

Reviewing Normative Detachment Theory

Reviewer Guide 1.11 — September 2026 Status: Unratified RC4 author draft; not an approved review protocol.

Normative Detachment Theory (NDT) is a model of how someone forms normative judgments and of what distinguishes moral judgment from other kinds of normative cognition. It is easy to misread because several familiar questions sit just outside its explanatory target: whether a judgment is true, whether an action is right, whether a person acts on a judgment, whether a motive has moral worth, and whether ordinary speakers use the word *moral* in exactly the same way.

This guide identifies the minimum reading needed for a fair review, the documents that govern different kinds of claims, and the distinctions a critique should preserve.

A short route through the project

For a first encounter, begin with **Reader-Facing Article 0.23** and the **Normative Cognition Explorer**. They explain the central problem, the architecture, the moral-relationship account, and the main examples without requiring the reader to begin with notation.

For a substantive critique of the theory, read **Formal Specification 1.6.3** and **Scope and Commitments 1.16** in full. The Specification fixes the formal definitions, while Scope and Commitments states what the model does and does not claim.

Then choose the source that matches the criticism:

- For a claim about originality, ancestry, omission of prior work, or relation to another theory, consult **Literature Review 0.13**, **Master Bibliography 1.8**, **NDT in Context Dataset 1.20**, **Data Contract 1.13**, and **Source Register 2.15**. Dataset 1.20 contains 48 canonical records supported by the governed source units and assigns each record one exact source-governed relation. The re-adjudication ledger records every retained or changed relation.
- For a claim about empirical testability, measurement, calibration, research design, or the current evidence base, consult **Evidence and Open Questions 1.10** and **Empirical Audit 2.3**.
- For a claim about theorem status or inferential dependence, consult **Proposition Registry 2.9** as well as the Specification. **Executable Model 2.2** is a partial companion rather than an independent authority.
- For a claim about how the visual map encodes the theory, consult the Explorer's Architecture and Axioms & Theorems modules. The visualization is an explanatory interface, not an independent source of theoretical authority.

Five distinctions a review must preserve

1. Judgment is not behavior

NDT explains the formation of a judgment occurrence. It does not infer action from judgment. A person may form a moral judgment and fail to act on it; a person may also perform a morally conforming action for habit, fear, convenience, or another nonmoral reason. Evidence about behavior therefore bears on NDT only when a bridge from behavior to judgment formation is independently defended.

2. Normative cognition is broader than moral cognition

Normative cognition includes judgments about what should be done, permitted, required, preferred, praised, blamed, or appraised. The current NDT grammar models six output families and does not yet treat primitive permission as a separate formal output; that limitation should not be confused with the breadth of ordinary normative thought. NDT treats moral cognition as a structured subset of this wider class. Instrumental, prudential, institutional, aesthetic, and epistemic judgments may be normative without being moral.

3. Judgment content is not its producing provenance

Two judgment occurrences can use the same words and recommend the same action while being formed through different cognitive routes. NDT classifies the occurrence by what actually helped produce it, not merely by which beliefs, rules, values, or representations were available. A critique that infers cognitive type from verbal content alone has not yet addressed the theory's provenance claim.

4. Constitutive roles are not measurement proposals

NDT's constitutive claims identify the minimal functional organization of the proposed cognitive kind. Its scoring functions, thresholds, indicators, and identification procedures remain open to refinement. Showing that one proposed measure is poor does not by itself show that the functional role is unnecessary; defending a role does not validate a particular measure.

5. Comparative explication, formal countermodel structure, operational identification, and calibrated measurement are different achievements

Comparative explication asks whether NDT proposes a formally and conceptually noncircular architecture and preserves paradigm and contrast distinctions better than alternatives. Formal countermodel structure asks whether the definitions expose explicit cases in which registered conclusions hold or fail. Operational identification requires independently measured evidence that real episodes instantiate the proposed organization rather than a rival; whether NDT's boundary can be identified this way remains open. Calibrated measurement additionally requires validated indicators, weights, thresholds, uncertainty estimates, and a declared domain of use. A review should state which standard an objection reaches.

The relationship core: what the necessity claim amounts to

NDT proposes three independently variable roles:

- **Inherent valence:** independent positive and negative ways the object is represented as mattering in itself. Positive-only, negative-only, mixed, and neither profiles are distinct; opposite signs do not cancel.
- **Relational Positioning:** an operative bridge that locates this exact object relative to this exact subject. Its routes are identity-mediated positioning, intimacy, and direct encounter. Identity-mediated forms are shared, contrastive, and complementary-role.
- **Perceived agency:** the represented capacity of this subject to affect, answer to, respond to, or treat this object within a declared horizon.

The finite companion derives separate positive-valence-by-agency and negative-valence-by-agency branches before a profile-preserving relationship-strength gate, which reviewers should treat as derived rather than as a fourth builder.

Removing a whole role tests a different distinction: inherent significance versus merely instrumental relevance; abstract evaluation versus subject-relative positioning; or passive appraisal versus a represented capacity to respond. Removing one positioning route or identity form is narrower and does not remove the whole positioning role when another route remains.

Functional access is separate enabling machinery. The former standing-mediated category is retired: institutional assignment can be complementary-role identity-mediated positioning when the exact bridge is operative; physical or social distance is absence under an adequate route scheme; opposition is contrastive positioning, not negative valence.

Reviewers must also distinguish the core from downstream bending: a **strict bender** preserves the exact relationship-core snapshot while acting at one fixed downstream locus through the required live contributing path. An operation that changes the core is core-changing. Builder input, realization, and modulation are sufficient but nonexhaustive ways of changing the core. One representation may support separate positioning and bender edges, but evidence for one does not establish the other.

This is a contrastive claim for conditional nonredundancy, not a proof that NDT's labels, finite gates, or modules are uniquely correct. NDT asserts no universal numerical threshold. A critic can show that a registered removal test fails, that a distinction is unnecessary, or that a better account preserves the same work with less structure.

How to formulate a useful objection

A strong critique identifies four things:

1. **The exact claim.** Quote or cite the relevant section, definition, proposition, comparison record, or implementation clause.
2. **The claim type.** State whether the target is constitutive, formal, empirical, operational, quantitative, historical, comparative, or rhetorical.
3. **The rival or counterexample.** Explain which condition fails, which distinction survives without it, or which alternative organization explains the case better.
4. **The consequence.** State whether the objection requires a correction, a narrower claim, a new comparison, an empirical test, or a different exposition.

Evidence can count against NDT before a critic supplies a complete superior theory. Sustained failure on an independently specified contrast or consequence test can require restriction, revision, or abandonment of the affected construct. Where adequate measurement is not yet available, that is a limitation rather than evidence in the theory's favor.

The following are valuable forms of criticism:

- a formal inconsistency or countermodel under the stated definitions;
- a rival explication that preserves paradigm cases and explanatory distinctions with fewer or better conditions;
- an independently sourced literature comparison that changes the stated relation between constructs;
- an identification strategy that cannot distinguish NDT from a live rival;
- a proposed measure that is circular, confounded, or insensitive to the relevant organization;
- a robust episode in which pre-specified measures rule out explicit and implicit subject representation while independent process, transfer, and contrast evidence groups it with paradigm moral rather than tragic or axiological appraisal;
- sustained failure to identify any complete framed-detachment episode under mature component measures;
- pre-specified provenance partitions that repeatedly fail their registered intervention, transfer, process, or explanatory consequences; such repeated failure alone counts against provenance-based kind unity, while consistent rival outperformance strengthens the evidence but is not required;
- a reader-facing ambiguity that persists after the governing definitions are read.

Mechanistic diversity alone is insufficient to defeat the provenance-based kind-unity claim because a cognitive kind may be multiply realized.

By contrast, a claim that a topic is absent should follow a full-text check of the document in question. A familiar label, similar verdict, shared example, or neighboring philosophical theme does not by itself establish identity, contradiction, or lineage.

The canonical comparison layer and the frozen historical-development audit answer different questions. Admission to Dataset 1.20 establishes current comparison governance; it does not imply that a record was included in the earlier 27-record pre-2026 corpus pass. A reviewer challenging development history should therefore inspect the current development status and migration audit rather than infer history from a current formal mapping or from the existence of a comparison card.

A two-stage protocol for AI-assisted review

AI-assisted reviews should separate comprehension from criticism.

Stage 1: reading certificate

The reviewer reads the current Formal Specification and Scope and Commitments in full, then states—in its own words—its understanding of the five distinctions above. The certificate should identify the document versions and disclose whether the reading was continuous, performed through a complete text extraction, or based partly on a verified line-level comparison with a fully read prior version.

The reviewer should then stop. No substantive criticism belongs in Stage 1.

Stage 2: targeted analysis

After the certificate is accepted, the reviewer evaluates a stated target. It should cite the exact passage or formal clause, identify the claim type, and distinguish what follows from the objection from what remains untouched. When alleging a missing source, missing distinction, or missing caveat, the reviewer should search every public document in which that material would reasonably appear and report the search method.

An AI review is evidence about the work only after its textual and conceptual premises have been checked. Confidence, repetition across models, and fluent analogy do not substitute for primary-source verification.

Reading NDT's claim about “is” and “ought”

NDT describes how someone derives a normative judgment from an evaluatively structured representation of a situation. It does not claim that an evaluatively empty description logically generates an ought. The model's answer to the cognitive is-ought problem is therefore architectural: identify how evaluation is already present, attributed, or encoded; identify the working representation and viable judgment route; and explain how a particular normative conclusion is formed.

A review may challenge the adequacy of that architecture, its account of encoded evaluation, or the empirical identification of its components. It should not replace the claim with the stronger and rejected thesis that bare descriptive content produces normativity from nothing.

What a fair review can—and cannot—settle

NDT's value does not depend on treating falsifiability as the single standard of theoretical merit. The project also aims to clarify the explanandum, distinguish moral from nonmoral normativity, expose hidden assumptions in prevailing research designs, and provide a formal vocabulary in which rival accounts and empirical programs can be compared.

Some parts of that contribution are conceptual or constitutive. Some admit formal countermodels. Some generate empirical contrasts but still lack validated measures. Some are open development problems. A precise review improves the work by locating pressure under the correct standard; constitutive status does not exempt a descriptive explication from revision.

Compact checklist

Before submitting a substantive review, verify that it can answer yes to each question:

- Have I read the current Specification and Scope in full?
- Can I distinguish judgment from behavior, normativity from morality, content from provenance, constitution from implementation, and comparative explication from formal countermodels, operational identification, and calibrated measurement?
- Have I identified the exact claim and its status?
- If I invoke another theory, have I checked the primary source and compared bounded constructs at the same explanatory level?
- If I challenge a relationship role, route, identity form, or strict bender, have I used the matching removal test and held the other layers fixed?
- If I treat evidence as adverse, have I identified the pre-specified contrast or consequence that failed, rather than demanding a complete rival as a prerequisite?
- If I allege an absence, have I checked the complete relevant document rather than a sample?
- Have I stated the narrowest consequence the criticism warrants?

Review standard. A review that meets these conditions can be severe without being misplaced. That is the standard the project invites.