Review Article

From Circuit Descriptions to Testable Mechanism Space in PTSD: A Minimal, Identifiability-Aware Computational Framework for Heterogeneous Threat Inference

Regio Marcos Pinto Abreu Filho[1]

1. PMERJ, Brazil

Post-traumatic stress disorder (PTSD) is commonly described in terms of frontolimbic circuit alterations and stress-system dysregulation. Yet the central translational problem remains mechanistic: patients with similar symptom severity can differ in why fear persists, why safety fails under stress, and why relapse follows context shifts. Here we synthesize convergent evidence across fear circuitry, noradrenergic and endocrine gain control, and context-dependent learning, and we propose a minimal computational framework that is explicitly designed to be (i) hypothesis-generating rather than confirmatory, and (ii) constrained by identifiability limits in realistic multimodal human datasets.

The framework separates fast threat-expression dynamics from slower latent-context inference and learning, and compresses heterogeneity into four composite dimensions: *Control* (stress-fragile regulation), *Context* (imprecise or biased context inference), *Gain* (arousal-amplified expression), and *Recovery* (feedback and return-to-baseline). We emphasize that "attractor" language is used as an inference-level coarse-graining, not as a directly demonstrated property of human PTSD circuitry. We then derive discriminative, falsifiable predictions and provide a translational mapping that distinguishes plausible causal levers from state-dependent modulators and correlational markers. The goal is not a PTSD biomarker, but an adjudicable mechanism space: a disciplined way to ask which process is dominant in a given person, under which conditions, and with what decision-relevant consequences.

 doi.org/10.32388/76FBO7.3

Corresponding author: Regio Marcos Abreu Filho, regiomarcosabreu@gmail.com

Epistemic status (read first)

What this paper is: a theory-driven synthesis plus a minimal *phenomenological* formalism that translates circuit-and-stress concepts into *testable* predictions and study designs.

What it is not: an empirically fitted model of PTSD; not a claim that human PTSD circuitry has proven multistable attractors; not a proposal that biomarkers can diagnose PTSD or replace clinical formulation.

Design principle: when parameters are not identifiable in practice, we prioritize *identifiable predictions* (and composite dimensions) over precise parameter estimation.

1. Introduction

PTSD affects a substantial minority of trauma-exposed individuals and is characterized by intrusive re-experiencing, avoidance, negative alterations in cognition and mood, and hyperarousal [1][2]. Contemporary neurobiological accounts consistently implicate distributed systems spanning amygdala-centered threat processing, medial prefrontal regulation, hippocampal context encoding, and stress physiology [3][4]. Yet two persistent translational gaps remain.

First, circuit descriptions often specify *where* differences are observed but less often specify *how* multi-timescale interactions generate hallmark clinical phenomena: (i) dissociation between explicit safety knowledge and persistent defensive responding, (ii) relapse under stress despite "knowing better," and (iii) context-dependent return of fear after extinction [5][6]. Second, PTSD is heterogeneous: similar symptom totals can emerge from different dominant mechanisms, and the same patient may shift mechanisms across contexts and states.

This paper contributes a minimal computational–clinical synthesis aimed at producing *discriminative predictions* rather than comprehensive biophysical realism. The framework is deliberately low-dimensional, but it is also *identifiability-aware*: we explicitly address the risk that richly parameterized models can become empirically underdetermined (degenerate) in realistic human multimodal datasets.

2. Empirical substrate: what is stable, what is heterogeneous

2.1. Fear circuitry as a distributed control problem (not a lesion model)

Across species, fear learning and extinction can be modeled as acquisition of threat value plus context-dependent inhibitory learning, rather than erasure [5][6]. Human PTSD studies implicate distributed abnormalities across amygdala, medial prefrontal regions, hippocampus, and their interactions, but effect sizes and directions can be moderated by trauma type, comorbidity, medication, and measurement choices [3][4]. The appropriate inference is therefore *process-level*: PTSD is plausibly a disorder of threat inference and its regulation, not a unitary circuit defect.

2.2. Structural findings: informative but not determinative

Large-scale consortium work supports smaller hippocampal volume in PTSD, with important moderators and interpretive constraints [7][8]. Structural findings are best treated as partial constraints on plausible mechanisms, not as direct mechanistic proofs.

2.3. Stress and arousal reshape expression and learning

Arousal changes do not merely "add symptoms"; they can alter how threat signals are amplified, how control is deployed, and how context is inferred. The locus coeruleus–noradrenaline (LC–NE) system is a canonical gain-control architecture that modulates signal-to-noise and behavioral policy under uncertainty [9]. Stress also impairs prefrontal function via well-characterized pathways, producing state-dependent reductions in cognitive control [10]. For the HPA axis, meta-analytic evidence supports heterogeneity rather than a single cortisol phenotype; altered glucocorticoid sensitivity and feedback dynamics are plausible dimensions of variation [11][12].

3. A minimal multi-timescale framework

3.1. Why "minimal" is a methodological commitment

Mechanistic ambition must be balanced against identifiability. Complex dynamical models frequently exhibit "sloppiness," where many parameter combinations yield similar outputs, limiting the interpretability of fitted parameters without strong experimental constraints [13]. We therefore construct a minimal model that (i) yields discriminative predictions, (ii) can be linked to tasks and perturbations,

and (iii) can be reduced to composite dimensions even when individual parameters cannot be uniquely recovered.

3.2. Layer A (fast): threat expression with stress-fragile control

Let x_t denote a coarse threat-expression state and u_t a coarse regulatory-control state, both at behavioral timescales (seconds–minutes). We model their interaction phenomenologically:

$$x_{t+1} = x_t + f(x_t;\theta_x) - \underbrace{c_t u_t}_{\text{effective control}} + \underbrace{g_t}_{\text{gain}} + \epsilon_t^x, \tag{1}$$

$$u_{t+1} = u_t + h(u_t;\theta_u) - \underbrace{d_t x_t}_{\text{interference}} + \eta_t. \tag{2}$$

Here g_t is a state-dependent gain term (LC–NE / arousal / endocrine influences), and c_t captures *control availability* which can drop under stress (state-dependent control collapse). This is not a claim about specific synapses; it is a compact representation of an empirically testable idea: *the same learned safety can fail to express when gain is high or control is fragile*.

3.3. Layer B (slow): latent context inference and learning

Let C_t denote a latent context state and o_t observed cues/outcomes. Context inference is Bayesian:

$$P(C_t \mid o_{1:t}) \propto P(o_t \mid C_t;\pi) P(C_t \mid C_{t-1}), \tag{3}$$

where π is *context precision*. Threat value learning is context-sensitive (any standard RL variant suffices at this level of abstraction):

$$V_{t+1} = V_t + \alpha(C_t)\,(r_t - V_t). \tag{4}$$

Expression is a function of learned threat value plus current state:

$$\mathrm{FearExpression}_t = \Phi\,(V_t, x_t, g_t). \tag{5}$$

Low π (imprecise context inference) and/or biased priors over context can yield overgeneralization and renewal-like relapse after context shifts [5][6].

3.4. Composite mechanism space (four dimensions)

To avoid overparameterization, we compress to four composite dimensions that can be operationalized even when micro-parameters are not identifiable:

- **Control (CTRL)**: stress-fragile regulation (effective c_t falls when arousal rises; interference d_t rises).
- **Context (CTX)**: low precision or biased priors in context inference (low π and/or biased $P(C)$).

- **Gain (GAIN)**: arousal-amplified expression (high g_t; LC–NE / autonomic state).
- **Recovery (REC)**: return-to-baseline dynamics (feedback and damping in stress physiology).

These are not diagnostic categories; they are *axes* along which patients (and states within patients) can vary.

4. Identifiability and degeneracy: constraints, not footnotes

4.1. Structural vs practical identifiability

Even if a model is identifiable in principle, real datasets can render parameters practically non-identifiable due to noise, limited sampling, and correlated measurements [14]. In sloppy regimes, the recommended stance is to focus on predictions that are stable across broad parameter ranges [13].

4.2. Design strategies to make the framework testable

We propose three strategies aligned with model-based cognitive neuroscience:

1. **Composite inference:** infer CTRL/CTX/GAIN/REC rather than a high-dimensional parameter vector.
2. **Orthogonal perturbations:** pair learning tasks with state manipulations (e.g., stress induction, sleep restriction, adrenergic challenge) to separate GAIN from CTX and CTRL.
3. **Hierarchical modeling:** fit within-person, across-state variation and between-person differences jointly, improving identifiability via repeated measures [15].

> **Interpretation rule:** if two mechanistic stories fit the same data, the correct conclusion is not "both are true," but "the study did not adjudicate them." The point of this framework is to specify what would adjudicate them.

5. Translational mapping without biomarker determinism

A frequent failure mode in translational writing is to map markers directly to interventions as if markers were causal levers. We instead distinguish three roles:

- **Causal leverage:** changing this process is hypothesized to change outcomes.
- **State modulation:** changing this process changes expression/engagement conditions (and thus treatment access), not necessarily the learned threat itself.

- **Correlational marker:** informative about subgrouping or prognosis but not necessarily a lever.

Axis	Operational signature	Primary role	Treatment implication (hypothesis)
CTRL	safety learned in-session, fails under stress	leverage/modulation	pace exposure; add stress-inoculation; reduce interference
CTX	poor discrimination; relapse after context shift	leverage	contextualized exposure; discrimination training; vary contexts
GAIN	hyperarousal, startle, sleep disruption	modulation	stabilize arousal to enable learning; symptom-targeted adjuncts
REC	slow return to baseline after stressors	leverage/modulation	target recovery (sleep, rhythms, stress physiology)

Table 1. Mechanism-space mapping: levers vs modulators vs markers

5.1. Pharmacology example: prazosin as a symptom-conditional modulator

Prazosin illustrates why mechanistic claims must respect heterogeneity. Early RCTs suggested benefit for trauma-related nightmares in certain populations [16], whereas a large multisite trial in veterans did not show benefit on nightmares or sleep quality [17]. Contemporary VA/DoD guidance reflects this nuance: prazosin is not a universal core-PTSD treatment but may be considered for nightmare-focused targets in appropriate contexts [18][19]. Within the present framework, prazosin primarily modulates **GAIN** (expression conditions), which may improve sleep-linked symptom expression and engagement with psychotherapy, without implying it rewrites threat learning.

5.2. Rapid-acting plasticity agents: conditional leverage via coupling to learning

Ketamine shows evidence for rapid symptom reduction in PTSD in RCTs [20][21], with broader meta-analytic syntheses emphasizing heterogeneity and design constraints [22]. The most conservative mechanistic claim is conditional: ketamine may transiently increase network flexibility, which becomes clinically durable primarily when coupled to structured learning or psychotherapy. That is a hypothesis to be tested, not assumed.

6. Discriminative predictions (what would falsify the story)

We present predictions in an adjudication-friendly form: mechanism ⇒ manipulation ⇒ measurable consequence.

1. **CTRL-dominant:** under acute stress induction, regulatory signatures (behavioral control metrics and/or circuit-level control proxies) will degrade disproportionately, predicting relapse despite intact in-session extinction.
2. **CTX-dominant:** context shift will produce renewal-like relapse even when arousal is controlled; discrimination training should preferentially reduce relapse probability.
3. **GAIN-dominant:** arousal metrics (startle/pupil/autonomic indices) will predict fear expression conditional on equivalent learned threat value, and arousal stabilization will increase treatment tolerability and adherence.
4. **REC-dominant:** slow stress recovery will predict symptom volatility and relapse probability over time, beyond baseline severity.

Importantly, failure of any prediction does not "disprove PTSD neurobiology"; it refines which axis dominates in which subgroup, and whether the coarse-graining is adequate.

7. Practical roadmap: what datasets and analyses would make this real

We recommend an explicitly staged validation program:

1. **Stage 1:** demonstrate that CTRL/CTX/GAIN/REC can be estimated with acceptable reliability from tasks + repeated measures (not necessarily fMRI-first).
2. **Stage 2:** show incremental predictive utility for clinically relevant outcomes (dropout, relapse, remission) beyond symptom totals.
3. **Stage 3:** embed mechanism estimates into adaptive trials where augmentation is assigned by dominant axis.

Open resources can accelerate this program, including curated trial repositories and shared fear-learning datasets [23][24].

8. Limitations

This is a minimal, phenomenological framework. It omits important contributors (dissociation subtypes, inflammation, social threat, sleep-memory microstructure) and collapses multiple interacting mechanisms into composite axes. The payoff is adjudicability: clear predictions, explicit uncertainty, and a roadmap for falsification.

9. Conclusion

PTSD is best approached as a heterogeneous disturbance of threat inference and its regulation across contexts and states, constrained by stress physiology and arousal gain. The contribution here is not a new circuit diagram but a disciplined mechanism space that (i) is explicit about what is evidence versus hypothesis, (ii) respects identifiability limits, and (iii) generates discriminative predictions and translational designs without biomarker determinism. If the framework fails, it should fail informatively: by telling us which coarse-grainings were wrong, in which subgroups, and under which perturbations.

Statements and Declarations

Funding

No specific funding was received for this work.

Potential competing interests

No potential competing interests to declare.

Data Availability

This is a conceptual synthesis; no new data were generated.

References

1. ^*Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB (1995). "Posttraumatic Stress Disorder in the National Comorbidity Survey." Arch Gen Psychiatry. **52**:1048–1060.*
2. ^*Koenen KC, Ratanatharathorn A, Ng L, et al. (2017). "Posttraumatic Stress Disorder in the World Mental Health Surveys." Psychol Med. **47**:2260–2274.*

3. a, b *Pitman RK, Rasmusson AM, Koenen KC, et al. (2012). "Biological Studies of Post-Traumatic Stress Disord er." Nat Rev Neurosci. **13**:769–787.*

4. a, b *Shin LM, Liberzon I (2010). "The Neurocircuitry of Fear, Stress, and Anxiety Disorders." Neuropsychopha rmacology. **35**:169–191.*

5. a, b, c *Milad MR, Quirk GJ (2012). "Fear Extinction as a Model for Translational Neuroscience: Ten Years of Pr ogress." Annu Rev Psychol. **63**:129–151.*

6. a, b, c *Bouton ME (2004). "Context and Behavioral Processes in Extinction." Learn Mem. **11**:485–494.*

7. ^ *Logue MW, van Rooij SJH, Dennis EL, et al. (2018). "Smaller Hippocampal Volume in Posttraumatic Stress Disorder: A Multisite ENIGMA–PGC Study." Biol Psychiatry. **83**:244–253.*

8. ^ *O'Doherty DCM, Chitty KM, Saddiqui S, Bennett MR, Lagopoulos J (2015). "A Systematic Review and Meta -Analysis of MRI Measurement of Structural Volumes in PTSD." Psychiatry Res Neuroimaging. **232**:1–33.*

9. ^ *Aston-Jones G, Cohen JD (2005). "An Integrative Theory of Locus Coeruleus–Norepinephrine Function: Ad aptive Gain and Optimal Performance." Annu Rev Neurosci. **28**:403–450.*

10. ^ *Arnsten AFT (2009). "Stress Signalling Pathways That Impair Prefrontal Cortex Structure and Function." Nat Rev Neurosci. **10**:410–422.*

11. ^ *Morris MC, Compas BE, Garber J (2012). "Relations Among Posttraumatic Stress Disorder, Trauma Exposur e, and HPA Axis Function: A Meta-Analysis." Psychoneuroendocrinology. **37**:224–240.*

12. ^ *Schumacher S, Niemeyer H, Engel S, Cwik JC, Knaevelsrud C (2019). "HPA Axis Regulation in PTSD: A Meta -Analysis Focusing on Potential Moderators." Neurosci Biobehav Rev. **100**:35–57.*

13. a, b *Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP (2007). "Universally Sloppy Para meter Sensitivities in Systems Biology Models." PLoS Comput Biol. **3**:e189.*

14. ^ *Chis OT, Banga JR, Balsa-Canto E (2011). "Structural Identifiability of Systems Biology Models: A Critical C omparison of Methods." PLoS One. **6**:e27755.*

15. ^ *Turner BM, Forstmann BU, Love BC, Palmeri TJ, Van Maanen L (2017). "Approaches to Analysis in Model-B ased Cognitive Neuroscience." J Math Psychol. **76**:65–79.*

16. ^ *Raskind MA, Peterson K, Williams T, et al. (2013). "A Trial of Prazosin for Combat Trauma PTSD With Nigh tmares in Active-Duty Soldiers." Am J Psychiatry. **170**:1003–1010.*

17. ^ *Raskind MA, Peskind ER, Chow B, et al. (2018). "Trial of Prazosin for Post-Traumatic Stress Disorder in Mil itary Veterans." N Engl J Med. **378**:507–517.*

18. ^ *Department of Veterans Affairs & Department of Defense (n.d.). "VA/DoD Clinical Practice Guideline for the Management of Posttraumatic Stress Disorder and Acute Stress Disorder." Department of Veterans Affairs & Department of Defense.*
19. ^ *Lang AJ, Hamblen JL, Holtzheimer PE, et al. (2024). "A Clinician's Guide to the 2023 VA/DoD Clinical Practice Guideline for PTSD and Acute Stress Disorder." J Trauma Stress.*
20. ^ *Feder A, Parides MK, Murrough JW, et al. (2014). "Efficacy of Intravenous Ketamine for Treatment of Chronic PTSD: A Randomized Clinical Trial." JAMA Psychiatry. **71**:681–688.*
21. ^ *Feder A, Costi S, Rutter SB, et al. (2021). "A Randomized Controlled Trial of Repeated Ketamine Infusions for Chronic PTSD." Am J Psychiatry. **178**:193–202.*
22. ^ *Du R, Xu Y, Shen L, et al. (2022). "The Multivariate Effect of Ketamine on PTSD: Systematic Review and Meta-Analysis." Front Psychiatry. **13**:813103.*
23. ^ *O'Neil ME, Harik JM, McDonagh M, et al. (2020). "Development of the PTSD-Repository: A Resource for Evidence Synthesis and Trial Tracking." J Trauma Stress. **33**:410–419.*
24. ^ *FearBase Consortium (2026). "FearBase: A Cooperative, Dynamically Growing Database for Human Fear-Conditioning Research." FearBase Consortium. https://fearbase.org. Accessed 2026-02-10.*

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.