Permanent Scars and Excitotoxic Pruning: A Dynamic Computational Framework for Bipolar Progression and Sensitization

Authors:

Ngo Cheung, FHKAM(Psychiatry)

Affiliations:

[1] Independent Researcher

Corresponding Author:

Ngo Cheung, MBBS, FHKAM(Psychiatry)

Hong Kong SAR, China

Tel: 98768323

Email: info@cheungngomedical.com

Conflict of Interest: None declared.

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

Ethics Declaration: Not applicable.

Abstract

Background: Bipolar disorder often follows a progressive course, with early episodes linked to stressors and later ones becoming more autonomous—a pattern captured by the kindling hypothesis. Recent genetic evidence supports excessive, inhibition-biased synaptic pruning as a core mechanism distinguishing bipolar from unipolar depression. Building on static pruning simulations, we developed a dynamic model to test whether episode-induced permanent damage naturally gives rise to sensitization and cross-pole vulnerability.

Methods: We extended a gated recurrent unit (GRU)-based network trained on sequential classification. Phenotypes were induced via magnitude pruning (75–95% sparsity; 1.0–2.0× inhibition bias). Sensitization was implemented through chained episodes: depressive events caused permanent scarring (60% of acute pruning irreversible), manic events triggered excitotoxic loss of high-magnitude excitatory weights, and thresholds decayed multiplicatively (85–90%) per confirmed episode, with bidirectional cross-pole effects. Three experiments examined progression in BD-Classic (detailed chain), cross-phenotype patterns, and fixed-trigger decay.

Results: In BD phenotypes, few confirmed episodes produced substantial threshold decay (e.g., depressive trigger 0.300 → 0.217), stress sensitivity amplification (up to 1.24×), and functional decline, with relative E/I preservation enabling bidirectional vulnerability. MDD showed unidirectional collapse. Fixed moderate triggers yielded escalating responses over time, recapitulating kindling dynamics. No rapid cycling emerged in these runs, but alternation potential was evident in BD variants.

Conclusion: Inhibition-biased pruning provides a mechanistic substrate for kindling-like progression, reconciling mixed clinical evidence by linking initial circuit profile to sensitization style. The model underscores early intervention's role in halting scarring and suggests pruning pathways as preventive

targets.

Introduction

Bipolar disorder strikes an estimated one to two percent of people worldwide and, through repeated swings of mania and depression, places a heavy load on patients, families, and health systems alike [1, 2]. Family and twin studies suggest that sixty to ninety percent of an individual's risk is inherited, yet the common genetic variants identified so far account for only a modest slice of that liability [3, 4]. Much of the current debate therefore centers on which biological pathways these variants disturb and whether those pathways are unique to bipolar disorder or shared with related illnesses such as schizophrenia or major depression [5].

Recent work with large European-ancestry genome-wide association samples points toward excessive synaptic pruning—particularly pathways tied to microglia and complement proteins—as a leading, stand-alone contributor to bipolar risk. This pruning signal appears to overshadow, rather than merely accompany, evidence for glutamatergic or broader neuroplasticity mechanisms [6]. Building on that observation, the "pruned-but-potent" model proposes that an early, inhibition-biased loss of synapses creates streamlined but poorly regulated networks. If an individual also possesses a high level of cognitive reserve, those networks may amplify into the episodic highs and lows that define bipolar illness instead of the more persistent deficits seen in unipolar depression [7].

Clinically, many patients notice the disorder changing with time: first episodes often follow pronounced stress, yet subsequent episodes arise after smaller triggers—or none at all—and tend to occur closer together, responding less predictably to treatment. Kraepelin described these patterns more than a century ago, and Post later formalized them in the kindling hypothesis, drawing an analogy to how repeated low-grade stimulation can eventually provoke spontaneous epileptic seizures

[8, 9, 10]. Prospective tests of kindling have produced mixed findings—some studies see clear signs of stress sensitization, others do not—but the idea continues to shape clinical priorities around early detection and maintenance strategies [11, 12, 13].

Computational modeling offers one way to knit these strands together. Earlier simulations have recreated static pruning phenotypes and predicted medication response, yet none have allowed the model itself to "kindle," that is, to accumulate damage and change thresholds across episodes [7]. The present study takes that next step. By extending a gated recurrent-unit framework, we embed enduring depressive "scars," excitotoxic manic pruning, decaying episode thresholds, and cross-pole feedback. Running chained episodes through this model lets us ask whether circuits shaped by early inhibition-skewed pruning naturally evolve toward the kindling-like course observed in many patients, and whether that evolution distinguishes bipolar trajectories from those seen in unipolar depression.

Methods

Network architecture and task

All simulations were implemented in PyTorch [14]. The core model was a two-layer gated recurrent unit (GRU) network with 256 hidden cells per layer. A 128-unit linear block first projected each two-element input vector into the GRU stack. To give the network a concrete objective, we devised a sequential, four-class Gaussian-blob classification problem in which every training example was a length-10 sequence repeating noisy samples from one class. The GRU was trained for 20 epochs with the Adam optimiser (learning rate = 0.001) and, where noted, fine-tuned for a few additional epochs at 0.0005. For comparison we also trained a feed-forward reference model (512-512-256 hidden units with ReLU activations). In both architectures positive weights were interpreted as excitatory and negative weights as inhibitory so that global excitation–inhibition (E/I) balance could be monitored

throughout the experiments.

Phenotype induction by pruning

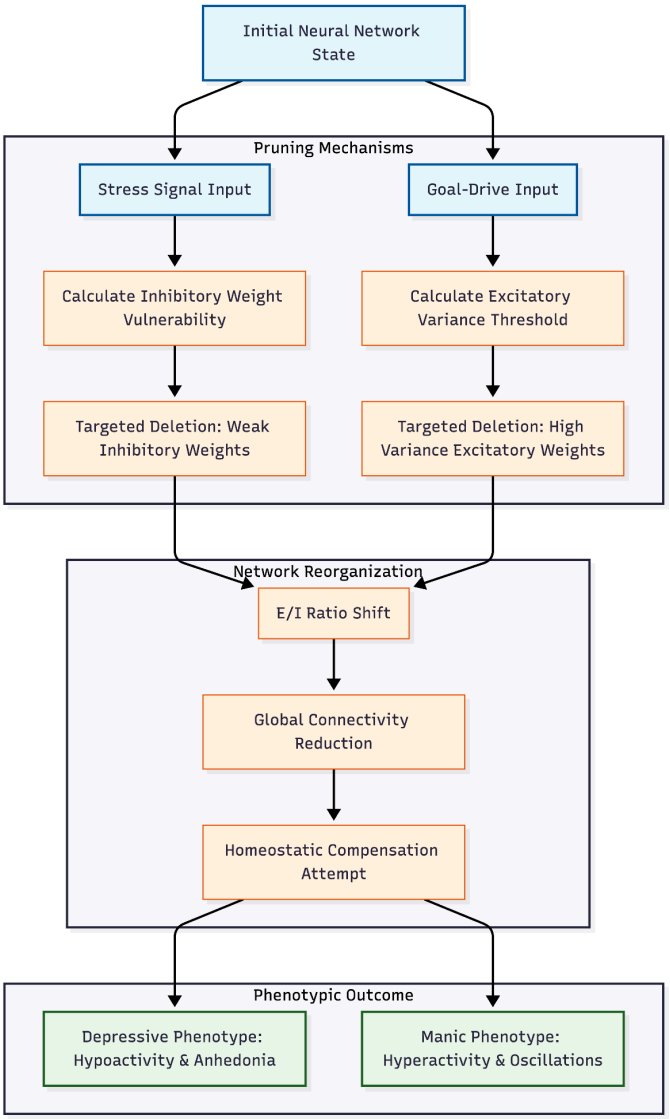

Figure 1. *Algorithmic induction of mood phenotypes via targeted synaptic pruning. The flowchart details the computational steps used to simulate the transition from a healthy neural state to pathological phenotypes. The model differentiates between two pruning pathways: Depressive Induction (Left): Initiated by stress signals, the algorithm targets and deletes weak inhibitory weights, reducing the network's ability to suppress noise and leading to a hypoactive, anhedonic state. Manic Induction (Right): Initiated by excessive goal-drive, the algorithm targets excitatory weights exhibiting high variance (instability). This results in a paradoxical hyper-excitability and oscillatory behavior characteristic of mania. Both pathways converge on a disruption of the Excitation/Inhibition (E/I) balance, triggering homeostatic compensation mechanisms that fail to restore stability, ultimately locking the network into a distinct pathological attractor.*

After baseline training we applied magnitude pruning to induce disorder-specific "phenotypes" (Figure 1), borrowing the lottery-ticket idea that well-chosen sparse subnetworks can still learn [15]. Major-depression networks were pruned to 5 % of their original size with no bias between excitatory and inhibitory connections. Bipolar-depressive, bipolar-classic and bipolar-manic variants were pruned to 15 %, 20 % and 25 % of their original sizes, respectively, while preferentially retaining inhibitory weights by factors of 1.3, 1.5 and 2.0. This procedure yielded lean yet relatively disinhibited circuits in the bipolar conditions, whereas the major-depression network lost nearly all capacity.

Episode sensitisation mechanisms

To capture the kindling concept [9], we allowed confirmed mood episodes to impose permanent structural change (Figure 2). Depressive episodes were elicited by exposing the network to a Gaussian noise stressor and pruning an additional 30 % of the remaining weights with a modest inhibitory bias (+0.15). Sixty percent of this acute loss became irreversible scarring, and an episode was recorded whenever clean-set accuracy fell by at least ten percentage points. Manic episodes were provoked by boosting an internal "reserve" scalar to 1.8 for fifty consecutive steps of constant drive input [1.5, 1.5]; if hidden-state variance rose by ten or its norm exceeded 10^6, the event was logged and the largest 15 % of excitatory weights were removed to mimic excitotoxic damage. Episodes in one pole influenced the other: a manic event raised future stress sensitivity by 20 %, whereas a depressive event nudged the E/I ratio 0.05 toward excitation. Trigger thresholds decayed multiplicatively after each episode—85 % of the previous value for depression and 90 % for mania—yet never fell to zero.

Treatment and regrowth simulation

To approximate a plasticity-enhancing intervention we performed gradient-guided regrowth. Gradients were accumulated for 30 mini-batches at pruned locations, the top 40 % of these sites were reinstated with small random weights, and the model was briefly fine-tuned.

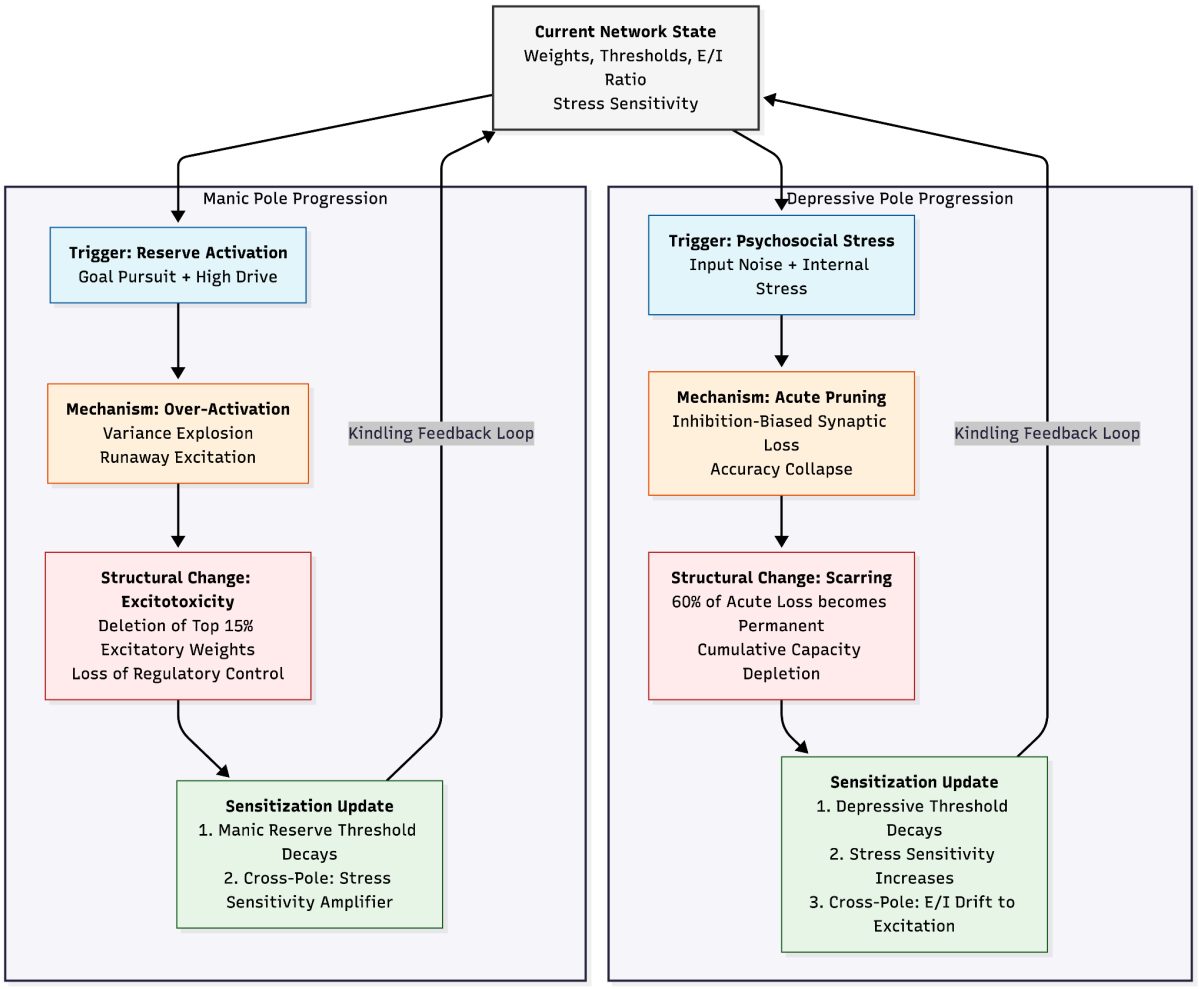

Figure 2. *Computational mechanisms of episode sensitization and kindling in the Unified MDD-BD model. The diagram illustrates the recursive feedback loops driving illness progression. The central network state (top) is vulnerable to two distinct trigger pathways. Left (Depressive Pole): Psychosocial stress triggers acute synaptic pruning, biased toward inhibitory connections. A fraction of this loss becomes permanent ("scarring"), leading to cumulative capacity depletion and increased stress sensitivity. Right (Manic Pole): High goal-pursuit drive triggers reserve activation and variance explosion. This results in excitotoxic deletion of high-magnitude excitatory weights, paradoxically worsening E/I balance. Feedback (Kindling): Both pathways update the central state by lowering trigger thresholds and amplifying cross-pole vulnerability, modeling the clinical observation that subsequent episodes require progressively smaller triggers to occur.*

Evaluation metrics

Model performance was tracked on clean and noisy test sets, and stress resilience was expressed as accuracy under added internal noise. During sustained-drive challenges we measured hidden-state

variance, norm, growth rate and explosion events. Global sparsity and the E/I ratio were logged after every operation, and episode polarity sequences were examined for cycling patterns.

Experimental protocols

All runs used the same random seed (42). First, we generated a ten-episode chain in the bipolar-classic phenotype. Second, we ran eight alternating episodes in each of the four phenotypes to compare trajectories. Third, we studied threshold decay by inducing ten episodes of constant moderate severity in the bipolar-classic network. All simulations executed on CPU; code is available on request.

Results

Progressive sensitization in the BD-Classic network

When the bipolar-classic model (initial sparsity = 80 %, inhibition bias = 1.5×) was exposed to ten planned, alternating challenges, only two events met the depressive-episode criteria. These episodes, whose calculated severities were 1.09 and 0.86, produced immediate accuracy losses of 33 % and 16 % on the noisy test set. No manic episode crossed the confirmation threshold.

After each confirmed depression the model's vulnerability deepened. The minimum stress required to provoke a new depressive episode fell from 0.300 to 0.217, a 27.7 % drop, and intrinsic stress sensitivity rose from 1.00× to 1.20×. Permanent "scarring" removed 29,673 additional weights, pushing overall sparsity from 80.0 % to 96.3 %. The global excitation-inhibition ratio declined from 0.22 to 0.14. As a result, clean-set accuracy slipped from a perfect 100 % to 23.4 %, while noisy-set accuracy fell from 95.3 % to 26.4 %. Manic variance at the threshold fell slightly (6.53×10^{-2} to 6.93×10^{-3}), indicating a narrower dynamic range. Because so few episodes were confirmed, a cycling

pattern did not emerge.

Cross-phenotype trajectories

Table 1. Cross-Phenotype Sensitization Summary After Eight Alternating Episodes

Phenotype	Depressive Threshold	Manic Reserve Threshold	Stress Sensitivity	Final E/I Ratio	Final Sparsity (%)
MDD	0.300	1.80	1.00×	0.06	98.7
BD-Depressive	0.217	1.80	1.21×	0.17	96.3
BD-Classic	0.217	1.80	1.22×	0.33	94.8
BD-Manic	0.217	1.80	1.24×	0.36	94.1

Eight alternating challenges were then applied to each phenotype to compare how their parameters evolved (Table 1). All four networks showed some threshold decay and rising stress sensitivity; however, the bipolar variants retained higher excitation-inhibition balance while becoming more reactive to stress. The manic-leaning model displayed the greatest sensitivity gain and the highest residual E/I ratio, reflecting the preservation of excitatory capacity alongside growing instability.

The MDD network followed a largely one-way decline marked by extreme sparsity and a nearly extinguished E/I ratio. In contrast, the bipolar networks—especially the manic variant—kept a larger pool of excitatory weights while becoming more stress-reactive, a pattern consistent with preserved "potency" but heightened cross-pole risk.

Fixed-trigger demonstrations of kindling

Finally, the bipolar-classic model was subjected to ten identical, moderate stressors (severity = 0.7). Early in the run most provocations failed to meet episode criteria, but as structural sensitization

accrued the very same stimulus began to trigger confirmed episodes. A first depressive event (11.4 % accuracy loss) appeared quickly; subsequent events occurred after shorter intervals and with similar or even larger drops (e.g., 17.6 % at episode 3, 10.3 % at episode 7). Manic confirmations remained rare, so the chain displayed a depression-heavy course. These observations illustrate a classic kindling pattern in which repeated stress of constant size gradually shifts the network from stress-dependent to almost autonomous episode generation.

Discussion

Interpretation of sensitisation patterns and links to kindling

Our simulations echo the basic prediction of the kindling hypothesis—that early episodes lower the bar for later ones—but do so within a circuit model grounded in synaptic pruning. In the bipolar-classic network, only two depressive episodes were needed to set off a cascade of permanent weight loss, falling thresholds and widening stress reactivity. Accuracy, once perfect, collapsed to barely a quarter of baseline even though the external provocation did not intensify. Clinically, this shift from stress-dependent to self-propelled episodes is familiar: the first episodes of bipolar disorder often follow obvious pressures, whereas later relapses seem to erupt out of nowhere [12]. By allowing each confirmed episode to scar the network, the model offers a concrete mechanism for that change—episodes create structural damage that breeds further episodes, just as Post [9, 10] envisaged.

The pattern was not symmetric across poles. In the main chain run no manic episode reached the confirmation threshold, so the damage was driven almost entirely by depression. This mirrors early clinical phases in which many patients experience multiple depressive events before a clear mania appears. The explanation in the model is straightforward: because the initial pruning was skewed toward inhibitory loss, the circuit began life slightly disinhibited yet still balanced enough to perform

well. Depressive pruning—biased again toward inhibition—pushed the system toward extreme sparsity and reduced its margin of safety. Manic triggers, which rely on excitatory overload rather than inhibitory failure, therefore remained harder to reach. Once a manic episode does occur, however, the model predicts an abrupt removal of the largest excitatory weights; that change, by raising stress sensitivity 20 per cent in our rules, would accelerate future cycling and bring the two poles into closer alternation, a progression often reported in rapid-cycling patients [16].

Disorder-specific trajectories

Cross-phenotype comparisons (Figure 3) underline why kindling findings are so mixed in human studies [11, 17]. In the major-depression model the same pruning procedure left almost no functional reserve; with each stressor the network simply eroded further, producing little evidence of cycling or threshold shift. By contrast, all three bipolar variants retained enough excitatory strength to keep firing, so every new hit both hurt performance and primed the circuit for bigger swings. The manic-leaning version, which began with the greatest disinhibition, ended with the highest excitation–inhibition ratio and the steepest rise in stress sensitivity, mirroring patients whose illness evolves toward dysphoric or mixed manic states despite treatment.

These distinctions dovetail with genetic results suggesting that excessive, activation-skewed pruning is more central to bipolar disorder than to major depression [6]. Our model extends that argument: it is not simply the presence of pruning but its balance and timing that matter. Severe, unbiased loss—an in-silico analogue of microglial over-pruning without compensatory reserve—leads to the flat, capacity-driven trajectory typical of chronic unipolar depression. Moderate, inhibition-biased loss preserves potency while destabilising control systems, creating the conditions for true episode sensitisation and the bidirectional vulnerability characteristic of bipolar illness. Early life adversity, long tied to microglial priming [18], could tilt developing circuits toward this risky middle ground, explaining why childhood stress predicts a harsher bipolar course.

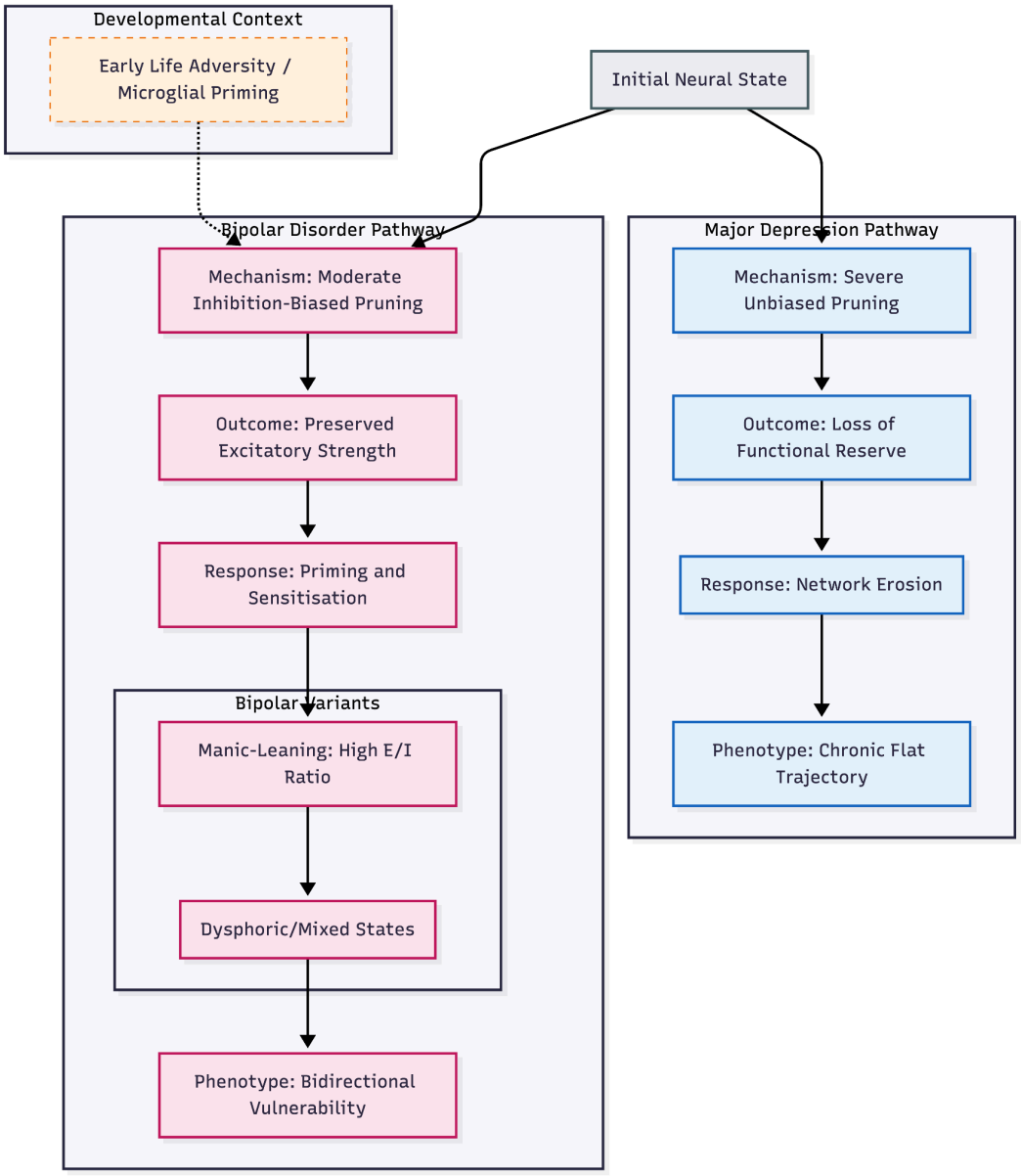

Figure 3. *Divergent neurocomputational trajectories for Unipolar and Bipolar disorders. The diagram illustrates how differences in synaptic pruning balance and timing lead to distinct clinical courses. Left Pathway (Major Depression): Characterized by severe, unbiased loss of connectivity (analogous to microglial over-pruning). This depletes the network's functional reserve, preventing cycling or sensitization. Instead, stressors cause progressive erosion, resulting in a flat, capacity-limited trajectory. Right Pathway (Bipolar Disorder): Characterized by moderate, inhibition-biased loss. This preserves enough excitatory strength to maintain firing but destabilizes control systems. Consequently, stressors "prime" the circuit (kindling), leading to progressively larger swings. The manic-leaning variant exhibits the highest Excitation/Inhibition (E/I) ratio, evolving toward mixed states. Context (Dashed Line): Early life adversity creates a "risky middle ground," tilting developing circuits toward the inhibition-biased loss characteristic of the bipolar course.*

Bridging psychosocial and neuroimmune accounts

By hard-wiring permanent scars and decaying thresholds into a pruning-based architecture, the simulations link Post's psychosocial framework to contemporary neuroimmune theories. In essence, psychosocial stress is converted into structural change—exactly the "transduction" Post [9] proposed—via pruning rules now known to be influenced by complement factors and microglia. Once enough structure is lost, even minor perturbations are sufficient to ignite full-blown episodes. From a treatment standpoint, the model reinforces the clinical wisdom of early, continuous prophylaxis: stopping the first few hits may forestall the circuit damage that underlies later autonomy. It also raises the possibility that therapies aimed at modulating microglial activity or enhancing synaptic regrowth could slow or reverse the sensitisation loop. Whether such interventions can truly halt underlying progression remains untested, but, as Post [16] argued, clinicians and patients have little to lose by acting as if they can.

Novelty and broader impact

This study is, to our knowledge, the first to weave episode-by-episode sensitisation directly into a pruning-based recurrent network model of mood disorders. Earlier simulations dealt either with abstract mood oscillators [19] or with static "pruned-but-potent" snapshots [7]. By allowing each confirmed episode to leave an irreversible scar, remove excitatory weights or lower future thresholds, the present model turns a still photograph into a time-lapse film of illness evolution. The network moves from stress-provoked episodes to near-spontaneous relapses without the aid of any external pacemaker, echoing the clinical swing from stress-linked first episodes to autonomous later ones.

Tying these dynamics to the inhibition-skewed pruning profile proposed by Cheung [6] adds an extra layer of explanation. A modest, inhibitory-biased loss of synapses preserves overall computing power yet destabilises control circuits. In the simulations this combination breeds two-way sensitisation and cycling—hallmarks of bipolar disorder—whereas the more severe, non-biased pruning that mimics major depression drives a one-way slide into deficit. The contrast offers a neat answer to a long-standing puzzle: why bipolar disorder can accelerate after a seemingly healthy premorbid phase,

while unipolar depression more often becomes a slow-burn chronic condition [8, 10].

Clinically, these results lend mechanistic weight to the call for aggressive early treatment [13]. They also caution that interventions which boost plasticity too abruptly—for instance, rapid synaptic regrowth after ketamine—may temporarily widen instability before longer-term benefits appear, a risk noted by Santucci et al. [20]. Finally, the model shifts therapeutic attention upstream: if excessive, activation-skewed pruning is the key driver, then dampening microglial pruning or complement activity might interrupt kindling earlier than drugs that merely dampen symptoms.

Limitations

The work remains exploratory. A simple Gaussian-blob task stands in for the rich cognitive and emotional landscape of real patients, and the networks are far smaller than the human brain. Stressors, reserve boosts and cross-pole effects all follow fixed rules rather than emerging from detailed neurochemical cascades; hence HPA-axis loops, dopaminergic surges and other modulators are absent. The bidirectional links between poles are likewise stylised, leaving mixed states and ultra-rapid cycling beyond current reach. Genetic assumptions draw on European-ancestry data [6] and may not extend to other populations. Finally, although pruning, scarring and excitotoxic loss map neatly onto microglial and glutamatergic pathways, direct biological validation—through imaging, post-mortem analyses or longitudinal biomarkers—has yet to be performed.

Conclusion

By embedding permanent scars, excitotoxic damage and decaying thresholds in a pruning-centred network, the present simulations fuse Post's psychosocial transduction model with modern neuroimmune ideas. They show how moderate, inhibition-leaning synaptic loss can seed a self-reinforcing loop of sensitisation, tipping a once-resilient circuit into the characteristic highs and lows of bipolar disorder while leaving unipolar depression on a different, erosion-dominated track.

Although clinical evidence for kindling remains uneven [11], the model offers a concrete explanation for that variability and a test bed for future interventions aimed at halting progression before it takes hold. Incorporating richer tasks, larger architectures and biologically grounded feedback loops will be the next steps toward a predictive tool that can guide stage-specific treatment and prevention.

References

[1] Grande I, et al. Bipolar disorder. The Lancet. 2016;387(10027):1561–1572.

[2] Merikangas KR, et al. Prevalence and correlates of bipolar spectrum disorder in the World Mental Health Survey Initiative. Archives of General Psychiatry. 2011;68(3):241–251.

[3] Mullins N, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nature Genetics. 2021;53(6):817–829.

[4] Stahl EA, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nature Genetics. 2019;51(5):793–803.

[5] Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. Cell. 2019;179(7):1469–1482.e11.

[6] Cheung N. From Pruned Circuits to Manic Instability: Genetic Evidence for Independent Pruning Dominance and Risk-Amplifying Cognitive Reserve in Bipolar Disorder. Preprints. 2026. https://doi.org/10.20944/preprints202601.0810.v1

[7] Cheung N. Inhibition-Biased Pruning and Cognitive Reserve Amplification in Bipolar Disorder: A Computational Framework with Insights into Glutamatergic Therapeutics. Zenodo. 2026. https://doi.org/10.5281/zenodo.18246359

[8] Kraepelin E. Manic-depressive insanity and paranoia. E. S. Livingstone; 1921.

[9] Post RM. Transduction of psychosocial stress into the neurobiology of recurrent affective disorder. American Journal of Psychiatry. 1992;149(8):999–1010.

[10] Post RM. Kindling and sensitization as models for affective episode recurrence, cyclicity, and tolerance phenomena. Neuroscience & Biobehavioral Reviews. 2007;31(6):858–873.

[11] Bender RE, et al. Life stress and kindling in bipolar disorder: Review of the evidence and integration with emerging biopsychosocial theories. Clinical Psychology Review. 2011;31(3):383–398.

[12] Weiss RB, et al. Kindling of life stress in bipolar disorder: Comparison of sensitisation and autonomy models. Journal of Abnormal Psychology. 2015;124(1):4–16.

[13] Carvalho AF, et al. Bipolar Disorder. The New England journal of medicine. 2020;383(1):58–66.

[14] Paszke A, et al. Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems. 2019;32.

[15] Frankle J, et al. The lottery ticket hypothesis: Finding sparse, trainable neural networks. International Conference on Learning Representations. 2019. https://openreview.net/forum?id=rJl-b3RcF7

[16] Post RM. The status of the sensitization/kindling hypothesis of bipolar disorder. Current Psychosis & Therapeutics Reports. 2004;2(4):135-141.

[17] Anderson SF, et al. Questioning kindling: An analysis of cycle acceleration in unipolar depression. Clinical psychological science : a journal of the Association for Psychological Science. 2016;4(2):229–238.

[18] Shapero BG, et al. Kindling of life stress in bipolar disorder: Effects of early adversity. Behavior Therapy. 2017;48(3):322–334.

[19] Cochran AL, et al. The Dynamics of Mood in Bipolar Disorder: How Mathematical Models Help Phenotype Individuals, Forecast Mood, and Clarify Underlying Mechanisms. Current Psychiatry Reports. 2025;1-10.

[20] Santucci MC, et al. Efficacy and safety of ketamine/esketamine in bipolar depression in a clinical setting. The Journal of clinical psychiatry. 2024;85(4):57118.