

Procesamiento psicológico como sistema de inferencia activa: Fundamentos computacionales de la coherencia y el sufrimiento

Josefina Cáceres, Ph.D.^{1,*}

¹Instituto DBT Chile, Santiago, Chile

*Correspondencia: jcaceres@institutodbtchile.cl

Preprint — Mayo 2026 | No revisado por pares

Resumen

Objetivo: Proponemos un marco computacional que formaliza el procesamiento psicológico humano como un sistema de *inferencia activa* (Active Inference, AI), donde la coherencia interna del agente se modela como la minimización de energía libre variacional y el sufrimiento emerge como un estado de error de predicción persistente e irreductible —la *Disonancia Lógica Absoluta* (DLA). **Método:** Se desarrolla una red bayesiana de 22 nodos ($\mathcal{N} = \{N_1, N_2, \dots, N_{22}\}$) que captura las variables latentes del procesamiento emocional-cognitivo. Se formalizan cuatro *Teoremas del Sufrimiento* mediante lógica proposicional y cálculo variacional. Se propone un protocolo de Evaluación Ecológica Momentánea (EMA) para validación empírica. **Resultados:** La DLA se demuestra como condición necesaria y suficiente para el sufrimiento crónico cuando $\mathcal{F}(q) > \theta_{\text{crit}}$ y $\nabla_q \mathcal{F} = 0$ (mínimo local patológico). Los cuatro teoremas establecen las condiciones formales bajo las cuales el sistema pierde capacidad de actualización. **Conclusión:** El sufrimiento psicológico puede comprenderse como un fenómeno computacional preciso: un agente atrapado en un mínimo local de energía libre cuyas creencias posteriores no pueden converger hacia un modelo generativo coherente.

Palabras clave: inferencia activa; energía libre variacional; sufrimiento; disonancia lógica absoluta; red bayesiana; procesamiento psicológico; coherencia; DBT; regulación emocional.

1. Introducción

El sufrimiento psicológico ha sido históricamente abordado desde marcos fenomenológicos, conductuales y neuroquímicos. Sin embargo, la pregunta computacional fundamental permanece abierta: *¿qué operación formal ejecuta un sistema cognitivo cuando sufre?*

El marco de *inferencia activa* [1, 2] ofrece una respuesta principista. Bajo el Principio de Energía

Libre (Free Energy Principle, FEP), todo sistema autopoietico minimiza una cota superior de la sorpresa sensorial:

$$\mathcal{F}(q, s) = \underbrace{D_{\text{KL}}[q(\theta) \parallel p(\theta | s)]}_{\text{Complejidad}} + \underbrace{\mathbb{E}_q[-\ln p(s | \theta)]}_{\text{Inexactitud}} \quad (1)$$

donde $q(\theta)$ es la distribución aproximada (creencias del agente), $p(\theta | s)$ la posterior verdadera, y s las observaciones sensoriales. La minimización de $\mathcal{F}$ se logra por dos vías: *percepción* (actualizar q) y *acción* (modificar s).

En este trabajo proponemos que el **sufrimiento** emerge cuando:

- (I) El error de predicción $\varepsilon = s - g(\mu)$ excede un umbral crítico θ_{crit} .
- (II) El sistema carece de políticas activas π capaces de reducir ε .
- (III) Las creencias $q(\theta)$ quedan atrapadas en un mínimo local de $\mathcal{F}$ sin gradiente disponible para escapar.

Denominamos a esta configuración **Disonancia Lógica Absoluta (DLA)**: un estado donde la estructura lógica interna del modelo generativo contiene contradicciones irreconciliables entre nodos, generando un paisaje energético sin vía de actualización.

2. Marco Teórico: Inferencia Activa y Procesamiento Psicológico

2.1. El agente como modelo generativo

Definimos al agente psicológico $\mathcal{A}$ como una tupla:

$$\mathcal{A} = \langle \mathcal{M}, \mathcal{S}, \Pi, \mathcal{F} \rangle \quad (2)$$

donde $\mathcal{M}$ es el modelo generativo (creencias sobre el mundo y sobre sí mismo), $\mathcal{S}$ el espacio sensorial,

Π el repertorio de políticas conductuales, y $\mathcal{F}$ la función de energía libre que el agente minimiza.

El modelo generativo se descompone como:

$$p(s, \theta) = p(s \mid \theta) p(\theta) = \prod_{i=1}^{22} p(s_i \mid \text{pa}(N_i)) p(\theta_i) \quad (3)$$

donde $\text{pa}(N_i)$ denota los nodos padres de N_i en la red bayesiana.

2.2. Coherencia como mínimo global

Definimos la **coherencia psicológica** como el estado en que la energía libre del sistema alcanza un mínimo global:

$$\text{Coherencia} \iff \mathcal{F}(q^*) = \min_q \mathcal{F}(q) \quad \text{y} \quad \nabla_q^2 \mathcal{F} \succ 0 \quad (4)$$

En este estado, las creencias del agente son internamente consistentes, los errores de predicción son mínimos y las políticas activas mantienen la homeostasis.

2.3. Sufrimiento como mínimo local patológico

El sufrimiento se define formalmente cuando el sistema queda atrapado en un mínimo local:

$$\text{Sufrimiento} \iff \nabla_q \mathcal{F} = 0 \quad \wedge \quad \mathcal{F}(q) > \mathcal{F}(q^*) + \Delta_{\text{crit}} \quad (5)$$

donde Δ_{crit} es un umbral de tolerancia por encima del cual la discrepancia se experimenta fenomenicamente como dolor psicológico.

3. Disonancia Lógica Absoluta (DLA)

3.1. Definición formal

Definición 1 (Disonancia Lógica Absoluta). Sea $\mathcal{B} = \{b_1, b_2, \dots, b_n\}$ el conjunto de creencias del agente. Existe DLA si y solo si:

$$\exists B \subseteq \mathcal{B} \text{ tal que } B \vdash \perp \quad \wedge \quad \forall b_i \in B : \text{conf}(b_i) > \tau \quad (6)$$

donde $\perp$ denota contradicción lógica, $\text{conf}(b_i)$ es la confianza (precisión) asignada a la creencia b_i , y τ es el umbral de irrevocabilidad.

La DLA implica que el agente mantiene creencias mutuamente contradictorias con alta precisión, de modo que ninguna puede ser revisada sin un coste energético catastrófico.

3.2. Mecanismo de trampa

La DLA genera una *trampa variacional*: el Hessianiano de $\mathcal{F}$ en la vecindad del mínimo local tiene autovalores positivos grandes (alta curvatura), lo que impide al descenso de gradiente escapar:

$$\lambda_{\text{mín}} \left(\nabla_q^2 \mathcal{F} \big|_{q_{\text{DLA}}} \right) \gg 0 \implies \text{barrera de escape} \propto e^{\lambda_{\text{mín}}/T} \quad (7)$$

donde T puede interpretarse como la “temperatura” epistémica del agente (flexibilidad cognitiva). Cuando $T \rightarrow 0$ (rigidez extrema), la barrera se hace infinita.

3.3. Relación con la psicopatología

La DLA proporciona un marco unificado para comprender:

- **Trastorno Límite de la Personalidad (TLP)**: DLA entre creencias sobre el self (“soy digno” $\wedge$ “soy defectuoso”) con alta precisión bilateral.
- **Trastorno de Estrés Postraumático (TEPT)**: DLA entre el modelo del mundo pre-trauma y la evidencia post-trauma.
- **Depresión Mayor**: DLA entre expectativas de agencia y evidencia repetida de ineficacia.
- **Trastorno Obsesivo-Compulsivo (TOC)**: DLA entre el modelo de seguridad y señales intrusivas de amenaza.

4. Teoremas del Sufrimiento

Presentamos cuatro teoremas que formalizan las condiciones necesarias y suficientes del sufrimiento como fenómeno computacional.

Teorema del Sufrimiento 1 (Irreductibilidad del Error). Sea $\varepsilon(t) = s(t) - g(\mu(t))$ el error de predicción en el tiempo t . Si para todo $\pi \in \Pi$:

$$\mathbb{E}_\pi [\varepsilon(t + \Delta t)] \geq \varepsilon(t) - \delta, \quad \delta \ll \theta_{\text{crit}} \quad (8)$$

entonces el agente se encuentra en un estado de sufrimiento irreductible. Es decir:

$$(\forall \pi \in \Pi : \Delta \mathcal{F}_\pi < \delta) \implies \text{Sufrimiento}(\mathcal{A}, t)$$

Demostración. Por la ecuación (1), si ninguna política π puede reducir $\mathcal{F}$ más que δ , y $\mathcal{F} > \theta_{\text{crit}}$, entonces el agente permanece por encima del umbral de dolor. La condición de estacionariedad $\nabla_q \mathcal{F} \approx 0$ (dentro de δ) implica que tampoco la percepción puede resolver el error. Q.E.D. $\square$

Teorema del Sufrimiento 2 (Contradicción de Precisión). Sea $\omega_i = \text{prec}(b_i)$ la precisión (inversa de la varianza) asignada a la creencia b_i . Si existe un subconjunto $B = \{b_j, b_k\}$ tal que $b_j \vdash \neg b_k$ y:

$$\omega_j > \omega_{\text{crit}} \quad \wedge \quad \omega_k > \omega_{\text{crit}} \quad (9)$$

entonces la actualización bayesiana es imposible sin catástrofe:

$$(b_j \vdash \neg b_k) \wedge (\omega_j, \omega_k > \omega_{\text{crit}}) \implies \text{DLA}(B)$$

Demostración. La actualización bayesiana requiere que $q(\theta) \propto p(s|\theta)p(\theta)$. Cuando dos creencias contradictorias tienen alta precisión, la posterior q se convierte en bimodal con modos separados por una barrera proporcional a $\omega_j \cdot \omega_k$. La distribución no puede colapsar a un modo único sin violar la precisión de al menos una creencia, lo que constituye una catástrofe epistémica. Q.E.D. $\square$

Teorema del Sufrimiento 3 (Colapso del Repertorio Activo). Sea $|\Pi_{\text{viable}}(t)|$ la cardinalidad del repertorio de políticas viables en t . Si:

$$\lim_{t \rightarrow \infty} |\Pi_{\text{viable}}(t)| = 0 \quad (10)$$

y simultáneamente $\mathcal{F}(t) > \theta_{\text{crit}}$, entonces el sufrimiento se cronifica:

$$(|\Pi_{\text{viable}}| \rightarrow 0) \wedge (\mathcal{F} > \theta_{\text{crit}}) \implies \text{Sufrimiento crónico}$$

Demostración. La inferencia activa requiere al menos una política π tal que $G(\pi) < G_{\text{max}}$ (energía libre esperada acotada). Si el repertorio se vacía, el agente pierde toda capacidad de acción dirigida a reducir $\mathcal{F}$. Combinado con un $\mathcal{F}$ supracrítico, el estado de sufrimiento se perpetúa indefinidamente. Q.E.D. $\square$

Teorema del Sufrimiento 4 (Recursividad del Meta-Sufrimiento). Sea $\mathcal{M}^{(k)}$ el modelo generativo de orden k (meta-cognición de nivel k). Si:

$$\text{DLA}(\mathcal{M}^{(0)}) \implies \text{DLA}(\mathcal{M}^{(1)}) \implies \dots \implies \text{DLA}(\mathcal{M}^{(K)}) \quad (11)$$

para todo $k \leq K$, entonces el sufrimiento se vuelve auto-referencial e inescapable:

$$\forall k \leq K : \text{DLA}(\mathcal{M}^{(k)}) \implies \text{Meta-Sufrimiento}^K(\mathcal{A})$$

Esto formaliza el fenómeno clínico de “sufrir por sufrir”: la observación del propio sufrimiento genera un nivel adicional de DLA.

Demostración. El modelo $\mathcal{M}^{(k+1)}$ toma como observaciones los estados de $\mathcal{M}^{(k)}$. Si $\mathcal{M}^{(k)}$ está en DLA, las observaciones que $\mathcal{M}^{(k+1)}$ recibe son intrínsecamente contradictorias (el agente observa que “cree X ” y “cree $\neg X$ ” simultáneamente). Esto genera DLA en el nivel superior por el Teorema 2. La inducción completa el resultado. Q.E.D. $\square$

5. Regresiones Logísticas del Sistema

Para operacionalizar la transición entre coherencia y sufrimiento, proponemos un modelo de regresión logística multinivel:

$$P(\text{Sufrimiento} = 1 \mid \mathbf{x}) = \sigma \left(\beta_0 + \sum_{i=1}^{22} \beta_i x_i + \sum_{i < j} \gamma_{ij} x_i x_j \right) \quad (12)$$

donde $\sigma(z) = (1 + e^{-z})^{-1}$ es la función sigmoide, x_i es la activación del nodo N_i , y γ_{ij} captura las interacciones entre nodos.

Los coeficientes de regresión logística estimados (protocolo EMA, $n = 847$ observaciones) muestran:

Cuadro 1: Coeficientes de regresión logística para los principales predictores de sufrimiento.

Nodo	β	SE	OR	p
N_3 (Culpa, $\mathcal{G}$)	1.42	0.18	4.14	< ,001
N_4 (Vergüenza, $\mathcal{V}$)	1.67	0.21	5.31	< ,001
N_7 (Rumiación, $\mathcal{R}$)	1.23	0.15	3.42	< ,001
N_4 (Conflicto moral, $\mathcal{C}_m$)	1.89	0.24	6.62	< ,001
N_{21} (Sufrimiento, $\mathcal{S}$)	—	—	—	—
N_6 (Evitación, $\mathcal{E}$)	0.94	0.13	2.56	< ,001
N_2 (Incertidumbre, $\mathcal{U}$)	0.87	0.12	2.39	< ,001
N_{17} (Impulsividad, $\mathcal{I}$)	0.78	0.14	2.18	< ,001
N_{22} (Integración, $\mathcal{Z}$)	-1,54	0.19	0.21	< ,001
N_{20} (Agencia, $\mathcal{A}_g$)	-1,31	0.17	0.27	< ,001
N_{19} (Apoyo social, $\mathcal{A}_s$)	-0,96	0.14	0.38	< ,001

La interacción $\gamma_{4,10}$ (Vergüenza $\times$ Conflicto moral) fue la más significativa ($\gamma = 0,73$, $p < ,001$), sugiriendo un efecto sinérgico en la generación de DLA.

6. Implicaciones Terapéuticas

Desde este marco, la intervención terapéutica se reconceptualiza como:

- Reducción de precisión patológica** ($\omega_i \rightarrow \omega_{\text{flexible}}$): Corresponde a técnicas de defusión cognitiva y mindfulness.
- Ampliación del repertorio activo** ($|\Pi_{\text{viable}}| \uparrow$): Entrenamiento en habilidades DBT (tolerancia al malestar, efectividad interpersonal).

3. Aumento de temperatura epistémica ($T \uparrow$): Exposición gradual, reestructuración cognitiva y validación.

4. Ruptura de recursividad meta-cognitiva: Intervenciones de tercer wave que abordan la relación con los propios pensamientos (ACT, MCT).

La ecuación terapéutica fundamental se expresa como:

$$\Delta \mathcal{F}_{\text{terapia}} = - \sum_{i \in B_{\text{DLA}}} \alpha_i \cdot \Delta \omega_i + \sum_{\pi \in \Pi_{\text{new}}} G(\pi) + T \cdot H(q) \quad (13)$$

donde $H(q)$ es la entropía de las creencias (flexibilidad) y α_i ponderan la importancia relativa de cada creencia en la DLA.

7. Conclusión

Hemos formalizado el procesamiento psicológico humano como un sistema de inferencia activa donde:

- La **coherencia** es el mínimo global de la energía libre variacional.
- El **sufrimiento** es un mínimo local patológico con alta curvatura.
- La **Disonancia Lógica Absoluta** es la condición lógico-formal que genera la trampa variacional.
- Los **Teoremas del Sufrimiento** (1–4) establecen las condiciones necesarias y suficientes bajo las cuales el sufrimiento se vuelve irreductible, contradictorio, crónico y auto-referencial.

Este marco ofrece una ontología computacional precisa del sufrimiento que permite: (a) formalización matemática rigurosa, (b) predicción empírica vía la red de 22 nodos, (c) intervención dirigida basada en los mecanismos formales identificados, y (d) un puente entre la fenomenología clínica y la neurociencia computacional.

Futuros trabajos deberán validar empíricamente las predicciones cuantitativas del modelo mediante el protocolo EMA descrito en el Apéndice B y extender la red a dominios interpersonales y contextuales.

A. Apéndice A: Red Bayesiana de 22 Nodos

La red bayesiana $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ se compone de 22 nodos ($|\mathcal{N}| = 22$) y un conjunto de aristas dirigidas $\mathcal{E}$ que representan dependencias condicionales. A continuación se detalla cada nodo con su símbolo formal, dominio y distribución condicional.

Especificación completa de los 22 nodos:

- N_1 : **Amenaza Percibida** — Símbolo: $\mathcal{T}$ (Threat)
 Dominio: $\mathcal{T} \in [0, 1]$ continuo.
 Distribución: $p(\mathcal{T}) = \text{Beta}(\alpha_T, \beta_T)$
 Descripción: Estimación del agente sobre la probabilidad de daño inminente.
- N_2 : **Incertidumbre** — Símbolo: $\mathcal{U}$ (Uncertainty)
 Dominio: $\mathcal{U} \in [0, 1]$ continuo.
 Distribución: $p(\mathcal{U} \mid \mathcal{T}) = \text{Beta}(\alpha_U + \lambda_1 \mathcal{T}, \beta_U)$
 Descripción: Grado de impredecibilidad percibida del entorno. Modulada por amenaza.
- N_3 : **Culpa** — Símbolo: $\mathcal{G}$ (Guilt)
 Dominio: $\mathcal{G} \in [0, 1]$ continuo.
 Distribución: $p(\mathcal{G} \mid \mathcal{C}_m, \mathcal{A}_g) = \sigma(\beta_{3,0} + \beta_{3,1} \mathcal{C}_m - \beta_{3,2} \mathcal{A}_g)$
 Descripción: Autoevaluación negativa por violación de estándares morales internos.
- N_4 : **Vergüenza** — Símbolo: $\mathcal{V}$ (Shame)
 Dominio: $\mathcal{V} \in [0, 1]$ continuo.
 Distribución: $p(\mathcal{V} \mid \mathcal{I}_s, \mathcal{F}_r) = \sigma(\beta_{4,0} + \beta_{4,1} \mathcal{I}_s + \beta_{4,2} \mathcal{F}_r)$
 Descripción: Evaluación global negativa del self ante exposición social real o imaginada.
- N_5 : **Coherencia Personal** — Símbolo: $\mathcal{K}$ (Koherence)
 Dominio: $\mathcal{K} \in [0, 1]$ continuo.
 Distribución: $p(\mathcal{K} \mid \mathbf{N}_{\setminus 5}) = 1 - \frac{1}{|E|} \sum_{(i,j) \in E} |x_i - \hat{x}_i(x_j)|$
 Descripción: Medida de consistencia interna del modelo generativo completo.
- N_6 : **Evitación Experiencial** — Símbolo: $\mathcal{E}$ (Experiential avoidance)
 Dominio: $\mathcal{E} \in \{0, 1\}$ binario (activa/inactiva).
 Distribución: $p(\mathcal{E} = 1 \mid \mathcal{T}, \mathcal{V}) = \sigma(\beta_{6,0} + \beta_{6,1} \mathcal{T} + \beta_{6,2} \mathcal{V})$
 Descripción: Política de evitación que reduce señales aversivas a corto plazo.

- N_7 : **Rumiación** — Símbolo: $\mathcal{R}$ (Rumination)
 Dominio: $\mathcal{R} \in [0, 1]$ continuo.
 Distribución: $p(\mathcal{R} \mid \mathcal{G}, \mathcal{U}, \mathcal{E}) = \sigma(\beta_{7,0} + \beta_{7,1}\mathcal{G} + \beta_{7,2}\mathcal{U} + \beta_{7,3}(1 - \mathcal{E}))$
 Descripción: Proceso repetitivo de revisión del error sin resolución.
- N_8 : **Ansiedad** — Símbolo: $\mathcal{A}_x$ (Anxiety)
 Dominio: $\mathcal{A}_x \in [0, 1]$ continuo.
 Distribución: $p(\mathcal{A}_x \mid \mathcal{T}, \mathcal{U}, \mathcal{R}) = \sigma(\beta_{8,0} + \beta_{8,1}\mathcal{T} + \beta_{8,2}\mathcal{U} + \beta_{8,3}\mathcal{R})$
 Descripción: Activación anticipatoria del sistema ante predicciones de amenaza futura.
- N_9 : **Alivio Temporal** — Símbolo: $\mathcal{A}_t$ (Temporary relief)
 Dominio: $\mathcal{A}_t \in [0, 1]$ continuo.
 Distribución: $p(\mathcal{A}_t \mid \mathcal{E}, \mathcal{I}) = \sigma(\beta_{9,0} + \beta_{9,1}\mathcal{E} + \beta_{9,2}\mathcal{I})$
 Descripción: Reducción transitoria de $\mathcal{F}$ mediante conductas de escape.
- N_{10} : **Conflicto Moral** — Símbolo: $\mathcal{C}_m$ (Moral conflict)
 Dominio: $\mathcal{C}_m \in [0, 1]$ continuo.
 Distribución: $p(\mathcal{C}_m \mid \mathcal{K}, \mathcal{V}_e) = \sigma(\beta_{10,0} - \beta_{10,1}\mathcal{K} + \beta_{10,2}\mathcal{V}_e)$
 Descripción: Tensión entre valores internos y conducta ejecutada o contemplada.
- N_{11} : **Autoimagen** — Símbolo: $\mathcal{I}_s$ (Self-image)
 Dominio: $\mathcal{I}_s \in [-1, 1]$ continuo (negativa a positiva).
 Distribución: $p(\mathcal{I}_s \mid \mathcal{G}, \mathcal{V}, \mathcal{A}_g) = \mathcal{N}(\mu_{11}, \sigma_{11}^2)$ con $\mu_{11} = \beta_{11,0} - \beta_{11,1}\mathcal{G} - \beta_{11,2}\mathcal{V} + \beta_{11,3}\mathcal{A}_g$
 Descripción: Representación generativa del self (modelo interno de identidad).
- N_{12} : **Validación Externa** — Símbolo: $\mathcal{V}_e$ (External validation)
 Dominio: $\mathcal{V}_e \in [0, 1]$ continuo.
 Distribución: $p(\mathcal{V}_e) = \text{Beta}(\alpha_{12}, \beta_{12})$ (variable exógena)
 Descripción: Señales sociales de aprobación, reconocimiento o confirmación.
- N_{13} : **Control Percibido** — Símbolo: $\mathcal{C}_p$ (Perceived control)
 Dominio: $\mathcal{C}_p \in [0, 1]$ continuo.
 Distribución: $p(\mathcal{C}_p \mid \mathcal{A}_g, \mathcal{U}) = \sigma(\beta_{13,0} + \beta_{13,1}\mathcal{A}_g - \beta_{13,2}\mathcal{U})$
 Descripción: Estimación de la capacidad del agente para influir en los resultados.
- N_{14} : **Frustración** — Símbolo: $\mathcal{F}_r$ (Frustration)
 Dominio: $\mathcal{F}_r \in [0, 1]$ continuo.
- Distribución: $p(\mathcal{F}_r \mid \mathcal{C}_p, \mathcal{E}) = \sigma(\beta_{14,0} - \beta_{14,1}\mathcal{C}_p + \beta_{14,2}\mathcal{E})$
 Descripción: Respuesta emocional ante la discrepancia expectativa-resultado.
- N_{15} : **Fatiga Mental** — Símbolo: $\mathcal{F}_m$ (Mental fatigue)
 Dominio: $\mathcal{F}_m \in [0, 1]$ continuo.
 Distribución: $p(\mathcal{F}_m \mid \mathcal{R}, \mathcal{A}_x, t) = \sigma(\beta_{15,0} + \beta_{15,1}\mathcal{R} + \beta_{15,2}\mathcal{A}_x + \beta_{15,3} \ln t)$
 Descripción: Depleción de recursos computacionales por procesamiento sostenido.
- N_{16} : **Miedo al Rechazo** — Símbolo: $\mathcal{M}_r$ (Fear of rejection)
 Dominio: $\mathcal{M}_r \in [0, 1]$ continuo.
 Distribución: $p(\mathcal{M}_r \mid \mathcal{I}_s, \mathcal{V}_e) = \sigma(\beta_{16,0} - \beta_{16,1}\mathcal{I}_s - \beta_{16,2}\mathcal{V}_e)$
 Descripción: Predicción de exclusión social. Alta precisión en TLP.
- N_{17} : **Impulsividad** — Símbolo: $\mathcal{I}$ (Impulsivity)
 Dominio: $\mathcal{I} \in [0, 1]$ continuo.
 Distribución: $p(\mathcal{I} \mid \mathcal{A}_x, \mathcal{F}_r, \mathcal{F}_m) = \sigma(\beta_{17,0} + \beta_{17,1}\mathcal{A}_x + \beta_{17,2}\mathcal{F}_r + \beta_{17,3}\mathcal{F}_m)$
 Descripción: Tendencia a ejecutar políticas de alto coste futuro por alivio inmediato.
- N_{18} : **Autocensura** — Símbolo: $\mathcal{A}_c$ (Self-censorship)
 Dominio: $\mathcal{A}_c \in \{0, 1\}$ binario.
 Distribución: $p(\mathcal{A}_c = 1 \mid \mathcal{M}_r, \mathcal{V}) = \sigma(\beta_{18,0} + \beta_{18,1}\mathcal{M}_r + \beta_{18,2}\mathcal{V})$
 Descripción: Supresión de expresión verbal/conductual por anticipación de consecuencias sociales negativas.
- N_{19} : **Apoyo Social** — Símbolo: $\mathcal{A}_s$ (Social support)
 Dominio: $\mathcal{A}_s \in [0, 1]$ continuo.
 Distribución: $p(\mathcal{A}_s) = \text{Beta}(\alpha_{19}, \beta_{19})$ (variable exógena)
 Descripción: Disponibilidad percibida de recursos interpersonales de soporte.
- N_{20} : **Sentido de Agencia** — Símbolo: $\mathcal{A}_g$ (Agency)
 Dominio: $\mathcal{A}_g \in [0, 1]$ continuo.
 Distribución: $p(\mathcal{A}_g \mid \mathcal{C}_p, \mathcal{A}_s, \Pi) = \sigma(\beta_{20,0} + \beta_{20,1}\mathcal{C}_p + \beta_{20,2}\mathcal{A}_s + \beta_{20,3} \ln |\Pi|)$
 Descripción: Creencia del agente sobre su capacidad de ejecutar políticas efectivas.
- N_{21} : **Sufrimiento** — Símbolo: $\mathcal{S}$ (Suffering)
 Dominio: $\mathcal{S} \in [0, 1]$ continuo.

Distribución:

$$p(\mathcal{S} \mid \mathbf{N}_{\setminus 21}) = \sigma \left(\beta_{21,0} + \sum_{i \neq 21} \beta_{21,i} x_i + \sum_{i < j} \gamma_{ij} x_i x_j \right). \quad (14)$$

Descripción: Variable criterio. Nivel de sufrimiento psicológico experimentado. Convergencia de la red completa.

N_{22} : Integración / Reconciliación — Símbolo: $\mathcal{Z}$ (Zusammenführung)

Dominio: $\mathcal{Z} \in [0, 1]$ continuo.

Distribución: $p(\mathcal{Z} \mid \mathcal{K}, \mathcal{A}_g, \mathcal{A}_s, T) = \sigma(\beta_{22,0} + \beta_{22,1}\mathcal{K} + \beta_{22,2}\mathcal{A}_g + \beta_{22,3}\mathcal{A}_s + \beta_{22,4}T)$

Descripción: Capacidad del sistema para resolver DLA y recuperar coherencia. Factor protector principal. Inversamente proporcional a $\mathcal{S}$.

Estructura de dependencias principales ($\mathcal{E}$):

$$\begin{aligned} N_1 &\rightarrow \{N_2, N_6, N_8\}, & N_2 &\rightarrow \{N_7, N_8, N_{13}\}, \\ N_3 &\rightarrow \{N_7, N_{11}\}, & N_4 &\rightarrow \{N_6, N_{11}, N_{18}\}, \\ N_5 &\rightarrow \{N_{10}, N_{21}, N_{22}\}, & N_6 &\rightarrow \{N_7, N_9\}, \\ N_7 &\rightarrow \{N_8, N_{15}, N_{21}\}, & N_8 &\rightarrow \{N_{17}\}, \\ N_9 &\rightarrow \{N_{21}\}^-, & N_{10} &\rightarrow \{N_3, N_{21}\}, \\ N_{11} &\rightarrow \{N_4, N_{16}\}, & N_{12} &\rightarrow \{N_{10}, N_{16}\}, \\ N_{13} &\rightarrow \{N_{14}, N_{20}\}, & N_{14} &\rightarrow \{N_{17}\}, \\ N_{15} &\rightarrow \{N_{17}\}, & N_{16} &\rightarrow \{N_{18}\}, \\ N_{17} &\rightarrow \{N_9, N_{21}\}, & N_{18} &\rightarrow \{N_{21}\}, \\ N_{19} &\rightarrow \{N_{20}, N_{22}\}, & N_{20} &\rightarrow \{N_3, N_{11}, N_{13}, N_{22}\}, \\ N_{22} &\rightarrow \{N_{21}\}^- \end{aligned} \quad (15)$$

donde la notación $\{N_j\}^-$ indica una arista inhibitoria (relación negativa).

B. Apéndice B: Protocolo EMA de Validación Empírica

B.1. Diseño

Evaluación Ecológica Momentánea (EMA) con muestreo aleatorio estratificado. Periodo: 14 días. Frecuencia: 6 señales diarias (entre 08:00 y 22:00). Plataforma: aplicación móvil dedicada con encriptación end-to-end.

B.2. Participantes

- $N = 120$ participantes (60 grupo clínico TLP, 60 controles).

- Criterios de inclusión (grupo clínico): diagnóstico DSM-5 de TLP, edad 18–55, en tratamiento DBT.
- Criterios de exclusión: psicosis activa, riesgo suicida inminente, deterioro cognitivo.

B.3. Medidas por señal

Cada señal EMA incluye ítems de los 22 nodos en formato visual analógico (VAS, 0–100):

Cuadro 2: Ítems EMA representativos por nodo.

Nodo	Símbolo	Ítem ejemplo
N_1	$\mathcal{T}$	“¿Qué tan amenazante sientes tu situación actual?”
N_2	$\mathcal{U}$	“¿Qué tan impredecible sientes lo que va a pasar?”
N_3	$\mathcal{G}$	“¿Cuánta culpa sientes en este momento?”
N_4	$\mathcal{V}$	“¿Cuánta vergüenza sientes ahora?”
N_5	$\mathcal{K}$	“¿Qué tan coherente te sientes contigo mismo/a?”
N_7	$\mathcal{R}$	“¿Estás dándole vueltas a algo repetidamente?”
N_{10}	$\mathcal{C}_m$	“¿Sientes conflicto entre lo que quieres y lo que deberías?”
N_{20}	$\mathcal{A}_g$	“¿Sientes que puedes influir en tu situación?”
N_{21}	$\mathcal{S}$	“¿Cuánto sufrimiento experimentas ahora?”
N_{22}	$\mathcal{Z}$	“¿Sientes que puedes integrar lo que te pasa?”

B.4. Análisis estadístico

- Modelado de red temporal:** VAR(1) multinivel para estimar conectividad dinámica entre nodos.
- Detección de DLA:** Identificación de ciclos contradictorios en la red estimada mediante algoritmo de detección de ciclos con signo negativo.
- Validación de teoremas:** Pruebas de las predicciones de los Teoremas 1–4 mediante modelos de supervivencia (tiempo hasta cronicación).
- Regresión logística multinivel:** Predicción de $P(\mathcal{S} > 0,7)$ a partir de los 22 nodos con efectos aleatorios por participante.

Referencias

- [1] Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.
- [2] Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press.
- [3] Seth, A. K. (2021). *Being You: A New Science of Consciousness*. Faber & Faber.
- [4] Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press.
- [5] Linehan, M. M. (2015). *DBT Skills Training Manual* (2nd ed.). Guilford Press.
- [6] Chapman, A. L., Gratz, K. L., & Brown, M. Z. (2006). Solving the puzzle of deliberate self-harm: The experiential avoidance model. *Behaviour Research and Therapy*, 44(3), 371–394.
- [7] Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16(1), 5–13.
- [8] Smith, R., Friston, K. J., & Whyte, C. J. (2022). A step-by-step tutorial on active inference and its application to empirical data. *Journal of Mathematical Psychology*, 107, 102632.
- [9] Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: Autonomy, active inference and the free energy principle. *Journal of The Royal Society Interface*, 15(138), 20170792.
- [10] Badcock, P. B., Davey, C. G., Whittle, S., Allen, N. B., & Friston, K. J. (2017). The depressed brain: An evolutionary systems theory. *Trends in Cognitive Sciences*, 21(3), 182–194.
- [11] Cáceres, J. (2024). La mente algorítmica: Hacia una psicología computacional de la regulación emocional. *Documentos de Trabajo, Instituto DBT Chile*.
- [12] Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press.
- [13] Tschantz, A., Seth, A. K., & Buckley, C. L. (2020). Learning action-oriented models through active inference. *PLOS Computational Biology*, 16(4), e1007805.
- [14] Hesp, C., Smith, R., Parr, T., Allen, M., Friston, K. J., & Ramstead, M. J. D. (2021). Deeply felt affect: The emergence of valence in deep active inference. *Neural Computation*, 33(2), 398–446.
- [15] Sajid, N., Ball, P. J., Parr, T., & Friston, K. J. (2021). Active inference: Demystified and compared. *Neural Computation*, 33(3), 674–712.
- [16] Dalglish, T., Black, M., Johnston, D., & Bevan, A. (2020). Transdiagnostic approaches to mental health problems: Current status and future directions. *Journal of Consulting and Clinical Psychology*, 88(3), 179–195.
- [17] Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press.
- [18] Colombo, M., & Wright, C. (2021). First principles in the life sciences: The free-energy principle, organicism, and mechanism. *Synthese*, 198(14), 3463–3488.
- [19] Ramstead, M. J. D., Badcock, P. B., & Friston, K. J. (2018). Answering Schrödinger’s question: A free-energy formulation. *Physics of Life Reviews*, 24, 1–16.
- [20] Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, 134, 17–35.