TECHNICAL AND EMPIRICAL COMPANION

NDT Empirical Hypothesis and Identification Audit

From formal commitment to discriminating test

Version 2.3 - September 2026

STATUS This document is an unratified, unmerged RC4 candidate and remains non-authoritative. Within the candidate, Formal Specification 1.6.3 and Proposition Registry 2.9 govern the proposed theory and claim status. The audit maps possible empirical development, circularity risks, identification strategies, evidential pressure, and present maturity. It does not treat falsifiability as the sole or central standard of NDT's value, and it does not commit the project to completing every program listed here.

Executive summary

NDT is not one empirical hypothesis. It is a layered architecture containing claims about evaluatively structured representation, normative detachment, relationship construction, moral subjecthood, actual contribution, defeat, appraisal, framing, and polyphony. Each layer requires its own antecedent measure, intervention logic, consequence measure, and rival model. A successful provenance study would bear directly on one central MDT thesis; it would not by itself confirm CET, the detachment engine, the relationship builders, the subject gate, or the full account of conflict.

The earlier audit identified and repaired one authority-level circularity risk: ControlRecruited may not be defined by the action or judgment later counted as its consequence. The current Formal Specification retains that independent control-state requirement and the local evaluative contribution predicates that prevent mere route reachability from substituting for actual constitutive relevance. Executable Model 2.2 is the finite RC4 companion and retains guards for evaluable-domain binding, local contribution, determinate satisfaction, output-source provenance, and bounded relationship/subject/object integration. Gate 2 additionally requires separate positive and negative inherent-value measurement, preservation of the four profile states, Relational Positioning identification across its governed routes and identity forms, functional-access separation, and a strict-bender/core-change test. No candidate label supplies its functional classification.

The current CET program is decomposed rather than treated as one result. H1 and H14 are semantic-control association claims. H1b is the open practical control bridge and requires independently identified control recruitment. H15a concerns encoded-evaluativity existence; H15b concerns non-negligible occurrence at the same representation-token unit. The universal strong-semantic-CET argument is external and supplies no NDT assumption or empirical conclusion. No one of these claims entails the others.

This audit also registers two architecture-level bridge hypotheses. H17 asks whether at least some real episodes instantiate the complete framed-detachment architecture rather than isolated compatible components. H18 asks whether independently identified provenance partitions earn their proposed unifying role through held-out intervention, transfer, process, or explanatory stability. These are open development targets, not claims of present confirmation.

1. Audit standards

The audit distinguishes four levels. Comparative explication adequacy evaluates noncircularity, paradigm and contrast coverage, integration, and discriminating consequences. Formal countermodel structure asks whether finite conditions of failure are specified for registered formal claims. Operational identification supplies evidence that can discriminate the proposed organization from rivals. Calibrated measurement fixes validated indicators, weights, thresholds, uncertainty, and a population or operating range. Precision at one level does not establish success at another.

Research feasibility cuts across all four levels. It concerns whether available designs can manipulate or measure the target with adequate fidelity and without changing its rivals. Present intractability is a limitation, not confirmation. The three builders--the independent-component inherent-valence profile, perceived agency, and non-scalar Relational Positioning--and the represented-subject condition remain constitutive role-claims in the current explication, but constitutive status does not immunize them from sustained adverse evidence. A disputed label is not a self-interpreting falsifier; neither is a fully developed rival theory required before an independently specified pattern becomes adverse.

Criterion	Requirement
Antecedent independence	The construct named in the hypothesis must be identified without using the predicted output as its criterion.

Criterion	Requirement
Pre-specification	Judging system, task index, comparison class, profile scheme, indicators, aggregation, intervention family, thresholds, and exclusion rules are fixed before target-token classification. Different genuine classes may use different schemes, but neither class nor scheme may be tailored after a token verdict.
Discriminating intervention	At least one manipulation changes the antecedent while preserving plausible rivals or makes rival predictions diverge.
Outcome separation	Constitutive target, observable proxy, and behavioral consequence are not collapsed.
Comparison class	The domain within which invariance, sufficiency, or convergence is claimed is declared.
Evidential pressure	The program states what pattern would pressure the local hypothesis, a quantitative instance, an implementation choice, or the adequacy of the constitutive explication.
Revision rule	Sustained failure on independently specified contrast or consequence tests must be allowed to restrict, revise, or abandon a hypothesis, quantitative instance, implementation family, or constitutive explication at the level actually tested. Lack of an adequate current measure is recorded as a limitation, not favorable evidence.
Research feasibility	Report intervention feasibility, manipulation fidelity, measurement burden, and likely collateral changes across all four evaluative levels; do not treat feasibility as a separate warrant category.
Maturity	Status is distinguished as registered hypothesis, developed identification program, pilot design, or open extension.

2. Empirical commitment map

Layer: Semantic-control association, control bridge, and encoded evaluativity

Registry: H1; H1b; H14; H15a-H15b; IP6

Empirical question: Is semantic differentiation associated with independently measured control organization; when do independently recruited representations realize a stake map; and do encoded cases exist or occur non-negligibly at the representation-token unit?

Maturity: Decomposed registered claims; control measurement and encoded-token identification remain open

---

Layer: Normative detachment

Registry: NDT core; IP7

Empirical question: Do represented alternatives, outcome maps, standards, feasibility, priorities, and challenges jointly predict token-level guidance?

Maturity: Moderate; distributed tests needed

Layer: Moral relationship architecture

Registry: H2a-H2b; IP9

Empirical question: Do the independent positive/negative inherent-value profile, perceived agency, and non-scalar Relational Positioning each do nonredundant work, while access, route, identity-form, and contribution measures remain distinct?

Maturity: Moderate; valence-branch, positioning, access, route/form, and contribution measures incomplete

---

Layer: Moral subject gate

Registry: H3

Empirical question: Does attributed integrative recognition, rather than entity kind, predict third-person subjecthood, and can pre-specified subject measures identify robust paradigm-patterning episodes with no explicit or implicit represented subject?

Maturity: Strong conceptual design; empirical measures needed

---

Layer: Moral provenance

Registry: H4; IP1-IP3

Empirical question: Can relationship-grounded and policy-only same-content tokens be distinguished by intervention, transfer, process, lesion/rescue, or model comparison?

Maturity: Most developed program

Layer: Stored and habitual judgment provenance

Registry: Open extension; F021/F127

Empirical question: When a stored or habitual judgment is retrieved, reactivated, endorsed, or expressed, can present operative provenance be identified separately from any recoverable original formation history?

Maturity: Open; original-history recovery is not a constitutive gate

---

Layer: Challenge admission

Registry: H5

Empirical question: Do independently admitted challenges predict route blocking or suspension under the selected semantics?

Maturity: Moderate

Layer: Frame variation and individuation

Registry: H6; H13

Empirical question: Can frame tokens and their boundaries be identified independently of the output difference they explain?

Maturity: Foundational but underdeveloped

Layer: Polyphony and alignment

Registry: H7; H12

Empirical question: Does manipulating represented bridges move locally fixed outputs between unaligned plurality and comparison-ready conflict?

Maturity: Clear prediction; difficult intervention

Layer: Collective subjecthood

Registry: H8

Empirical question: Does attributed collective recognition shift institution-directed relative to member-directed blame under fixed conduct and harm?

Maturity: Testable; undeveloped

Layer: Strict benders and core-changing factors

Registry: H9; IP5

Empirical question: Under matched core preservation, do candidate representations selectively alter downstream action or outcome evaluations, action descriptions, requirements, permissions, priorities, appraisals, or responsibility attributions; and when the core changes, can realization/input/modulation be distinguished?

Maturity: Adopted construct split; measures and interventions undeveloped

---

Layer: Violation-sensitive fallback

Registry: H10

Empirical question: When every option violates something, do priority/violation profiles guide choice beyond outcome score?

Maturity: Clear prediction; no dedicated program

Layer: Target-sort appraisal

Registry: H11

Empirical question: Does appraisal grammar vary with represented target sort under controlled valence and content?

Maturity: Clear prediction; ontology incomplete

Layer: Full-architecture instantiation

Registry: H17

Empirical question: Do any independently reconstructed real episodes instantiate the complete framed-detachment architecture rather than isolated compatible modules?

Maturity: Registered bridge hypothesis; whole-architecture identification undeveloped

---

Layer: Provenance-based kind unity

Registry: H18

Empirical question: Do independently identified provenance partitions provide greater held-out intervention, transfer, process, or explanatory stability than rival partitions?

Maturity: Registered; local provenance methods exist, kind-level comparison undeveloped

---

Layer: Independent edge credit

Registry: G23; C19; IP5

Empirical question: Can valence, positioning, strict-bender, and attractor effects of one representation be separately identified and selectively lesioned?

Maturity: Registered design; undeveloped

3. Noncircularity audit

Construct: ControlRecruited

Circularity risk: Risk: the policy state is inferred from the same behavior later predicted.

Required independence: Use pre-output internal/control-state measures, independent task transfer, or a separately measured policy state. Target behavior may corroborate but cannot define the antecedent.

Status: Authority repair in 1.4.4

Construct: Inherent-valence profile

Circularity risk: Risk: positive and negative inherent value are collapsed into net, maximum, sum, mean, generic affect, or target valence; downstream duties are used to define the builder.

Required independence: Validate the positive and negative components separately, retain the ordered pair and four category states through every analysis, and use only separately registered branch predicates. Neither component may be defined by a requirement, permission, priority, appraisal, blame, avoidance, punishment, or target verdict.

Status: Constitutive profile fixed; measurement needed

Construct: Functional access

Circularity risk: Risk: attention, awareness, or memory availability is treated as intrinsically moral, as positioning, or inferred from the judgment it is meant to explain.

Required independence: Pre-specify interface, availability, attention, and awareness measures; test whether blocking the affected path prevents contribution while leaving the three-builder profile and alternative access paths intact.

Status: Contribution machinery, not a positioning subtype; measurement needed

---

Construct: Relational Positioning

Circularity risk: Risk: positioning is inferred from moral language, valence, downstream duty, generic association, salience, or the eventual output.

Required independence: Before target classification identify identity-mediated, intimacy, or direct-encounter organization with exact bindings. Use positive/negative controls holding valence, agency, access, content, and response fixed. Under an adequate scheme, failure of every governed route is positioning absence rather than an unnamed rescue.

Status: Constitutive role and closed current taxonomy fixed; IP9 governs measurement

---

Construct: Identity-mediated positioning

Circularity risk: Risk: category membership, role possession, institutional assignment, or conventional duty association is counted as an operative bridge.

Required independence: Independently identify the identity/role schema, its operative uptake, exact subject/object location, and bridge contribution. Distinguish shared, contrastive, and complementary-role forms; test the bridge separately from every downstream consequence.

Status: IP9 identity-form target

Construct: Strict bender

Circularity risk: Risk: any correlated feature or anything that changes the moral relationship is called a bender.

Required independence: Hold the exact subject/object tuple, inherent-valence profile, AgencyCap, and positioning fixed; preregister a downstream locus and require actual contribution there. If the core changes, classify that operation as core-changing. Builder realization, input, and modulation are sufficient but nonexhaustive ways of changing the core. Candidate labels never determine the classification.

Status: IP5 adopted; empirical tests needed

Construct: Profile-sensitive contribution

Circularity risk: Risk: any accessible object information, or profile-level redundancy alone, is counted as local contribution by the relationship token.

Required independence: At tier one require route/process-specific access, reachability, local minimal-sufficiency membership, and the relationship token's local difference-maker. At tier two require either a controlled same-chain effect or a profile-level NESS witness for the exact bound profile. Profile NESS cannot repair a false tier-one difference-maker. Hold nonprofile relationship features and unrelated evaluative drivers fixed; reject cross-chain and coincidental-covariation witnesses.

Status: Formal two-tier guard registered; empirical identification needed

---

Construct: Challenge admission

Circularity risk: Risk: a consideration is called “unadmitted” because it did not change the output.

Required independence: Measure recognition, target identification, credibility, and relevance before the target output; supplement with attention, memory, or interference measures.

Status: Existing safeguard

Construct: Constitutive profile scheme

Circularity risk: Risk: a comparison class or scheme is introduced or re-indexed after a token verdict to obtain the preferred contribution classification.

Required independence: Fix judging system, task index, declared comparison class, scheme, resolution, and invariance domain prospectively; define adequacy by independently specified functional organization; classify unresolved rivalry rather than force a token verdict.

Status: Gate-1 authority repair in 1.5.3; empirical governance needed

---

Construct: Frame individuation

Circularity risk: Risk: frames are segmented wherever outputs differ, then output difference is explained by frame difference.

Required independence: Pre-specify boundary indicators in object/agent/action construal, question, action space, or control regime; test stability and predictive gain.

Status: Open foundational problem

Construct: Represented-subject presence

Circularity risk: Risk: an implicit individual, collective, institutional, or other agent is supplied after a verdict is labeled moral.

Required independence: Pre-specify subject indicators and exclusion rules; measure subject-to-target binding before the target classification; compare independently with tragic and axiological appraisal profiles.

Status: Direct pressure test registered; measurement undeveloped

Construct: Moral subject gate

Circularity risk: Risk: moral vision is inferred from blame or moral language.

Required independence: Use independent attribution tasks about reason registration, integration, standing recognition, and conduct revision; cross referent kind with attributed profile.

Status: Conceptually clear; measurement needed

3.1 Maturity and futility for contribution-scheme competition

IP2 and IP3 use the same discipline already registered for the full-architecture program. Before comparing schemes, fix the admissible judging system, task index, comparison class, scheme set, declared resolution, intervention family and coverage, holdout unit, equivalence margin, reliability threshold, and evidence that can retire a candidate. A scheme introduced only after observing a preferred token verdict is inadmissible.

The program must also state a stopping rule. Unresolvedness remains the correct local classification while genuinely live, independently fixed schemes survive and the registered evidence cannot discriminate them. After the intervention family and discriminator set satisfy the preregistered maturity threshold, persistent verdict divergence cannot be deferred indefinitely as mere measurement incompleteness. It then counts against the identification program's viability or, where a larger hypothesis requires stable identification, against that hypothesis. Equal or better held-out performance by a simpler independently fixed partition strengthens the adverse result.

4. Program designs beyond provenance

The following are candidate development routes, not a mandatory research contract. Each becomes probative only after its antecedent measures and rival interpretations are independently specified.

P1 Semantic-control association, control recruitment, and encoded evaluation

Design. Treat the claims as separate tests. For H1/H14, manipulate or measure semantic differentiation while holding reward and explicit instruction fixed and assess its association with independently measured control-sensitive organization. For H1b, identify consumption by standing control and intervention-sensitive policy state before testing whether a realized stake map exists. For H15a and H15b, independently delimit a population of representation tokens, identify EncOutcomeEval at that same unit, and govern sampling, uncertainty, and any non-negligibility rule before data inspection.

Key rivals. Separately tokened evaluator, attributed evaluation, associative priming, simple performance learning, output-defined control recruitment, unit shifting from tokens to persons, and post-hoc prevalence thresholds.

Discriminating prediction. H1/H14 predict the registered association; H1b predicts the practical bridge only when its full antecedent is independently present; H15a predicts at least one encoded token; H15b predicts non-negligible occurrence under future pre-specified governance. Control recruitment alone establishes none of encodedness, stable order preservation, or action guidance. The optional evolutionary or learning rationale is not itself tested by these associations.

P2 Detachment and fallback

Design. Cross action feasibility, represented prospect ordering, hard requirement, priority, satisfaction determinacy, and violation profile. Include missing and unresolved satisfaction states as negative controls, plus cases where every action determinately violates something. Test singleton and all-tied action sets separately from ordinary-language ought and permission judgments.

Key rivals. Outcome-only choice, rule recitation without detachment, heuristic conflict.

Discriminating prediction. Choice and appraisal follow the registered guide/fallback structure, including outcome-inferior lower-priority violations.

P3 Relationship architecture

Design. First validate the positive and negative inherent-value components separately across the intended population and frames. Preserve the ordered pair and the positive-only, negative-only, mixed, and neither states in every stored and

analyzed profile. Preregister separate branch criteria; a finite implementation may disjoin their Boolean results only at ValenceGate. It may not use net, maximum, sum, mean, broad affect, lexical polarity, or downstream response as a substitute for the pair. Cross that decision boundary with the independently measured perceived-agency floor.

Next identify Relational Positioning independently while holding the valence profile, agency, functional access, content, and overt response fixed. The governed route families are identity-mediated positioning, intimacy, and direct encounter. The identity-mediated forms are shared, contrastive, and complementary-role. For an identity case, separately identify an independently specified schema, operative uptake, exact subject/object location, and bridge contribution. Category membership, role possession, institutional assignment, or the presence of a conventional duty is not enough. A physician/patient pairing and an impersonal public role are useful complementary-role fixtures; each requires evidence that the modeled subject took the schema up in the operative chain.

IP9 fixes negative evidence before data collection. A domain-specific study declares route/form coverage, positive and negative controls, intervention and transfer coverage, reliability and equivalence thresholds, and a stopping rule. Under an adequate scheme, failure of every governed route counts as Relational Positioning absence rather than represented distance or an unnamed route. Standing-mediated is not a residual escape category. Local actual contribution is assessed only after the three-builder profile has been independently identified.

Perceived agency receives a parallel boundary test. The modal-temporal horizon is fixed before classification, and candidate present, prospective, historical, counterfactual, collective, role-mediated, reparative, testimonial, or expressive response affordances are identified separately from present feasibility. Bare awareness or evaluation is a negative control. Episodes represented as allowing no object- or relationship-relevant response across the declared horizon test the outer boundary rather than automatically expanding the meaning of response.

Key rivals. Single-factor moral visibility, net or maximum valence, one-dimensional affect, post-hoc concern measures, access-as-morality, identity labels without operative bridges, duty-to-identity reverse inference, and mere object-level reachability.

Discriminating prediction. Each builder does independent work. Lesioning one valence component can move mixed to positive-only or negative-only without erasing the other component; lesioning both produces neither. Removing every positioning witness eliminates subject-relative positioning even when valence, agency, and access remain. Blocking access prevents the affected route from contributing without becoming a positioning verdict. Route/form lesions have selective signatures: shared identity changes common-category transfer; contrastive identity changes self/other or ally/opponent transfer; complementary roles change paired-role transfer; intimacy changes relationship-history persistence; direct encounter changes direct-presentation uptake. One loss need not eliminate positioning when another witness remains.

P3b Strict benders, core-changing factors, and multi-edge credit

Design. Begin with a fully identified core relationship. Manipulate a candidate source while holding the exact subject/object tuple, inherent-valence profile, perceived agency, and positioning fixed. Preregister one downstream locus, such as an action or outcome evaluation, action description, requirement, permission, priority, override, feasibility representation, alignment relation, appraisal, or responsibility attribution. A core-preserving effect with actual contribution supports StrictBender. If a builder or core construal changes, the same manipulation tests a core-changing operation. Builder realization, input, and modulation are sufficient but nonexhaustive ways of changing the core; each role requires its own evidence.

Test candidate labels symmetrically. Urgency, vulnerability, power, sentience, ownership, stewardship, authority, kinship, and institutional role are source descriptions only. They do not establish strict bending. For multi-edge cases, use selective lesions or mediation models to test valence, positioning, strict-bender, and attractor credits separately. Identity-mediated positioning and a duty derived from that identity are the central noncollapse control.

Key rivals. Semantic proximity to the outcome, broad affect, generic salience, builder change mislabeled as bending, downstream duty used to infer identity, and one latent attitude explaining every edge.

Discriminating prediction. Holding the core fixed, changing urgency may alter a downstream priority while leaving complementary-role positioning intact. Removing the priority edge should not remove the bridge; removing the bridge should not be scored merely from continued duty language. When a manipulation changes positive/negative valence, perceived agency, positioning, or a core construal, StrictBender must fail for that operation even if a downstream effect also occurs.

P4 Represented subjects and the third-person gate

Design. After builder measures are validated, run two linked tests. First, cross human/nonhuman referent with high/low attributed integrative recognition while holding object, act, harm, and relationship profile fixed. Second, assemble a pre-registered subject-absence challenge set in which independent measures test explicit and implicit individual, collective,

institutional, or other agent representation before the target judgment. Include first-person prospective controls, separately coded first-person reconstructive cases, and independent appraisal-target measures.

Key rivals. Entity-kind heuristic, answerability proxy, anthropomorphism, hidden or implicit subject representation, tragic or axiological appraisal, and the bounded character/moral-worth interpretation of some reconstructive cases.

Discriminating prediction. Third-person gate verdicts track represented recognition rather than referent kind. A robust episode for which the pre-specified model rules out subject representation is adverse evidence against the subject condition when independent process, transfer, and contrast measures place it with paradigm moral judgments rather than tragic or axiological appraisal. The result is evidence, not an automatic one-case verdict.

P5 Challenge admission

Design. Measure challenge registration before the target judgment; manipulate credibility/relevance and use competing skeptical, credulous, graded, or probabilistic models.

Key rivals. Output-defined admission, generic uncertainty, demand compliance.

Discriminating prediction. Only independently admitted challenges show the predicted blocking pattern.

P6 Frame individuation

Design. Use change-point, representational-similarity, task-switch, and question/action-space manipulations to identify frame events before conflict classification.

Key rivals. Continuous drift, response strategy, experimenter-imposed labels.

Discriminating prediction. Pre-specified frame boundaries improve out-of-sample prediction of token and bridge structure.

P7 Target-sort appraisal

Design. Hold broad valence and event content fixed while varying whether the target is an act, agent, institution, outcome, relation, or event.

Key rivals. Lexical association, valence-only model.

Discriminating prediction. Judgment grammar and responsibility requirements vary by represented target sort.

P8 Collective subjecthood

Design. Hold institutional outcome, policy, harm, and member conduct fixed; manipulate attributed institutional recognition and revision capacity.

Key rivals. Grammar/personification, organization size, diffuse responsibility.

Discriminating prediction. Institution/member blame allocation shifts with attributed collective recognition.

P9 Alignment and polyphony

Design. Hold two local outputs fixed; add/remove act, execution, target, common-question, or dimension bridges.

Key rivals. Simple inconsistency detection, semantic similarity.

Discriminating prediction. Pair classification moves between unaligned plurality, aligned plurality, and conflict as the appropriate bridge changes.

P10 Strict-bender qualifier test

Design. Hold core relationship and overt content fixed; independently manipulate stewardship, sovereignty, dominion, guardianship, authority, or kinship.

Key rivals. Legal status alone, intimacy, generic ownership wording.

Discriminating prediction. Downstream duties, permissions, priorities, overrides, or appraisal change at the predicted locus without direct moralization and while the exact core profile remains fixed. A core change is reported separately rather than counted as strict bending.

P11 Full-architecture instantiation

Design. Identify frame boundaries, evaluative grounding, decision or appraisal structure, candidate formation, active contribution, defeat, and output independently, then compare a joint framed-detachment model with disconnected component models on held-out episodes. Moral cases additionally require an independently identified three-builder relationship--including the two-component inherent-valence profile, perceived agency, and a governed Relational Positioning route--plus subject, object, and same-chain contribution. Pre-register a staged maturity rule: component measures, binding measures, and the joint reconstruction must each meet declared reliability and contrast criteria before H17 is assessed. Also declare convergence and futility rules so persistent failure to obtain an identifiable joint episode after mature measurement cannot be deferred indefinitely as mere measurement incompleteness.

Key rivals. Retrospective route reconstruction, compatible but causally disconnected modules, content-only prediction, and behavioral fit without token-level provenance.

Discriminating prediction. The joint architecture predicts held-out token formation and intervention effects better than disconnected component models. Repeated component-wise fit without identifiable joint episodes pressures H17.

P12 Provenance-based kind unity

Design. Define provenance partitions before measuring the target consequences. Pre-register a primary comparison metric, holdout unit, complexity controls, uncertainty rule, and the result counted as substantively meaningful. Compare provenance partitions with content-, lexical-, behavior-, character-, policy-source, and strong hybrid partitions on held-out intervention, transfer, process, and explanatory-stability measures.

Key rivals. Content identity, lexical moral labeling, behavioral conformity, character appraisal, policy-source classification, and hybrids that combine those partitions with limited provenance features.

Discriminating prediction. Provenance partitions exhibit greater held-out stability or discrimination. Mere internal heterogeneity is not failure under multiple realization; repeated absence of comparative advantage at or above the preregistered substantively meaningful margin pressures H18. Consistent rival outperformance strengthens the adverse result but is not required.

5. Moral-provenance identification: developed but local

The existing provenance program compares relationship-constructed, relationship-retaining general-norm, and provenance-free policy pathways while holding content and overt response constant. It uses intervention, transfer, process measures, lesion/rescue, and held-out model comparison. This program addresses whether the represented relationship actually helped construct or sustain the operative evaluation. It is a direct test of one central MDT thesis. It is not a general empirical test of CET, normative detachment, builder thresholds, the subject gate, frame individuation, challenge admission, or polyphony.

Some third-person and first-person reconstructive provenance intuitions may partly concern the agent's character or the moral worth of conduct. Treat that as a local rival, not a global alternative to MDT: it does not explain first-person prospective judgment, and its predicted variance should track independently measured appraisal target. Character assessment is itself available within NDT's agent-appraisal grammar. A probabilistic relation among behavior evidence, expected conduct, character attribution, and judgment type remains an open future extension rather than a current formal commitment.

Strategy	Role
Intervention	Alter the relationship representation while preserving the explicit rule; alter the rule while preserving the relationship route.
Transfer	Compare generalization by relational similarity with transfer by formal rule membership.
Process	Measure latency, confidence, attention, affective intensity, persistence, memory, and reframing sensitivity.
Lesion/rescue	Weaken one route while preserving the other, then restore it.
Model comparison	Compare relationship-provenance, policy-only, outcome-only, and mixed-route models on held-out cases.

H18 asks a broader question than local pathway discrimination: whether provenance supplies a useful unifier of the target kind. That claim earns support only if independently identified provenance partitions improve held-out intervention, transfer, process, or explanatory stability relative to registered rival partitions. If they repeatedly do not achieve the preregistered substantively meaningful advantage, NDT's provenance-based unity claim is undermined even if individual provenance reconstructions remain possible. Consistent rival outperformance makes that conclusion stronger but is not a necessary condition.

6. Existing evidence and current warrant

NDT begins from a scoped literature diagnosis: representative moral-psychology programs are methodologically developed but operationalize different explananda and do not share a token-level criterion separating moral from nonmoral normativity. Moral Cognition Explanandum Audit 1.2 records the comparison and its limits. The current evidential posture is strongest at conceptual, architectural, and formal levels: NDT identifies a foundational classification problem, supplies formally non-self-referential definitions, and proposes a conceptually noncircular token-level distinction, distinguishes moral from nonmoral normativity, and states consequences that rival models can contest. Phenomenology, literary and naturalistic fit, philosophical unification, neighboring findings, and finite conformance support taking that architecture seriously, but they do not substitute for direct operational identification or calibrated measurement. Validated, operationally independent identification remains open; examples of the required discipline include antecedent validation outside the target verdict and held-out consequence or transfer tests. Falsifiability matters where NDT advances empirical or quantitative claims, but it is not the master criterion of theoretical value: clarifying the explanandum, articulating a noncircular kind proposal, and exposing hidden dependencies are substantive achievements even before a calibrated measurement program exists.

Beal and Rottman (2024): executed study and bounded warrant

Across two preregistered U.S. adult studies (combined final N = 361), issue-targeted nonmoral-belief items outperformed the studies' broad value measures in the reported regressions across eight controversial judgments. The result warrants a narrow evidence claim and the limits below.

Audit question	Finding and evidential consequence
Narrow supported result	Broad value endorsements left issue-level variation unexplained; targeted beliefs carried substantial additional information in the reported samples and models.
Specificity and semantic proximity	Candidate-bender items were tailored to particular judgments and predicted directions, while broad values were reused. The design does not isolate nonmoral status or strict-bender function from target specificity and semantic proximity. Generic attractors in Study 2 weaken, but do not eliminate, a wholly specificity-based account.
Distinctness from the verdict	The validation vignette supports non-synonymy: changing another premise could change the anticipated verdict while the candidate item was held fixed. It does not establish psychometric discriminant validity, exclude shared latent attitude or method variance, identify actual contribution, or demonstrate core preservation.
Item construction	Item families were asymmetric at the outset; Study 2 also reworded low-performing belief and value items. Study 2 is a tuned extension rather than a fully frozen replication, and retained-item counts are not exchangeable opportunities across differently sized and targeted item families.
Model selection and generalization	Both studies used backward p-value elimination; Study 2 recombined selected predictors. Preregistration and Bonferroni correction do not supply post-selection inference, model-stability evidence, or held-out prediction.
Direction of influence	Cross-sectional associations measured after the moral judgments do not distinguish belief-to-judgment influence from rationalization, consistency maintenance, reciprocal influence, or a common cause. Exploratory conservatism controls narrow one rival but do not settle direction.

Audit question	Finding and evidential consequence
NDT evidence status	Preliminary consilience for distinguishing candidate downstream beliefs from broader attractors and for motivating H16. Not evidence that any item satisfies StrictBender; not a matched-perturbation test of H16 or a test of the builders, subject condition, actual contribution, MDT, H17, H18, or the full architecture.

- Keep consilience separate from the Proposition Registry's guarantees and empirical hypotheses.
- Name the exact local claim to which a cited study is relevant.
- State that the study was not designed to test NDT unless it was.
- Avoid prevalence claims where the model presently specifies only possibility or conditional structure.
- When evidence is ambiguous, leave it out rather than convert compatibility into confirmation.

7. One possible development sequence

This ordering records dependencies among possible studies; it does not claim that the project must execute the sequence or that empirical completion is the theory's primary aim.

1. Validate the independent ControlRecruited measurement architecture, then keep H1/H14 association, H1b practical bridging, H15a encoded existence, and H15b non-negligibility as separate analyses at their registered units.
2. Run a crossed builder study that separately exercises the positive and negative valence branches, preserves the four profile states, varies perceived agency, identifies each governed positioning route and identity form, and keeps functional access separate.
3. After builder and attributed-recognition measures are validated, run the crossed subject-gate design with referent kind orthogonal to attributed recognition and include first-person prospective controls.
4. Extend the existing moral-provenance program with lesion/rescue and held-out model comparison.
5. Develop and validate a frame-individuation protocol before population-level polyphony claims.
6. Pilot violation-sensitive fallback and target-sort appraisal as tractable downstream tests.
7. Test strict benders, core-changing factors, and multi-edge credit with matched core-preservation and selective-lesion designs; include collective subjecthood and relation qualifiers as later-stage programs.
8. Treat H17 and H18 as integrative tests after the relevant component measures are available; early absence of whole-architecture data is neither confirmation nor refutation.

8. Evidential pressure and revision ledger

Local claim	Evidential pressure or revision question
H1/H14 semantic-control association	Semantic differentiation shows no registered association with independently measured control-sensitive organization across the declared range, or a simpler rival explains the association better. This result does not by itself settle H1b or H15a-H15b.
H1b practical control bridge	The full independently measured antecedent is present but no realized stake map is found across mature registered tests, or attributed/separable evaluators consistently explain the cases better. Control recruitment alone is not a positive H1b result.
H15a/H15b encoded evaluativity	Mature token-level measurement finds no EncOutcomeEval instance in the delimited population, pressuring H15a; or encoded cases occur but fail the pre-specified non-negligibility rule, pressuring H15b without erasing existence. Person-level existential or prevalence evidence may not substitute for the registered representation-token unit.

Local claim	Evidential pressure or revision question
Builder condition	Under independently specified contrast tests, removing either valence component, agency, or Relational Positioning does not erase the distinction assigned to that builder; the four valence states cannot be reliably distinguished; independently measured paradigm-patterning episodes systematically lack one role; or a redistributed rival yields substantially better noncircular coverage and discrimination. Under a mature adequate positioning scheme, absence of all three governed routes is adverse rather than an unnamed-route deferral. Repeated failure to add intervention or transfer information pressures CE1 or H2b without treating raw verdict counts as decisive.
Represented subject and gate	Validated measures provide sufficiently strong negative evidence for explicit and implicit subject representation under a declared detection model in robust episodes that independent process, transfer, and contrast measures place with paradigm moral judgments rather than tragic or axiological appraisal; or third-person classification systematically tracks entity kind rather than attributed recognition. The first pressures the subject condition; the second pressures the gate model or its implementation.
Provenance	Relationship-grounded and policy-only pathways remain indistinguishable under every feasible intervention/process model, or policy-only cases consistently receive moral type despite absent relationship contribution.
Fallback	When every option violates a requirement, choice follows only outcome score despite stable represented priority and violation structure.
Target grammar	Target sort adds no predictive information beyond valence and lexical content across pre-specified cases.
Alignment/polyphony	Adding or removing the registered bridge does not change comparison readiness or conflict classification while local outputs remain fixed.
Strict bender/core change	Under exact core preservation, candidate-source manipulations do not selectively alter a downstream locus; apparent benders change a builder or construal; or source labels predict classification without edge-specific evidence. The first pressures H9; the second requires retyping; the third indicates construct leakage.
Full architecture	Independently measured components fit separately, but no real episode exhibits the registered joint framed-detachment organization, or disconnected rival models consistently predict held-out formation and intervention effects better.
Provenance-based kind unity	Independently identified provenance partitions repeatedly fail to achieve the preregistered substantively meaningful improvement in held-out intervention, transfer, process, or explanatory stability. Consistent rival outperformance strengthens the adverse result but is not required. Heterogeneity alone is insufficient because multiple realization remains possible.

9. Authority and limits

This audit does not amend the formal architecture. It tracks the four evaluative levels distinguished in Formal Specification 1.6.3 and registers optional empirical work that could strengthen, refine, restrict, or undermine local applications and the theory's descriptive bridge claims. Failure of a local hypothesis may disconfirm a quantitative instance or implementation family. Sustained failure on independently specified contrast or consequence tests may also pressure a constitutive explication before a fully developed superior rival has been supplied, although disputed lexical verdicts are not automatically decisive. Conversely, success in one program - including moral-provenance identification - does not confirm the whole theory. NDT's conceptual value is not suspended until this open program is completed.

Executable Model 2.2 supplies a passing 158-method finite regression suite, including the crossed-comparison-class regression, and a passing 92-mutation sensitivity audit for selected formal seams. Those results are executed finite-model results, not proposed studies, empirical findings, psychological validation, or proof of prose-code equivalence.