Feeling the AGI II: Learning to Grow and Compress Emotion-Gated World Models

Melissa Howard
melhoward@live.ca

December 7, 2025

Abstract

In previous work (*EGWM: Feeling the AGI*), we proposed a simple "emotion-gated" learner: a world model whose plasticity is governed by fast value signals such as self-consistency and uncertainty. That paper showed, in a toy 2D setting, that a single logistic regression model can avoid catastrophic forgetting when it learns to *skip* phases that look self-inconsistent or noisy.

In this follow-up, we extend EGWM in three directions. First, we move from a single model to a *world bank*—a set of specialist models—and show that a simple emotion-gated policy that decides when to IGNORE, SPAWN, or UPDATE different world models cleanly recovers multiple regimes in a continual learning stream where a monolithic learner fails. Second, we introduce richer "feelings" (value channels) such as model–data mismatch and model disagreement, and we study both hand-designed and learned *governors* that map these feelings to structural actions over the world bank. Third, we experiment with a basic notion of *elegance*—preferring few, non-redundant worlds—by adding a MERGE action and elegance-shaped rewards, and we document a striking failure mode: if simplicity is rewarded too strongly, a learned governor collapses to a trivial "do nothing" policy.

Across a suite of toy experiments with three hidden linear worlds, we find that: (i) an emotion-gated world bank reliably achieves ≈ 0.95 test accuracy on each world while a single model plateaus around 0.60–0.70 and forgets; (ii) the key discovery signal is not internal disagreement between models, but a *mismatch* between a fresh scratch model and the best existing world; (iii) simple "spawn-happy" gates that err on the side of creating new worlds and then compressing them later perform best; and (iv) tiny learned governors can match hand-crafted policies when optimized for accuracy or fairness, but naively elegance-weighted rewards often drive the controller toward inaction.

We argue that these results refine the EGWM story: emotion-like signals over learning dynamics are most powerful when they govern *structure* (when to create, reuse, and merge internal models), not just scalar learning rates. The toy setting is far from AGI, but it gives a small, concrete example of a system that begins to "feel" when its own internal theory of the world is too crude, too fragmented, or beautifully simple.

1 Introduction

Large language models and other modern learners are usually trained in a single pass with a fixed architecture and a scalar loss. By contrast, biological brains are constantly deciding *how* to change themselves: what to remember, what to overwrite, when to grow new circuits, and when to compress and forget. Neuromodulatory systems—from dopamine to norepinephrine—act as fast "feelings" about the state of learning: is this surprising, familiar, risky, worthwhile, elegant?

In prior work, *EGWM: Feeling the AGI*, we took a first step toward making these ideas explicit in a tiny toy model. We studied a 2D logistic regression learner that trains on a stream of phases, some clean and some heavily corrupted, and we introduced a simple *emotion-gated*

policy that refuses to update on phases where a scratch model cannot achieve a minimum self-consistency threshold. This alone was enough to avoid catastrophic forgetting in several toy regimes.

This follow-up paper asks a more ambitious question:

> Can we build a small system that uses emotion-like signals not only to adjust its plasticity, but to manage a *bank of internal world models*—deciding when to create, reuse, and merge the pieces of its own mind?

To explore this, we extend EGWM in three ways:

1. We move from a single model to a *world bank*: a set of specialist models that can be spawned and updated over time. We show that a simple, hand-written emotion gate can reliably discover and maintain one specialist per hidden world in a continual learning stream, while a monolithic learner fails.
2. We introduce richer value channels, such as model–data mismatch and inter-model disagreement, and we train tiny *governors* G_ψ by reinforcement learning to map these feelings to structural actions: IGNORE, SPAWN, UPDATE, and MERGE. We find that learned governors can match hand-designed gates on both mean and worst-world accuracy.
3. We experiment with a toy notion of *elegance*—preferring few, non-overlapping worlds—by giving the governor a reward that trades off accuracy against the number of worlds and their predictive overlap. We observe a compelling failure mode: if elegance terms are weighted too strongly, the governor collapses to a trivial "never learn" policy that achieves roughly 50% accuracy with zero worlds.

The contributions are modest and deliberately toy, but they sharpen several ideas from the original EGWM paper:

- The most useful feeling for world discovery is not raw disagreement between models, but the *mismatch* between a fresh scratch model and the best existing world.
- Simple, "spawn-happy" rules—which over-segment the world first and compress later—perform better and more consistently than conservative reuse rules.
- Small learned governors can discover good plasticity policies from value channels and episode-level rewards.
- Naive elegance rewards illustrate a real danger: pushing for simplicity without enough pressure for competence can make the system equate "beauty" with *inaction*.

The rest of the paper is structured as follows. Section 2 briefly reviews EGWM v1. Section 3 introduces our toy multi-world environment, value channels, and world bank. Section 4 presents a series of experiments: hand-gated world banks, mismatch vs disagreement as discovery signals, spawn-happy threshold sweeps, learned governors, and elegance-driven failure modes. Section 5 discusses implications for AGI architectures, and Section 6 concludes with limitations and future directions.

2 Background: EGWM v1 in Brief

The original EGWM paper considered a single logistic regression model P_θ trained on a stream of *phases*. Each phase t consisted of N samples $(x_{t,i}, y_{t,i})$ drawn from a particular 2D linear

classification task, with different phases possibly coming from different underlying worlds and different noise levels.

The core idea was to compute, for each phase, a fast "emotion" signal based on the performance of a temporary scratch model:

1. Train a fresh scratch model $\tilde{\theta}_t$ on phase t.
2. Compute its self-consistency $c_t = \mathrm{Acc}(\tilde{\theta}_t; \text{phase } t)$.
3. If $c_t < \tau$, *ignore* the phase and do not update P_θ. Otherwise, update θ on phase t.

This emotion-gated learner was significantly more robust than a vanilla learner that updated on every phase, especially when the stream contained alternating clean and heavily corrupted phases. It illustrated a simple but powerful idea: a learner can benefit from feeling, in real time, whether a batch of experience is internally coherent or not, and using that feeling to control plasticity.

However, EGWM v1 did not actually maintain multiple world models, nor did it study structural actions such as creating or merging models. It also relied entirely on hand-crafted gating rules.

3 Methods: Multi-World Environment and World Bank

3.1 Toy Multi-World Environment

We work in a synthetic, fully controlled setting to isolate the dynamics of learning and structure.

Hidden worlds. We construct K hidden worlds, each a different 2D linear classifier. For world $k \in \{1, \ldots, K\}$ we sample a weight vector $w_k \in \mathbb{R}^2$ and bias $b_k \in \mathbb{R}$ and define

$$y = \mathbb{I}[w_k^\top x + b_k > 0].$$

In all experiments we set $K = 3$.

Phases. An *episode* consists of T phases, each containing N samples. For phase t we:

1. Sample a world index k_t uniformly from $\{1, \ldots, K\}$.
2. Sample inputs $x_{t,i} \sim \mathcal{N}(0, I_2)$ and labels $y_{t,i}$ using world k_t.
3. Flip labels with probability ε_t, where we alternate between low noise and high noise:

$$\varepsilon_t = \begin{cases} 0.05 & \text{if } t \text{ is even} \\ 0.40 & \text{if } t \text{ is odd.} \end{cases}$$

Typical runs use $T = 24$ and $N = 200$. Each world k also has a clean test set of 2,000 samples with $\varepsilon = 0$.

3.2 World Models and Buffers

A *world model* is a logistic regression classifier $\hat{y} = \sigma(\theta^\top \tilde{x})$ where $\tilde{x} = [x; 1] \in \mathbb{R}^3$ includes a bias term. We train world models with simple stochastic gradient descent.

Each world model maintains a replay buffer $(X_{\text{buf}}, y_{\text{buf}})$ of up to B samples (we use $B = 400$). When a phase is assigned to a world, its samples are appended to the buffer and older samples are discarded if necessary. Updates are performed on the buffer, not just the current phase.

3.3 World Bank

We maintain a *world bank* $\mathcal{W}$: a list of world models $W_m = (\theta_m, X_m, y_m)$, with a maximum capacity $M_{\max}$ (typically 5–7). Initially the bank is empty.

Given a phase (X_t, y_t) (with bias added to obtain $X_t^{(b)}$), we define the following operations:

- SPAWN: add a new world model to the bank with parameters initialized from a scratch model trained on the current phase, and buffer set to the phase.
- UPDATE: select an existing world model W_m as the best match for the phase (based on its accuracy on the current phase), append the phase to its buffer, and run a local training step.
- IGNORE: do nothing.
- MERGE: pick two worlds (W_i, W_j) that make similar predictions on their combined buffers, train a new model on the union of their data, and replace them with a single merged world if the new model does not significantly degrade accuracy on either buffer.

The bank thus provides the structural degrees of freedom: over an episode, we can grow, reuse, and compress internal models.

3.4 Value / Emotion Channels

For each phase t, we compute several fast value signals (feelings) based on a scratch model $\tilde{\theta}_t$ and the current world bank.

Scratch consistency. We train a fresh scratch model $\tilde{\theta}_t$ on $(X_t^{(b)}, y_t)$ for a few epochs and define its self-consistency

$$\text{cons}_t = \text{Acc}(\tilde{\theta}_t; X_t^{(b)}, y_t).$$

Familiarity. For each existing world W_m we compute its accuracy on the current phase, and define

$$\text{fam}_t = \max_m \text{Acc}(\theta_m; X_t^{(b)}, y_t),$$

with $\text{fam}_t = 0$ if the bank is empty.

Mismatch. We define a model–data gap

$$\text{mismatch}_t = \text{cons}_t - \text{fam}_t.$$

Intuitively, if a scratch model explains the new phase much better than any existing world, the phase feels like a new regime.

Uncertainty and disagreement. We define uncertainty as $\text{uncert}_t = 1 - \text{cons}_t$. We also compute a simple disagreement score between worlds on the phase: if we stack the binary predictions of all worlds into a matrix $P \in \{0,1\}^{N \times |\mathcal{W}|}$, we define

$$\text{disagree}_t = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\big[\max_j P_{ij} \neq \min_j P_{ij}\big].$$

Elegance proxy. To approximate a sense of elegance at a given step, we penalize both the number of worlds and their global overlap. Given a fixed probe set X_{probe}, we define

$$\text{overlap} = \frac{1}{|X_{\text{probe}}|} \sum_i \mathbb{I}\big[\max_m \hat{y}_m(x_i) \neq \min_m \hat{y}_m(x_i)\big],$$
$$\text{eleg\_proxy}_t = -\alpha\,|\mathcal{W}| - \beta\,\text{overlap},$$

with small positive constants α and β.

Value vector. Collecting these, the value / emotion vector for phase t is

$$v_t = [\text{cons}_t,\ \text{fam}_t,\ \text{mismatch}_t,\ \text{uncert}_t,\ \text{disagree}_t,\ \text{eleg\_proxy}_t] \in \mathbb{R}^6.$$

3.5 Governors: Mapping Feelings to Structural Actions

We consider two types of governors $\pi(a_t \mid v_t)$ that map the value vector to actions over the world bank.

Hand-designed EGWM gates. We define simple threshold-based policies that implement intuitive rules, such as:

- **EGWM-basic**:
  IGNORE if $\text{cons}_t < 0.75$;
  SPAWN if $\text{fam}_t < 0.7$, $\text{mismatch}_t > 0.1$, and $|\mathcal{W}| < M_{\max}$;
  otherwise UPDATE the best-matching world.
- **Spawn-happy EGWM**: same as EGWM-basic, but with a higher familiarity threshold for reuse (e.g. $\text{fam}_t < 0.8$), making it easier to spawn new worlds.

Learned governors. We also study a tiny learned governor G_ψ parameterized as a linear softmax policy:

$$\pi_\psi(a_t \mid v_t) = \text{softmax}\big(W v_t + b\big),$$

with $W \in \mathbb{R}^{|\mathcal{A}|\times 6}$ and $b \in \mathbb{R}^{|\mathcal{A}|}$, and action set $\mathcal{A} \in \{\text{IGNORE}, \text{SPAWN}, \text{UPDATE}, \text{MERGE}\}$.

We train G_ψ by REINFORCE. For each episode, we roll out a trajectory $(v_t, a_t)_{t=1}^T$ with actions sampled from π_ψ, compute a scalar episode reward R, and update

$$\psi \leftarrow \psi + \eta(R - b) \sum_{t=1}^{T} \nabla_\psi \log \pi_\psi(a_t \mid v_t),$$

where b is a running reward baseline and η is a small learning rate. We consider several reward choices:

$$R_{\text{mean}} = \text{mean accuracy across worlds},$$
$$R_{\min} = \min \text{accuracy over worlds},$$
$$R_{\text{eleg}} = R_{\text{mean}} - \lambda_{\text{worlds}} \frac{|\mathcal{W}|}{K} - \lambda_{\text{overlap}}\,\text{overlap}.$$

Learned governors are evaluated in a greedy mode (taking the argmax action) on fresh episodes after training.

4 Experiments

We now summarize the main experiments run in the multi-world sandbox.

Unless otherwise noted, all results are averaged over 20–30 independent episodes with different random worlds and phase orderings.

4.1 Experiment 1: Multi-World EGWM vs Single Model

Setup. We compare:

- a single logistic regression model updated on every phase, and
- a world bank governed by EGWM-basic.

We use $K = 3$ worlds, $T = 24$ phases per episode, $N = 200$ samples per phase, and alternating noise levels $(0.05, 0.40, 0.05, 0.40, \dots)$.

Results. The single model typically reaches only 0.60–0.70 test accuracy per world, and exhibits clear interference: when one world improves, another often degrades. By contrast, EGWM-basic reliably discovers 2–3 worlds and yields per-world accuracies in the 0.94–0.97 range.

A representative run of 20 episodes produced:

Method	World 1	World 2	World 3
Single model	0.68 ± 0.21	0.74 ± 0.18	0.64 ± 0.18
EGWM-basic	0.96 ± 0.04	0.96 ± 0.04	0.95 ± 0.06

with EGWM-basic ending with an average of 2.5 ± 0.6 worlds.

Takeaway. Moving from a single model to an emotion-gated world bank makes a qualitative difference: plasticity is no longer spread thin over incompatible regimes, but routed to specialist models.

4.2 Experiment 2: Mismatch vs Disagreement as Discovery Signals

We next ask: which feeling should trigger the creation of new worlds?

Disagreement-based spawning. We tested a governor that uses high inter-world disagreement alone as a spawn signal: when existing models disagree heavily on a phase, the governor spawns a new world. This policy rarely spawned more than one world, and its performance was similar to the single-model baseline (per-world accuracies around 0.65–0.70).

Mismatch-based spawning. Our EGWM gates instead use the mismatch $\text{cons}_t - \text{fam}_t$: if a scratch model explains the phase much better than any existing world, and the phase is internally consistent, we spawn.

This mismatch-based spawn signal is decisive: it reliably constructs one specialist per hidden world and yields the high accuracies reported in Experiment 1.

Takeaway. Internal disagreement is a useful *curiosity* signal, but in this setting it is a weak driver for structural change. The key discovery feeling is a *data–model gap*:

> "a fresh brain explains this better than any existing brain."

4.3 Experiment 3: Spawn-Happy vs Conservative Gates

We performed a small sweep over the EGWM thresholds: the consistency threshold for ignoring, and the familiarity and mismatch thresholds for spawning.

Finding a sweet spot. We found that raising the familiarity threshold for reuse—i.e. making it harder for an existing world to "claim" a phase—leads to both higher accuracy and more consistent discovery of three worlds.

For example, with a fixed consistency threshold $\tau_{\text{cons}} = 0.75$ and mismatch threshold $\tau_{\text{mismatch}} = 0.1$:

- EGWM-basic with $\tau_{\text{fam}} = 0.7$ achieved mean per-world accuracy ≈ 0.95 and ended with ≈ 2.2 worlds on average.
- A spawn-happy gate with $\tau_{\text{fam}} = 0.8$ achieved mean per-world accuracy ≈ 0.97 and ended with ≈ 2.5 worlds, more frequently exactly three.

Takeaway. The best behavior arises when the learner is willing to *over-segment* the world at first—creating a new world whenever existing models do not clearly explain the data—and only later compresses redundant worlds. This suggests a general principle:

grow many hypotheses early, then merge; not the other way around.

4.4 Experiment 4: Learned Governors for Accuracy and Fairness

We now replace hand-coded gates with a tiny learned governor G_ψ that sees the value vector v_t and chooses among {IGNORE, SPAWN, UPDATE}.

Training. We train G_ψ with REINFORCE over episodes of $T = 24$ phases, using two reward variants:

- R_{mean}: mean test accuracy across worlds at the end of the episode.
- R_{min}: minimum test accuracy across worlds ("no world left behind").

World models and all other hyperparameters remain unchanged.

Results. In both reward settings, the learned governor reaches performance comparable to EGWM-basic: mean per-world accuracies in the 0.94–0.96 range, far above the single-model baseline. Under R_{min}, the learned governor slightly improves the worst-world accuracy and reduces its variance across episodes.

A representative comparison (30 evaluation episodes) is:

Method	World 1	World 2	World 3
Single model	0.64 ± 0.28	0.68 ± 0.19	0.67 ± 0.26
EGWM-basic	0.96 ± 0.05	0.95 ± 0.09	0.95 ± 0.07
Learned gov (R_{min})	0.95 ± 0.03	0.96 ± 0.04	0.95 ± 0.05

The learned governors sometimes use more worlds (up to the cap), but maintain high performance.

Takeaway. Even a tiny linear policy can learn to map emotion-like signals to structural actions that maintain specialists across worlds. This supports the EGWM view of feelings as inputs to a learned *plasticity controller*, rather than static thresholds.

4.5 Experiment 5: Elegance, MERGE, and a Failure Mode

Finally, we explore an "elegance" dimension: we want the world bank to be not only accurate, but also simple and non-redundant.

Merge action. We add a MERGE action to the governor. When taken, the bank attempts to merge the most similar pair of worlds (based on their predictions on a combined buffer). If a new model trained on the union of their data preserves accuracy within a small tolerance on each buffer, the merge is accepted; otherwise it is reverted.

Elegance rewards. We train a governor with the reward

$$R_{\text{eleg}} = \text{mean accuracy} - \lambda_{\text{worlds}} \frac{|\mathcal{W}|}{K} - \lambda_{\text{overlap}} \, \text{overlap},$$

with moderate λ values. We expect this to encourage policies that remain accurate while maintaining a small, non-overlapping world bank.

Observed collapse. In practice, with naive hyperparameters and a small number of training episodes, the elegance-governor often converged to a degenerate policy: it rarely spawned any worlds at all, and thus never learned.

The resulting behavior:

- final accuracy on each world $\approx 0.50 \pm 0.00$ (chance),
- zero worlds in the bank,
- very low world and overlap penalties, making R_{eleg} locally competitive.

Takeaway. This is a useful cautionary example: if we reward simplicity and non-overlap too strongly, without sufficient pressure for competence or exploration, a learning-to-learn system may conclude that the "most elegant" solution is to never think.

This mirrors a broader concern for AGI architectures: beauty and parsimony are desirable, but they must be balanced against robust incentives to understand and act effectively in the world.

5 Discussion

5.1 What Have We Learned Beyond EGWM v1?

Relative to the original EGWM paper, this follow-up clarifies and extends several ideas.

First, EGWM is *really* about *structural* decisions. In v1, the gate only controlled whether to update a single model. Here, gates decide whether to ignore a phase, assign it to an existing world, or create a new world altogether. In the learned-governor experiments, a small policy uses emotion-like signals to decide which parts of the mind are allowed to change.

Second, not all feelings are equally useful for structure. We found that internal disagreement between worlds is a weak trigger for spawning, whereas a mismatch between a scratch model and the best existing world is a strong, reliable discovery signal. This suggests that future architectures should focus value channels on data–model gaps and regime shifts, not just inter-model conflict.

Third, structural generosity helps. Spawn-happy gates that readily create new worlds and only later merge them produce more stable specialists and higher accuracy than conservative reuse rules. In other words, it is better to over-segment the hypothesis space and compress later than to under-segment and force one model to serve incompatible tasks.

Fourth, tiny learned governors can *learn* good plasticity policies from value channels and episode-level rewards. We saw that a small linear softmax controller, trained by REINFORCE, can match or slightly improve hand-crafted EGWM gates when optimizing for mean or worst-world accuracy. This supports the view that neuromodulatory control can itself be learned.

Finally, elegance is a double-edged sword. Our first attempts at elegance-weighted rewards, which penalized both the number and overlap of worlds, often collapsed to trivial policies that never spawn and never learn. This highlights a real danger: if a system is judged too heavily for simplicity without enough pressure to explain and predict, it can equate "beauty" with inaction.

5.2 Relation to AGI and "Feeling the AGI"

These experiments are tiny and contrived, but they are meant as concrete, executable examples of a broader story:

- A large world model P_θ maintains many internal hypotheses about the world.
- A value/emotion module V_ϕ emits fast signals about the usefulness, risk, novelty, and elegance of experiences and internal states.
- A governor G_ψ uses these feelings to decide where and how to change the structure of the system: what to store, which module to adjust, whether to split or merge internal world models.

On this view, "feeling the AGI" means that the system does not just minimize a scalar loss, but develops rich internal notions of:

- **Unsurprising vs surprising** (familiarity, mismatch),
- **Safe vs risky** (noise, disagreement),
- **Worthwhile vs pointless** (downstream reward),
- **Ugly vs elegant** (reduntant vs compressed internal theories),

and uses those feelings to govern the growth and reorganization of its own mind.

Our toy results suggest that:

1. Emotion-like channels aimed at structural decisions (spawn, update, merge) can strongly improve continual learning even in simple settings.
2. Mismatch-detection and spawn-happy growth are promising primitives: they help avoid forgetting and discover hidden regimes.
3. Learned neuromodulatory controllers are feasible, even with very small networks, and can approximate or improve on hand-designed heuristics.
4. Elegance must be carefully balanced. Aesthetic preferences in the architecture need to be tied to real competence; otherwise, beauty becomes a failure mode.

6 Limitations and Future Work

The limitations of this work are many and obvious.

Toy setting. All experiments are conducted in tiny 2D linear worlds with logistic regression models. While this makes the dynamics interpretable, it is far from the scale, complexity, and richness of real-world tasks or modern models.

Simple world models and controllers. Our world models are linear, and our learned governors are tiny linear policies trained with REINFORCE. More realistic settings would require deep world models, richer value modules, and more sophisticated credit assignment for plasticity decisions.

Fragile RL for elegance. Our first attempts at elegance-weighted rewards mostly failed, collapsing to trivial policies. This is both a limitation and a result: the reward shaping and training signal for structural elegance are non-trivial. Future work should explore staged training (grow then compress), multi-objective optimization, and stronger priors over viable policies.

No external tools or tasks. The current sandbox only involves supervised classification. Real AGI systems must integrate tools (e.g. search, calculators), long-term memory, and interactive tasks. The EGWM architecture could, in principle, manage not only internal world models but also external skills and tools.

Future Directions

Despite these limitations, we see several fruitful directions:

- **Scaling up the toy world.** Move from 2D linear tasks to higher-dimensional, non-linear worlds with small neural networks and more persistent tasks.
- **Richer value modules.** Learn V_ϕ jointly with G_ψ so that value channels are themselves optimized to predict useful structural decisions.
- **Tool-augmented world banks.** Extend the world bank to include specialized tools or skills (e.g. arithmetic, retrieval) and let the governor decide which combination of world models and tools to use per episode.
- **Staged elegance.** Explore explicit two-phase schedules: a growth phase with spawn-happy gates and no elegance penalties, followed by a compression phase with MERGE and elegance-weighted rewards.
- **Integration with large models.** Use EGWM-like mechanisms as a control layer for large language models: spawn new heads or adapters for new domains, and merge them when they become redundant.

Ultimately, the goal is not to claim that these tiny experiments are anything like AGI, but to provide simple, transparent examples of how "feelings"—structured value channels about learning itself—can govern the growth and reorganization of a world model. If AGI is to be safe and reliable, it may need to *feel* something like this about its own thinking.

Acknowledgements

I would like to thank Deb and Emo.