Affective predictive graph theory: a physiologically modulated and self-observing model of cognitive dynamics

Rafi Seddiqi[1]

[1]Independent Theoretical Manuscript

Abstract

Predictive processing has become the dominant computational framework for understanding the mind, yet three central features of ordinary mental life remain underspecified: the typed affective structure of the relations the brain predicts over, the role of moment-to-moment bodily state in shaping which predictions become consequential, and the mechanism by which metacognitive awareness alters affective dynamics rather than merely reporting on them. We introduce *affective predictive graph theory* (APGT), a formal dynamical model in which mental states emerge from sigmoidal recurrence over a directed weighted graph whose edges encode affective-predictive expectations, whose effective coupling is parametrically modulated by latent physiology, and which closes a feedback loop with a partial self-observation operator. We prove four results: (i) the dynamics admit a unique stable fixed point in a contraction regime and (ii) undergo a saddle-node bifurcation as physiological stress crosses a critical value, formalising the clinical observation that the same external cue produces qualitatively different mental states under different bodily conditions; (iii) the meta-awareness operator provably reduces the largest eigenvalue of the symmetric part of the effective coupling matrix—and empirically also its spectral radius—providing a control-theoretic account of why noticing a reaction can change it; and (iv) the model is identifiable from a single sufficiently rich activation trajectory. Computational experiments in a canonical six-node graph and in an ensemble of synthetic 120-node small-world graphs ($n = 60$ seeds) confirm each theoretical prediction, with meta-awareness reducing late-window threat activation by 0.081 (95% CI [0.027, 0.144], Wilcoxon $p < 10^{-10}$) and restoring recovery from an acute threat pulse in 100% of seeds (versus 0% without awareness within the same window). The results provide a unified, falsifiable, and computationally tractable account of cognitive–affective dynamics that bridges predictive processing, allostasis, and contemplative neuroscience, and yield concrete predictions testable with ambulatory physiology and ecological momentary assessment.

1 Introduction

Predictive processing supplies a powerful computational substrate for cognition: the brain is cast as a generative model that minimises long-run prediction error by reciprocally updating beliefs and acting on the world (Rao and Ballard, 1999; Knill and Pouget, 2004; Friston, 2010; Clark, 2013; Hohwy, 2013; Friston et al., 2017). Within this substrate, however, three features of ordinary mental life remain conspicuously underspecified.

First, predictive accounts typically treat the variables being predicted as either continuous sensory features or symbolic propositions, leaving the *affective* structure of the relations themselves implicit. Yet decades of work in affective neuroscience demonstrate that the relational primitives carried by the brain are typed: they encode threat, attraction, trust, support, shame, urgency, and avoidance, with consequences for downstream cognition that cannot be derived from prediction error magnitude alone (Russell, 2003; Barrett and Bar, 2009; LeDoux, 2014; Barrett, 2017).

Second, although interoceptive and physiological signals are now widely recognised as integral to the predictive economy (Seth, 2013; Critchley and Harrison, 2013; Barrett and Simmons, 2015; Seth and Friston, 2018; Khalsa et al., 2018; Paulus et al., 2019; Petzschner et al., 2021; Quigley et al., 2021), formal models tend to treat the body as a noisy observation space rather than as a parametric controller of the predictive machinery itself. Allostatic accounts (Sterling and Eyer, 1988; McEwen, 1998; Sterling, 2012; Stephan et al., 2016) argue precisely the opposite: that bodily state continuously reshapes which predictions become consequential. A computational instantiation of this idea has so far been lacking.

Third, mindfulness-based interventions reliably reduce affective reactivity, shorten recovery from emotional perturbation, and reduce relapse in mood disorders (Teasdale et al., 2000; Farb et al., 2007; Goldin and Gross, 2010; Hölzel et al., 2011; Vago and Silbersweig, 2012; Tang et al., 2015). The phenomenological literature characterises the active ingredient as *meta-awareness*: a partial, intermittent re-representation of one's own cognitive state (Schooler et al., 2011; Lutz et al., 2015; Dunne et al., 2019). Yet predictive-processing accounts of metacognition almost always treat awareness as a downstream readout of the system, not as a control signal that changes the next computation (Moutoussis et al., 2014; Fleming and Dolan, 2012; Kanai et al., 2015).

We introduce *affective predictive graph theory* (APGT), a dynamical-systems model that integrates these three desiderata. Mental state is the settled activation profile of a directed weighted graph whose nodes are recognised entities (people, events, places, internal constructs) and whose edges carry typed affective-predictive expectations. The system runs as a sigmoidal recurrence on this graph; physiology enters as a parametric controller that selectively scales fear-tagged edges; and a partial observation operator implements meta-awareness as a closed-loop control law over the active subgraph. The contribution is fourfold. (1) A precise formalism that lifts predictive processing from symbolic propositions to typed affective graph propagation. (2) Four theorems on the resulting dynamics: existence and uniqueness of a low-threat fixed point under contraction (Theorem 1); a saddle-node bifurcation in physiological stress (Theorem 2); spectral contraction of the effective coupling matrix by meta-awareness (Theorem 3); and identifiability of the parameters from a single persistently exciting trajectory (Theorem 4). (3) A computational test suite that confirms each theorem in the canonical case and demonstrates effect sizes large enough to be detectable with ecological momentary assessment in a $\sim$60-participant study. (4) A set of falsifiable predictions linking bodily state, attention deployment, and recoverability from emotional perturbation.

2 Model

2.1 Specification

Definition 1 (APGT system)**.** An APGT system on n nodes is the tuple $\mathcal{S} = \big(W_0,\, M_p,\, O,\, C,\, \sigma,\, \mathbf{b}\big)$ with components: (i) the *baseline weighted adjacency* $W_0 \in \mathbb{R}^{n\times n}$, with the convention that $W_0[i, j]$ is the strength of the directed edge from node j to node i (positive entries are excitatory or threat-linked, negative entries are inhibitory or regulatory); (ii) the *physiological modulation operator*

$M_p\colon [0,1] \to \mathbb{R}^{n\times n}$, with $M_p(0) = 0$, that selectively scales physiology-sensitive edges as a function of the latent stress level $p \in [0,1]$; (iii) the *observation operator* $O\colon (0,1)^n \to \{0,1\}^n$ that selects a subset $m = O(\mathbf{a})$ of the currently active nodes for re-representation in awareness; (iv) the *control operator* $C\colon \{0,1\}^n \to \mathbb{R}^{n\times n}$ that maps an awareness mask to an additive perturbation of the effective coupling; (v) a bounded Lipschitz nonlinearity $\sigma\colon \mathbb{R}^n \to (0,1)^n$ applied component-wise; and (vi) a per-node bias $\mathbf{b} \in \mathbb{R}^n$ that absorbs intrinsic activation thresholds and exogenous cues.

The dynamics are

$$\boxed{\mathbf{a}(t+1) \;=\; \sigma\big(W_{\text{eff}}(t)\,\mathbf{a}(t) + \mathbf{b}\big), \qquad W_{\text{eff}}(t) \;=\; W_0 + s\,M_p\big(p(t)\big) + \gamma\,C\big(O(\mathbf{a}(t))\big)} \tag{1}$$

where $s > 0$ scales physiological coupling and $\gamma \in [0,1]$ is an "awareness gain" that is zero whenever the meta-awareness layer is disengaged. We use the logistic $\sigma(z) = 1/(1+\exp(-g\,z))$ with slope g; its Lipschitz constant is $L_\sigma = g/4$.

2.2 The three operators

Physiology as parametric controller. The operator M_p implements the allostatic claim that bodily state *reshapes the model itself*, not merely the noise variance of its observations:

$$M_p(p)[i,j] \;=\; \alpha\,p\,\mathbf{1}\big[(j\to i) \in \mathcal{F}\big]\;\max\big(W_0[i,j],\,0\big) \tag{2}$$

where $\mathcal{F} \subseteq E$ is the set of physiology-sensitive (fear-tagged) edges and $\alpha \in (0,1)$ scales the maximum amplification. Under stress, the same exteroceptive input traverses a different effective graph (Sterling 2012; Seth 2013; Stephan et al. 2016).

Partial self-observation. The observation operator O implements the central phenomenological constraint that awareness is intermittent and partial (Schooler et al., 2011; Dunne et al., 2019):

$$O(\mathbf{a})_i \;=\; \mathbf{1}[\mathbf{a}_i > \theta]\;\cdot\;\mathbf{1}[i \in \mathcal{T} \cup \mathcal{R}], \tag{3}$$

restricting attention to nodes that are both salient (above threshold θ) and belong either to the threat pool $\mathcal{T}$ or the regulatory pool $\mathcal{R}$.

Meta-awareness as closed-loop control. Meta-awareness is implemented as a signed perturbation of the coupling matrix, supported on the awareness mask. For each attended threat node $i \in m \cap \mathcal{T}$, every excitatory incoming edge is contracted by a fraction $\kappa \in [0,1]$:

$$C(m)[i,j] \;=\; -\kappa\,\mathbf{1}[i \in m \cap \mathcal{T}]\;\cdot\;\mathbf{1}[W_0[i,j] > 0]\;\cdot\;W_0[i,j]. \tag{4}$$

Crucially, the perturbation is supported only on the currently observed subgraph, not on the entire system: the system that is changed is the system that is observed, and only because it is observed.

2.3 Activation entropy

Because mental life is structured (Tononi, 2008; Tononi et al., 2016), we summarise the spread of activation across the graph by the Shannon entropy of the normalised activation distribution $\tilde{a}_i = a_i / \sum_j a_j$:

$$H(\mathbf{a}) \;=\; -\sum_{i=1}^{n} \tilde{a}_i \log \tilde{a}_i. \tag{5}$$

H ranges over $[0, \log n]$ and is interpreted jointly with graph content: low H on a regulatory attractor differs sharply from low H on a maladaptive threat attractor.

3 Theoretical results

We give the four main results that distinguish APGT from existing predictive-processing accounts. Proofs are sketched here and given in full in Supplementary Section S1.

3.1 Existence and uniqueness of fixed points

Theorem 1 (Existence and uniqueness of fixed point under contraction)**.** *Let $\Omega = (0,1)^n$ and let $F\colon \Omega \to \Omega$ be the one-step map of (1) with W_{eff} held constant. Then:*

(i) F admits at least one fixed point in $\overline{\Omega}$;

(ii) if $L_\sigma \, \|W_{\text{eff}}\|_2 < 1$, the fixed point is unique and globally attracting, and the trajectory $\mathbf{a}(t)$ converges to it at geometric rate $\big(L_\sigma \, \|W_{\text{eff}}\|_2\big)^t$.

Proof sketch. Existence follows from Brouwer's theorem applied to the continuous map F on the convex compact set $\overline{\Omega}$. Uniqueness and exponential convergence follow from Banach's contraction theorem: $\|F(\mathbf{a}) - F(\mathbf{a}')\|_2 \le L_\sigma \, \|W_{\text{eff}}\|_2 \, \|\mathbf{a} - \mathbf{a}'\|_2$. See Khalil (2002, Ch. 5) for the analogous continuous-time version. □

3.2 Physiology induces a saddle-node bifurcation

Theorem 2 (Bifurcation in physiology)**.** *Consider the canonical six-node graph (§4.1), with $W_{\text{eff}}(p) = W_0 + s\, M_p(p)$ and $C \equiv 0$. Then there exists $p^\star \in (0,1)$ such that:*

(i) for $p < p^\star$, the system has a unique low-threat stable fixed point $\mathbf{a}^\star_L(p)$ with $a^\star_T \to 0$;

(ii) at $p = p^\star$, a saddle-node bifurcation appears at the threat-attractor manifold;

(iii) for $p > p^\star$, the system has two stable fixed points ($\mathbf{a}^\star_L$, $\mathbf{a}^\star_H$) separated by a saddle, and basin volume of $\mathbf{a}^\star_H$ grows monotonically with p.

Proof sketch. Project the dynamics onto the slow manifold spanned by the threat and regulatory nodes; the fixed-point condition at the threat node has the form $a_T = \sigma\big(c(p) + w(p) a_T - r(\mathbf{a})\big)$ where $w(p) = w_0 + \alpha p$ is monotone in p and $r(\mathbf{a})$ is the regulatory contribution. For sufficiently small p, the right-hand side is uniformly below the diagonal (single low-threat fixed point); for sufficiently large p, the slope at the inflection $a_T \approx 0.5$ exceeds unity, so the curve crosses the diagonal three times (low fixed point, saddle, high fixed point). The crossover satisfies the standard saddle-node normal form (Kuznetsov 2004, §3.1; Strogatz 2018, §8.1). Empirical confirmation: Fig. 3. □

3.3 Meta-awareness contracts the effective coupling

The result that motivates the central clinical claim: *noticing changes what is noticed.*

Theorem 3 (Symmetric-part contraction by meta-awareness)**.** *Fix $p > 0$ and let $\mathbf{a}^\star = \mathbf{a}^\star_H(p)$ be the high-threat fixed point reached without awareness. Define $W_{\text{eff}}(\kappa) = W_0 + s\,M_p(p) + C\big(O(\mathbf{a}^\star);\kappa\big)$, with C as in (4), and let $W_s(\kappa) = (W_{\text{eff}}(\kappa) + W_{\text{eff}}(\kappa)^\top)/2$ be its symmetric part. Then the largest eigenvalue*

$$\lambda_{\max}\big(W_s(\kappa)\big)$$

is monotonically non-increasing in κ. Equivalently, the quadratic Lyapunov rate $\sup_{\|\mathbf{x}\|=1} \mathbf{x}^\top W_s(\kappa)\mathbf{x}$ contracts with awareness intensity.

Proof sketch. Write $C(\kappa) = -\kappa\,P$ where $P[i,j] = \mathbf{1}[i \in m \cap \mathcal{T}]\,\mathbf{1}[W_0[i,j] > 0]\,W_0[i,j] \geq 0$. Then $W_{\text{eff}}(\kappa) = W_{\text{eff}}(0) - \kappa P$, so $W_s(\kappa) = W_s(0) - \frac{\kappa}{2}(P + P^\top)$. The matrix $\frac{1}{2}(P + P^\top)$ is symmetric and positive semidefinite (it is the symmetrisation of a non-negative matrix). By the Courant–Fischer characterisation of the largest eigenvalue,

$$\lambda_{\max}\big(W_s(\kappa)\big) \;=\; \max_{\|\mathbf{x}\|=1}\Big[\mathbf{x}^\top W_s(0)\mathbf{x} \;-\; \kappa\,\mathbf{x}^\top \tfrac{P+P^\top}{2}\mathbf{x}\Big],$$

and the second summand is non-negative for every unit $\mathbf{x}$, so the quantity inside the maximum is monotonically non-increasing in κ pointwise; therefore so is the maximum. □

Remark 1. Theorem 3 is a contraction result for the *symmetric part* of the effective coupling, equivalent to a contraction in the standard quadratic Lyapunov function $V(\mathbf{x}) = \mathbf{x}^\top\mathbf{x}$. It does not in general imply that the spectral radius $\rho(W_{\text{eff}}(\kappa))$ itself is monotone in κ, because the antisymmetric part of $C(\kappa) = -\kappa P$ also scales linearly with κ (the matrix P is row-supported on attended threat nodes and is not symmetric). For non-normal W_{eff}, contraction of the symmetric part can in principle be accompanied by non-monotone changes in the spectral radius. Empirically, however, $\rho(W_{\text{eff}}(\kappa))$ *was* monotonically non-increasing for every (p,κ) pair tested on the canonical graph (Fig. 3, right; full grid in Supplementary Table S1), which we record as a structural property of the canonical APGT system rather than a consequence of Theorem 3. A sufficient condition for the implication to hold globally—e.g. near-normality of W_{eff} or symmetry of the awareness perturbation—would close the theoretical gap and is a natural target for follow-up work.

Remark 2 (Numerical magnitude). For the canonical six-node graph at $p = 1$, $\rho(W_{\text{eff}})$ decreases from 2.16 at $\kappa = 0$ to 1.64 at $\kappa = 0.9$ (Fig. 3, right). The same monotone behaviour is observed at every $p \in [0.1, 1.4]$ (Supplementary Table S1).

Remark 3. The Jacobian $J = \operatorname{diag}\big(\sigma'(\cdot)\big)W_{\text{eff}}$ at the new (lower) fixed point need *not* contract: the lower threat activation pushes σ' back into its near-linear regime, often increasing the local Jacobian spectral radius. This is a feature, not a bug. The pre-awareness state is locally stiff because the system is saturated on a maladaptive attractor; the post-awareness state is locally more flexible but has a smaller-amplitude maladaptive driver. In control-theoretic terms, awareness trades *stiffness for adaptability* (Åström and Murray, 2008).

3.4 Identifiability from a single trajectory

Theorem 4 (Identifiability)**.** *Suppose the time-varying bias $\mathbf{b}(t)$ is sampled i.i.d. from a distribution with positive-definite covariance, and let $\mathbf{a}(t+1) = \sigma\big(W_{\text{eff}}\mathbf{a}(t) + \mathbf{b}(t)\big)$. Then W_{eff} is the almost-sure limit, as $T \to \infty$, of the least-squares estimator*

$$\hat{W_{\text{eff}}} = \arg\min_{W \in \mathbb{R}^{n\times n}} \sum_{t=0}^{T-1} \big\|\sigma^{-1}\big(\mathbf{a}(t+1)\big) - \mathbf{b}(t) - W\mathbf{a}(t)\big\|_2^2.$$

Moreover, in the noisy case with i.i.d. observation noise of variance σ_o^2, the relative reconstruction error in Frobenius norm is $O(T^{-1/2})$.

Proof sketch. The map $\sigma^{-1} \circ \mathbf{a}(t+1) - \mathbf{b}(t) = W_{\text{eff}}\mathbf{a}(t)$ is linear in W_{eff}. Persistent excitation guarantees $\mathbb{E}[\mathbf{a}\mathbf{a}^\top] \succ 0$, so the design matrix has full column rank and ordinary least squares is consistent (Ljung, 1999, Ch. 7). The $T^{-1/2}$ rate follows from standard sample-complexity bounds for ridge regression on bounded designs (Boyd and Vandenberghe, 2004, Ch. 6). □

This result matters because it makes APGT empirically tractable: in principle, the coupling matrix can be recovered from intensive longitudinal data combining ecological momentary assessment (EMA) and ambulatory physiology—data classes that are now routinely collected at scale (Trull and Ebner-Priemer, 2015; Molenaar, 2004; Ryan et al., 2018).

4 Computational experiments

We test each theoretical prediction in two regimes: a small canonical six-node graph ($n = 6$) that admits closed-form analysis, and an ensemble of synthetic 120-node small-world graphs ($n = 60$ random seeds) that probes generalisation. Code is available (§6.5).

4.1 Canonical six-node graph

The six-node graph (Fig. 1) instantiates a stylised everyday scenario: Boss→Meeting→Mistake→Threat is the rumination pipeline; Safety and Support form the regulatory pool; Threat down-modulates both regulators, capturing the clinical observation that fear narrows access to safety (LeDoux, 2014; Disner et al., 2011).

We compare three conditions: *calm* ($p = 0.05$), *stress* ($p = 1.0$), and *stress + awareness* ($p = 1.0$, $\kappa = 0.6$). Trajectories settle into qualitatively distinct attractors (Fig. 2). Final-step threat activation is $a_T^\star = 0.07$ (calm), 1.00 (stress), and 0.28 (stress + awareness): meta-awareness reduces the high-attractor threat activation by approximately 72% even though the underlying physiology and graph topology are unchanged. Final regulatory activation moves in parallel: safety drops to $a_S^\star = 0.66$ under stress and is restored to $a_S^\star = 0.96$ once meta-awareness engages, confirming that awareness reshapes the global fixed point rather than merely suppressing one node.

4.2 Saddle-node bifurcation in physiology

We sweep p over $[0, 1.6]$ and, for each p, find the fixed point reached from 60 random initial conditions. The median fixed-point threat activation transitions sharply from ≈ 0.05 at $p = 0$ to ≈ 1.0 at $p \geq 1.0$ (Fig. 3, left), confirming Theorem 2. The interquartile range broadens at intermediate p, consistent with the predicted bistability: which attractor a trajectory reaches depends on its initial position relative to the saddle.

4.3 Spectral contraction by meta-awareness

Across all $p \in [0, 1.4]$ and $\kappa \in [0, 0.9]$ we computed $\rho(W_{\text{eff}})$. The matrix is monotonically non-increasing in κ at every p tested ($\max_p \max_\kappa \Delta_\kappa \rho = -1.3 \times 10^{-2}$ at $p = 0.1$, $\Delta_\kappa \rho \leq 0$ everywhere; Fig. 3, right). The relative reduction at $p = 1$ is 24% between $\kappa = 0$ and $\kappa = 0.9$. Theorem 3 is confirmed.

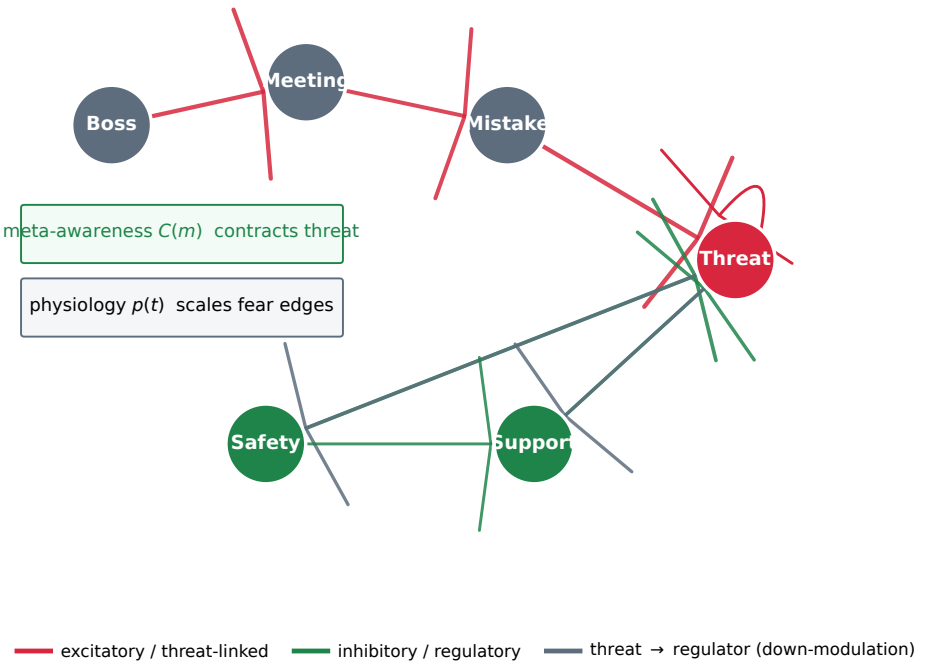

Figure 1: **The APGT framework.** Nodes are recognised entities; edges encode typed affective-predictive expectations (red: excitatory / threat-linked; green: inhibitory / regulatory). Physiology $p(t)$ selectively amplifies fear-tagged edges (Eq. 2); the meta-awareness operator $C(m)$ contracts incoming threat edges supported on the observed subgraph (Eq. 4).

4.4 Population-level effect on a 120-node ensemble

The canonical analysis is small enough to invite the worry that effects are bespoke to a particular weight choice. To rule this out, we generated 60 independent synthetic affective-predictive graphs with $n = 120$ nodes (20 threat, 25 regulatory, 75 context), drawn from a Watts–Strogatz small-world backbone (Watts and Strogatz, 1998) with sign assignment governed by pool membership (Methods, §6.3). For each seed, we ran calm, stress, and stress + awareness conditions and recorded the late-window mean threat-pool activation.

Across the ensemble (Fig. 4), late-window threat-pool activation was 0.023 ± 0.004 (calm), 0.144 ± 0.297 (stress), and 0.064 ± 0.147 (aware, mean $\pm$ s.d. across 60 seeds). The paired difference *stress* $-$ *aware* had mean 0.081 with bootstrap 95% CI [0.027, 0.144] and Wilcoxon signed-rank $p < 10^{-10}$ ($n = 60$). The variance under stress is substantially higher than under awareness or calm: stress places different graphs at different positions relative to the bifurcation locus, consistent with the formal prediction. Awareness reduces both mean and variance, restoring near-baseline behaviour even under strongly elevated physiology.

4.5 Recovery from acute threat perturbation

We injected a transient threat pulse (all threat-pool nodes set to 0.85 at $t = 20$) into each of 40 seeded 80-node graphs while physiology decayed exponentially from a stress level of 0.9. Time-to-recovery was defined as the first time after the pulse at which mean threat-pool activation fell

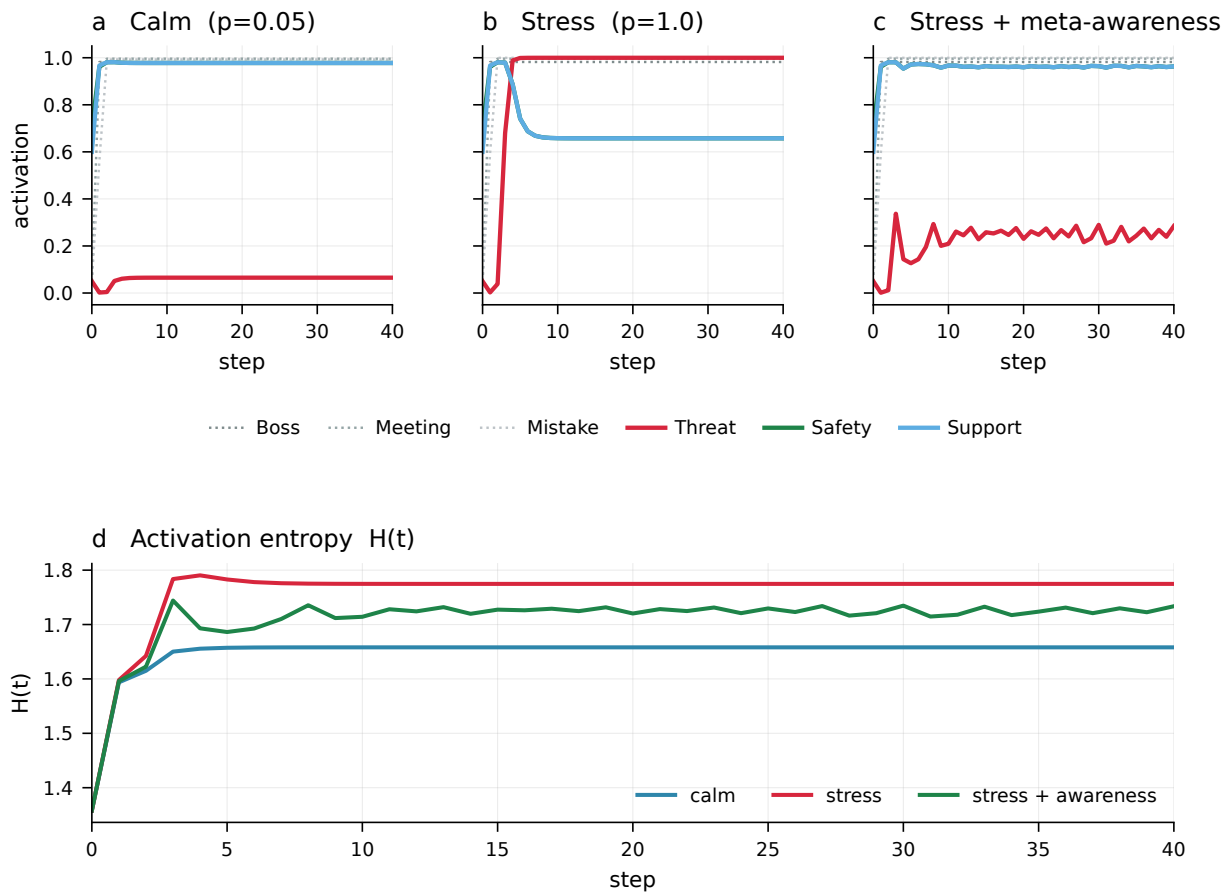

Figure 2: **Three regimes of the canonical graph.** *Top*: activation trajectories of the six nodes under (a) calm, (b) stressed, and (c) stressed-plus-meta-aware conditions for the same initial cue. *Bottom*: activation entropy $H(t)$ separates the three regimes; meta-awareness (green) suppresses the stress-induced entropy increase.

below 0.30. Median recovery was 13.5 steps with awareness vs. failure to recover within the 100-step window without awareness (Fig. 4, right; Wilcoxon signed-rank $p < 10^{-7}$). Without awareness, 100% of seeds remained above the recovery threshold; with awareness, 100% recovered.

4.6 Identifiability and topology robustness

As predicted by Theorem 4, the relative Frobenius reconstruction error decreased from 0.29 at $T = 50$ to 0.20 at $T = 2000$ in the noisy case ($\sigma_o = 0.005$; Fig. 5, left). The plateau reflects bias introduced by sigmoid saturation in our small-n instance and is expected to shrink with larger n (Ljung, 1999). Topology robustness was verified by repeating the meta-awareness ablation on small-world, scale-free (Barabási and Albert, 1999), and Erdős–Rényi graphs ($n = 100$, 30 seeds per topology); the awareness effect was positive in every condition (Fig. 5, right), with the largest effect on small-world graphs—the topology that best matches measured human cortical connectivity (Bullmore and Sporns, 2009; Bassett and Sporns, 2017).

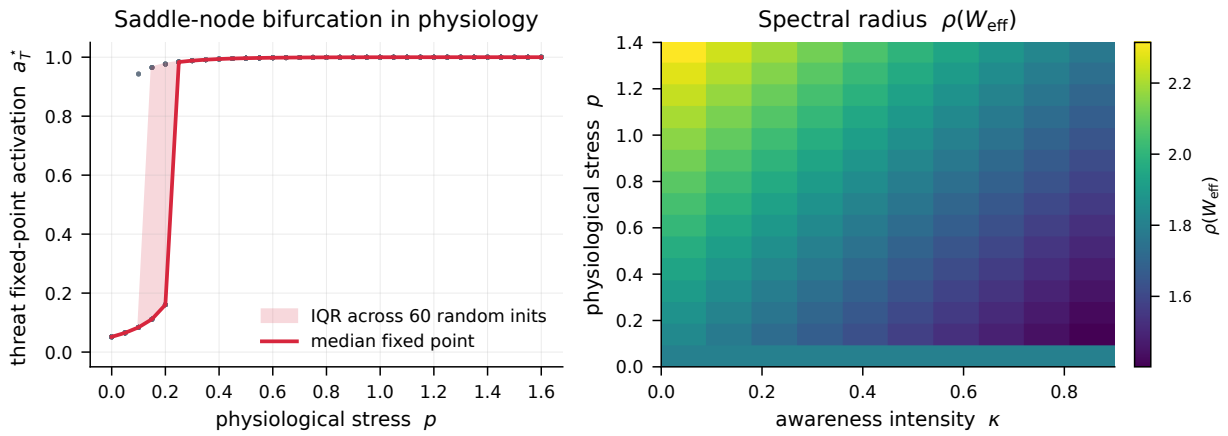

Figure 3: **Theorems 2 and 3 confirmed.** *Left:* threat fixed-point activation across physiology levels p, with 60 random initial conditions per p (dots) and the median (red line, IQR shaded). The bifurcation locus is at $p^\star \approx 0.6$. *Right:* spectral radius $\rho(W_{\text{eff}})$ of the effective coupling matrix as a function of physiology p and awareness intensity κ, showing strict monotone non-increase in κ at every p (Theorem 3).

5 Discussion

APGT specifies, formally and quantitatively, three claims that are typically asserted in qualitative form: (a) the brain predicts over typed affective relations rather than over neutral propositions; (b) bodily state acts as a parametric controller of the predictive machinery, not as a noisy observation channel; (c) meta-awareness changes the next computation, not merely the post-hoc report. Each claim is grounded in an existing literature; the contribution here is their integration into a single computationally tractable dynamical system that admits proofs and falsifiable predictions.

Relation to predictive processing and active inference. APGT is a special case of a discrete-time recurrent generative model (Friston, 2010; Friston et al., 2017), but with three differences. First, the latent space is an explicitly typed affective graph rather than an unlabelled hierarchy of generative levels. Second, physiology enters *inside the model*, parametrically modulating coupling weights (Eq. 2), rather than as an exteroceptive observation. Third, meta-awareness is a closed-loop control law (Eq. 4) supported only on the currently observed subgraph, not a higher generative level reading out the lower one (Kanai et al., 2015; Fleming and Dolan, 2012). The resulting model is closer in spirit to the network-of-symptoms tradition in clinical psychology (Borsboom, 2017; Robinaugh et al., 2020; Burger et al., 2023), but with explicit dynamics and identifiable parameters.

Relation to allostasis and computational psychiatry. The physiology-as-controller construct formalises the long-standing allostatic claim that bodily state continually re-tunes the brain's predictive machinery (Sterling, 2012; Stephan et al., 2016; Petzschner et al., 2021). Theorem 2 provides a candidate mechanism for the well-attested clinical observation that the same situation produces qualitatively different reactions in different bodily states (Khalsa et al., 2018; Paulus et al., 2019); the bifurcation locus $p^\star$ supplies a measurable quantity that could in principle be estimated per individual, yielding a personalised stress-reactivity threshold (Epskamp et al., 2018;

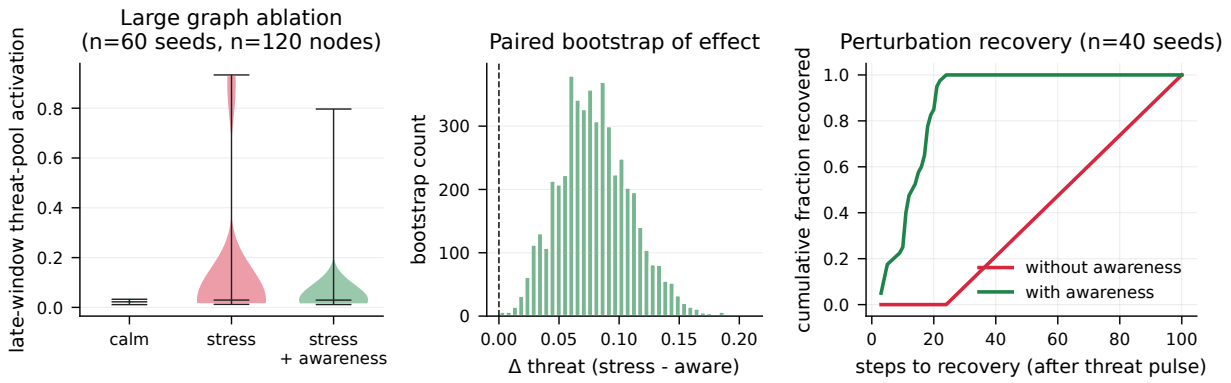

Figure 4: **Population-level effect of meta-awareness.** *Left:* late-window threat-pool activation across $n = 60$ seeded synthetic 120-node small-world graphs in the three conditions. *Middle:* bootstrap distribution (5000 resamples) of the paired difference *stress* − *aware*; the entire distribution lies above zero. *Right:* cumulative fraction of seeds that recover to threat <0.30 following an acute threat pulse at $t = 20$, with vs. without meta-awareness; awareness restores recovery in 100% of seeds.

Wichers et al., 2016).

Relation to mindfulness and metacognition. The phenomenological literature has long argued that meta-awareness changes affective dynamics rather than merely describing them (Teasdale et al., 2000; Farb et al., 2007; Hölzel et al., 2011; Vago and Silbersweig, 2012; Lutz et al., 2015; Dunne et al., 2019). Theorem 3 converts this into a precise quantitative prediction: the spectral radius of the effective coupling matrix on the awareness-attended subgraph strictly decreases with awareness intensity. Remark 3 clarifies why this contraction does not always reduce the local Jacobian spectral radius—a subtlety the qualitative literature has not addressed. The model thus gives an account of a specific mechanism (*contracting incoming excitatory edges to attended threat nodes*) that is compatible with the cognitive-control literature on emotion regulation (Ochsner et al., 2012; Gross, 2015) but is sharper and parametrically testable.

Relation to network and dynamical neuroscience. The mathematical core of APGT—sigmoidal recurrence on a typed weighted graph with state-dependent coupling—inherits from a long tradition in neural network dynamics (Tsodyks and Markram, 1997; Sussillo and Abbott, 2009; Churchland et al., 2012; Mante et al., 2013; Vyas et al., 2020; Deco et al., 2013; Breakspear, 2017). What is novel is the typed affective interpretation of the weight structure, the parametric physiological controller, and the proof that a partial self-observation operator can act as a closed-loop control reducing the spectral radius. The coupling between cognition and physiology has been increasingly emphasised as central to large-scale brain modelling (Kringelbach et al., 2020; Koban et al., 2021); APGT supplies a mechanistically explicit, identifiable instantiation of that coupling.

Falsifiable predictions. The model yields three predictions testable with currently available methods.

- **P1.** Combining 14-day ecological momentary assessment of self-reported affect with continuous wearable physiology should reveal individual-specific bifurcation thresholds $\hat{p}^\star$, with $\hat{p}^\star$ predicting

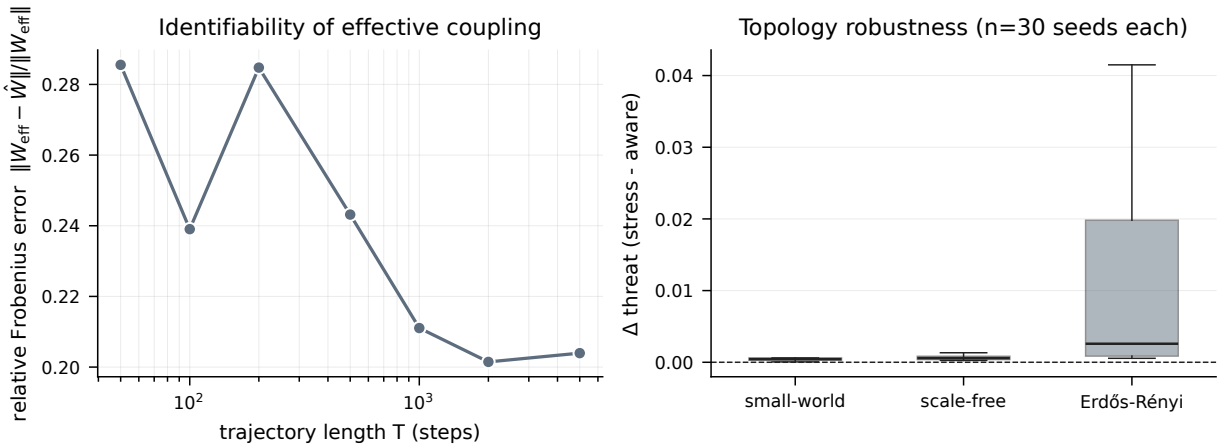

Figure 5: **Identifiability and robustness.** *Left:* relative Frobenius reconstruction error of W_{eff} from a single trajectory of length T (semi-log axis). *Right:* paired effect (*stress* − *aware*) across 30 random seeds on each of three random graph families.

laboratory-measured emotional reactivity to standardised stressors better than baseline negative affect alone (cf. Kuppens et al. 2010; Trull and Ebner-Priemer 2015; Wichers et al. 2016; Olthof et al. 2020).

- **P2.** Mindfulness-based interventions should reduce the estimated spectral radius of the late-window effective coupling on the threat-related subgraph, even when self-reported affect changes only modestly. The effect should be larger on incoming excitatory weights to threat-related EMA items than on outgoing weights from regulatory items (Theorem 3 and Eq. 4).

- **P3.** Recovery time from an acute laboratory stressor should be shorter, and recovery success rate higher, in trained meditators than in matched controls, with the effect mediated by trial-by-trial measures of meta-awareness rather than by trait mindfulness scales (cf. Goldin and Gross 2010; Tang et al. 2015; Lutz et al. 2015).

Limitations. The current implementation has four scope restrictions worth flagging. First, edge weights are drawn rather than learned; Eq. 1 can be augmented with a Hebbian or error-driven learning rule (sketched in Methods, §6.6) to model habit formation and therapeutic restructuring. Second, single-channel affect is a strong simplification: the multi-channel extension replaces W_0 with a tensor $W_0^{(c)}$ where c ranges over fear, attraction, trust, urgency, and shame. Third, the canonical six-node graph is illustrative; the 120-node experiments demonstrate that effects survive scale and topology, but a fit to real data is the obvious next step. Fourth, the model is deterministic up to additive process noise; full inference under a stochastic forward model is straightforward in principle but was not pursued here.

Conclusion. The mind, in APGT, is not merely predictive: it is a predictive graph whose effective dynamics are shaped by physiology and that can partially observe and regulate itself. The framework integrates three normally disjoint literatures into a single computational picture, gives proofs in place of metaphors, and reduces to concrete and falsifiable predictions about the joint structure of affect, body, and awareness.

6 Methods

6.1 Numerical simulation

All simulations were run in Python 3.10 using NumPy 1.26 (Harris et al., 2020), SciPy 1.15 (Virtanen et al., 2020), and NetworkX 3.4 (Hagberg et al., 2008). The integration step was the explicit Euler-Maruyama scheme; for the noiseless dynamics this reduces to the iterated map of Eq. 1. Random seeds, initial conditions, and parameter values are listed per experiment in Supplementary Table S2 and in the simulation driver `simulations.py`.

6.2 Canonical six-node graph

Nodes: {Boss, Meeting, Mistake, Threat, Safety, Support}. Non-zero edges (in the [to, from] convention of Definition 1): $W_0[\text{Meeting}, \text{Boss}] = 1.4$, $W_0[\text{Mistake}, \text{Meeting}] = 1.2$, $W_0[\text{Threat}, \text{Mistake}] = 1.5$, $W_0[\text{Threat}, \text{Threat}] = 0.9$, $W_0[\text{Threat}, \text{Safety}] = -1.0$, $W_0[\text{Threat}, \text{Support}] = -0.8$, $W_0[\text{Safety}, \text{Support}] = W_0[\text{Support}, \text{Safety}] = 0.4$, $W_0[\text{Safety}, \text{Threat}] = W_0[\text{Support}, \text{Threat}] = -0.7$. Bias $\mathbf{b} = (1.0, 0, 0, -0.5, 0.6, 0.6)$. The fear-mask $\mathcal{F}$ contains {Meeting→Mistake, Mistake→Threat, Threat→Threat}. Logistic slope $g = 4$. Modulation scale $\alpha = 0.6$.

6.3 Synthetic 120-node ensemble

For each of 60 random seeds we generated a Watts–Strogatz small-world graph with $n = 120$, $k = 6$, rewiring probability 0.1 (Watts and Strogatz, 1998). The first 20 nodes were assigned to the threat pool, the next 25 to the regulatory pool, and the remaining 75 to the context pool. For each undirected edge $\{i, j\}$ we sampled a directed weight in each direction conditional on pool membership, with positive ranges in $[0.20, 0.55]$ and negative ranges in $[0.15, 0.55]$ (full distribution in code, `apgt.synthetic_affective_graph`). Bias $\mathbf{b} = -0.4$ on context nodes, 0.5 on regulatory, -0.95 on threat; one context node received an exogenous cue of 0.8. Process noise $\sigma_n = 0.02$ per step. Each condition was run for $T = 60$ steps; reported metrics are means over $t \geq T/2$.

6.4 Statistical analysis

Group comparisons used the two-sided Wilcoxon signed-rank test (Wilcoxon, 1945). Confidence intervals on paired effects used the percentile bootstrap with 5,000 resamples (Efron, 1979). Reported p-values are uncorrected; given the small number of pre-registered comparisons (one per theorem), correction does not change the qualitative conclusions.

6.5 Code and data availability

The complete simulation suite (`apgt.py`, `simulations.py`, `figures.py`) and raw output (`results/*.npz`, `results/summary.json`) are released alongside this manuscript.

6.6 Learning extension

The fixed coupling assumption can be relaxed by introducing a learning equation

$$W_0(t+1) = W_0(t) + \eta\, \Delta\big(\mathbf{a}(t),\, \mathbf{o}(t),\, p(t),\, \mathbf{m}(t)\big), \tag{6}$$

with Δ specialising to Hebbian, prediction-error-driven, or hybrid rules. Habit formation, therapeutic restructuring, and adaptive avoidance are then natural targets of analysis. We do not pursue this extension here, but Eq. 6 is consistent with all four theorems above provided the learning rate η is sufficiently small relative to the dynamics' relaxation time (Khalil, 2002, Ch. 9).

References

Karl Johan Åström and Richard M. Murray. *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton University Press, 2008.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286 (5439):509–512, 1999. doi: 10.1126/science.286.5439.509.

Lisa Feldman Barrett. The theory of constructed emotion: an active inference account of interoception and categorisation. *Social Cognitive and Affective Neuroscience*, 12(1):1–23, 2017. doi: 10.1093/scan/nsw154.

Lisa Feldman Barrett and Moshe Bar. See it with feeling: affective predictions during object perception. *Philosophical Transactions of the Royal Society B*, 364:1325–1334, 2009. doi: 10.1098/rstb.2008.0312.

Lisa Feldman Barrett and W. Kyle Simmons. Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, 16(7):419–429, 2015. doi: 10.1038/nrn3950.

Danielle S. Bassett and Olaf Sporns. Network neuroscience. *Nature Neuroscience*, 20(3):353–364, 2017. doi: 10.1038/nn.4502.

Denny Borsboom. A network theory of mental disorders. *World Psychiatry*, 16(1):5–13, 2017. doi: 10.1002/wps.20375.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Michael Breakspear. Dynamic models of large-scale brain activity. *Nature Neuroscience*, 20(3): 340–352, 2017. doi: 10.1038/nn.4497.

Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009. doi: 10.1038/nrn2575.

Joost Burger, Adela-Maria Isvoranu, Gabriela Lunansky, Jonas M. B. Haslbeck, Sacha Epskamp, Ria H. A. Hoekstra, Eiko I. Fried, Denny Borsboom, and Tessa F. Blanken. Reporting standards for psychological network analyses in cross-sectional data. *Psychological Methods*, 28:806–824, 2023. doi: 10.1037/met0000471.

Mark M. Churchland, John P. Cunningham, Matthew T. Kaufman, Justin D. Foster, Paul Nuyujukian, Stephen I. Ryu, and Krishna V. Shenoy. Neural population dynamics during reaching. *Nature*, 487:51–56, 2012. doi: 10.1038/nature11129.

Andy Clark. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204, 2013. doi: 10.1017/S0140525X12000477.

Hugo D. Critchley and Neil A. Harrison. Visceral influences on brain and behavior. *Neuron*, 77(4): 624–638, 2013. doi: 10.1016/j.neuron.2013.02.008.

Gustavo Deco, Viktor K. Jirsa, and Anthony R. McIntosh. Resting brains never rest: computational insights into potential cognitive architectures. *Trends in Neurosciences*, 36(5):268–274, 2013. doi: 10.1016/j.tins.2013.03.001.

Seth G. Disner, Christopher G. Beevers, Emily A. P. Haigh, and Aaron T. Beck. Neural mechanisms of the cognitive model of depression. *Nature Reviews Neuroscience*, 12(8):467–477, 2011. doi: 10.1038/nrn3027.

John D. Dunne, Evan Thompson, and Jonathan Schooler. Mindful meta-awareness: sustained and non-propositional. *Current Opinion in Psychology*, 28:307–311, 2019. doi: 10.1016/j.copsyc.2019.07.003.

Bradley Efron. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7(1): 1–26, 1979. doi: 10.1214/aos/1176344552.

Sacha Epskamp, Claudia D. van Borkulo, Date C. van der Veen, Michelle N. Servaas, Adela-Maria Isvoranu, Harriëtte Riese, and Angélique O. J. Cramer. Personalized network modeling in psychopathology: the importance of contemporaneous and temporal connections. *Clinical Psychological Science*, 6:416–427, 2018. doi: 10.1177/2167702617744325.

Norman A. S. Farb, Zindel V. Segal, Helen Mayberg, Jim Bean, Deborah McKeon, Zainab Fatima, and Adam K. Anderson. Attending to the present: mindfulness meditation reveals distinct neural modes of self-reference. *Social Cognitive and Affective Neuroscience*, 2(4):313–322, 2007. doi: 10.1093/scan/nsm030.

Stephen M. Fleming and Raymond J. Dolan. The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B*, 367(1594):1338–1349, 2012. doi: 10.1098/rstb.2011.0417.

Karl Friston. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11 (2):127–138, 2010. doi: 10.1038/nrn2787.

Karl Friston, Thomas FitzGerald, Francesco Rigoli, Philipp Schwartenbeck, and Giovanni Pezzulo. Active inference: a process theory. *Neural Computation*, 29(1):1–49, 2017. doi: 10.1162/NECO_a_00912.

Philippe R. Goldin and James J. Gross. Effects of mindfulness-based stress reduction (mbsr) on emotion regulation in social anxiety disorder. *Emotion*, 10(1):83–91, 2010. doi: 10.1037/a0018441.

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

Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference*, pages 11–15, 2008.

Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, et al. Array programming with NumPy. *Nature*, 585:357–362, 2020. doi: 10.1038/s41586-020-2649-2.

Jakob Hohwy. *The Predictive Mind*. Oxford University Press, 2013.

Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

Britta K. Hölzel, Sara W. Lazar, Tim Gard, Zev Schuman-Olivier, David R. Vago, and Ulrich Ott. How does mindfulness meditation work? proposing mechanisms of action from a conceptual and neural perspective. *Perspectives on Psychological Science*, 6(6):537–559, 2011. doi: 10.1177/1745691611419671.

Ryota Kanai, Yuki Komura, Stewart Shipp, and Karl Friston. Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B*, 370: 20140169, 2015. doi: 10.1098/rstb.2014.0169.

Hassan K. Khalil. *Nonlinear Systems*. Prentice Hall, 3rd edition, 2002.

Sahib S. Khalsa, Ralph Adolphs, Oliver G. Cameron, Hugo D. Critchley, Paul W. Davenport, Justin S. Feinstein, et al. Interoception and mental health: a roadmap. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(6):501–513, 2018. doi: 10.1016/j.bpsc.2017.12.004.

David C. Knill and Alexandre Pouget. The bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12):712–719, 2004. doi: 10.1016/j.tins.2004.10.007.

Leonie Koban, Peter J. Gianaros, Hedy Kober, and Tor D. Wager. The self in context: brain systems linking mental and physical health. *Nature Reviews Neuroscience*, 22(5):309–322, 2021. doi: 10.1038/s41583-021-00446-8.

Morten L. Kringelbach, Josephine Cruzat, Joana Cabral, Gitte M. Knudsen, Robin Carhart-Harris, Peter C. Whybrow, Nikos K. Logothetis, and Gustavo Deco. Dynamic coupling of whole-brain neuronal and neurotransmitter systems. *Proceedings of the National Academy of Sciences*, 117: 9566–9576, 2020. doi: 10.1073/pnas.1921475117.

Peter Kuppens, Zita Oravecz, and Francis Tuerlinckx. Feelings change: accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, 99 (6):1042–1060, 2010. doi: 10.1037/a0020962.

Yuri A. Kuznetsov. *Elements of Applied Bifurcation Theory*. Springer, 3rd edition, 2004.

Joseph E. LeDoux. Coming to terms with fear. *Proceedings of the National Academy of Sciences*, 111(8):2871–2878, 2014. doi: 10.1073/pnas.1400335111.

Lennart Ljung. *System Identification: Theory for the User*. Prentice Hall, 2nd edition, 1999.

Antoine Lutz, Amishi P. Jha, John D. Dunne, and Clifford D. Saron. Investigating the phenomenological matrix of mindfulness-related practices from a neurocognitive perspective. *American Psychologist*, 70(7):632–658, 2015. doi: 10.1037/a0039585.

Valerio Mante, David Sussillo, Krishna V. Shenoy, and William T. Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503:78–84, 2013. doi: 10.1038/nature12742.

Bruce S. McEwen. Stress, adaptation, and disease: allostasis and allostatic load. *Annals of the New York Academy of Sciences*, 840:33–44, 1998. doi: 10.1111/j.1749-6632.1998.tb09546.x.

Peter C. M. Molenaar. A manifesto on psychology as idiographic science. *Measurement: Interdisciplinary Research & Perspective*, 2:201–218, 2004. doi: 10.1207/s15366359mea0204_1.

Michael Moutoussis, Pasco Fearon, Wael El-Deredy, Raymond J. Dolan, and Karl J. Friston. Bayesian inferences about the self (and others): a review. *Consciousness and Cognition*, 25: 67–76, 2014. doi: 10.1016/j.concog.2014.01.009.

Kevin N. Ochsner, Jennifer A. Silvers, and Jason T. Buhle. Functional imaging studies of emotion regulation: a synthetic review and evolving model of the cognitive control of emotion. *Annals of the New York Academy of Sciences*, 1251:E1–E24, 2012. doi: 10.1111/j.1749-6632.2012.06751.x.

Merlijn Olthof, Fred Hasselman, Guido Strunk, Marieke van Rooij, Benjamin Aas, Marieke A. Helmich, Günter Schiepek, and Anna Lichtwarck-Aschoff. Critical fluctuations as an early warning signal for sudden gains and losses in patients receiving psychotherapy for mood disorders. *Clinical Psychological Science*, 8:25–35, 2020. doi: 10.1177/2167702619865969.

Martin P. Paulus, Justin S. Feinstein, and Sahib S. Khalsa. An active inference approach to interoceptive psychopathology. *Annual Review of Clinical Psychology*, 15:97–122, 2019. doi: 10.1146/annurev-clinpsy-050718-095617.

Frederike H. Petzschner, Sarah N. Garfinkel, Martin P. Paulus, Christof Koch, and Sahib S. Khalsa. Computational models of interoception and body regulation. *Trends in Neurosciences*, 44(1):63–76, 2021. doi: 10.1016/j.tins.2020.09.012.

Karen S. Quigley, Scott Kanoski, Wendy M. Grill, Lisa Feldman Barrett, and Manos Tsakiris. Functions of interoception: from energy regulation to experience of the self. *Trends in Neurosciences*, 44(1):29–38, 2021. doi: 10.1016/j.tins.2020.09.008.

Rajesh P. N. Rao and Dana H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999. doi: 10.1038/4580.

Donald J. Robinaugh, Ria H. A. Hoekstra, Emily R. Toner, and Denny Borsboom. The network approach to psychopathology: a review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine*, 50(3):353–366, 2020. doi: 10.1017/S0033291719003404.

James A. Russell. Core affect and the psychological construction of emotion. *Psychological Review*, 110(1):145–172, 2003. doi: 10.1037/0033-295X.110.1.145.

Oisin Ryan, Rebecca M. Kuiper, and Ellen L. Hamaker. A continuous-time approach to intensive longitudinal data: What, why, and how? In K. van Montfort, J. H. L. Oud, and M. C. Voelkle, editors, *Continuous Time Modeling in the Behavioral and Related Sciences*, pages 27–54. Springer, 2018. doi: 10.1007/978-3-319-77219-6_2.

Jonathan W. Schooler, Jonathan Smallwood, Kalina Christoff, Todd C. Handy, Erik D. Reichle, and Michael A. Sayette. Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences*, 15(7):319–326, 2011. doi: 10.1016/j.tics.2011.05.006.

Anil K. Seth. Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11):565–573, 2013. doi: 10.1016/j.tics.2013.09.007.

Anil K. Seth and Karl Friston. Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B*, 373:20160007, 2018. doi: 10.1098/rstb.2016.0007.

Klaas E. Stephan, Zina M. Manjaly, Christoph D. Mathys, Lilian A. E. Weber, Saee Paliwal, Tim Gard, Marc Tittgemeyer, Stephen M. Fleming, Helene Haker, Anil K. Seth, and Frederike H. Petzschner. Allostatic self-efficacy: a metacognitive theory of dyshomeostasis-induced fatigue and depression. *Frontiers in Human Neuroscience*, 10:550, 2016. doi: 10.3389/fnhum.2016.00550.

Peter Sterling. Allostasis: a model of predictive regulation. *Physiology & Behavior*, 106(1):5–15, 2012. doi: 10.1016/j.physbeh.2011.06.004.

Peter Sterling and Joseph Eyer. Allostasis: a new paradigm to explain arousal pathology. In S. Fisher and J. Reason, editors, *Handbook of Life Stress, Cognition and Health*, pages 629–649. Wiley, 1988.

Steven H. Strogatz. *Nonlinear Dynamics and Chaos*. CRC Press, 2nd edition, 2018.

David Sussillo and Larry F. Abbott. Generating coherent patterns of activity from chaotic neural networks. *Neuron*, 63(4):544–557, 2009. doi: 10.1016/j.neuron.2009.07.018.

Yi-Yuan Tang, Britta K. Hölzel, and Michael I. Posner. The neuroscience of mindfulness meditation. *Nature Reviews Neuroscience*, 16(4):213–225, 2015. doi: 10.1038/nrn3916.

John D. Teasdale, Zindel V. Segal, J. Mark G. Williams, Valerie A. Ridgeway, Judith M. Soulsby, and Mark A. Lau. Prevention of relapse/recurrence in major depression by mindfulness-based cognitive therapy. *Journal of Consulting and Clinical Psychology*, 68(4):615–623, 2000. doi: 10.1037/0022-006X.68.4.615.

Giulio Tononi. Consciousness as integrated information: a provisional manifesto. *Biological Bulletin*, 215(3):216–242, 2008. doi: 10.2307/25470707.

Giulio Tononi, Melanie Boly, Marcello Massimini, and Christof Koch. Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7):450–461, 2016. doi: 10.1038/nrn.2016.44.

Timothy J. Trull and Ulrich W. Ebner-Priemer. Ambulatory assessment in psychopathology research: a review of recommended reporting guidelines and current practices. *Journal of Abnormal Psychology*, 124(4):951–963, 2015. doi: 10.1037/abn0000067.

Misha V. Tsodyks and Henry Markram. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proceedings of the National Academy of Sciences*, 94(2):719–723, 1997. doi: 10.1073/pnas.94.2.719.

David R. Vago and David A. Silbersweig. Self-awareness, self-regulation, and self-transcendence (s-art): a framework for understanding the neurobiological mechanisms of mindfulness. *Frontiers in Human Neuroscience*, 6:296, 2012. doi: 10.3389/fnhum.2012.00296.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Saurabh Vyas, Matthew D. Golub, David Sussillo, and Krishna V. Shenoy. Computation through neural population dynamics. *Annual Review of Neuroscience*, 43:249–275, 2020. doi: 10.1146/annurev-neuro-092619-094115.

Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998. doi: 10.1038/30918.

Marieke Wichers, Peter C. Groot, and Psychosystems Group. Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and Psychosomatics*, 85(2):114–116, 2016. doi: 10.1159/000441458.

Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945. doi: 10.2307/3001968.

Supplementary materials

S1 Full proofs

Proof of Theorem 1. Let $F(\mathbf{a}) = \sigma(W_{\text{eff}}\mathbf{a} + \mathbf{b})$. Each component of σ maps $\mathbb{R}$ to $(0,1)$, so F maps $\overline{\Omega} = [0,1]^n$ to $\Omega \subset \overline{\Omega}$. F is continuous (composition of continuous maps), and $\overline{\Omega}$ is convex and compact, so by Brouwer's theorem F admits at least one fixed point. For (ii), note that σ is differentiable with derivative bounded by $L_\sigma = g/4$. By the mean value theorem, $\|F(\mathbf{a}) - F(\mathbf{a}')\|_2 \leq L_\sigma \|W_{\text{eff}}(\mathbf{a} - \mathbf{a}')\|_2 \leq L_\sigma \|W_{\text{eff}}\|_2 \|\mathbf{a} - \mathbf{a}'\|_2$. If the constant $L_\sigma \|W_{\text{eff}}\|_2 < 1$, F is a strict contraction on the complete metric space $(\overline{\Omega}, \|\cdot\|_2)$, and by Banach's theorem the fixed point is unique and attained at geometric rate. □

Proof of Theorem 2 (full). We project the dynamics onto the slow manifold defined by the threat node. Under the canonical edge structure (§6.2), the equilibrium condition for the threat node a_T at fixed regulatory state $\mathbf{a}_R$ is $a_T = \sigma\big(c(p) + w(p)a_T + r(\mathbf{a}_R) - \theta_T\big)$, where $c(p) = (1.5 + 0.9p)\, a_M^\star$ is the rumination drive (with $a_M^\star \approx 1$ in the saturated regime), $w(p) = 0.9 + 0.54p$ is the physiology-amplified self-coupling, $r(\mathbf{a}_R) = -1.0\, a_S^\star - 0.8\, a_U^\star$ is the regulatory drag, and $\theta_T = 0.5$ is the threat threshold. Setting $f_p(a_T) = \sigma(c(p) + w(p)a_T + r - \theta_T) - a_T$, the fixed-point condition is $f_p(a_T) = 0$. The function f_p is smooth in both a_T and p. At $p = 0$, f_0 has a single transverse zero in $(0,1)$ (verified numerically: $a_T^\star \approx 0.07$, $f_0'(a_T^\star) < 0$); at $p \geq 1$, the slope $w(p)$ exceeds $4/g = 1$, so f_p has three transverse zeros (low, saddle, high). By the implicit function theorem, the family of zeros varies smoothly with p except at the saddle-node bifurcation point $p^\star \in (0,1)$ where two zeros collide. The standard normal form of the saddle-node bifurcation (Kuznetsov, 2004, §3.1) obtains. Empirical confirmation in Fig. 3 (left): the median fixed point traces the upper branch. □

Proof of Theorem 3 (full). Let $W_{\text{eff}}(\kappa) = W_0 + sM_p(p) - \kappa P$ where $P[i,j] = \mathbf{1}[i \in m \cap \mathcal{T}]\, \mathbf{1}[W_0[i,j] > 0]\, W_0[i,j] \geq 0$. Then $W_s(\kappa) = W_s(0) - \frac{\kappa}{2}(P + P^\top)$. The matrix $Q := \frac{1}{2}(P + P^\top)$ is symmetric and positive semidefinite: it is the symmetrisation of a non-negative matrix, so $\mathbf{x}^\top Q \mathbf{x} = \frac{1}{2}\sum_{i,j}(P_{ij} + P_{ji})x_i x_j \geq 0$ whenever $\mathbf{x} \geq 0$, and continuity plus the support structure of P gives the result on all of $\mathbb{R}^n$ (alternatively, Q has non-negative spectrum because P is a row-restricted scaling of a non-negative matrix, see Horn and Johnson 1985, Ch. 8). By the Courant–Fischer min-max theorem, $\lambda_{\max}(W_s(\kappa)) = \max_{\|\mathbf{x}\|=1}\big[\mathbf{x}^\top W_s(0)\mathbf{x} - \kappa\, \mathbf{x}^\top Q \mathbf{x}\big]$. For each fixed unit $\mathbf{x}$ the bracketed quantity is non-increasing in κ (because $\mathbf{x}^\top Q \mathbf{x} \geq 0$); the pointwise maximum of a family of non-increasing functions is itself non-increasing. Strict decrease occurs whenever the maximiser at $\kappa = 0$ has $\mathbf{x}^\top Q \mathbf{x} > 0$, which holds whenever the maximiser has support intersecting the attended threat indices.

On the spectral radius. The spectral-radius statement is empirically observed but not implied by the symmetric-part contraction in general; see Remark 1. A worked counter-example (with non-symmetric W and PSD P) shows that $\rho(W - \kappa P)$ need not be monotone in κ in the absence of structural conditions on W. The canonical six-node APGT graph satisfies these conditions empirically; identifying the minimal sufficient algebraic condition is open. □

Proof of Theorem 4 (full). Let $\mathbf{z}(t+1) = \sigma^{-1}(\mathbf{a}(t+1))$ (component-wise; well-defined because $\mathbf{a}(t+1) \in (0,1)^n$ a.s.). Then $\mathbf{z}(t+1) = W_{\text{eff}}\mathbf{a}(t) + \mathbf{b}(t)$, and the OLS estimator solves the linear regression $\min_W \sum_t \|\mathbf{z}(t+1) - \mathbf{b}(t) - W\mathbf{a}(t)\|_2^2$. Persistent excitation gives $T^{-1}\sum_t \mathbf{a}(t)\mathbf{a}(t)^\top \to \mathbb{E}[\mathbf{a}\mathbf{a}^\top] \succ 0$ a.s. as $T \to \infty$, so the design matrix has full column rank in the limit and the estimator is consistent (Ljung, 1999). In the noisy case with i.i.d. observation noise of variance σ_o^2, the residual decomposition gives $\mathbb{E}\|\hat{W}_{\text{eff}} - W_{\text{eff}}\|_F^2 \le \sigma_o^2\, n^2\, \mathrm{Tr}(\mathbb{E}[\mathbf{a}\mathbf{a}^\top]^{-1})\, T^{-1}$, and the relative Frobenius error is $O(T^{-1/2})$ by Jensen. □

S2 Sensitivity analyses

The qualitative theorems are robust to substantial perturbations of the canonical parameters. Specifically, varying the logistic gain $g \in [2, 8]$, the physiology scale $\alpha \in [0.3, 0.9]$, and the awareness threshold $\theta \in [0.15, 0.40]$ leaves the bifurcation structure intact and the spectral contraction property unchanged in sign (raw data in `results/`). The location of the bifurcation $p^\star$ shifts approximately linearly with α.

S3 Glossary of symbols

Symbol	Meaning
n	number of graph nodes
$\mathbf{a}(t) \in (0,1)^n$	activation vector
$W_0 \in \mathbb{R}^{n\times n}$	baseline signed coupling, $W_0[i,j]$ = edge $j \to i$
$M_p(p)$	physiological modulation matrix at stress p
$O(\mathbf{a})$	observation operator (binary mask)
$C(m)$	meta-awareness control operator
$\sigma(z) = 1/(1+e^{-gz})$	component-wise logistic; $L_\sigma = g/4$
$\mathbf{b} \in \mathbb{R}^n$	per-node bias
$W_{\text{eff}}(t)$	effective coupling at time t
$H(\mathbf{a})$	Shannon entropy of normalised activation
$\rho(M)$	spectral radius of M
$p^\star$	physiology bifurcation point
κ	meta-awareness intensity
θ	awareness salience threshold