

What Makes a Judgment Moral?

The Case for a Computational Account

Bree Beal · September 2026 · Discussion paper 0.1

Abstract

People can reach the same practical conclusion for reasons they distinguish as moral and nonmoral. The difference raises a problem for attempts to identify moral judgment from its topic, wording, or outward consequences. If two judgments differ in moral character while sharing the features used to classify them, a classification based only on those features must miss the difference. This conditional argument motivates attention to how considerations function in forming a judgment. It does not establish that every moral judgment has one common organization, or that a particular computational theory supplies the correct account. Normative Detachment Theory makes the stronger proposal: otherwise nonmoral capacities combine in a distinctive relationship that contributes to moral judgment, while variations in that organization explain further familiar distinctions. A sustained example develops the proposal and separates it from claims about moral correctness, conscious deliberation, and the recovery of actual cognitive processes. The resulting research wager is that diversity in moral cognition can be explained through structure, rather than treated as sufficient reason to abandon a unified account.

One action and different reasons

Mara has prepared an analysis for a team meeting. Her supervisor repeatedly interrupts her, then presents one of her findings without acknowledging her contribution. Mara can ask to finish her explanation and clarify which work she contributed. She expects that speaking up will improve the chance that her work receives credit. She concludes, “I should speak up.”

Imagine two ways Mara could arrive at the conclusion. In one, her attention is directed to an upcoming promotion. Receiving credit is useful because it improves her prospects, and speaking up appears to be the available means. In the other, the treatment bears on Mara's sense of her own standing: her contribution deserves recognition because she matters as a person, not only because recognition might secure an advantage. Mara understands herself as the person being dismissed and as someone able to answer. Speaking up becomes a way of standing up for herself.

The workplace, available action, expected credit, and words of the conclusion can remain the same. What differs is the basis on which speaking up matters. The first imagined judgment is supported entirely by a career objective. In the second, Mara's sense of her own worth helps make the treatment objectionable and supports her judgment that she should respond. Protecting an advantage and responding from self-respect are intelligibly different reasons, even when they favor the same action.

The comparison does not divide people into those who have self-respect and those who seek promotions. Mara can have both concerns. She can also retain her self-respect while a particular practical question is settled entirely by its career implications. The contrast concerns what helps form one judgment, not Mara's complete personality or the whole history of her commitments. A third possibility is that both concerns contribute to the judgment. Their combination preserves the distinction between them rather than making it disappear.

Calling the career-only judgment nonmoral does not mean that pursuing a promotion is bad or that Mara ought to turn every workplace decision into moral deliberation. Nor does identifying a moral basis certify that speaking up is justified under every circumstance. The immediate question is what kind of judgment Mara is forming, not whether an observer endorses it.

The imagined reasoning is specified for comparison. It is not inferred from Mara's speed, tone, observable conduct, or later account of her motives. An actual person's reasoning may be mixed, implicit, or unavailable to inspection. The example makes a difference intelligible before asking how reliably that difference could be identified in an actual episode.

The boundary problem

A theory of normative cognition concerns judgments about what to do and how to appraise people, actions, or situations. Moral judgment belongs within that larger field: prudential advice, technical requirements, aesthetic appraisals, and moral obligations all involve evaluation or guidance, but they are not interchangeable. A theory that seeks to distinguish moral judgment must explain the difference without defining every kind of guidance as moral.

Mara's example makes several tempting shortcuts inadequate. The topic is workplace treatment in both cases. The proposed behavior is speaking up. The anticipated benefit is recognition. Even "I should" occurs in both judgments. None of those shared features, considered alone, explains why one judgment concerns an advantage and another concerns self-respect.

The same difficulty can recur with apparently more promising topics. Cooperation can serve a person's advantage or express a responsibility to others. Compliance with a rule can be guided by concern, convenience, or fear of a penalty. A familiar action does not carry one cognitive character into every context. The point is not that topics and outcomes are irrelevant. They become explanatory through the role they have in the person's understanding.

An appeal to a term such as *justice* does not automatically resolve the problem. If *justice* already means a moral requirement, the definition borrows the category it was meant to explain. If the term is given an independently specified meaning, the further question is how the represented requirement functions in the judgment. This is a demand for an explanation, not an objection to using ordinary moral vocabulary once its meaning is clear.

It would be a mistake, however, to infer that a criterion is mental merely because someone thinks about it. A chemist thinks about a sample's composition, but the composition is not thereby a thought. The stronger point in the Mara comparison is that the role of a consideration is part of the difference being explained. The same expected recognition can function as a means to promotion, as an

expression of personal standing, or as both. Explaining the judgment requires preserving that difference in function.

What the argument establishes

The core argument can be stated without assuming NDT's proposed components. Suppose a classifier receives only an episode's action, stated conclusion, and expected consequences. If two episodes have identical descriptions along those dimensions, the classifier must treat them alike. If the episodes nevertheless differ in the target distinction, the available description is insufficient. Adding more detail helps only if the added detail separates the relevant cases.

The conclusion is conditional. Readers must accept that the compared episodes are possible and differ in the respect being explained. The argument then rules out an exact classification that ignores every difference between them. It does not establish the moral classification of the episodes by mathematical fiat. Nor does one comparison establish that every proposed noncognitive criterion fails in every possible domain.

The productive next question is what an adequate repair must preserve. Adding Mara's reasons helps only if *reason* means more than a phrase attached to her conclusion. She could recite a self-respect explanation that did not help form her judgment. Conversely, self-respect could inform a rapid judgment without an internal speech. The explanatory target is the current organization through which the considerations support the judgment, whether or not Mara puts that organization into words.

Computational is used in that functional sense. An account specifies what is represented, how representations are related, which evaluations are operative, and how those relations support a prescription or appraisal. It need not describe conscious calculation or a literal diagram in the brain. A stored principle can retain and reactivate relevant relationships; the account does not require replaying the entire episode in which the principle was first learned.

The argument therefore supports a demanding constraint: an adequate general explanation cannot erase distinctions in cognitive organization that make a difference to the kind of judgment produced. A description in another vocabulary could preserve the same distinctions. A sufficiently informative neural description would not refute the argument simply by avoiding psychological terms. What would challenge the argument is an explanation that recovers the relevant contrasts while showing that their supposed cognitive dependencies are dispensable.

Two stronger claims remain separate. One is that moral judgments share an informative organization across different contents and contexts. The other is that NDT identifies that organization correctly. The first is a unification conjecture; the second is a particular theoretical proposal. Neither follows merely from the fact that outwardly similar actions can involve different thinking.

A relationship that contributes to judgment

NDT develops an approach introduced in *What Are the Irreducible Basic Elements of Morality?* and *The Nonmoral Conditions of Moral Cognition* (Beal, 2020, 2021). The proposal seeks unity in an organization of capacities rather than in an exclusively moral topic or ingredient. To understand what the proposal contributes, return to Mara's self-respect judgment.

Three aspects of the represented relationship can be distinguished. First, Mara matters in her own right, rather than only as a means to another benefit. NDT calls the positive version of that evaluation *positive inherent value*. Second, Mara connects the person being dismissed to herself as the relevant responder. NDT calls the more general connection *Relational Positioning*. Third, Mara represents a relevant capacity to respond: she can challenge the interruption and claim recognition. NDT calls that represented capacity *agency*.

The three aspects answer different questions. Mattering does not specify who is related to whom. A connection does not establish value. Being able to do something does not establish either. NDT calls the aspects *builders* because their combination helps constitute the proposed moral relationship. The current formal account additionally constrains their joint qualification; merely recording three labels or arbitrarily tiny values does not establish a qualifying relationship.

The combination must also concern the same represented people, object, situation, and judgment. Mara's valuation of a friend cannot be combined with an unrelated person's opportunity to act and counted as the relationship supporting Mara's response to her supervisor. In the self-directed example, Mara is both the responder and the person whose standing is at issue. The more general architecture also represents relations to other people and objects.

Finally, the relationship must actually contribute to the evaluation from which the particular judgment forms. In Mara's self-respect case, her own standing makes the dismissal objectionable, which gives speaking up its bearing as a response. A background belief that people have worth would not, by its mere presence, establish that connection. The relationship must do work in the same reasoning that produces the judgment, not be appended afterward as a description of Mara's character.

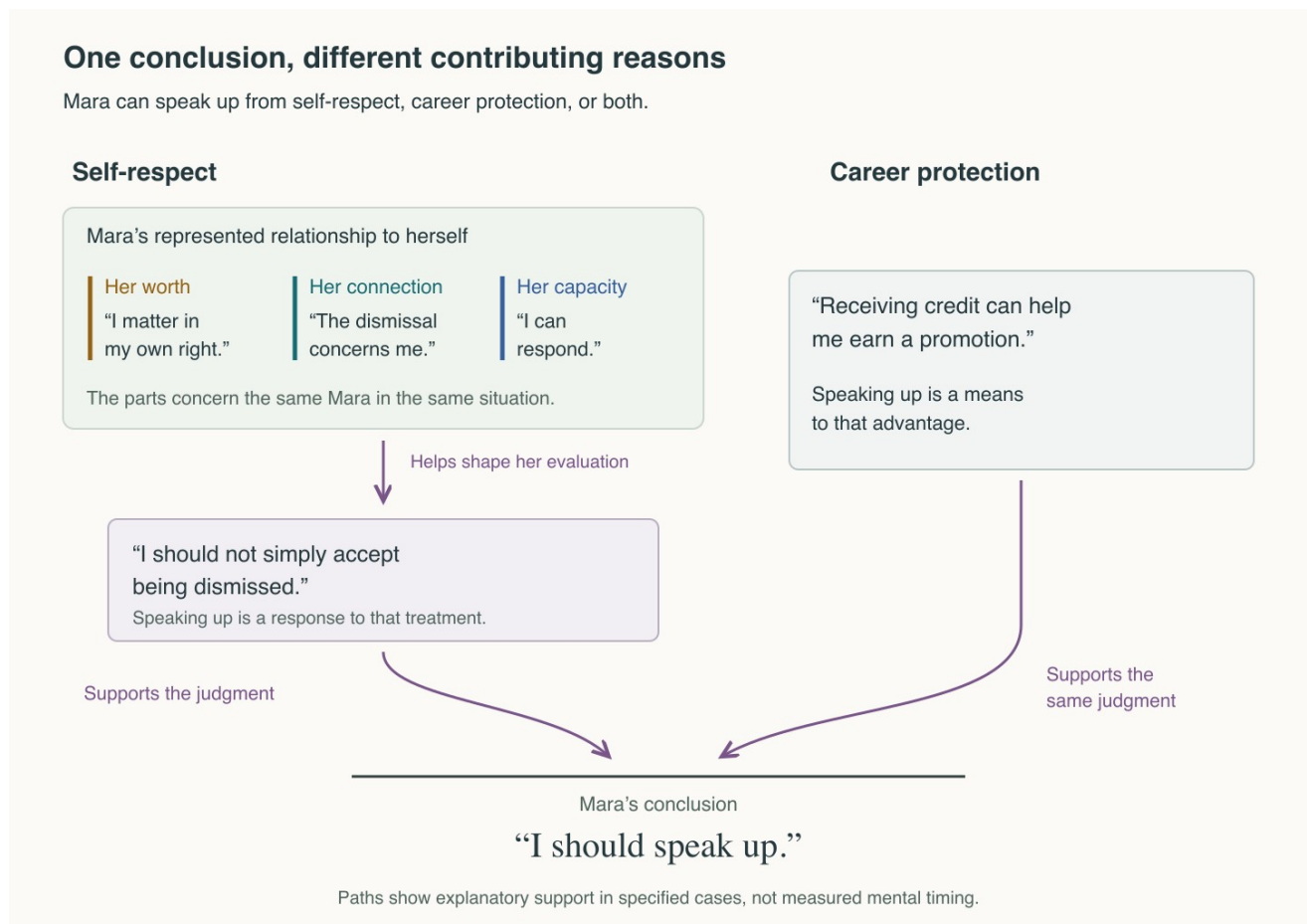


Figure 1. Similar conclusions can have different contributing bases. The paths summarize the stipulated cases, not a measured sequence of mental events or an execution of all formal conditions. The relationship-based path includes both the represented relationship and its contribution to the judgment. A retained but noncontributing concern is not an additional active path.

The career-only comparison is now more precise. Mara may continue to represent herself as a person of worth. For the specified episode, however, that relationship does not help organize her evaluation of speaking up; the expected promotion benefit supplies the operative reason. No other qualifying moral relationship contributes. When the self-respect relationship and the promotion concern both contribute, the judgment has a moral basis without ceasing to have a prudential basis. NDT does not require moral reasoning to be free of self-interest.

Why further distinctions matter

The value of a proposed architecture grows when established components clarify another contrast. A removal comparison asks what changes in a specified case when a particular condition is removed, while retaining the conditions needed to understand the contrast. It is a conceptual comparison within the account, not a procedure for deleting a feature from a person's mind.

Removing the contribution of Mara's self-respect relationship leaves that relationship represented but inactive in the particular judgment. Removing her represented inherent worth is a different change. In a case with only positive inherent value, removing that positive evaluation leaves instrumental

interests possible but removes a constituent of the relationship. The two interventions must not be pictured as the same disappearance. One concerns what Mara represents; the other concerns whether a retained representation helps form her judgment.

Value itself contains a further distinction. Suppose Mara is considering a colleague whom she once admired. She may now feel no inherent regard for the colleague while recognizing the colleague's usefulness to a project. Or she may regard the colleague as objectionable in their own right. The absence of positive inherent value and the presence of negative inherent value are different possibilities. Neither follows simply from the disappearance of admiration.

NDT accordingly represents positive and negative inherent value separately. Both may also be present: within one understanding of the colleague, Mara can regard that same person as both good and bad in their own right. The example concerns the person under the same construal, not merely useful and inconvenient features or a switch between different interpretations. The combination differs from indifference. Removing positive value from a mixed profile leaves negative value; removing it from a positive-only profile leaves neither. A single net score would hide a distinction that the separate components preserve.

The value profiles do not determine complete moral judgments. A represented relationship, relevant capacity, actual contribution, and the remaining conditions still matter. Nor does regarding a person as inherently bad determine whether the response is resistance, avoidance, punishment, or something else. Those consequences require additional structure. The immediate gain is that one organization can preserve distinctions that a coarser description might collapse.

Recognizable difference	What the comparison changes	What the architecture must preserve
Seeking an advantage and standing up for oneself	Which considerations support the judgment	The same sensible conclusion can have a prudential basis, a moral basis, or both
Having self-respect and drawing on it now	Contribution of a retained relationship	A relationship can remain represented without helping form this judgment
No inherent regard and inherent badness	Positive and negative value separately	Removing positive value does not automatically introduce negative value
Mixed regard and indifference	Which value components remain	Positive and negative value together do not collapse into neither

Table 1. Selected dependencies, not an exhaustive inventory or evidence of universal coverage. Builder removal demonstrates nonredundancy conditional on NDT's definitions. It does not prove that every adequate rival must use NDT's variables. A competing account may represent the distinctions differently; the question is whether the differences remain intelligible and their dependencies are explained.

The challenge from disunity

The unification conjecture encounters a serious alternative. Sinnott-Armstrong and Wheatley (2014) examine candidates for unity across content, phenomenology, force, form, function, and brain

mechanisms. They conclude that moral judgments are not unified at those levels and recommend studying more carefully distinguished groups. Their challenge is not answered by noting that cognition can have several neural realizations. It extends to the proposed organization and function of judgment.

Stich's later discussion identifies another difficulty: different research programs use different procedures to decide which judgments count as moral. In *The Cognitive Science of Moral Judgment and the Tower of Babel Problem* (2024, §4), he argues that neither intuitive classifications nor familiar identification procedures have established the required unity. He leaves open a successful alternative identification procedure, but regards that prospect as unlikely and recommends abandoning the unqualified category as a useful scientific term.

NDT can accept the need to distinguish phenomena and still investigate whether those distinctions belong within a common account. Failure of a proposed surface signature does not establish failure of every possible organization. Yet the converse is equally important: drawing one diagram that can accommodate many cases does not establish a discovered unity. A sufficiently unconstrained diagram could accommodate almost anything.

The substantive reply must therefore be architectural. The proposed conditions should be independently intelligible, jointly informative, and capable of excluding cases. Their relations should explain differences beyond those initially used to introduce them. A shared organization should do more than assign the same label to several examples; it should show why changes to specified features alter some distinctions while leaving others intact.

Disagreement between ordinary classification and NDT cannot be settled by automatic deference in either direction. An apparent counterexample may reveal that NDT misdescribes the reasoning, omits a relevant condition, or exposes an ambiguity in the ordinary category. Independent reasons must decide among those possibilities. Declaring every difficult case outside morality would protect a definition at the cost of its explanatory purpose.

There is a further risk that the proposed organization captures only a subset of what needs explaining. A broad normative architecture could encompass several genuinely different families without identifying a distinctive moral configuration. Generality of the surrounding architecture is not enough; the proposed moral boundary must earn its place within it. A persuasive failure to account for a relevant family would narrow or undermine the unification conjecture even if NDT remained useful elsewhere.

The disagreement is thus not between appreciating diversity and denying it. It concerns whether the diversity supports an organized explanation, and whether NDT supplies enough constraint to make that explanation informative. Fine-grained distinctions are part of the task of unification, not an alternative to precision that NDT must reject.

Precision appropriate to the subject

The chemical analogy used in Beal (2021) concerns an explanatory achievement, not a guarantee about cognition. Material described as water occurs in different appearances and arrangements.

Chemical understanding supplies distinctions that a classification equating water with liquid cannot express. Its usefulness does not depend on preserving every ordinary use of *water*.

Weisberg's *Water Is Not H₂O* challenges a simple one-to-one correspondence between ordinary-language kinds and scientific kinds. His account also shows that chemical descriptions themselves depend on the distinctions relevant to the inquiry (Weisberg, 2006). The relevant lesson is not that structural explanation fails when vocabulary is context-sensitive. It is that scientific usefulness and exact correspondence with an everyday word are different demands.

For moral cognition, the analogous aspiration is to find a structure precise enough to explain consequential differences. The analogy does not establish that such a structure exists, and the theory should not borrow chemistry's evidential standing. Aristotle's recommendation to seek the precision a subject permits is compatible with a demanding inquiry into what further precision is available. It need not be a reason to stop at inherited ambiguity (Aristotle, *Nicomachean Ethics*, I.3, 1094b).

Exact formal conditions and uncertainty about an actual person can coexist. A theory may state its conditions clearly while available information leaves it uncertain whether a particular episode satisfies them. Conversely, exact notation cannot establish that the conditions capture the intended phenomenon. The two problems require different work: examining the adequacy of the proposed organization and developing warranted ways to identify its application.

A research wager with consequences

The conditional argument requires an account to preserve differences in cognitive organization whenever those differences change the kind of judgment being explained. The unification conjecture goes further: a common, informative organization may distinguish moral cognition across its diversity. NDT supplies a particular candidate for that organization and makes its dependencies available for inspection. The conditional constraint motivates the search without deciding its outcome.

The proposal can be challenged without pretending that present methods recover complete cognitive provenance. An internally inconsistent specification, an ingredient whose definition secretly presupposes morality, or a well-developed case exposing a mistaken dependency would put pressure on the account. An alternative that explains the same contrasts with fewer independent assumptions would challenge its explanatory advantage. Whether the proposed organization describes actual cognition at the claimed breadth remains a further question; matching a person's reported answer does not settle it.

The initial benefit is already a concrete one to inspect. Mara's career concern and self-respect can lead to the same prescription without becoming the same reason. Self-respect can remain part of her understanding without contributing to every judgment. Someone can cease to seem good in their own right without coming to seem bad in their own right, and mixed regard can remain different from indifference. NDT connects those distinctions through specified relations rather than leaving them as an unrelated list.

The central question for a competing account is therefore not whether it reproduces NDT's diagram. It is whether it explains the distinctions and dependencies the diagram makes explicit, or supplies good reasons to revise the explanatory target. The ambition is a common structure that clarifies difference. Its success must be judged by what the structure explains.

Further exploration

Explore the paired cases in the interactive lesson. The lesson presents specified cases and their explanatory dependencies; it is not an execution of the complete formal theory.

Read the broader introduction to NDT. The Formal Specification, especially §§4–4b, supplies the complete conditions. The Scope and Commitments distinguishes the theory's commitments from the questions still open.

References

Aristotle. *Nicomachean Ethics*, Book I, chapter 3, 1094b. [English translation by W. D. Ross.](#)

Beal, B. (2020). What are the irreducible basic elements of morality? A critique of the debate over monism and pluralism in moral psychology. *Perspectives on Psychological Science*, 15(2), 273–290. <https://doi.org/10.1177/1745691619867106>

Beal, B. (2021). The nonmoral conditions of moral cognition. *Philosophical Psychology*, 34(8), 1097–1124. <https://doi.org/10.1080/09515089.2021.1942811>

Beal, B. (2026). *Normative Detachment Theory Formal Specification*, version 1.6.5, especially §§4–4b and 8.5. [Current formal specification.](#)

Sinnott-Armstrong, W., & Wheatley, T. (2014). Are moral judgments unified? *Philosophical Psychology*, 27(4), 451–474. <https://doi.org/10.1080/09515089.2012.736075>

Stich, S. (2024). *The cognitive science of moral judgment and the Tower of Babel problem*. Author manuscript for *Proceedings and Addresses of the American Philosophical Association*, especially §4. [Author manuscript.](#)

Weisberg, M. (2006). Water is not H₂O. In D. Baird, E. Scerri, & L. McIntyre (Eds.), *Philosophy of Chemistry: Synthesis of a New Discipline* (pp. 337–345). Springer. https://doi.org/10.1007/1-4020-3261-7_18