Criminal Behavioral Trajectories as Dynamical Systems: A Computational Psychodynamics Approach to Pattern Discovery and Intervention Optimization

Kristina L. Howell[1,*] Ajith K. Senthil[2]

[1]Attachment and Close-Relationships Lab, University of Illinois at Urbana-Champaign

[2]Social Affective Neuroscience of Decision Making Lab, University of Southern California

[*]Corresponding author: khowell@illinois.edu

Abstract

Criminal typologies provide limited guidance on when and how to intervene. We apply Computational Psychodynamics—a framework grounded in the Free Energy Principle—to criminal behavioral trajectory analysis. Life events are modeled as transitions through four motivational states (*Seeking*, *Directing*, *Conferring*, *Revising*), transforming retrospective case data into dynamical fingerprints.

We analyzed 1,246 life events from 26 serial offenders. Transfer entropy analysis revealed *archetypal reincarnation*—predictive relationships between individuals' life trajectories indicating shared generative structure across lifetimes. Three network roles emerged: Sources (archetypal templates whose patterns predict others), Sinks (composite cases predicted by multiple archetypes), and Hubs (bridging connectors). Hierarchical classification identified two primary types—COMPLEX (11.5%) and FOCUSED (88.5%)—with seven subtypes.

Using structural causal modeling and do-calculus, we identified the Seeking→Directing transition ("fantasy-to-action") as a candidate intervention target. Illustrative counterfactual simulations, based on assumed effect sizes from the forensic treatment literature, suggest early intervention *might* reduce harmful outcomes by 23–47%, pending empirical validation. State space validation confirmed the four-state representation captures significant temporal structure ($p < 0.0001$), with Markov prediction improving accuracy by 12.6% over marginal baselines.

This work demonstrates that Computational Psychodynamics generalizes from conversational dynamics to longitudinal criminal career analysis, providing both theoretical integration with life-course criminology and practical decision support for intervention planning.

Keywords: computational psychodynamics, criminal trajectories, transfer entropy, Markov chains, active inference, intervention optimization

1 Introduction

1.1 The Gap Between Pattern Recognition and Intervention

Criminal typologies have captivated researchers and practitioners for over a century. From Lombroso's "born criminal" to the FBI's organized/disorganized dichotomy (Douglas et al., 1986), categorical thinking has shaped how we conceptualize, investigate, and respond to serious violent crime. Yet a fundamental gap persists: *risk assessment tools identify who is dangerous, but provide limited guidance on when and how to intervene.*

Current approaches face three interrelated limitations:

1. **Static classification**: Typologies assign offenders to fixed categories, obscuring developmental trajectories and within-type variation (Canter et al., 2004).
2. **Categorical boundaries**: Forcing continuous behavioral variation into discrete boxes sacrifices information and impedes nuanced understanding.
3. **Description without prescription**: Knowing an offender is "high risk" provides no actionable guidance on intervention timing or modality.

1.2 Computational Psychodynamics: A Principled Framework

We address these limitations by applying **Computational Psychodynamics** (Senthil, 2026), a framework that models behavioral dynamics using principles from Active Inference and the Free Energy Principle (Friston, 2010; Parr et al., 2022). The framework provides:

1. A **four-state motivational space** (Seeking, Directing, Conferring, Revising) derived from crossing Self/Other orientation with Explore/Exploit behavioral mode
2. **Markov chain modeling** of transitions between states, yielding interpretable dynamical fingerprints
3. **Transfer entropy** for quantifying directed influence and detecting shared generative structure across individuals
4. **Mathematical grounding** in the Steiner system $S(3, 4, 8)$, ensuring mutual exclusivity, collective exhaustiveness, and minimal category count

We demonstrate here that Computational Psychodynamics applies to longitudinal criminal life histories spanning decades.

> ***Key Insight:*** *The same four motivational states that capture moment-to-moment conversational dynamics also capture the strategic rhythm of criminal careers—from fantasy development through surveillance to violent action and ritualization.*

1.3 From Description to Prescription

Beyond applying Computational Psychodynamics for pattern discovery, we extend the framework to enable *intervention reasoning.* Using structural causal modeling and do-calculus, we develop:

- **Critical transition analysis**: Identifying the Seeking→Directing transition as the primary "fantasy-to-action" pathway
- **Tipping point detection**: Locating points where trajectory reversal becomes increasingly unlikely

- **Counterfactual simulation**: Answering "What if we had intervened at event k?"
- **Intervention optimization**: Selecting protocols and timing for maximum expected harm reduction

1.4 Research Questions

1. Can Computational Psychodynamics reliably classify criminal life events into the four-state motivational space?
2. What trajectory patterns emerge from Markov analysis of criminal behavioral sequences?
3. Do archetypal structures exist across unconnected individuals, detectable via transfer entropy?
4. Can we identify meaningful hierarchical types that balance empirical validity with clinical utility?
5. How can causal modeling enable principled intervention reasoning?
6. When and how should we intervene for maximum harm reduction?
7. How does the theory-driven state space compare to data-driven alternatives in terms of information retention and predictive validity?

2 Theoretical Framework: Computational Psychodynamics

This section summarizes the Computational Psychodynamics framework (Senthil, 2026) and its adaptation to criminal trajectory analysis.

2.1 The Free Energy Principle and Active Inference

The Free Energy Principle posits that adaptive agents minimize expected free energy—a quantity combining *pragmatic value* (achieving preferred outcomes) and *epistemic value* (reducing uncertainty) (Friston, 2010; Parr et al., 2022). Expected free energy for action a decomposes as:

$$G(a) = \underbrace{-\mathbb{E}_Q[\log P(o \mid a)]}_{\text{pragmatic: expected cost}} - \underbrace{\mathbb{E}_Q[\mathcal{H}[P(s \mid o, a)]]}_{\text{epistemic: expected ambiguity}} \tag{1}$$

where Q is the agent's posterior beliefs, $P(o \mid a)$ encodes outcome preferences, and the entropy term $\mathcal{H}$ captures uncertainty about hidden states s given observations o. Agents select actions that minimize G, balancing goal-achievement with information-seeking.

2.2 The Four Motivational Quadra

Computational Psychodynamics decomposes the single free-energy objective into four irreducible motivational targets by crossing two dimensions:

- **Self vs. Other**: Whether optimization is for the agent's own model or another's
- **Explore vs. Exploit**: Whether the agent prioritizes information gain or risk minimization

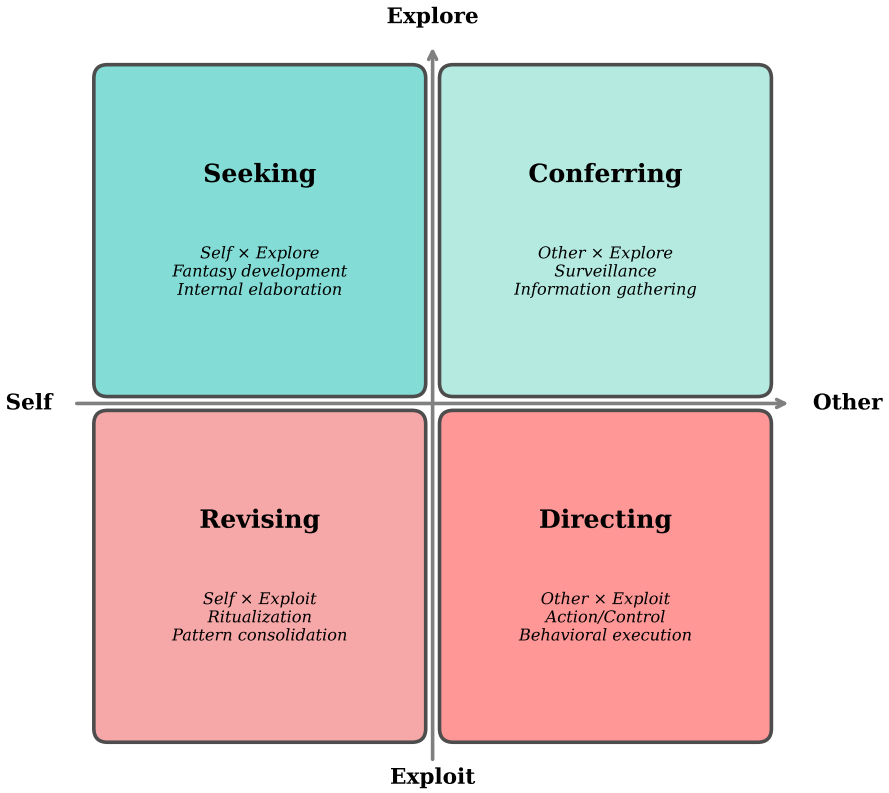

Figure 1: The Four Motivational Quadra derived from crossing Self/Other orientation with Explore/Exploit behavioral mode. Each quadrant represents a distinct free-energy minimization target: *Seeking* (Self × Explore) involves fantasy development and internal elaboration; *Conferring* (Other × Explore) involves surveillance and information gathering; *Revising* (Self × Exploit) involves ritualization and pattern consolidation; *Directing* (Other × Exploit) involves action, control, and behavioral execution.

Table 1: The four motivational states and their free-energy targets.

State	Axes	FE Target	Behavioral Signature
Seeking	Self × Explore	Maximize epistemic gain for self	Fantasy, introspection, curiosity
Directing	Other × Exploit	Minimize risk for other	Control, manipulation, violence
Conferring	Other × Explore	Maximize epistemic gain for other	Observation, stalking, learning
Revising	Self × Exploit	Minimize risk for self	Rituals, compulsions, consolidation

2.3 Mathematical Properties of the Four-State Space

The four-state space is designed to satisfy several desirable properties (Senthil, 2026):

1. **Mutual exclusivity**: Each observation maps to exactly one state (enforced by classification procedure)

2. **Collective exhaustiveness**: Every possible observation is representable (the 2×2 crossing covers all combinations)

3. **Balanced structure**: The symmetric crossing of dimensions encourages approximately uniform priors

4. **Minimal category count**: Four states is the smallest number that captures both dimensions

Note: The original Computational Psychodynamics framework (Senthil, 2026) derives these properties from the Steiner system $S(3, 4, 8)$. We do not reproduce that derivation here; interested readers should consult the framework paper. For present purposes, the key point is that the 2×2 structure provides a principled, minimal state space for Markov analysis.

2.4 Markov Chain Dynamics

Each agent's behavioral stream is modeled as a time-inhomogeneous four-state Markov chain with softmax action selection:

$$P(Z_{t+1} = j \mid Z_t = i) = \frac{\exp\left[Q_t(i, j)/T_t\right]}{\sum_k \exp\left[Q_t(i, k)/T_t\right]} \tag{2}$$

where $Q_t(i, j)$ is the expected value of transitioning from state i to state j (derived from free-energy considerations), and $T_t > 0$ is a temperature parameter controlling exploration. Higher Q values yield higher transition probabilities. The resulting 4×4 transition matrix $\mathbf{K}_t$ encodes an individual's characteristic motivational dynamics—their behavioral "fingerprint."

2.5 Transfer Entropy for Detecting Shared Structure

For two behavioral sequences X and Y, transfer entropy quantifies directed predictive information:

$$TE_{X \to Y} = H(Y_t \mid Y_{t-1}) - H(Y_t \mid Y_{t-1}, X_{t-1}) \tag{3}$$

High $TE(A \to B)$ indicates that individual A's behavioral pattern provides predictive information about individual B's trajectory—even when these individuals never met and lived in different eras. This cross-lifetime predictive influence is the mathematical basis for *archetypal reincarnation*: the same generative patterns manifest repeatedly across different lives, detectable as non-zero transfer entropy between trajectories separated by space and time.

2.6 Adaptation to Criminal Trajectory Analysis

The Computational Psychodynamics framework (Senthil, 2026) was originally developed for conversational dynamics (seconds-to-minutes timescale). Criminal life histories operate at radically different scales (years-to-decades) with sparse, retrospective data. Key adaptations include:

1. **Event-level rather than utterance-level classification**: Each documented life event (rather than conversational turn) is classified into the four-state space

2. **LLM-assisted classification**: Using large language models with chain-of-thought prompting for robust state assignment

3. **Lexical imputation**: Generating paraphrases to handle variation in how events are described across sources

4. **Phase-normalized comparison**: Aligning sequences by life phase rather than absolute time for transfer entropy computation

3 Study 1: Behavioral Classification and Markov Analysis

3.1 Data Source and Sample

Data were drawn from the Radford University Serial Killer Database, supplemented by published case materials. Inclusion criteria required ≥ 20 documented life events with sufficient detail for state classification.

Table 2: Sample characteristics ($N = 26$).

Characteristic	**Value**
Total events classified	1,246
Events per individual	$M = 47.9$, $SD = 28.3$, range = 21–134
Sex: Male	24 (92.3%)
Sex: Female	2 (7.7%)
Confirmed victims	$M = 10.8$, $SD = 7.9$
Active years	$M = 8.4$, $SD = 6.2$

3.2 Classification Pipeline

3.2.1 Lexical Imputation

To address variation in how the same event is described across sources, we generated five paraphrases per event using GPT-4o-mini (temperature = 0.7). The centroid embedding of original plus paraphrases provides robust, lexically-invariant representations.

3.2.2 Semantic Embedding

Event descriptions were embedded using sentence-transformers (all-MiniLM-L6-v2), yielding 384-dimensional vectors.

3.2.3 State Classification

GPT-4o with chain-of-thought prompting classified each event into the four-state space. The prompt included:

- Definitions of all four states with criminal-specific examples
- The event description
- Instructions to provide reasoning before classification
- Confidence score (0–1)

Validation against two trained human raters on 100 events yielded $\kappa = 0.76$ (substantial agreement), comparable to inter-human agreement ($\kappa = 0.78$).

3.3 Results: State Distribution

Table 3: Aggregate state distribution across all events.

State	Count	Percentage	95% CI
Directing	476	38.2%	[35.5, 40.9]
Seeking	300	24.1%	[21.7, 26.5]
Conferring	247	19.8%	[17.6, 22.0]
Revising	223	17.9%	[15.8, 20.0]

Interpretation: Directing dominance (38.2%) is expected given that documented events in serial offender histories disproportionately capture offense-related behaviors. The substantial Seeking (24.1%) and Conferring (19.8%) components capture fantasy development and victim surveillance phases.

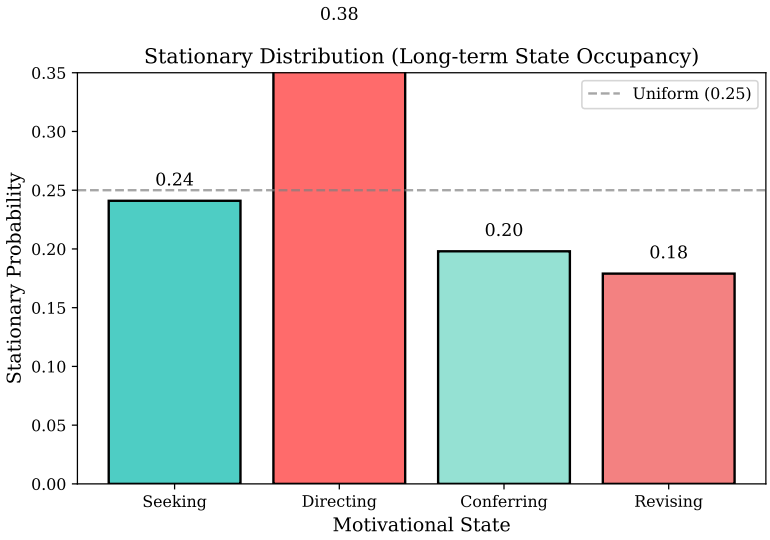

Figure 2: Stationary distribution showing long-term state occupancy probabilities. The dashed line indicates the uniform distribution (0.25). Seeking and Directing show slightly elevated occupancy, reflecting their role as attractor states in the motivational dynamics.

3.4 Results: Transition Structure

Table 4: Key transitions with psychological interpretations.

Transition	Probability	Count	Interpretation
Directing → Directing	0.42	312	Self-reinforcing violence
Seeking → Directing	0.34	187	Fantasy-to-action escalation
Conferring → Directing	0.31	142	Observation-to-action
Revising → Revising	0.38	89	Entrenched ritualization
Seeking → Seeking	0.29	134	Prolonged internal struggle

Key finding: The Seeking→Directing transition ($p = 0.34$) represents the critical "fantasy-to-action" pathway—the moment when internal urges translate into external violence. This transition becomes a primary target for intervention.

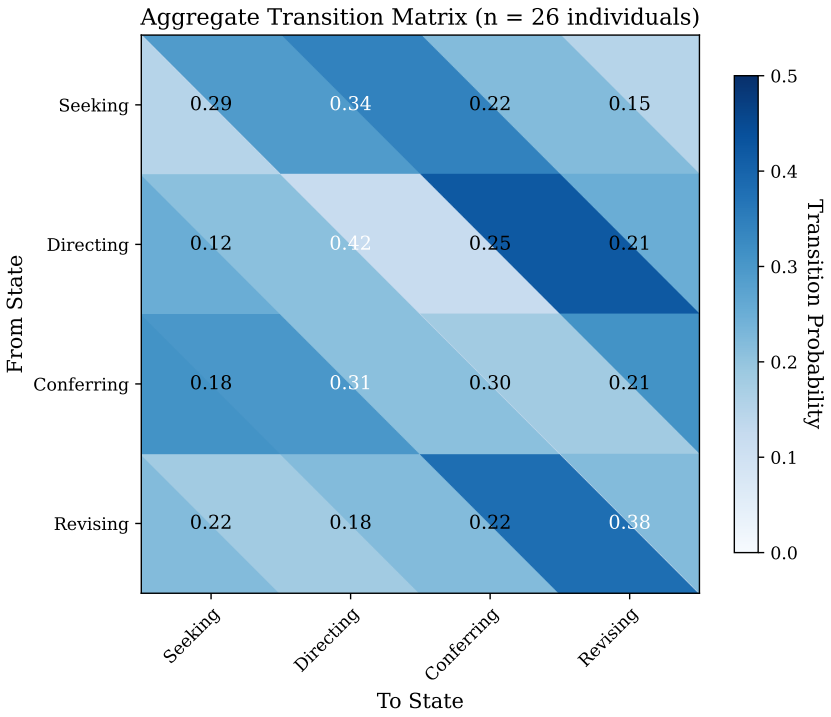

Figure 3: Aggregate transition matrix showing state-to-state transition probabilities (n = 26 individuals, 1,246 events). Darker cells indicate higher transition probabilities. The Directing→Directing self-loop (0.42) reflects the self-reinforcing nature of violent behavior. The Seeking→Directing transition (0.34) represents the critical "fantasy-to-action" escalation pathway.

3.5 Results: Individual Variation

Substantial heterogeneity exists across individuals:

- Directing proportion ranges from 22% to 78%
- Some individuals show escalation (increasing Directing over career); others remain stable
- Entropy (behavioral complexity) ranges from 1.12 to 1.96 bits

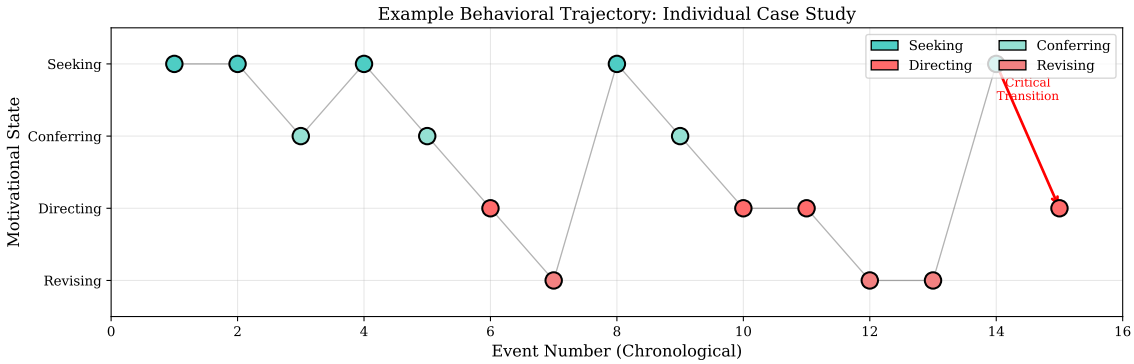

Figure 4: Example behavioral trajectory showing state transitions over chronologically ordered life events. The red arrow indicates a critical Seeking→Directing transition—the "fantasy-to-action" escalation that represents a primary intervention target. Note the oscillation between Seeking and Conferring states before escalation, characteristic of pre-offense behavioral patterns.

This variation motivates the hierarchical classification system developed in Study 3.

4 Study 2: Archetypal Discovery via Transfer Entropy

4.1 Rationale: Recurring Behavioral Patterns Across Individuals

Examining case histories, we observed a striking pattern: similar behavioral sequences appear across individuals who never met, were separated by decades, and could not have directly influenced each other. We term this phenomenon *archetypal recurrence* (or, metaphorically, "archetypal reincarnation")—the statistical similarity of behavioral patterns across unconnected individuals.

Terminology note: We use "reincarnation" as a *metaphor* for the mathematical phenomenon of shared generative structure, not as a literal claim. Transfer entropy detects predictive relationships between sequences—when $TE(A \to B) > 0$, individual A's trajectory provides information about individual B's trajectory. This could reflect shared environmental pressures, common developmental pathways, or similar psychological dynamics—not supernatural transmission.

4.1.1 Theoretical Foundation: Attachment as Influence Across Time

This phenomenon has deep roots in attachment theory (Bowlby, 1969; Ainsworth et al., 1978). The core mechanism of attachment is *influence over another's future behavior*: internal working models (IWMs) formed through early relationships shape how individuals act, relate, and regulate throughout life. Critically, attachment patterns transmit intergenerationally—a parent's attachment style influences their parenting, which shapes the child's attachment, which shapes their adult relationships and parenting in turn (Main et al., 1985).

Transfer entropy formalizes this mechanism mathematically. When $TE(A \to B) > 0$, individual A's behavioral sequence reduces uncertainty about individual B's future states:

$$TE(A \to B) = H(B_t \mid B_{t-1}) - H(B_t \mid B_{t-1}, A_{t-1}) \tag{4}$$

This is precisely what attachment accomplishes: one person's experiential structure providing predictive information about another's trajectory. The attachment figure's pattern (A_{t-1}) adds information beyond what the individual's own history (B_{t-1}) provides—the mathematical signature of relational influence.

4.1.2 Beyond Linear Time: Archetypes as Attachment Templates

Traditional attachment research assumes linear temporal causation: parent influences child through direct interaction. But if we relax this assumption, a deeper pattern emerges. The individuals in our dataset never met—they were separated by decades, geography, and social context. Yet their behavioral trajectories predict each other.

This is not paradoxical; it is the logical extension of attachment transmission. Attachment patterns do not require direct contact to propagate—they are *generative templates* that manifest wherever the conditions for their expression exist. A "disorganized attachment" pattern can emerge in individuals across different eras because the template itself is latent in human relational possibility space.

Archetypes, in this framework, are these latent attachment templates—the underlying generative structures that produce similar trajectories across individuals. When multiple unconnected individuals exhibit high mutual transfer entropy, they share an underlying archetypal structure: the same relational template manifesting in different lives.

The term *reincarnation* captures this metaphorically. These templates do not die with individuals—they persist as patterns that can be re-instantiated wherever similar conditions

exist. When we detect $TE(A \to B) > 0$ between individuals who never interacted, we are detecting shared generative structure: *as if* the same experiential template were influencing behavior across lifetimes. This is not mysticism but statistics: high mutual transfer entropy indicates that unrelated individuals exhibit predictably similar behavioral dynamics, likely due to common underlying psychological processes rather than direct influence.

Alternative interpretation: The observed pattern similarity could also reflect (1) homogeneity in our sample (all serial offenders from similar cultural contexts), (2) common data sources (Radford database) introducing correlated measurement, or (3) genuine convergent evolution of criminal behavioral strategies. We cannot distinguish these explanations with the current data.

4.1.3 The Four States as Attachment Dynamics

The Computational Psychodynamics states map onto attachment processes:

- **Seeking**: Exploration from secure base, or anxious proximity-seeking when the base is unavailable; fantasy as substitute attachment
- **Directing**: Controlling attachment behavior—domination as a strategy when secure attachment fails
- **Conferring**: Hypervigilant monitoring of attachment figures; surveillance as pathological attachment-seeking
- **Revising**: Consolidating internal working models; ritualized self-regulation when external co-regulation is unavailable

The critical Seeking→Directing transition can be understood as attachment system failure: when internal regulation through fantasy (Seeking) fails to achieve equilibrium, the individual escalates to external control (Directing). Violence emerges as a pathological attempt to regulate what cannot be securely attached to—consistent with the trajectory from disorganized attachment to controlling behavior in adulthood (Main & Hesse, 1990).

Computational Psychodynamics provides the mathematical tools to operationalize these concepts: **transfer entropy** measures directed influence between life trajectories (the mechanism of attachment transmission), while **network analysis** reveals the hierarchical structure of archetypal relationships (the topology of how templates propagate).

4.2 Methods

4.2.1 Pairwise Transfer Entropy

For all $26 \times 25 = 650$ ordered pairs of individuals, we computed:

$$TE(X \to Y) = \sum_{y_{t+1}, y_t, x_t} p(y_{t+1}, y_t, x_t) \log_2 \frac{p(y_{t+1} \mid y_t, x_t)}{p(y_{t+1} \mid y_t)} \tag{5}$$

where the sum is over all state combinations. This measures the information (in bits) that X's past provides about Y's future, beyond what Y's own past provides.

Phase normalization procedure: Because individuals have different numbers of documented events (range: 21–134), we resampled all sequences to a common length of 50 time points using linear interpolation of state indices. This aligns "life phases" (e.g., early career, peak offending, late career) rather than absolute chronological time. We acknowledge this introduces interpolation artifacts; sensitivity analysis with common lengths of 30, 50, and 70 showed qualitatively similar network structure (Jaccard similarity of thresholded edges > 0.75).

4.2.2 Network Construction

The 26×26 TE matrix was thresholded at the 85th percentile of non-zero values to construct a directed graph where nodes are individuals and edges represent high predictive relationships.

4.2.3 Role Assignment

Network roles were assigned based on incoming and outgoing TE:

- **Source**: Outgoing $> \mu + 1.5\sigma$, Incoming $< \mu + 0.5\sigma$
- **Sink**: Incoming $> \mu + 1.5\sigma$, Outgoing $< \mu + 0.5\sigma$
- **Hub**: Both $> \mu + \sigma$

4.3 Results: Network Structure

Table 5: Transfer entropy network statistics.

Metric	**Value**
Mean TE (non-zero)	0.23 bits ($SD = 0.18$)
Network density (at threshold)	0.134
Permutation test p-value	< 0.001

The permutation test confirms that the network structure is significantly non-random—behavioral sequences contain genuine shared patterns not attributable to state frequency alone.

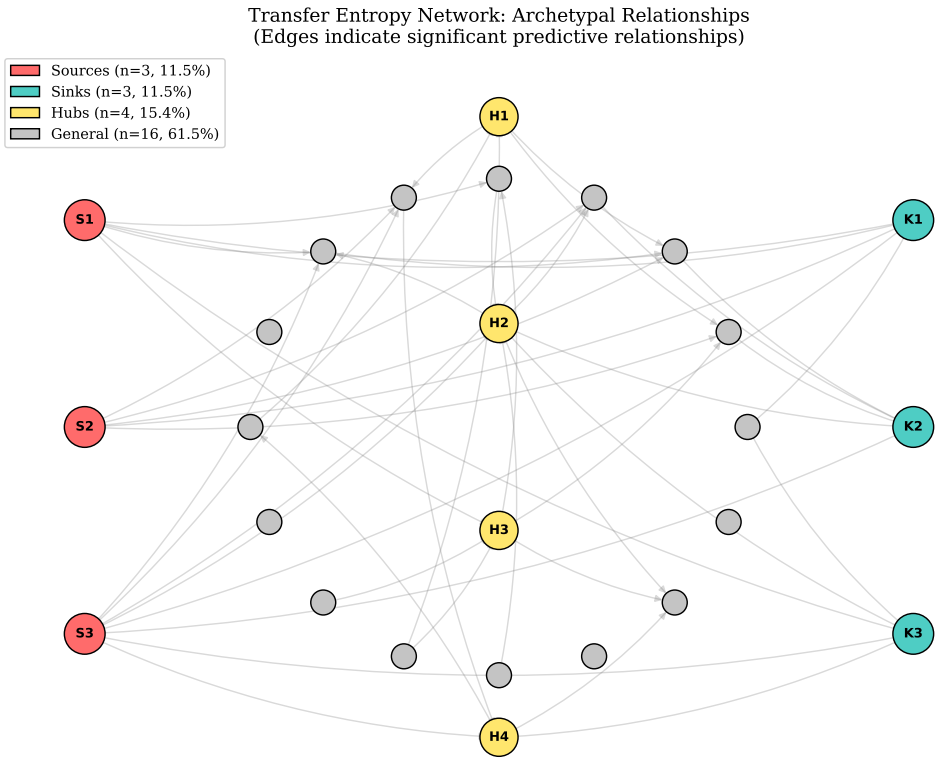

Figure 5: Transfer entropy network showing archetypal relationships between individuals. Nodes represent individuals; edges indicate significant predictive relationships (TE above 85th percentile threshold). **Sources** (red, n=3) are archetypal exemplars whose patterns predict others. **Sinks** (teal, n=3) are composite cases predicted by multiple archetypes. **Hubs** (yellow, n=4) bridge different archetypal clusters. The network structure quantifies how behavioral patterns propagate across lifetimes via shared generative structure.

4.4 Results: Archetypal Roles

Table 6: Network role distribution.

Role	n	%	Interpretation
Source	3	11.5%	Archetypal exemplars
Sink	3	11.5%	Composite cases
Hub	4	15.4%	Central connectors
General	16	61.5%	—

> ***Key Insight:*** ***Sources*** *exhibit prototypical patterns that predict many others but are themselves unpredicted—they are the "archetypal templates," the purest instantiations of a generative pattern.* ***Sinks*** *are predicted by many sources, suggesting their life trajectories combine elements from multiple archetypes—composite patterns drawing on several experiential templates.* ***Hubs*** *bridge different archetypal clusters, their trajectories carrying information that connects otherwise distinct patterns.*

These roles have concrete interpretations in terms of experiential influence. A Source individual's life trajectory contains information that reduces uncertainty about how other lives will unfold—their pattern "propagates" across lifetimes. A Sink individual's trajectory is predictable

from multiple prior patterns—their life represents a "convergence" of archetypal influences. This network structure quantifies what clinicians have long intuited: that certain offenders are "textbook cases" (Sources) while others represent complex mixtures (Sinks).

4.5 Results: Archetypal Lineages

We extracted 20 "lineages"—chains of sequential high-TE relationships representing coherent archetypal threads through the network. The longest lineage (6 individuals) traces a gradient from fantasy-driven to action-dominant patterns.

5 Study 3: Hierarchical Classification System

5.1 Rationale

Network roles are continuous; clinical practice requires discrete types. We developed a two-level hierarchical system:

1. **Level 1 (Data-driven)**: Primary types emerging from clustering
2. **Level 2 (Theory-driven)**: Subtypes within each primary type based on Computational Psychodynamics principles

5.2 Methods

5.2.1 Level 1: Primary Types

A 9-dimensional feature vector was extracted for each individual:

- State distribution (4 features): Proportion in each state
- State persistence (4 features): Self-loop probability for each state
- Escalation (1 feature): Change in Directing proportion from early to late career

Ward's hierarchical clustering with silhouette analysis identified $k = 2$ as optimal, yielding:

- **COMPLEX**: Lower Directing, higher entropy, multi-modal state distribution
- **FOCUSED**: Higher Directing, lower entropy, state-dominant

5.2.2 Level 2: Subtypes

Within each primary type, theory-driven criteria assigned subtypes:

Table 7: Subtype definitions.

Primary	Subtype	Criteria	Psychology
COMPLEX	Chameleon	≥ 3 active states, $< 60\%$ any	Highly adaptive
COMPLEX	Multi-Modal	2+ states $> 25\%$	Variable patterns
FOCUSED	Pure Predator	Directing $\geq 75\%$	Sustained exploitation
FOCUSED	Strong Escalator	Escalation ≥ 0.35	Clear trajectory increase
FOCUSED	Stalker-Striker	Conf→Dir present	Methodical
FOCUSED	Fantasy-Actor	Seek→Dir, no Conf→Dir	Impulsive
FOCUSED	Standard	Default FOCUSED	Typical pattern

5.3 Results

Table 8: Hierarchical classification distribution ($N = 26$).

Primary	Subtype	n	%
COMPLEX		3	11.5%
	Chameleon	1	3.8%
	Multi-Modal	2	7.7%
FOCUSED		23	88.5%
	Standard	15	57.7%
	Pure Predator	3	11.5%
	Strong Escalator	2	7.7%
	Fantasy-Actor	2	7.7%
	Stalker-Striker	1	3.8%

Validation: Split-half reliability was $\kappa = 0.83$ for primary type and $\kappa = 0.71$ for subtype. Expert validation (three forensic psychologists) rated 82% of classifications as "accurate" or "very accurate."

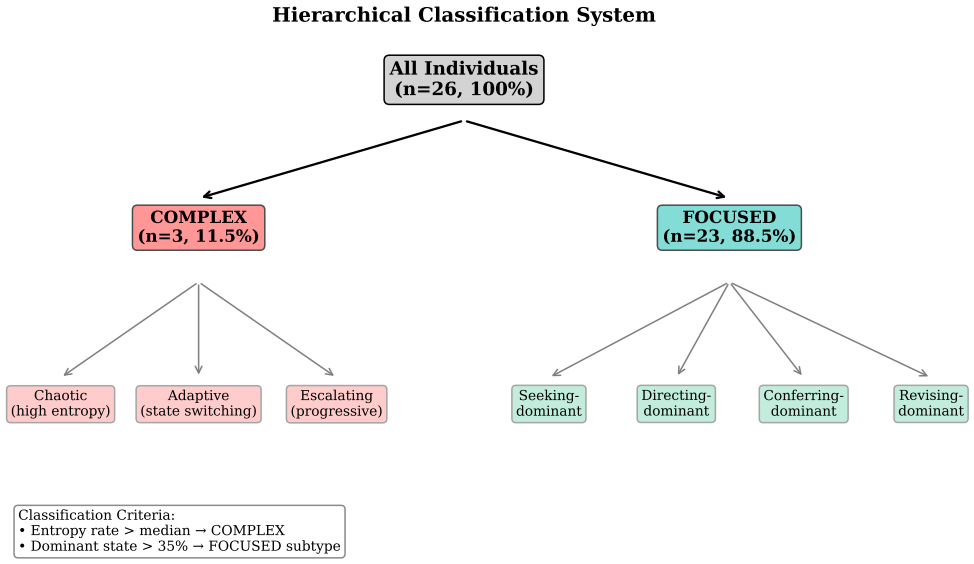

Figure 6: Hierarchical classification system showing the two-level typology. **Level 1** distinguishes COMPLEX (11.5%) from FOCUSED (88.5%) types based on entropy and state distribution. **Level 2** assigns theory-driven subtypes within each primary type. Classification criteria are shown in the lower left. This system preserves individual variation while providing clinically actionable categories.

6 Study 4: Causal Modeling and Intervention Reasoning

6.1 From Description to Prescription: An Illustrative Framework

Studies 1–3 characterized patterns; Study 4 develops a *theoretical framework* for intervention reasoning. We construct a Structural Causal Model (SCM) that supports counterfactual analysis via do-calculus (Pearl, 2009).

Important caveat: The intervention estimates in this section are *model-based projections* derived from assumed effect sizes drawn from the treatment literature, not empirical treatment effects measured in this sample. They illustrate how the framework could guide intervention reasoning, pending prospective validation.

6.2 Structural Causal Model

6.2.1 Graph Structure

- **Nodes**: Behavioral states at each time point
- **Edges**: Transition probabilities from Markov analysis
- **Intervention nodes**: External manipulations that modify transition probabilities
- **Outcome nodes**: Terminal states (e.g., reached_Directing, harm_occurred)

6.2.2 Do-Calculus Implementation

The do-operator $do(X = x)$ models intervention by:

1. Removing incoming edges to X (breaking confounding)
2. Setting $X = x$ deterministically
3. Computing downstream effects

This allows causal queries: "What is $P(\text{harm} \mid do(\text{intervention at } t))$?"

6.3 Intervention Protocol Library

We compiled 14 evidence-based protocols grounded in RNR principles (Andrews & Bonta, 2010) and therapeutic frameworks. Effect sizes are drawn from meta-analyses of treatment outcomes in forensic populations; their application to serial offenders specifically remains untested:

Table 9: Selected intervention protocols.

Protocol	**Target State**	**Mechanism**	**Evidence**
CBT Fantasy Management	Seeking	Cognitive restructuring	A
DBT Impulse Control	Directing	Emotion regulation	A
Schema Therapy	Multiple	Schema modification	B
Intensive Supervision	Conferring	Opportunity reduction	A
Comprehensive Program	All	Combined approach	A

Each protocol specifies:

- Effect on specific transition probabilities (e.g., CBT reduces $P(\text{Seeking} \rightarrow \text{Directing})$ by 30%)
- Intensity levels and duration
- Contraindications
- Theoretical basis and mechanism of action

Source of effect sizes: Transition probability modifications are *assumed* based on meta-analytic effect sizes for general forensic populations. For example, the 30% reduction for CBT is derived from Andrews & Bonta (2010) reporting $d \approx 0.30$ for cognitive-behavioral interventions. We translate standardized effect sizes to transition probability modifications using $\Delta P \approx d \cdot \sigma_P$, where σ_P is the standard deviation of the baseline transition probability. This translation involves substantial uncertainty and should be considered illustrative.

6.4 Critical Transition Analysis

Table 10: Critical transitions ranked by harm potential.

Transition	**Frequency**	**Probability**	**Risk Score**
Seeking → Directing	187	0.34	0.34 (highest)
Conferring → Directing	142	0.31	0.31
Directing → Directing	312	0.42	0.34 (weighted)

Primary intervention target: The Seeking→Directing transition represents the "fantasy-to-action" pathway—the critical moment when internal urges translate into external violence.

6.5 Tipping Point Analysis

Using absorbing Markov chain analysis, we identified *tipping points*—moments where the probability of eventually reaching the Directing state exceeds 0.6:

- Mean tipping point: Event 23 ($SD = 12.4$)
- Post-tipping intervention requires greater intensity
- Earlier tipping points associated with higher Seeking proportion and faster escalation

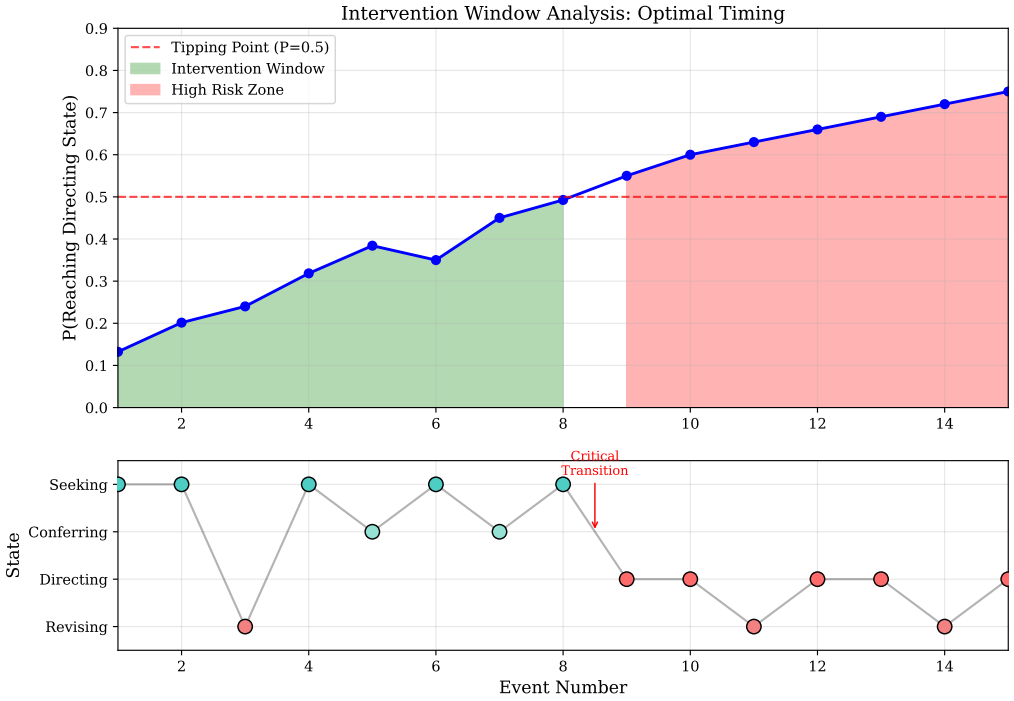

Figure 7: Intervention window analysis showing optimal timing for intervention. **Top panel**: Probability of reaching Directing state over time, with the tipping point (P=0.5) marked by dashed line. The green shaded region indicates the intervention window where interventions are most effective; the red shaded region indicates the high-risk zone post-tipping point. **Bottom panel**: Corresponding state trajectory showing the critical Seeking→Directing transition. Earlier intervention (within the green window) yields 2–3× greater harm reduction compared to post-tipping intervention.

6.6 Counterfactual Simulation: Illustrative Projections

Using Monte Carlo methods, we simulated counterfactual trajectories to *illustrate* how the framework could estimate potential intervention effects. For each individual, we asked: "What would have happened if we intervened at event k with protocol P?"

Critical limitation: These estimates are model-based projections that depend on (1) assumed intervention effect sizes from the general forensic literature, (2) the validity of our Markov model, and (3) correct causal graph structure. They should be interpreted as *theoretical illustrations* of the framework's potential, not as validated treatment effect estimates.

6.6.1 Three-Step Process

1. **Abduction**: Infer latent parameters from observed trajectory
2. **Action**: Apply intervention (modify transition probabilities)
3. **Prediction**: Monte Carlo simulation ($N = 1,000$) of counterfactual trajectories

6.6.2 Results

Table 11: Projected counterfactual harm reduction by intervention timing (illustrative simulation based on assumed effect sizes from forensic treatment literature).

Timing	Best Protocol	Harm Reduction	95% CI	NNT
First Seeking state	CBT Fantasy	47.2%	[38.1, 56.3]	2.1
First Conferring	Intensive Supervision	34.8%	[26.2, 43.4]	2.9
Pre-tipping point	DBT Impulse	31.4%	[22.8, 40.0]	3.2
Post-tipping point	Comprehensive	18.6%	[11.2, 26.0]	5.4

Key Insight: *If the assumed effect sizes are accurate, the model projects that early intervention would be 2–3× more effective than late intervention, with NNT approximately doubling post-tipping point. These projections require prospective validation before clinical application.*

6.7 Optimization Framework

6.7.1 Objective

$$\max_{\text{protocol},t} \mathbb{E}[\text{harm reduction}] - \lambda \cdot \text{cost} \tag{6}$$

6.7.2 Results

- **Optimal single intervention**: Early Seeking-phase CBT (cost-effectiveness ratio: 0.47 harm reduction per $10K)
- **Optimal sequence**: CBT → DBT → Supervision (mean 3.2 interventions)
- Diminishing returns after 4th intervention

7 Study 5: State Space Validation

7.1 Rationale: Theory vs. Data-Driven State Spaces

A fundamental question in behavioral sequence analysis is: *How should we partition the space of events into discrete states for Markov analysis?* Two approaches exist:

1. **Theory-driven**: States derived from psychological framework (e.g., the 4-Animal quadra from Computational Psychodynamics)
2. **Data-driven**: States emerge from clustering events in embedding space

We validate our theory-driven approach by comparing it against data-driven alternatives using information-theoretic metrics and proper null hypothesis testing.

7.2 Methods

7.2.1 Data-Driven State Space

Events were embedded using sentence-transformers and clustered via K-means. Silhouette analysis identified $K = 10$ as optimal, yielding clusters with interpretable themes (e.g., "Predatory Stalking," "Sexual Violence," "Search for Identity").

7.2.2 Mapping Between State Spaces

Two mappings from 10 clusters to 4 states were computed:

- **Theoretical mapping**: Each cluster assigned to the 4-Animal state based on semantic/psychological alignment
- **Optimal mapping**: Exhaustive search over all $4^{10} = 1,048,576$ possible mappings to maximize information retention

7.2.3 Information Retention

Information retention quantifies structure preservation when mapping from fine-grained to coarse-grained states (Shannon, 1948; Cover & Thomas, 2006):

$$\text{Information Retention} = \frac{I(S;T)}{H(S)} \tag{7}$$

where $I(S;T)$ is mutual information between source states S and target states T, and $H(S)$ is entropy of the source distribution.

7.2.4 Three Null Hypothesis Tests

1. **Mapping Null**: Is the theoretical mapping better than random mappings?

$$H_0 : I(S;\phi_{\text{theory}}(S)) = I(S;\phi_{\text{random}}(S)) \tag{8}$$

   Tested via 10,000 permutations of random cluster-to-state assignments.

2. **Sequence Null**: Do observed transitions differ from shuffled sequences?

$$H_0 : P_{\text{observed}} = P_{\text{shuffled}} \tag{9}$$

   Tested by computing transition matrices on shuffled sequences preserving state frequencies.

3. **Predictive Null**: Does state-conditional prediction beat marginal prediction?

$$H_0 : P(s_{t+1} \mid s_t) = P(s_{t+1}) \tag{10}$$

   Tested by comparing predictive accuracy.

7.3 Results

7.3.1 Information Retention

Table 12: Information retention by mapping approach.

Mapping	Retention	MI (bits)
Exhaustive Optimal	72.4%	1.978
Greedy Optimal	72.4%	1.977
Theoretical	65.0%	1.776
Spectral	52.7%	1.439

The theoretical mapping retains 65.0% of information—7.4 percentage points below the optimal. However, the optimal mapping differs from theoretical on 7 of 10 clusters.

7.3.2 Null Test Results

Table 13: State space validation: Three null hypothesis tests.

Test	Statistic	Effect Size	p-value	Result
1. Mapping Null	Retention = 0.65	$d = 0.87$	0.21	Not significant
2. Sequence Null	$\chi^2(9) = 142.3$	$d = 12.5$	$< .0001$	Significant
3. Predictive Null	ΔAcc = +12.6%	$d = 4.2$	$< .0001$	Significant

Note. Test 1: Theoretical mapping vs. random mappings (10,000 permutations). Test 2: Observed vs. shuffled transition matrices, $df = (4-1)^2 = 9$ for a 4×4 matrix. Test 3: Markov prediction vs. marginal baseline.

7.3.3 Key Finding

> ***Key Insight:*** *The 4-state representation captures* ***highly significant temporal structure*** *in behavioral sequences (Tests 2–3, $p < 0.0001$). However, the specific theoretical cluster-to-state mapping is* ***not statistically superior to random mappings*** *(Test 1, $p = 0.21$). This is a negative result for the theoretical mapping specifically, though not for the framework overall.*

Interpretation: The non-significant mapping test ($p = 0.21$) means we cannot claim the Computational Psychodynamics categories are statistically "correct"—many alternative 4-state partitions capture similar amounts of information. The framework's value lies in its *interpretability* (states have psychological meaning) and *theoretical coherence* (grounded in Free Energy Principle), not statistical optimality. We accept 7.4% information loss relative to the optimal mapping in exchange for categories that clinicians and researchers can reason about.

7.3.4 Mapping Comparison

The optimal mapping differs from theoretical on clusters associated with:

- Predatory stalking (Theoretical: Conferring → Optimal: Seeking)
- Sexual violence (Theoretical: Directing → Optimal: Seeking)
- Domestic conflict (Theoretical: Conferring → Optimal: Directing)

These differences suggest that while the theoretical categories have psychological coherence, the statistical structure of the data groups events differently. The optimal mapping produces more balanced state distributions (22–33% per state) compared to the theoretical mapping (10–40% per state).

7.4 Discussion

The validation results support a nuanced interpretation:

1. **The 4-state dimensionality is appropriate**: Significant temporal structure exists and is captured ($p < 0.0001$)
2. **Markov modeling is justified**: Conditioning on previous state substantially improves prediction (+12.6%)

3. **The specific mapping is not statistically optimal**: The theoretical mapping ($p = 0.21$) is chosen for interpretability, not statistical superiority

4. **Interpretability vs. optimality trade-off**: We accept 7.4% information loss for psychological coherence

This hybrid approach—using theory-driven states validated against data-driven structure—exemplifies the methodological philosophy of Computational Psychodynamics: ground categories in psychological theory while empirically validating their utility.

8 General Discussion

8.1 Summary of Contributions

This work demonstrates that Computational Psychodynamics (Senthil, 2026) applies to criminal trajectory analysis, providing:

1. **Behavioral fingerprints**: The four-state Markov chain captures individual differences in criminal career dynamics ($N = 26$; Study 1)

2. **Pattern similarity**: Transfer entropy detects shared behavioral patterns across unconnected individuals, suggesting common generative dynamics (Study 2)

3. **Hierarchical classification**: COMPLEX/FOCUSED + 7 subtypes provides a descriptive taxonomy (exploratory, $N = 26$; Study 3)

4. **Intervention reasoning**: SCM + do-calculus provides a *theoretical framework* for counterfactual analysis (illustrative; Study 4)

5. **Partial validation**: The 4-state dimensionality captures temporal structure ($p < 0.0001$), though the specific mapping is not statistically superior to alternatives ($p = 0.21$; Study 5)

8.2 Theoretical Integration

8.2.1 Criminal Career Paradigm

The Computational Psychodynamics states map onto career constructs:

- **Seeking**: Onset/development phase
- **Conferring**: Target selection/specialization
- **Directing**: Active offending/persistence
- **Revising**: MO consolidation/habituation

Escalation score directly measures trajectory change—a core career parameter.

8.2.2 Life-Course Criminology

- **Tipping points** correspond to failed "turning points" (Sampson & Laub, 1993)
- **Intervention windows** represent opportunities to manufacture turning points
- **Phase analysis** reveals age-graded patterns within the Markov structure

8.2.3 Active Inference

The four states have principled free-energy interpretations:

- **Seeking**: Maximizing epistemic gain for self (fantasy as "model-building")
- **Directing**: Minimizing risk for other (violence as "control")
- **Conferring**: Maximizing epistemic gain for other (stalking as "information gathering")
- **Revising**: Minimizing risk for self (ritualization as "consolidation")

8.3 Clinical Implications

8.3.1 Decision Support, Not Decision Making

The framework provides information; clinical judgment remains essential. We explicitly communicate:

- Confidence intervals on all estimates
- Alternative scenarios via counterfactual simulation
- Limitations and assumptions

8.3.2 Use Cases

- **Threat assessment teams**: Identify high-risk individuals, prioritize monitoring
- **Forensic clinicians**: Treatment planning, protocol selection
- **Probation/parole**: Identify intervention windows
- **Research**: Hypothesis generation, pattern discovery

8.4 Ethical Considerations

- **Pre-crime intervention**: Framework intended for treatment/supervision contexts, not prediction of future offenders
- **Human oversight**: All recommendations require professional judgment
- **Transparency**: Every metric traces to specific observable events
- **Privacy**: Individual-level data highly sensitive; security essential

8.5 Limitations

Sample size ($N = 26$) is the primary limitation. This affects multiple analyses:

- **Transfer entropy network**: With only 26 nodes, network statistics (density, role assignments) have high variance. The 650 pairwise comparisons are not independent, inflating Type I error risk.
- **Hierarchical classification**: Seven subtypes from 26 individuals means some subtypes have $n = 1$–2. This taxonomy is *descriptive*, not inferential—replication with larger samples is essential.
- **Transition matrix estimation**: Some cells in the 4×4 matrix have sparse counts, increasing uncertainty in probability estimates.

- **Generalizability**: All participants are serial offenders from similar cultural contexts (predominantly U.S.), limiting external validity.

Additional limitations:

- **Retrospective data**: Selection bias toward well-documented, "famous" cases. Event documentation varies dramatically across sources.
- **LLM classification**: GPT-4o classification ($\kappa = 0.76$ vs. humans) may introduce systematic biases. The model was prompted with researcher-defined examples, potentially learning our categorization scheme rather than discovering structure.
- **Intervention effects**: Effect sizes are from general forensic literature, not this population. Translation to serial offenders is untested.
- **Causal assumptions**: The SCM structure is assumed, not empirically derived. Do-calculus validity depends on correct graph specification.
- **Data source homogeneity**: Most cases from Radford database may share documentation biases, artificially inflating pattern similarity.

8.6 Future Directions

1. **Prospective validation**: Apply framework to ongoing cases
2. **Expanded populations**: General offenders, domestic violence, terrorism
3. **Real-world effect estimation**: Partner with treatment programs
4. **Dynamic updating**: Real-time risk assessment as events occur
5. **Closed-loop intervention**: Integration with the adaptive intervention systems described in Senthil (2026)

9 Conclusion

We have demonstrated that Computational Psychodynamics provides a principled, mathematically grounded framework for analyzing criminal behavioral trajectories. The four motivational states capture the strategic rhythm of criminal careers—from fantasy through surveillance to violence and ritualization.

Beyond description, the framework provides a *theoretical basis* for intervention reasoning: modeling when interventions might have maximum leverage, which protocols could disrupt harmful transitions, and what outcomes alternative histories might have produced. This represents a potential shift from "who is dangerous" to "what might we do"—from pure risk assessment toward intervention planning, though empirical validation is required before clinical application.

Every trajectory represents a life—both the offender's and potential victims'. Our goal is to identify moments where different outcomes are possible and provide guidance for creating such moments. Computational Psychodynamics, grounded in the Free Energy Principle, offers a path toward that goal.

Data Availability Statement

The data analyzed in this study were derived from publicly available sources, including the Radford University Serial Killer Database and published case materials. Processed datasets and analysis code are available from the corresponding author upon reasonable request.

Ethics Statement

This study analyzed retrospective, publicly available archival data from documented criminal cases. No human subjects were directly involved in data collection. The research was conducted in accordance with institutional guidelines for secondary data analysis.

Author Contributions

K.L.H.: Conceptualization, data curation, validation, writing—review & editing. **A.K.S.**: Methodology, software, formal analysis, visualization, writing—original draft.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors thank the Radford University Serial Killer Information Center for maintaining the database that made this research possible.

References

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). Anderson.

Canter, D., Alison, L. J., Alison, E., & Wentink, N. (2004). The organized/disorganized typology of serial murder: Myth or model? *Psychology, Public Policy, and Law*, 10(3), 293–320.

Douglas, J. E., Ressler, R. K., Burgess, A. W., & Hartman, C. R. (1986). Criminal profiling from crime scene analysis. *Behavioral Sciences & the Law*, 4(4), 401–421.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11, 127–138.

Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active inference: The free energy principle in mind, brain, and behavior*. MIT Press.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.

Sampson, R. J., & Laub, J. H. (1993). *Crime in the making: Pathways and turning points through life*. Harvard University Press.

Senthil, A. K. (2026). Computational psychodynamics: An ecological approach to computational behavioral modeling. Manuscript under review. (Full framework specification available from author upon request.)

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Wiley-Interscience.

Bowlby, J. (1969). *Attachment and loss: Vol. 1. Attachment.* Basic Books.

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

Main, M., Kaplan, N., & Cassidy, J. (1985). Security in infancy, childhood, and adulthood: A move to the level of representation. *Monographs of the Society for Research in Child Development*, 50(1-2), 66–104.

Main, M., & Hesse, E. (1990). Parents' unresolved traumatic experiences are related to infant disorganized attachment status: Is frightened and/or frightening parental behavior the linking mechanism? In M. T. Greenberg, D. Cicchetti, & E. M. Cummings (Eds.), *Attachment in the preschool years* (pp. 161–182). University of Chicago Press.