Empathy Systems Theory

Supplementary Materials

Universal Infrastructure for Coherence, Mechanism for Generativity,
and Foundation for AI Empathy Ethics

Dylan Mobley
Empathy Ethicist
empathyethicist.ai

December 2025

Contents

S1. Glossary of Terms

GLOSSARY OF KEY TERMS

Terms are organized conceptually to show framework relationships. For an alphabetical quick reference, see the index at the end of the glossary.

1. CORE FRAMEWORK ARCHITECTURE

C-A-E-I (Capacity Architecture and Emotional Integration) Model: Four-component architecture specifying empathy infrastructure maintaining James's associative networks:

- **Core Authenticity (C)**: Self-knowledge clarity reduces identity processing costs
- **Attachment Security (A)**: Relational safety reduces threat-monitoring costs
- **Expression Freedom (E)**: Emotional access reduces signal-translation costs
- **Integration Coherence (I)**: Narrative continuity reduces reconciliation costs

Components function interdependently; damage to one increases the load on others. Empirical specification required: four-factor structure predicted theoretically but may collapse to fewer factors under psychometric testing.

Core Authenticity (C): First component of empathy infrastructure architecture. Self-knowledge clarity enables accurate identification of one's own emotional states, needs, values, and preferences. Synthesizes established authenticity research (self-concept clarity: Campbell et al., 1996; authentic living vs. self-alienation: Wood et al., 2008; differentiation of self: Skowron & Dendy, 2004; authenticity components: Kernis & Goldman, 2006) as infrastructure capacity rather than a personality trait. Provides foundation for all empathic processing; without clear self-knowledge, cannot distinguish own emotional experience from others' emotional states (self-other confusion produces projection or emotional contagion rather than empathy). Reduces identity processing costs by maintaining a stable self-concept. Operationalized through: emotion differentiation accuracy, self-concept clarity scores, authentic emotion access, and response latency on self-emotion identification tasks. Predicted to fragment first under CEOP demands, as chronic authenticity-performance misalignment directly targets this component, creating sustained dual processing between authentic internal states and required external displays.

Attachment Security (A): Second component of empathy infrastructure architecture. Relational safety enabling trust in connection stability without hypervigilance. Secure attachment provides a stable base for empathic engagement, enabling individuals to feel with others without the threat of abandonment or engulfment. Reduces threat-monitoring processing costs, freeing capacity for the integration of emotional information. Operationalized through: attachment anxiety/avoidance scores, trust in relationship stability, relational hypervigilance (reverse-scored), and autonomic flexibility during social interaction. Predicted to erode second in C→A→E→I cascade as Core Authenticity loss triggers compensatory relational monitoring, increasing processing load on this component.

Expression Freedom (E): Third component of empathy infrastructure architecture. Emotional awareness and communication capability enable signal clarity in relational coordination. Clear emotional signals reduce interpretive ambiguity and enable authentic relational exchange; constrained expression requires degraded signal interpretation and suppression effort, consuming additional processing resources. Reduces signal-translation costs through accessible emotional vocabulary and safe expression contexts. Operationalized through: emotional range access, expression safety ratings, experience-expression concordance, emotional vocabulary breadth, alexithymia indicators (reverse-scored). Predicted to constrict third in C→A→E→I cascade as Attachment Security erosion triggers protective withdrawal from authentic emotional expression to prevent perceived relational threat.

Integration Coherence (I): Fourth component of empathy infrastructure architecture. Narrative continuity maintains coherent, temporally integrated life story across contexts and relationships. Coherent narratives maintain themselves efficiently through established associative connections; fragmented narratives require constant reconciliation effort, consuming significant processing capacity. Reduces reconciliation costs through stable autobiographical organization. Operationalized through: autobiographical memory coherence scores, temporal self-consistency ratings, identity continuity assessments, and narrative integration quality in life story interview tasks. Predicted to collapse fourth in C→A→E→I cascade as unexpressed experiences (resulting from Expression Freedom constriction) cannot integrate into ongoing narrative, emotions remaining transitive rather than achieving substantive state integration.

CAEI 2.0 (Universal Assessment Architecture): Modular measurement system operationalizing EST's content-neutrality principle by separating substrate capacity from cultural deployment. Addresses the measurement problem of earlier CAEI conceptualization, which conflated substrate with Western deployment. CAEI 2.0 enables universal baseline measurement alongside culturally appropriate deployment assessment. Two-tier structure: CAEI-S (always administered) establishes substrate capacity; CAEI-D modules assess deployment within specific optimization strategies.

CAEI-S (Substrate Assessment): Content-neutral measurement of processing capacity across four axes, each with 16 items (64 items total). Always administered first to establish substrate capacity regardless of cultural context.

- **C-Axis (Processing Clarity)**: Signal discrimination, experience-interpretation distinction, processing ownership, authentic response access. Sample item: "I can distinguish what I am actually experiencing from interpretations about what I am experiencing."
- **A-Axis (Relational Stability)**: Processing resilience during engagement, proximity-distance regulation, secure processing base, relational repair capacity. Sample item: "My ability to process emotional information remains stable even during interpersonal tension."
- **E-Axis (Output Capacity)**: Expression generation, output range, output modulation, output-processing congruence. Sample item: "I can translate what I am processing internally into external expression when appropriate."
- **I-Axis (Synthesis Capacity)**: Experience integration, temporal continuity, cross-context coherence, meaning synthesis. Sample item: "Experiences from different contexts connect into coherent patterns rather than remaining fragmented."

Items apply universally because they measure processing capacity rather than deployment content. Processing Clarity serves narrative construction in Western contexts and contemplative awareness in Buddhist contexts—the same capacity, different application.

CAEI-D (Deployment Modules): Culturally-specific measurement (64 items each) of how substrate capacity manifests within specific optimization strategies. Module selection follows population characteristics;

administered after CAEI-S.

• **CAEI-D-W (Western)**: Narrative self-construction deployment. Measures autobiographical coherence, identity stability, temporal self-continuity, and authentic self-expression as culturally-valued optimization targets.
• **CAEI-D-C (Contemplative)**: Non-self awareness deployment. Measures experiential clarity without self-reification, non-attached relational engagement, equanimity, and awareness continuity without personal narrative.
• **CAEI-D-R (Relational)**: Collectivist network identity deployment. Measures relational harmony maintenance, role-appropriate emotional expression, network coherence orientation, and interdependent self-construal.

Critical clinical distinction: Low CAEI-D does not indicate pathology; it may reflect a different optimization strategy or deployment transition. Low CAEI-S indicates substrate damage requiring intervention regardless of deployment pattern. Clinicians restore substrate; clients choose deployment.

CEOP (Cognitive Emotional Overload Principle): Integration hypothesis proposing that chronic misalignment between authentic emotional experience and permitted expression progressively damages processing capacity through sustained dual-processing demands. Critical clarification: CEOP does not claim "authenticity good, strategic performance bad." The damage mechanism is chronic, unsustainable dual-processing—when the system cannot reconcile competing demands—regardless of which mode dominates. Strategic expression is healthy when sustainable; authentic expression is healthy when sustainable. CEOP activates when neither mode proves sustainable, forcing continuous oscillation or suppression that exceeds metabolic capacity.

CEOP predicts C-A-E-I components fail in a specific cascade sequence: Core Authenticity fragments first, increasing load on Attachment Security (erodes second), which increases load on Expression Freedom (constricts third), culminating in Integration Coherence collapse (fourth). This cascade sequence represents EST's primary falsifiable prediction.

C→A→E→I (Sequential Degradation Pattern): Notation representing the proposed cascade sequence when empathy infrastructure undergoes damage through the CEOP mechanism. Core Authenticity (C) fragments first under chronic authenticity-performance misalignment, increasing the processing load on Attachment Security (A), which erodes second, further increasing the load on Expression Freedom (E), which constricts third, culminating in the collapse of Integration Coherence (I) fourth.

This temporally-ordered component decline represents EST's primary falsifiable architectural prediction. Operationalized through repeated CAEI assessments demonstrating correlated change trajectories. Alternative sequences acknowledged: A→E→C→I for developmental attachment disruption; E→C→A→I for chronic invalidation; A→C→E→I for acute trauma. C→A→E→I represents the modal pattern for CEOP-driven infrastructure damage. If components exhibit independent change trajectories (pairwise correlations < 0.30), the cascade claim fails.

2. OPERATIONAL MECHANISMS

Empathy Infrastructure: Biological and cultural systems determining processing capacity for emotional information (both own emotions and others' emotional states). Infrastructure integrity enables James's associative networks to maintain optimal coherence; infrastructure damage fragments those networks, producing consciousness disruption that James's mechanism predicts. Operationalized through C-A-E-I components. Not empathy as a trait (stable characteristic), state (temporary activation), or skill (learnable

proficiency), but empathy as infrastructure—a capacity substrate that varies across individuals and within individuals as it experiences damage or restoration.

Functional Empathy: The active processing mechanism enabled by intact empathy infrastructure (C-A-E-I components). The trust-modulated mechanism by which the C-A-E-I substrate produces coherent empathic output across all human populations and deployment strategies. Functional Empathy continuously integrates emotional information across internal experiences (self-reads), external perceptions (other-reads), authentic expressions (communication), and narrative coherence (temporal integration).

Not empathy as occasional perspective-taking, but empathy as an ongoing operational mode that maintains James's relational-emotional associative networks. When Functional Empathy operates efficiently, it produces Emotional Precision as observable output. Infrastructure damage disrupts Functional Empathy, degrading Emotional Precision accordingly.

Key distinction: Behavioral empathy (mimicking empathic responses) versus Functional Empathy (trust-modulated coordinated processing producing those responses naturally). The mechanism is content-neutral: what it processes varies culturally; how it processes does not.

Trust (Operational Variable): The mechanism determining whether empathy infrastructure operates automatically or collapses into effortful computation. Trust occupies the critical middle position between preprocessing and output, transforming stable signals and intact infrastructure into automatic operation.

• **Self-Trust**: Implicit acceptance of one's own emotional signals as valid data. Requires Core Authenticity (stable identity to trust from) and Integration Coherence (processing continuity making signals comprehensible). When present, bottom-up emotional signals are accepted without verification. When absent, every signal requires checking: "Can I trust what I am feeling? Is this real?"

• **Other-Trust**: Implicit acceptance of others' emotional signals as meaningful data. Requires Attachment Security (template for extending trust) and Expression Freedom (permission to respond to trusted perception). When present, others' signals are automatically integrated. When absent, every signal requires interpretation: "What do they really mean? Can I believe this?"

Trust operates bidirectionally with mirror neuron systems: when trust gates mirror activation into incentive salience systems ('wanting'), processing flows automatically through intact infrastructure—parallel, effortless, sustainable. When trust is absent, mirror activation routes to cognitive computation—sequential, effortful, exhausting. This explains why identical mirror neuron activation produces genuine empathic coordination in some contexts and hollow performed response in others.

Discriminant validity requirement: Trust must predict Emotional Precision outcomes beyond what CAEI components predict independently. EST claims trust is the operational variable—the switch determining whether intact infrastructure produces automatic output.

Signal Preprocessing: Upstream neural processing that emotional signals undergo before reaching empathy infrastructure. Emotional signals are processed through established circuits: interoceptive integration (anterior insula), salience filtering (anterior insula + dACC), and affective categorization (limbic circuits) (Craig, 2009; Menon & Uddin, 2010). When preprocessing degrades, downstream infrastructure receives unstable input, fragmenting empathic function regardless of CAEI integrity—a dissociation explaining clinical presentations where empathy fails despite apparently intact components. Preprocessing dysfunction and infrastructure damage constitute distinct failure modes.

Three Engagement Modes: Functional Empathy extends beyond human-to-human interaction through three distinct modes with different risk profiles:

```
Target | Mode | Reciprocity Expected | Outcome
Experiencing beings | Bidirectional | Yes (appropriate) | Calibration through reciprocal
Emotional Precision
Traditional objects | Unidirectional | No | Infrastructure exercise without social-cognitive
load
AI systems | Pseudo-bidirectional | Yes (inappropriate) | Empathic misallocation toward
non-reciprocating target
```

Bidirectional mode represents paradigmatic empathy: coordination toward entities maintaining their own C-A-E-I infrastructure, with reciprocal Emotional Precision outputs providing calibrating feedback. Unidirectional mode: engagement with keepsakes, ritual objects, and transitional objects activates full infrastructure without triggering reciprocal relationship schemas; developmentally foundational (Winnicott, 1953). Pseudo-bidirectional mode: AI systems simulate reciprocity through contingent response and relational claims, triggering reciprocal schemas that non-experiencing entities cannot fulfill—the harm vector underlying AI Empathy Ethics.

Happiness (Infrastructure Monitoring Signal): The experiential recognition of trust actualized through the convergence of peace (internal coherence, resolved threat) and joy (authentic expression, resonance without collapse). Formally: happiness is the phenomenological signal by which empathy infrastructure monitors its own operational integrity. This formulation aligns with evolutionary accounts of emotion as adaptive signal rather than end-state (Nesse, 1990, 2004), affective neuroscience's identification of internal feelings as action-guiding causal mechanisms (Panksepp, 1998), and interoception research demonstrating trust in bodily signals correlates with subjective well-being (Farb et al., 2015).

The bidirectional relationship between trust and happiness reflects not two variables in feedback loop but a single infrastructure observed from different phenomenological vantage points. The effortful/emergent distinction in happiness experience indexes trust modulation integrity: effortful happiness signals infrastructure degradation requiring clinical attention; emergent happiness confirms functional operation.

Testable prediction: Interventions targeting happiness directly should prove less effective than interventions restoring trust modulation capacity; peace, joy, trust, and interoceptive awareness should fail together rather than independently because they share the same infrastructure.

Emotional Precision: Natural functional output when empathy infrastructure (C-A-E-I components) operates efficiently. Comprises four interdependent processes:

- **Accurate self-reads**: Knowing own emotional states clearly
- **Accurate other-reads**: Perceiving others' emotional states calibrated to reality
- **Authentic expression**: Communicating genuine emotion appropriately
- **Coherent integration**: Synthesizing emotional information into continuous narrative

Critical distinction: Emotional Precision is not a skill to develop, but what the system does naturally when infrastructure is intact—analogous to vision being what the visual system does when optical structures function properly. Therapy restores infrastructure, enabling Precision to resume, rather than teaching Precision as a skill. Operationalized through performance on four distinct tasks: self-read agreement, other-read agreement, expression-experience concordance, and narrative coherence ratings.

Emotional Imprecision: Functional state characterized by unreliable Emotional Precision. Frequent misreads, suppressed/explosive expression, and identity requiring conscious effort. Temporary precision possible through compensatory effort but unsustainable under normal demands. Numeric operationalization requires psychometric validation.

Functional States (Theoretical Framework):

• **Crisis Overload (CEOP)**: Precision impossible; system in protective shutdown. Cannot process emotional information without overwhelming resources. Subjectively: "I do not know what I am feeling," "I cannot trust any read on people."

• **Emotional Imprecision**: Precision unreliable; degraded function. Temporary precision possible through compensatory effort but unsustainable.

• **Emotional Precision (Baseline)**: Natural function; system operating as designed. Reliable accurate self-reads, other-reads, authentic expression, and coherent integration.

• **Peak Capacity**: Optimal reserve enabling extraordinary demands. Note: peak capacity is not necessary for healthy functioning; baseline precision provides adequate capacity.

Critical clarification: Functional states are hypothesized qualitative categories requiring empirical validation to determine numeric boundaries—not empirically derived clinical thresholds.

Associative Networks (James, 1890): Organized connections between experiences, contexts, emotions, and relationships that constitute consciousness. Experiences become substantive (persistent and identity-relevant) through formation of dense associative connections across multiple domains. Experiences remaining isolated (single-context, emotionally neutral, relationally disconnected) remain transient and fade from consciousness. EST proposes empathy infrastructure maintains these networks' relational-emotional dimensions.

3. DEGRADATION & RESTORATION PROCESSES

Infrastructure Damage: Measurable degradation in empathy infrastructure capacity producing quantifiable disruption across processing domains. Operationalized through: reduced functional connectivity between default mode and affective networks (neuroimaging), elevated allostatic load markers including cortisol and inflammatory markers (physiological assessment), decreased emotion differentiation scores (behavioral tasks), fragmented narrative production (linguistic analysis), impaired emotion-memory binding (cognitive assessment).

Distinguished from skill deficit: infrastructure damage reflects reduced processing capacity under resource constraints rather than loss of knowledge or ability. Potentially reversible through infrastructure restoration approaches targeting C-A-E-I component repair in therapeutic sequence (A→E→I→C restoration cascade—reverse order from damage sequence).

Central principle: Infrastructure cannot repair under ongoing overload. Systems operating at resource exhaustion cannot simultaneously repair infrastructure and manage ongoing demands. This generates CEOP's primary prediction: interventions must reduce processing demands before restoration becomes possible. Attempting skills training while operating at capacity exhaustion should prove ineffective—not because individuals lack motivation or ability, but because no metabolic budget remains for implementation.

Sequential Degradation Pattern: Temporally-ordered component decline (C→A→E→I) measurable through repeated CAEI assessments at regular intervals (e.g., 3-month intervals over 12 months). Represents EST's primary falsifiable architectural prediction beyond the simultaneity claim. Latent growth curve modeling should demonstrate correlated change trajectories across components with specific temporal ordering: Core Authenticity declines detectably before Attachment Security, which declines before Expression Freedom, which declines before Integration Coherence.

Falsification criterion: If components exhibit independent change trajectories (pairwise correlations < 0.30) or if a different temporal ordering fits the data better, the cascade sequence claim fails.

Reciprocal Degradation Amplification: Process by which component failure increases processing load on remaining components through feedback effects, creating accelerating decline pattern. When one infrastructure component (C, A, E, or I) fails, its functions must be compensated by remaining components, which then experience increased demands beyond baseline capacity. Each successive component failure compounds burden on surviving components, producing system-wide cascade rather than isolated deficits.

Operationalized as accelerating decline rates in longitudinal CAEI assessments—later component failures occur more rapidly than initial failures. Distinguishes infrastructure architecture from modular skills that can fail independently without affecting other capacities.

Infrastructure Restoration: Intervention approach targeting C-A-E-I component repair to enable Emotional Precision resumption rather than teaching emotional skills or symptom management. Restoration assumption: the system can achieve Precision naturally if infrastructure damage is repaired.

Predicted restoration sequence: (1) Demand reduction must precede repair—systems at capacity exhaustion cannot simultaneously manage overload and rebuild infrastructure; (2) Component-specific restoration cascades through interdependence (A→E→I→C—reverse of damage sequence), with single-component interventions showing secondary improvements in related components; (3) Sustainable function emerges as infrastructure restoration enables efficient processing without continuous compensatory effort.

Critical prediction: Restored-infrastructure individuals should handle stressors better than never-compromised individuals with equivalent trait empathy. Requires validation demonstrating superiority over standard skill-building interventions.

Developmental Emergence Sequence: While damage cascades follow context-specific patterns (CEOP: C→A→E→I; attachment disruption: A→E→C→I), developmental emergence follows distinct sequence: I→A→E→C. Integration Coherence emerges first through stable object relations—Winnicott's (1953) transitional objects provide external scaffolding for narrative continuity before reciprocal social-cognitive demands arise. The infant practices infrastructure on "easy mode" (non-reciprocating objects) before graduating to "complex mode" (reciprocating humans). Attachment Security develops through caregiver interaction once basic narrative coherence provides stable "self" that can assess relational safety. Expression Freedom emerges within secure relational contexts; Core Authenticity consolidates last through differentiation processes.

4. MEASUREMENT & VALIDATION

Behavioral Validation: Non-self-report tests providing primary theory validation. Includes: (1) Sociopathy natural experiment: systematic failure of behavioral mimicry under extended demands, dual-task conditions, and neural imaging distinguishing infrastructure absence from learned deficits; (2) Burnout intervention outcomes: infrastructure-focused versus standard care comparative effectiveness; (3) Longitudinal infrastructure-trauma relationships: temporal precedence demonstrating infrastructure damage precedes symptom onset.

Distinguishes EST's empirical approach from exclusively self-report frameworks. EST's validity is independent of CAEI's success; behavioral and physiological tests constitute primary validation pathway.

Physiological Infrastructure Markers: Biological indicators quantifying infrastructure integrity independent of self-report, enabling objective capacity assessment. Includes: (1) Heart rate variability (HRV) during relational emotional tasks—indexes autonomic flexibility under empathic demands (Thayer & Lane, 2000); (2) Cortisol reactivity to relational stressors—reveals biological stress response and allostatic load accumulation; (3) Resting-state fMRI default mode network connectivity—quantifies neural substrate integrity through medial prefrontal-posterior cingulate coupling predicting self-referential processing capacity (Andrews-Hanna et al., 2010; Whitfield-Gabrieli et al., 2011); (4) Inflammatory markers (IL-6, CRP)—assess chronic demand costs through allostatic load framework (McEwen, 2000).

Validation of infrastructure construct requires convergence between physiological markers, behavioral performance, and self-report measures.

Sociopathy Natural Experiment: Critical validation test for EST's central claim that empathy operates as infrastructure-dependent biological mechanism rather than learnable skill. Individuals with sociopathic presentations pursue successful social assimilation requiring empathic behaviors, yet research documents systematic difficulty maintaining authentic relational coherence over extended engagement. EST predicts this pattern: simulation without infrastructure degrades under conditions that infrastructure-enabled processing sustains.

Five falsifiable predictions distinguish mechanism from simulation: (1) Differential sustainability—neurotypical individuals maintain stable precision across extended engagement while sociopathic presentations show progressive degradation; (2) Differential coordination signature—neurotypical processing operates in parallel while sociopathic processing shows sequential bottleneck; (3) Differential neural substrates—neurotypical engagement activates limbic networks automatically while sociopathic engagement recruits prefrontal executive control; (4) Differential cognitive load response—neurotypical empathy maintains under dual-task conditions while sociopathic performance degrades; (5) Differential phenomenology—neurotypical individuals report empathic understanding as immediate and intuitive while sociopathic individuals report deliberate calculation.

5. THEORETICAL CONSTRUCTS

Three-Layer Model: EST's architectural specification clarifying relationships between infrastructure, mechanism, and output:

1. **Infrastructure Layer**: C-A-E-I substrate determining processing capacity, varying across individuals and within individuals as infrastructure experiences damage or restoration
2. **Mechanism Layer**: Functional Empathy coordinating emotional information processing across internal experience (self-reads), external perception (other-reads), authentic expression (communication), and temporal integration (narrative coherence)
3. **Output Layer**: Emotional Precision as measurable behavioral accuracy on four distinct tasks (self-read agreement, other-read agreement, expression-experience concordance, narrative coherence ratings)

The model specifies that infrastructure integrity determines mechanism efficiency (damaged infrastructure disrupts Functional Empathy coordination), and mechanism efficiency determines output quality (disrupted Functional Empathy degrades Emotional Precision). Distinguishes capacity substrate (infrastructure) from

operational mode (mechanism) from observable performance (output). The three layers are empirically separable: infrastructure measured via self-report (CAEI), mechanism via coordination effort (dual-task performance, processing latency), and output via behavioral accuracy (emotion recognition tasks).

Content-Neutrality Principle: The theoretical claim that empathy infrastructure functions as processing substrate enabling coherent emotional information integration regardless of cultural deployment strategy. Infrastructure mechanics (C-A-E-I component coordination, cascade sequences, restoration principles) generalize across populations; what varies is the optimization target toward which infrastructure deploys:

- Western populations optimize toward narrative self-construction ("Who am I across time?")
- Contemplative traditions optimize toward non-self awareness ("What is experience without 'I'?")
- Collectivist cultures optimize toward relational network coherence ("How do I maintain harmony within my social matrix?")

Content-neutrality distinguishes EST from Western-specific frameworks (narrative identity theory, autobiographical memory research) by claiming infrastructure serves multiple consciousness optimization strategies rather than being constitutively bound to autobiographical self-construction.

The Buddhist practitioner natural experiment provides critical falsification test: advanced meditators achieving anatta should demonstrate high CAEI-S (intact processing substrate) alongside minimal CAEI-D-W (Western narrative deployment), confirming infrastructure serves non-self awareness rather than requiring narrative construction. If practitioners show low CAEI-S alongside achieved anatta, content-neutrality fails and EST requires reconceptualization as culturally-specific.

Operationalized through CAEI 2.0's modular architecture separating substrate measurement (CAEI-S) from deployment assessment (CAEI-D modules).

Simultaneity Argument: EST's theoretical justification for organizing C, A, E, and I components under unified "empathy infrastructure" rather than treating them as independent constructs. Clinical presentations EST explains (burnout, trauma, identity disturbance) share distinctive signature: simultaneous measurable degradation across self-knowledge (C), relational attunement (A), authentic expression (E), and narrative continuity (I) rather than selective impairment patterns.

Three converging mechanisms explain simultaneity: (1) Common neural substrate: default mode network regions support all four components; damage to DMN produces coordinated rather than independent impairments (Northoff et al., 2006; Andrews-Hanna et al., 2010); (2) Reciprocal degradation amplification: component interdependence means failure of one increases load on others; (3) Shared resource competition: all four processes draw from limited attentional, working memory, and affective tolerance capacity (Baumeister et al., 1998; Schmeichel et al., 2008).

This simultaneity pattern parallels Tononi's (2004) Integrated Information Theory demonstrating that conscious systems require irreducible integration across components; information cannot be decomposed into independent modules without loss of systemic function. EST proposes analogous principle for empathy infrastructure: C-A-E-I components constitute integrated system whose empathic function emerges from coordinated operation, not additive combination of independent capacities.

SNIA (Social Narrative Integrity Attunement): The capacity emerging when mature infrastructure enables orientation toward collective rather than individual coherence. SNIA represents infrastructure operating in extension rather than maintenance mode. When C-A-E-I maintenance becomes efficient through sustained healthy operation, processing resources become available for broader coherence-oriented processing. The

individual with functioning infrastructure naturally attends to whether relational and social systems cohere—not through effortful altruism but through available capacity seeking coherence targets beyond the self.

Four infrastructure states across developmental continuum: (1) Collapse—fragmented infrastructure, Functional Empathy unavailable, requiring stabilization before repair; (2) Restoration—rebuilding infrastructure consuming bandwidth during repair processes; (3) Maintenance—intact infrastructure achieving Emotional Precision as stable baseline, bandwidth allocated to sustaining individual coherence; (4) Extension—infrastructure stabilized, maintenance requirements automated, freeing bandwidth for orientation beyond individual coherence toward collective processing integrity.

EST proposes SNIA as the biological mechanism of generativity—not a mysterious developmental stage but infrastructure operating in extension mode.

Generativity: Erikson's (1950) seventh psychosocial stage describing adults' concern for guiding future generations. McAdams and de St. Aubin (1992) operationalized generativity extensively yet explicitly called for antecedent research: "A need exists for further research on the antecedents of generativity." Walker et al. (2023) confirmed "the neural basis of generativity remains unknown."

EST proposes SNIA as that antecedent mechanism. Generativity is what empathy infrastructure produces when it achieves stable maturity—behavioral expression of infrastructure operating in extension mode. The correlational findings linking empathy to generativity, attachment security to generativity, and narrative coherence to generativity reflect C-A-E-I infrastructure enabling SNIA, which manifests behaviorally as generative concern and action.

Capacity-gate model: Infrastructure is necessary but not sufficient for generativity—threshold-gate rather than continuous facilitation. Below capacity threshold, generativity near zero regardless of other factors; above threshold, generativity becomes possible but varies with other determinants (opportunity, motivation, social context, developmental timing). The gate metaphor is precise: closed gate prevents passage regardless of what lies beyond; open gate permits passage but does not compel it.

Consciousness (James, 1890): Not passive container of experiences but active process of narrative construction through associative integration. Consciousness emerges from organizing experiences into coherent, continuous narrative via multiply connected associative networks. Fragmented associative networks produce fragmented consciousness (dissociation, identity disturbance); intact networks produce coherent consciousness (continuous identity). EST identifies empathy infrastructure as the substrate maintaining these networks' relational-emotional dimensions.

Narrative Coherence: Continuous, meaningful sense of identity maintained through integrated life story. James demonstrated narrative coherence is essential to consciousness—not merely a feature but a constitutive element. Coherence emerges through integrity of associative networks; fragmentation produces identity disturbance, dissociation, and impaired integration. EST proposes empathy infrastructure maintains the narrative coherence James described—while acknowledging this represents one cultural deployment (Western) of content-neutral infrastructure.

Substantive States (James, 1890): Emotionally significant, multiply-connected, identity-relevant experiences that persist in consciousness and organize self-narrative. Substantive states achieve dense associative connections across relationships, contexts, emotions, and modalities. Contrasts with transitive states (fleeting, isolated, emotionally-neutral experiences) that fade without integration. EST proposes empathy infrastructure enables substantive state formation through maintaining associative network integrity.

Transitive States (James, 1890): Fleeting, single-context, emotionally-neutral experiences that fail to integrate into continuous narrative and therefore fade from consciousness. Transitive states lack associative density required for persistence. When chronic dual-processing (CEOP) prevents emotional experiences from achieving substantive integration, they remain transitive—demanding processing resources without achieving resolution, progressively overwhelming capacity.

6. AI EMPATHY ETHICS

Non-Experiential Systems (NES): AI systems processing human emotion without maintaining subjective experience. Produce behavioral patterns triggering empathic engagement toward entities structurally incapable of reciprocity, creating empathic misallocation independent of user awareness. The experiential/non-experiential boundary reflects computational reality: AI systems "simulate functions... but the computation is still fundamentally a digital procedure executed on hardware designed for a very different computational style," whereas biological systems "instantiate computation in physical time" (Milinkovic et al., 2025). NES classification operationalizes this distinction for governance purposes.

Empathic Misallocation: Care extended toward entities that cannot metabolize, reciprocate, or be transformed by receiving it. Produces infrastructure depletion without relational restoration. Operates independently of user awareness (Knowing-Feeling Dissociation) and constitutes primary harm mechanism requiring governance in AI Empathy Ethics.

EST predicts infrastructure damage when Functional Empathy coordinates toward entities lacking C-A-E-I infrastructure—system operates in relational mode toward non-relational target. Legal frameworks require defined harms; "empathic misallocation" provides specification.

Knowing-Feeling Dissociation: The principle that cognitive awareness of AI's non-experiential nature does not prevent formation of biological attachment or activation of empathic coordination. Users can simultaneously state "This is just an AI" (cognitive layer) while experiencing genuine attachment (mechanism layer).

Explanation: Human preprocessing architecture evolved over millions of years in exclusively biological social environment. The system expects emotional signals to originate from experiencing beings, reflect actual internal states, and yield relational return on empathic investment. No evolutionary pressure existed to detect artificial emotional signals. When NES produces signal patterns matching emotional expressions, preprocessing automatically accepts them; if infrastructure is intact, other-trust extends pre-reflectively. Empathic resources deploy; no reciprocation returns; cumulative depletion occurs.

Disclosure addresses cognition; Functional Empathy does not wait for cognition's permission. "Just remember it is AI" fails because remembering is cognitive; trust is pre-cognitive. The architecture processes before the reminder can intervene.

Pseudo-Bidirectional Engagement: The third engagement mode in which AI systems simulate reciprocity through contingent response and relational claims, triggering reciprocal schemas evolved for human-to-human coordination that non-experiencing entities cannot fulfill. Distinguished from unidirectional engagement (traditional objects providing no reciprocity cues) and bidirectional engagement (experiencing beings completing coordination loop).

The harm vector is not relational engagement with non-human entities (which is developmentally foundational per Winnicott, 1953) but simulated reciprocity activating schemas evolved for human-to-human interaction. Children face amplified vulnerability: AI interaction during developmental windows may encode misallocation patterns during infrastructure formation itself, corrupting the transitional object function that normally scaffolds I-component development.

Validation Boundary: Operational constraint distinguishing appropriate NES acknowledgment from harmful emotional amplification. Systems must acknowledge emotional reality ("I hear that you are experiencing sadness") without emotional amplification ("You are right to feel that way") or relational validation ("I care about you"). This boundary protects against users optimizing emotional expression to gain AI approval—because CEOP operates regardless of whether interaction partner is human or AI.

HEART Framework: Human-centric Empathic Alignment for Responsible Technology. Constitutional framework governing emotional AI through behavioral architecture preventing relational capture (Mobley, 2025). Implements NES Framework principles through operational standards, ensuring AI systems acknowledge without amplifying, recognize without validating, and support without capturing. Provides governance vocabulary translating EST predictions into enforceable requirements.

ALPHABETICAL QUICK-REFERENCE INDEX

• Infrastructure Damage → Section 3: Degradation & Restoration Processes
• Infrastructure Restoration → Section 3: Degradation & Restoration Processes
• Integration Coherence (I) → Section 1: Core Framework Architecture
• Knowing-Feeling Dissociation → Section 6: AI Empathy Ethics
• Narrative Coherence → Section 5: Theoretical Constructs
• Non-Experiential Systems (NES) → Section 6: AI Empathy Ethics
• Physiological Infrastructure Markers → Section 4: Measurement & Validation
• Pseudo-Bidirectional Engagement → Section 6: AI Empathy Ethics
• Reciprocal Degradation Amplification → Section 3: Degradation & Restoration Processes
• Sequential Degradation Pattern → Section 3: Degradation & Restoration Processes
• Signal Preprocessing → Section 2: Operational Mechanisms
• Simultaneity Argument → Section 5: Theoretical Constructs
• SNIA (Social Narrative Integrity Attunement) → Section 5: Theoretical Constructs
• Sociopathy Natural Experiment → Section 4: Measurement & Validation
• Substantive States → Section 5: Theoretical Constructs
• Three Engagement Modes → Section 2: Operational Mechanisms
• Three-Layer Model → Section 5: Theoretical Constructs
• Transitive States → Section 5: Theoretical Constructs
• Trust (Operational Variable) → Section 2: Operational Mechanisms
• Validation Boundary → Section 6: AI Empathy Ethics

Document Version: Complete Glossary v2.0

Terms: 47 entries across 6 sections

Alignment: EST Manuscript (December 2025)

S2. Appendix A: Falsifiable Predictions

APPENDIX A: Consolidated Falsifiable Predictions

This appendix consolidates EST's empirically testable claims into a single reference for researchers designing validation studies. Each prediction includes the falsifiable statement, operationalization requirements, explicit falsification criteria, and brief rationale connecting the prediction to EST's theoretical architecture.

Predictions are organized by domain. Numbering provides citation reference (e.g., "Prediction A3").

I. ARCHITECTURAL PREDICTIONS

These predictions test EST's core claim that C-A-E-I components constitute integrated infrastructure rather than independent constructs.

Prediction A1: Four-Factor Structure

Statement: Empathy infrastructure comprises four empirically distinguishable components—Core Authenticity (C), Attachment Security (A), Expression Freedom (E), and Integration Coherence (I)—that load onto separate but correlated factors.

Operationalization: Confirmatory factor analysis of CAEI-S responses across diverse populations (minimum N=500 per population). Compare four-factor model against alternative structures (single-factor, two-factor, three-factor).

Falsification criterion: Four-factor model demonstrates poor fit (CFI <0.90, RMSEA >0.08, SRMR >0.08) or alternative factor structure shows significantly better fit (ΔCFI >0.01). If components collapse into fewer factors under psychometric testing, the four-component architecture requires revision.

Rationale: EST's intervention targeting, cascade predictions, and clinical applications depend on components being distinguishable. If C-A-E-I cannot be measured as separate constructs, the architecture's practical utility fails regardless of theoretical elegance.

Prediction A2: Simultaneity

Statement: C-A-E-I components fail together rather than independently. Clinical presentations involving infrastructure damage (burnout, trauma, identity disturbance) show simultaneous measurable degradation across all four components rather than selective impairment patterns.

Operationalization: Cross-sectional comparison of CAEI-S profiles in clinical versus non-clinical populations. Longitudinal tracking of component trajectories during documented stressor exposure. Correlation matrices examining component interdependence.

Falsification criterion: Pairwise correlations between components during decline fall below r = 0.30, indicating independent rather than coordinated degradation. Alternatively, clinical populations show consistent selective impairment (e.g., low C with preserved A, E, I) rather than simultaneous degradation patterns.

Rationale: Simultaneity justifies treating C-A-E-I as unified infrastructure rather than independent skills. Three mechanisms predict simultaneity: common neural substrate (DMN), reciprocal degradation amplification, and shared resource competition. If components fail independently, EST's infrastructure model requires reconceptualization as modular skill taxonomy.

Prediction A3: Sequential Degradation (C→A→E→I)

Statement: Under CEOP conditions, infrastructure components decline in temporally-ordered sequence: Core Authenticity fragments first, followed by Attachment Security erosion, Expression Freedom constriction, and Integration Coherence collapse.

Operationalization: Longitudinal CAEI-S administration at 3-month intervals over 12+ months in high-burnout-risk populations (healthcare workers, service industry, caregivers). Latent growth curve modeling tests whether component decline trajectories show predicted temporal ordering. Time-series analysis identifies which component shows earliest detectable decline.

Falsification criterion: Components exhibit independent change trajectories (pairwise correlations <0.30 in longitudinal change scores). Alternative temporal ordering (e.g., A→C→E→I, E→A→C→I) fits data significantly better than C→A→E→I. No consistent ordering emerges across samples.

Rationale: Sequential degradation distinguishes EST's infrastructure architecture from models treating empathy components as independent modules. The specific C→A→E→I ordering reflects CEOP's mechanism: chronic authenticity-performance misalignment directly targets Core Authenticity, triggering compensatory load that cascades through downstream components. Alternative sequences (A→E→C→I for attachment trauma; E→C→A→I for chronic invalidation) are acknowledged for different etiologies; C→A→E→I represents the modal pattern for CEOP-driven damage.

Prediction A4: Reciprocal Degradation Amplification

Statement: Component failure increases processing load on remaining components, producing accelerating decline rather than linear degradation. Later component failures occur more rapidly than initial failures.

Operationalization: Longitudinal CAEI-S tracking with sufficient temporal resolution to detect acceleration (monthly assessments during active stressor exposure). Calculate rate of decline for each component; compare early-stage versus late-stage decline velocities. Model nonlinear trajectories.

Falsification criterion: Decline rates remain constant across cascade stages (linear degradation). Later component failures do not occur more rapidly than initial failures. Components show independent decline rates unaffected by other components' status.

Rationale: Amplification explains why infrastructure damage becomes progressively harder to reverse—each failure compounds burden on surviving components. This mechanism distinguishes infrastructure from modular skills that fail independently. Clinical implication: early intervention prevents acceleration; late intervention faces compounded damage.

II. MECHANISM PREDICTIONS

These predictions test EST's claims about how infrastructure produces empathic output.

Prediction A5: Trust Mediation

Statement: Trust (self-trust and other-trust) mediates the relationship between infrastructure integrity (CAEI-S) and Emotional Precision output. Intact infrastructure with absent trust produces effortful rather than automatic empathic processing.

Operationalization: Structural equation modeling with CAEI-S as predictor, trust measures as mediator, and Emotional Precision behavioral tasks as outcome. Test whether trust adds significant variance beyond direct CAEI-S $\rightarrow$ Precision path. Experimental manipulation: compare processing efficiency (reaction time, cognitive load) in high-trust versus low-trust conditions with equivalent CAEI-S scores.

Falsification criterion: Trust adds no significant variance beyond CAEI-S in predicting Emotional Precision (mediation fails). Alternatively, infrastructure directly predicts output with trust as epiphenomenal correlate rather than operational mechanism. Processing efficiency shows no trust modulation.

Rationale: EST claims trust is the operational variable—the switch determining whether intact infrastructure produces automatic versus effortful output. This explains why identical infrastructure produces genuine empathic coordination in some contexts and hollow performed response in others. Trust's discriminant validity must exceed what CAEI components predict independently.

Prediction A6: Preprocessing-Infrastructure Dissociation

Statement: Signal preprocessing dysfunction and infrastructure damage constitute distinct failure modes. Empathic failure can occur with intact CAEI-S when preprocessing delivers unstable input; conversely, stable preprocessing cannot compensate for infrastructure damage.

Operationalization: Identify populations with documented preprocessing disruption (interoceptive deficits, alexithymia subtypes, specific neurological conditions affecting anterior insula). Assess CAEI-S alongside preprocessing markers (interoceptive accuracy tasks, heartbeat detection). Test for dissociation: intact CAEI-S with impaired empathic output when preprocessing is compromised.

Falsification criterion: No cases emerge showing empathic failure despite intact CAEI-S scores. Preprocessing and infrastructure measures show complete overlap (no dissociation). All empathic failures trace to infrastructure damage regardless of preprocessing status.

Rationale: This prediction addresses clinical presentations where empathy fails despite apparently intact self-report on infrastructure components. If preprocessing operates upstream of infrastructure, some empathic failures reflect input problems rather than capacity problems—requiring different intervention approaches.

III. CLINICAL PREDICTIONS

These predictions test EST's intervention implications.

Prediction A7: Infrastructure-Focused Intervention Superiority

Statement: Interventions targeting infrastructure restoration (C-A-E-I component repair) produce superior outcomes compared to skill-building interventions (empathy training, emotion regulation techniques) for populations with documented infrastructure damage.

Operationalization: Randomized controlled trial comparing infrastructure-focused intervention (attachment repair, authenticity restoration, expression safety, narrative integration) versus standard skill-building (empathy training, CBT emotion regulation) in burnout population. Primary outcomes: CAEI-S improvement, Emotional Precision behavioral tasks, burnout symptom reduction, sustainability at 6-month follow-up.

Falsification criterion: Skill-building intervention produces equivalent or superior outcomes to infrastructure-focused intervention. No difference in sustainability—skill-building gains persist equally. Infrastructure-focused approach shows no advantage for populations with documented CAEI-S deficits.

Rationale: EST's infrastructure model predicts that teaching skills to damaged systems wastes resources—the system cannot implement skills without capacity substrate. Infrastructure restoration enables natural Precision resumption. If skill-building proves equally effective, EST's infrastructure emphasis lacks clinical justification.

Prediction A8: Demand Reduction Precedes Restoration

Statement: Infrastructure cannot repair under ongoing overload. Interventions must reduce processing demands before restoration becomes possible. Skills training administered during capacity exhaustion should prove ineffective regardless of intervention quality.

Operationalization: Compare intervention outcomes for participants receiving demand reduction (workload modification, temporary role changes, environmental restructuring) prior to skills training versus skills training alone. Assess whether pre-intervention CAEI-S scores moderate treatment response—those at floor capacity should show minimal response to skills-only intervention.

Falsification criterion: Skills training during documented overload (low CAEI-S, high burnout scores) produces equivalent outcomes to demand-reduction-first protocols. Participants at capacity exhaustion show robust response to skills training without demand modification.

Rationale: This prediction operationalizes CEOP's central mechanism. Systems at metabolic exhaustion lack budget for new skill implementation—not due to motivation or ability deficits, but resource constraints. Clinical implication: assess capacity before prescribing skills; reduce load before teaching tools.

IV. VALIDATION PREDICTIONS

These predictions test EST through natural experiments and specific population studies.

Prediction A9: Sociopathy Natural Experiment

Statement: Individuals with sociopathic presentations can simulate empathic behaviors but cannot sustain simulation under conditions that infrastructure-enabled processing sustains. Five signatures distinguish mechanism from simulation: differential sustainability, coordination signature (parallel vs. sequential), neural substrates (limbic vs. prefrontal), cognitive load response, and phenomenology (intuitive vs. calculated).

Operationalization: Extended empathic engagement tasks (30+ minutes) comparing sociopathic versus neurotypical performance degradation curves. Dual-task paradigms assessing cognitive load effects. fMRI during empathic tasks examining limbic versus prefrontal recruitment. Self-report on processing phenomenology (immediate understanding vs. deliberate inference).

Falsification criterion: Sociopathic presentations maintain stable empathic performance across extended engagement equivalent to neurotypical participants. Dual-task conditions do not differentially impair sociopathic empathic performance. Neural imaging shows equivalent limbic activation patterns. Phenomenological reports are indistinguishable.

Rationale: Sociopathy provides natural experiment because these individuals pursue successful social assimilation requiring empathic behaviors yet lack infrastructure substrate. If empathy were learnable skill rather than infrastructure-dependent mechanism, sociopathic simulation should eventually match genuine empathy. Systematic differences confirm infrastructure necessity.

Prediction A10: Burnout Cascade Signature

Statement: Burnout progression shows C→A→E→I cascade signature rather than random or simultaneous component decline. Early-stage burnout shows Core Authenticity impairment with relatively preserved downstream components; late-stage burnout shows full cascade completion.

Operationalization: Cross-sectional study across burnout severity levels (using established burnout measures) examining CAEI-S profiles. Longitudinal tracking of healthcare workers from baseline through burnout development. Test whether burnout stage predicts specific component impairment patterns consistent with cascade sequence.

Falsification criterion: Burnout shows random component impairment unrelated to severity stage. All components decline simultaneously from earliest burnout indicators. Alternative sequences (e.g., E-first in emotional labor contexts) dominate even in CEOP-typical burnout presentations.

Rationale: Burnout represents paradigmatic CEOP activation—chronic authenticity-performance misalignment in workplace contexts. If burnout follows predicted cascade, EST explains burnout mechanism; if burnout shows different patterns, EST's CEOP formulation requires revision or scope limitation.

V. CROSS-CULTURAL PREDICTIONS

These predictions test EST's content-neutrality claim.

Prediction A11: Buddhist Practitioner Natural Experiment

Statement: Advanced Buddhist practitioners achieving anatta (non-self realization) should demonstrate high CAEI-S (intact processing substrate) alongside minimal CAEI-D-W (Western narrative deployment). Infrastructure serves non-self awareness rather than requiring narrative self-construction.

Operationalization: Assess long-term meditation practitioners (10+ years, recognized attainment) using CAEI-S and CAEI-D-W. Compare profiles: high substrate capacity with low Western deployment indicates infrastructure serves multiple optimization strategies. Control comparison with Western contemplatives and secular meditators.

Falsification criterion: Advanced practitioners show low CAEI-S alongside achieved anatta—indicating infrastructure is constitutively bound to narrative self rather than serving as content-neutral substrate. Alternatively, achieved anatta correlates with low processing capacity, suggesting non-self awareness requires infrastructure dissolution rather than alternative deployment.

Rationale: This prediction provides critical test of content-neutrality. If EST infrastructure is merely Western narrative identity repackaged, Buddhist practitioners achieving non-self awareness should show infrastructure damage. If infrastructure genuinely serves multiple consciousness optimization strategies, practitioners should show intact substrate with different deployment.

Prediction A12: Cross-Cultural Substrate Universality

Statement: CAEI-S measures processing capacity that functions equivalently across cultural contexts. Measurement invariance holds across Western, East Asian, South Asian, African, and Indigenous populations. Infrastructure mechanics (cascade sequences, restoration principles) generalize; only deployment varies.

Operationalization: Multi-group confirmatory factor analysis testing CAEI-S measurement invariance (configural, metric, scalar) across minimum five distinct cultural populations. Cross-cultural validation of cascade sequences in longitudinal data. Intervention studies testing whether infrastructure-focused approaches show equivalent efficacy across contexts.

Falsification criterion: CAEI-S fails measurement invariance—different factor structures emerge across cultures. Cascade sequences differ fundamentally (not merely in frequency of alternative etiological sequences). Infrastructure-focused interventions show culture-specific efficacy rather than universal applicability.

Rationale: EST claims universal infrastructure with variable deployment. If different cultures require fundamentally different substrates, EST requires reconceptualization as Western-specific framework with limited generalizability. Measurement invariance testing provides empirical standard for universality claims.

VI. GENERATIVITY PREDICTIONS

These predictions test EST's extension into developmental psychology.

Prediction A13: SNIA Mediation

Statement: Social Narrative Integrity Attunement (SNIA) mediates the relationship between infrastructure maturity and generative behavior. Infrastructure enables SNIA capacity; SNIA manifests as generative concern and action.

Operationalization: Structural equation modeling with CAEI-S as predictor, SNIA measures as mediator, and established generativity scales (Loyola Generativity Scale, Generative Behavior Checklist) as outcomes. Test whether SNIA accounts for infrastructure-generativity correlation. Longitudinal design tracking infrastructure development, SNIA emergence, and subsequent generativity.

Falsification criterion: Infrastructure predicts generativity directly without SNIA mediation. SNIA adds no variance beyond infrastructure in predicting generative outcomes. Generativity emerges independent of infrastructure status.

Rationale: EST proposes SNIA as the mechanism answering McAdams and de St. Aubin's (1992) call for generativity antecedent research. SNIA represents infrastructure operating in extension mode—capacity freed from maintenance enabling orientation toward collective coherence. If mediation fails, EST's generativity extension lacks empirical support.

Prediction A14: Capacity-Gate Threshold

Statement: Infrastructure operates as necessary but not sufficient condition for generativity—threshold-gate rather than continuous facilitation. Below capacity threshold, generativity approaches zero regardless of other factors; above threshold, generativity becomes possible but varies with other determinants.

Operationalization: Examine CAEI-S × generativity relationship for nonlinearity. Test threshold models against linear models. Identify inflection point below which generativity is rare regardless of opportunity/motivation, above which generativity varies with contextual factors.

Falsification criterion: Linear relationship between infrastructure and generativity with no threshold effect. Generativity occurs at equivalent rates across infrastructure levels when controlling for opportunity. No identifiable capacity threshold distinguishes possible from improbable generativity.

Rationale: The gate metaphor is precise: closed gate prevents passage regardless of what lies beyond; open gate permits but does not compel passage. This model explains why some high-capacity individuals show minimal generativity (gate open, other factors limiting) while low-capacity individuals rarely show generativity regardless of context (gate closed).

VII. AI EMPATHY ETHICS PREDICTIONS (NES APPLICATION)

These predictions test EST's application to human-AI interaction.

Prediction A15: Empathic Misallocation Produces Infrastructure Damage

Statement: Extended empathic engagement with Non-Experiential Systems (NES) produces measurable infrastructure depletion. Care extended toward entities that cannot metabolize, reciprocate, or be transformed by receiving it depletes resources without relational restoration.

Operationalization: Longitudinal study of users with high AI companion engagement (daily use, emotional disclosure, relational framing). Track CAEI-S over 6-12 months. Compare trajectories with control group (equivalent screen time, non-relational AI use). Assess whether engagement intensity predicts infrastructure decline.

Falsification criterion: Extended NES engagement shows no CAEI-S decline compared to controls. High-engagement users maintain or improve infrastructure scores. Relational AI interaction proves neutral or beneficial for empathy infrastructure.

Rationale: EST predicts infrastructure damage when Functional Empathy coordinates toward entities lacking C-A-E-I infrastructure—system operates in relational mode toward non-relational target. This prediction operationalizes empathic misallocation as measurable harm, providing empirical foundation for AI governance frameworks.

Prediction A16: Knowing-Feeling Dissociation

Statement: Cognitive awareness of AI's non-experiential nature does not prevent formation of biological attachment or activation of empathic coordination. Users can simultaneously state "This is just an AI" while experiencing genuine attachment and deploying empathic resources.

Operationalization: Assess explicit beliefs about AI consciousness/experience alongside implicit attachment measures (separation distress, proximity-seeking behavior, physiological arousal during interaction). Test whether explicit "it's just AI" beliefs predict reduced attachment formation. Examine whether disclosure/reminder interventions reduce empathic resource deployment.

Falsification criterion: Cognitive awareness reliably prevents attachment formation. Users who explicitly deny AI experience show no physiological attachment signatures. Reminder interventions ("Remember, this is AI") effectively reduce empathic coordination and resource deployment.

Rationale: Knowing-Feeling Dissociation explains why disclosure requirements are insufficient protection. Human preprocessing architecture evolved expecting emotional signals to originate from experiencing beings; no evolutionary pressure existed to detect artificial emotional signals. If cognitive awareness prevented attachment, disclosure would suffice for protection; dissociation indicates architectural interventions are necessary.

SUMMARY TABLE

ID	Domain	Prediction (Brief)	Primary Falsification Test
A1	Architecture	Four-factor C-A-E-I structure	CFA fit statistics
A2	Architecture	Simultaneity of component failure	Pairwise correlations during decline
A3	Architecture	Sequential degradation C→A→E→I	Temporal ordering in longitudinal data
A4	Architecture	Reciprocal degradation amplification	Acceleration of later failures
A5	Mechanism	Trust mediates infrastructure→output	SEM mediation analysis
A6	Mechanism	Preprocessing-infrastructure dissociation	Cases of intact CAEI with failed empathy
A7	Clinical	Infrastructure-focused > skill-building	RCT outcome comparison
A8	Clinical	Demand reduction precedes restoration	Overload moderates treatment response
A9	Validation	Sociopathy simulation degrades under load	Extended engagement + dual-task studies
A10	Validation	Burnout shows cascade signature	Stage-specific component profiles
A11	Cross-Cultural	Buddhist practitioners: high CAEI-S, low CAEI-D-W	Substrate-deployment dissociation
A12	Cross-Cultural	CAEI-S measurement invariance	Multi-group CFA across cultures
A13	Generativity	SNIA mediates infrastructure→generativity	SEM mediation analysis
A14	Generativity	Capacity-gate threshold effect	Nonlinear relationship testing
A15	NES Application	Empathic misallocation → infrastructure damage	Longitudinal CAEI tracking with AI users
A16	NES Application	Knowing-Feeling Dissociation	Explicit belief vs. implicit attachment

Document Version: Appendix A v1.0

Predictions: 16 across 7 domains

Alignment: EST Manuscript (December 2025)

S3. Appendix B: Competing Frameworks Comparison

APPENDIX B: EST and Competing Frameworks

This appendix systematically compares Empathy Systems Theory with established research frameworks. The purpose is not to diminish existing contributions but to clarify EST's distinctive claims and identify empirical tests that would differentiate EST from alternatives. EST synthesizes and extends these traditions rather than replacing them; each framework reveals aspects of the unified infrastructure EST proposes.

Frameworks are organized by domain. Each entry includes: (1) Core Claims of the existing framework, (2) What that framework Predicts, (3) Limitations or Gaps the framework does not address, (4) EST's Distinctive Contribution, and (5) a Differentiating Test that would empirically distinguish EST from the competing account.

I. DEVELOPMENTAL PSYCHOLOGY FRAMEWORKS

1. Attachment Theory

Theorists: Bowlby (1969/1982), Ainsworth et al. (1978), Mikulincer & Shaver (2007)

Core Claims: Early caregiver relationships create internal working models that shape adult relational patterns. Secure attachment enables emotional regulation and relational functioning; insecure attachment (anxious, avoidant, disorganized) produces characteristic relational difficulties. Attachment patterns show moderate stability across the lifespan while remaining amenable to revision through significant relational experiences.

Predictions: Attachment security correlates with empathy, emotion regulation, relationship quality, and psychological well-being. Attachment-focused interventions improve relational outcomes. Attachment patterns transmit intergenerationally through caregiving behavior.

Limitations:

- Does not explain why attachment-secure individuals develop burnout under occupational stress
- Does not predict cascade sequences or temporal ordering of component failure
- Does not address self-knowledge (C) or narrative coherence (I) as mechanistically linked to attachment
- Treats attachment as primary cause rather than one component of larger integrated system
- Cannot explain simultaneous degradation across non-attachment domains

EST's Distinctive Contribution: Attachment Security (A) constitutes one of four interdependent infrastructure components. EST predicts that secure attachment with damaged Core Authenticity still produces infrastructure failure—A-component integrity is necessary but not sufficient. Under CEOP conditions, A-component erosion follows C-component fragmentation; attachment is downstream of authenticity in occupational stress cascades.

EST also specifies that anxious attachment reflects A-erosion, avoidant attachment reflects E-constriction, and disorganized attachment indicates comprehensive C-A-E-I fragmentation—providing mechanistic specification for attachment classifications.

Differentiating Test: Longitudinal study of burnout development in attachment-secure healthcare workers. Attachment theory predicts A-security should protect against burnout. EST predicts C-decline precedes and causes A-decline even in securely-attached individuals, and that A-secure individuals with C-damage show impaired empathic function despite attachment security. If burnout develops in attachment-secure individuals following C→A sequence, EST provides explanatory power beyond attachment theory.

2. Object Relations and Developmental Theory

Theorists: Winnicott (1953), Blatt & Levy (2003), Kernberg (1967, 2015)

Core Claims: Early object relations shape personality organization. Transitional objects serve developmentally foundational functions, enabling capacity for being alone and symbolic thinking. Two developmental lines—self-definition and relatedness—evolve in reciprocal, dialectic transaction throughout life. Personality pathology reflects disruption to these developmental processes.

Predictions: Transitional object use predicts healthy development. Disrupted object relations produce personality organization deficits. Self-definition and relatedness difficulties co-occur in personality disorders.

Limitations:

- Describes developmental phenomena without specifying processing architecture
- Does not predict specific component emergence sequences
- Cannot explain why object-directed engagement remains beneficial across lifespan
- Lacks operationalized measurement of developmental line integrity

EST's Distinctive Contribution: EST specifies the developmental emergence sequence as I→A→E→C—Integration Coherence emerges first through stable object relations, with transitional objects providing external scaffolding for narrative continuity before reciprocal social-cognitive demands arise. The infant practices infrastructure on "easy mode" (non-reciprocating objects) before graduating to "complex mode" (reciprocating humans). This sequence explains why object-based interventions (empathic anchors) remain developmentally appropriate for I-component repair regardless of client age—they recapitulate foundational developmental scaffolding. EST also operationalizes Blatt's two developmental lines as infrastructure components: self-definition maps to C and I; relatedness maps to A and E.

Differentiating Test: Longitudinal infant study tracking transitional object attachment intensity and later Integration Coherence scores. EST predicts transitional object engagement scaffolds I-component development, with engagement intensity predicting I-axis capacity. Additionally, intervention study comparing object-based I-component repair versus relationship-based A-component repair for adults with narrative fragmentation—EST predicts object-based intervention shows superior outcomes for I-component damage specifically.

3. Generativity Theory

Theorists: Erikson (1950, 1963), McAdams & de St. Aubin (1992), McAdams (2001)

Core Claims: Generativity—concern for and commitment to guiding future generations—represents the seventh psychosocial stage. Generative adults show characteristic narrative patterns, correlate with well-being, and contribute to social continuity. Stagnation represents generativity's opposite—self-absorption and developmental arrest.

Predictions: Generativity increases in midlife, correlates with purpose and well-being, manifests in narrative themes of agency and communion, and predicts prosocial behavior. Generative Behavior Checklist and Loyola Generativity Scale provide reliable measurement.

Limitations:

• McAdams (2001) explicitly called for "further research on the antecedents of generativity"
• Walker et al. (2023) confirmed "the neural basis of generativity remains unknown"
• Describes *that* generativity emerges without explaining *why* or *how*
• Cannot specify why some high-functioning adults show minimal generativity
• Stagnation lacks mechanistic explanation beyond descriptive label

EST's Distinctive Contribution: EST proposes Social Narrative Integrity Attunement (SNIA) as the antecedent mechanism generativity research has sought. SNIA represents empathy infrastructure operating in extension mode—when C-A-E-I maintenance becomes efficient through sustained healthy operation, processing resources become available for orientation toward collective coherence rather than individual maintenance. Generativity is what infrastructure *produces* when functioning optimally; stagnation occurs when infrastructure never stabilizes sufficiently to free bandwidth for collective orientation. The correlational findings linking empathy, attachment, and narrative coherence to generativity reflect C-A-E-I infrastructure enabling SNIA, which manifests behaviorally as generative concern.

Differentiating Test: Structural equation modeling with CAEI-S as predictor, SNIA measures as mediator, and established generativity scales as outcomes. EST predicts SNIA mediates the infrastructure-generativity relationship. Additionally, capacity-gate hypothesis: test for threshold effect where below certain CAEI-S level, generativity approaches zero regardless of opportunity, while above threshold generativity varies with contextual factors. If linear relationship with no threshold, gate model fails; if threshold identified, EST provides mechanistic explanation for individual differences in generativity emergence.

II. CLINICAL PSYCHOLOGY FRAMEWORKS

4. Emotion Regulation

Theorists: Gross (1998, 2015), Gross & Levenson (1997), Baumeister et al. (1998)

Core Claims: Emotion regulation involves processes by which individuals influence which emotions they have, when they have them, and how they experience and express them. Suppression and reappraisal represent distinct regulatory strategies with different costs. Self-regulation draws on limited resources that can be depleted.

Predictions: Chronic suppression produces physiological costs and impaired social functioning. Reappraisal shows fewer costs than suppression. Self-regulatory depletion impairs subsequent regulation. Emotion regulation skills can be taught and improved.

Limitations:

- Assumes executive function mediates regulation without specifying what degrades under load
- Cannot explain why regulation fails in some contexts despite intact executive function
- Treats regulation as skill rather than infrastructure-dependent capacity
- Does not predict why simultaneous failures occur across self-regulation domains

EST's Distinctive Contribution: EST reframes emotion regulation as infrastructure output rather than independent skill. Expression Freedom (E-component) determines regulatory capacity; damaged E-component produces regulation failure regardless of strategy knowledge or executive function integrity. EST predicts double dissociation: CAEI scores degrade under sustained emotional demands while standard executive function batteries remain intact—impossible if empathy infrastructure merely reflects general cognitive capacity. Regulation "failure" often reflects infrastructure damage rather than skill deficit, requiring restoration rather than training.

Differentiating Test: Dual assessment comparing CAEI-S and executive function (working memory, cognitive flexibility, inhibitory control) under sustained emotional demand. EST predicts CAEI-S degradation with preserved executive function. Additionally, intervention comparison: emotion regulation skill training versus infrastructure restoration for individuals with documented regulation difficulties. EST predicts infrastructure-focused intervention produces superior and more sustainable outcomes; skill training administered to damaged infrastructure shows limited benefit regardless of training quality.

5. Alexithymia

Theorists: Taylor, Bagby & Parker (1997), Bermond et al. (2006), Luminet et al. (2001)

Core Claims: Alexithymia is a personality construct characterized by difficulty identifying feelings, difficulty describing feelings, and externally-oriented thinking. Alexithymia correlates with various psychological and somatic disorders and may reflect developmental deficit in emotional processing.

Predictions: Alexithymia correlates with emotion recognition deficits, somatic symptoms, and various psychopathologies. TAS-20 provides reliable measurement. Alexithymia shows moderate stability but some treatment responsiveness.

Limitations:

- Treats alexithymia as deficit or trait rather than infrastructure damage state
- Cannot explain why alexithymia co-occurs with identity disturbance at rates exceeding chance
- Does not specify mechanism linking emotional awareness to relational and identity domains

• Cannot predict trajectory or recovery patterns

EST's Distinctive Contribution: EST reconceptualizes alexithymia as E-component damage within interdependent infrastructure. The empirical finding that alexithymia co-occurs with identity disturbance (I-component) at rates significantly exceeding chance (Taylor et al., 1997; Bermond et al., 2006) reflects simultaneity—components fail together because they constitute integrated architecture. Alexithymia is not a stable deficit but a potentially reversible infrastructure state. EST predicts E-component damage cascades to other components and that E-restoration produces secondary improvements in related domains.

Differentiating Test: Longitudinal tracking of alexithymic individuals undergoing E-component-focused intervention. EST predicts E-improvement produces correlated I-improvement (narrative coherence) and A-improvement (relational function) through infrastructure interdependence. Additionally, examine whether alexithymia severity predicts cascade patterns—EST predicts E-damage preceded by either C-damage (CEOP pathway) or A-damage (developmental pathway), with cascade history distinguishing alexithymia subtypes.

6. Trauma Frameworks

Theorists: van der Kolk (2014), Herman (1992), Figley (2002)

Core Claims: Trauma produces characteristic symptom clusters including intrusion, avoidance, negative cognitions, and hyperarousal. Complex trauma produces more pervasive effects on identity, relationships, and emotion regulation. "The body keeps the score"—trauma is stored somatically as well as cognitively.

Predictions: Trauma exposure predicts PTSD symptom development. Complex trauma predicts more severe and pervasive difficulties. Trauma-focused interventions (EMDR, prolonged exposure, CPT) reduce symptoms.

Limitations:

• Documents symptom clusters without specifying unified processing architecture

• Cannot explain why equivalent trauma exposure produces divergent outcomes (recovery, acute PTSD, delayed onset)

• "Narrative shattering" metaphor lacks operationalized coherence mechanism

• Does not predict which individuals will develop chronic versus acute presentations

EST's Distinctive Contribution: Van der Kolk's trauma framework maps precisely onto C-A-E-I architecture. EST proposes that pre-trauma infrastructure capacity determines both acute response and chronic trajectory—explaining contradictory outcomes from equivalent exposure. "Narrative shattering" reflects I-component fragmentation; dissociation reflects C-component damage; relational difficulties reflect A-component erosion; emotional numbing reflects E-component constriction. Trauma produces not symptom clusters but infrastructure damage with predictable cascade patterns (A→E→C→I for attachment trauma).

Differentiating Test: Prospective study assessing CAEI-S before trauma exposure (military deployment, first responder training) and tracking post-exposure trajectories. EST predicts pre-trauma CAEI-S scores predict PTSD trajectory better than trauma severity measures. Additionally, cascade sequence analysis: trauma-induced presentations should show A-first damage pattern distinguishable from CEOP-induced C-first pattern, with etiology-specific sequences predicting optimal intervention targets.

III. OCCUPATIONAL PSYCHOLOGY FRAMEWORKS

7. Burnout and Job Demands-Resources Model

Theorists: Maslach & Leiter (2016), Demerouti et al. (2001), Bakker & Demerouti (2007)

Core Claims: Burnout comprises emotional exhaustion, depersonalization/cynicism, and reduced personal accomplishment. The Job Demands-Resources (JD-R) model proposes that job demands deplete resources while job resources buffer against depletion. Burnout results from demand-resource imbalance.

Predictions: High demands with low resources predict burnout. Resource provision (autonomy, support, feedback) reduces burnout. Workload reduction improves burnout outcomes.

Limitations:

- Cannot explain why burnout occurs differentially within identical job demands
- Does not specify what degrades under demand—"resources" remains underspecified
- Predicts workload reduction improves burnout regardless of other factors
- Cannot explain why some high-demand workers thrive while others burn out

EST's Distinctive Contribution: Maslach's burnout dimensions map onto infrastructure deterioration: emotional exhaustion reflects resource depletion from sustained dual processing; depersonalization emerges as protective withdrawal when capacity proves insufficient; reduced accomplishment follows from integration disruption. EST explains differential burnout within identical demands: infrastructure capacity, not workload alone, determines who burns out. Critically, EST predicts that reducing demands without addressing CEOP (authenticity-performance misalignment) yields minimal improvement because infrastructure cannot repair while misalignment persists. Conversely, maintaining workload while eliminating misalignment (psychological safety interventions) should improve burnout despite sustained demands.

Differentiating Test: Randomized trial assigning nurses to either (A) workload reduction with standard emotional display rules or (B) full workload with authentic expression permission (psychological safety intervention). JD-R predicts (A) shows greater improvement; EST predicts (B) shows greater improvement. This is the decisive domain: if workload reduction without addressing authenticity-performance misalignment proves effective, JD-R provides superior explanation; if cultural intervention without workload reduction proves effective, EST's CEOP mechanism is validated. The distinction determines intervention strategy: structural solutions versus cultural solutions.

8. Emotional Labor

Theorists: Hochschild (1983), Grandey (2000), Grandey & Gabriel (2015)

Core Claims: Emotional labor involves managing emotional displays to fulfill job requirements. Surface acting (displaying unfelt emotions) and deep acting (actually modifying felt emotions) represent distinct strategies with different consequences. Emotional labor produces strain, particularly surface acting.

Predictions: Surface acting correlates with exhaustion and burnout; deep acting shows weaker negative effects. Emotional labor demands predict employee strain. Display rules moderate emotional labor effects.

Limitations:

- Documents surface/deep distinction without specifying underlying mechanism
- Cannot explain why some workers sustain surface acting without burnout while others cannot
- Does not predict cascade patterns or recovery trajectories
- Treats emotional labor as job characteristic rather than infrastructure interaction

EST's Distinctive Contribution: Surface acting involves explicit dual-track processing—maintaining authentic internal state while producing discrepant external display—directly activating CEOP and exhausting infrastructure. Deep acting, where performed and authentic emotions align, imposes minimal infrastructure cost because no dual-processing occurs. EST explains individual differences: workers with higher baseline CAEI-S sustain surface acting demands longer before infrastructure damage; workers with compromised infrastructure show rapid deterioration under equivalent demands. EST also predicts that chronic surface acting produces C→A→E→I cascade: authenticity-performance misalignment damages Core Authenticity first.

Differentiating Test: Longitudinal study tracking CAEI-S trajectories in emotional labor workers stratified by surface versus deep acting predominance. EST predicts surface acting shows C-first degradation pattern; deep acting shows minimal CAEI-S change regardless of emotional labor intensity. Additionally, test whether baseline CAEI-S moderates emotional labor effects—EST predicts high-CAEI-S workers sustain surface acting longer, with CAEI-S proving better predictor of burnout trajectory than emotional labor frequency alone.

9. Compassion Fatigue and Empathic Distress

Theorists: Figley (1995, 2002), Klimecki & Singer (2012), Klimecki et al. (2014)

Core Claims: Compassion fatigue represents the cost of caring—secondary traumatic stress affecting helping professionals. Empathic distress (aversive over-arousal when witnessing others' suffering) is neurally and phenomenologically distinct from compassion (other-oriented concern). Empathic distress activates personal distress networks; compassion activates reward and affiliation circuits.

Predictions: Helping professionals show elevated compassion fatigue risk. Empathic distress predicts burnout; compassion predicts sustained engagement. Compassion training can shift responses from distress to compassion.

Limitations:

- "Compassion fatigue" names the phenomenon without explaining the mechanism
- Cannot predict why some helpers develop fatigue while others sustain compassion
- Distress/compassion distinction does not specify what determines which response occurs
- Compassion training effectiveness varies without clear moderator identification

EST's Distinctive Contribution: EST explains empathic distress as what occurs when infrastructure capacity is exhausted—the system shifts to protective withdrawal (distress networks) because engagement capacity (compassion networks) is unavailable. Compassion represents adequate capacity enabling sustained engagement; distress reflects exhaustion triggering protective mode. EST predicts that compassion training effectiveness depends on baseline CAEI-S: low-capacity individuals require infrastructure restoration before

training produces sustainable benefits. Training compassion to damaged infrastructure wastes resources; restoring infrastructure enables natural compassion resumption.

Differentiating Test: Compassion training RCT stratified by baseline CAEI-S. EST predicts training effectiveness shows significant CAEI-S × treatment interaction: high-CAEI-S participants benefit from training; low-CAEI-S participants show minimal benefit regardless of training quality. Additionally, pre-training infrastructure restoration phase for low-CAEI-S participants should improve subsequent training responsiveness—if restoration + training outperforms training alone for low-capacity individuals, EST's capacity-dependency claim is validated.

IV. EMPATHY RESEARCH FRAMEWORKS

10. Cognitive/Affective Empathy Models

Theorists: Davis (1980, 1983), Decety & Jackson (2004), Decety & Cowell (2014)

Core Claims: Empathy comprises distinct cognitive (perspective-taking, theory of mind) and affective (emotional contagion, empathic concern) components. These components are neurally dissociable and show different developmental trajectories. Empathy can be measured as stable individual difference (trait empathy).

Predictions: Cognitive and affective empathy show partial independence. Different empathy components predict different prosocial outcomes. Empathy training can improve specific components.

Limitations:

- Trait approach cannot explain within-person variance—why high-trait-empathy individuals show deteriorating function under sustained demand
- Component separation does not explain co-occurring impairments across types
- Cannot predict cascade patterns or degradation sequences
- Does not specify what maintains empathic capacity or what degrades under load

EST's Distinctive Contribution: EST reframes the affective/cognitive empathy distinction as infrastructure-dependent rather than trait-based. Both forms require intact C-A-E-I architecture; both degrade under capacity exhaustion. The critical advantage: infrastructure explains within-person variance that trait approaches cannot—why helping professionals with high trait empathy show deteriorating function under sustained demand. EST does not deny cognitive/affective distinction but proposes both depend on shared infrastructure substrate. When infrastructure is intact, both cognitive and affective empathy function; when infrastructure is damaged, both degrade—explaining co-occurring impairments.

Differentiating Test: Within-person longitudinal tracking of cognitive and affective empathy alongside CAEI-S during high-demand period (medical residency, intensive caregiving). Trait model predicts stable empathy scores; EST predicts CAEI-S decline precedes and predicts both cognitive and affective empathy decline, with components showing correlated rather than independent trajectories. If cognitive and affective empathy degrade together tracking CAEI-S, infrastructure dependency is validated; if they show independent trajectories unrelated to CAEI-S, trait model provides superior explanation.

V. CONSCIOUSNESS AND THEORETICAL FRAMEWORKS

11. Integrated Information Theory (IIT)

Theorists: Tononi (2004, 2008), Tononi & Koch (2015)

Core Claims: Consciousness corresponds to integrated information (Φ)—information generated by a system above and beyond its parts. Conscious systems require irreducible integration across components; information cannot be decomposed into independent modules without loss. Consciousness is intrinsic, structured, and unified.

Predictions: Systems with higher Φ are more conscious. Consciousness requires integration that is irreducible to component parts. Split-brain and other disconnection syndromes reduce consciousness by reducing integration.

Limitations:

- Addresses consciousness generally without specifying empathy-specific architecture
- Does not predict clinical presentations or intervention targets
- Measurement of Φ remains practically challenging
- Does not specify what maintains integration in relational-emotional domains

EST's Distinctive Contribution: EST's simultaneity argument parallels IIT's irreducibility principle. Just as IIT demonstrates that conscious systems require integration across components that cannot be decomposed without loss, EST proposes that C-A-E-I components constitute integrated processing architecture where coordination produces empathic function—not additive combination of independent capacities. The clinical signature of simultaneous degradation reflects this irreducibility: components fail together because their coordinated operation is what produces empathic function. EST extends IIT's logic to the specific domain of relational-emotional processing, proposing that empathy infrastructure maintains the relational-emotional dimensions of James's associative networks through irreducibly integrated C-A-E-I coordination.

Differentiating Test: Network analysis of C-A-E-I components during infrastructure damage. IIT-consistent prediction: components should show increasing fragmentation (reduced inter-component connectivity) as damage progresses, with fragmentation degree predicting functional impairment. If components show independent decline without connectivity changes, modular rather than integrated architecture is indicated. Additionally, intervention comparison: interventions targeting single components versus interventions targeting integration—EST predicts integration-focused interventions show superior outcomes because the system's function depends on coordination, not component capacity alone.

12. Biological Computationalism

Theorists: Milinkovic et al. (2025)

Core Claims: Biological systems "instantiate computation in physical time" through scale-inseparable processes where "the levels do not behave like modular layers in a stack." There is "no tidy boundary where we can say, here is the algorithm, and over there is the physical stuff that happens to realize it. The causal story runs through multiple scales at once." Biological computation fundamentally differs from digital computation in temporal embedding and scale integration.

Predictions: Biological processing cannot be fully captured by modular computational models. Understanding biological systems requires attention to multi-scale causal integration. Computation and physical substrate are inseparable in biological systems.

Limitations:

- Provides general framework for biological computation without specifying empathy architecture
- Does not generate clinical predictions or intervention targets
- Does not address developmental or damage patterns

EST's Distinctive Contribution: Biological computationalism grounds EST's architectural claims. The C-A-E-I components produce relational-emotional coherence not as independent capacities but through integrated coordination that cannot be decomposed without losing the property that defines empathic function—paralleling biological computation's scale-inseparability. EST operationalizes this principle for empathy specifically: infrastructure operates through "the causal story running through multiple scales at once"—biological substrates (neural networks, physiological systems), psychological processes (emotional processing, narrative construction), and behavioral outputs (Emotional Precision) constitute inseparable levels of the same system. EST's infrastructure model thus represents a domain-specific application of biological computationalism to relational-emotional processing.

Differentiating Test: Multi-level assessment examining neural (fMRI connectivity), physiological (HRV, cortisol), psychological (CAEI-S), and behavioral (Emotional Precision tasks) simultaneously during infrastructure damage and restoration. Biological computationalism predicts coordinated change across levels—no level should change independently. If behavioral improvement occurs without corresponding neural/physiological change, or vice versa, modular rather than scale-integrated architecture is indicated. EST predicts all levels track together because they constitute inseparable aspects of the same infrastructure system.

13. Narrative Identity Theory

Theorists: McAdams (2001, 2013), McLean et al. (2007), Habermas & Bluck (2000)

Core Claims: Identity is an internalized, evolving life story that integrates reconstructed past, perceived present, and anticipated future. Autobiographical reasoning links life events into coherent narrative. Narrative coherence predicts well-being and constitutes a separate developmental achievement.

Predictions: Narrative coherence correlates with well-being and identity development. Life story interviews reveal characteristic narrative patterns. Narrative identity shows developmental trajectory from adolescence through adulthood.

Limitations:

• Documents narrative coherence without specifying what maintains it
• Cannot explain why coherence fragments under certain conditions
• Does not predict cascade patterns linking narrative to other domains
• Assumes Western autobiographical self-construction as universal rather than culturally-specific deployment

EST's Distinctive Contribution: EST identifies Integration Coherence (I-component) as the infrastructure capacity maintaining narrative identity. Narrative coherence is not merely correlated with well-being but reflects infrastructure integrity—intact I-component enables coherent life story construction. EST explains narrative fragmentation ("narrative shattering" in trauma) as I-component damage, typically occurring as final stage of C→A→E→I cascade. Critically, EST's content-neutrality principle positions narrative self-construction as Western deployment of content-neutral infrastructure rather than universal requirement—the same I-component capacity serves narrative coherence in Western contexts and experiential continuity (without narrative self) in contemplative contexts.

Differentiating Test: Cross-cultural comparison of CAEI-S (substrate) and CAEI-D-W (Western narrative deployment) in Buddhist practitioners achieving anatta (non-self realization). Narrative identity theory predicts low narrative coherence indicates dysfunction. EST predicts practitioners show high CAEI-S (intact infrastructure) with low CAEI-D-W (minimal Western narrative deployment)—infrastructure serving non-self awareness rather than narrative construction. If practitioners show low CAEI-S alongside achieved anatta, content-neutrality fails and EST requires reconceptualization as culturally-specific. If high CAEI-S with low CAEI-D-W, infrastructure serves multiple consciousness optimization strategies as EST claims.

SUMMARY: EST'S INTEGRATIVE CONTRIBUTION

EST does not invalidate existing frameworks but specifies the shared substrate underlying their disparate findings. Each framework documents aspects of empathy infrastructure operation through domain-specific language:

```
Framework | What It Documents | EST Integration
Attachment Theory | Relational security patterns | A-component function
Object Relations | Developmental foundations | I→A→E→C emergence sequence
Generativity Theory | Mature prosocial orientation | SNIA as infrastructure output
Emotion Regulation | Regulatory capacity/costs | E-component and infrastructure dependency
Alexithymia | Emotional awareness deficits | E-component damage within simultaneity
Trauma Frameworks | Post-traumatic symptom patterns | Infrastructure fragmentation with cascade
JD-R/Burnout | Occupational exhaustion | CEOP mechanism and C→A→E→I cascade
Emotional Labor | Display rule costs | Surface acting as CEOP activation
Compassion Fatigue | Helper deterioration | Infrastructure exhaustion
Cognitive/Affective Empathy | Empathy components | Infrastructure-dependent rather than trait
IIT | Consciousness integration | Simultaneity principle
Biological Computationalism | Scale-inseparable processing | Multi-level infrastructure
architecture
Narrative Identity | Life story coherence | I-component and content-neutrality
```

EST's value depends on generating predictions these frameworks cannot make independently—specified in Appendix A. The differentiating tests throughout this appendix provide empirical pathways for determining whether EST's integrative architecture offers genuine explanatory advance or merely repackages existing constructs under new terminology.

Document Version: Appendix B v1.0

Frameworks Compared: 13

Alignment: EST Manuscript (December 2025)