NEURAL POPULATION EMBEDDINGS AS A TEACHER FOR BELIEF-GUIDED DECISION MAKING

Anonymous authors
Paper under double-blind review

ABSTRACT

Low-dimensional structure in neural population activity often resembles the representational geometry that emerges in artificial networks trained on the same task, which suggests that recorded neural activity could serve as a teacher for artificial agents. Prior brain-tuning work, including its single application to decision-making, has transferred representations tied to the stimulus presented on each trial. To our knowledge, no study has asked whether neural activity can transmit a latent context that is absent from the input and must be inferred across trials from the history of choices and outcomes. Here we address this question using the International Brain Laboratory (IBL) task, in which mice choose the side of a visual stimulus of varying contrast while the prior probability of that side switches between uncued biased blocks, such that on zero-contrast trials the animal must rely on its inferred belief alone. Specifically, we train a recurrent agent on a simulated version of the same protocol and, building on the Brain2Model paradigm, add a transfer loss that aligns a low-dimensional belief embedding within the agent to a teacher constructed from CEBRA embeddings of brain-wide IBL population activity averaged by trial condition. We compare this brain teacher against a noiseless Bayesian ideal-observer teacher and a shuffled control that retains the same embedding vectors under a permuted condition assignment. On zero-contrast trials, the brain teacher matches the ideal-observer teacher across the mid-range of transfer weights despite receiving only a coarse belief label rather than the observer's posterior, and it outperforms the shuffled control by a margin that grows with transfer weight while also accelerating learning. Importantly, this advantage cannot be attributed to regularization, since all three teachers produce the same small gain at low transfer weight, and it replicates with a self-supervised embedding constructed without any belief labels. Moreover, within the neural embedding itself, the separation between low- and high-prior states is largest at zero contrast, where the prior is the only information available to the animal. Together, these results provide the first demonstration that neural population data can substitute for a hand-designed teacher when the task-relevant variable is latent and must be inferred across trials.

1 INTRODUCTION

Past work indicates that sensorimotor representations in neural populations during decision-making are often restricted to low-dimensional trajectories, relative to the total dimensionality of the data (Churchland et al., 2012; Mante et al., 2013; Cunningham & Yu, 2014; Gallego et al., 2017; Langdon et al., 2023). More recent work has suggested that deep neural networks trained in a variety of tasks from different domains display similar low-dimensional representation geometry (Hosseini et al., 2024; Huh et al., 2024), while neural representations may converge onto a specific subset of representations that overlaps with artificial models as a function of task complexity (Cao & Yamins, 2024; Yamins & Nayebi, 2026). Similar representational convergence may be exploited in representation transfer and knowledge distillation (Hinton et al., 2015; Romero et al., 2014; Tian et al., 2020; Gou et al., 2021), where new student models can be bootstrapped from pre-existing representations in a teacher model that is known to perform a task. By utilizing the brain as a teacher model, it is conceptually possible to train new models to higher performance, or in fewer epochs, in a variety of tasks, as demonstrated in machine vision (Fong et al., 2018; Li et al., 2019; Federer et al., 2020;

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Dapello et al., 2022), as well as in the domains of language (Negi et al., 2025; Moussa et al., 2025), reasoning (Xiao et al., 2026), and decision-making (Aquino et al., 2025). Importantly, nearly all of this work has targeted sensory or linguistic representations, in which the variable being transferred is tied to the stimulus presented on each trial, and the single application to decision-making (Aquino et al., 2025) provided the current context to the learner as an input. To our knowledge, no brain-tuning study has used neural activity as a teacher for a latent context that is absent from the input and must be inferred across trials from the history of choices and outcomes. This is the setting in which such a teacher should matter most, since an agent trained by reinforcement alone receives only sparse and delayed feedback about a hidden variable, and it is also the setting in which the ideal teacher, a Bayesian observer equipped with the generative model of the task, is rarely available outside of simulation.

Here we test whether neural population recordings can fill that role. Specifically, we adapt B2M, previously validated in memory-based decision-making, to a recurrent neural network (RNN) trained on the International Brain Laboratory (IBL) two-alternative task (International Brain Laboratory et al., 2021; Angelaki et al., 2025), in which the prior over stimulus side switches between uncued blocks. The rodent data were collected under the same protocol on which the agent is trained, such that neural and agent conditions can be matched one-to-one. This work differs from prior brain-tuning studies in four respects. First, the transferred variable is a hidden context rather than a stimulus encoding, and the student is a recurrent policy trained by reinforcement rather than a feedforward encoder. Second, the teacher is constructed from brain-wide rodent population recordings rather than human intracranial data. Third, we replace the convex trade-off between task and transfer losses with an additive transfer term whose weight α is swept, which allows the effect of alignment to be traced continuously. Fourth, we compare the brain teacher against both a noiseless Bayesian ideal observer, which bounds what any belief teacher can provide, and a shuffled control that holds identical embedding vectors under a permuted condition assignment, which isolates the contribution of the brain's condition-to-embedding binding from that of the vectors themselves. Two of these, the reinforcement learning student and the permuted-label control, address limitations identified in the original B2M work (Aquino et al., 2025). We demonstrate that brain alignment improves network performance on zero-contrast trials, in which the decision must be made in the absence of sensory evidence, that the brain teacher matches the ideal observer across the mid-range of α and exceeds the shuffled control by a margin that grows with α, and that this advantage replicates with a self-supervised embedding constructed without belief labels. Together, these results indicate that representations of top-down context belief can be transmitted to artificial agents from neural geometry alone.

1.1 The task and IBL dataset

We use the IBL dataset (International Brain Laboratory et al., 2021; Angelaki et al., 2025), which pairs brain-wide Neuropixels recordings with a standardized two-alternative forced choice task. On each trial, a visual stimulus of a given contrast appears to the left or right of center, and the mouse reports its choice by turning a wheel that moves the stimulus to the center (Figure 1 (a)). The contrast takes one of five discrete values, 0, 0.0625, 0.125, 0.25 and 1.0, where a contrast of 0 corresponds to the absence of any visual stimulus while the mouse is still required to make a choice.

In each session, after an initial block of unbiased trials of fixed length, the prior probability that the stimulus appears on the right becomes 0.8 (right block) or 0.2 (left block) and remains constant over a block of trials whose length is drawn from a truncated exponential distribution with a mean of 51 trials on the interval $[20, 100]$. The block identity is never cued, so the animal must infer it from the recent history of stimulus sides and outcomes. Prior expectations are represented across a broad set of regions in these recordings (Findling et al., 2025), which motivates using the population activity as a teacher for the block belief.

This dataset is well suited to the question above for two reasons. It offers population recordings of a scale and standardization that few datasets match, and the task itself is a partially observable Markov decision process, since the block prior is hidden and shifts over time. On zero-contrast trials in particular, the correct choice depends on the prior alone, so accuracy on these trials is a direct readout of how well an agent has inferred the hidden block. The question we ask is therefore whether an agent's representation of context belief, measured where sensory evidence is absent,

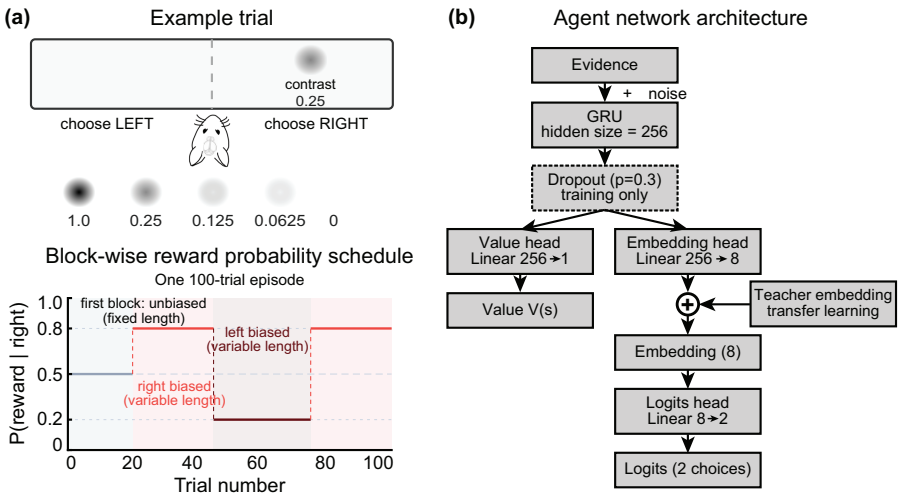

Figure 1: (a) Top: in each trial a visual stimulus of a given contrast appears to the left or right of center, and the task is to report the correct side. Middle: the five possible contrasts, including 0, which corresponds to a trial with no visual stimulus. Bottom: the block schedule of one simulated 100-trial episode. The first fixed number of trials have equal probability for each side. Thereafter the environment alternates between right-biased ($p_{right} = 0.8$) and left-biased ($p_{right} = 0.2$) blocks of varying lengths. (b) The agent network architecture. The input is a four-dimensional observation consisting of the noise-corrupted signed evidence, a one-hot encoding of the previous action, and the previous reward. It is then passed to a GRU with 256 hidden units, whose output feeds a value head and a linear 8-dimensional belief embedding. The policy reads only from this embedding. During training a transfer loss pulls the embedding toward the teacher's target for the current condition.

can be improved by aligning it to recordings from rodents that experienced exactly the same task structure.

1.2 BRAIN2MODEL VARIATION

The agent architecture is shown in Figure 1 (b). On each trial the agent receives noisy sensory evidence (see Methods) along with its previous action and reward, and the block prior is never provided as an input. These inputs are integrated by a gated recurrent unit (GRU) with 256 hidden units. The value head reads directly from the GRU state, whereas the policy does not. Instead, the GRU state is projected linearly onto an 8-dimensional belief embedding, and the policy is a linear map from this embedding to the two choice logits, such that every decision must pass through the low-dimensional bottleneck. During training, a transfer loss pulls this belief embedding toward the teacher's target for the current trial condition. The teacher therefore shapes the agent's internal representation without ever appearing in its input, and it is not consulted at test time. We depart from the original B2M formulation (Aquino et al., 2025) in how the two losses are combined. Rather than trading task loss against transfer loss through a convex mixture, we retain the full task loss and add the transfer term with a weight α that is swept across runs, such that the total loss becomes:

$$L_{total} = L_{task} + \alpha * L_{transfer}$$

This choice is motivated by the structure of the task. Trials vary in ambiguity, and the relative importance of the two contributing losses depends on that ambiguity. At high contrast the decision should follow the evidence and the task loss should dominate, whereas at zero contrast the decision should follow the prior and the transfer term is the only signal that speaks to it. Retaining the full task loss at every α also allows all three teachers to be compared at identical settings of the objective.

1.3 DIFFERENT TEACHERS

We compare three teachers that differ only in what the target embedding encodes (see Methods for construction). The Bayes teacher serves as a benchmark that establishes what alignment to the true belief can achieve when that belief is hand-coded. It encodes the posterior of an ideal observer that filters the true rewarded sides with the nominal generative parameters of the block process ($p_{high} = 0.8$, switch hazard $1/50$), and it supplies this posterior as a continuous scalar. The brain teacher is a lookup table that maps each trial condition to the mean CEBRA embedding (Schneider et al., 2023) of IBL population activity for that condition. The table has 27 entries, one for each combination of 3 belief bins and 9 signed-contrast buckets, so the brain teacher receives only a coarse belief label together with the signed contrast, and the geometry of its targets, both within and across bins, is inherited from the recordings. The shuffle teacher is the control. It uses the identical set of 27 brain embeddings but randomly permutes their assignment to conditions, with a new permutation drawn for each seed, so that it holds the same vectors, scale and covariance as the brain teacher and differs only in the correspondence between condition and embedding.

1.4 BRAIN EMBEDDINGS

To construct the brain teacher we trained two CEBRA models on the same recordings, one conditioned on the 27 behavioral condition labels (supervised) and one conditioned on time alone (self-supervised), each with held-out trials reserved for building the table (see Methods). We use the supervised table as the primary teacher because it encodes the belief variable more strongly (η^2 of 2.21% vs 0.21%, belief SNR of 3.06 vs 1.54, measured on the condition table alone). Importantly, because the supervised embedding is trained on the same condition labels the teacher later looks up, we treat the self-supervised table as the critical control. The brain advantage over shuffle replicates on it under strong alignment, with no belief label used anywhere in its construction (Figure 3). All teachers reported here use tables built from trials on which the mouse chose correctly.

2 RESULTS

Figure 2 (a) shows the final accuracy of the agent on 0-contrast trials (i.e., no sensory evidence) as a function of α for all 3 teachers. We can see that brain and Bayes are statistically indistinguishable across the mid-range of α. The largest difference at $\alpha = 3$ is 0.0034 ± 0.0017 (paired, $p = 0.095$). And at $\alpha = 30$ Bayes is ahead of brain (-0.0039 ± 0.0006, $p = 0.0005$). Against shuffle the brain teacher holds a larger and growing margin (0.0124 ± 0.0016, 0.0223 ± 0.0032 and 0.0397 ± 0.0133 at $\alpha = 1, 3$ and 10; $p \leq 0.02$ at every $\alpha \geq 0.3$). Figure 2 (b) shows these differences. At small α all three teachers, including shuffle perform equally well, and all three raise the 0-contrast accuracy by about 0.02 over $\alpha = 0$ baseline (0.0230 ± 0.0031, 0.0235 ± 0.002 and 0.0196 ± 0.0030 at $\alpha = 0.3$). This shows that the higher α advantage over shuffle cannot be attributed to the transfer loss only acting as a regularizer or adding capacity: this effect is present for every teacher and is saturated at $\alpha = 0.3$.

Figure 2 (c) shows the agent's learning curves under the brain teacher across α: stronger alignment both accelerates early learning and raises the final level. Figure 2 (d) depicts it as area under the 0-contrast accuracy learning curve (AULC) for all three teachers. The learning speed of the brain teacher sits above that of Bayes throughout the mid-range of α, although the difference is not significant at any α (largest at $\alpha = 3 : 0.0079 \pm 0.0063$, $p = 0.25$). The shuffle teacher moves the other way: its AULC falls monotonically beyond $\alpha = 0.3$, and its target is fit less well than the brain target at every α (final transfer loss 0.105 vs 0.073 at $\alpha = 30$).

To test whether transfer requires an embedding that is conditioned on the task labels, we ran the experiment with a teacher built from time-conditioned CEBRA of the same recordings. The two builds are based on the same 356 sessions resulting from the correct-only filter, and the same hold-out split, the only difference is the conditioning variable. With this embedding the brain outperforms shuffle under strong alignment, while falling slightly below the Bayesian observer (Figure 3). In both embeddings the gap to shuffle widens with α, and the gap with Bayesian teacher remains flat: near zero for the label-conditioned embedding and about -0.01 for the time-conditioned embedding.

The brain teacher using supervised CEBRA matches the Bayesian observer across the mid-range of α despite receiving only a three-level belief bin and the signed contrast, whereas Bayes has access

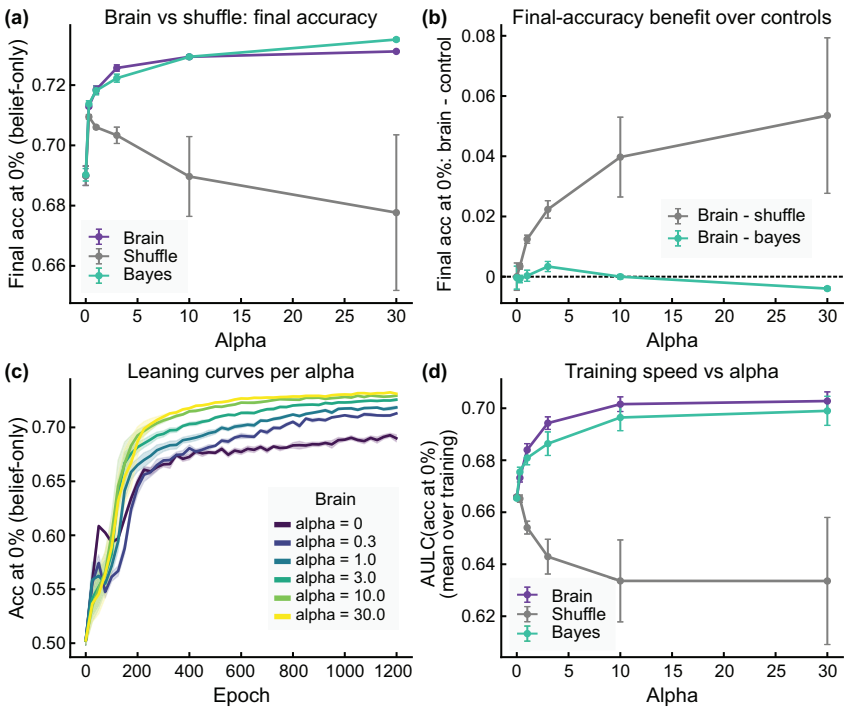

Figure 2: Zero-contrast (belief-only) performance across transfer weight α for the supervised brain teacher (8 seeds per condition, error bars SEM). (a) Final accuracy for the brain, Bayes and shuffle teachers. The brain teacher matches Bayes for $\alpha \leq 10$ and exceeds shuffle at every $\alpha \geq 0.3$. (b) Final-accuracy difference between the brain teacher and each control. (c) Learning curves under the brain teacher. Larger α both accelerates learning and raises the final level. (d) Area under the zero-contrast learning curve. The brain teacher lies slightly above Bayes across the mid-range, not significantly at any α, and the shuffle teacher falls below its own $\alpha = 0$ baseline by $\alpha = 10$.

to p_{high}, the hazard rate and the generative model that the observer is given. It also exceeds the shuffled control, which supplies the same vectors with scrambled labels. Since we have looked at performance on trials where contrast is 0, in which the agent has to rely on prior belief alone, we investigated the representational separability of priors in the brain embedding condition table. We found that the belief separation for the low and high priors is the highest for 0 contrast in the table (Figure 4 (a)).

To test if this is a reliable result rather than just a coincidence we resampled sessions with replacement to see if the result holds. The LOW - HIGH separation shown on Figure 4 (b) is the difference between the separation between the 2 priors at 0 (LOW) and 1 (HIGH) contrasts. We compared 2 tables trained with supervised CEBRA: one is trained on all the trials available in IBL dataset (labeled unfiltered), and the other is the embedding used for the brain teacher on Figure 2 trained on correct trials only. While the separation between LOW and HIGH contrast belief encodings for the first table falls on either side of 0, for the second dataset it is decidedly positive, and the 95% CI does not include 0.

Figure 4 (c) shows the correlation between separation and contrast. We see that the correlation is spread out around zero for the embedding obtained from the unfiltered data, while that from the correct data shows a negative correlation. So the separation between priors is higher for lower contrast, where the mouse needs to rely on the prior more to make a correct choice.

Thus we establish that in the brain embeddings the separation between low and high priors grows as the contrast falls, reaching its maximum at 0 contrast. To further ensure that this is a reproducible result, we created a different CEBRA embedding with the same filtered data using a different seed and ensured that the results still hold.

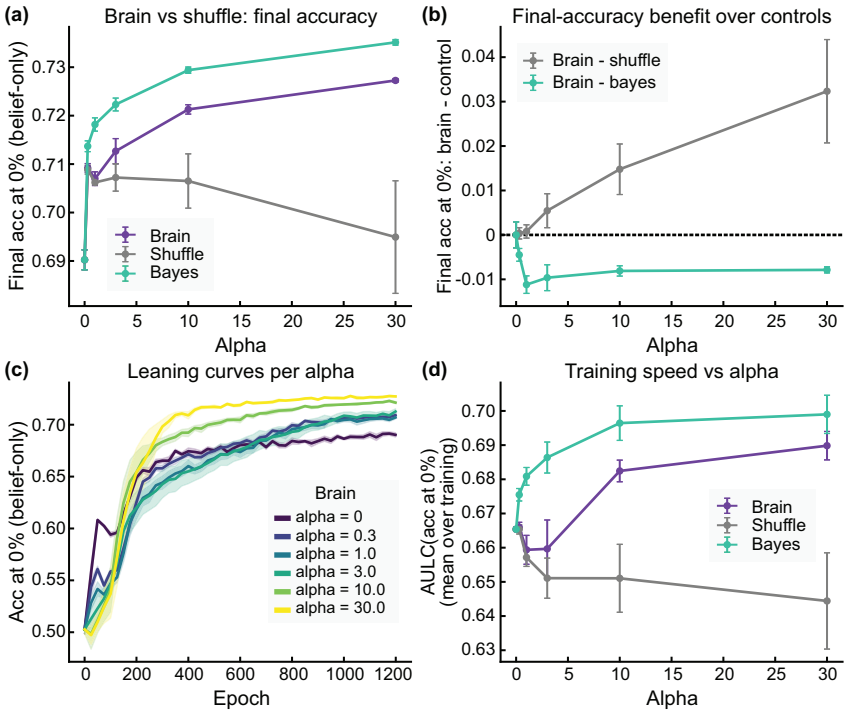

Figure 3: A brain teacher built without task labels still outperforms the shuffled control. (a) Final zero-contrast accuracy for a brain teacher built from the time-conditioned (self-supervised) CEBRA embedding, which never sees condition labels, alongside the Bayes and shuffle teachers. (b) Final-accuracy difference between this brain teacher and each control. The margin over shuffle grows with α and is positive for $\alpha \geq 10$, while the gap to the noiseless Bayes teacher remains near 0.01 throughout. (c) Learning curves at each α. Alignment initially slows learning, with the $\alpha = 0$ curve above all aligned curves for the first few hundred epochs. The $\alpha = 1$ and 3 curves recover later in training, while $\alpha \geq 10$ overtake the baseline sooner and stay ahead for the remainder of training. Final accuracy rises monotonically with α. (d) Area under the zero-contrast learning curve. The brain teacher dips at small α and rises sharply between $\alpha = 3$ and $\alpha = 10$, whereas Bayes increases monotonically and shuffle declines throughout.

3 Discussion

In this work we test whether representations recorded across the mouse brain, used only as an alignment target for a belief embedding of an artificial agent can teach an agent to infer a hidden block prior. We show that this can be accomplished with a condition table built from behavior-labeled CEBRA embeddings of recordings from IBL, and that the benefit is specific to the correspondence between embedding and condition, rather than to the target simply being neural-shaped. The brain teacher exceeds the shuffle control on 0-contrast accuracy at every $\alpha \geq 0.3$. Since the shuffle control holds the identical embedding vectors with permuted labels, this gap cannot be attributed to any generic property of targets that are shaped like neural data.

We also compare the brain teacher to optimal Bayes teacher, and find that it matches the performance of this teacher at 0 contrast trials in mid range α, even though the two teachers are not given the same information. The Bayes teacher receives the observer's continuous posterior, computed with $p_{high} = 0.8$, the hazard rate and the generative model of the block structure, whereas the brain teacher receives only a three-level belief bin together with the signed contrast, and everything else it supplies, including the geometry within and across bins, comes from the recorded population activity. We also observe that at small α every teacher, including shuffle, raises 0-contrast accuracy by about 0.02 over the no-transfer baseline. We interpret this gain as caused by regularization rather than what the target encodes.

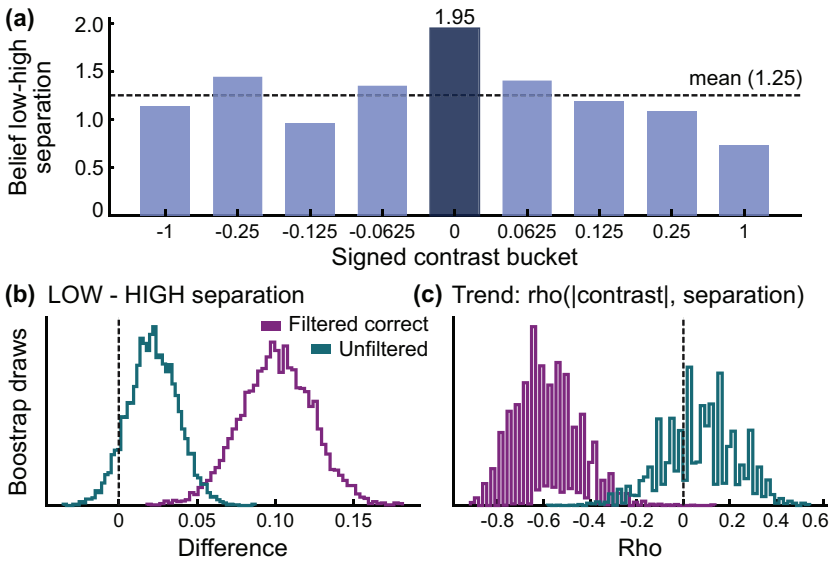

Figure 4: (a) Separation between low- and high-prior embeddings by signed contrast in the correct-trial condition table. Separation is largest at zero contrast, where the mouse must rely on the prior alone. (b) Statistics on the belief separation depending on low vs high contrast. The embedding of the correct trials only shows the separability between beliefs is consistently higher for zero (LOW) contrast than for $|contrast|$ of 1 (HIGH) compared to the control embedding trained on all trials. (c) The correlation between $|contrast|$ and separation is consistently negative for the embedding trained on correct trials only compared the that trained on all the trials. Dashed line marks the no effect line, while the unfiltered run goes right through it, the correct only run is clear of it.

We build condition tables from two CEBRA embeddings of the same recordings, one conditioned on time, and the other on behavior. In this work we report the behavior-conditioned (supervised) fit as primary since belief accounts for 2.21% of condition variance in that fit, compared to 0.21% of the time-conditioned fit. This criterion was measured on the condition table alone, without reference to the agent's performance. Thus the scope of the claims in this paper is limited by the fact the behavior-conditioned embedding is supervised on the same 27 condition labels the teacher later looks up. However, to argue against the natural objection that the transfer result is created by this supervision we present the two following features. First, as explained in the methods section the table is built entirely out of sample: CEBRA is fit on 80% of each session's contiguous 50-trial chunks, and each table entry is the mean embedding of the remaining trials which the fit never saw. The ability to recover the belief geometry on unseen block segments indicates that the information is present in the population activity. Second, in Figure 3 we show that the effect does not require label conditioning. Repeating the experiment with a teacher that uses the time-conditioned embeddings table, which never sees the condition labels, the brain teacher still outperforms shuffle although at larger αs. The margin is smaller here, and the performance is lower than that of Bayes teacher, but the fact that the method works on an embedding built with no knowledge of the task conditions at all shows that it can exploit the modest belief structure present in the representation of an embedding that was not built to carry it. All the supervision determines is which subspace of the activity the embedding expresses.

We emphasize that unlike the Bayes teacher, the brain teacher gets no explicit generative model beyond a coarse belief-bin index for an ideal observer, and this role is filled by the recorded representations. We interpret this as evidence that neural data carries transferable inductive bias that can serve as a substitute for the posterior on a task, where the relevant variable is latent and must be inferred across trials.

The separability analysis in Figure 4 offers one account of why the brain teacher helps where it does. In the correct-trial embedding, the separation between low- and high-prior states grows as contrast falls and is largest at zero contrast, which is consistent with the broad representation of

prior expectations reported across these recordings (Findling et al., 2025). This pattern should be read with one caveat. Because the table is built from correct trials only, and because the correct choice at zero contrast follows the rewarded side, the two belief bins in the zero-contrast cell differ in choice as well as in belief, whereas in every nonzero-contrast cell the correct choice is fixed by the stimulus. The elevated separation at zero contrast may therefore include choice-related activity within the analysis window in addition to prior-related activity.

Several limitations should be noted. Since in this work we are particularly interested in the role of the brain teacher in learning latent variables, we have kept the scope deliberately narrow. The hidden prior we are interested in can only be measured on 0-contrast trials in biased blocks, where sensory evidence is not available and the agent can only make its best guess based on a prior. All results also come from a single task and dataset, and the effect sizes on zero-contrast trials are on the order of a few percentage points. While we did succeed in improving model belief accuracy with brain fine-tuning, we incurred a trade-off: as α rises, the accuracy on non-zero-contrast trials falls, indicating that evidence-driven accuracy is traded for prior-driven accuracy. Improving this by either gating the transfer term on evidence reliability, using a teacher that encodes both variables, or aligning only a subspace of the embedding with the teacher lies outside the scope of this paper and is left as future work.

4 METHODS

4.1 THE TASK AND THE AGENT

The simulated task was deliberately matched to the IBL 2AFC protocol (International Brain Laboratory et al., 2021), so that the labels used to build the neural embeddings for each condition and those produced by the agent's environment can be matched one-to-one. Within a block, a stimulus is more likely to be found at the biased side with probability 0.8, blocks alternate between the 2 biases, and the length of each block is drawn from a geometric variate of mean 50 clipped to [20, 80]. This is different from IBL's truncated exponential with a mean of 51 on [20,100]; the clipping binds on 54% of draws, so the realized mean block length is 43.4 trials (median 35). In the simulated environment the episodes open with 20 unbiased (50/50) trials as opposed to the 90 used in IBL and run for 100 trials. On each trial the stimulus side and absolute contrast are drawn from the current blocks environment and the uniform contrast distribution $\{0, 0.0625, 0.125, 0.25, 1\}$ respectively. The stimulus side is coded by adding a sign to contrast (positive for right, negative for left). The signed contrast is then corrupted by additive Gaussian noise $\mathcal{N}(0, 0.6)$ resulting in a scalar evidence value, which is then fed to the agent.

The agent received a 4-dimensional observation: the signed, noise-corrupted evidence, a two-unit one hot encoding of its previous action, and the previous trials reward. The block state is not passed to the agent and has to be inferred. Because the combination of the agent's previous action and the reward reveal the correct side, the agent has the same information as the Bayesian observer described below. The observation is passed to a single layer GRU consisting of 256 hidden units (Figure 1 (b)). Dropout (p=0.3) is applied during training, and the value head receives only the hidden state of the GRU. This 256-size hidden state is also mapped linearly to an 8-dimensional belief embedding and the policy is a linear map from this embedding to two choice logits. Thus the policy has no direct access to the hidden state of the recurrent network and has to go through this 8-dimensional embedding. Meanwhile the value head bypasses this bottleneck so that the transfer loss constrains only the policy. The teacher acts only through the training objective: the transfer loss is the mean squared distance between the belief embedding and the teacher's embedding for the current condition. The total loss is the sum of the task loss and a weighted transfer loss: $L_{total} = L_{task} + \alpha * L_{transfer}$, where $L_{task} = L_{policy} + 0.5 * L_{value} - 0.01 * entropy$. The total gradient is clipped to 1.0 and the step is taken using Adam optimizer (lr=0.001). Each agent is trained for 1200 epochs. Each epoch is a fresh batch of 64 episodes of 100 trials (64000 trials), drawn with a per-epoch environemnt seed. Final accuracy is computed on 256 simulated episodes generated at a fixed evaluation seed, using the single evaluationat the last epoch (1199) with dropout disabled. Zero-contrast accuracy is the fraction of zero-contrast trials in biased blocks on which the chosen side matches the rewarded side. AULC is the trapezoidal area under the zero-contrast accuracy curve across 49 evaluation checkpoints (every 25 epochs), normalized by the epoch span. Each teacher

and each value of α is run with 8 random seeds. Comparisons between teachers at a given α use two-sided paired t-tests across seeds, and all error bars show SEM across seeds.

4.2 THE THREE TEACHERS.

To construct the teachers we first use a Bayesian observer which calculates posterior belief about the current block. The Bayesian observer is a two-state hidden Markov filter over block identity: on each trial it updates the posterior over {left-biased, right-biased} from the current side, assuming the true stimulus probability $p_{high} = 0.8$ and a constant switch hazard of 1/50 per trial. It then emits the posterior mean P(right) as a scalar belief.

For all three teachers the target embeddings are based on quantities withheld from the agent: the Bayesian observer's posterior, computed from the true stimulus sequence with a generative model of the block structure, and (for the brain and shuffle teachers) the noiseless signed contrast, which the agent observes only through sensory noise. This supervision from privileged information is only present during training: the teacher is not consulted at test time and does not appear anywhere in the forward pass, so the trained agent must reconstruct any such structure from its own observations.

All three teachers are placed on a common scale before training. This is important because the transfer term is a squared distance: multiplying a teacher's targets by c multiplies it by approximately c^2, which is indistinguishable from multiplying α by c^2. The brain and shuffle condition tables are centered and rescaled so that the root-mean-square target norm across the 27 conditions equals a fixed value (1.0 throughout). The Bayes teacher has no table, so its gain is set by the same average applied to its own targets: $w_0 = 1/RMS_{cond}$, where RMS_{cond} is the root mean square of $2*(< b > -0.5)$ across the 27 condition means and $< b >$ is the mean posterior within a condition. RMS_{cond} is computed from the task rather than stored, so it tracks p_{high}, block length and episode length; under our settings $RMS_{cond} = 0.4130$ and $w_0 = 2.4210$. Measured on held-out simulated episodes, all three teachers then sit at a per-condition target RMS of 1.00 (brain 1.0000, shuffle 1.0000, Bayes 0.9992), and so we can compare teachers at the same αs.

Bayes teacher: This teacher encodes only the ideal-observer belief b: target $= 2(b - 0.5)\mathbf{w}$, on dimension 0 only. It has no contrast or evidence component, so at every contrast its target is identical, and b is continuous not binned. The choice of axis is irrelevant since the transfer loss is rotation-invariant and the belief layer is learnable.

Brain teacher: The brain teacher is a lookup table from trial condition to a mean CEBRA embedding of IBL neural activity, with one entry per condition. The CEBRA embeddings were obtained by first discarding all the trials where the mouse made the incorrect choice then dividing each session into fit and holdout trials, holding out one fifth of its contiguous 50-trial chunks (23.5% of trials in practice), because Bayesian belief is autocorrelated at 0.95 and a randomly held-out trial would sit between two fit trials of the same block. The CEBRA model was fit on the train fit trials only, conditioned on 27 discrete condition labels (3 beliefs x 9 contrasts), and the table was then built from the held-out trials, with population activity averaged over -0.3 to +0.2s relative to stimulus onset. Because CEBRA's multi-session sampler draws each positive sample from another session carrying the same class, all train sessions must contain all 27 conditions. Thus 103 sessions were excluded because after filtering for correct only trials they were missing at least one condition, leaving 356 sessions out of 459.

Shuffle teacher: The shuffle teacher uses the same brain table with condition to embedding assignment randomly permuted. The permutation is applied after centering and rescaling, so the two teachers hold an identical multiset of embedding vectors with the same scale, covariance and marginal distribution, and the same per-condition RMS; only the correspondence between condition and embedding differs. The permutation is drawn uniformly from the run's seed and redrawn for each seed, so the brain teacher is tested against a distribution of label shuffles rather than one arbitrary permutation. Being uniform rather than a derangement, the permutation leaves the exact condition correctly bound in 1 of 27 cells on average and the belief bin alone in 9 of 27. The latter carries no usable signal, since the retained cell receives an embedding from a different contrast within its bin and the belief axis is not consistent across contrasts.

AI USE STATEMENT

In this work, we used generative AI tools for helping develop conceptual frameworks and implementing methods. We have not used generative AI tools to generate synthetic data sets, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, propose or refine hypotheses, design or provide feedback on research methodology or experiments, support qualitative and thematic data analysis, or interpret results, and assisting with translation, cleaning and reformatting datasets are not applicable to this work. Additionally, we used generative AI tools for editing software code and brainstorming how to refine initial drafts of artificial belief agent architectures. We have reviewed all AI-assisted work. LLM-generated code was verified and tested for correctness by 2 authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

We do not report any ethics concerns for this submission. We utilize an open dataset which was originally collected in accordance with the IRB guidelines of the International Brain Laboratory.

REPRODUCIBILITY STATEMENT

We have taken care to make our results reproducible from the paper alone. All experiments use the publicly available IBL Brain-wide Map dataset (Angelaki et al., 2025); Section 4.2 specifies the session-inclusion criterion (356 of 459 sessions retaining all 27 conditions after filtering to correct trials), the trial window over which population activity is averaged (-0.3 to +0.2 s relative to stimulus onset), and the chunked fit/hold-out split used to build the condition tables out of sample. The simulated task is fully specified in Section 4.1 ($p_{high} = 0.8$, block lengths drawn from a geometric distribution of mean 50 clipped to [20, 80], 20 unbiased trials per 100-trial episode, the contrast set, and the N(0, 0.6) evidence noise), as is the agent architecture (Figure 1 (b): a single-layer GRU with 256 hidden units, dropout p = 0.3, a linear 8-dimensional belief embedding, and linear value and policy heads) and the training scheme. Section 4.2 gives the construction of all three teachers, including the Bayesian observer's parameters, the CEBRA conditioning scheme (3 belief bins × 9 signed-contrast buckets), and the common rescaling procedure that makes teachers comparable at equal α. CEBRA was run with output dimension 8 to match the agent's belief embedding, using the offset10-model architecture, 20000 iterations, batch size 32, learning rate 0.0003, and temperature initialized at 1.0 with a floor of 0.1, using the multi-session sampler across sessions. Belief bins were defined by thresholding the Bayesian observer's posterior $P(\text{right})$ at 0.35 and 0.65. Population activity was taken from all recorded units in each session, pooled across brain regions, with 142 units per session. Each condition is run across 8 random seeds, with the shuffle permutation redrawn per seed. Source code cannot be released at submission because the implementation is subject to a pending intellectual-property filing. The neural data are public, CEBRA is open source, and every quantity needed to re-implement the environment, agent, teachers and training loop is specified in Sections 4.1 and 4.2. The three condition tables and the simulated environment will be made available upon request under an appropriate use agreement, and we intend to release the full code once the filing permits.

REFERENCES

Dora Angelaki, Brandon Benson, Julius Benson, Daniel Birman, Niccolò Bonacchi, Kcénia Bougrova, Sebastian A Bruijns, Matteo Carandini, Joana A Catarino, et al. A brain-wide map of neural activity during complex behaviour. *Nature*, 645(8079):177–191, 2025.

Tomas Gallo Aquino, Victoria Liu, Habiba Azab, Raissa Mathura, Andrew J Watrous, Eleonora Bartoli, Benjamin Y Hayden, Paul Sajda, Sameer A Sheth, and Nuttida Rungratsameetaweemana. Brain2model transfer: Training sensory and decision models with human neural activity as a teacher. *arXiv preprint arXiv:2506.20834*, 2025.

Rosa Cao and Daniel Yamins. Explanatory models in neuroscience, part 2: Functional intelligibility and the contravariance principle. *Cognitive Systems Research*, 85:101200, 2024.

Mark M Churchland, John P Cunningham, Matthew T Kaufman, Justin D Foster, Paul Nuyujukian, Stephen I Ryu, and Krishna V Shenoy. Neural population dynamics during reaching. *Nature*, 487 (7405):51–56, 2012.

John P Cunningham and Byron M Yu. Dimensionality reduction for large-scale neural recordings. *Nature neuroscience*, 17(11):1500–1509, 2014.

Joel Dapello, Kohitij Kar, Martin Schrimpf, Robert Geary, Michael Ferguson, David D Cox, and James J DiCarlo. Aligning model and macaque inferior temporal cortex representations improves model-to-human behavioral alignment and adversarial robustness. *BioRxiv*, pp. 2022–07, 2022.

Callie Federer, Haoyan Xu, Alona Fyshe, and Joel Zylberberg. Improved object recognition using neural networks trained to mimic the brain's statistical properties. *Neural Networks*, 131:103–114, 2020.

Charles Findling, Felix Hubert, International Brain Laboratory, Luigi Acerbi, Brandon Benson, Julius Benson, Daniel Birman, Niccolò Bonacchi, E Kelly Buchanan, Sebastian Bruijns, et al. Brain-wide representations of prior information in mouse decision-making. *Nature*, 645(8079): 192–200, 2025.

Ruth C Fong, Walter J Scheirer, and David D Cox. Using human brain activity to guide machine learning. *Scientific reports*, 8(1):5397, 2018.

Juan A Gallego, Matthew G Perich, Lee E Miller, and Sara A Solla. Neural manifolds for the control of movement. *Neuron*, 94(5):978–984, 2017.

Jianping Gou, Baosheng Yu, Stephen J Maybank, and Dacheng Tao. Knowledge distillation: A survey. *International journal of computer vision*, 129(6):1789–1819, 2021.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

Eghbal Hosseini, Colton Casto, Noga Zaslavsky, Colin Conwell, Mark Richardson, and Evelina Fedorenko. Universality of representation in biological and artificial neural networks. *bioRxiv*, 2024.

Minyoung Huh, Brian Cheung, Tongzhou Wang, and Phillip Isola. The platonic representation hypothesis. *arXiv preprint arXiv:2405.07987*, 2024.

International Brain Laboratory, Valeria Aguillon-Rodriguez, Dora Angelaki, Hannah Bayer, Niccolo Bonacchi, Matteo Carandini, Fanny Cazettes, Gaelle Chapuis, Anne K Churchland, Yang Dan, et al. Standardized and reproducible measurement of decision-making in mice. *Elife*, 10:e63711, 2021.

Christopher Langdon, Mikhail Genkin, and Tatiana A Engel. A unifying perspective on neural manifolds and circuits for cognition. *Nature Reviews Neuroscience*, 24(6):363–377, 2023.

Zhe Li, Wieland Brendel, Edgar Walker, Erick Cobos, Taliah Muhammad, Jacob Reimer, Matthias Bethge, Fabian Sinz, Zachary Pitkow, and Andreas Tolias. Learning from brains how to regularize machines. *Advances in neural information processing systems*, 32, 2019.

Valerio Mante, David Sussillo, Krishna V Shenoy, and William T Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *nature*, 503(7474):78–84, 2013.

Omer Moussa, Dietrich Klakow, and Mariya Toneva. Improving semantic understanding in speech language models via brain-tuning. In *International conference on learning representations*, volume 2025, pp. 59823–59849, 2025.

Anuja Negi, Subba Reddy Oota, Anwar O Nunez-Elizalde, Manish Gupta, and Fatma Deniz. Brain-informed fine-tuning for improved multilingual understanding in language models. *bioRxiv*, pp. 2025–07, 2025.

A Romero, N Ballas, S Ebrahimi Kahou, A Chassang, C Gatta, and Y Bengio. Fitnets: Hints for thin deep nets. arxiv 2014. *arXiv preprint arXiv:1412.6550*, 10, 2014.

Steffen Schneider, Jin Hwa Lee, and Mackenzie Weygandt Mathis. Learnable latent embeddings for joint behavioural and neural analysis. *Nature*, 617(7960):360–368, 2023.

Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. In *European conference on computer vision*, pp. 776–794. Springer, 2020.

Mingqing Xiao, Kai Du, and Zhouchen Lin. Beyond representational alignment with brain-guided language models for robust reasoning. *Nature Machine Intelligence*, pp. 1–15, 2026.

Dan Yamins and Aran Nayebi. Contravariance theory: Strong alignment for minimal solutions to hard tasks. *arXiv preprint arXiv:2607.08561*, 2026.