Part 7 – EGWM: validating the student and teacher

Mel Howard

December 2025

1 Introduction

This section will introduce the motivation and context for Part 7 of EGWM: validating the student (proposal network) and teacher (structural governor) on a simple continual-learning benchmark.

1.1 Experiment 6: Hindsight-Learned Structural Controller (HR-EGWM Toy)

In Experiments 1–5, EGWM treated the structural governor as a hand-designed rule: a fixed mapping from value channels ("feelings") to structural actions (SPAWN, UPDATE, etc.). In this experiment, we take the first concrete step toward the Hierarchical Reflective EGWM (HR-EGWM) architecture by asking a more ambitious question:

> Can a fast, learned proposal network P_θ reliably imitate a slower, hand-crafted "System 2" structural governor, purely from hindsight traces of value channels and chosen structural actions?

We construct a small but non-trivial continual learning world where a teacher governor (System 2) uses EGWM feelings to decide whether to **SPAWN** a new expert head or **UPDATE** an existing one. We then train a logistic proposal network P_θ on those hindsight decisions and test whether it can replace the hand-crafted rule without loss in competence.

Environment and tasks. We define four distinct linear classification "worlds" in $\mathbb{R}^2$. Each world $w \in \{0, 1, 2, 3\}$ is specified by a weight vector $W^{(w)} \in \mathbb{R}^2$ and bias $b^{(w)} \in \mathbb{R}$ and induces binary labels

$$y = \begin{cases} 1, & \text{if } W^{(w)} \cdot x + b^{(w)} > 0, \\ 0, & \text{otherwise,} \end{cases}$$

for inputs $x \in [-2, 2]^2$.

An episode consists of 80 phases. For the first 20 phases we sample from world 0, then from worlds 1, 2, and 3 in blocks of 20 phases each. Thus the agent experiences a non-stationary stream that gradually introduces new worlds it has not seen before. Each phase t provides a batch of 64 i.i.d. samples from the active world. For evaluation we sample a separate batch of 200 points from the same world and measure classification accuracy.

Structural heads and feelings. The agent maintains a bank of logistic heads $\{h_i\}$, each parameterized by weights and bias $\phi_i = (w_i, b_i)$ and trained with gradient descent. For a given phase with active world w_t, we compute:

- A **scratch model** h_{scratch}: a fresh logistic regressor trained only on the current phase.
- The **best existing head** h_{best} by accuracy on the phase.
- **Phase accuracies** acc_{best} and $\text{acc}_{\text{scratch}}$ on the phase batch.
- **Mismatch** $\Delta_{\text{mismatch}} = \text{acc}_{\text{scratch}} - \text{acc}_{\text{best}}$.

We collect a low-dimensional feeling vector

$$f_t = \left[\text{acc}_{\text{best}},\ \text{acc}_{\text{scratch}},\ \Delta_{\text{mismatch}},\ \frac{|\mathcal{H}_t|}{10}\right],$$

where $|\mathcal{H}_t|$ is the number of heads currently in the bank, normalized by 10. These channels correspond, respectively, to a familiarity signal, a local learnability signal, a novelty or mismatch signal, and a crude capacity load signal.

Teacher structural policy (System 2). We first define a hand-crafted teacher governor G_{teacher} that maps feelings to structural actions:

1. At the very first phase, when there are no heads present, always **SPAWN** a head from the scratch model.
2. For each later phase, compute acc_{best}, $\text{acc}_{\text{scratch}}$, and Δ_{mismatch}.
3. If
$$\text{acc}_{\text{scratch}} > 0.7 \quad \text{and} \quad \Delta_{\text{mismatch}} > 0.1,$$
then **SPAWN**: add a new head initialized from the scratch model.
4. Otherwise, **UPDATE**: further train the current best head h_{best} on the phase data.

To isolate structural decisions, we evaluate competence using an oracle router: for each evaluation batch we report the accuracy of the best-performing head on that world. This measures how well the structural decisions (SPAWN vs UPDATE) maintain a useful expert for each world.

Controller	Mean phase acc.	Std. dev.	Final # heads	SPAWNs / episode
Teacher (hand-crafted rule)	0.985	0.002	3.6	2.6
Student (P_θ proposal)	0.984	0.002	3.5	2.5

Table 1: HR-EGWM toy experiment. A fast, learned proposal network P_θ imitates a hand-crafted System 2 structural governor using hindsight traces of feelings and actions. When used as the sole structural controller, P_θ matches the teacher's competence while using a slightly more conservative number of SPAWN operations and expert heads.

We run 20 episodes under this teacher policy, collecting for each phase

$$(f_t, a_t),$$

where $a_t \in \{\text{UPDATE} = 0,\ \text{SPAWN} = 1\}$. This yields 1600 labeled examples of feelings and structural actions, with SPAWN chosen on roughly 4% of steps.

Learning the proposal network P_θ. We now train a small logistic proposal network P_θ to imitate G_{teacher}. The network takes the standardized feeling vector f_t as input and outputs a Bernoulli probability

$$P_\theta(a_t = \text{SPAWN} \mid f_t),$$

trained with cross-entropy to predict the teacher's action a_t.

After standardizing each feature of f_t (zero mean and unit variance), a simple logistic regressor trained with gradient descent achieves high accuracy on the teacher labels, indicating that the feelings are sufficient to reconstruct the teacher's structural decisions.

Student regime: replacing the teacher by P_θ. Next we run new episodes in the same environment, but replace the teacher governor with the learned structural controller:

1. At the first phase, we spawn a head from the scratch model, as before.
2. At every subsequent phase, we compute f_t and choose an action

$$a_t = \begin{cases} \text{SPAWN}, & \text{if } P_\theta(\text{SPAWN} \mid f_t) > 0.5, \\ \text{UPDATE}, & \text{otherwise.} \end{cases}$$

3. SPAWN and UPDATE are applied exactly as in the teacher regime.

We again evaluate competence using oracle routing to the best head per world. Table 1 summarizes performance over 20 teacher episodes and 20 student episodes.

Takeaway. This toy HR-EGWM experiment demonstrates that:

1. A simple, low-dimensional feeling vector f_t is sufficient for a slow, hand-crafted governor to make effective structural decisions (SPAWN vs UPDATE) in a non-stationary world.

2. The same feelings can be used as inputs to a fast proposal network P_θ that learns the structural policy from hindsight traces, achieving near-perfect imitation of System 2.

3. When we replace the hand-crafted governor by P_θ, the agent maintains the same level of competence (oracle-routed phase accuracy) with a very similar number of heads and SPAWN operations.

In other words, even in this minimal setting, we see a complete HR-EGWM loop: a reflective, slow controller generates structural experience; this experience trains a fast proposal network P_θ that can then autonomously control model structure using only felt value channels.