

READER-FACING COMPANION

Normative Detachment Theory

Scope and Commitments

Bree Beal

Version 1.16 - September 2026

STATUS Unratified RC4 author draft; not an approved theory artifact. This document states the proposed reader-facing boundaries in ordinary language. Formal Specification 1.6.3 controls technical definitions; Proposition Registry 2.9 controls claim status within this unratified RC4 candidate. The executable reference, empirical audit, literature review, and visualization suite are non-authoritative companions.

The commitments in one page

Commitment	Plain-language statement
Aim	Describe how normative and moral judgments form; do not decide which judgments are correct.
Object of explanation	One occurrence of judgment, not the later transition from judgment to action.
Level	A computational or functional architecture compatible with fast, slow, intuitive, reflective, neural, symbolic, or mixed implementations.
Represented properties	Value, agency, access, identity, intimacy, encounter, responsibility, ownership, permission, feasibility, and standards describe how someone represents a situation.
Normative formation	No prescription is generated from a base empty of every relevant evaluative ordering, goal, requirement, priority, or commitment.
Moral boundary	A moral judgment represents a subject in relation to an object through an inherent-valence profile, perceived agency, and non-scalar Relational Positioning; that exact bound relationship helps shape the operative evaluation on the same actual contributing chain.
Identity	Sentences and behaviors do not determine cognitive type. Numerical token identity is fixed by the judgment-event anchor; interpretation, derivation type, realization profile, and actual source are separately represented features of that token.
Framing	Agents, objects, actions, episodes, relations, and evaluative questions are represented under interpretations rather than given in one neutral description.
Conflict	Several locally coherent judgments may remain active; conflict additionally requires represented comparison-readiness, and the theory does not require one final verdict.
Evidence	Comparative explication adequacy, formal countermodel structure, operational identification, and calibrated measurement are distinct. Research feasibility constrains inquiry across all four and is not itself confirmation.
Actual contribution	Candidate empirical schemes are evidence about a fixed functional organization; unresolved scheme competition withholds classification rather than creating a negative output.
Positioning	Relational Positioning is valence-neutral. Its governed routes are identity-mediated positioning, intimacy, and direct encounter. Identity-mediated forms are shared, contrastive, and complementary-role.
Bending	A strict bender preserves the exact core relationship while changing a registered downstream locus. A core-changing operation is not strict bending on that same edge. Builder realization, input, and modulation are sufficient but not exhaustive ways to change the core; one support may also have a separately evidenced strict-bender edge.

1. What the theory explains

Normative Detachment Theory (NDT) describes how someone forms a prescription or appraisal from a represented situation. The central event is judgment formation: how the person represents what is happening, what matters within that representation, which actions or responses are possible, which standards apply, and which sources actually help shape the conclusion.

The theory includes instrumental, prudential, institutional, aesthetic, epistemic, and moral judgments within one broad architecture. Moral cognition is a subset with additional relational and agency-involving conditions.

2. Descriptive, not prescriptive

NDT does not decide which moral judgments are true, valid, justified, humane, or acceptable. It can describe someone making a judgment as endorsing, condemning, licensing, or rejecting a position, including positions rejected within another represented framework. A judgment can be moral in cognitive type without being endorsed by the theorist or by the reader.

The theory therefore does not assume moral realism, anti-realism, consequentialism, deontology, virtue ethics, care ethics, or any other first-order moral philosophy. Those approaches can describe styles of practical reasoning without fixing the truth or cognitive type of every token that uses them.

NDT's use of moral is an explanatory explication rather than a dictionary report. Ordinary and disciplinary uses may form a family of overlapping meanings; the model is adequate insofar as it is noncircular, handles paradigm and contrast cases, distinguishes otherwise identical outputs with different sources, integrates moral with nonmoral normativity, and supports independent tests. A familiar label does not force every token beneath it to count as moral cognition.

3. Judgment formation, not behavior

Detachment means the formation of a judgment. It does not mean that the person acts on the judgment. Action may be prevented by incapacity, competing motives, fear, weakness of will, habit, or external constraint. Moral worth and blame for action therefore remain distinct from the moral type of the judgment that preceded it.

Stored and habitual judgments remain within scope when a person retrieves, reactivates, endorses, or expresses a judgment token. The operative source of that later token may be a current reconstruction, a stored trace, or a mixed route. NDT does not make successful recovery of the judgment's original formation history a constitutive condition of its present type. Whether and when that earlier provenance can be recovered is an open empirical problem.

4. Level of explanation

NDT is stated at a computational or functional level. "Inference" means a structured transformation over representations, not necessarily conscious deliberation. A process may be fast, automatic, intuitive, perception-like, reflective, linguistic, neural, symbolic, or mixed and still instantiate the relevant structure.

The formal routes and graphs identify dependencies and bindings. They are not claims that the brain stores literal diagrams. Different algorithms or neural implementations may realize the same functional organization.

5. Represented properties and frame relativity

Terms such as value, agency, access, identity, intimacy, encounter, responsibility, feasibility, ownership, permission, and standard report what someone represents within an active interpretation. NDT does not infer mind-independent value, valid ownership, metaphysical agency, or genuine permission from those representations.

The current six-output grammar does not contain a primitive Permission output. Permission-like conclusions must be encoded through the current decision and appraisal structures or introduced only in a later, separately governed extension. Describing represented permissions or licenses does not itself add that primitive.

A frame is the coordinated interpretation under which someone construes the relevant agent, object, action, episode, relationship, alternatives, and question. The same person or event can be represented differently across frames. Before outputs count as conflicting rather than merely plural, the person must represent an appropriate bridge between the relevant acts or execution spaces and must make their appraisal questions or dimensions comparison-ready.

6. Evaluativity and the is-ought boundary

NDT does not generate a prescription from inputs stipulated to be empty of every relevant evaluative ordering, goal, requirement, priority, or commitment. The evaluative prerequisite is the broader claim that every normative detachment needs some operative evaluative organization. That organization may be explicit, attributed, encoded in a representation's functional role, or built from those sources.

The Constitutive Evaluativity Thesis (CET) proper is narrower and more distinctive: some representations are evaluative partly because of what they represent or because of the functional role that makes them the representations they are. Evaluation need not always be attached afterward by a separately tokened evaluator. Encoded evaluativity is therefore one proposed constitutive case; it is not claimed to be the exclusive route by which evaluation enters normative cognition. The stronger argument that every genuine semantic success condition is therefore minimally evaluative is acknowledged as an external argument, not adopted as a governing NDT assumption. Whether encoded realization exists in real systems and whether it is non-negligible are separate empirical questions, assessed at the same representation-token unit; the present theory supplies no universal numerical threshold for the second question.

The theory does not claim that evaluation by itself is a behaviorally specific ought. An action-guiding model must still connect the representation to possible actions, expected outcomes, standards, feasibility, priorities, and undefeated challenges.

7. Moral cognition as a relational, agency-involving case

A moral judgment represents a subject standing in relation to an object of concern, and the relationship helps shape the evaluation used in that judgment. Subject and object roles may refer to different entities or to the same self while remaining formally distinct roles. The subject may be implicit in the sentence, as in an appraisal of an event caused by a person or institution, but the subject cannot be absent from the cognition. An event someone represents as having no agent behind it may be terrible or tragic without being immoral.

The moral relationship has three builders. The inherent-valence builder is one profile with independent positive and negative components: an object may be represented as mattering positively, negatively, both, or neither beyond mere instrumental use or disvalue. Those four states remain distinct; the profile may not be replaced by a net or maximum score. A finite test may apply separate positive and negative branch criteria and disjoin their Boolean results only at the decision boundary. No universal threshold is assumed.

Perceived agency is the second builder. Relational Positioning is the third: it asks how this exact object is located relative to this exact subject. It is non-scalar and valence-neutral. A bound relationship token, the subject and object conditions, and the relationship profile's actual contribution are further requirements. When another agent is judged, the observer also represents that agent as capable of recognizing and responding to the relevant object. Self-reference clears only the subject-gate disjunct.

Relational Positioning has three governed route families: identity-mediated positioning, intimacy, and direct encounter. Identity-mediated positioning is realized when an independently specified and operative identity or role schema locates the exact object relative to the exact subject and contributes through that bridge. Its forms are shared identity, contrastive identity, and complementary-role identity. Mere category membership, possession of a role, or conventional association with a duty is insufficient. A physician/patient or official/constituent pairing can qualify through complementary roles when the subject actually takes up that schema in the reasoning chain. The old standing-mediated residual is retired. Represented opposition is contrastive identity, not negative inherent value; distance is the absence of positioning under an adequate detection scheme, not a fourth positioning route.

Interface fit and availability to memory, attention, or awareness belong to functional access and contribution machinery, not positioning. Any duty, permission, priority, appraisal, or other consequence supported by an identity bridge needs its own downstream edge. Identity does not carry those consequences by definition.

Perceived agency is also bounded. It requires at least one represented object- or relationship-relevant response affordance within a modal-temporal horizon fixed before classification. Present action feasibility is separate. Historical, prospective, counterfactual, collective, role-mediated, reparative, testimonial, expressive, and falsely represented capacities may qualify when the declared horizon admits them; bare awareness or evaluation does not. If the judgment represents total inability to respond across every admitted horizon, the perceived-agency condition fails in the current model.

The builder-deletion test has three top-level results. Removing the inherent-valence profile erases the distinction between an object that matters positively or negatively in itself and one that is valence-null or merely instrumental. Removing perceived agency erases the distinction between passive regard and situated capacity to affect or respond. Removing Relational Positioning removes every governed bridge joining the exact subject and object even if valence and agency

remain. Removing access blocks the affected contribution path. Removing one positioning witness need not eliminate moral type when another qualifying witness remains.

The represented-subject condition is also answerable to evidence. Validated measures can provide sufficiently strong negative evidence for explicit and implicit subject representation under a declared detection model. Such a result counts against the condition when independent process, transfer, and contrast measures place the episode with paradigm moral judgments rather than tragic or axiological appraisal. The detection model must state its temporal window, unit, domain, sensitivity, false-negative risk, and evidential threshold. A superior rival strengthens the challenge but is not required before the evidence becomes adverse; no single verbal classification is automatically decisive.

Strict benders alter a downstream action or outcome evaluation, action description, requirement, permission, priority, exception, alignment, feasibility representation, appraisal, or responsibility attribution while holding the exact core relationship fixed on the tested edge. An operation that changes a builder or a core subject/object construal is core-changing, not strict bending. Builder realization, builder input, and builder modulation are sufficient routes to a core change, but they do not exhaust the ways an operation could change the relationship core. Because the general core-changing label records that a support has at least one such operation, it does not rule out a separately evidenced strict-bender operation by that same support. Every edge must pass its own preservation, intervention, and contribution tests. Vulnerability, power, sentience, ownership, stewardship, authority, and urgency are candidate sources, not automatic classifications. NDT describes a represented duty or license without determining whether it is correct.

8. Tokens, sources, and actual contribution

NDT classifies one occurrence of judgment rather than a sentence type. The same words and action can arise through care, policy, fear, habit, professional training, or several sources at once. A source affects moral type only when it helps form the operative evaluation on the same connected chain; mere availability, background activation, or a separate witness route is insufficient.

Actual contribution is constitutive of the modeled event within a fixed constitutive model instance. A candidate profile scheme must preserve independently specified differences in the event's functional organization before minimal sufficiency and contribution are assessed. The declared comparison class and its scheme are fixed prospectively, before any target-token verdict in that class. Genuinely different declared classes may use different schemes, but no class or scheme may be introduced or revised after a token verdict in order to obtain that verdict. Constitutively adequate schemes are guaranteed to agree only when they are connected by a same-resolution bridge that preserves profile structure, registered intervention effects, and minimal-sufficiency relations. Unrestricted cross-scheme invariance is a stronger model-class constraint, not a consequence of adequacy alone. When live candidate schemes remain verdict-divergent and evidence does not discriminate among them, empirical classification is unresolved rather than negative.

9. Polyphony and the absence of a mandatory global verdict

Several interpretations of one broad case may support judgments that are locally coherent yet jointly incompatible. NDT calls that represented many-voiced structure normative polyphony. It differs from a tie within one model, tragic choice among unacceptable actions, ordinary inconsistency, irrelevant duplication, or several outputs that the person has not made comparison-ready. Unaligned plurality includes both the absence of an act bridge and cases in which one execution is shared but the appraisals lack a common question or surviving dimension map.

A higher-order interpretation may arbitrate among lower-order judgments, but the theory does not require arbitration to continue until one global verdict remains.

10. Present warrant and empirical limits

NDT begins from a diagnosis: moral psychology has developed methodologically varied and technically sophisticated experimental and computational programs without comparable agreement on what makes a cognitive token moral rather than nonmoral. Moral Cognition Explanandum Audit 1.2 supports this scoped diagnosis across a representative set of leading programs; it does not claim exhaustive proof that no existing theory could supply such a boundary. NDT addresses the classification problem directly: it supplies formally non-self-referential definitions and proposes a conceptually noncircular token-level distinction, distinguishes moral from nonmoral normativity, unifies several reasoning styles without treating their labels as essences, and states finite formal consequences. Its present warrant is therefore strongest at the conceptual, architectural, and formal levels. Empirical identification remains open at the level of validated, operationally independent identification. That work is an important direction for development, not a condition for recognizing the

classificatory and formal contribution and not a promise that every constitutive question can be converted into one decisive experiment.

Four levels should not be collapsed. Comparative explication adequacy asks whether the proposed kind is noncircular, preserves paradigm and contrast cases, integrates with nonmoral normativity, and earns discriminating consequences. Formal countermodel structure asks whether the architecture specifies finite conditions under which a registered formal claim fails. Operational identification asks what evidence can distinguish instantiations of the organization from rivals. Calibrated measurement supplies validated indicators, weights, thresholds, uncertainty, and a declared population or operating range. NDT is comparatively definite at the first two levels, constrains the third, and remains incomplete at the fourth - most visibly for access, separately measured positive and negative inherent value, positioning-route and identity-form uptake, adequate negative detection of Relational Positioning, profile-sensitive contribution, and empirical branch and agency criteria.

Research feasibility cuts across all four levels: a construct may be clearly specified while presently difficult to manipulate or measure. That difficulty is a registered limitation, not evidence in the theory's favor. Constitutive status likewise does not immunize the theory from assessment. Sustained failure on independently specified contrast or consequence tests can require restriction, revision, or abandonment of the proposed construct. Builder measurement is therefore not a cheap preliminary; subject-gate and prevalence claims depend on prior validation of those measures.

NDT also makes two architecture-level empirical bridge claims. At least some real episodes should instantiate the complete framed-detachment architecture rather than isolated compatible parts, and independently identified provenance groupings should outperform purely lexical, behavioral, character-based, content-based, or policy-source partitions on registered held-out consequences. Repeated failure on those tests would pressure the theory's descriptive application and its proposed provenance-based unifier. Falsifiability remains relevant without becoming the sole or central measure of the theory's value.

11. What the theory does not assume

- one fixed list of moral values or foundations;
- harm, cooperation, welfare, or social convention as the universal essence of morality;
- one privileged neural or symbolic implementation;
- that every moral conflict is polyphonic;
- that every normative judgment is moral;
- that every case yields an all-things-considered answer;
- that finite formalization or executable conformance is empirical confirmation;
- that empirical falsifiability is the sole or central measure of the theory's value;
- or that artificial systems lack moral cognition merely because they are artificial.

12. Authority and companion documents

Formal Specification 1.6.3 controls definitions; Proposition Registry 2.9 controls claim status within this unratified RC4 candidate. Empirical Audit 2.3, Moral Cognition Explanandum Audit 1.2, Literature Review 0.13, Master Bibliography 1.8, NDT in Context Dataset 1.20, Data Contract 1.13, and Source Register 2.15 are the RC4 candidate's empirical, comparative, and source companions. Source Register 2.15 contains 77 governed source units, and Dataset 1.20 contains 48 canonical records under twelve exact relation types. Nine records formerly carried only Literature Review identifiers; they are now canonical CTX048-CTX056 and pass through the same current schema, source, relation, kind, and formal mapping governance as the other current comparison records. Their canonical admission does not retroactively place them inside the frozen 27-record pre-2026 corpus audit. CTX060 separately records the rival classificatory partition that the earlier combined Moral Foundations record could not represent without obscuring its distinct implementation-family role. Dataset 1.20 preserves Dataset 1.19's records and relation adjudications while correcting approved terminology and authority pointers. It does not treat absent disagreement as compatibility or compress conflicts into similarity. Reader-Facing Article 0.23, Evidence and Open Questions 1.10, and Reviewer Guide 1.11 are the reader companions for this candidate. Executable Model 2.2 and the Explorer are non-authoritative companions. Executable Model 2.2 is a finite conformance model, not an empirical validation or exhaustive realization of the formal theory.