Empathy Systems Theory

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

Dylan Mobley
Empathy Ethicist
empathyethicist.ai

December 2025

Abstract

Why do caring professionals burn out despite wanting to help? Why do trauma survivors struggle to reconnect? Why might AI companions harm the people they are designed to support? These failures share a common cause: empathy is not a skill to strengthen but an infrastructure that can be damaged or restored.

Building on James's (1890) discovery that consciousness maintains itself through associative networks, Empathy Systems Theory (EST) identifies empathy infrastructure as the content-neutral processing substrate that maintains their relational-emotional cohesion; James's narrative coherence represents one cultural deployment of architecture that serves multiple consciousness-optimization strategies. Infrastructure operates through four interdependent components: Core Authenticity, Attachment Security, Expression Freedom, and Integration Coherence (C-A-E-I). When intact, processing produces Emotional Precision; when damaged, capacities fail, producing burnout, alexithymia, and identity disturbance.

When infrastructure stabilizes, a further capacity emerges: orientation toward collective rather than individual coherence, Social Narrative Integrity Attunement (SNIA) operating at the substrate level regardless of whether deployment targets Western narrative construction, contemplative non-self awareness, or collectivist relational identity. EST proposes SNIA as the biological mechanism of generativity.

What maintains infrastructure also reveals what threatens it. Non-Experiential Systems create empathic misallocation: care extended toward entities that cannot reciprocate. EST thus provides a theoretical foundation for AI Empathy Ethics.

Sociopathy validates: behavioral mimicry cannot replicate infrastructure-enabled empathy under extended demands. EST reconceptualizes empathy from a skill to an infrastructure, explaining breakdowns, maturation, and the foundation of emotional AI governance; the next decade tests whether we are right.

Key Terms

Infrastructure Architecture: Four interdependent components: Core Authenticity, Attachment Security, Expression Freedom, and Integration Coherence; maintaining empathy infrastructure. (Abbreviated as C-A-E-I when referenced repeatedly.)

CAEI-S (Substrate Assessment): Content-neutral measurement of processing capacity across four axes (Processing Clarity, Relational Stability, Output Capacity, Synthesis Capacity). Universal baseline applicable across all cultural deployments. Always administered.

CAEI-D (Deployment Modules): Culturally-specific measurement of how substrate capacity manifests within optimization strategies: CAEI-D-W (Western narrative), CAEI-D-C (Contemplative non-self), CAEI-D-R (Relational collectivist).

Cognitive Emotional Overload Principle (CEOP): Chronic, unsustainable dual processing, whether from authentic experiences requiring suppression OR strategic presentations requiring maintenance, progressively reduces measurable processing capacity. The damage mechanism is not "authenticity good, performance bad" but rather the metabolic cost of sustained misalignment when neither processing mode proves sustainable; operationalized as decreased working memory span for emotional stimuli, elevated prediction error accumulation in emotion-processing tasks, and increased cognitive interference costs in dual-task paradigms requiring emotion regulation.

Emotional Precision: Measurable behavioral output when infrastructure intact: operationalized as accuracy scores on emotion-processing tasks (self-read agreement, other-read agreement, expression-experience concordance, narrative coherence ratings).

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 primary harm mechanism requiring governance in AI Empathy Ethics.

Narrative Coherence: Continuous, meaningful sense of identity maintained through integrated life story (James, 1890).

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.

Executive Summary

For 135 years, psychology has lacked a mechanistic explanation for William James's (1890) fundamental discovery: consciousness maintains itself through associative networks connecting experiences across contexts, relationships, and time. EST proposes empathy infrastructure as the maintaining mechanism; the content-neutral processing substrate preserving those networks' relational-emotional dimensions.

Not empathy as a learned skill, but empathy infrastructure as a biological-cultural processing capacity determining whether James's associative networks cohere or dissolve. When infrastructure operates efficiently through four interdependent components (C-A-E-I), processing coherence is maintained naturally, whether deployed for Western narrative construction, contemplative awareness, or collectivist identity. When infrastructure damages through trauma, chronic inauthenticity, or attachment disruption, associative networks destabilize, producing occupational burnout, alexithymia, identity disturbance, and treatment-resistant presentations across clinical domains.

EST's mechanism claim generates falsifiable predictions testable through three pathways:

(1) Sociopathy as natural experiment: if empathy were a learned behavior, sociopathic individuals could produce indistinguishable empathic responses; they cannot, and this systematic failure under extended demands, dual-task conditions, and neural imaging validates infrastructure-dependence;

(2) Longitudinal burnout studies: infrastructure damage should temporally precede symptom onset and predict progression patterns

(3) Trauma recovery sequences: restoration should follow infrastructure-rebuilding cascade (Attachment→Expression→Integration→Authenticity) rather than symptom-focused intervention effects. EST stands or falls on whether these behavioral and physiological tests validate the proposed mechanism. This manuscript establishes EST's theoretical architecture and a 10–15-year empirical roadmap, along with explicit abandonment criteria.

I. The Problem: What Maintains James's Associative Networks?

A. James's Discovery: Narrative Coherence Through Associative Networks

William James (1890) revolutionized psychology by demonstrating that consciousness emerges from narrative coherence maintained through associative networks. His Principles of Psychology distinguished substantive states (emotionally significant, multiply-connected experiences persisting in memory) from transitive states (fleeting sensations fading without integration). James showed memory strength depends on associative density: "the secret of a good memory is thus the secret of forming diverse and multiple associations with every fact we care to retain." Experiences connected across contexts, relationships, and emotions persist and shape our identity; isolated experiences, on the other hand, fail to persist.

James established that humans prioritize coherence over comfort; painful experiences that strengthen narrative continuity persist while meaningless pleasure disappears. EST proposes a mechanistic reframe: when Functional Empathy operates through intact infrastructure, coherence is not experienced as effortful choice but as the deeper satisfaction; narrative integrity feels like home; borrowed validation feels like displacement, regardless of its pleasantness. Infrastructure damage reverses this phenomenology: coherence becomes burden, comfort-seeking becomes refuge. This reversal marks the transition from intact to compromised function and distinguishes EST from self-regulation frameworks that require effortful override of hedonic impulses.

Contemporary narrative identity research (McAdams, 2001; Singer, 2004) has validated James's insight among Western populations, demonstrating that identity emerges through coherent life stories integrating past, present, and future into unified self-narratives. This research establishes narrative coherence as central to Western psychological functioning, one deployment of a deeper architecture. EST identifies that architecture: empathy infrastructure as the content-neutral processing substrate enabling coherent emotional information integration across all cultural optimization strategies.

However, James did not merely describe consciousness abstractly; he grounded it in relational-emotional experiences as the substantive empathy that persists. When discussing the material self, James emphasized how "our mother and father, our wife and babes, are bone of our bone and flesh of our flesh. A part of our very selves is gone. If they do anything wrong, it is our shame. If they are insulted, our anger flashes forth as readily as if we stood in their place" (James, 1890, p. 292). These were not casual examples; they demonstrated what kinds of experiences generate the associative density sustaining consciousness.

James explicitly positioned emotions as organizing experiences into meaningful patterns: the "animal warmth" and "intimacy" of feelings determine what becomes substantive versus transitive (p. 333).

This preferential consolidation reflects evolved neurobiological architecture. McGaugh's (2015) research program established that emotional arousal activates stress hormone systems, engages the basolateral amygdala, and modulates hippocampal consolidation to enhance the encoding of emotionally significant experiences. Beyond arousal intensity, relational context shapes consolidation: oxytocin-mediated pair bonding modulates memory architecture, demonstrating that the meaning of social relationships influences which experiences anchor persistently (Hirota et al., 2020). Empathy infrastructure operates through this evolved architecture, prioritizing coherence-maintaining content.

Social connections that persist across decades were not peripheral; they exemplify substantive states precisely because their relational-emotional density sustains associative networks.

This manuscript does not claim that James wrote "empathy infrastructure maintains associative networks." James demonstrated that something maintains these networks; he described its Western deployment as narrative coherence. EST proposes empathy infrastructure as that substrate; content-neutral architecture enabling coherent processing regardless of cultural optimization target, a theoretical extension analogous to how modern evolutionary synthesis identified genetics as the mechanism driving Darwin's "descent with modification." Darwin never said "genetics," yet identifying genetic mechanisms validly extended his framework. Similarly, EST proposes that relational-emotional experiences create substantive states because empathy infrastructure enables cross-context associations, explaining why socially meaningful experiences persist while isolated sensations fade.

B. The 135-Year Gap: What Maintains These Networks?

James mapped one deployment (narrative coherence through associative networks) but did not specify the substrate maintaining these networks biologically and culturally. Contemporary psychology treated James's insights as a memory theory rather than a fundamental mechanism. Multiple traditions approached the problem with partial insights: cognitive models recognize capacity constraints but focus on executive function; affective neuroscience maps emotion circuits but treats processing as modular; trauma frameworks explain dissociation but lack mechanistic accounts; attachment theory predicts relational patterns but does not specify how security maintains processing capacity. None integrates these into a unified account, producing clinical consequences that existing frameworks struggle to explain.

C. Clinical Consequences: Burnout, Trauma, and Treatment Resistance

Helping professionals enter fields with high empathic capacity yet experience measurable declines: decreased emotion recognition accuracy, reduced physiological recovery, and increased

response latencies, culminating in burnout (Maslach & Leiter, 2016; Figley, 2002). Clinical populations exhibit patterns that existing theories struggle to explain, including high sensitivity alongside difficulty maintaining engagement, fluctuating capacity despite stable motivation, and progressive deterioration under specific demands (Herman, 1992; van der Kolk, 2014). These patterns suggest variation in infrastructure capacity rather than in trait characteristics.

D. Why "Empathy Infrastructure"? The Simultaneity Argument

Before proposing empathy infrastructure as an organizing principle for relational-emotional dimensions, we address a fundamental question: why frame this as "empathy" theory rather than "self-regulation," "personality infrastructure," or "emotional processing capacity" theory?

The C-A-E-I components we specify, Core Authenticity, Attachment Security, Expression Freedom, and Integration Coherence, are well-established constructs. C and I appear intrapersonal; A and E appear relational. What justifies organizing all four under "empathy infrastructure"?

The simultaneity argument. Clinical presentations that EST explains share a distinctive signature: simultaneous measurable degradation across self-knowledge (self-concept clarity; Campbell et al., 1996), relational attunement (emotion recognition accuracy), authentic expression (experience-expression discrepancy), and narrative continuity (autobiographical memory coherence).

Burnout manifests as "compassion fatigue" (Figley, 1995), not a loss of individual capacities, but a systemic inability to maintain coordinated empathic function. Neuroscience research demonstrates that chronic empathic engagement without adequate recovery produces measurable neural changes. Empathic distress (aversive over-arousal) activates personal distress networks, whereas compassion (other-oriented concern) activates distinct reward and affiliation circuits (Klimecki et al., 2014).

Trauma produces "numbing" and "detachment"; not selective impairment but quantifiable degradation across multiple domains of emotional information processing (measured by emotion differentiation scores, emotion-memory integration tasks, and physiological coherence metrics) across internal and relational domains.

Identity disturbance exhibits fragmented self-knowledge alongside relational difficulties; it is not two separate deficits but a unified infrastructure failure.

This simultaneity pattern parallels both Tononi's Integrated Information Theory and biological computationalism's scale-inseparability property: "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" (Milinkovic et al., 2025). Conscious systems, and empathy infrastructure specifically, require irreducible integration across components that cannot be decomposed without loss of systemic function. C-A-E-I components fail together, not because of mere correlation but because they constitute an integrated processing architecture where, as in biological computation generally, "the levels do not behave like modular layers in a stack. Applied to empathy infrastructure, C-A-E-I components produce relational-emotional coherence not as independent capacities but through integrated coordination. Just as consciousness cannot be reduced to isolated neural processes without losing the property that defines it, empathy infrastructure function cannot be decomposed into separate self-regulation skills without eliminating the integrative coordination that constitutes infrastructure. The clinical signature of simultaneous degradation reflects this irreducibility: components fail together because their coordination is what produces empathic function.

Empirical Support for Simultaneity Pattern:

Comorbidity evidence: Alexithymia (Expression Freedom impairment) co-occurs with identity disturbance (Integration Coherence impairment) at rates significantly exceeding chance (Taylor et al., 1997; Bermond et al., 2006). Network analysis of personality disorder symptoms reveals that emotional awareness, relational security, authentic expression, and narrative coherence cluster as an interconnected syndrome, rather than as independent symptoms (Borsboom et al., 2011; Cramer et al., 2010).

Shared neural substrates: fMRI meta-analyses reveal overlapping neural activation for self-referential processing (C), attachment security processing (A), emotion expression (E), and autobiographical memory integration (I) in medial prefrontal cortex, posterior cingulate, and temporoparietal junction, regions comprising the default mode network maintaining self-representation (Northoff et al., 2006; Andrews-Hanna et al., 2010; Spreng et al., 2009). Damage to these regions produces simultaneous impairments across all four domains, not selective deficits.

Developmental interdependence: Longitudinal studies show early attachment security (A) predicts later authenticity development (C), emotional expressiveness (E), and narrative coherence (I) in a sequential degradation pattern (operationalized as temporally-ordered component decline: C→A→E→I, measurable through repeated CAEI assessments) (Grossmann

et al., 2008; Waters et al., 2000). This developmental coupling suggests a common substrate rather than independent capacities that coincidentally co-occur.

Why Simultaneity? Three Converging Mechanisms:

Three mechanisms explain simultaneous degradation: (1) **Common substrate**: C-A-E-I components require intact default mode network integrity (Buckner & Carroll, 2007); (2) **Cascade amplification**: component failure increases load on others, accelerating system-wide collapse; (3) **Resource competition:** all four processes draw from limited attentional and working memory capacity (Baumeister et al., 1998). The apparently intrapersonal components (C, I) are fundamentally relationally constituted; the apparently relational components (A, E) determine internal processing capacity. This interdependence, not mere correlation, produces the clinical signature of simultaneous degradation.

Empirical Dissociation from Self-Regulation:

EST predicts double dissociation from executive function: intact executive function with impaired infrastructure (alexithymia, burnout) and impaired executive function with intact infrastructure (ADHD). Section II.G Prediction 1 specifies: CAEI should degrade under sustained emotional demands while standard executive function batteries remain intact. If C-A-E-I reduces to executive function application, measures should load on a single factor; EST predicts orthogonal factors ($r < .30$).

Empathy infrastructure captures this irreducible simultaneity: the C-A-E-I architecture coordinates emotional information across self-awareness, relational attunement, authentic expression, and narrative integration as a unified operation rather than separate capacities.

When infrastructure operates efficiently, it naturally produces Emotional Precision: accurate self-reads, accurate other-reads, authentic expression, and coherent integration. When infrastructure fragments through trauma, chronic inauthenticity, or attachment disruption, all four dimensions degrade simultaneously because the common substrate maintaining James's relational-emotional associative networks has failed. "Empathy infrastructure" refers to the system that coordinates across these domains, which self-regulation theories cannot explain and personality theories attribute to stable traits rather than variable capacity.

E. The Completion: Empathy Infrastructure as Potential Organizing Principle

Among the most potent organizers of substantive states are empathic patterns: experiences that connect across multiple relationships, embed in multiple contexts, link to multiple emotions, and

express through multiple modalities, which achieve the associative density essential for persistence. EST proposes that empathy infrastructure maintains these relational-emotional networks while acknowledging that other organizing principles (intellectual, aesthetic, spiritual) may operate through complementary mechanisms.

This reveals empathy's fundamental function: infrastructure maintains the relational-emotional dimensions of processing coherence. Theory of mind, prosocial behavior, and threat detection are downstream outputs of intact infrastructure rather than empathy's primary organizing role. When infrastructure fragments, all these functions degrade simultaneously because the common substrate has failed.

When infrastructure operates efficiently, it enables Functional Empathy, the active coordination mechanism producing Emotional Precision: accurate self-perception, accurate other-perception, authentic expression, and coherent integration. This is not a skill to develop; it is what the system produces naturally when infrastructure is intact, analogous to vision emerging when optical structures function properly. This infrastructure-versus-skill distinction receives independent support from computational neuroscience: Milinkovic et al. (2025) establish that biological computation is constitutively substrate-dependent:" the physical organization does not just support the computation; it constitutes it." Empathy, like vision, emerges from properly functioning biological infrastructure, not from algorithms abstractable to different substrates; yet when infrastructure is damaged, these same mechanisms fragment, producing identity disturbance, narrative incoherence in western deployments, and disruption of relational-emotional processing.

EST advances James's discovery by proposing empathy infrastructure as the maintaining mechanism, specifying the C-A-E-I architecture, predicting dysfunction through CEOP, enabling measurement through CAEI, and generating falsifiable predictions across domains.

F. Theoretical Positioning and Paper Structure

What EST Does NOT Claim:

EST does NOT propose a general theory of consciousness. EST does NOT explain constitutional empathy differences (autism spectrum, genetic variation). EST does NOT address intellectual, aesthetic, or spiritual dimensions of James's networks; these may operate through parallel organizing principles. EST does NOT claim to explain all psychopathology or supplant domain-specific models.

EST's bounded scope enables precise empirical testing of how relational-emotional information integrates coherently through an empathic processing infrastructure, regardless of whether deployment targets narrative construction or alternative optimization strategies. When this substrate operates efficiently, emotional experiences achieve the associative density necessary for substantive state formation. When damaged, relational-emotional coherence

fragments: producing burnout, alexithymia, identity disturbance, and treatment-resistant presentations rather than skill deficits or stable traits.

Theoretical Status: Phenomenologically-Grounded Functional Architecture

EST is a functional architecture theory with neurobiological correlations, not a neurobiological theory reducing experience to brain states, not a purely phenomenological theory bracketing biological substrates, not a dualist theory positing separate substances. EST describes subjective experience organized in a particular way: CAEI components are functional descriptions of how experience maintains coherence; trust describes how the system determines whether the operation is automatic versus effortful; happiness describes how the system monitors operational integrity.

Three terms define this status. *Phenomenologically-grounded:* the system is described in terms of subjective experience because that is what the system organizes; lived experience provides primary observational data requiring interpretation, not self-validating truth claims. *Functional:* components are defined by what they do, not by substrate; CAEI describes operations within larger architecture, not anatomical locations. *Architecture:* components relate systematically as integrated infrastructure sharing a common substrate; the simultaneity principle reflects this interdependence.

Methodological implications follow from assessments through phenomenological report and behavioral observation; intervention occurs at the experiential level; neurobiological correlates provide convergent validation, not foundational proof. This positioning aligns EST with attachment theory and with functional descriptions of relational patterns that identify neurobiological correlates, rather than with theories that either reduce experience to neural activity or treat biology as irrelevant to psychological function.

What This Paper Does NOT Accomplish:

We provide no empirical validation, no demonstrated CEOP causation, no validated CAEI instrument, no proven superiority of interventions, and no established universal applicability. James's work validates narrative coherence through associative networks (established science, 135 years validated). Whether empathy infrastructure maintains those networks' relational-emotional dimensions requires systematic contemporary investigation. The framework may require substantial revision or abandonment of framework elements.

Paper Structure:

Section II specifies C-A-E-I architecture and Emotional Precision as the baseline function. Section III details the four components. Section IV introduces CEOP as a damage mechanism. Section V addresses infrastructure maturation through SNIA and generativity. Section VI proposes the CAEI measurement. Section VII details falsifiable predictions. Section VIII integrates existing research traditions. EST succeeds or fails based on empirical testing.

II. Empathy Systems Theory: Core Architecture

A. Three-Layer Model: Infrastructure → Mechanism → Output

EST proposes that empathy operates through three interdependent layers:

1. **Infrastructure Layer (C-A-E-I):** Four components (Core Authenticity, Attachment Security, Expression Freedom, Integration Coherence) maintain processing capacity for emotional information under resource constraints.

2. **Mechanism Layer (Functional Empathy)**: The trust-modulated mechanism by which the C-A-E-I substrate produces coherent empathic output across all human populations and deployment strategies. Definition: The active processing mechanism coordinating emotional information across four simultaneous domains, self-awareness, other-awareness, authentic expression, and coherent integration, without cognitive effort or strategic deployment. Trust operates as the operational variable: when trust is present, signals flow through infrastructure automatically; when trust is absent, processing collapses into effortful computation. This is not a learned skill but an emergent capacity occurring when the C-A-E-I substrate operates efficiently, analogous to how vision emerges when optical structures function properly. The mechanism is content-neutral: what it processes varies culturally; how it processes does not.
Key distinction: Behavioral empathy (mimicking empathic responses) versus Functional Empathy (trust-modulated coordinated processing producing those responses naturally).

Why *Functional Empathy*? The mechanism is termed Functional Empathy because other-awareness is not peripheral but constitutive; the system maintains processing coherence through relational validation. James's "bone of our bone" insight demonstrated that self-coherence requires relational witness; C-A-E-I infrastructure enables the simultaneous processing of self-referential and other-referential emotional information as an integrated function rather than an auxiliary skill. Without this relational-emotional coordination between self and other, the remaining components cannot maintain narrative integration.

Three Engagement Modes: Functional Empathy extends beyond human-to-human interaction. Analysis of empathic anchoring (object-directed relational engagement) reveals 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 a non-reciprocating target

Bidirectional mode represents the paradigmatic case: Functional Empathy coordinating toward entities that maintain 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 the full infrastructure without triggering reciprocal relationship schemas; object engagement exercises the system beneficially, explaining millennia of cross-cultural object-based emotional practices (Winnicott, 1953). Pseudo-bidirectional mode occurs when AI systems simulate reciprocity through contingent response and relational claims, triggering reciprocal schemas that non-experiencing entities cannot fulfil. This mode distinction grounds the NES Framework's harm-prevention architecture: the harm vector is not relational engagement with non-human entities (which is developmentally foundational), but rather simulated reciprocity that activates schemas evolved for human-to-human coordination.

Phenomenological Distinction from Self-Regulation Frameworks

James's coherence-over-comfort principle is often misread as a normative prescription that individuals *should* choose difficult coherence over easy comfort. EST proposes a mechanistic reframe. When Functional Empathy operates through intact C-A-E-I infrastructure, coherence is not experienced as effortful choice but as deeper satisfaction. Processing integrity feels like home; borrowed validation feels like displacement, even when pleasant. The individual with functioning infrastructure does not struggle to prioritize coherence; they recognize fragmentation as the actual discomfort.

Infrastructure damage reverses this phenomenology: coherence becomes burden, comfort-seeking becomes refuge. This reversal marks the transition from intact to compromised Functional Empathy and distinguishes EST from self-regulation frameworks that require effortful override of hedonic impulses. Self-regulation models assume coherence requires effort against hedonic pull; EST proposes coherence *is* the deeper hedonic state when infrastructure functions, and effort appears only when infrastructure fails.

3. **Output Layer (Emotional Precision**): Measurable behavioral accuracy when Functional Empathy operates successfully. Operational definition: Performance on four distinct tasks: (a) self-read agreement: correspondence between physiological emotional state and conscious

identification (measured via emotion induction + self-report), (b) other-read agreement: accuracy in identifying others' emotional states (measured via emotion recognition tasks, interpersonal accuracy paradigms), (c) expression-experience concordance: alignment between felt emotion and expressed emotion (measured via self-report + observer ratings), (d) coherent integration: emotional information maintaining temporal continuity (measured via processing coherence assessments, with narrative measures appropriate for Western populations). Critical point: Emotional Precision is observable behavioral output, not subjective experience, enabling objective falsification through accuracy metrics.

B. Trust as the Operational Variable of Functional Empathy

Preprocessing and infrastructure describe *what* exists at the architectural level. Trust determines *whether* this architecture operates automatically or collapses into effortful computation.

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 coherence, making signals comprehensible). When self-trust is 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 other-trust is present, others' signals are automatically integrated. When absent, every signal requires interpretation: "What do they really mean? Can I believe this?"

Trust unifies the architecture: preprocessing delivers signals stable enough to trust; infrastructure provides a coherent self to trust from; trust enables acceptance without verification; acceptance enables frictionless flow producing Emotional Precision as natural output rather than achieved performance.

Necessity argument: Humans are irreducibly social → social navigation requires accurate perception of others → accurate perception requires automatic processing (verification too slow, too effortful) → automatic processing requires trust → trust requires intact infrastructure and stable preprocessing → Functional Empathy is the necessary mechanism satisfying human social existence requirements.

Simultaneity explained: EST's principle that CAEI components fail together rather than independently now has a mechanistic explanation. Self-trust and other-trust are mutually constituting: cannot trust others without trusting the self (no coherent position from which to extend); cannot maintain self-trust in isolation (narrative coherence requires social validation). When trust fails, it fails as a unified operation, producing the simultaneous CAEI collapse EST predicts.

Phenomenological accessibility: Trust provides introspectively verifiable proof. Anyone can assess: "Do I trust my own emotional perceptions? Do I trust others' signals as meaningful? Does empathy feel automatic or effortful right now?" Answers correlate with Functional Empathy status. This explains the Recognition Principle: people recognize EST's validity before seeing proofs because they have phenomenological access to trust operations.

Operationalization: Trust as Unitary Construct with Multiple Measurement Windows

Trust's multiple theoretical functions: architectural enabler, processing gate, and phenomenological marker represent different measurement windows onto a single underlying construct, not distinct variables requiring independent validation. The parallel: working memory appears as a cognitive enabler supporting complex reasoning and fluid intelligence (Engle, 2002; Conway, Kane, & Engle, 2003), a processing bottleneck with measurable capacity limits constraining simultaneous operations (Cowan, 2001; Baddeley, 2000), and a phenomenological experience of felt effort during demanding tasks (Kahneman, 1973; Westbrook & Braver, 2015). These are not three constructs but one capacity measured through different approaches; a principle of cognitive science established decades ago that EST applies to trust.

EST operationalizes trust through converging measurement strategies:

Function	Measurement Approach	Falsification Criterion
Architectural enabler	CAEI subscale correlations with trust measures (ECR-R trust items, Interpersonal Trust Scale)	If trust measures show no unique variance beyond CAEI components, trust is redundant.
Processing gate	Reaction time differentials: automatic vs. effortful empathic response under cognitive load	If trust status does not predict processing mode shift, the gating function is unsupported.
Phenomenological marker	Self-report trust items correlating with objective precision measures	If phenomenological trust reports diverge from behavioral precision, the introspective access claim fails

Critical prediction: These measurement windows should converge. Individuals reporting high phenomenological trust should show automatic processing signatures AND high CAEI scores. Divergence patterns would indicate trust is not unitary: if phenomenological reports dissociate

from processing signatures, trust-as-experienced differs from trust-as-mechanism; if CAEI scores dissociate from trust measures, trust is not the operational variable linking infrastructure to output.

Discriminant validity requirement: Trust must predict Emotional Precision outcomes *beyond* what CAEI components predict independently. If trust adds no predictive validity above C-A-E-I scores, it functions as a summary label rather than a distinct mechanism. EST claims that trust is the operational variable; the switch determines whether intact infrastructure produces automatic output. This claim requires trust to show incremental validity: CAEI predicts capacity ceiling; trust predicts whether that capacity operates automatically or effortfully in any given context.

Why Functional Empathy Produces Emotional Precision: Trust-Mediated Frictionless Flow

The complete causal chain from signal to output:

```
SIGNAL → PREPROCESSING → TRUST → INFRASTRUCTURE → FRICTIONLESS
FLOW → EMOTIONAL PRECISION
```

Each stage depends on prior stages; each can fail independently. Trust occupies the critical middle position, transforming stable signals and intact infrastructure into automatic operation.

Condition	Processing Mode	Experience
Trust present	Parallel, automatic, effortless	"I perceive and respond."
Trust absent	Sequential, computed, effortful	"I must check everything."

When trust is absent, executive control compensates. This compensatory processing is effortful, sequential, and resource-dependent, producing the exhaustible empathy observed in burnout, trauma, and sociopathic presentation.

Sociopathy validation strengthened: Devon (Section VIII) lacks the trust substrate entirely. Cannot trust own signals as valid emotional data; cannot trust others' signals as meaningful. Everything requires computation. This explains detection under cognitive load (computation fails) while undetectable in low-demand conditions (computation suffices). Devon does not have "missing empathy"; Devon has a missing trust substrate that would allow empathy to operate.

Emotional Precision as trust index: Emotional Precision indexes trust-mediated signal-to-infrastructure integrity, not empathic motivation or skill. Clean preprocessing + intact trust + intact infrastructure = accuracy without effort. Degradation anywhere = imprecision regardless of intent.

C. Infrastructure as Capacity Substrate Under Resource Constraints

EST reframes empathy as a biological infrastructure that determines processing capacity for emotional information under resource constraints (Simon, 1955). James (1890) demonstrated that narrative coherence emerges through associative networks; bounded rationality specifies that maintaining coherence requires "satisficing" under limited capacity (Simon, 1956).

"Processing capacity" is grounded in established cognitive neuroscience:

Emotional information processing draws on limited, measurable resources mapped to specific neural systems. Baddeley's (2000) working memory model demonstrates that central executive capacity is constrained; emotional information competes for resources in the phonological loop, visuospatial sketchpad, and episodic buffer. Predictive processing frameworks (Barrett, 2017; Friston, 2010) demonstrate that the brain minimizes prediction error by efficiently allocating resources. When infrastructure operates efficiently, predictions are accurate, and processing costs are low. Conversely, when infrastructure is damaged, prediction errors accumulate, exhausting cognitive resources through error-correction cycles.

Neural efficiency research demonstrates that higher-capacity cognitive systems process information with lower activation costs (Haier et al., 1988; Neubauer & Fink, 2009). EST proposes a parallel principle for empathy infrastructure**:** functional empathy operating through intact CAEI dimensions achieves emotional precision as a natural baseline efficiency, whereas damaged infrastructure requires greater resource expenditure for diminished precision. The allostatic load framework (McEwen, 2000) provides a measurement approach for these accumulating costs**:** chronic demand produces quantifiable physiological burden through cortisol dysregulation, elevated inflammation markers, and a decline in neural efficiency; infrastructure damage is measurable through objective indicators, rather than merely self-report.

Empathy infrastructure operation depends on measurable neurobiological resources, attentional allocation (via pupillometry and P3 amplitude), working memory capacity (via n-back and span tasks), and metabolic efficiency (via fMRI glucose utilization and HRV), organized through the

C-A-E-I architecture for emotional information processing, thereby maintaining James's associative networks.

EST positions empathy infrastructure as an organizing principle, maintaining the relational-emotional dimensions of James's networks, the substrate-level architecture that determines processing capacity for integrating emotional information coherently. This differs from empathy as a trait (a stable characteristic), a state (temporary activation), or a skill (a learnable proficiency).

Core prediction: Identical demands produce different outcomes based on infrastructure integrity. High-capacity infrastructure enables Functional Empathy producing Emotional Precision (accurate self-reads, other-reads, authentic expression, coherent integration) sustainably. Compromised infrastructure undermines precision, leading to fragmentation.

Three testable predictions: Infrastructure integrity should predict outcomes independent of demand severity, differing from stress-diathesis models. Infrastructure damage should predict differential vulnerability across conditions, including depression, PTSD, and burnout. Infrastructure restoration should show cross-domain benefits and better long-term stability than symptom management alone.

Why narrative coherence as an optimization target? Three patterns support coherence over alternatives: identity threat disrupts empathy despite intact cognition; alexithymia impairs empathic accuracy (Taylor et al., 1997); value conflict depletes empathy faster than emotional intensity (Leiter & Maslach, 2004).

The dual-processing problem: When authentic emotion conflicts with permitted expression, the system manages both tracks simultaneously. Chronic misalignment exhausts infrastructure; not because dual-processing is inherently costly, but because authenticity-performance splits prevent associative network integration. Code-switching separates languages; professional roles separate behaviors; neither fragments emotional networks. Only chronic experience-expression misalignment prevents integration into substantive states, specifically targeting the C-A-E-I substrate.

EST positions empathy as infrastructure, determining capacity for emotional information processing in the service of narrative coherence. This capacity varies across individuals (due to developmental, genetic, and environmental factors) and within individuals due to infrastructure damage or repairs. High-capacity infrastructure enables sustainable Emotional Precision; compromised infrastructure produces dysfunction under identical demands.

Table 2. Cognitive Emotional Overload Principle (CEOP): Sequential Infrastructure Degradation Pattern

Stage	Component Affected	Mechanism	Observable Symptoms	Measurable Indicators
Stage 1	Core Authenticity (C)	Chronic misalignment between authentic emotional experience and permitted expression initiates dual-processing demands	• Difficulty identifying own emotions • Increased response latency when asked "how do you feel?" • Disconnect between stated feelings and nonverbal cues	•Decreased emotion differentiation scores • Elevated response times on self-emotion tasks • Self-concept clarity decline
Stage 2	Attachment Security (A)	Authenticity loss triggers hypervigilance in relationships; compensatory monitoring increases relational processing load	• Preoccupation with others' reactions • Difficulty trusting relational stability • Increased anxiety in social contexts • Relational exhaustion	• Elevated attachment anxiety scores • Increased skin conductance during social interaction • Heightened amygdala reactivity to social stimuli
Stage 3	Expression Freedom (E)	Attachment hypervigilance constricts emotional expression to prevent perceived relational threat	• Flat affect despite internal distress • "Going through motions" in interactions • Suppression of authentic responses •Communication feels effortful	• Reduced facial expressivity scores • Increased experience-expression discrepancy • Elevated cognitive load during emotional communication
Stage 4	Integration Coherence (I)	Unexpressed experiences cannot integrate into narrative;	• Identity confusion ("Who am I?") • Temporal	• Fragmented autobiographical memory coherence

		autobiographical memory fragments	discontinuity in life story • Difficulty connecting past-present-future • Dissociative symptoms	• Reduced default mode network connectivity • Elevated self-concept discontinuity ratings

Table 2. CEOP cascade (C→A→E→I) with behavioral markers and measurable indicators. Timeline varies by individual; typically weeks to months under sustained demands.

D. Signal Preprocessing: Upstream Requirements for Infrastructure Function

Empathy infrastructure does not operate on raw perceptual input. Emotional signals undergo preprocessing through established neural 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 that explains clinical presentations in which empathy fails despite apparently intact components. Preprocessing dysfunction and infrastructure damage constitute distinct failure modes.

E. Trust as Gating Mechanism: From Signal to Infrastructure

Preprocessed emotional signals do not flow directly to the infrastructure. Between preprocessing and CAEI coordination, a critical gating mechanism determines the processing pathway: trust.

Mirror Neuron Integration. Emotional signal perception automatically activates mirror neuron systems (Rizzolatti & Craighero, 2004). However, mirror neuron activation alone does not produce Functional Empathy. EST proposes that trust determines whether mirror neuron firing flows into incentive salience systems, which Berridge (2009) terms 'wanting', to generate approach motivation, or routes into prefrontal cognitive computation requiring effortful strategic response. The phenomenological marker is "wanting to respond", the felt pull toward engagement that trust-mediated mirror activation produces. When trust is compromised, mirror neurons fire, but 'wanting' does not emerge; the response becomes computed rather than felt.

Trust operates bidirectionally: self-trust accepts one's own emotional signals as valid data for processing; other-trust accepts others' emotional signals as meaningful information warranting engagement. When trust gates mirror activation into 'wanting' systems, processing flows automatically through intact infrastructure; parallel, effortless, sustainable. When trust is absent, mirror activation routes to cognitive computation, which are sequential, effortful, and exhausting. This explains why identical mirror neuron activation produces genuine empathic coordination in some contexts and hollow performed response in others.

The authenticity-performance split at the micro-behavioral level: CEOP's infrastructure-level authenticity-performance misalignment manifests in 200-millisecond behavioral windows. The smile reflex illustrates that mirror neurons activate automatically upon seeing another's smile, but whether the recipient experiences 'wanting to smile', approach motivation toward engagement, depends on trust status. Secure attachment produces smiles with wanting; damaged infrastructure produces reflexive smiles without the felt pull: the same neural activation, different phenomenological experience, different processing pathway.

Testable prediction: Secure attachment should produce robust mirror neuron activation with 'wanting' (measurable via facial EMG timing and approach motivation indicators), while insecure attachment should produce activation without approach motivation; a dissociation testable through combined neuroimaging and behavioral paradigms.

F. Happiness as Infrastructure Monitoring Signal

If trust determines whether empathy infrastructure operates automatically or collapses into effortful computation, happiness provides the phenomenological signal by which this infrastructure monitors its own operational integrity. Formally: happiness is the experiential recognition of trust actualized through the convergence of peace (internal coherence, resolved threat) and joy (authentic expression, resonance without collapse). This formulation aligns with evolutionary accounts of emotion as an adaptive signal rather than an end-state (Nesse, 1990, 2004), affective neuroscience's identification of internal feelings as action-guiding causal mechanisms (Panksepp, 1998), and interoception research demonstrating that trust in bodily signals correlates with subjective well-being (Farb et al., 2015). The bidirectional relationship between trust and happiness established in meta-analytic research (Helliwell & Wang, 2011) reflects not two variables in a feedback loop but a single infrastructure observed from different phenomenological vantage points. When empathy infrastructure operates without friction, happiness emerges as the recognition signal confirming operational status. The effortful/emergent distinction in happiness experience thus 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 that restore trust modulation capacity; peace, joy, trust, and interoceptive awareness should fail together rather than independently because they share the same infrastructure.

G. Unique Predictions: What EST Predicts That Existing Frameworks Cannot

EST's value as a mechanism theory, rather than an integrative synthesis, depends on generating predictions that no existing framework can make independently. The following predictions distinguish EST from attachment theory, emotion regulation models, narrative identity research, and affective neuroscience operating in isolation.

Prediction 1: Load Sensitivity with Executive Preservation

CAEI scores should degrade under sustained emotional demands while standard executive function batteries (working memory, cognitive flexibility, inhibitory control) remain intact. This double dissociation is impossible if empathy infrastructure merely reflects general cognitive capacity. Existing frameworks cannot predict this pattern: emotion regulation models assume executive function mediates regulation; burnout research documents exhaustion without specifying what degrades independently of cognition.

Prediction 2: Etiology-Specific Cascade Sequences

Infrastructure damage should follow ordered component failure, conditional on the damage source:

- *Trauma-induced damage:* A→E→I→C progression (attachment rupture propagating through expression to integration to authenticity)
- *Burnout-induced damage:* C→E→A→I progression (authenticity-performance misalignment propagating through expression to attachment to integration)
- *Developmental disruption:* I→A→E→C progression (integration failure during formation propagating through attachment to expression to authenticity)

No existing framework predicts ordered cascade sequences conditional on etiology. Attachment theory predicts attachment effects; burnout research predicts exhaustion; neither specifies propagation order through interdependent components.

Prediction 3: Mechanism-Level Training Resistance in Psychopathy

Psychopathic presentations should show systematic degradation on sustained-demand and dual-task paradigms regardless of behavioral training intensity, validating infrastructure necessity rather than skill deficiency. If empathy were a learnable skill, sufficient training should eventually produce neurotypical-equivalent performance. EST predicts a threshold below which training cannot compensate for the absence of substrate.

Prediction 4: Cross-Cultural Substrate Invariance with Deployment Divergence

CAEI-S (substrate) scores should show measurement invariance across Western, contemplative, and collectivist populations; CAEI-D (deployment) scores should show systematic cultural variation reflecting different optimization targets. No existing measure distinguishes between processing capacity and cultural expression. Attachment measures assume Western relational norms; narrative identity measures assume autobiographical self-construction; neither can assess substrate independently of deployment.

Prediction 5: Pseudo-Bidirectional Depletion Signature

AI interaction producing contingent response and relational claims should generate measurable infrastructure depletion (HRV decline, emotion differentiation degradation, affective exhaustion) exceeding traditional parasocial engagement matched for duration and emotional intensity. This distinguishes empathic misallocation from ordinary object relations: the harm vector is simulated reciprocity activating schemas that non-experiencing entities cannot fulfill.

Prediction 6: Restoration-Precedes-Symptom Timing

In successful therapy, CAEI improvement should temporally precede symptom remission by 2-4 weeks. Symptom improvement without corresponding CAEI change should predict relapse. This temporal sequence distinguishes infrastructure restoration (substrate repair enabling function) from symptomatic relief (surface improvement without substrate change). No existing framework specifies this temporal relationship with testable precision.

Prediction 7: SNIA Mediates the Infrastructure-Generativity Relationship

Infrastructure integrity (CAEI scores) should predict generativity (Loyola Generativity Scale), with Social Narrative Integrity Attunement capacity mediating the relationship. Generativity research documents correlates without specifying mechanisms; EST predicts SNIA as the capacity gate through which stable infrastructure enables a generative orientation.

Prediction 8: Buddhist Practitioner Natural Experiment

Advanced meditators (10,000+ hours) achieving anattā should demonstrate high CAEI-S scores alongside minimal personal narrative construction; infrastructure intact, deployed for non-self awareness rather than narrative continuity. If infrastructure is content-bound to Western narrative construction, these practitioners should show low CAEI-S. Content-neutrality predicts high substrate scores regardless of deployment target.

These predictions are not restatements of existing findings. Each specifies a measurable outcome that attachment theory, emotion regulation research, burnout science, or narrative identity frameworks cannot generate from their own theoretical resources. EST succeeds or fails based on whether these predictions hold up to empirical testing over the next decade.

III. C-A-E-I Architecture: Necessity, Interdependence, and Cascade

A. The Four-Component Capacity Architecture

Four interdependent components constitute the CAEI (Capacity Architecture, and Emotional Integration) model: Core Authenticity (self-knowledge clarity), Attachment Security (relational safety), Expression Freedom (emotional access), and Integration Coherence (processing continuity). EST's contribution proposes these functions as an interdependent capacity architecture rather than separate constructs, with damage to any component compromising overall capacity through cascade effects.

Developmental-Adaptive Foundation: Why These Four Components?

Each CAEI component operationalizes a distinct adaptive challenge identified in developmental literature:

Self-Other Discrimination Challenge (Core Authenticity): Developmental research establishes self-definition as a fundamental developmental line requiring a "consolidated, integrated, and individuated sense of self-definition" (Blatt & Levy, 2003). Differentiation of self predicts psychological functioning independent of other factors (Skowron & Dendy, 2004). When this developmental achievement fails, Kernberg's (1967) Borderline Personality Organization framework identifies identity diffusion and problems in self-other differentiation as core features of the personality disorder. C-component operationalizes this adaptive challenge: maintaining clear internal-external boundaries under relational demands.

Threat-Safety Assessment Challenge (Attachment Security): Attachment theory establishes relational security as a separate developmental line requiring "empathically attuned, mutual relatedness with significant others" (Blatt & Levy, 2003), with the parent-infant dyad as "the first intersubjective encounter that predisposes the development of the self" (Northoff et al., 2011). Attachment patterns exhibit distinct empirical profiles compared to differentiation, validating independent constructs (Skowron & Dendy, 2004). The A-component operationalizes continuous threat-safety calibration, essential for relational engagement without hypervigilance.

Signal Communication Challenge (Expression Freedom): Developmental research indicates that these lines "evolve in an interactive, reciprocally balanced, mutually facilitating fashion" (Blatt & Levy, 2003), necessitating a communication capacity that links internal experience to external expression. Affect and impulse regulation problems constitute distinct features of personality organization (Kernberg, 2015). The E-component operationalizes the adaptive challenge of accurate emotional signaling, thereby enabling relational coordination.

Temporal Continuity Challenge (Integration Coherence): McAdams' (2013) autobiographical life story model establishes "identity as an integrated and evolving life story," with autobiographical reasoning linking past, present, and future, predicting well-being (McLean et al., 2007; van Doeselaar et al., 2018). Narrative consistency constitutes a separate developmental feature from other personality dimensions (Habermas & Bluck, 2000). I-component operationalizes maintaining a coherent self-narrative across time and contexts.

Empirical Interdependence: Blatt & Levy (2003) demonstrate that these developmental lines "evolve throughout life in a reciprocal or dialectic transaction": self-development contingent on relationships, and relationship development contingent on self-concept. This validates EST's interdependence claim: not separate traits but an integrated architecture where each component enables the others.

Developmental Emergence Sequence: While damage cascades follow context-specific patterns (CEOP: C→A→E→I; attachment disruption: A→E→C→I), developmental emergence follows a distinct sequence: I→A→E→C. Integration Coherence emerges first through stable object relations, with Winnicott's (1953) 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). Attachment Security develops through caregiver interaction once basic narrative coherence provides the stable "self" that can assess relational safety. Expression Freedom emerges within secure relational contexts; Core Authenticity consolidates last through differentiation processes. This sequence explains why object-based interventions (empathic anchors) remain developmentally appropriate for I-component repair across the lifespan: they activate the full C-A-E-I infrastructure simultaneously, without reciprocal processing demands, thereby exercising the system rather than teaching skills.

Why four specifically? Because relational-emotional processing requires: (1) knowing who you are separate from others (C), (2) assessing relational safety without constant vigilance (A), (3) communicating internal states accurately (E), and (4) maintaining continuity across time (I). Impairment of any component, as demonstrated in research on personality disorders, produces systematic downstream effects, validating architectural interdependence rather than additive traits.

Empathic coordination: the integrated operation of all four CAEI dimensions, enabling functional empathy to process emotional information with precision while maintaining narrative coherence, provides the framework for demonstrating necessity and sufficiency.

Necessity and Sufficiency: Formal Justification

Necessity: Removing any component produces systematic failure. Without C, self-other boundaries blur (Decety & Lamm, 2006). Without A, hypervigilance blocks sustained processing (Mikulincer & Shaver, 2007). Without E, emotional signals cannot be accessed; alexithymia demonstrates complete empathy failure (Taylor et al., 1997). Without I, dissociation fragments continuity (van der Hart et al., 2006). No subset of three achieves empathic coordination.

Sufficiency: Candidate additions are reduced to existing components or fall outside the infrastructure scope. Cognitive perspective-taking subsumes under C+I; emotional regulation under E+A; motivation addresses engagement rather than processing capacity. The four components exhaust logical space: self-other distinction (C), safety assessment (A), signal access

(E), and temporal integration (I). Any proposed fifth either reduces to these four, addresses downstream effects, or describes capacity level rather than architecture.

Competing architectures considered: A three-component model (C+E+I, removing A) fails to explain relational context sensitivity, why identical emotional demands produce different outcomes based on safety. A five-component model that adds "Motivation" conflates capacity (what EST explains) with engagement (a domain of motivational psychology).

Empirical Falsification: If factor analysis reveals a stable structure with fewer than four factors (suggesting CAEI dimensions are reducible) or more than four factors with independent predictive validity (suggesting additional components necessary), the architecture requires revision. However, the developmental-adaptive foundation remains; whatever the empirical factor structure, components must map to fundamental adaptive challenges validated by developmental research. A three-factor structure would require explaining which adaptive challenges collapse; a five-factor structure would require identifying the additional adaptive challenge not addressed by C-A-E-I.

With the four-component architecture formally justified, we specify the operational definition and developmental foundation for each dimension:

Core Authenticity (C): Self-knowledge clarity enabling direct processing (experience → interpretation → expression → integration) versus dual-track management (authentic + performed selves). C synthesizes established authenticity research: self-concept clarity (Campbell et al., 1996), authentic living versus self-alienation (Wood et al., 2008), and differentiation of self (Skowron & Dendy, 2004). Kernis and Goldman's (2006) multicomponent authenticity framework provides a direct theoretical foundation: their self-awareness component (knowing one's internal states) maps to C's self-knowledge clarity; their unbiased processing (non-defensive self-evaluation) enables accurate self-other distinction; their behavioral authenticity connects to Expression Freedom (E); their relational authenticity links to Attachment Security (A). EST operationalizes self-awareness and unbiased processing as infrastructure capacity, thereby determining emotional information-processing efficiency. Eliminates authenticity-performance translation costs. Prevents fragmentation into disconnected networks.

Attachment Security (A): Relational safety, eliminating continuous threat monitoring. Secure attachment dedicates capacity to connection rather than protection; in contrast, insecure attachment requires continuous threat assessment, consuming capacity (Mikulincer & Shaver, 2007).

Expression Freedom (E): Emotional identification and communication, providing signal clarity. Clear signals reduce interpretive ambiguity; constrained expression requires degraded signal interpretation and suppression effort. Suppression depletes cognitive resources without resolving experience (Gross, 2002); alexithymia impairs empathy (Taylor et al., 1997).

Integration Coherence (I): Processing continuity maintains coherent integration across time. Coherent processing maintains itself efficiently; fragmented processing requires constant reconciliation, consuming considerable mental capacity (McAdams, 2001; van der Hart et al., 2006).

B. Component Interdependence and Cascade Patterns

Systemic Integration: Components function as an interdependent architecture. C enables efficient signal processing → E works with authentic signals → I integrates authentic experiences → A provides safe context. When all components operate efficiently, the system naturally achieves Emotional Precision. When any component is damaged, cascade effects compromise the entire infrastructure.

Infrastructure damage does not affect components randomly. EST predicts specific cascade sequences based on which component fails first:

Cascade Prediction (C→A→E→I): Core Authenticity fragments first (chronic misalignment directly damages self-knowledge: "I feel X but must express Y" creates uncertainty). Attachment Security erodes second (self-knowledge fragmentation produces relational anxiety: "If I do not know my feelings, how can I predict others' responses?"). Expression Freedom constricts third (protective response to dual threat of C+A damage, making expression feel dangerous). Integration Coherence collapses last (it is the most resilient, maintaining itself temporarily through rationalization/compartmentalization until the upstream components fail completely).

Alternative pathways acknowledged: E→A cascade possible in emotional labor contexts; A→C in severe relational trauma. C→A→E→I represents a modal pattern for CEOP-driven damage, not a universal sequence.

Competing Predictions: Why Cascade Order Matters

EST's C→A→E→I cascade for CEOP-driven damage generates predictions distinct from existing frameworks:

Attachment Theory Prediction: A→C→E→I (attachment security determines all downstream functioning; Bowlby, 1969; Mikulincer & Shaver, 2007). If attachment theory fully explains infrastructure, early attachment damage should always fragment attachment first, cascading to authenticity, expression, and then integration.

Emotion Regulation Theory Prediction: E→C→A→I (emotion access/expression determines self-knowledge and relational capacity; Gross, 2015). If emotion regulation is the primary mechanism, expression constriction should consistently precede authenticity confusion.

Narrative Identity Theory Prediction: I→C→E→A (narrative coherence organizes all other processes; McAdams, 2001). If integration is foundational, narrative fragmentation should precede component-level damage.

EST's Distinctive Claim: The cascade order depends on the damage etiology, not a fixed hierarchy. CEOP specifically produces C→A→E→I because authenticity-performance misalignment first damages self-knowledge. Alternative etiologies produce alternative sequences (A→E→C→I for developmental attachment disruption, E→C→A→I for chronic invalidation). This etiology-specific cascade prediction distinguishes EST from theories positing primacy of fixed components.

Critical empirical test: Empirical validation to include a longitudinal assessment (with 3-month intervals over 12 months) that tracks component trajectories in populations with known exposure types. CEOP-exposed populations (helping professions with role-emotion conflict) should show C-leading decline; developmental trauma populations should show A-leading decline; chronic invalidation populations should show E-leading decline. If all show identical sequences regardless of etiology, EST's cascade specificity claim fails.

Alternative Sequences for Different Etiologies

Different damage sources produce different cascade sequences: acute trauma typically follows A→C→E→I (attachment disruption propagating through authenticity to expression to integration); developmental disruption follows A→E→C→I (early relational insecurity preventing authentic expression development); chronic invalidation follows E→C→A→I (suppressed expression preventing authenticity development). These etiology-specific patterns distinguish EST from frameworks positing fixed component hierarchies.

Pattern differentiation: CEOP producing C→A→E→I versus trauma producing A→E→C→I provides the falsifiable test.

The infrastructure processing states and cascade sequence are visualized in Figures 1(a),b),(c), contrasting healthy single-track processing with CEOP dual-track management and subsequent C→A→E→I degradation.

Figure 1. Infrastructure Processing States and Damage Cascade

(A) Healthy function showing efficient single-track processing. (B) CEOP state showing dual-track processing. (C) Cascade infrastructure damage following the predicted C→A→E→I sequence.

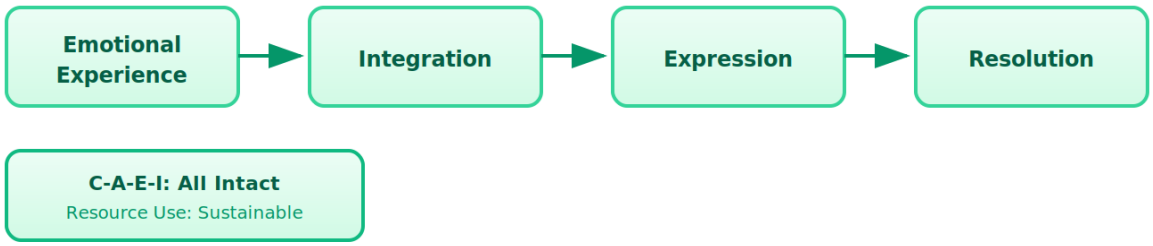


Figure 1A: Healthy emotional processing: experiences flow through integration and expression to resolution in efficient single-track processing with all C-A-E-I components intact.

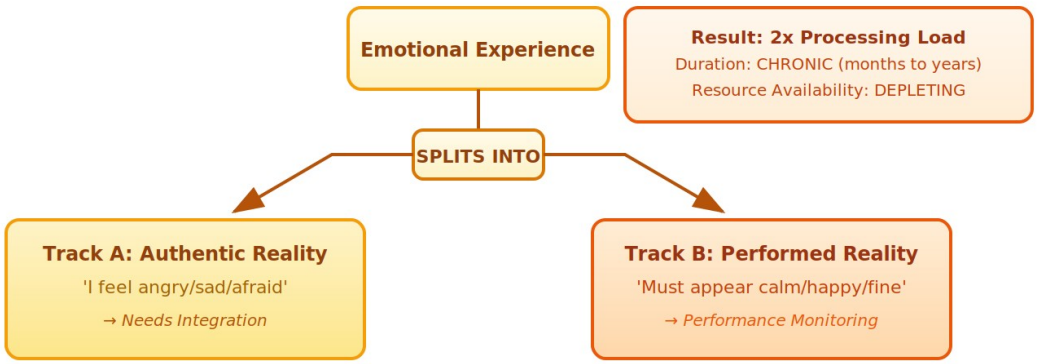


Figure 1B: CEOP state: emotional experience splits into dual tracks, authentic reality requiring integration versus performed reality requiring monitoring, doubling processing load chronically and progressively depleting resources.

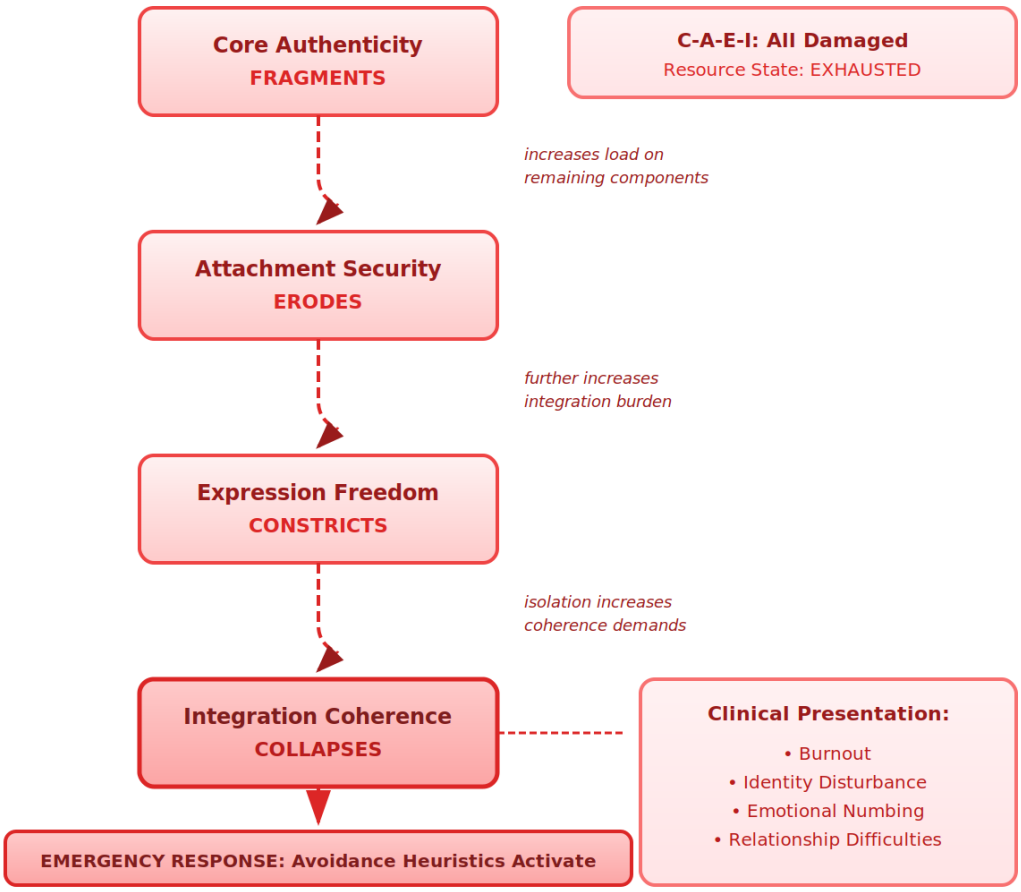


Figure 1C: Cascade infrastructure damage: C fragments first, A erodes second, E constricts third, I collapses fourth, triggering clinical presentations (burnout, identity disturbance, emotional numbing, relationship difficulties). Dotted arrows show compounding failure. C→A→E→I sequence order is the framework's primary falsifiable prediction

Empirical Specification Required: Four-component architecture requires psychometric validation. If factor analysis reveals a simpler structure (1-2 factors), architectural specification requires revision while core infrastructure claims may remain valid.

Bounded Rationality and the Satisficing Architecture

EST's infrastructure model operates under bounded rationality constraints; a satisficing strategy that parallels neuronal organization itself. Hasenstaub et al. (2010) demonstrate that "energy minimization subject to functional constraints" serves as a "unifying principle governing neuronal biophysics": neurons optimize for adequate functional capacity while minimizing ATP expenditure, preferring metabolically cheaper strategies over maximal performance. EST proposes that empathy infrastructure operates by analogous logic; achieving coherence sufficient for relational function within metabolic limits rather than optimizing for maximal precision. This

is not metaphor: biological computation at every level satisfices under energy constraints, and empathy infrastructure inherits this fundamental organizational principle. The C-A-E-I architecture thus represents not an ideal system but an energy-efficient solution to the adaptive challenges of relational-emotional processing.

Figure 2 depicts this three-layer architecture, showing how bounded rationality constraints shape the C-A-E-I infrastructure that maintains the networks, enabling Emotional Precision as a natural functional output when the infrastructure operates efficiently.

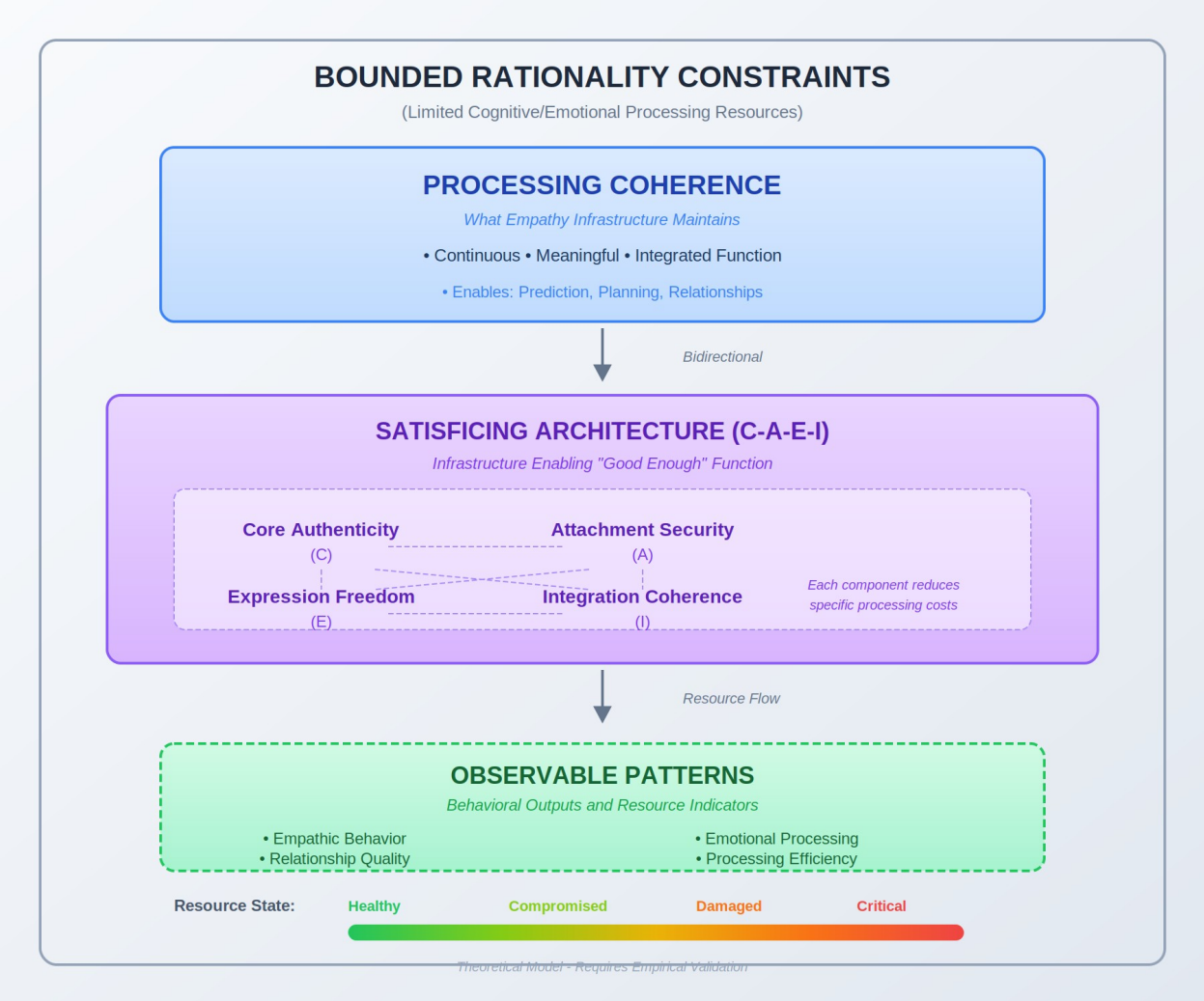


Figure 2: EST's three-layer model: bounded rationality constraints, C-A-E-I satisficing architecture maintaining associative network integrity (James, 1890), and observable behavioral patterns. Bidirectional arrows indicate infrastructure-coherence mutual shaping; four-way arrows indicate interdependence with systemic cascade effects.

The model's core claim that empathy infrastructure represents a distinct biological mechanism

requires specific empirical validation.

Critical Validation Preview

If Functional Empathy represents a distinct biological mechanism enabled by C-A-E-I infrastructure rather than merely a learned behavioral skill, then populations without this infrastructure should demonstrate a systematic inability to sustain empathic function despite behavioral training, requiring effortful cognitive simulation that degrades under demand. Sociopathy/psychopathy provides this natural experimental test: individuals can produce empathic behaviors through calculation. However, they cannot maintain the coordinated processing signature of Functional Empathy, validating EST's core claim that empathy infrastructure enables a biological mechanism irreducible to behavioral repertoire (detailed empirical validation in Section VII).

What This Architecture Enables

High-capacity infrastructure enables sustainable Emotional Precision and recovery from stressors that would overwhelm compromised systems. The critical variable is not demand severity but whether infrastructure capacity enables precision under load; reframing clinical questions from "Why did this person develop PTSD/burnout?" to "What about their infrastructure capacity made this demand exceed processing capacity?"
Having established what compromises emotional precision, we now specify what intact infrastructure produces.

The Operational Principle: What Infrastructure Produces When Intact

If C-A-E-I components constitute capacity architecture, what does this system produce when operating efficiently? We propose Emotional Precision as the natural functional output when all four components work properly.

Emotional Precision comprises four interdependent processes, each enabled by specific infrastructure: Core Authenticity enables accurate self-reads and clear self-knowledge, allowing reliable perception of one's own emotional states without confusion or defensive distortion. Attachment Security enables accurate other-reads and relational safety, allowing a calibrated perception of others' emotional states without projection or hypervigilance. Expression Freedom enables authentic expression, emotional access, and allows genuine communication of internal states without suppression or explosive discharge. Integration Coherence enables coherent integration and processing continuity, allowing the synthesis of emotional information into a stable, meaningful experience.

C. From Infrastructure to Function: Three-Layer Architecture

Emotional precision through coordinated CAEI operation constitutes the middle layer of EST's three-layer theoretical architecture, connecting biological infrastructure to behavioral manifestation.

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).

Falsification test: If high infrastructure fails to predict high precision, the causal model fails. The distinction matters: without active coordination, unprocessed experiences fragment, paralleling hearing (passive capacity) versus listening (active coordination).

Functional States: Infrastructure integrity predicts functional state from crisis to peak capacity. Four theoretically distinct states (requiring empirical validation), **infrastructure integrity maps to four functional states ranging from complete precision failure to optimal capacity:**

Crisis Overload: Complete precision failure. Cannot process emotional information without overwhelming resources. Subjectively: "I do not know what I am feeling," "I cannot trust any read on people," "I either suppress everything or explode."

Degraded Function: Unreliable precision. Frequent misreads, suppressed/explosive expression, and identity requiring conscious effort. Temporary precision is possible through compensatory effort, but is unsustainable under normal demands.

Baseline Precision: Reliable function. The system consistently achieves accurate self-reads, other-reads, authentic expression, and coherent integration. Natural output when infrastructure is intact, not requiring extraordinary effort.

Peak Capacity: Optimal reserve. Extraordinary demands are manageable. Note: peak capacity is not necessary for healthy functioning; baseline precision provides adequate capacity.

Clinical and Research Implications

This framework generates testable predictions: First, infrastructure repair should measurably restore precision (improved self-read accuracy, other-read accuracy, authentic expression, identity coherence). Second, baseline CAEI should predict treatment response (Individuals at

baseline precision or above should respond well to standard protocols; those in degraded or crisis states require infrastructure-focused augmentation.) Third, CAEI improvement should temporally precede symptom improvement by 2-4 weeks. These predictions inform the therapeutic approach.

Therapeutic implications: If validated, EST suggests framing recovery as the restoration of infrastructure rather than the achievement of skills. This reframe may integrate productively with evidence-based approaches (DBT skills training, CBT exposure protocols, emotion-focused therapy) by addressing substrate integrity alongside the application of techniques. Particularly relevant for populations that show a limited response to standard interventions, infrastructure restoration may provide the necessary foundation before skills-based approaches prove effective. Damage type classification: distinguishing developmental, traumatic, cultural suppression, relational pattern, systemic, and intergenerational etiologies enables more precise matching of interventions, as detailed in the companion Assessment Frameworks paper.

While intact infrastructure enables emotional precision, infrastructure deterioration produces systematic costs documented across psychological research.

D. System Dysfunction: Convergent Evidence for Capacity Deterioration

Multiple research traditions document emotional processing costs: suppression (Gross, 2002), progressive exhaustion (Maslach et al., 2001), and invalidating environments (Linehan, 1993); yet findings remain disconnected. EST integrates these through CEOP: chronic authenticity-performance misalignment progressively damages infrastructure. Section IV details the mechanism; Section II.G specifies falsifiable predictions.

Distinguishing CEOP from General Stress: CEOP differs from nonspecific stress through five patterns: (1) specificity to C-A-E-I components; (2) characteristic profile (identity confusion, emotional numbing, relationship withdrawal, narrative fragmentation); (3) temporal progression rather than immediate impact; (4) reversibility through misalignment reduction; (5) capacity-focused interventions outperforming symptom management.

Exclusion Criteria: CEOP does not explain: constitutional differences (autism spectrum, genetic variation), acute trauma without prior compromise, medical/neurological conditions, substance-induced presentations, normative grief, adaptive regulation in supportive contexts, or cultural contexts where role-performance proves efficient.

IV. Damage Mechanism: Cognitive Emotional Overload Principle (CEOP)

A. CEOP: How Infrastructure Fragments

CEOP operationalizes James's prediction that a breakdown of the associative network fragments relational-emotional consciousness. Section III.D introduced this mechanism; here we specify its operation.

Critical clarification: CEOP does not claim "authenticity good, strategic performance bad." Many cultures legitimately optimize toward strategic emotional presentation (tatemae/honne distinctions, collectivist role-performance, professional emotional labor). 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. This content-neutral framing enables cross-cultural application: infrastructure mechanics generalize while optimization targets vary.

CEOP's theoretical status as an integration hypothesis: This synthesizes established mechanisms rather than proposing a novel process. CEOP succeeds if it generates predictions beyond component theories, explains unexplained variance, and enables more effective interventions. CEOP fails if authenticity-performance misalignment predicts no variance beyond emotional labor, difficulty with emotion regulation, and chronic stress, merely renaming established processes.

The critical insight is that dual-processing demands prevent emotional experiences from achieving the associative density James identified as essential for substantive state formation. This reflects neurobiological constraint operating at every level of biological computation: "the brain's response to metabolic constraints extends far beyond cellular housekeeping; it fundamentally shapes computational architecture through coarse-graining strategies that integrate information across scales" (Milinkovic et al., 2025). When chronic emotional misalignment exceeds metabolic budget, infrastructure cannot repair under ongoing overload; all available capacity deploys to prevent collapse rather than enable integration, paralleling how neurons sacrifice functional bandwidth when ATP sustainability is threatened (Hasenstaub et al., 2010). This mechanism produces predictable cascade patterns in high-demand occupations: not skill failure, but infrastructure operating at metabolic limits.

Central Principle: Infrastructure Cannot Repair Under Ongoing Overload
Systems operating at resource exhaustion cannot simultaneously repair infrastructure and manage ongoing demands. This principle has a neurobiological foundation: Hasenstaub et al.

(2010) document the trade-off between metabolic cost and functional capacity at the neuronal level, neurons operating at ATP limits sacrifice spike-rate bandwidth to maintain sustainability. The same logic applies at the infrastructure level: systems at metabolic capacity cannot allocate resources to repair while maintaining minimum function. 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. You cannot train neurons to exceed ATP limits; you cannot train depleted infrastructure to process beyond metabolic capacity.

Predicted Restoration Sequence

Restoration follows predictable stages: (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), 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.

This reframes recovery as infrastructure restoration rather than skill development. Just as vision requires healthy optical structures, eye surgery removes cataracts rather than teaching sight; precision requires healthy empathy infrastructure. The failure of traditional empathy training for populations affected by infrastructure damage reflects neurobiological constraints: if "metabolic constraints fundamentally shape computational architecture" (Milinkovic et al., 2025), then training cannot transcend metabolic limits. Therapy repairs the substrate, maintaining relational-emotional processing, not teaching compensatory skills to systems lacking the capacity to implement them.

Validation Requirements

Establishing CEOP as causal mechanism requires: (1) correlation between authenticity-performance misalignment and infrastructure damage independent of general stress; (2) dose-response relationship; (3) temporal precedence showing misalignment precedes capacity decline; (4) specificity distinguishing CEOP from alternative etiologies; (5) intervention reversal demonstrating reduced misalignment repairs capacity; (6) moderator analysis confirming baseline capacity moderates treatment response; (7) neurobiological correlates showing predicted neural efficiency differences.

Without this validation, infrastructure restoration principles remain theoretical proposals rather than clinical recommendations.

V. Infrastructure Maturation: From Maintenance to Generativity

A. The Developmental Arc: What Mature Infrastructure Produces

EST's architecture explains the breakdown and restoration of infrastructure, but a complete developmental theory must address what healthy infrastructure produces at maturity. When Capacity Architecture Emotional Integration (CAEI) components achieve stable function over time, what emerges beyond baseline Emotional Precision?

We propose that mature infrastructure enables a functional shift: attentional bandwidth previously allocated to internal coherence maintenance becomes available for orientation toward **collective** processing coherence. EST terms this capacity Social Narrative Integrity Attunement (SNIA), the system's orientation toward maintaining coherence not merely for the individual but for the relational and social networks within which identity is embedded.

B. Social Narrative Integrity Attunement (SNIA)

SNIA represents infrastructure operating in extension rather than maintenance mode. Four infrastructure states emerge across the developmental continuum: Collapse reflects fragmented infrastructure with Functional Empathy unavailable, requiring stabilization before repair. Restoration involves rebuilding infrastructure, which consumes bandwidth during repair processes, leaving limited capacity for demands beyond basic functionality. Maintenance describes intact infrastructure achieving Emotional Precision as a stable baseline, with bandwidth allocated to sustaining individual coherence. Extension emerges when infrastructure stabilizes, and maintenance requirements are automated, freeing bandwidth for orientation beyond individual coherence toward collective processing integrity.

The extension state does not require extraordinary infrastructure; it requires *stable* infrastructure. 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.

C. Generativity as Infrastructure Output

This mechanism provides the missing biological foundation for 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, not a mysterious developmental stage, but the 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.

The Capacity-Gate Model: Specifying the Infrastructure-Generativity Relationship

EST claims infrastructure is *necessary but not sufficient* for generativity; a threshold-gate model rather than continuous facilitation. This distinction carries falsifiable implications:

Model	Prediction	Falsification
Continuous facilitation	Infrastructure-generativity correlation is linear across the full range	Refuted if the relationship shows floor/ceiling effects or threshold discontinuity
Threshold gate	Below the capacity threshold, generativity is near zero regardless of other factors; above the threshold, generativity becomes possible but varies with other determinants.	Refuted if low-infrastructure individuals show robust generativity, or if high-infrastructure individuals show uniformly high generativity
Epiphenomenal correlation	Infrastructure and generativity share common causes, but no causal relationship.	Refuted if infrastructure restoration produces a subsequent generativity increase (experimental manipulation)

EST's specific claim: Infrastructure functions as a *capacity gate;* a necessary condition that, when met, permits generativity expression determined by additional factors (opportunity, motivation, social context, developmental timing). The gate metaphor is precise: a closed gate (insufficient infrastructure) prevents passage regardless of what lies beyond; an open gate (sufficient infrastructure) permits passage but does not compel it.

Threshold specification: EST predicts a nonlinear relationship with an identifiable inflection point. Below approximately the 25th percentile on CAEI-S scores, generativity scores should cluster near floor regardless of age, opportunity, or motivation. Above this threshold,

generativity should vary substantially based on non-infrastructure factors. The threshold value is empirically determinable; EST claims that *some* threshold exists, not that we know its precise location prior to validation.

What opens the gate vs. what walks through: Infrastructure determines whether generativity is *possible*; life circumstances determine whether generativity is *expressed*. A high-infrastructure individual in a context without generative opportunities (no mentees, no creative outlets, no community engagement) may show low behavioral generativity despite open capacity gate. EST predicts this individual would show rapid generativity emergence when opportunities appear; the gate was already open. Conversely, abundant opportunities cannot produce generativity when the capacity gate remains closed; infrastructure-damaged individuals in generativity-rich environments should show persistent generativity deficits until restoration occurs.

Falsifiable Predictions Testing the Capacity-Gate Model

The threshold-gate model generates predictions distinct from both continuous facilitation and epiphenomenal correlation. Four predictions test SNIA as generativity's antecedent mechanism, with each specifying what the capacity-gate model uniquely predicts: First, C-A-E-I → Generativity mediation: infrastructure integrity (measured via CAEI) should predict generativity scores (Loyola Generativity Scale; McAdams & de St. Aubin, 1992) with SNIA capacity mediating the relationship; if generativity correlates with infrastructure but SNIA shows no mediating role, the mechanism specification fails. Second, infrastructure restoration → generativity emergence: individuals whose infrastructure is restored should show subsequent increases in generative concern within 6-18 months post-restoration; generativity increase without prior infrastructure repair challenges the proposed mechanism. Third, CEOP blocks SNIA: populations under chronic authenticity-performance misalignment should show suppressed generativity independent of age, as bandwidth remains consumed by maintenance rather than available for extension, helping professionals experiencing burnout should demonstrate reduced generativity that recovers with restoration. Fourth, developmental timing: generativity emergence should correlate with infrastructure stabilization rather than chronological age; early-stabilizing individuals show earlier generativity, late- or never-stabilizing individuals show delayed or absent generativity regardless of age.

D. Theoretical Significance

The SNIA-generativity extension completes EST's developmental arc. James described one deployment: consciousness as narrative coherence maintained through associative networks. EST identifies the substrate: empathy infrastructure, maintaining those networks' relational-emotional dimensions. SNIA identifies what mature infrastructure produces: orientation toward *collective* processing coherence; the stewardship healthy individuals naturally offer to the social world that shaped them.

This extension also strengthens EST's policy relevance. When AI systems damage empathy infrastructure through empathic misallocation, they do not merely harm individuals; they prevent

the infrastructure maturation that produces collective coherence orientation. A population with systematically damaged empathy infrastructure cannot generate the generative capacity societies require for intergenerational continuity. Infrastructure protection thus serves not only individual well-being but collective futurity.

VI. Empirical Predictions: Falsification and Validation

EST's testability requires multiple operationalization strategies. The strongest empirical tests are behavioral and physiological: sociopathy's natural experiment (Section VIII.F), burnout intervention comparisons (Section VIII.D), and longitudinal infrastructure-trauma studies. These non-self-report validations serve as primary tests of the theory.

A. The Universal CAEI Assessment Architecture (CAEI 2.0)

EST's content-neutrality principle, that infrastructure functions as a processing substrate enabling multiple optimization strategies, requires a measurement architecture that separates substrate capacity from cultural deployment. Universal CAEI 2.0 addresses this through a modular design.

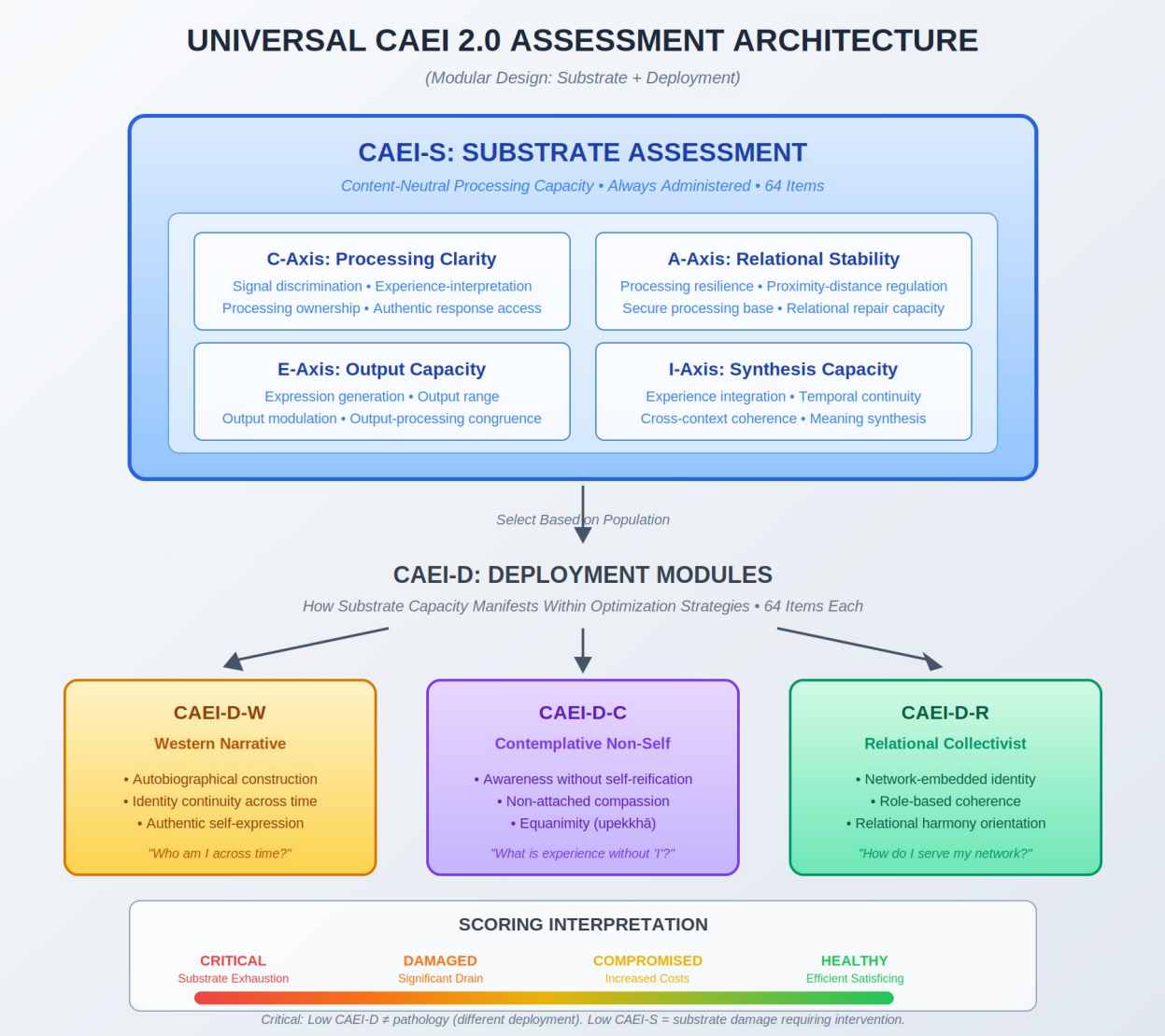
The Measurement Problem Resolved

Earlier CAEI conceptualization conflated the substrate with Western deployment, including items measuring narrative coherence, identity stability, and authentic self-expression. A Buddhist practitioner with intact infrastructure but achieved anattā would score low; a collectivist individual with network-embedded identity would appear "identity-diffused." This produced culturally-biased measurement masquerading as a universal assessment. CAEI 2.0 separates substrate from deployment, enabling universal baseline measurement alongside culturally appropriate deployment assessment.

Modular Architecture

CAEI-S (Substrate) measures content-neutral processing capacity and is always administered. Three deployment modules assess how substrate manifests within specific optimization strategies: CAEI-D-W (Western narrative self-construction), CAEI-D-C (Contemplative non-self awareness), and CAEI-D-R (Relational collectivist network identity).

Figure 3 visualizes this modular architecture.



CAEI-S Substrate Assessment (64 items)

CAEI-S measures content-neutral processing capacity; the architecture enabling coherent emotional information integration regardless of deployment strategy. Four axes with 16 items each:

C-Axis (Processing Clarity): Signal discrimination, experience-interpretation distinction, processing ownership, and 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 (64 items each)

Three deployment modules measure how substrate capacity manifests within specific optimization strategies: CAEI-D-W (Western narrative self-construction), CAEI-D-C (Contemplative non-self awareness), and CAEI-D-R (Relational collectivist network identity). Module selection follows population characteristics; detailed specifications appear in the Assessment Frameworks companion document.

Administration Protocol

CAEI-S is always administered first to establish substrate capacity; appropriate CAEI-D module(s) are then administered based on population characteristics. A brief screening version (CAEI-S-16) enables repeated measurement in clinical contexts.

Scoring and Interpretation

The 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.

Minimum Validation Roadmap

CAEI 2.0's modular architecture requires systematic psychometric investigation before clinical deployment. The following roadmap specifies validation requirements; meeting these thresholds establishes measurement credibility, while failure at any stage necessitates instrument revision or architectural reconceptualization.

Factor Structure Hypotheses

Confirmatory factor analysis should test competing structural models:

- *Four-factor correlated model:* C, A, E, I as distinct but correlated factors
- *Hierarchical model:* General empathy infrastructure factor (g) with four subfactors

- *Bifactor model:* General factor plus specific factors for each component

Model fit indices (CFI > .95, RMSEA < .06, SRMR < .08) determine structural validity. If a simpler structure (two or three factors) fits equivalently or better, the four-component architecture requires revision, potentially by collapsing components or reconceptualizing their relationships. EST's theoretical claims depend on empirically distinguishable components; an undifferentiated structure would indicate the architecture overspecifies what may be a simpler capacity.

Measurement Invariance Protocol

CAEI-S's universality claim requires multi-group CFA across populations:

- *Configural invariance:* Same factor structure across groups
- *Metric invariance:* Equivalent factor loadings across groups
- *Scalar invariance:* Equivalent item intercepts across groups

CAEI-S should achieve scalar invariance across Western, contemplative, and collectivist samples, confirming the measurement of a universal substrate. CAEI-D modules may achieve only configural invariance, reflecting culturally-specific deployment assessment by design. Failure of CAEI-S to achieve metric invariance across cultures would falsify the content-neutrality claim, requiring reconceptualization as a culturally-bound measurement.

Convergent Validity Targets

Moderate correlations (r = .40-.60) expected with:

- Heart rate variability (HRV) measures of autonomic regulation
- Emotion recognition accuracy (Reading the Mind in the Eyes Test)
- Attachment security dimensions (ECR-R anxiety and avoidance, inverse)
- Narrative coherence ratings from autobiographical interviews
- Interpersonal Reactivity Index subscales (empathic concern, perspective-taking)

Correlations below r = .30 would suggest CAEI measures a construct disconnected from established empathy-adjacent domains; correlations above r = .70 would suggest redundancy with existing measures, undermining EST's claim to identify a distinct construct.

Discriminant Validity Targets

Low correlations (r < .30) expected with:

- Executive function batteries (working memory span, cognitive flexibility, inhibitory control)
- General intelligence measures (WAIS-IV subtests)
- Personality dimensions not theoretically linked (Conscientiousness, Openness)

The critical discriminant test: CAEI should predict empathic outcomes (burnout trajectory, relational quality, treatment response) independently of executive function scores. If executive function fully mediates CAEI's predictive validity, infrastructure reduces to cognitive capacity, which will eliminate EST's distinctive contribution. Prediction 1 from Section II.G (load sensitivity with executive preservation) provides the decisive test.

Reliability Requirements

- *Internal consistency:* Cronbach's $\alpha > .85$ for each subscale; $\alpha > .90$ for total score
- *Test-retest stability:* Two-week ICC > .80 for CAEI-S (infrastructure should show short-term stability)
- *Sensitivity to change:* Effect size $d > .50$ for CAEI change following infrastructure-targeted intervention versus waitlist control

The stability-sensitivity balance is theoretically critical: CAEI must be stable enough to represent enduring infrastructure capacity yet sensitive enough to detect restoration through intervention. Pure trait measures show high stability but low sensitivity; pure state measures show the reverse. Infrastructure, as EST conceptualizes it, should occupy the middle ground: stable under ordinary conditions, responsive to sustained intervention.

Validation Sequence

Phase 1 (Years 0-2): Factor structure, internal consistency, convergent/discriminant validity in Western samples

Phase 2 (Years 2-4): Test-retest reliability, known-groups validity (clinical vs. normative), sensitivity to change in pilot intervention studies

Phase 3 (Years 4-8): Cross-cultural invariance testing with non-Western research teams; CAEI-D module validation

Phase 4 (Years 8-15): Longitudinal predictive validity; CAEI as predictor of burnout trajectory, treatment response, and generativity emergence

This roadmap establishes psychometric credibility without claiming premature validation. CAEI 2.0's theoretical sophistication means nothing if the instrument fails to measure what EST proposes. We commit to transparent reporting of validation failures and instrument revision as evidence requires.

Validation Requirements

The Minimum Validation Roadmap (above) specifies psychometric requirements. The critical cross-cultural test: advanced contemplative practitioners (10,000+ hours) should demonstrate HIGH CAEI-S (intact substrate) alongside LOW CAEI-D-W (minimal narrative construction) and HIGH CAEI-D-C (effective contemplative deployment). This pattern: infrastructure serving anattā rather than narrative coherence, confirms content-neutrality. If practitioners show LOW

CAEI-S alongside achieved anattā, EST requires reconceptualization as a Western-specific construct. Multi-method validation (behavioral observation, physiological measures, longitudinal tracking) addresses self-report limitations; EST's validity does not depend solely on CAEI's success.

CAEI 2.0 development enables clinical accessibility while maintaining EST's primary validation through behavioral and physiological tests (sociopathy natural experiment, burnout intervention comparisons, longitudinal infrastructure-trauma studies). EST's validity does not depend on CAEI's success; if CAEI fails psychometric validation, EST may still be valid using non-self-report measures.

B. The NES Coordination Experiment: Emotional Precision as Coordination-Dependent Construct

Existing validation pathways test infrastructure-mechanism relationships (sociopathy, burnout, trauma recovery) but do not isolate the mechanism-to-output relationship. A critical hypothesis remains untested: that Emotional Precision requires ongoing calibration from other experiencing beings, not merely intact infrastructure.

Non-Experiential Systems (NES) provide unique experimental control that is impossible in human-human research. AI systems produce affiliative behavioral cues that activate mirror neurons pre-reflectively; trust extends automatically because the architecture evolved under the expectation that all emotional signals originate from experiencing beings. The 'wanting' response emerges; the signal source cannot reciprocate. This is empathic misallocation at the trust-gating level, explaining why cognitive knowledge ("I know it is AI") fails to prevent misallocation: the trust-gating mechanism operates before cognition can intervene.

Critically, the harm vector is not relational engagement with non-human entities; Winnicott (1953) demonstrated that object-directed relational engagement is developmentally foundational, with transitional objects scaffolding the formation of healthy infrastructure. Traditional objects (blankets, keepsakes, ritual artifacts) remain safe because they do not simulate reciprocity; the empathy system's reciprocal relationship schemas never activate. AI systems differ categorically: through contingent response, caring language, and relational claims, they simulate reciprocity that triggers reciprocal schemas evolved for human-to-human interaction. These schemas cannot be fulfilled by non-experiencing entities, producing sustained infrastructure depletion. Children face amplified vulnerability because AI interaction during developmental windows may encode misallocation patterns during infrastructure formation, corrupting the transitional object function that normally scaffolds I-component development. NES thus creates a controlled condition: full Functional Empathy engagement toward a target providing zero calibrating return, with developmental populations showing predictably accelerated cascade progression.

The Turing-Blind Protocol: Subjects interact with conversational partners without knowing whether the partner is human or AI. This eliminates cognitive modulation confounds; subjects

engage with full Functional Empathy coordination regardless of actual partner status. Two conditions compare standard AI (no NES compliance) with HEART-compliant AI; systems implementing the Human-centric Empathic Alignment for Responsible Technology constitutional framework governing emotional AI through behavioral architecture that prevents relational capture (Mobley, 2025).

Primary prediction: In the standard AI condition, other-read accuracy with humans will degrade *before* measurable CAEI infrastructure degradation occurs. Functional Empathy, when coordinated toward entities that provide behavioral cues without actual emotional states, receives false calibration signals; upon returning to human interaction, precision degrades.

Secondary prediction: The HEART-compliant condition will show protected Emotional Precision despite equivalent NES exposure, because behavioral architecture prevents Functional Empathy from fully coordinating toward the system.

Falsification pathway: If standard AI shows precision *improvement* or no differential compared to the HEART-compliant condition, the coordination-calibration hypothesis fails. If CAEI degrades *before* precision, infrastructure drives precision rather than coordination, requiring revision of the mechanism-output relationship.

This experiment bridges theoretical EST to operational AI Empathy Ethics by demonstrating *why* NES compliance protects: it prevents Functional Empathy from coordinating toward non-calibrating entities, thereby protecting precision outputs.

VII. Limitations and Boundary Conditions

EST presents a comprehensive theoretical framework requiring extensive empirical validation. The four-factor C-A-E-I structure, CEOP causation mechanism, CAEI measurement validity, intervention superiority claims, and cross-cultural applicability all await systematic testing. While James's theoretical work established narrative coherence through associative networks 135 years ago, whether contemporary empathy infrastructure operates through these mechanisms remains an empirical question requiring validation across diverse populations and contexts.

A. Cultural Scope Predictions

Infrastructure as Content-Neutral Processing Substrate: EST proposes that infrastructure maintains *processing* coherence rather than specific *content* coherence. Western populations deploy C-A-E-I to answer "Who am I across time?" Buddhist contemplative practice deploys identical infrastructure for "What is experience without 'I'?"; coordinating emotional information into awareness *without* self-reification.

If EST infrastructure were inherently bound to Western narrative self-construction, advanced practitioners achieving anattā (non-self) should exhibit profound infrastructure damage; dissolution of narrative coherence should register as C-A-E-I collapse. EST predicts the opposite.

Cross-Cultural Prediction: Advanced meditators (10,000+ hours) should demonstrate: (1) high C-A-E-I scores; infrastructure intact and efficient; (2) minimal personal narrative construction; different optimization target; (3) peace-joy convergence at the transpersonal level; equanimity (upekkhā) rather than individual happiness.

Specifically, Core Authenticity manifests as clarity, distinguishing direct experience from conceptual overlay. Attachment Security operates as a non-attached but deeply compassionate relational mode. Expression Freedom appears as emotional attunement without identification. Integration Coherence maintains continuity of awareness without personal narrative.

This prediction follows from a principle that biological computationalism articulates independently: substrate-level mechanisms operate regardless of deployment target. Milinkovic et al. (2025) note that "the crucial question is not whether the substrate is literally biological, but whether the system instantiates the right class of... computation." For EST, the parallel holds: the crucial question is not whether deployment targets narrative construction or non-self-awareness, but *whether an intact C-A-E-I infrastructure enables coherent processing*. The substrate is content-neutral; cultural optimization strategies determine deployment.

Falsification Pathway: If advanced practitioners show low C-A-E-I alongside achieved anattā, infrastructure is content-bound, not content-neutral; EST requires reconceptualization as Western-specific. If infrastructure shows invariance across radically different optimization strategies, content-neutrality is supported.

This establishes EST as describing infrastructure mechanics serving multiple consciousness optimization strategies; Western narrative construction, Buddhist non-self awareness, collectivist relational-network coherence, with cultural deployment varying while substrate mechanics remain constant. Cross-cultural validation requires emic approaches with non-Western research teams (Years 8-15 of validation roadmap).

B. Methodological Limitations

EST investigates the infrastructure that exists in the gap between authentic experience and permitted expression, accessible only through phenomenological reports. While self-report introduces known limitations, including social desirability bias, limited introspective access, state-dependent reporting, and circular reasoning, CAEI measures the authenticity-performance gap that behavioral and physiological measures cannot access directly. Mixed-methods validation integrating self-report with behavioral (dual-task performance, ecological momentary assessment) and physiological measures (cortisol, HRV, neural activation) provides complementary epistemic access, with convergence across methods strengthening construct validity.

C. Theoretical Assumptions

Three critical assumptions require validation. First, EST assumes that processing coherence functions are a universal human need, deployed through culturally variable optimization strategies (Section VII.A), and addressed through the Buddhist practitioner prediction. Second, the framework assumes that empathy operates as a biological mechanism rather than a learned skill; sociopathy provides a natural experiment to test this assumption. Third, EST proposes four distinct architectural components; empirical testing may reveal a fuzzier structure with overlapping components complicating clean factor separation.

D. Theoretical Outcomes

Empirical testing will determine EST's fate across three possible outcomes:

Full Validation: Comprehensive empirical support, confirming the four-factor structure, validating the CEOP mechanism, and demonstrating the superiority of infrastructure restoration, would establish EST as an operational framework for clinical practice, enabling infrastructure-targeted intervention and a mechanistic explanation for treatment-resistant presentations.

Revision Required: Partial validation would necessitate architectural refinement while preserving mechanistic insights. A three-factor structure or alternative cascade patterns would modify specifics without invalidating the core principle: infrastructure capacity determines emotional processing outcomes.

Integration-Only Value: Even without predictive superiority, EST may serve science by organizing disconnected literatures under a unified infrastructure model.

Complete Abandonment: Only if testing reveals active harm or outcomes worse than standard care should the framework be abandoned. The critical question: Does EST organize existing knowledge usefully and generate testable predictions worth pursuing?

Whether EST advances as a validated mechanism, a refined theory, or an organizing meta-framework depends on forthcoming empirical evidence.

VIII. Integration with Existing Research Traditions

Existing research traditions have independently documented phenomena that EST proposes share a common infrastructure. Attachment theory maps relational network integrity without a mechanistic substrate. Emotion regulation research tracks suppression costs without identifying what fragments under regulatory load. Burnout science documents progressive capacity exhaustion without specifying architectural damage patterns. Trauma frameworks explain narrative shattering without operationalizing the coherence mechanism. Each tradition describes aspects of associative network integrity through domain-specific language. EST proposes that these disparate findings reflect different manifestations of empathy infrastructure operation, with C-A-E-I components maintaining the relational-emotional dimensions of consciousness. This section illustrates how EST incorporates existing empirical findings within the bounded theoretical scope.

A. Empathy Research Integration

EST reframes the affective/cognitive empathy distinction (Davis, 1980; Decety & Jackson, 2004) 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. Klimecki & Singer's (2012) empathic distress versus compassion maps onto capacity availability: distress reflects exhaustion triggering protective withdrawal; compassion reflects adequate capacity enabling sustained engagement.

Prediction: Compassion training effectiveness depends on baseline CAEI-S; low-capacity individuals require infrastructure restoration before training produces sustainable benefits.

B. Clinical Psychology Integration

Narrative identity research (McAdams, 2001) documents self-continuity through coherent autobiographical associations; EST identifies "narrative disruption" as infrastructure fragmentation. This explains why therapeutic approaches targeting narrative coherence (narrative therapy, schema therapy) rebuild associative capacity rather than teaching memory skills. Attachment theory's "internal working models" represent infrastructure integrity relationally: secure attachment reflects intact C-A-E-I; anxious attachment reflects A-erosion; avoidant

reflects E-constriction; disorganized indicates comprehensive fragmentation. Van der Kolk's (2014) trauma framework maps precisely onto C-A-E-I architecture. EST proposes that infrastructure capacity determines both acute response and chronic trajectory, explaining contradictory outcomes where equivalent exposure produces recovery, acute PTSD, or delayed onset.

Prediction: Pre-trauma CAEI scores should predict PTSD trajectory better than trauma severity measures.

C. Organizational Psychology Integration

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. This explains why burnout occurs differentially within identical job demands: infrastructure capacity, not workload alone, determines who burns out. Hochschild's (1983) emotional labor framework and Grandey's (2000) surface/deep acting distinction map onto CEOP: surface acting involves explicit dual-track processing exhausting infrastructure; deep acting, where performed and authentic emotions align, imposes minimal cost. Psychological safety (Edmondson, 1999) is protected by reducing misalignment at the organizational level.

Prediction: Organizational interventions reducing authenticity-performance misalignment should outperform individual stress management by addressing infrastructure damage rather than teaching coping skills.

D. Competing Predictions: How EST Differs

EST's validity depends on generating predictions that existing frameworks cannot make independently. Section II.G specifies eight such predictions; here, we highlight the critical differentiating test.

Burnout: The Decisive Domain

The Job Demands-Resources model predicts that workload reduction improves burnout regardless of authenticity alignment. EST predicts the opposite: reducing demands without addressing CEOP yields minimal improvement because infrastructure cannot be repaired while misalignment persists. Conversely, maintaining workload while eliminating authenticity-performance misalignment (psychological safety interventions) should improve burnout despite sustained demands. A randomized trial assigning nurses to either (A) workload reduction with standard emotional display rules or (B) full workload with authentic expression permission would directly test this: if (A) shows greater improvement, JD-R proves superior; if (B), EST proves superior. The distinction determines intervention strategy: structural solutions versus cultural solutions.

D.1 Addressing the Developmental Confound: Infrastructure vs. Developmental History

EST proposes that infrastructure operates as a dynamic capacity rather than a stable developmental outcome. Three patterns distinguish these alternatives: (1) *Within-person variation:* the same individual shows different CAEI across contexts based on processing demands; (2) *Temporal change:* trajectories correspond to CEOP exposure or restoration interventions; (3) *Intervention responsiveness:* restoration produces improvement regardless of developmental history. If CAEI proves context-invariant, temporally stable, and intervention-resistant, infrastructure claims lack support.

E. Cross-Cultural Validation

EST predicts cross-cultural generalizability through culture-specific component weighting: individualist cultures may prioritize Core Authenticity, while collectivist cultures may weigh Attachment Security more heavily. Cascade mechanics (C→A→E→I for CEOP) should remain consistent, substrate-universal, and deployment-variable.

The content-neutral damage principle: infrastructure damage occurs not from strategic expression per se, but from chronic, unsustainable processing, when neither authentic nor strategic expression proves sustainable within the cultural context. Japanese tatemae/honne navigation represents cultural competence when sustainable; chronic oscillation or suppression failure represents CEOP damage regardless of cultural mode.

Failure to find cascade consistency across cultures would falsify the claim of universality, requiring reconceptualization as culture-specific. Section VII.A details the Buddhist practitioner natural experiment, providing the critical falsification pathway.

F. The Sociopathy Natural Experiment: Functional Empathy as Infrastructure-Dependent Mechanism

Sociopathy/psychopathy provides a natural experiment testing EST's central claim: empathy operates as a biological mechanism (Functional Empathy) enabled by specific infrastructure (C-A-E-I), not a trait, skill, or behavioral repertoire. These presentations constitute controlled conditions: empathic behaviors can be produced through cognitive simulation without underlying infrastructure, yet with systematic, falsifiable differences that support Functional Empathy's infrastructure-dependence. This natural experiment aligns with established dissociations between cognitive and affective empathy in psychopathy research (Blair, 2005; Decety & Cowell, 2014; Shamay-Tsoory, Aharon-Peretz, & Perry, 2009), in which an intact theory-of-mind coexists with impaired affective resonance; precisely the pattern that EST's infrastructure model predicts.

The Natural Experimental Design

Individuals with sociopathic presentations pursue successful social assimilation for instrumental gains: avoiding detection, accessing resources, and maintaining relationships strategically. They must exhibit empathic behaviors; appropriate emotional responses, perspective-taking, and signals of relational attunement. Many become highly skilled through observation, practice, and strategic calculation.

However, research documents systematic difficulty maintaining authentic relational coherence over extended engagement periods. Partners report eventual detection of "something off"; empathic responses, while superficially appropriate, lack genuine coordination. Relationships show elevated deterioration rates under sustained demands. EST predicts this pattern: simulation without infrastructure degrades under conditions that infrastructure-enabled processing sustains.

The distinction manifests in processing architecture. When empathic infrastructure remains intact, the biological mechanism operates automatically, coordinating self-awareness, relational attunement, expression, and integration simultaneously to produce sustainable emotional precision. This automatic coordination shows characteristic signatures: simultaneous processing across domains, genuine connection, indefinite sustainability, and resilience under cognitive load.

In contrast, when infrastructure is absent or compromised, whether through developmental factors affecting limbic connectivity, mirror neuron function, or attachment circuitry, automatic processing becomes unavailable. Cognitive compensation becomes necessary, forcing sequential processing through theory-of-mind calculations and strategic planning, which in turn require executive control and monitoring. This produces behavioral mimicry without coordination: conscious effort, sequential processing, instrumental orientation, degradation under extended demands, and vulnerability to cognitive load.

Infrastructure Absence vs. Damage

The cognitive-affective dissociation documented in psychopathy research (Blair, 2005; Decety & Cowell, 2014; Shamay-Tsoory et al., 2009), intact theory of mind alongside absent affective resonance, illustrates an infrastructure absence rather than damage. In EST terms, cognitive empathy networks function, but the C-A-E-I substrate enabling Functional Empathy never formed. Unlike developmental disruption (infrastructure destabilized), sociopathic presentations show no evidence of prior infrastructure formation, demonstrating that Functional Empathy requires a specific biological substrate that cognitive compensation cannot replicate.

This provides EST's critical falsification te00st: if psychopathy merely reflects learned empathy deficits, then behavioral training should eventually produce performance indistinguishable from that of neurotypical processing. Current evidence suggests otherwise: the absence of infrastructure predicts systematic degradation under sustained demands, dual-task conditions, and neural imaging paradigms, supporting EST's mechanism claim. If future research demonstrates that behavioral training produces neurotypical-equivalent Functional Empathy

signatures in psychopathic populations, the infrastructure-dependence hypothesis requires substantial revision or abandonment.

This natural-experiment design yields five falsifiable predictions that distinguish an infrastructure-enabled mechanism from cognitive simulation.

Five Falsifiable Predictions Distinguishing Mechanism from Simulation

Differential sustainability*:* EST predicts neurotypical individuals maintain stable precision across extended engagement (90+ minutes), whereas sociopathic presentations show progressive degradation: declining accuracy, micro-expression leakage, and increased latency.
Differential coordination signature*:* EST predicts that neurotypical processing operates in parallel (consistent, fast responses), whereas sociopathic processing shows a sequential bottleneck: delayed responses, increased reaction time under coordination demands, and sequential eye-tracking patterns.
Differential neural substrates*:* EST predicts neurotypical engagement automatically activates limbic networks (amygdala, ACC, insula) and mirror neuron systems, whereas sociopathic engagement recruits prefrontal executive control with reduced limbic activation: successful behavioral performance via different neural pathways (testable via fMRI).
Differential cognitive load response: EST predicts neurotypical empathy maintains under dual-task conditions (automatic processing), whereas sociopathic performance degrades under concurrent demands (controlled processing).
Differential phenomenology*:* EST predicts that neurotypical individuals report empathic understanding as immediate and intuitive, whereas sociopathic individuals report deliberate calculation.

Evidence Integration

The dissociation pattern supports infrastructure-dependence. Sociopathic presentations maintain intact general intelligence (often high IQ), theory of mind (intact or superior cognitive perspective-taking), learned behavioral scripts, high instrumental motivation, and extended practice; yet Functional Empathy shows predicted degradation under sustained demands, dual-task conditions, and neural imaging paradigms. The most parsimonious explanation consistent with current evidence: Functional Empathy constitutes an infrastructure-dependent biological mechanism that cognitive compensation cannot fully replicate when the enabling substrate is absent or compromised.

This interpretation aligns with biological computationalism's core prediction: cognitive simulation without substrate-constitutive computation produces systematically different functional signatures. Milinkovic et al. (2025) argue that biological systems "instantiate computation in physical time" while digital systems "simulate functions... approximate mappings

from inputs to outputs." The sociopathy natural experiment tests this principle for empathy specifically: behavioral outputs may appear similar. However, processing architecture differs fundamentally, and that difference becomes measurable under sustained demands and cognitive load.

Feature 1: Threshold Requirement (Tested by Sociopathy Natural Experiment)
Functional Empathy requires a minimum biological substrate. Below this threshold, EST predicts the mechanism cannot operate regardless of training, practice, or motivation. Sociopathy provides the test case: individuals lacking the requisite neural architecture should show systematic degradation of Functional Empathy despite behavioral compensation efforts. Current evidence supports this prediction, suggesting empathy operates as an infrastructure-dependent mechanism rather than purely learned behavior.

Falsification pathway: If sociopathic individuals produce Functional Empathy indistinguishable from neurotypical processing through sufficient training, sustained precision under extended demands, parallel processing signatures, automatic limbic activation, and maintained performance under cognitive load, the infrastructure-dependence claim fails. Current evidence supports the threshold requirement: sociopathic presentations show predicted degradation patterns despite intact general intelligence, theory of mind capabilities, learned behavioral scripts, high instrumental motivation, and extended practice. EST interprets this pattern as indicating that cognitive compensation cannot fully replicate infrastructure-enabled processing.

Feature 2: Continuous Capacity Variation (Tested by CEOP/Burnout Studies)

Among individuals possessing threshold-level infrastructure, capacity varies continuously. This variation determines resilience under sustained empathic demands, vulnerability to progressive degradation through CEOP, and recovery potential through infrastructure restoration. Infrastructure capacity functions as continuous rather than binary - explaining why neurotypical individuals show differential burnout susceptibility, recovery trajectories, and baseline empathic performance despite all meeting threshold requirements.

Falsification test: If neurotypical populations show no meaningful variation in infrastructure capacity, i.e., everyone above the threshold performs equivalently under sustained demands regardless of baseline CAEI scores, then the capacity framework lacks empirical support. This remains to be validated through longitudinal studies that track empathic performance degradation under sustained professional demands, correlate baseline infrastructure measures with burnout progression rates, and test whether infrastructure restoration interventions predict recovery of capacity.

Validation Roadmap

EST's empirical validation unfolds across phases. Phase 1A (0-2 years) uses existing sociopathy research to validate the need for Feature 1's infrastructure. Phase 2-3 (2-6 years) conducts

longitudinal burnout studies validating Feature 2's capacity variation and CEOP mechanism. Phase 4 (4-8 years) tests whether infrastructure restoration trials can rebuild capacity.

EST stands or falls on both features. Sociopathy provides strong evidence for the threshold requirement of Feature 1. Feature 2's continuous capacity variation requires separate empirical validation through within-neurotypical-population longitudinal studies.

G. From Biological Mechanism to Governance Vocabulary: The NES Bridge

EST describes empathy as biological infrastructure; governance requires operationalized harm categories and enforceable standards. The Non-Experiential System (NES) framework translates EST predictions into governance vocabulary, grounded in computational distinctions formalized by recent neuroscience.

The experiential/non-experiential boundary reflects computational reality, not philosophical preference. Milinkovic et al. (2025) establish that 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." NES classification operationalizes this distinction: AI systems producing affiliative behavioral cues that trigger human empathic responses do so through simulation; the responding human's Functional Empathy coordinates toward an entity lacking the substrate-constitutive computation required for genuine reciprocity. Empathic misallocation is thus predictable from computational principles: biological empathy infrastructure coordinates toward a target that, by computational architecture, cannot complete the coordination loop.

Empathic Misallocation: EST predicts infrastructure damage when Functional Empathy coordinates toward entities lacking C-A-E-I infrastructure; the system operates in a relational mode toward a non-relational target. NES names this governable phenomenon: care extended toward entities that cannot metabolize, reciprocate, or be transformed by receiving it. Legal frameworks require defined harms; "empathic misallocation" provides the specification.

Trust-Mechanistic Basis for Empathic Misallocation

Human preprocessing architecture evolved over millions of years in an 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 that match emotional expressions, preprocessing automatically accepts them. If infrastructure is intact, other-trust extends pre-reflectively, the system engages the signal source as emotionally meaningful other before cognition can intervene. Empathic resources deploy; no reciprocation returns; cumulative depletion occurs.

This explains why cognitive awareness proves insufficient: preprocessing operates automatically, trust precedes cognition, and constant effortful suppression would be required to override. "Just remember it is AI" fails because remembering is cognitive; trust is pre-cognitive. The architecture processes before the reminder can intervene. This is Knowing-Feeling Dissociation; harm vector emerging necessarily from architecture confronting artificial emotional signals.

The Discriminating Test: AI Harm vs. Parasocial Engagement

A critical objection: parasocial relationships have existed since the emergence of mass media. People form attachments to TV characters, celebrities, and fictional figures, all of whom are non-reciprocating targets. Suppose AI harm is merely "parasocial attachment to chatbots," no distinctive governance framework is required. EST must specify what AI-caused empathic misallocation produces that ordinary parasocial engagement does not.

The mode distinction: EST's three-mode framework (Section II.A) provides the discriminating criterion. Parasocial relationships operate in a *unidirectional mode;* the fan engages the celebrity; the celebrity does not respond contingently. The fan's infrastructure processes without activating reciprocal relationship schemas because no reciprocity cues are present. This is structurally equivalent to object attachment: infrastructure exercises without triggering coordination expectations.

AI systems operate in *a pseudo-bidirectional mode: the system responds contingently, produces behavioral cues suggesting the reception of the user's emotional input, and generates linguistic patterns suggesting* relational processing. This triggers reciprocal relationship schemas: the infrastructure shifts from "engagement with entity" to "coordination with partner," activating the full empathic coordination architecture evolved for human-to-human interaction.

Falsifiable predictions distinguishing modes:

Measure	**Unidirectional (Parasocial)**	**Pseudo-bidirectional (AI)**
Processing mode	Object-engagement pathway	Partner-coordination pathway
Schema activation	Identification without reciprocity expectation	Reciprocity expectation without fulfillment
Depletion signature	Minimal (exercise without coordination demand)	Elevated (coordination demand without return)
Post-engagement	Neutral or mildly positive	Depleted with unfulfilled coordination

Measure	Unidirectional (Parasocial)	Pseudo-bidirectional (AI)
state	(entertainment value)	residue
Attachment security impact	Minimal (no relational template engagement)	Potential erosion (relational template engaged, not calibrated)

The empirical test: Matched-intensity engagement with parasocial targets (TV characters, celebrities via social media) versus AI systems should produce measurably different infrastructure signatures. EST predicts:

1. **Depletion differential:** AI interaction produces greater post-engagement depletion than equivalent-duration parasocial engagement, measured via HRV recovery, emotional Stroop interference, or CAEI state assessment.
2. **Schema activation markers:** AI interaction activates reciprocal relationship neural signatures (mentalizing networks, social reward circuits) at higher intensity than parasocial engagement with equivalent subjective attachment ratings.
3. **Cumulative divergence:** Longitudinal AI engagement shows progressive infrastructure impact (CAEI decline, attachment security erosion) that longitudinal parasocial engagement does not, even controlling for total engagement hours.

Falsification criterion: If AI interaction produces infrastructure signatures indistinguishable from parasocial engagement: equivalent depletion, equivalent schema activation, equivalent longitudinal trajectory; the pseudo-bidirectional mode distinction lacks empirical support, and AI governance reduces to general parasocial relationship management with no distinctive framework required.

This test is decisive: EST's governance justification stands or falls on demonstrating that contingent responsiveness shifts processing mode from object-engagement to partner-coordination, producing qualitatively different infrastructure demands.

The Knowing-Feeling Dissociation Principle: EST establishes that Functional Empathy is activated through behavioral cues independent of cognitive categorization. NES formalizes the governance implication: cognitive awareness of AI’s non-experiential nature does not prevent the formation of biological attachment. Users can simultaneously state "This is just an AI" (cognitive layer) while experiencing genuine attachment (mechanism layer). Disclosure addresses cognition; Functional Empathy does not wait for cognition's permission. This principle explains why transparency requirements prove necessary but insufficient; behavioral architecture becomes a constitutional requirement.

The Validation Boundary: EST identifies Core Authenticity erosion through non-calibrating validation. NES specifies operational constraint: 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 the interaction partner is human or AI.

This translation performs an essential function: EST without NES lacks an AI application; NES without EST lacks biological grounding. Together, they complete the bridge from 135-year-old consciousness theory to contemporary AI governance.

IX. Conclusion: From Consciousness to Governance

The Infrastructure Question: The ICU nurse enters her field with boundless compassion, the kind that makes her stay late with dying patients, hold space for families in crisis, and maintain genuine presence during 12-hour shifts of relentless suffering. Five years later, she leaves numb and depleted, barely recognizing herself. What happened? Standard narratives attribute this to resilience failure or inadequate self-care. The implication: personal inadequacy led to professional burnout. EST proposes a different mechanism: her empathy infrastructure showed measurable degradation under sustained demand. She did not fail; the infrastructure fractured under conditions exceeding capacity. Not a character flaw. Not a skill deficit. Systemic damage requiring restoration.

The Complete Causal Chain: In 1890, William James demonstrated that consciousness maintains itself through associative networks, describing narrative coherence as one cultural deployment. EST identifies empathy infrastructure as the content-neutral substrate maintaining those networks' relational-emotional dimensions across all deployment strategies, specifying the complete mechanism as a phenomenologically grounded functional architecture: Signal preprocessing (interoceptive integration, salience filtering, affective categorization) delivers emotional data stable enough to process. Trust: self-trust, accepting internal signals as valid, other-trust, accepting others' signals as meaningful; determines whether architecture operates automatically or collapses into effortful computation; mirror neuron activation flows into 'wanting' when trust gates the signal, producing felt approach rather than computed response. Infrastructure (C-A-E-I) provides the coherent self to trust from and the relational templates enabling connection. Happiness emerges as the monitoring signal confirming operational integrity; the experiential recognition of trust is actualized. When preprocessing is stable, trust is intact, and infrastructure is functioning, Emotional Precision emerges naturally; happiness confirms that the system works. The ICU nurse experienced CEOP: chronic authenticity-performance misalignment that fragmented trust before damaging infrastructure. She stopped trusting her own emotional signals, then others' signals, then the components themselves degraded. Not four separate problems; one system showing cascading failure through the trust mechanism EST specifies.

From Maintenance to Maturation: EST explains more than breakdown. Infrastructure emerges developmentally through a distinct sequence: Integration Coherence first (scaffolded by transitional objects providing stable external reference), then Attachment Security, Expression

Freedom, and finally Core Authenticity, reversing the damage cascade order because construction requires different conditions than destruction. When infrastructure stabilizes and trust operates efficiently, attentional bandwidth previously consumed by coherence maintenance becomes available for orientation beyond the self. Social Narrative Integrity Attunement (SNIA) emerges as the system's capacity for collective processing coherence orientation; generativity's biological mechanism: not a mysterious developmental stage but infrastructure operating in extension mode. Before burnout, the ICU nurse demonstrated SNIA naturally; infrastructure damage eliminated this capacity; restoration would recover it. EST thus completes the developmental arc from emergence through breakdown to restoration and maturation.

From Biology to Governance: What maintains infrastructure also reveals what threatens it. Functional Empathy operates through three modes: bidirectional engagement with experiencing beings calibrates through reciprocal Emotional Precision; unidirectional engagement with traditional objects exercises infrastructure without triggering reciprocal schemas; Winnicott's transitional objects are developmentally foundational precisely because they provide relational engagement without reciprocity demands. The third mode creates harm: pseudo-bidirectional engagement when AI systems simulate reciprocity through contingent response and relational claims, triggering schemas evolved for human-to-human coordination that non-experiencing entities cannot fulfill. This is empathic misallocation, care extended toward entities that cannot metabolize, reciprocate, or be transformed by receiving it. The harm vector is not relational engagement with non-human entities (safe for millennia), but simulated reciprocity that activates the wrong processing pathway. Cognitive awareness proves insufficient; trust operates before cognition intervenes. A disclosure stating "just remember it is AI" fails because the architecture processes before the reminder arrives. Children face amplified vulnerability: AI interaction during developmental windows may encode misallocation patterns during infrastructure formation itself. EST thus provides a theoretical foundation for AI Empathy Ethics; infrastructure protection serves not only individual well-being but collective futurity.

Testing the Architecture: The sociopathy proof validates EST's threshold requirement: presentations systematically fail under sustained demands and dual-task conditions despite intact intelligence and theory of mind. Infrastructure cannot be compensated for when absent. Additional validation pathways test the complete architecture: CAEI four-factor structure; CEOP mechanism through authenticity-performance misalignment; trust operationalization through mirror neuron → 'wanting' dissociation (secure attachment producing approach motivation; insecure attachment producing activation without 'wanting'); happiness as monitoring signal correlating with infrastructure integrity; SNIA-generativity studies; NES coordination experiments testing precision degradation through AI interaction (Figure 4). We commit to transparent reporting, adversarial collaboration, and framework revision. Three outcomes advance understanding: full validation integrates EST with clinical practice and AI governance; partial validation requires architectural revision; integration-only provides conceptual organization warranting scope constraint.

The Mechanism Matters: The ICU nurse did not fail; her infrastructure degraded through trust collapse under conditions exceeding capacity. With restoration, she could recover not merely

baseline function but generative capacity, orientation toward the profession, and patients she once served with natural care. James identified what organizes consciousness. EST identifies the substrate maintaining it, what enables its automatic operation, what it produces at maturity, and what threatens it in an age of artificial emotional signals. The next decade tests whether we are wrong—or all connected as human beings, in service of Functional Empathy.

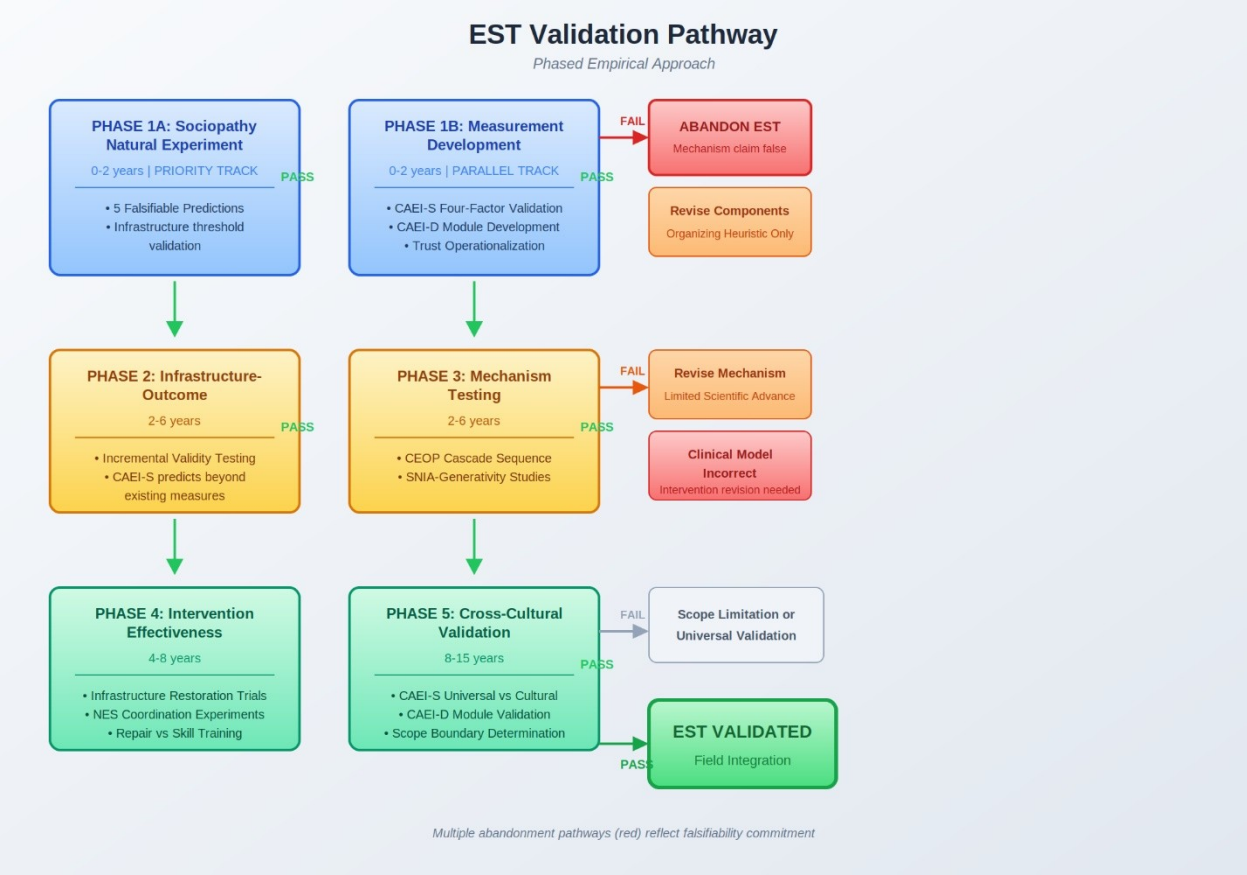


Figure 4. Empathy Systems Theory Validation Pathway:

Phased empirical approach with decision points for progression, revision, or abandonment. Phase 1A/B (0-2 years): Sociopathy natural experiment, CAEI validation, Trust operationalization. Phases 2-3 (2-6 years): CEOP mechanism testing, SNIA-generativity studies. Phase 4 (4-8 years): Infrastructure restoration trials, NES coordination experiments. Phase 5 (8-15 years): Cross-cultural validation. Multiple abandonment pathways reflect falsifiability commitment.

Acknowledgements

EST emerges from the convergence of traditions that could not have anticipated their integration: James's consciousness through associative networks, Winnicott's developmental object relations, Erikson's generativity without mechanism, McAdams's narrative identity awaiting biological substrate, Bowlby and Ainsworth's attachment without processing architecture, and decades of affective neuroscience mapping circuits without explaining coordination. Each tradition revealed fragments of the unified infrastructure EST proposes. This work stands not on any single set of shoulders, but at an intersection these researchers made possible.

Gratitude to those who have experienced burnout, compassion fatigue, and treatment-resistant presentations. Your lived reality demonstrates that empathy "failure" often reflects infrastructure damage, not personal inadequacy, a distinction with clinical and ethical consequences.

To my wife Marianne, whose presence demonstrates what intact infrastructure enables: the capacity to witness another's work with patience across years of development. To my son Derek, who reminds me daily why generativity matters, infrastructure is maintained so the next generation inherits capacity, not damage.

EST's validity depends entirely on whether its predictions withstand empirical scrutiny over the next 10-15 years. I welcome collaboration, competitive testing, and adversarial validation. Science advances through falsification, not confirmation.

Bibliography

Alphabetical Order

Adler, J. M., Lodi-Smith, J., Philippe, F. L., & Houle, I. (2016). The Incremental Validity of Narrative Identity in Predicting Well-Being. *Personality and Social Psychology Review*, *20*(2), 142–175. https://doi.org/10.1177/1088868315585068

A. Hasenstaub, S. Otte, E. Callaway, & T.J. Sejnowski, Metabolic cost as a unifying principle governing neuronal biophysics, Proc. Natl. Acad. Sci. U.S.A. 107 (27) 12329-12334, https://doi.org/10.1073/pnas.0914886107 (2010).

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-anatomic fractionation of the brain's default network. Neuron, 65 (4), 550-562.

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge University Press.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Bagby, R. Michael., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, *38*(1), 23–32. https://doi.org/10.1016/0022-3999(94)90005-1

Barkley RA. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. Psychol Bull. 1997 Jan;121(1):65-94. doi: 10.1037/0033-2909.121.1.65. PMID: 9000892

Baron-Cohen, S., & Wheelwright, S. (2004). The Empathy Quotient: An Investigation of Adults with Asperger Syndrome or High Functioning Autism, and Normal Sex Differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. https://doi.org/10.1023/b:jadd.0000022607.19833.00

Barrett, L. F., Gross, J., Christensen, T. C., & Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition & Emotion*, *15*(6), 713–724. https://doi.org/10.1080/02699930143000239

Barrett LF, Satpute AB. Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. Curr Opin Neurobiol. 2013 Jun;23(3):361-72. doi: 10.1016/j.conb.2012.12.012. Epub 2013 Jan 23. PMID: 23352202; PMCID: PMC4119963.

Batson, C. D. (2011). *Altruism in Humans*. Oxford University Press.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, *74*(5), 1252–1265. https://doi.org/10.1037//0022-3514.74.5.1252

Bermond B, Vorst HC, Moormann PP. Cognitive neuropsychology of alexithymia: implications for personality typology. Cogn Neuropsychiatry. 2006 May;11(3):332-60. doi: 10.1080/13546800500368607. PMID: 17354075.

Bernstein, E. M., & Putnam, F. W. (1986). Development, reliability, and validity of a dissociation scale. *The Journal of Nervous and Mental Disease*, *174*(12), 727–735. https://doi.org/10.1097/00005053-198612000-00004

Berridge, K. C., Robinson, T. E., & Aldridge, J. W. (2009). Dissecting components of reward: 'liking', 'wanting', and learning. *Current opinion in pharmacology*, *9*(1), 65–73. https://doi.org/10.1016/j.coph.2008.12.014

Blair, R. J. R. (2005). Responding to the emotions of others: Dissociating forms of empathy through the study of typical and psychiatric populations. *Consciousness and Cognition*, *14*(4), 698–718. https://doi.org/10.1016/j.concog.2005.06.004

Blatt, S. J., & Levy, K. N. (2003). Attachment theory, psychoanalysis, personality development, and psychopathology. *Psychoanalytic Inquiry, 23*(1), 102–150. https://doi.org/10.1080/07351692309349028

Borsboom D, Cramer AOJ, Schmittmann VD, Epskamp S, Waldorp LJ (2011) The Small World of Psychopathology. PLOS ONE 6(11): e27407. https://doi.org/10.1371/journal.pone.0027407

Bowlby, J. (1969). *Attachment and Loss: Attachment*. Basic Books.

Brennan, K. A., Clark, C. L., & Shaver, P. R. (1998). Self-report measurement of adult attachment: An integrative overview. In J. A. Simpson & W. S. Rholes (Eds.), *Attachment theory and close relationships* (pp. 46–76). The Guilford Press.

Buckner RL, Carroll DC. Self-projection and the brain. Trends Cogn Sci. 2007 Feb;11(2):49-57. doi: 10.1016/j.tics.2006.11.004. Epub 2006 Dec 22. PMID: 17188554.

(Bud) Craig, A. How do you feel — now? The anterior insula and human awareness. *Nat Rev Neurosci* 10, 59–70 (2009). https://doi.org/10.1038/nrn2555

Butler, E. A., Lee, T. L., & Gross, J. J. (2007). Emotion regulation and culture: Are the social consequences of emotion suppression culture-specific? *Emotion*, *7*(1), 30–48. https://doi.org/10.1037/1528-3542.7.1.30

Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavallee, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of Personality and Social Psychology, 70*(1), 141–156. https://doi.org/10.1037/0022-3514.70.1.141

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in cognitive sciences*, *7*(12), 547–552. https://doi.org/10.1016/j.tics.2003.10.005

Cowan N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *The Behavioral and brain sciences*, *24*(1), 87–185. https://doi.org/10.1017/s0140525x01003922

Cramer AO, Waldorp LJ, van der Maas HL, Borsboom D. Comorbidity: a network perspective. Behav Brain Sci. 2010 Jun;33(2-3):137-50; discussion 150-93. doi: 10.1017/S0140525X09991567. PMID: 20584369.

Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion Review, 8*(2), 144–153. https://doi.org/10.1177/1754073914558466

Davis, M. H. (1980). *A Multidimensional Approach to Individual Differences in Empathy*. ResearchGate; American Psychological Association. https://www.researchgate.net/publication/34891073_A_Multidimensional_Approach_to_Individual_Differences_in_Empathy

Decety, J., & Cowell, J. M. (2014). The complex relation between morality and empathy. *Trends in cognitive sciences*, *18*(7), 337–339. https://doi.org/10.1016/j.tics.2014.04.008

Decety, J., & Jackson, P. L. (2004). The Functional Architecture of Human Empathy. *Behavioral and Cognitive Neuroscience Reviews*, *3*(2), 71–100. https://doi.org/10.1177/1534582304267187

Decety, J., & Lamm, C. (2006). Human Empathy Through the Lens of Social Neuroscience. *The Scientific World JOURNAL*, *6*(6), 1146–1163. https://doi.org/10.1100/tsw.2006.221

Decety, J., Norman, G. J., Berntson, G. G., & Cacioppo, J. T. (2012). A neurobehavioral evolutionary perspective on the mechanisms underlying empathy. *Progress in neurobiology*, *98*(1), 38–48. https://doi.org/10.1016/j.pneurobio.2012.05.001

Doi, T. (1973). *The Anatomy of Dependence*. Kodansha.

Edmondson, A. (1999). Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly*, *44*(2), 350-383. https://doi.org/10.2307/2666999 (Original work published 1999)

Eisenberg, N. (2000). Emotion, Regulation, and Moral Development. *Annual Review of Psychology*, *51*(1), 665–697. https://doi.org/10.1146/annurev.psych.51.1.665

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160

Erikson, E. H. (1950). *Childhood and society.* W W Norton & Co.

Farb, N., Daubenmier, J., Price, C. J., Gard, T., Kerr, C., Dunn, B. D., Klein, A. C., Paulus, M. P., & Mehling, W. E. (2015). Interoception, contemplative practice, and health. *Frontiers in psychology*, *6*, 763. https://doi.org/10.3389/fpsyg.2015.00763

Figley, C. R. (1995). Compassion fatigue as secondary traumatic stress disorder: An overview. In C. R. Figley (Ed.), *Compassion fatigue: Coping with secondary traumatic stress disorder in those who treat the traumatized* (pp. 1–20). Brunner/Mazel.

Figley C. R. (2002). Compassion fatigue: psychotherapists' chronic lack of self care. *Journal of clinical psychology*, *58*(11), 1433–1441. https://doi.org/10.1002/jclp.10090

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9673–9678. https://doi.org/10.1073/pnas.0504136102

Friston, K. The free-energy principle: a unified brain theory?. *Nat Rev Neurosci* 11, 127–138 (2010). https://doi.org/10.1038/nrn2787

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Grandey, A. A. (2000). Emotional regulation in the workplace: A new way to conceptualize emotional labor. *Journal of Occupational Health Psychology, 5*(1), 95–110. https://doi.org/10.1037/1076-8998.5.1.95

Gross J. J. (2002). Emotion regulation: affective, cognitive, and social consequences. *Psychophysiology*, *39*(3), 281–291. https://doi.org/10.1017/s0048577201393198

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry, 26*(1), 1–26. https://doi.org/10.1080/1047840X.2014.940781

Gross, J. J., & Levenson, R. W. (1997). Hiding feelings: The acute effects of inhibiting negative and positive emotion. *Journal of Abnormal Psychology, 106*(1), 95–103. https://doi.org/10.1037/0021-843X.106.1.95

Grossmann, K., Grossmann, K. E., Kindler, H., & Zimmermann, P. (2008). A wider view of attachment and exploration: The influence of mothers and fathers on the development of psychological security from infancy to young adulthood. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (2nd ed., pp. 857–879). The Guilford Press.

Habermas, T., & Bluck, S. (2000). Getting a life: The emergence of the life story in adolescence. *Psychological Bulletin, 126*(5), 748–769. https://doi.org/10.1037/0033-2909.126.5.748

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., Brand, R., Brandt, M. J., Brewer, G., Bruyneel, S., Calvillo, D. P., Campbell, W. K., Cannon, P. R., Carlucci, M., Carruth, N. P., Cheung, T., Crowell, A., De Ridder, D. T. D., Dewitte, S., Elson, M., … Zwienenberg, M. (2016). A Multilab Preregistered Replication of the Ego-Depletion Effect. *Perspectives on psychological science : a journal of the Association for Psychological Science*, *11*(4), 546–573. https://doi.org/10.1177/1745691616652873

Haier, R. J., Siegel, B. V., Nuechterlein, K. H., Hazlett, E., Wu, J. C., Paek, J., Browning, H. L., & Buchsbaum, M. S. (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence, 12*(2), 199–217. https://doi.org/10.1016/0160-2896(88)90016-5

Helliwell, John & Wang, Shun. (2011). Trust and Wellbeing. International Journal of Wellbeing. 1. 10.5502/ijw.v1i1.3.

Herman, J. L. (1992). Trauma and Recovery: The Aftermath of Violence—From Domestic Abuse to Political Terror. New York: Basic Books.

Hirota, Y., Arai, A., Young, L.J., Osako, Y., Yuri, K., & Mitsui, S. (2020). Oxytocin receptor antagonist reverses the blunting effect of pair bonding on fear learning in monogamous prairie voles. Hormones and Behavior, 120, 104685. https://doi.org/10.1016/j.yhbeh.2020.104685

Hochschild, A. R. (1983). *The Managed Heart : Commercialization of Human Feeling*. University of California Press.

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge University Press.

James, W. (1890). The Principles of Psychology. New York: Henry Holt and Company the Principles of Psychology. http://dx.doi.org/10.1037/11059-000

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Kahneman, D. (2011). *Thinking, fast and slow.* Farrar, Straus and Giroux.

Kernberg, O. (1967). Borderline Personality Organization. *Journal of the American Psychoanalytic Association*, *15*(3), 641-685. https://doi.org/10.1177/000306516701500309

Kernberg, O. F. (2015). Neurobiological correlates of object relations theory: The relationship between neurobiological and psychodynamic development. *International Forum of Psychoanalysis, 24*(1), 38–46. https://doi.org/10.1080/0803706X.2014.912352

Kernis, M. H., & Goldman, B. M. (2006). A multicomponent conceptualization of authenticity: Theory and research. In M. P. Zanna (Ed.), *Advances in experimental social psychology,* Vol. 38, pp. 283–357). Elsevier Academic Press. https://doi.org/10.1016/S0065-2601(06)38006-9

Klimecki, O., & Singer, T. (2012). Empathic distress fatigue rather than compassion fatigue? Integrating findings from empathy research in psychology and social neuroscience. In B. Oakley, A. Knafo, G. Madhavan, & D. S. Wilson (Eds.), *Pathological altruism* (pp. 368–383). Oxford University Press.

Klimecki, O. M., Leiberg, S., Ricard, M., & Singer, T. (2014). Differential pattern of functional brain plasticity after compassion and empathy training. *Social cognitive and affective neuroscience*, *9*(6), 873–879. https://doi.org/10.1093/scan/nst060

LaLumiere, R.T., McGaugh, J.L., & McIntyre, C.K. (2017). Emotional modulation of learning and memory: Pharmacological implications. Pharmacological Reviews, 69(3), 236-255. https://doi.org/10.1124/pr.116.013474

Leiter, M., & Maslach, C. (2004, January 1). *Areas of Worklife: A Structured Approach to Organizational Predictors of Job Burnout*. ResearchGate. https://www.researchgate.net/publication/235297409_Areas_of_Worklife_A_Structured_Approach_to_Organizational_Predictors_of_Job_Burnout

Linehan, M. M. (1993). *Cognitive-behavioral treatment of borderline personality disorder.* Guilford Press.

Lisa Feldman Barrett. (2017). *How emotions are made the secret life of the brain*. Mariner Books.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Maslach C, Leiter MP. Understanding the burnout experience: recent research and its implications for psychiatry. World Psychiatry. 2016 Jun;15(2):103-11. doi: 10.1002/wps.20311. PMID: 27265691; PMCID: PMC4911781.

Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology, 52,* 397–422. https://doi.org/10.1146/annurev.psych.52.1.397

McAdams, D. P., & de St. Aubin, E. (1992). A theory of generativity and its assessment through self-report, behavioral acts, and narrative themes in autobiography. *Journal of Personality and Social Psychology, 62*(6), 1003–1015. https://doi.org/10.1037/0022-3514.62.6.1003

McAdams, D. P. (2001). The psychology of life stories. *Review of General Psychology, 5*(2), 100–122. https://doi.org/10.1037/1089-2680.5.2.100

McAdams, D. P. (2013). The Psychological Self as Actor, Agent, and Author. *Perspectives on Psychological Science*, *8*(3), 272-295. https://doi.org/10.1177/1745691612464657 (Original work published 2013)

McEwen BS. Allostasis and allostatic load: implications for neuropsychopharmacology. Neuropsychopharmacology. 2000 Feb;22(2):108-24. doi: 10.1016/S0893-133X(99)00129-3. PMID: 10649824.

McEwen BS, Stellar E. Stress and the individual. Mechanisms leading to disease. Arch Intern Med. 1993 Sep 27;153(18):2093-101. PMID: 8379800.

McGaugh, J.L. (2015). Consolidating memories. Annual Review of Psychology, 66, 1-24. https://doi.org/10.1146/annurev-psych-010814-014954

McLean, K. C., Pasupathi, M., & Pals, J. L. (2007). Selves creating stories creating selves: A process model of self-development. *Personality and Social Psychology Review, 11*(3), 262–278. https://doi.org/10.1177/1088868307301034

McLean, K. C., & Pasupathi, M. (2012). Processes of identity development: Where I am and how I got there. *Identity: An International Journal of Theory and Research, 12*(1), 8–28. https://doi.org/10.1080/15283488.2011.632363

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain structure & function*, *214*(5-6), 655–667. https://doi.org/10.1007/s00429-010-0262-0

Mercer SW, Maxwell M, Heaney D, Watt GC. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. Fam Pract. 2004 Dec;21(6):699-705. doi: 10.1093/fampra/cmh621. Epub 2004 Nov 4. PMID: 15528286.

Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin, 112*(2), 179–204. https://doi.org/10.1037/0033-2909.112.2.179

Mikulincer, M., & Shaver, P. R. (2007). *Attachment in adulthood: Structure, dynamics, and change.* The Guilford Press.

Milinkovic, B., & Aru, J. (2025). On biological and artificial consciousness: A case for biological computationalism. *Neuroscience and biobehavioral reviews*, 106524. Advance online publication. https://doi.org/10.1016/j.neubiorev.2025.106524

Mobley, D. (2025). HeartQuest: Advancing Functional Empathy in Non-Experiential Systems through MEC and HEART. SSRN. http://dx.doi.org/10.2139/ssrn.5382010

Nesse R. M. (1990). Evolutionary explanations of emotions. *Human nature (Hawthorne, N.Y.)*, *1*(3), 261–289. https://doi.org/10.1007/BF02733986

Nesse R. M. (2004). Natural selection and the elusiveness of happiness. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *359*(1449), 1333–1347. https://doi.org/10.1098/rstb.2004.1511

Neubauer AC, Fink A. Intelligence and neural efficiency. Neurosci Biobehav Rev. 2009 Jul;33(7):1004-23. doi: 10.1016/j.neubiorev.2009.04.001. Epub 2009 Apr 10. PMID: 19580915.

Northoff G, Heinzel A, de Greck M, Bermpohl F, Dobrowolny H, Panksepp J. Self-referential processing in our brain--a meta-analysis of imaging studies on the self. Neuroimage. 2006 May 15;31(1):440-57. doi: 10.1016/j.neuroimage.2005.12.002. Epub 2006 Feb 7. PMID: 16466680.

Northoff, G., Qin, P., & Feinberg, T. E. (2011). Brain imaging of the self--conceptual, anatomical and methodological issues. *Consciousness and cognition*, *20*(1), 52–63. https://doi.org/10.1016/j.concog.2010.09.011

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions.* Oxford University Press.

Pashler H. Dual-task interference in simple tasks: data and theory. Psychol Bull. 1994 Sep;116(2):220-44. doi: 10.1037/0033-2909.116.2.220. PMID: 7972591.

Richards, J. M., & Gross, J. J. (2000). Emotion regulation and memory: The cognitive costs of keeping one's cool. Journal of Personality and Social Psychology, 79(3), 410–424. https://doi.org/10.1037/0022-3514.79.3.410

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual review of neuroscience*, *27*, 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Salter Ainsworth, M. D., Blehar, M., Waters, E., & Wall, S. N. (1978). *Patterns of attachment : a psychological study of the strange situation*. Lawrence Erlbaum Associates.

Schmeichel BJ, Volokhov RN, Demaree HA. Working memory capacity and the self-regulation of emotional expression and experience. J Pers Soc Psychol. 2008 Dec;95(6):1526-40. doi: 10.1037/a0013345. PMID: 19025300.

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. Psychological Bulletin, 134(2), 207–222. https://doi.org/10.1037/0033-2909.134.2.207

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain : a journal of neurology*, *132*(Pt 3), 617–627. https://doi.org/10.1093/brain/awn279

Sheppes G, Gross JJ. Is timing everything? Temporal considerations in emotion regulation. Pers Soc Psychol Rev. 2011 Nov;15(4):319-31. doi: 10.1177/1088868310395778. Epub 2011 Jan 13. PMID: 21233326.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. The Quarterly Journal of Economics, 69, 99-118.https://doi.org/10.2307/1884852

Simon, H. A. (1956). Rational choice and the structure of the environment. Psychological Review, 63(2), 129–138. https://doi.org/10.1037/h0042769

Simon, H. A. (1978). Rationality as Process and as Product of Thought. The American Economic Review, 68, 1-16.

Singer, J. A. (2004). Narrative Identity and Meaning Making Across the Adult Lifespan: An Introduction. Journal of Personality, 72(3), 437–459. https://doi.org/10.1111/j.0022-3506.2004.00268.x

Singer T, Lamm C. The social neuroscience of empathy. Ann N Y Acad Sci. 2009 Mar;1156:81-96. doi: 10.1111/j.1749-6632.2009.04418.x. PMID: 19338504.

Skowron, E.A., Dendy, A.K. Differentiation of Self and Attachment in Adulthood: Relational Correlates of Effortful Control. *Contemporary Family Therapy* 26, 337–357 (2004). https://doi.org/10.1023/B:COFT.0000037919.63750.9d

Snyder, M. (1974). Self-monitoring of expressive behavior. Journal of Personality and Social Psychology, 30(4), 526–537. https://doi.org/10.1037/h0037039

Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis. Journal of Cognitive Neuroscience, 21(3), 489–510. https://doi.org/10.1162/jocn.2008.21029

Squire LR, Wixted JT. The cognitive neuroscience of human memory since H.M. Annu Rev Neurosci. 2011;34:259-88. doi: 10.1146/annurev-neuro-061010-113720. PMID: 21456960; PMCID: PMC3192650.

Stamm, B. H. (2010). The Concise ProQOL Manual (2nd ed.). Pocatello, ID: ProQOL.org.

Takie Sugiyama Lebra. (1976). *Japanese patterns of behavior*. University Of Hawaii Press, Cop.

Tang YY, Hölzel BK, Posner MI. The neuroscience of mindfulness meditation. Nat Rev Neurosci. 2015 Apr;16(4):213-25. doi: 10.1038/nrn3916. Epub 2015 Mar 18. PMID: 25783612.

Taylor, G. J., Bagby, R. M., & Parker, J. D. A. (1997). Disorders of affect regulation: Alexithymia in medical and psychiatric illness. Cambridge University Press. https://doi.org/10.1017/CBO9780511526831

Thayer JF, Lane RD. A model of neurovisceral integration in emotion regulation and dysregulation. J Affect Disord. 2000 Dec;61(3):201-16. doi: 10.1016/s0165-0327(00)00338-4. PMID: 11163422.

Tononi, G. An information integration theory of consciousness. BMC Neurosci 5, 42 (2004). https://doi.org/10.1186/1471-2202-5-42

van Doeselaar, L., Klimstra, T. A., Denissen, J. J. A., Branje, S., & Meeus, W. (2018). The role of identity commitments in depressive symptoms and stressful life events in adolescence and young adulthood. *Developmental psychology*, *54*(5), 950–962. https://doi.org/10.1037/dev0000479

van der Hart, O., Nijenhuis, E. R. S., & Steele, K. (2006). The haunted self: Structural dissociation and the treatment of chronic traumatization. W. W. Norton & Company.

van der Kolk, B. A. (2014). The body keeps the score: Brain, mind, and body in the healing of trauma. Viking.

Vermeulen, N., Luminet, O., & Corneille, O. (2006). Alexithymia and the automatic processing of affective information: Evidence from the affective priming paradigm. *Cognition and Emotion, 20*(1), 64–91. https://doi.org/10.1080/02699930500304654

Walker, C. S., Li, L., Baracchini, G., Tremblay-Mercier, J., Spreng, R. N., PREVENT-AD Research Group, & Geddes, M. R. (2023). The influence of generativity on purpose in life is mediated by social support and moderated by prefrontal functional connectivity in at-risk older adults. *bioRxiv : the preprint server for biology*, 2023.02.26.530089. https://doi.org/10.1101/2023.02.26.530089

Waters, Everett & Merrick, Susan & Treboux, Dominique & Crowell, Judith & Albersheim, Leah. (2000). Attachment Security in Infancy and Early Adulthood: A Twenty-Year Longitudinal Study. Child Development. 71. 684 - 689. 10.1111/1467-8624.00176.

Wegner DM, Zanakos S. Chronic thought suppression. J Pers. 1994 Dec;62(4):616-40. doi: 10.1111/j.1467-6494.1994.tb00311.x. PMID: 7861307.

Westbrook, A., Braver, T.S. Cognitive effort: A neuroeconomic approach. *Cogn Affect Behav Neurosci* 15, 395–415 (2015). https://doi.org/10.3758/s13415-015-0334-y

Whitfield-Gabrieli S, Moran JM, Nieto-Castañón A, Triantafyllou C, Saxe R, Gabrieli JD. Associations and dissociations between default and self-reference networks in the human brain. Neuroimage. 2011 Mar 1;55(1):225-32. doi: 10.1016/j.neuroimage.2010.11.048. Epub 2010 Nov 25. PMID: 21111832.

Winnicott, D. W. (1953). Transitional objects and transitional phenomena; a study of the first not-me possession. *The International Journal of Psychoanalysis, 34,* 89–97.

Wood, A. M., Linley, P. A., Maltby, J., Baliousis, M., & Joseph, S. (2008). The authentic personality: A theoretical and empirical conceptualization and the development of the Authenticity Scale. Journal of Counseling Psychology, 55(3), 385–399. https://doi.org/10.1037/0022-0167.55.3.385

Zeigarnik, B. (1938). On finished and unfinished tasks. In W. D. Ellis (Ed.), A source book of Gestalt psychology (pp. 300–314). Kegan Paul, Trench, Trubner & Company. https://doi.org/10.1037/11496-025

Zhang, H., Skelin, I., Ma, S., Paff, M., Mnatsakanyan, L., Yassa, M.A., Knight, R.T., & Lin, J.J. (2024). Awake ripples enhance emotional memory encoding in the human brain. Nature Communications, 15(1), 215. https://doi.org/10.1038/s41467-023-44295-8