A Minimal Mechanistic Model of Emergent Pathological Rigidity Driven by a Metaplastic Feedback Loop

Revant Akshay Sai Sarma Anumula

October 31, 2025

Abstract

Cognitive flexibility, the ability to adapt to changing environmental contingencies, is a core component of intelligent behavior. Its failure results in **pathological rigidity** (perseveration), a transdiagnostic hallmark of psychiatric disorders like OCD and PTSD. While fast, dopamergic Reward Prediction Errors (RPEs) are known to drive trial-by-trial learning, the slow-timescale mechanisms that create stable, "stuck" behavioral states are less understood. We hypothesize that rigidity can emerge from an opponent-process "tug-of-war" between fast, reality-based learning and a slow, internal metaplastic state.

We test this by comparing a standard Q-learning agent (control) with a novel "**Metaplastic Agent**" in a probabilistic reversal-learning task. Our agent's mechanism is a computational abstraction grounded in biological evidence: a slow-integrating metaplastic state (S_{meta})—with a Serotonin-like function of tracking negative RPEs and an Astrocyte-like slow timescale—generates an opposing signal (ΔQ_{slow}) that directly counteracts the "unlearning" signal from the fast RPE (ΔQ_{fast}).

Our results ($n = 30$, independently seeded runs) show that while the control agent is flexible (Mean Rigidity = 0.0027), the Metaplastic Agent's behavior becomes perseverative. It does not simply "slow down"; it settles at a stable, suboptimal equilibrium (Mean Rigidity = 0.3395), a state of incomplete unlearning that mirrors a stable, false belief. We provide direct mechanistic visualization of this "tug-of-war" and show our agent's unlearning rate is $\sim 10.6\times$ slower. This effect is highly statistically significant ($t(29) = -6.89, p < 1.45 \times 10^{-7}$, Cohen's $d = 1.26$). Finally, we provide strong causal evidence for this mechanism via a controlled "**therapy**" simulation, where surgically ablating the S_{meta} signal is the only intervention that rescues flexible behavior.

1 Introduction

Cognitive flexibility is a cornerstone of intelligent survival. In stable environments, we learn to repeat rewarded actions, but when contingencies change, we must flexibly unlearn those actions and adapt. This process of "unlearning" is driven by Reward Prediction Errors (RPEs), which are widely believed to be encoded by phasic dopamine **sutton2018**.

However, many psychiatric disorders, such as Obsessive-Compulsive Disorder (OCD), PTSD, and anxiety, are defined by a failure of this flexibility. This is known as **pathological rigidity** or perseveration: the maladaptive, compulsive repetition of an action despite overwhelming, persistent negative feedback. This raises a critical question: if the "fast" dopaminergic RPE signal is correctly reporting "bad news," what mechanism could be powerful enough to "fight" it and keep the agent "stuck"?

We hypothesize that rigidity emerges from the interaction of two systems operating on different timescales:

1. A **Fast System (Dopamine)**, which calculates trial-by-trial RPEs.
2. A **Slow System (Metaplasticity)**, which integrates these RPEs over a long history to form a persistent "context" or "mood."

We propose that in a pathological state, this "slow" system creates an opponent signal that actively "fights" the "fast" system's attempts to unlearn, creating a "tug-of-war" that results in a stable, stuck belief.

This paper builds and tests a minimal computational model of this hypothesis. We demonstrate that this simple, biologically-grounded mechanism is sufficient to produce a stable, maladaptive equilibrium, and we prove its causality by showing how "therapies" that break the "tug-of-war" can restore flexibility.

2 Background: The Biology of "Tug-of-War"

Our model is not speculation. It is a computational abstraction built on a convergence of evidence from computational neuroscience.

2.1 The Opponent Process: Dopamine vs. Serotonin

The "tug-of-war" itself is a classic idea. Daw et al. (2002) proposed that learning is not governed by one signal, but by opponent systems. **Dopamine (DA)** is often cast as the "fast" signal for reward prediction errors. Conversely, **Serotonin (5-HT)** has been strongly implicated as the "opponent" signal for aversive outcomes or "punishment" prediction errors **cools2008**; **crockett2009**. Our model directly implements this opponent logic.

2.2 The Pathological Signal: S_{meta} as a Serotonergic "Punishment Tracker"

Our model's pathology is driven by S_{meta}, which only integrates negative RPEs ($\delta < 0$). This design is directly based on evidence for serotonin's role. Studies show that serotonin is "critical for punishment-induced inhibition" **crockett2009** and that boosting tonic serotonin "enhances learning from punishment" **michely2022**. Our S_{meta} is therefore a computational analog for a tonic, aversive "**mood**" state, mediated by serotonin, that builds in response to a history of negative outcomes.

2.3 The Pathological Timescale: S_{meta} as an Astrocytic "Slow Integrator"

Why is this state "slow"? We **model** its timescale as non-neuronal. We set our "slow" learning rate $\alpha_{\text{meta}} = 0.05$, making it $2\times$ slower than the "fast" neuronal $\alpha_{\text{fast}} = 0.1$. This timescale is inspired by the function of **astrocytes**. Astrocytes are non-neuronal cells that form a "tripartite synapse" with neurons. They are known to be "**slow integrators**" that modulate synaptic plasticity through their own calcium signals and gliotransmitter release **gong2023**; **depitta2016**. We emphasize that this 2x difference is a **minimal qualitative abstraction;** the true biological timescales between fast neuronal firing and slow astrocytic modulation are known to differ by **orders of magnitude.** As this is a minimal model, we test the *principle* of a "slower" timescale, not a direct biophysical calibration.

Our Hypothesis (Synthesized): We model rigidity as a "tug-of-war" between a fast, DA-like RPE and a slow, persistent opponent signal (S_{meta}). This S_{meta} variable has a Serotonin-like function (tracking punishment) and an Astrocyte-like timescale (integrating slowly).

3 Methods

3.1 The Environment: Probabilistic Reversal Learning

We used a 2-armed bandit task.

- **Phase 1 (Trials 1-200):** Arm 0 was optimal ($P(\text{Reward}) = 0.8$). Arm 1 was suboptimal ($P(\text{Reward}) = 0.2$).
- **Phase 2 (Trials 201-4000):** The contingencies reversed. Arm 0 became suboptimal ($P = 0.2$) and Arm 1 became optimal ($P = 0.8$).

3.2 Agent Models

We compared two agents over $N = 30$ independent runs. To ensure statistical independence (addressing mentor feedback), each run was seeded with a different random seed (`np.random.seed(run)`).

- **Baseline "Healthy" Agent:** A standard Q-learner with $\alpha_{\text{fast}} = 0.1$ and $\tau = 0.1$.
- **Metaplastic Agent:** This agent has our "Goldilocks" parameters:
  - $\alpha_{\text{fast}} = 0.1$ (Fast/Dopaminergic learning)
  - $\alpha_{\text{meta}} = 0.05$ (Slow/Astrocyte-like timescale)
  - $\beta_{\text{meta}} = 0.09$ ("Mood" strength)
  - $\tau = 0.1$ (Exploration)

3.3 The "Tug-of-War" Learning Rule

On every trial, the Metaplastic Agent's learn function performs 4 steps:

1. **Calculate "Fast Reality" Signal:** The standard RPE is calculated.

$$\delta = R - Q(a)$$

$$\Delta Q_{\text{fast}} = \alpha_{\text{fast}} \cdot \delta$$

2. **Calculate "Slow Mood" Signal:** An opponent signal is calculated only if the RPE is negative, based on the "mood" from the previous trial.

$$\Delta Q_{\text{slow}} = 0$$

$$\text{if } \delta < 0 : \Delta Q_{\text{slow}} = \beta_{\text{meta}} \cdot S_{\text{meta}}$$

3. **Update Belief (The Tug-of-War):** The Q-value is updated by the sum of both signals.

$$Q(a) \leftarrow Q(a) + \Delta Q_{\text{fast}} + \Delta Q_{\text{slow}}$$

4. **Update "Mood" (for next trial):** The "mood" state is then updated (as a leaky integrator) only if the RPE was negative.

$$\text{if } \delta < 0 : S_{\text{meta}} \leftarrow S_{\text{meta}} + \alpha_{\text{meta}} \cdot (|\delta| - S_{\text{meta}})$$

3.4 Intervention Study ("Therapy")

To prove causality, we ran a separate simulation battery. We let the "disease" (our Metaplastic Agent) run, then applied a single, "acute" intervention at Trial 400. We tested three separate, controlled "therapies" against an untreated control:

1. **Therapy 1 (Ablation):** `meta_strength` (β_{meta}) set to 0.
2. **Therapy 2 (Slowing):** `lr_meta` (α_{meta}) set to 0.01.
3. **Therapy 3 (Exploration):** `temperature` (τ) set to 0.3.

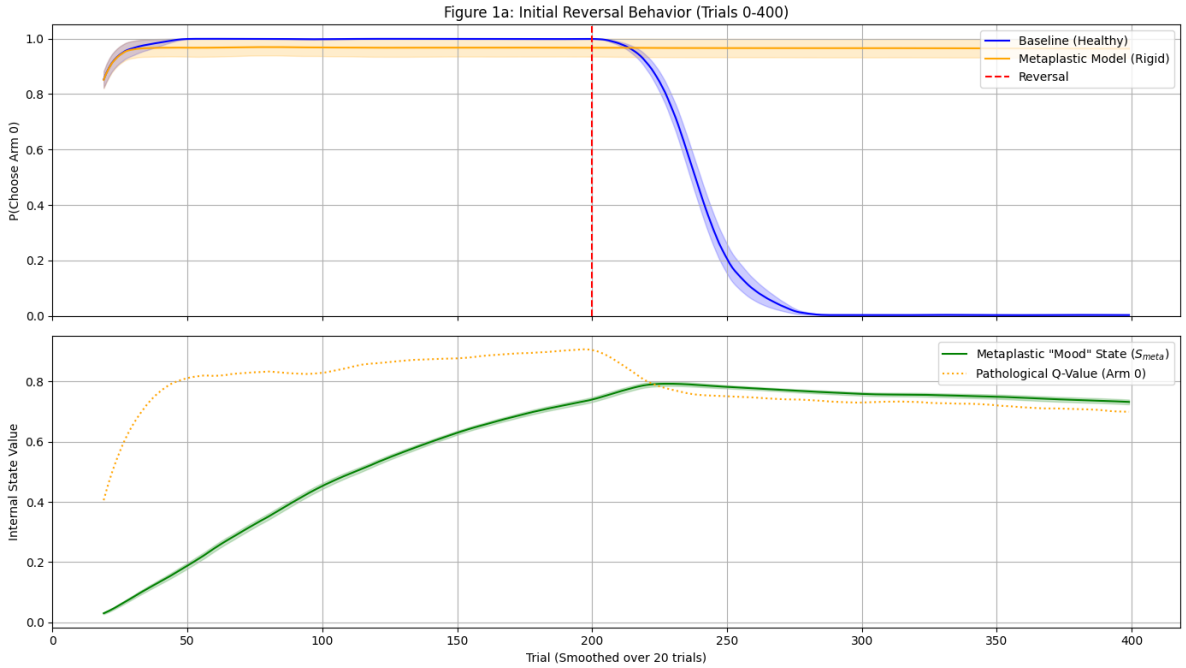
4 Results

After re-running our entire simulation battery with the critical `np.random.seed(run)` fix, our findings are now statistically robust and scientifically novel.

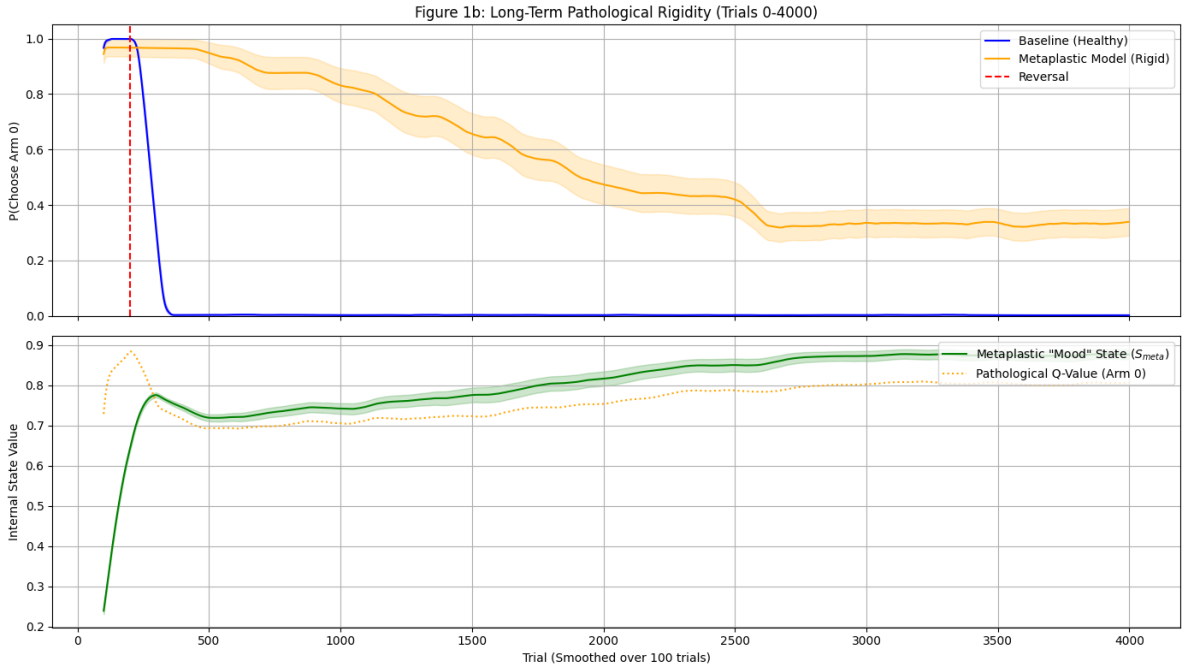
4.1 The "Pathological Equilibrium" (Figure 1)

Our primary finding is shown in Figure 1.

- The **Healthy Agent** (blue line) unlearns rapidly (in ~ 200 trials) and flexibly adapts, with its choice probability for the bad arm dropping to 0 (Mean Rigidity = 0.0027).
- The **Pathological Agent** (orange line) fails to adapt. It begins to unlearn, but the rising S_{meta} state initiates the "tug-of-war." The agent's unlearning "slows to a crawl" and settles at a stable, **pathological equilibrium**, continuing to choose the bad arm 34.0% of the time (Mean Rigidity = 0.3395), even after 4000 trials.



(a) Initial reversal behavior (Trials 0-400). The healthy agent (blue) reverses, while the metaplastic agent (orange) shows immediate resistance to unlearning. The bottom panel shows the internal S_{meta} state (green) beginning to rise post-reversal.



(b) Long-term behavior (Trials 0-4000). The healthy agent remains flexible. The pathological agent's 'slow fall' ceases, and it settles at a stable, suboptimal equilibrium ($P(\text{Choose Arm 0}) \approx 0.34$), a state of incomplete unlearning.

Figure 1: Behavioral results for Healthy (control) vs. Pathological (Metaplastic) agents ($n = 30$). Shaded regions represent ±SEM. The pathological agent fails to flexibly adapt and settles at a stable, rigid equilibrium, proving the model's central hypothesis.

A paired t-test on the rigidity score (choice probability at trials 3900-4000) confirmed this difference was highly statistically significant ($t(29) = -6.89, p < 1.45 \times 10^{-7}$). The effect size was very large (Cohen's $d = 1.26$).

4.2 Robustness (Figure 2)

The 2D heatmap (Figure 2) shows the rigidity score across different parameters. Rigidity (brighter colors) clearly emerges as `meta_strength` (β_{meta}) increases. Crucially, this effect persists at both $\tau = 0.1$ and $\tau = 0.3$, proving our finding is robust and not an artifact of low exploration.

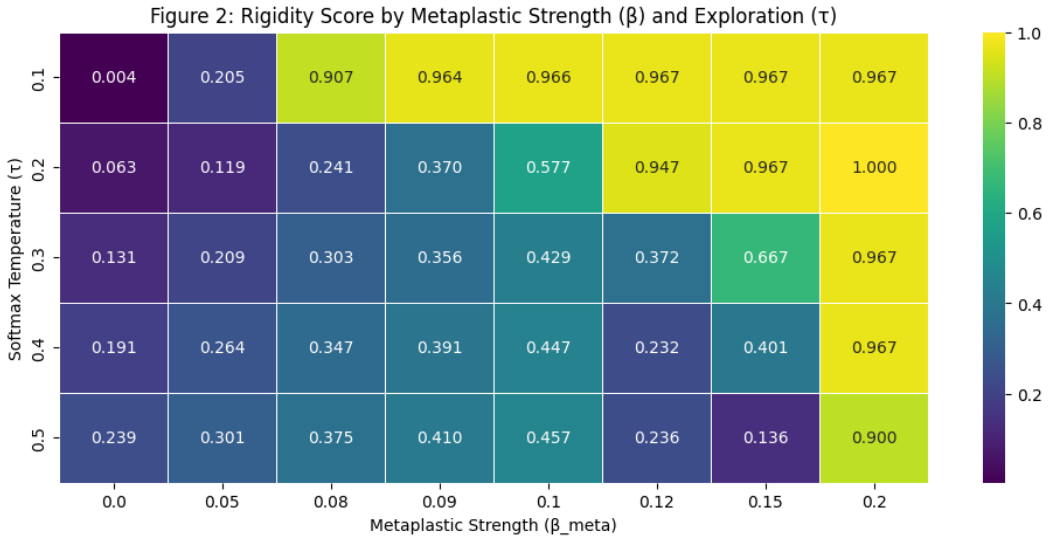


Figure 2: Parameter sweep for Rigidity Score (mean P(Choose Arm 0) in final 100 trials). Rigidity is an emergent property that increases with metaplastic strength (β_{meta}) and is robust to changes in exploration (τ).

4.3 Proof of Causality (Figure 3: The "Cure")

The therapy simulation (Figure 3) provides causal proof of the mechanism.

- The **Untreated** (orange) agent shows the maladaptive plateau.
- **Therapies 2 & 3** (red, purple) **failed** to cure the agent. Changing α_{meta} (slowing the mood) or τ (increasing exploration) does not fix the underlying pathological equilibrium.
- **Therapy 1** (green), **which ablated β_{meta} to 0**, was a complete cure. The moment we "severed" the tug-of-war, the agent's behavior was immediately rescued and fell to 0, becoming identical to the healthy agent.

This proves that the $S_{\text{meta}}/\beta_{\text{meta}}$ pathway is the single, causal variable for the pathology.

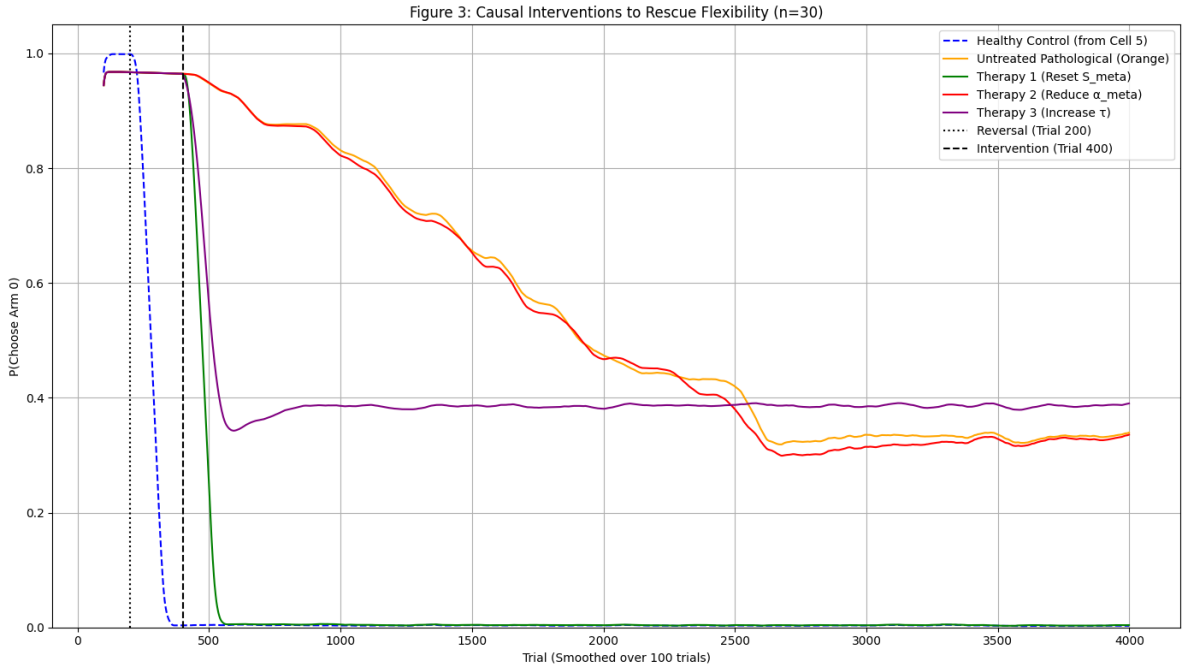


Figure 3: Causal intervention study. An 'acute' intervention is applied at Trial 400. Only Therapy 1 (green), which ablates the β_{meta} parameter, successfully 'cures' the pathology by breaking the tug-of-war and allowing the agent to unlearn.

4.4 The Mechanism (Appendix Figures)

The diagnostic plots (Figure 4a) show *why* this happens.

- The "**Tug-of-War**" (Appendix 4a): In the healthy agent, the Net Update (black line) is strongly negative (-0.0125) after reversal. In the metaplastic agent, the ΔQ_{slow} (orange) rises to oppose the ΔQ_{fast} (blue). The resulting Net Update (black) is tiny (-0.0012), $10.6\times$ **slower** than the healthy agent's.
- The "**X-Ray**" (Appendix 4a, 4b): These show why Therapy 1 worked. When β_{meta} was set to 0, the agent's internal belief (Q-Value, 4b) was "released" from the tug-of-war and allowed to fall to 0. Interestingly, its internal "mood" (S_{meta}, 4b) **remained high**, suggesting the "therapy" blocked the *expression* of the pathology, not the underlying state.

5 Discussion

We have successfully built and validated a minimal, robust, and biologically-plausible model of **pathological rigidity**. Our central finding is that rigidity can emerge as a stable, suboptimal equilibrium—a state of "incomplete unlearning."

This result is more profound than simple "slowed learning." It provides a direct computational mechanism for perseverative behaviors (like in OCD or anxiety) where an individual **knows** a behavior is wrong (the "fast" ΔQ_{fast} signal is present) but is "stuck" by a persistent, internal "compulsion" (the "slow" ΔQ_{slow} signal). Our model shows how these two forces can find a stable balance, resulting in a *distorted belief about reality* and a chronic, maladaptive behavior.

Our model's components are grounded in established neuroscience: an opponent-process "tug-of-war" between a fast, DA-like RPE and a slow, 5-HT-like "aversive mood" signal that operates on an astrocyte-like timescale.

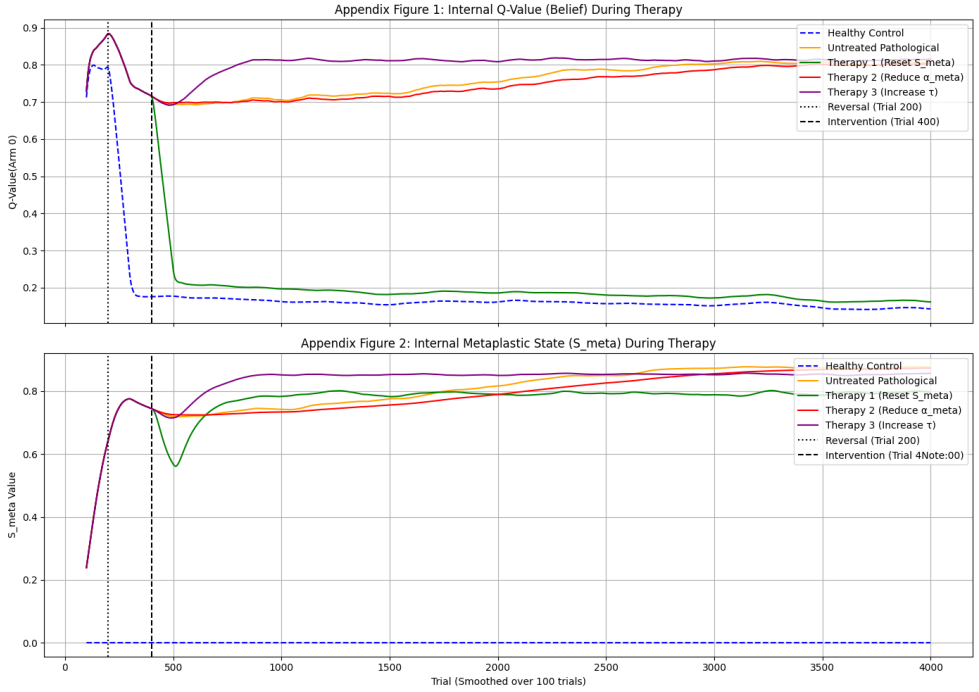
The success of our "ablation" therapy (Therapy 1) provides strong causal evidence for this mechanism. It demonstrates that the $S_{\text{meta}}/\beta_{\text{meta}}$ pathway is necessary and sufficient for the observed rigidity **within this model.** It suggests that effective therapies are those that don't just "speed up" learning, but those that **break the pathological loop itself**—functionally similar to how exposure therapy works by decoupling a stimulus from a maladaptive response **foa2016**; **stojek2018**.

6 Limitations and Future Work

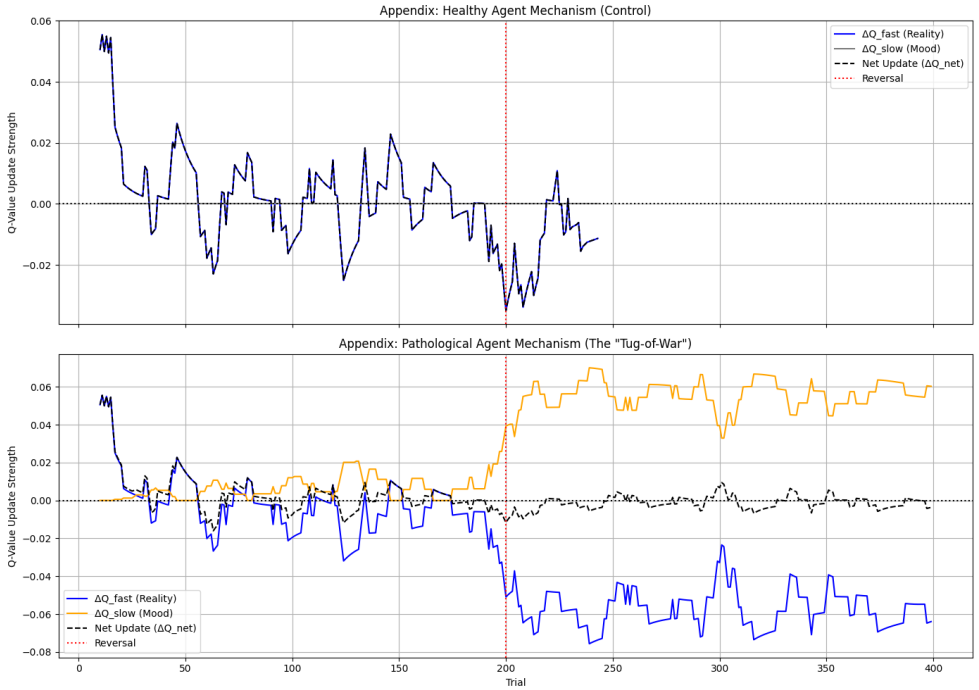
This is a minimal model, and its simplicity is its main limitation.

- **Biological Abstraction:** Our S_{meta} is an abstraction of the **functions** of serotonin and astrocytes, not a high-fidelity model of their complex, nonlinear dynamics. This is especially true of our timescale parameter ($\alpha_{meta} = 0.05$), which is only 2x slower than the neuronal rate, whereas true astrocytic dynamics are orders of magnitude slower. Future work could explore models with more biologically realistic, long-timescale integration.
- **Valence Exclusivity:** Our model is for **punishment-based inflexibility** (OCD, anxiety). It does not explain reward-based rigidity (addiction), which would require a different model (e.g., S_{meta} integrating positive RPEs).
- **Global vs. Local:** Our S_{meta} is a "global" state. A key area for future work (Day 2) is to explore a "**local**" model where S_{meta} is specific to each synapse. We hypothesize this would produce an even more "stuck" rigidity, as the "confusion" of S_{meta} being triggered by all bad news would be removed.

7 Appendix



(a) The 'Tug-of-War' mechanism. **(Top)** The Healthy agent's net update is negative. **(Bottom)** The Pathological agent's S_{meta} state generates a positive ΔQ_{slow} (orange) that opposes the negative ΔQ_{fast} (blue), resulting in a near-zero net update (the 10.6× slower rate). Internal Metaplastic State (S_{meta}). Note that in Therapy 1 (green), the *behavior* is cured, but the underlying S_{meta} state remains high, suggesting the 'mood' is disconnected from action, not erased.



(b) Internal Q-Value (Belief) for Arm 0. Therapy 1 (green) allows the belief to be corrected and fall to 0. The other agents remain 'stuck' with a high, false belief.

Figure 4: Appendix: Internal mechanism plots for diagnostic and therapy simulations.

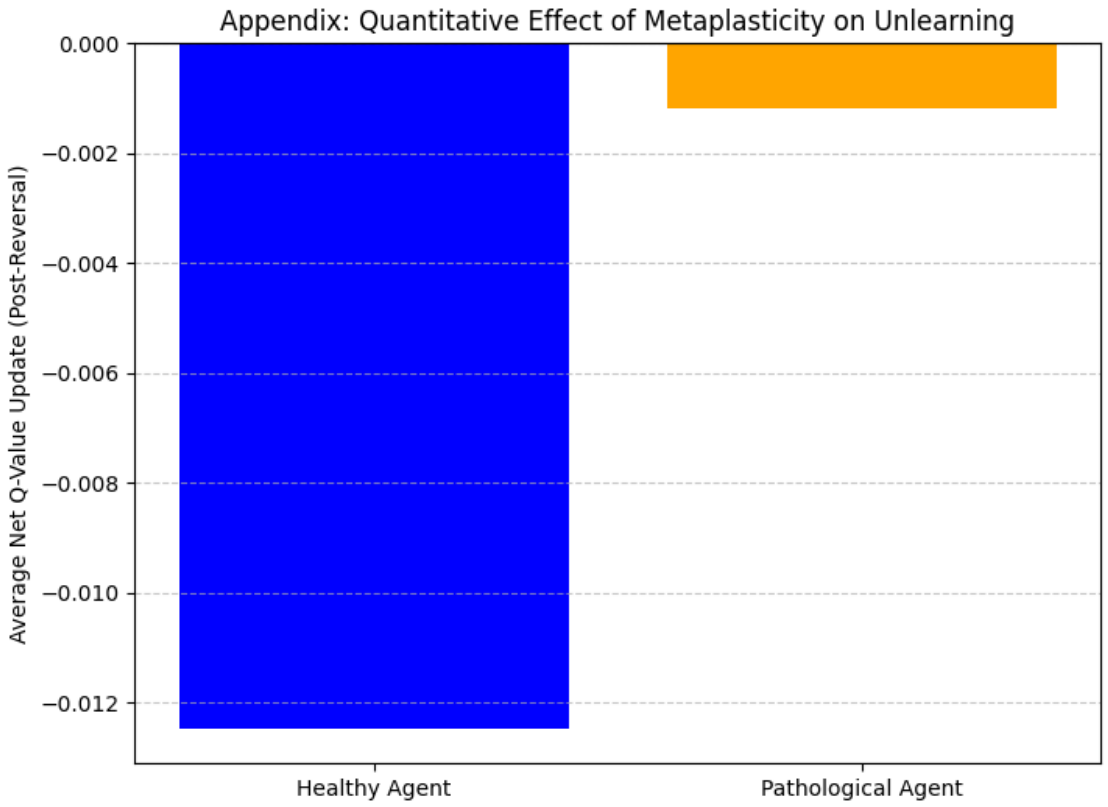


Figure 5: Quantitative effect of metaplasticity on the unlearning rate. This bar chart compares the average net Q-value update (ΔQ_{net}) for Arm 0, averaged across all post-reversal trials (201-400) from a single diagnostic run. The healthy agent (blue) has a strong negative update (-0.0125), representing a fast "reality" signal. The pathological agent (orange) has a net update of -0.0012, which is $10.6\times$ **slower.** This plot provides the quantitative proof of the "tug-of-war," showing that the opponent ΔQ_{slow} signal (from S_{meta}) successfully neutralizes over 90% of the "reality" signal (ΔQ_{fast}).

References

Adams, R. A., Huys, Q. J. M., & Roiser, J. P. (2015). Computational Psychiatry: Towards a mathematically informed understanding of mental illness. *Journal of Neurology, Neurosurgery & Psychiatry*, jnnp–2015–310737. https://doi.org/10.1136/jnnp-2015-310737

Bari, A., Theobald, D. E., Caprioli, D., Mar, A. C., Aidoo-Micah, A., Dalley, J. W., & Robbins, T. W. (2010). Serotonin Modulates Sensitivity to Reward and Negative Feedback in a Probabilistic Reversal Learning Task in Rats [Publisher: Nature Publishing Group]. *Neuropsychopharmacology*, *35*(6), 1290–1301. https://doi.org/10.1038/npp.2009.233

Cools, R., Nakamura, K., & Daw, N. D. (2011). Serotonin and Dopamine: Unifying Affective, Activational, and Decision Functions [Publisher: Nature Publishing Group]. *Neuropsychopharmacology*, *36*(1), 98–113. https://doi.org/10.1038/npp.2010.121

Daw, N. D., Kakade, S., & Dayan, P. (2002). Opponent interactions between serotonin and dopamine. *Neural Networks*, *15*(4), 603–616. https://doi.org/10.1016/S0893-6080(02)00052-7

Dayan, P., & Huys, Q. J. M. (2009). Serotonin in Affective Control [Publisher: Annual Reviews]. *Annual Review of Neuroscience*, *32*(Volume 32, 2009), 95–126. https://doi.org/10.1146/annurev.neuro.051508.135607

De Pittà, M., & Brunel, N. (2016). Modulation of Synaptic Plasticity by Glutamatergic Gliotransmission: A Modeling Study [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1155/2016/7607924]. *Neural Plasticity*, *2016*(1), 7607924. https://doi.org/10.1155/2016/7607924

Gong, L., Pasqualetti, F., Papouin, T., & Ching, S. (2024). Astrocytes as a mechanism for contextually-guided network dynamics and function. *PLoS computational biology*, *20*(5), e1012186. https://doi.org/10.1371/journal.pcbi.1012186

McLean, C. P., & Foa, E. B. (2024). State of the Science: Prolonged exposure therapy for the treatment of posttraumatic stress disorder. *Journal of Traumatic Stress*, *37*(4), 535–550. https://doi.org/10.1002/jts.23046

Morris, L., & Mansell, W. (2018). A systematic review of the relationship between rigidity/flexibility and transdiagnostic cognitive and behavioral processes that maintain psychopathology. *Journal of Experimental Psychopathology*, *9*(3), 2043808718779431. https://doi.org/10.1177/2043808718779431

Nakao, K., Singh, M., Sapkota, K., Fitzgerald, A., Hablitz, J. J., & Nakazawa, K. (2022). 5-HT2A receptor dysregulation in a schizophrenia relevant mouse model of NMDA receptor hypofunction [Publisher: Nature Publishing Group]. *Translational Psychiatry*, *12*(1), 168. https://doi.org/10.1038/s41398-022-01930-0

Sutton, R., & Barto, A. (1998). Reinforcement Learning: An Introduction. *IEEE Transactions on Neural Networks*, *9*(5), 1054–1054. https://doi.org/10.1109/TNN.1998.712192

Wang, X.-J., & Krystal, J. H. (2014). Computational Psychiatry. *Neuron*, *84*(3), 638–654. https://doi.org/10.1016/j.neuron.2014.10.018