be—right or wrong, good or bad, just or unjust, proper or improper, and virtuous or vicious. It typically refers to these evaluative aspects of human actions, dispositions, and relationships. However, although many theoretical frameworks for morality exist, it has proven difficult to define. The term morality can be used in at least two distinct senses. It can be used *normatively*, to refer to what actually is good or bad, or *descriptively*, to refer to what people judge or consider to be good or bad and how this influences people’s lives and actions. It is in the latter descriptive sense that morality has typically been studied in cognitive science. Empirical studies of people’s moral intuitions and judgments aim to uncover the underlying structure of the human moral sense and how it varies across individuals, societies, and history. Because naturalistic approaches to morality take what is right and wrong to be a product of human minds and societies, the descriptive study of human moral judgment and behavior can shed light on normative aspects of morality as well.

History

Historically, there are several major normative approaches to morality, each with associated theorists ([Skorupski, 2010](#)). *Virtue theory* or *virtue ethics* seeks to identify the principles or virtues that lead to a good life, or to “the good” more generally, including aspects of moral character and personal and community relationships. *Natural law* (the Stoics, Aquinas) attempts to ground moral life in human nature and the natural order, holding that right and wrong depend on facts about human needs and desires. *Deontology* (Kant) proposes that morality involves duties to follow rules and principles and that the moral character of an act depends on these and not the outcomes. This framework can be contrasted with *consequentialism* and its variant *utilitarianism* (Bentham), which judge the morality of an act based on the outcomes or effects it produces in the world.

Each of these approaches still exist today and have influenced the cognitive science of morality. The natural law framework resembles some evolutionary approaches to morality, which see moral values as grounded in human nature—for example, that cooperation is good and selfishness is bad or that people have a right to food, shelter, and family ([Nichols, 2004](#); [Tomasello, 2016](#)). The deontology/consequentialism distinction has informed experimental work in cognitive science that seeks to ask whether and when people’s moral judgments hinge on the reasons for or outcomes of actions ([Greene, 2014](#)). In anthropology, frameworks including virtue ethics, relational ethics, and ethics of care have increasingly been used to understand the ethics of everyday life around the world ([Mattingly & Throop, 2018](#); [Zigon, 2008](#)).

Although some scholars distinguish between *morality* as the individual sense of right and wrong and *ethics* as the standards of a community, others contest the distinction ([Mattingly & Throop, 2018](#)). Indeed, whether morality is fundamentally a product of individual psychology or of social interaction is one of the key questions in the cognitive science of morality, suggesting that there is no clear line between morality and ethics and perhaps no difference at all.

Descriptive approaches to morality (characterizing human judgments of right and wrong) in the different fields of cognitive science, from anthropology to psychology, have typically mirrored theoretical and methodological developments in each field. An influential early approach in anthropology and sociology was Émile Durkheim's framework, in which morality was simply the enactment of a society's social rules ([Durkheim, 1925](#)). His descriptive framework was culturally relativist and aligned with much of sociocultural anthropology today. At the same time, Durkheim's approach saw social rules as functioning to maintain social cohesion, reminiscent of current evolutionary approaches to morality. Today, the question of moral variation is still a major topic of anthropological research studied using ethnographic methods and, increasingly, cross-cultural experiments.

In psychology, the descriptive study of moral judgment began largely in developmental and social psychology, later expanding into neuroscience and experimental philosophy. Work in these fields has typically used experimental methods to test hypotheses about the rules and principles underlying moral judgment, how they develop in childhood, and how they vary across people and contexts. Influential early theorists included Jean Piaget, who argued for a constructivist view of moral development that began in obedience to authority and proceeded towards autonomous morality based on moral rules ([Piaget, 1932](#)). Lawrence Kohlberg developed this into a theory of three developmental stages: preconventional (based on reward and punishment), conventional (based on externally imposed rules), and postconventional (based on personal moral reasoning; [Kohlberg & Hersh, 1977](#)). In recent decades, the cognitive science of morality has expanded on these foundations, attempting to discover the mechanisms underlying human morality, including brain mechanisms, using experiments to test hypotheses based in moral philosophy.

Core concepts

Harm is an important concept in most normative theories of morality and a major focus of experimental studies of moral judgment in which participants are typically asked to make judgments about the actions and choices of people in hypothetical vignette scenarios. Inflicting intentional physical harm on another person is a prototypical case of a moral wrong that has been used in many such experimental scenarios. In the *trolley problem*, a philosophical thought experiment widely used in empirical studies of moral judgment, participants are asked whether they would intentionally kill a person (by pushing them or flipping a switch) to save the lives of several others ([Foot, 2002](#); [Greene, 2014](#)). This problem is sometimes seen as a test of whether people are consequentialists (save the most lives) or deontologists (do not kill). Other studies independently vary the *intention* or reason behind an act (the character in the vignette intends to cause harm, or not) and its *outcome* (someone is harmed, or not) to see whether and when intentions or outcomes matter more in people's moral judgments ([Young et al., 2007](#)). The relative importance of intentions versus outcomes appears to vary across contexts and people, and some scholars suggest that it is a major dimension of cultural difference ([Barrett et al., 2016](#)).

Because harm plays a role in most theories of morality, the question of what counts as a harm has been important in cognitive science research. Although physical harms (to body, property, and environment) are the prototype for theories of moral wrongdoing, debate surrounds cases of nonphysical harm (hurting someone's feelings) and cases in which no person or object seems to be harmed at all (eating disapproved food or privately thinking bad thoughts about someone). Some frameworks, such as the *moral foundations theory*, posit that there are different domains of moral judgment (harm, fairness, loyalty, authority, and purity), so that even acts that seem to cause no harm, such as eating disapproved food, can have moral valence ([Haidt, 2007](#)). According to the classic deontological framework (Kant), it is inner states or dispositions that carry moral weight, so lying can be wrong whether or not someone is harmed. In virtue ethics, ignorance, malice, and cynicism are vices (moral wrongs) that do not necessarily entail action. Sometimes causing harm can be judged as virtuous, as in cases of justified punishment or violence undertaken for righteous reasons ([Fiske & Rai, 2014](#)). The question of what, if anything, unites psychological judgments of moral wrongness remains an open one.

Psychologists, neuroscientists, and experimental philosophers have begun studying the mechanisms underlying moral judgment, including their brain bases ([Crockett et al., 2017](#); [Young et al., 2007](#)). *Dual process* theories posit that some moral judgments involve relatively slow conscious reflection and deliberation, whereas others are unconscious, fast, and driven by emotion. Experiments using the trolley problem suggest that the unconscious system is involved in deontological judgments, whereas the conscious, deliberative system underlies utilitarian judgments ([Greene, 2014](#)). The phenomenon of *moral dumbfounding*, in which people are unable to explain or rationalize the basis for strongly held and emotional moral judgments, supports the view that at least some moral judgment does not involve conscious reflection ([Haidt, 2001](#)).

Although much of cognitive science research has focused on moral wrongdoing, including harm, concepts of *virtue* and *good* also play an important role in peoples' moral judgments and ethical life, as emphasized by virtue theorists. Although virtue theory as developed by some Western philosophers such as Aristotle has been criticized as overly male centered and individualist, virtue-based ethical frameworks foregrounding relationality and the importance of community, including relational ethics and ethics of care, are widespread around the world in traditions including Confucianism and Buddhism as well as in feminist ethics ([Norlock, 2019](#); [Slingerland, 2011](#); [Van Norden, 2017](#)).

Questions, controversies, and new developments

The functions of morality

Because moral principles and judgments appear to have normative force—regarding what one should and should not do—some frameworks have suggested they might exist (e.g., that they evolved as part of human psychology) because of the functions they play in, for example, promoting social cohesion or otherwise regulating behavior. Evolutionary psychologists have linked moral judgment to the evolution of cooperation and the problem of altruism, the dilemma entailed when helping others requires

sacrificing one's own self-interest. Several evolutionary theories of morality posit that the human moral sense evolved to guide people towards cooperative outcomes, reminiscent of Durkheim's theory that a function of moral rules is to promote social cohesion ([Curry et al., 2019](#); [Nichols, 2004](#); [Tomasello, 2016](#)).

At the same time, morality can have a strongly arbitrary character, in which one person's or culture's virtue is another's vice. For example, in some countries today, there are people who consider it a moral imperative to wear clothing that reflects their religious identity, whereas others consider such displays to be socially harmful. A debate surrounds the degree to which morality is based on *social norms*, or *conventions*, which can vary across communities in a seemingly arbitrary way (for example, the nature, place, or time of a particular ritual or practice might be seen as extremely important but could easily have been otherwise; [Schmidt & Rakoczy, 2023](#)). Cultural evolution models show that arbitrary, group-beneficial yet culturally variable norms can evolve, consistent with Durkheim's proposal ([Richerson & Boyd, 2008](#)). That such norms frequently seem to be moralized appears to contradict theories that hold that conventional rules, because they are arbitrary, cannot be moral rules (e.g., theories that hold that morality must be based on universal principles of right and wrong; see [Kelly et al., 2007](#)).

Moral relativism

The question of normativity is related to debates about *moral relativism*. If one community holds that eating meat is bad and another that it is good, can they both be right? And might apparent cultural variation, for example, in what clothes are appropriate to wear to a wedding or funeral, be built on universal moral principles (don't be rude)? Normatively, some approaches to *moral realism* hold that there are moral facts independent of what people think, so that some communally held moral values might simply be wrong. Others, such as Durkheimian and cultural evolutionary views, hold that what is moral is whatever people agree on. Descriptively, the dimensions of variation in morality and moral judgment across cultures are still poorly understood, leaving open the question of whether any moral universals exist. Perhaps surprisingly, the question of how to identify the boundaries of the moral and the ethical—what morality is—remains unresolved ([Stich, 2018](#)).

Major questions remain concerning the degree to which our understanding of morality has been biased by who is doing the research, who is being studied, and where our theories of morality come from. The frameworks mentioned here largely come from European traditions because these have been the most influential in cognitive science research ([Van Norden, 2017](#)). There remain persistent biases in the representation of researchers and in which communities are studied. Ethnographic work suggests that peoples' everyday moral judgments might often resemble frameworks, such as relational ethics and ethics of care, that have been underexamined in cognitive science. Even now, theories of morality and ethics typically do not originate from the communities being studied, which presents a challenge to the ambition to understand and describe the morality of all humans, not just some.

Broader connections

Moral judgment is a uniquely human characteristic and makes use of many cognitive resources [see [Uniquely Human Cognition](#)]. Some kinds of moral judgment involve inferences about others' intentions, motivations, and beliefs and about the relationships between intentions, actions, and outcomes [see [Causal Reasoning](#); [Theory of Mind](#)]. Language, including words and concepts such as good, bad, permissible, and forbidden, are used in moral judgment and moral discourse and argumentation [see [Language](#)]. Morality may play a role in stabilizing cooperation, and morality has a strong normative aspect [see [Cooperation](#); [Normativity](#)]. Some kinds of moral behavior may involve signaling, such as signaling one's cooperativeness or moral virtue [see [Signaling](#)]. Moral systems may emerge through mechanisms of cultural evolution [see [Cultural Evolution](#)]. The systems composed of psychological mechanisms, culturally transmitted norms and values, and larger social structures can be understood as complex dynamical systems [see [Complex Dynamical Systems](#)]. Morality may play a role in niche construction, as when morally inflected political and economic systems structure and change social and physical environments [see [Niche Construction](#)]. Because much of cognitive science views itself as descriptive, the question of how the cognitive science of morality should inform social policy remains an open but important question for the future. As cognitive science grapples with and participates in the emergence of new technologies such as artificial intelligence and large language models, research on human morality and ethics are likely to, and should, play an important role in debates, policies, and orientations towards these unfolding technologies [see [Large Language Models](#)].

Further reading

- Greene, J. D. (2014). *Moral tribes: Emotion, reason, and the gap between us and them*. Penguin.
- Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. Oxford University Press.
- Tomasello, M. (2016). *A natural history of human morality*. Harvard University Press.
- Zigon, J. (2008). *Morality: An anthropological perspective*. Berg.

References

- Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M., Fitzpatrick, S., Gurven, M., Henrich, J., Kanovsky, M., Kushnick, G., & Pisor, A. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences*, 113(17), 4688–4693. <https://doi.org/10.1073/pnas.1522070113>

←

- Crockett, M. J., Siegel, J. Z., Kurth-Nelson, Z., Dayan, P., & Dolan, R. J. (2017). Moral transgressions corrupt neural representations of value. *Nature Neuroscience*, 20(6), 879–885. <https://doi.org/10.1038/nn.4557>

←

- Curry, O. S., Mullins, D. A., & Whitehouse, H. (2019). Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. *Current Anthropology*, 60(1), 47–69. <https://doi.org/10.1086/701478>

↩

- Durkheim, E. (1925). *Moral education*. Courier Corporation.

↩

- Fiske, A. P., & Rai, T. S. (2014). *Virtuous violence: Hurting and killing to create, sustain, end, and honor social relationships*. Cambridge University Press.

↩

- Foot, P. (2002). *Virtues and vices and other essays in moral philosophy*. Oxford University Press.

↩

- Greene, J. D. (2014). *Moral tribes: Emotion, reason, and the gap between us and them*. Penguin.

↩

- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834. <https://doi.org/10.1037/0033-295X.108.4.814>

↩

- Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316(5827), 998–1002. <https://doi.org/10.1126/science.1137651>

↩

- Kelly, D., Stich, S., Haley, K. J., Eng, S. J., & Fessler, D. M. T. (2007). Harm, affect, and the moral/conventional distinction. *Mind & Language*, 22(2), 117–131. <https://doi.org/10.1111/j.1468-0017.2007.00302.x>

↩

- Kohlberg, L., & Hersh, R. H. (1977). Moral development: A review of the theory. *Theory Into Practice*, 16(2), 53–59. <https://doi.org/10.1080/00405847709542675>

↩

- Mattingly, C., & Throop, J. (2018). The anthropology of ethics and morality. *Annual Review of Anthropology*, 47(1), 475–492. <https://doi.org/10.1146/annurev-anthro-102317-050129>

↩

- Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. Oxford University Press.

↩

- Norlock, K. (2019). *Feminist ethics*. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2019 ed.). <https://plato.stanford.edu/archives/sum2019/entries/feminism-ethics/>

↩

- Piaget, J. (1932). *The moral judgment of the child*. Kegan Paul.

↩

- Richerson, P. J., & Boyd, R. (2008). *Not by genes alone: How culture transformed human evolution*. University of Chicago press.

↩

- Schmidt, M. F. H., & Rakoczy, H. (2023). Children's acquisition and application of norms. *Annual Review of Developmental Psychology*, 5(1), 193–215. <https://doi.org/10.1146/annurev-devpsych-120621-034731>

↩

- Skorupski, J. (Ed.). (2010). *The Routledge companion to ethics*. Routledge.

↩

- Slingerland, E. (2011). The situationist critique and early Confucian virtue ethics. *Ethics*, 121(2), 390–419. <https://doi.org/10.1086/658142>

↩

- Stich, S. (2018). The quest for the boundaries of morality. In A. Zimmerman, K. Jones, & M. Timmons (Eds.), *The Routledge handbook of moral epistemology* (pp. 15–37). Routledge.

↩

- Tomasello, M. (2016). *A natural history of human morality*. Harvard University Press.

↩

- Van Norden, B. (2017). *Taking back philosophy: A multicultural manifesto*. Columbia University Press.

↩

- Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, 104(20), 8235–8240. <https://doi.org/10.1073/pnas.0701408104>

↩

- Zigon, J. (2008). *Morality: An anthropological perspective*. Berg.

↩