

Epistemic Neural Networks

Ian Osband*, Zheng Wen, Seyed Mohammad Asghari,
Vikranth Dwaracherla, Morteza Ibrahimi, Xiuyuan Lu, and Benjamin Van Roy
DeepMind, Efficient Agent Team, Mountain View

Abstract

Intelligence relies on an agent’s knowledge of what it does not know. This capability can be assessed based on the quality of *joint* predictions of labels across multiple inputs. Conventional neural networks lack this capability and, since most research has focused on marginal predictions, this shortcoming has been largely overlooked. We introduce the *epistemic neural network* (ENN) as an interface for models that represent uncertainty as required to generate useful joint predictions. While prior approaches to uncertainty modeling such as Bayesian neural networks can be expressed as ENNs, this new interface facilitates comparison of joint predictions and the design of novel architectures and algorithms. In particular, we introduce the *epinet*: an architecture that can supplement any conventional neural network, including large pretrained models, and can be trained with modest incremental computation to estimate uncertainty. With an epinet, conventional neural networks outperform very large ensembles, consisting of hundreds or more particles, with orders of magnitude less computation. We demonstrate this efficacy across synthetic data, ImageNet, and some reinforcement learning tasks. As part of this effort we open-source experiment code.

1 Introduction

Consider a conventional neural network trained to predict whether a random person would classify a drawing as either a ‘rabbit’ or a ‘duck’. As illustrated in Figure 1, given a single drawing, the network generates a marginal prediction that assigns probabilities to the two classes. If the probabilities are each 0.5, it remains unclear whether this is because opinions across the population are equally divided or the neural network would learn a definitive class if subsequently trained on additional data. These two possibilities would be distinguished by a joint prediction across two identical inputs. As illustrated in tables to the right of the neural network output, such a joint prediction takes the form of a two-by-two table. If opinions are divided due to ambiguity of the image, probabilities would be equally distributed across the four cells. On the other hand, if uncertainty would be completely resolved with further training, the probability of labels differing across identical images is zero.

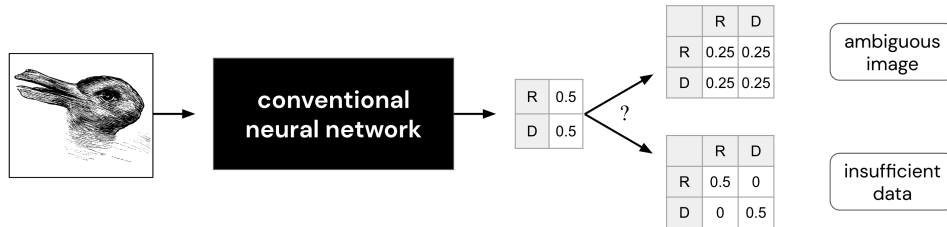


Figure 1: Conventional neural networks generate marginal predictions, which do not distinguish genuine ambiguity from insufficiency of data.

*Contact iosband@deepmind.com.

This example indicates that joint predictions express the extent to which uncertainty will be resolved by additional data, whereas marginal predictions do not. Understanding whether and how uncertainty can be resolved is critical to a number of intelligent behaviors such as efficient exploration and adaptation. As elucidated by Wen et al. (2022), such challenges call for the ability to produce accurate joint predictions.

In this paper, we introduce the *epistemic neural network* (ENN) – an interface for models that represent uncertainty as required to generate useful joint predictions. While prior approaches to uncertainty modeling such as Bayesian neural networks can be expressed as ENNs, the new interface facilitates assessment and comparison of joint predictions and the design of novel architectures and algorithms. In particular, we introduce the *epinet*: an architecture that can supplement any conventional neural network to estimate uncertainty more effectively than state-of-the-art approaches. For example, with an epinet, conventional neural networks can outperform very large ‘deep ensembles’ (Lakshminarayanan et al., 2017), consisting of hundreds or more particles, with orders of magnitude less computation.

Figure 2 offers a preview of results of Section 3.2, where we compare these approaches on ImageNet, evaluating each model on out-of-sample data after training. The quality of the ResNet’s marginal predictions – measured in terms of classification error or marginal log-loss – does not change much if supplemented with an epinet. However, as a function of the number of model parameters, the epinet-enhanced ResNet dramatically reduces joint log-loss.

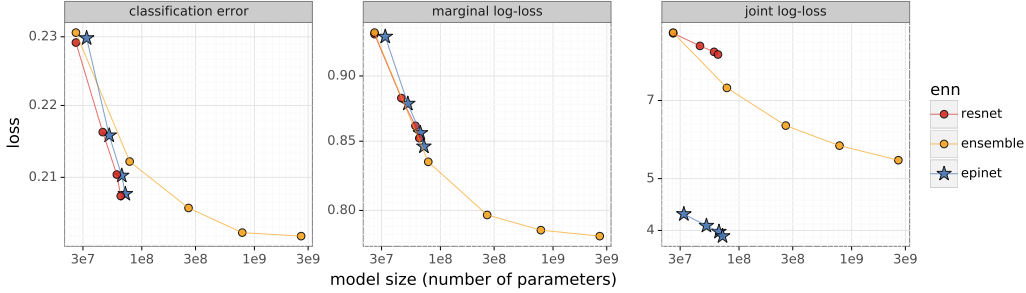


Figure 2: Quality of marginal and joint predictions across models on ImageNet (Section 3.2).

1.1 Epistemic neural networks

Given an input x and parameters θ , a conventional NN produces an output $f_\theta(x)$. The output $f_\theta(x, z)$ of an ENN depends additionally on an *epistemic index* z . An ENN specifies a fixed reference distribution P_Z for the epistemic index. Typical choices include a uniform distribution over a finite set or a standard Gaussian over a vector space. The index z is used to express epistemic uncertainty. In particular, variation of the network output with z indicates uncertainty that might be resolved by future data. As we will see, the introduction of an epistemic index allows us to represent the kind of uncertainty required to generate useful joint predictions.

In typical classification settings, a conventional neural network produces a marginal prediction that assigns a probability $\text{softmax}(f_\theta(x))_y$ to each class y . A joint prediction across inputs x_1, \dots, x_τ would assign a probability to each class combination y_1, \dots, y_τ . While conventional neural networks are not designed to provide joint predictions, joint predictions can be produced by multiplying marginal predictions:

$$\prod_{t=1}^{\tau} \text{softmax}(f_\theta(x_t))_{y_t}. \quad (1)$$

However, as discussed earlier, such use of marginal predictions fails to distinguish ambiguity from insufficiency of data. ENNs address this by enabling more expressive joint predictions through integrating over epistemic indices:

$$\int_z P_Z(dz) \prod_{t=1}^{\tau} \text{softmax}(f_\theta(x_t, z))_{y_t}. \quad (2)$$

This integration introduces dependencies so that joint predictions are not necessarily just the product of marginals. Figure 3 illustrates the difference between conventional and epistemic neural network architectures.

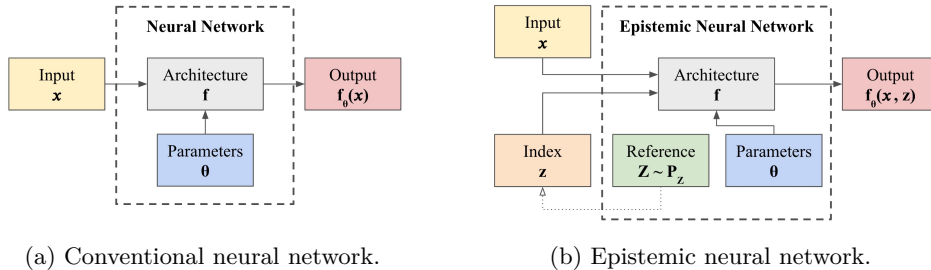


Figure 3: An ENN samples an epistemic index Z from a reference distribution P_Z . This enables expression of joint predictions beyond the product of marginal predictions.

Although the ENN framing and notation are new, all existing approaches to uncertainty estimation can be expressed as ENNs. In particular, any Bayesian neural network (BNN) can easily be written as an ENN. Where a conventional NN produces output $f_\theta(x)$, a BNN then maintains a posterior distribution over weights θ (Neal, 2012). For any posterior distribution, we can define reference distribution P_Z and parameterized class of functions $g_{\theta'}$ such that a prediction $f_{g_{\theta'}(Z)}(x)$, with Z sampled from P_Z , reflects predictive uncertainty equivalent to that of the original BNN. Concretely, a ‘deep ensemble’ can be expressed as an ENN by taking Z to index particles of the ensemble and the reference distribution P_Z to be uniform across these indices. For Monte Carlo dropout, we could take Z to identify the dropout mask (Gal, 2016). See Appendix B for full details and link to open-source code.

While ENNs and BNNs are closely related, there are some important differences. Where BNNs typically view neural network weights as unknown parameters to be inferred, ENNs focus on the resultant predictive distributions. Though ENNs maintain parameters and adapt them to data, the approach ascribes no semantic meaning to them. The sole role of ENN parameters is in driving effective predictions. While all BNNs can be naturally framed as ENNs, it is not the case that all ENNs are natural to describe as BNNs. The epinet, which we will introduce in Section 2, serves as an example of such an ENN.

1.2 Contributions

We introduce epistemic neural networks (ENNs), and we view this new perspective as an important contribution of this paper. This new interface brings focus to joint predictions rather than inference of network weight distributions, which is emphasized in Bayesian deep learning. Only data is *real*, and as such, joint predictions of unobserved data capture all relevant uncertainty. While the Bayesian deep learning literature views weights as unknowns to be inferred, ENNs attribute no semantic meaning to weights and treat them only as artifacts that facilitate computation. This ENN perspective sidesteps unnecessary restrictions that pose challenges for Bayesian deep learning algorithms.

While all BNNs can be expressed as ENNs, we develop an ENN that is not naturally expressed as a BNN but can produce accurate joint predictions with orders of magnitude less compute than BNN approaches. In Section 2, **we introduce the epinet: an ENN architecture that can supplement any conventional NN and be trained to estimate uncertainty**. At its core, the epinet training procedure uses randomized prior functions to generate approximate posterior samples (Osband et al., 2018). However, unlike previous work, which generates multiple samples independently via large ensembles, the epinet can be trained with modest incremental computation.

In the remainder of our paper, **we demonstrate the efficacy of the epinet across a range of experiments**. Section 3 shows that an epinet can match the performances of state-of-the-art approaches to joint predictions at orders of magnitude lower computational cost. These results are consistent across data generated by random neural networks (Section 3.1)

and large scale image benchmarks (Section 3.2). Section 4 shows that these improvements in the quality of joint predictions drive improved performance in some reinforcement learning tasks. We show that agents using an epinet can outperform existing BNN approaches at much lower computational cost. These results are consistent across tasks generated by random neural networks (Section 4.2) and a reinforcement learning benchmark (Section 4.3).

As part of this research effort, **we release the core open-source code necessary to reproduce our experimental results in Appendix A**. This includes an optimized Haiku-based library for epistemic neural networks, as well as high-quality implementations of baselines and experiments (Bradbury et al., 2018; Babuschkin et al., 2020). These research tools should enable the community to contribute to ENN technology and together make advances in uncertainty modeling.

1.3 Related work

Our work builds on the subject of Bayesian neural networks (BNNs) (MacKay, 1992; Neal, 2012), which has produced a rich literature. Research in this area has developed and analyzed methods that estimate uncertainty (Welling and Teh, 2011; Blundell et al., 2015; Mandt et al., 2018) and highlighted differing roles and interpretations of epistemic and aleatoric uncertainty (Der Kiureghian and Ditlevsen, 2009; Kendall and Gal, 2017).

Our work is motivated by the importance of joint predictions in driving decision, exploration, and adaptation (Wang et al., 2021; Osband et al., 2022a; Wen et al., 2022). Our experimental results rely on dyadic sampling to evaluate the quality of joint predictions in high dimensions (Osband et al., 2022b). There is a long history of research in reinforcement learning that develops representations that aim to track what the agent knows or does not know (see, e.g., Li et al., 2011; Lu et al., 2021). ENNs facilitate comparison and assessment of such representations, without worrying about ‘whether XYZ is Bayesian’ (Izmailov et al., 2021).

2 The epinet

This section introduces the *epinet*, which can supplement any conventional neural network to produce an effective ENN. Our approach reuses standard deep learning blocks and training algorithms. As we will see, it is straightforward to add an epinet to any existing model, even one that has been pretrained.

2.1 Architecture

Consider a conventional neural network, which we will refer to as the *base network*. Given base parameters ζ and an input x , denote the output by $\mu_\zeta(x)$. If this is a classification model, the associated class probabilities would be $\text{softmax}(\mu_\zeta(x))$.

An epinet is a neural network with privileged access to inputs and outputs of activation units in the base network. A subset of these inputs and outputs, denoted by $\phi_\zeta(x)$, are taken as input to the epinet along with an epistemic index z . For epinet parameters η , the epinet outputs $\sigma_\eta(\phi_\zeta(x), z)$. To produce an ENN, the output of the epinet is added to that of the base network, though with a “stop gradient”:

$$\underbrace{f_\theta(x, z)}_{\text{ENN}} = \underbrace{\mu_\zeta(x)}_{\text{base net}} + \underbrace{\sigma_\eta(\text{sg}[\phi_\zeta(x)], z)}_{\text{epinet}}. \quad (3)$$

The ENN parameters $\theta = (\zeta, \eta)$ include those of the base network and epinet. The “stop gradient” notation $\text{sg}[\cdot]$ indicates that when computing a gradient the argument is treated as fixed. For example,

$$\nabla_\theta f_\theta(x, z) = \begin{bmatrix} \nabla_\zeta \mu_\zeta(x) \\ \nabla_\eta \sigma_\eta(\phi_\zeta(x), z) \end{bmatrix}.$$

The epinet used to produce the ImageNet results of Figure 2 serves as an illustrative example. In that case, the base network is a ResNet and the input $\phi_\zeta(x)$ of the epinet is a concatenation of the ResNet’s input x and last-layer features. Each epistemic index is a 30-dimensional vector, and its reference distribution is standard normal.

Before training, variation of the ENN output $f_\theta(x, z)$ as a function of z reflects prior uncertainty. Since the base network does not depend on z , this variation must derive from the epinet. In our experiments, we induce this initial variation using *prior networks* (Osband et al., 2018). In particular, for $\tilde{x} = \text{sg}[\phi_\zeta(x)]$, our epinets take the form

$$\underbrace{\sigma_\eta(\tilde{x}, z)}_{\text{epinet}} = \underbrace{\sigma_\eta^L(\tilde{x}, z)}_{\text{learnable}} + \underbrace{\sigma^P(\tilde{x}, z)}_{\text{prior net}}. \quad (4)$$

The prior network σ^P represents prior uncertainty and has no trainable parameters. In our experiments, prior networks are suitably architected neural networks with weights randomly sampled such that outputs exhibit appropriate variation with the epistemic index z . For all the experiments in this paper, we use a simple MLP architecture for the learnable component σ_η^L . In general, the learnable component σ_η^L is designed to initially output near-zero values. After training on data, the resultant sum σ_η represents an estimate of posterior uncertainty.

2.2 Training

We describe in this section how we train the ENN in supervised learning tasks. In Section 4.3, we discuss an extension to reinforcement learning. Given a set \mathcal{D} of data pairs, consider a loss function

$$\mathcal{L}(\theta, \mathcal{D}) = \int_z P_Z(dz) \left(\overbrace{\sum_{(x,y) \in \mathcal{D}} \ell(\theta, x, y, z)}^{\text{data loss}} + \overbrace{\Psi(\theta, z)}^{\text{regularization}} \right). \quad (5)$$

This is similar to loss functions employed in the training of conventional neural networks, but with the addition of an integral over the epistemic index z . Parameters are adapted by stochastic gradient descent, but with the “stop gradient” introduced in Equation (3), which freezes the epinet input when computing gradients with respect to the base network parameters ζ . Each step is based on a minibatch $\tilde{\mathcal{D}} \subset \mathcal{D}$ of data along with a minibatch $\tilde{\mathcal{Z}}$ of indices sampled i.i.d. from the reference distribution P_Z . Each stochastic gradient sample takes the form

$$\nabla_\theta \left(\frac{1}{|\tilde{\mathcal{Z}}|} \sum_{z \in \tilde{\mathcal{Z}}} \left(\frac{|\mathcal{D}|}{|\tilde{\mathcal{D}}|} \sum_{(x,y) \in \tilde{\mathcal{D}}} \ell(\theta, x, y, z) + \Psi(\theta, z) \right) \right). \quad (6)$$

In our experiments, we use standard loss functions that are common to conventional neural network training. In particular, for regression problems, we use mean squared error $\ell^{\text{MSE}}(\theta, x, y, z) = (f_\theta(x, z) - y)^2$. For classification problems, we use the cross-entropy loss $\ell^{\text{XENT}}(\theta, x, y, z) = -\ln(\text{softmax}(f_\theta(x, z))_y)$. In both cases, we use ridge regularization: $\Psi^{L_2}(\theta, z) = \lambda \|\theta\|_2^2$. The coefficient λ is a tunable hyperparameter that weights regularization relative to data loss. It is natural to consider using different coefficients for base network versus epinet parameters. However, that difference did not play a significant role in our experiments and we used a single coefficient.

As elucidated by Wen et al. (2022), when the signal-to-noise ratio of the observed label y varies substantially with the input x , in addition to tuning the regularization coefficient λ , perturbing the data loss function in a manner that depends on the epistemic index z can improve out-of-sample joint predictions. Effective approaches to perturbation based on versions of the statistical bootstrap have been proposed in the literature (Osband and Van Roy, 2015; Lu and Van Roy, 2017). Such perturbations were not helpful in our experiments, however, perhaps because each of our domains did not present substantial variation in signal-to-noise ratios.

One useful feature of the epinet is that it can supplement a pretrained base network and be trained with frozen base network weights. This makes the epinet suitable as an addition to large models that are expensive to train, such as pretrained language models (Brown et al., 2020). Further, producing uncertainty estimates by sampling and evaluating outputs across many epistemic indices does not require multiple forward passes through the base network, which does not depend on the epistemic index, but only the epinet, which can be much smaller and thus economical.

3 Supervised learning

In this section, we study the performance of epinet architectures in supervised learning. We begin with a benchmark that produces synthetic data using a neural-network-based generative model. We then treat image classification benchmarks, including ImageNet. In each setting, we examine the empirical trade-off realized by a variety of agents between statistical loss and computational cost. We find the epinet to be particularly attractive. For example, it attains joint log-loss competitive with very large ensembles that include hundreds or thousands of particles, but with orders of magnitude less computation.

3.1 The neural testbed

The Neural Testbed is an open-source benchmark that evaluates the quality of predictive distributions on classification problems, with synthetic data produced by neural-network-based generative models (Osband et al., 2022a). It serves as a unit test for sanity-checking agents in a controlled environment. Table 1 lists agents that we study and compare as well as hyperparameters that we tune. In our experiments, we optimize these hyperparameters via grid search. We use the open-source github repository https://github.com/deepmind/neural_testbed and will submit our new **epinet** agent to this library.

Table 1: Summary of benchmark agents, full details in Appendix E.

agent	description	hyperparameters
mlp	Vanilla MLP	L_2 decay
ensemble	‘Deep Ensemble’ (Lakshminarayanan et al., 2017)	L_2 decay, ensemble size
dropout	Dropout (Gal and Ghahramani, 2016)	L_2 decay, network, dropout rate
bbb	Bayes by Backprop (Blundell et al., 2015)	prior mixture, network, early stopping
hypermodel	Hypermodel (Dwaracherla et al., 2020)	L_2 decay, prior, bootstrap, index dimension
ensemble+	Ensemble + prior functions (Osband et al., 2018)	L_2 decay, ensemble size, prior scale, bootstrap
sgmcmc	Stochastic Langevin MCMC (Welling and Teh, 2011)	learning rate, prior, momentum
epinet	MLP + MLP epinet (this paper)	L_2 decay, network, prior, index dimension

For our **epinet** agent, we initialize a random base network $\mu_\zeta(x)$ as per the baseline **mlp** agent. We take $\phi_\zeta(x)$ to be the concatenation of x and the last-layer features of the base network. Let C denote the number of classes and D_Z denote the index dimension. The learnable network $\sigma_\eta^L(\phi_\zeta(x), z) = g_\eta([\phi_\zeta(x), z])^T z$, where $g_\eta(\cdot)$ is a 2-layer MLP with outputs in $\mathbb{R}^{D_Z \times C}$, and $[\phi_\zeta(x), z]$ is concatenation of $\phi_\zeta(x)$ and z . The prior network σ^P is an ensemble of D_Z particles sampled from the distribution of the data generating model that act directly on the input x . After tuning, we chose epinet hidden layer widths (15, 15), with an index dimension of 8 and standard Gaussian reference distribution. Full details, together with open source code, are provided in Appendix E.

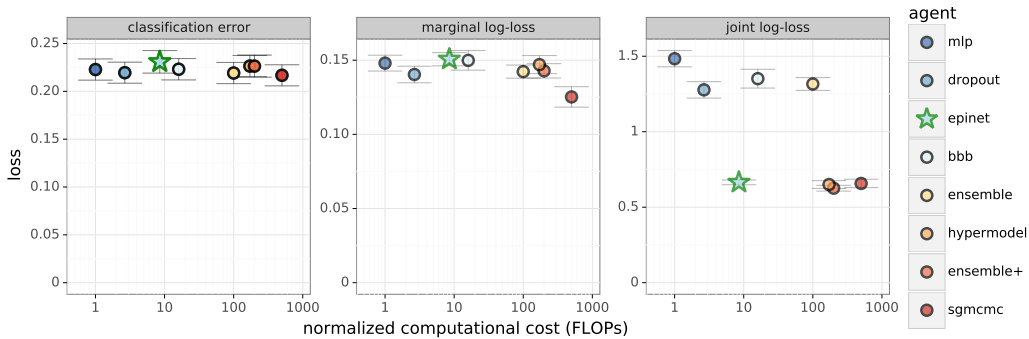


Figure 4: Comparison of agents on the neural testbed.

Figure 4 examines the trade-offs between statistical loss and computational cost for each of the agents after tuning. The bars indicate one standard error, estimated over the testbed’s random seeds. In terms of classification error, most of the agents perform similarly after tuning. Most agents also perform similarly in marginal log-loss, though **sgmcmc** agent reduces loss relative to others at a high computational cost. The key differences in performance come when examining the *joint* log-loss, which we evaluate over ten inputs selected via dyadic sampling (Osband et al. (2022b), Appendix D). In terms of joint log-loss, the **epinet** is competitive with top-performing agents at orders of magnitude lower computational cost.

Our results show that, in terms of joint predictions, the **epinet** can dramatically improve joint predictions and/or dramatically reduce computational cost relative to any of the other agents. These effects are huge, even when compared to popular Bayesian deep learning approaches: **bbb**, **dropout**, and **ensemble**. This highlights a blind-spot in the development of existing agents: the focus has been on marginal rather than joint predictions.

3.2 Image classification

The benefits of the **epinet** scale up to more complex image data. In fact, these benefits become even more substantial. This section focuses on experiments around the ImageNet dataset (Deng et al., 2009), although we produce qualitatively similar results for both CIFAR-10 and CIFAR-100 in Appendix F. This appendix also compares our agents against the *uncertainty baselines* of Nado et al. (2021). Even after tuning for joint log-loss, none of these agents match epinet performance.

For our experiments in this section, we first train several baseline ResNet architectures on ImageNet. In particular, we tune ResNet- L with $L \in \{50, 101, 152, 200\}$ over learning rate, weight decay and temperature rescaling (Wenzel et al., 2020) in the Jaxline framework (Babuschkin et al., 2020). The ensemble agent only uses the ResNet-50 architecture. After tuning hyperparameters, we independently initialize and train 100 ResNet-50 models to serve as ensemble particles. These models are then used to form ensembles of sizes 1, 3, 10, 30, and 100.

For our epinet agent, we take the pretrained ResNet as the base network and freeze its weights. The input to the learnable network σ_η^L includes the last-layer features of the base ResNet and the epistemic index. Similar to Section 3.1, the learnable network $\sigma_\eta^L(\phi, z) = g_\eta([\phi, z])^T z$, where ϕ is the last-layer features, z is the epistemic index, $[\phi, z]$ is the concatenation of ϕ and z , and $g_\eta(\cdot)$ is a 1-layer MLP with 50 hidden units and output $\in \mathbb{R}^{D_z \times C}$ for C classes. The fixed prior σ^P consists of a network with the same architecture and initialization as σ_η^L , together with an ensemble of small random convolutional networks that directly take the image as inputs. We choose the index dimension $D_z = 30$ and P_z as standard Gaussian. We train the epinet using cross-entropy loss with ridge regularization. We push the details on hyperparameters and ablations, together with open-source code, to Appendix F.

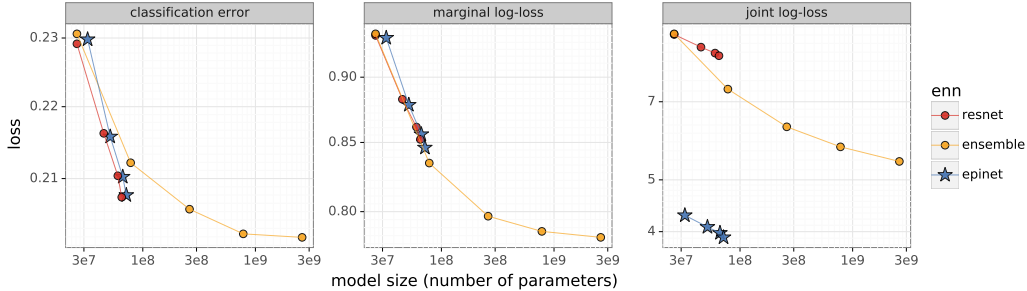


Figure 2: Quality of marginal and joint predictions on ImageNet. (repeated from page 2)

Figure 2 presents a key result of this paper: relative to large ensembles, epinets greatly improve joint predictions at orders of magnitude lower computational cost. Just as in the experiments of Section 3.1, we assess joint predictions over inputs selected by a dyadic sampling heuristic, which we review in Appendix D. The figure plots performance of three agents with respect to three notions of loss as a function of model size in the spirit of Kaplan et al. (2020). For these agents, the qualitative results are unchanged for alternative measures of computational cost such as FLOPs (Appendix F). The first two plots assess marginal prediction performance in terms of classification error and log-loss. Performance of the ResNet scales similarly whether or not supplemented with the epinet. Performance of the ensemble does not scale as well as that of the ResNet. The third plot pertains to performance of joint predictions and exhibits dramatic benefits afforded by the epinet. While joint log-loss incurred by the ensemble agent improves with model size more so than the ResNet, the epinet agent outperforms both alternatives by an enormous margin.

4 Reinforcement learning

Efficient exploration and adaptation in reinforcement learning rely on an agent’s ability to decide whether and how uncertainty can be resolved by additional data. Such abilities, as elaborated in Wen et al. (2022), are highly related to whether an agent can produce accurate joint predictions. In this section, we demonstrate that the improvements in the quality of joint predictions achieved by epinets in Section 3 translate to improved performance in some reinforcement learning tasks.

4.1 Enhanced DQN Agents

Common reinforcement learning agents, such as the deep Q-networks (DQNs) of Mnih et al. (2015), use conventional neural networks to estimate a value function. Such a value function generates a vector of predictions $f_\theta(s)$ as a function of state s . Each component $f_\theta(s)_a$ represents a prediction of discounted future rewards contingent on next executing action a . Use of an epistemic rather than conventional neural network enables an agent to infer whether particular predictions ought to improve with more data.

We fix an algorithm that selects each action based on the ENN used to represent the value function, and we compare performance across ENNs. We use a variant of DQN similar to that proposed in (Osband et al., 2016b), which applies to episodic environments. This approach selects actions using approximate Thompson sampling that draws a random value function at the start of each episode. In particular, the agent samples an epistemic index $Z_k \sim P_Z$ at the beginning of each k th episode. The agent then selects an action A_t that maximizes $f_\theta(S_t, Z_k)_{A_t}$ at each time t within the episode.

Similar to the DQN agent, we store state transition data tuples of the form $(S_t, A_t, R_{t+1}, S_{t+1})$ in a replay buffer, we update the network parameters via stochastic gradient steps, and we use a target network $f_{\theta^{\text{target}}}$ with periodically updated target parameters θ^{target} to stabilize gradients. Each gradient step is based on a minibatch \tilde{D} of samples from the replay buffer and a minibatch \tilde{Z} of epistemic indices drawn i.i.d. from the reference distribution:

$$\nabla_\theta \left(\frac{1}{|\tilde{Z}|} \sum_{z \in \tilde{Z}} \frac{1}{|\tilde{D}|} \sum_{(s,a,r,s') \in \tilde{D}} \left(f_\theta(s, z)_a - r - \gamma \max_{a'} f_{\theta^{\text{target}}}(s', z)_{a'} \right)^2 \right), \quad (7)$$

where $\gamma \in (0, 1)$ is a discount factor. Full details are presented in Appendix G.

4.2 Neural bandit

The neural bandit (Osband et al., 2022a) is an environment where rewards are generated by neural-network-based generating processes. We take the 2-layer MLP generative model from the Neural Testbed (Section 3.1). We consider $N = 1000$ actions, drawn i.i.d. from a 100-dimensional standard normal distribution. At each timestep, the reward of selecting an action a is generated by first forwarding the vector a through the MLP, which gives 2 logit outputs. The reward $\in \{0, 1\}$ is then sampled according to the class probabilities obtained from applying softmax to the logits. Our agents re-use the ENN architectures from Section 3.1 to estimate value functions that predict immediate rewards (i.e. apply discount factor 0). We run the agents for 50,000 timesteps and average results over 30 random seeds. The computational costs of each ENN scale similarly to Section 3.1, see Appendix G.1,

Figure 5 displays the average regret through time attained using different ENNs. We can see that all agents are able to learn from their experience, but the quality of their performance is strongly impacted by the choice of ENN. Epinet achieves the lowest average regret throughout time, while also using much less computation.

The scatter plots of Figure 6 report the correlation between prediction quality on the Neural Testbed and bandit performance, together with the 5th and 95th percentiles obtained by bootstrap resampling. The multiple points for any given agent represent results generated with different random seeds. Agents that produced accurate joint predictions performed well in the neural bandit. However, the quality of marginal predictions showed no strong relation with performance on bandit problems.

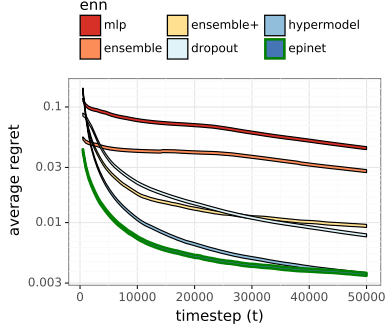


Figure 5: Bandit performance.

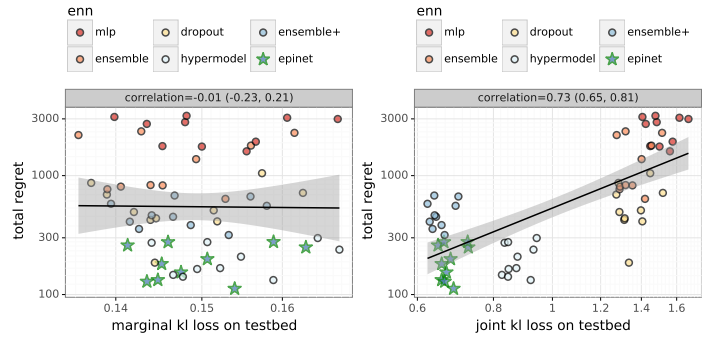


Figure 6: Relating bandit performance to prediction quality.

4.3 Behaviour suite for reinforcement learning

The behaviour suite for reinforcement learning, or **bsuite** for short, is a collection of environments carefully-designed to investigate core capabilities of RL agents (Osband et al., 2020). We repeat our analysis of ENNs applied to these environments. We use the ENNs from Section 3.1 to estimate value functions with discount $\gamma = 0.99$. For all agents using prior functions (**ensemble+**, **hypermodel**, and **epinet**) we scale the value prior to have mean 0 and variance 1 based on the observations over the first 100 timesteps under a random action policy. Full details are available in Appendix G.2.

In **bsuite**, an agent is assigned a score for each experiment. Figures 7 plots the “bsuite loss,” which we define to be one minus the the average score against computational cost. Once again, **epinet** performs similarly with large ensembles, but at orders of magnitude less computational cost. We include a more detailed breakdown of agent performance by competency in Appendix H.

Figure 8 reports the correlation between prediction quality and **bsuite** performance, together with the 5th and 95th percentiles obtained by bootstrap resampling. The multiple points for any given agent represent results generated with different random seeds. The results on mirror those for the neural bandit closely. Agents that produced accurate joint predictions performed well in **bsuite**. However, the quality of marginal predictions showed no strong relation with performance on **bsuite**.

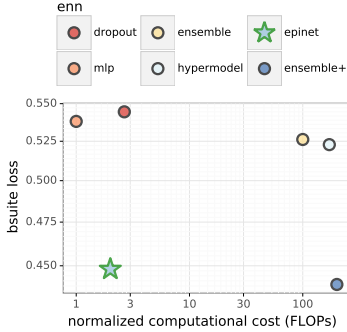


Figure 7: **bsuite** performance.

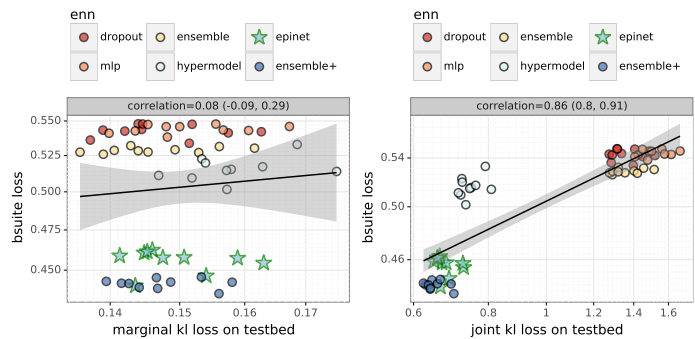


Figure 8: Relating **bsuite** performance to prediction quality.

5 Conclusion

We have introduced the ENN interface for uncertainty modeling. This facilitates design of new ENNs and comparison of joint predictions. We have also introduced the **epinet**, a novel ENN that greatly outperforms alternatives in terms of joint prediction quality, computational requirements, or both. In particular, for large models, the **epinet** enables joint predictions that are significantly better than an ensemble of 100 particles at computation cost marginally exceeding that of a single particle. The **epinet** improves performance across applications ranging from image classification to reinforcement learning.

Acknowledgments and Disclosure of Funding

We thank John Maggs for organization and management of this research effort and Rich Sutton, Yee Whye Teh, Geoffrey Irving, Koray Kavukcuoglu, and Satinder Singh for helpful discussions and feedback. We also extend a special thanks to Mark Collier, Mike Dusenberry, Jeremiah Luh, Balaji Lakshminarayanan and the rest of the Google Brain reliable deep learning team for their help in integrating our research and code with the Uncertainty Baselines project (Nado et al., 2021).

References

- Babuschkin, I., Baumli, K., Bell, A., Bhupatiraju, S., Bruce, J., Buchlovsky, P., Budden, D., Cai, T., Clark, A., Danihelka, I., Fantacci, C., Godwin, J., Jones, C., Hennigan, T., Hessel, M., Kaptunowski, S., Keck, T., Kemaev, I., King, M., Martens, L., Mikulik, V., Norman, T., Quan, J., Papamakarios, G., Ring, R., Ruiz, F., Sanchez, A., Schneider, R., Sezener, E., Spencer, S., Srinivasan, S., Stokowiec, W., and Viola, F. (2020). The DeepMind JAX Ecosystem.
- Blundell, C., Cornebise, J., Kavukcuoglu, K., and Wierstra, D. (2015). Weight uncertainty in neural networks. In *International Conference on Machine Learning*, pages 1613–1622. PMLR.
- Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., and Zhang, Q. (2018). JAX: composable transformations of Python+NumPy programs.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.
- Collier, M., Mustafa, B., Kokiopoulou, E., Jenatton, R., and Berent, J. (2020). A simple probabilistic method for deep classification under input-dependent label noise. *arXiv preprint arXiv:2003.06778*.
- Collier, M., Mustafa, B., Kokiopoulou, E., Jenatton, R., and Berent, J. (2021). Correlated input-dependent label noise in large-scale image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1551–1560.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.
- Der Kiureghian, A. and Ditlevsen, O. (2009). Aleatory or epistemic? does it matter? *Structural safety*, 31(2):105–112.
- Dwaracherla, V., Lu, X., Ibrahimi, M., Osband, I., Wen, Z., and Van Roy, B. (2020). Hypermodels for exploration. In *International Conference on Learning Representations*.
- Gal, Y. (2016). *Uncertainty in Deep Learning*. PhD thesis, University of Cambridge.
- Gal, Y. and Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *International Conference on Machine Learning*.
- Havasi, M., Jenatton, R., Fort, S., Liu, J. Z., Snoek, J., Lakshminarayanan, B., Dai, A. M., and Tran, D. (2020). Training independent subnetworks for robust prediction. *CoRR*, abs/2010.06610.
- Hoffman, M., Shahriari, B., Aslanides, J., Barth-Maron, G., Behbahani, F., Norman, T., Abdolmaleki, A., Cassirer, A., Yang, F., Baumli, K., Henderson, S., Novikov, A., Colmenarejo, S. G., Cabi, S., Gulcehre, C., Paine, T. L., Cowie, A., Wang, Z., Piot, B., and de Freitas, N. (2020). Acme: A research framework for distributed reinforcement learning. *arXiv preprint arXiv:2006.00979*.
- Izmailov, P., Vikram, S., Hoffman, M. D., and Wilson, A. G. (2021). What are Bayesian neural network posteriors really like? *arXiv preprint arXiv:2104.14421*.
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., and Amodei, D. (2020). Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.
- Kendall, A. and Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems*, volume 30.

- Kingma, D. and Ba, J. (2015). Adam: A Method for Stochastic Optimization. *Proceedings of the International Conference on Learning Representations*.
- Krizhevsky, A. (2009). Learning multiple layers of features from tiny images. Technical Report 0, University of Toronto, Toronto, Ontario.
- Lakshminarayanan, B., Pritzel, A., and Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems*, pages 6405–6416.
- Li, L., Littman, M. L., Walsh, T. J., and Strehl, A. L. (2011). Knows what it knows: a framework for self-aware learning. *Machine learning*, 82(3):399–443.
- Liu, J., Lin, Z., Padhy, S., Tran, D., Bedrax Weiss, T., and Lakshminarayanan, B. (2020). Simple and principled uncertainty estimation with deterministic deep learning via distance awareness. *Advances in Neural Information Processing Systems*, 33:7498–7512.
- Lu, X. and Van Roy, B. (2017). Ensemble sampling. In *Advances in Neural Information Processing Systems*, pages 3260–3268.
- Lu, X., Van Roy, B., Dwaracherla, V., Ibrahimi, M., Osband, I., and Wen, Z. (2021). Reinforcement learning, bit by bit. *arXiv preprint arXiv:2103.04047*.
- MacKay, D. J. (1992). A practical Bayesian framework for backpropagation networks. *Neural computation*, 4(3):448–472.
- Mandt, S., Hoffman, M. D., and Blei, D. M. (2018). Stochastic gradient descent as approximate Bayesian inference.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533.
- Nado, Z., Band, N., Collier, M., Djolonga, J., Dusenberry, M., Farquhar, S., Filos, A., Havasi, M., Jenatton, R., Jerfel, G., Liu, J., Mariet, Z., Nixon, J., Padhy, S., Ren, J., Rudner, T., Wen, Y., Wenzel, F., Murphy, K., Sculley, D., Lakshminarayanan, B., Snoek, J., Gal, Y., and Tran, D. (2021). Uncertainty Baselines: Benchmarks for uncertainty & robustness in deep learning. *arXiv preprint arXiv:2106.04015*.
- Neal, R. M. (2012). *Bayesian learning for neural networks*, volume 118. Springer Science & Business Media.
- Osband, I., Aslanides, J., and Cassirer, A. (2018). Randomized prior functions for deep reinforcement learning. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems 31*, pages 8617–8629. Curran Associates, Inc.
- Osband, I., Blundell, C., Pritzel, A., and Van Roy, B. (2016a). Deep exploration via bootstrapped DQN. In *Advances In Neural Information Processing Systems 29*, pages 4026–4034.
- Osband, I., Doron, Y., Hessel, M., Aslanides, J., Sezener, E., Saraiva, A., McKinney, K., Lattimore, T., Szepesvári, C., Singh, S., Van Roy, B., Sutton, R., Silver, D., and van Hasselt, H. (2020). Behaviour suite for reinforcement learning. In *International Conference on Learning Representations*.
- Osband, I. and Van Roy, B. (2015). Bootstrapped Thompson sampling and deep exploration. *arXiv preprint arXiv:1507.00300*.
- Osband, I., Van Roy, B., and Wen, Z. (2016b). Generalization and exploration via randomized value functions. In *Proceedings of The 33rd International Conference on Machine Learning*, pages 2377–2386.
- Osband, I., Wen, Z., Asghari, S. M., Dwaracherla, V., Hao, B., Ibrahimi, M., Lawson, D., Lu, X., O’Donoghue, B., and Van Roy, B. (2022a). The neural testbed: Evaluating predictive distributions.
- Osband, I., Wen, Z., Asghari, S. M., Dwaracherla, V., Lu, X., and Van Roy, B. (2022b). Evaluating high-order predictive distributions in deep learning.

- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958.
- Wang, C., Sun, S., and Grosse, R. (2021). Beyond marginal uncertainty: How accurately can Bayesian regression models estimate posterior predictive correlations? In *International Conference on Artificial Intelligence and Statistics*, pages 2476–2484. PMLR.
- Welling, M. and Teh, Y. W. (2011). Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 681–688. Citeseer.
- Wen, Z., Osband, I., Qin, C., Lu, X., Ibrahimi, M., Dwaracherla, V., Asghari, M., and Van Roy, B. (2022). From predictions to decisions: The importance of joint predictive distributions.
- Wenzel, F., Roth, K., Veeling, B. S., Światkowski, J., Tran, L., Mandt, S., Snoek, J., Salimans, T., Jenatton, R., and Nowozin, S. (2020). How good is the Bayes posterior in deep neural networks really? *arXiv preprint arXiv:2002.02405*.

A Open source code

Three related github repositories complement this paper:

1. **enn**: <https://github.com/deepmind/enn>
2. **enn_acme**: https://github.com/deepmind/enn_acme
3. **neural_testbed**: https://github.com/deepmind/neural_testbed

These libraries contain the code necessary to reproduce the key results in our paper, divided into repositories based on focus. Together with each repository, we include several ‘tutorial colabs’ – Jupyter notebooks that can be run in a browser without requiring any local installation. Each of these libraries is written in Python, and relies heavily on JAX for scientific computing (Bradbury et al., 2018). We view this open-source effort as a major contribution of our paper.

The first library, **enn**, focuses on the design of epistemic neural networks and their training. This includes all of our network definitions and loss functions. Our library is built around Haiku (Babuschkin et al., 2020). The library provides the basis for all of the computational work reported in this paper.

The second library, **enn_acme** combines the **enn** library with **acme**, a popular framework for reinforcement learning agent development (Hoffman et al., 2020). We use this library to run the RL experiments in Section 4.3.

The final library, **neural_testbed**, was introduced as part of *The Neural Testbed* (Osband et al., 2022a). We add the **epinet** agent, suitable for comparison with the existing agent implementations in that library.

B Examples of ENNs

Popular existing approaches to uncertainty estimation can be expressed as ENNs. In this section, we work through a few examples. In each case, there are multiple ways to express the approach in ENN terms, and what we present represents only one concrete option. We also release reference implementations of these agents in **neural_testbed/agents/factories**.

A deep ensemble represents uncertainty by retaining an ensemble of K particles, each of which is a conventional neural network (Osband and Van Roy, 2015; Lakshminarayanan et al., 2017). This representation can be framed as an ENN by taking the index set to be $\{1, \dots, K\}$, with each index identifying one of the K neural networks. The index reference distribution is taken to be uniform over these K particles. Suppose each k th particle of the ensemble has parameters θ^k and, given an input x , outputs $f_{\theta^k}(x)$. Then, given an index z , the deep ensemble outputs an indexed prediction $f_{\theta^z}(x)$. Averaging over indexed predictions produces the final prediction $\sum_{z=1}^K f_{\theta^z}(x)/K$.

Monte Carlo dropout offers another approach to estimating uncertainty in neural networks, which has received significant attention in the Bayesian deep learning literature (Srivastava et al., 2014; Gal and Ghahramani, 2016). To cast this as an ENN, take the index z to be the dropout mask. In particular, if the full neural network output is written as $f_{\theta}(x)$ then the output of the ENN can be written as $f_{\theta \odot z}(x)$.

Bayes by backprop (BBB) is a landmark approach to variational inference in neural networks (Blundell et al., 2015). In BBB, each scalar parameter θ_i^{NN} of a base neural network is viewed as an independent latent variable. The distribution of each is taken to be Gaussian, with mean μ_i and variance σ_i^2 . The joint parameter distribution induces a distribution over base neural networks, representing uncertainty about the mapping from input to prediction. Consider an epistemic index z with one component per neural network parameter θ_i^{NN} . If this epistemic index is sampled from a standard Gaussian reference distribution $N(0, I)$ then $\mu + \sigma \odot z$ represents a sample from the neural network parameter distribution. As such, the ENN represented by BBB can be expressed as $f_{\theta^{\text{ENN}}}^{\text{ENN}}(x, z) = f_{\mu + \sigma \odot z}^{\text{NN}}(x)$.

C The epinet

This section offers some intuition on how the epinet, as defined in Section 2, works. We begin with a ‘network diagram’ that illustrates data flow in epinet. Then, we consider didactic examples involving linear models. In this setting, epinets reduce to a linear hypermodels. Finally, we speculate on why epinets offer advantages over alternative ENNs in more complex settings.

C.1 Network diagram

Given an epinet input $\tilde{x} = \text{sg}[\phi_\zeta(x)]$ and epistemic index z , the ENN output can be written as

$$f_\theta(x, z) = \underbrace{\mu_\zeta(x)}_{\text{ENN}} + \underbrace{\sigma_\eta^L(\tilde{x}, z)}_{\text{base net}} + \underbrace{\sigma^P(\tilde{x}, z)}_{\text{epinet}}. \quad (8)$$

Figure 9 illustrates data flow within an epinet. The diagram is not meant to provide any new information beyond Equation (8), but may help to visualize how the epinet operates. Arrows leading into a module indicate dependence on data provided by other modules. The dashed line from $Z \sim P_Z$ indicates that predictive distributions are calculated with epistemic indices sampled from P_Z .

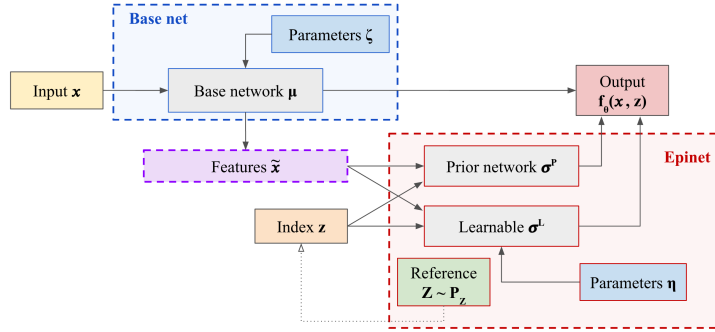


Figure 9: Visualizing the epinet network architecture as described by Equation (8).

C.2 Didactic examples

To offer some intuition for how epinets work and what they accomplish, we present a simple example specialized to a linear base model. A linear base model produces an output $\mu_\zeta(x) = \zeta^T x$ given an input x and model parameters ζ . It is natural to add to this a linear epinet $\sigma_\eta(\phi_\zeta(x), z) = z^T \eta x$. The combined architecture is equivalent to a linear *hypermodel* (Dwaracherla et al., 2020). To see this, note that,

$$f_\theta(x, z) = \mu_\zeta(x) + \sigma_\eta(\phi_\zeta(x), z) = \zeta^T x + z^T \eta x = (\zeta + \eta^T z)^T x = \mu_{\zeta + \eta^T z}(x). \quad (9)$$

As such, properties of linear hypermodels, such as their ability to implement exact Bayesian linear regression, carry over to such epinets.

Figure 10 illustrates predictive uncertainty estimates produced by linear epinets. Figure 10a presents posterior credible intervals of an epinet compared against exact Bayesian inference. Data is generated by a one-dimensional linear regression model with a Gaussian prior distribution and Gaussian noise. The loss function for this epinet follows prior work by includes Gaussian bootstrapping in regression (Lu and Van Roy, 2017; Dwaracherla et al., 2020),

$$\ell(\theta, x_i, y_i, z) = (f_\theta(x_i, z) - y_i + \sigma c_i^T z)^2.$$

Here c_i is a random signature drawn from the unit sphere in D_Z generated independently for each training example (x_i, y_i) and σ is the scale of the additive bootstrap noise. This figure indicates that the epinet outputs well-calibrated marginal predictive distributions.

We next consider classification with a two-dimensional input and two classes. Data is generated by a standard logistic regression model with parameters drawn from a Gaussian prior. Figure 10b presents standard deviations of marginal predictive distributions across the input space. We supplement a standard logistic regression model with a linear epinet. The plots compare results against those generated via SGMCMC, which we expect in this case to closely approximate exact Bayesian inference. While these figures bear qualitative similarities, significant differences arise because the linear epinet architecture imposes symmetries that are not respected by exact posterior distributions. In particular, this epinet can be thought of as representing parameter uncertainty as Gaussian. While our data generating process assumes a Gaussian prior distribution, the posterior distributions, which are conditioned on binary outcomes, are not Gaussian. More complex, nonlinear, epinets should be able to more accurately represent the posterior distribution over classifiers.

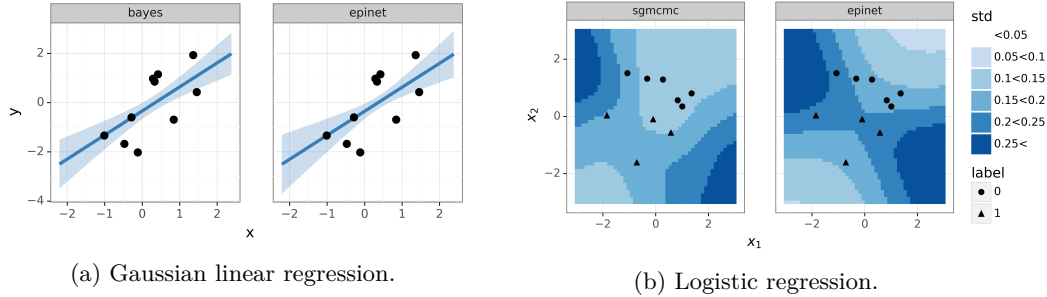


Figure 10: Visualizing the epinet predictions compared to benchmark approaches.

C.3 Advantages afforded by an epinet

Beyond linear models, epinets substantially differ from hypermodels. Empirical results presented in this paper suggest that these differences afford epinets considerable advantages over alternative ENNs. In this section, we speculate on potential sources of advantage.

All Bayesian approaches to neural network learning – such as hypermodels, ensembles, and dropout – can be cast in terms of ENNs. However, they are specialized to operate via a two-stage process: the epistemic index identifies base network parameters θ , and then, given an input x , the resultant network produces an output $f_{\theta}(x)$. While this separation is conceptually clean, it imposes a restrictive view of base network parameters as unknowns to be inferred. The ENN interface offers greater flexibility. For example, ENN parameters do not necessarily admit semantic meaning. Indeed, epinet parameters simply serve as computational artifacts that facilitate learning of a mapping that produces accurate joint predictions. Further, in an epinet, ENN parameters manifest within the main network and interact with the epistemic index through standard neural network building blocks. We believe that this allows for more efficient allocation of resources between learning point predictions and uncertainty estimates. Perhaps this is why epinets outperform ensembles several orders of magnitude larger.

Epinets also conveniently supplement pre-trained base networks. This is important for settings where expensive resources have been dedicated to training the base network. Without further training of the base network, the epinet can benefit from the features learned during pretraining and use them to represent uncertainty in a high-level latent space. This is in contrast other approaches, such as ensembles, which require retraining each particle from scratch, or hypermodels, which introduce complex nonlinearities into training dynamics.

D Dyadic sampling

To evaluate the quality of joint predictions, we sample batches of inputs (x_1, \dots, x_τ) and assess log-loss with respect to corresponding labels (y_1, \dots, y_τ) . With a high-dimensional input space, labels of inputs sampled uniformly at random are typically nearly independent. Hence, a large batch size τ is required to distinguish joint predictions that effectively reflect interdependencies among labels. However, this becomes impractical because computational requirements for testing grow exponentially in τ . Dyadic sampling serves as a practical heuristic that samples inputs more strategically so that effective agents are distinguished with a manageable batch size τ (Osband et al., 2022b) even when inputs are high-dimensional.

D.1 Basic version

The basic version of dyadic sampling, for each batch, first samples two independent random anchor points, \tilde{x}_1 and \tilde{x}_2 , from the input distribution. Then, to form a batch of size τ , the samples τ points independently and with equal probability from these two anchor points $\{\tilde{x}_1, \tilde{x}_2\}$. To assess an agent, its joint prediction of labels is evaluated for this batch of size τ . Even with a moderate value of $\tau = 10$, a batch produced by dyadic sampling gives rise to labels that are likely to correlate – in particular, labels assigned to the same anchor point. Osband et al. (2022b) demonstrate that, across many problems, this sampling heuristic is effective in distinguishing the quality of joint predictions with modest computation.

On the Neural Testbed, the input distribution is standard normal. Thus, for each test batch, we sample anchor points $x_1, x_2 \sim N(0, I)$ and then re-sample $\tau = 10$ points from these two anchor points to form a batch.

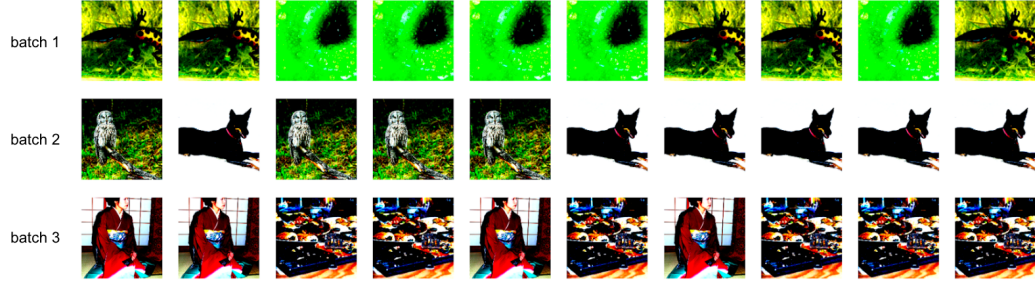
On ImageNet, for faster evaluation, rather than sampling anchor points from the evaluation set, we split the evaluation set into batches of size 2. We then iterate over these batches of size 2, re-sample $\tau = 10$ points from each input pair, and evaluate the log-loss of an agent’s joint predictions on these batches of size $\tau = 10$. Finally, we take the average of all the joint log-losses.

Figure 11a shows a few examples of these dyadic input batches of size $\tau = 10$. It may seem unsatisfactory that images are repeated exactly within a batch. Even though we view this metric as a unit test that an intelligent agent pass, it is conceivable to design a ‘cheating’ agent that takes advantage of this repeating structure. In the following section, we will consider a more robust version of dyadic sampling, where instead of repeating images exactly, we perturb images using standard data augmentation techniques to introduce more diversity within a dyadic batch.

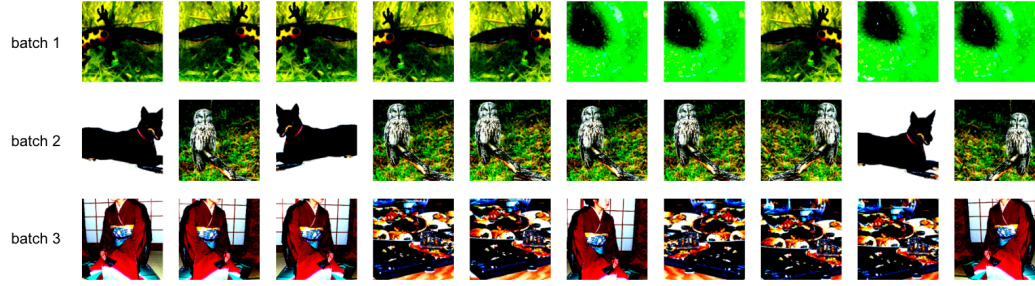
D.2 Augmented dyadic sampling

In vanilla dyadic sampling, each dyadic batch has two unique elements (the anchor points), which are repeated multiple times within the batch. To make the metric more robust at distinguishing agents, we independently perturb each input in a dyadic batch using standard data augmentation techniques, so each input within a batch differs from others. For ImageNet, the perturbation takes the form of random cropping and flipping. We take the label of each perturbed image to be the label of its original image. Effective joint predictions indicate that images perturbed from the same anchor image are likely to be the same. Joint log-loss penalizes agents that do not recognize this. We call this sampling scheme *augmented dyadic sampling*. Figure 11b presents a few examples of augmented dyadic batches for Imagenet.

We evaluate our trained ResNet, ensemble, and epinet agents in Section 3.2 using augmented dyadic sampling, and we compare the joint log-loss with that obtained from basic dyadic sampling. In Figure 12, we see that the results from using these two sampling schemes are qualitatively similar. While the overall joint log-loss is higher for augmented dyadic sampling due to added perturbations, all agents benefit from increasing the model size. The joint log-loss of the ensemble agent improves more so than the ResNet agent with increasing model size. More importantly, the epinet agent outperforms both baselines by a huge margin under both sampling schemes. These results give us further confidence in the quality of joint predictions produced by the epinet agent.



(a) Examples of input batches generated by basic dyadic sampling. Each row is a batch of size 10, on which the joint log-loss is evaluated.



(b) Examples of input batches generated using augmented dyadic sampling. Compared to Figure 11a, each image is randomly cropped and flipped, introducing more diversity to the batch.

Figure 11: Examples of input batches used for evaluating joint predictions. Figure 11a is generated through dyadic sampling and Figure 11b includes additional perturbations.

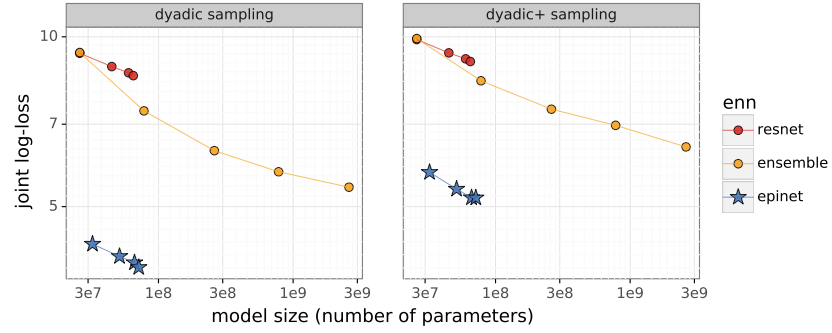


Figure 12: An agent's joint log-loss under dyadic (left) and augmented dyadic sampling (right).

E Testbed experiments

This section provides details about the Neural Testbed experiments in Section 3.1. We begin with a review of the neural testbed as a benchmark problem, and the associated generative models. We then give an overview of the baseline agents we compare against in our evaluation. Next, we provide supplementary details for the hyperparameters and implementation details for the epinet agent as outlined in Section 3.1. Finally, we investigate the sensitivity of our hyperparameter choices when evaluated across the testbed.

E.1 Neural testbed

The Neural Testbed (Osband et al., 2022a) is a collection of neural-network-based, synthetic classification problems that evaluate the quality of an agent’s predictive distributions. We make use of the open-source code at https://github.com/deepmind/neural_testbed. The Testbed problems use random 2-layer MLPs with width 50 to generate training and testing data. The specific version we test our agents on entails binary classification, input dimension $D \in \{2, 10, 100\}$, number of training samples $T = \lambda D$ for $\lambda \in \{1, 10, 100, 1000\}$, temperature $\rho \in \{0.01, 0.1, 0.5\}$ for controlling the signal-to-noise ratio, and 5 random seeds for generating different problems in each setting. The performance metrics are averaged across problems to give the final performance scores.

E.2 Benchmark agents

We follow Osband et al. (2022a) and consider the benchmark agents as in Table 1. We use the open-source implementation and hyperparameter sweeps at https://github.com/deepmind/neural_testbed/agents/factories. According to Osband et al. (2022a), the benchmark agents are carefully tuned on the Testbed problems, so we do not further tune these agents.

E.3 Epinet

We take the reference distribution of the epistemic index to be a standard Gaussian with dimension $D_Z = 8$. The base network μ_ζ has the same architecture as the baseline `mlp` agent, which is a 2-layer MLP with ReLU activation and 50 units in each hidden layer. The learnable part of the epinet σ_η^L takes $\phi_\zeta(x)$ and index z as inputs, where $\phi_\zeta(x)$ is the concatenation of x and the last-layer features of the base network. The learnable network has the form $\sigma_\eta^L(\phi_\zeta(x), z) = g_\eta([\phi_\zeta(x), z])^T z$ where $[\phi_\zeta(x), z]$ is the concatenation of $\phi_\zeta(x)$ and z , and $g_\eta(\cdot)$ is a 2-layer MLP with hidden width 15, ReLU activation, and outputs in $\mathbb{R}^{D_Z \times C}$ for number of classes $C = 2$.

For the fixed prior σ^P , we consider an ensemble of D_Z networks. Each member of the ensemble is a small MLP with 2 hidden layers and 5 units in each hidden layer. Each MLP takes x as input and returns logits for the two classes. Let $p^i(x) \in \mathbb{R}^C$ denote the output of the i^{th} member of the ensemble. We combine the outputs of the ensemble members by taking the weighted sum, $\sum_{i=1}^{D_Z} p^i(x) z_i$ and multiplying the sum by a tunable scaling factor α . Thus, we can write the prior function $\sigma^P(\phi_\zeta(x), z) = \alpha \sum_{i=1}^{D_Z} p^i(x) z_i$.

We combine the base network and the epinet by adding their outputs and applying a stop-gradient operation on $\phi_\zeta(x)$,

$$f_\theta(x, z) = \mu_\zeta(x) + \sigma_\eta^L(\text{sg}[\phi_\zeta(x)], z) + \sigma^P(\text{sg}[\phi_\zeta(x)], z),$$

where $\theta = (\zeta, \eta)$ denotes all the parameters of the base net and epinet. We train the parameters ζ and η jointly. The training loss takes the form as specified in (5), where we use cross-entropy loss for the data loss and ridge regularization. We update θ using stochastic gradient descent, as shown in (6), with a batch size of 100 and number of epistemic index samples equal to the index dimension. We use Adam optimizer with learning rate **1e-3**. The $L2$ weight decay and prior scaling factor α are roughly adjusted for different problem settings, taking in account the number of training samples and SNR.

Our implementation of the epinet agent can be found under the path `/agents/factories/epinet.py` in the anonymized neural testbed github.

E.4 Ablation studies

We run ablation experiments on the epinet agent that is trained simultaneously along with the base network. We sweep over various values for the index dimension, the number of hidden layers in the trainable epinet, the width of hidden layers of epinet, the ensemble prior’s prior scale, the width of the hidden layers in the prior network, and $L2$ weight decay. We keep all other hyperparameters fixed to the open-sourced default configuration while sweeping over one hyperparameter.

Our results are summarized in Figure 13. We see that a larger index dimension improves both joint and marginal kl estimates. The epinet performance is not sensitive to the number of hidden layers. We suspect that this is due to similar number of parameters across epinets with different number of hidden layers. We observe that epinet performance is not sensitive to width of the epinet hidden layers once the width is large enough. Performs of epinet seems sensitive to the prior scales. Smaller prior scale leads to better marginal kl, too small or too large prior scale degrades the joint kl. Increasing the width of the models in the ensemble prior network improves the epinet performance. However, the improvement seems marginal after a point. The epinet seems to be sensitive to $L2$ weight decay. A very small or large weight decay degrades the performance of epinet.

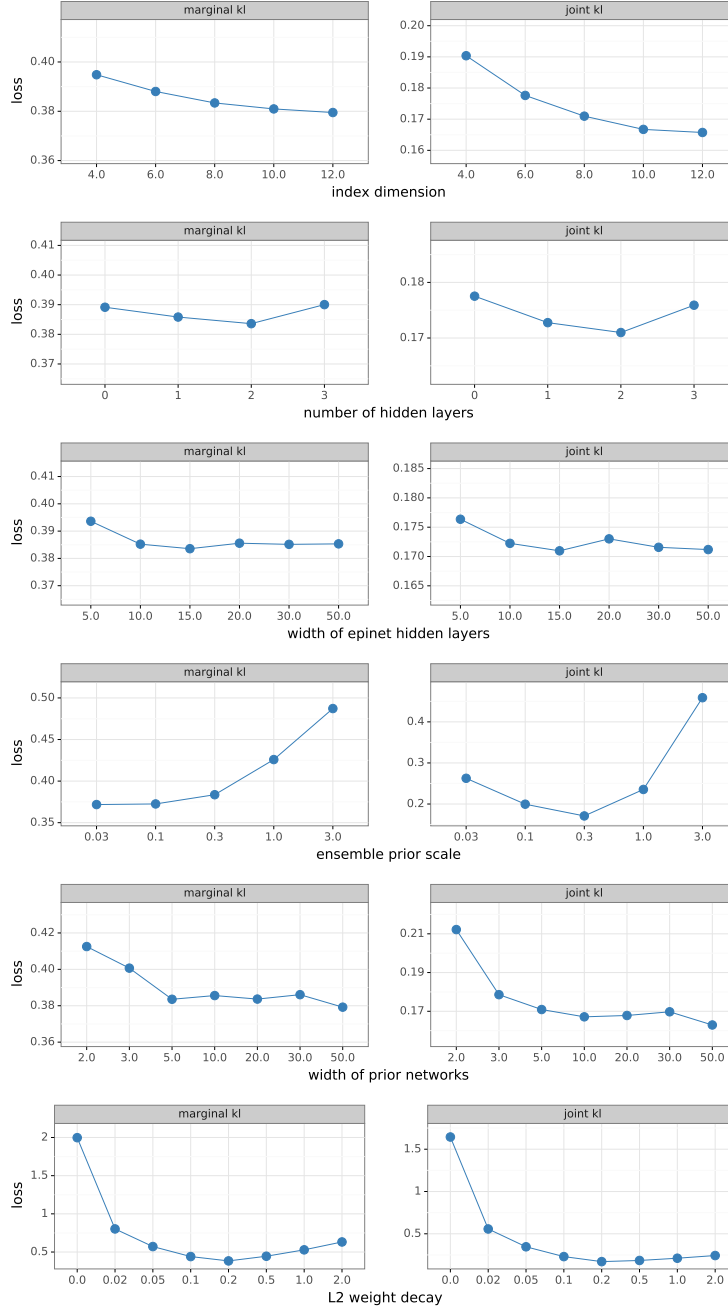


Figure 13: Ablation studies of epinet with 2-layer mlp base model on the neural testbed.

F Image agents

This section gives an overview of our experiments on image classification problems outlined in Section 3.2. We begin with a review of the hyperparameter choices and design details for the agents as implemented in our experiments. Then, we include a comparison of these agents to benchmark implementations in the field as embodied by ‘uncertainty baselines’ (Nado et al., 2021). Next, we present results for an evaluation on both CIFAR-10 and CIFAR-100, and find that the overall results match that on ImageNet. We complement these results with an analysis of the computational cost in terms of FLOPs as well as memory on modern TPU architectures. Finally, we perform a suite of ablations to investigate the sensitivity of our results across ImageNet.

F.1 Epinet details

We train one epinet for each ResNet- L baseline for $L \in \{50, 101, 152, 200\}$. The ResNet baselines are open sourced in the ENN library under the path `/networks/resnet/`, and the checkpoints are available under `/checkpoints/imagenet.py`. For each epinet agent, we take the pre-trained ResNet as the base network. We do not update the base network during epinet training.

As discussed in Section 3.2, we choose the reference distribution of the epistemic index to be a standard Gaussian with dimension $D_Z = 30$. The input to the learnable part of the epinet σ_η^L includes the last-layer features of the base ResNet and the epistemic index z . Let C denote the number of classes. For last-layer features ϕ and epistemic index z , the learnable network takes the form $\sigma_\eta^L(\phi, z) = g_\eta([\phi, z])^T z$, where $[\phi, z]$ is the concatenation of ϕ and z , and $g_\eta(\cdot)$ is a 1-layer MLP with 50 hidden units, ReLU activation, and output $\in \mathbb{R}^{D_Z \times C}$.

The fixed prior σ^P is made up of two components. The outputs of the two parts are summed together to produce the prior output. In general, we could have tunable scaling factors (which we refer to as ‘prior scales’ in the ablation studies) for the output of each component before we add them together. However, for ImageNet, we find that scaling factors of 1 already work well. The first component of the prior is a network with the same architecture and initialization as the learnable network σ_η^L . The second component is an ensemble of small convolutional networks that act directly on the input images. The number of networks in the ensemble is equal to the index dimension. Each convolutional network has the number of channels (4, 8, 8), kernel shapes $(10 \times 10, 10 \times 10, 3 \times 3)$, and strides (5, 5, 2). The outputs are flattened and taken through a linear layer to give a vector of dimension C . For input image x , let $p^i(x) \in \mathbb{R}^C$ denote the output of the i^{th} member of the ensemble. We combine the outputs of the ensemble members by taking the weighted sum $\sum_{i=1}^{D_Z} p^i(x) z_i$.

The epinet network architecture together with checkpoint weights can be found in the ENN library under the paths `/networks/epinet/` and `/checkpoints/imagenet.py`.

The ResNet baselines are trained using cross-entropy loss and ridge regularization. We optimize using SGD with a learning rate 0.1, a cosine learning rate decay schedule, and Nesterov momentum. We apply $L2$ weight decay of strength $1\text{e-}4$. We also incorporate label smoothing into the loss, where instead of one-hot labels, the incorrect classes receive a small weight of $0.1/C$. We train the ResNet agents for 90 epochs on 4×4 TPUs with a per-device batch size of 128.

We train the epinet using loss of the form (6) with cross-entropy loss with ridge regularization. Similar to ResNet training, we apply $L2$ weight decay of strength $1\text{e-}4$ and incorporate label smoothing into the loss. We draw 5 epistemic index samples for each gradient step. We optimize the loss using SGD with a learning rate 0.1, Nesterov momentum and decay 0.9. The epinet is trained on the same hardware with the same batch size for 9 epochs.

F.2 Uncertainty baselines

In this section we compare our results to the open source ‘uncertainty baselines’, which provides a reference implementation for much work on uncertainty estimation in Bayesian deep learning (Nado et al., 2021). As part of our development, we upstream an optimized

method for calculating joint log-loss, and contribute this to the community. We benchmark a few popular approaches to uncertainty estimation in terms of both marginal and joint predictions, and compare their results to ours. At a high level, our results mirror our own results of Figure 2. After tuning, most of the agents appear on a roughly similar tradeoff in terms of marginal quality. However, the approaches are widely separated in terms of their quality on *joint* prediction, and here *epinet* performs much better than the alternatives.

Figure 14 repeats Figure 2 but adding a few new agents from uncertainty baselines, which we

- **mimo**: Multi-input Multit-output ensemble (Havasi et al., 2020). This method appears to perform better than a baseline resnet in terms of marginal statistics, but provide no additional benefit to modeling joints. Note that the independent product of better marginals automatically does improve joints.
- **dropout**: Dropout as posterior approximation (Srivastava et al., 2014; Gal and Ghahramani, 2016). This approach seems to provide a slight improvement in marginal log-loss and a noticeable improvement in joint log-loss. However although dropout has a low computational cost in terms of *parameters*, these results required 10 forward passes of the network, and so actually *underperform* relative to an ensemble of similar inference cost (see Appendix F.4).
- **het**: Heteroscedastic loss (Collier et al., 2020, 2021). This approach performs well in terms of joint log-loss, and achieves performance close to the ensemble of size=100. Although the results are still significantly worse than those of our *epinet*, it is interesting to note that the functional form of the resultant *heteroscedastic* agent can actually be written as a particular form of *epinet*. In future work, we would like to understand better the commonalities between these two approaches, and see if/when the algorithms can borrow from each others’ strengths.
- **het_mimo**: Heteroscedastic loss with MIMO (Collier et al., 2020, 2021; Havasi et al., 2020). This approach combines both MIMO network with heteroscedastic loss, and seems to benefit from both. As above, this can be expressed as a particular form of *epinet*, although the result does not quite match the performance of our agents.
- **sngp**: Spectral-normalized Neural Gaussian Process (Liu et al., 2020). This approach produces leads to a joint log-loss that is significantly better than an ensemble of size=100 and approaching that of the the *epinet*. However, unlike *epinet* this *sngp* layer also leads to a significant degradation on the classification error. For many applications this trade off may may be unacceptable.

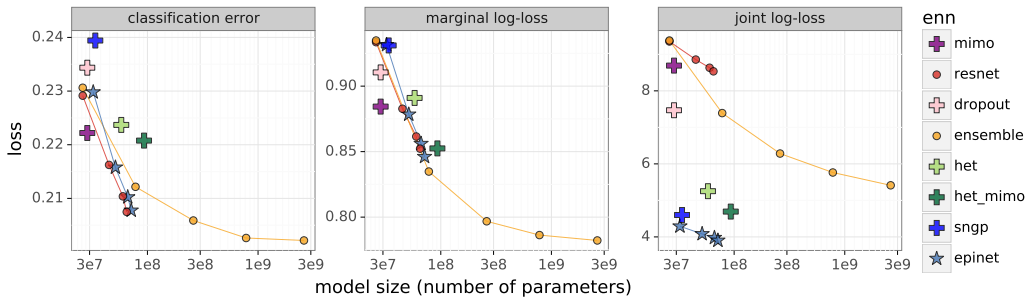


Figure 14: Quality of marginal and joint predictions across models on ImageNet. Reproduces the results of Figure 2 but now including ‘uncertainty baselines’ (Nado et al., 2021).

F.3 CIFAR-10 and CIFAR-100

In this section we reproduce a similar analysis to Section 3.2 applied to the CIFAR-10 and CIFAR-100 datasets. We find that, at a high level, our results mirror those when applying ResNet to ImageNet. In particular, we are able to produce results similar to Figure 2 for both CIFAR-10 and CIFAR-100. Relative to large ensembles, epinets greatly improve joint predictions at orders of magnitude lower computational cost.

For the experiments in this section, we mirror the ImageNet experiments but using the smaller ResNet architectures. The ResNet baselines are open sourced in the ENN library under the path `/networks/resnet/`, and the checkpoints are available under `/checkpoints/cifar10.py` and `/checkpoints/cifar100.py`. In particular, we tune ResNet- L for $L \in \{18, 32, 44, 56, 110\}$ over learning rate and weight decay. We did not include temperature rescaling in these sweeps, although this could further improve performance for all agents. After tuning hyperparameters we independently initialize and train 100 ResNet-18 models to serve as ensemble particles. These models are then used to form ensembles of size 1, 3, 10, 30, and 100.

For the epinet agent we take the pretrained ResNet as the base network and fix its weights. We then follow the same methodology as for ImageNet, described in Appendix F.1, but with a slightly smaller network. We use index dimension $D_Z = 20$ and alter the convolutional prior to have channels (4, 8, 4) each with kernel size 5×5 and stride 2 on account of the smaller image sizes in CIFAR-10 and CIFAR-100 (Krizhevsky, 2009).

The epinet network architecture together with checkpoint weights can be found in the ENN library under the paths `/networks/epinet/`, `/checkpoints/cifar10.py`, and `/checkpoints/cifar100.py`.

Figures 15 and 16 reproduce our scaling results for ImageNet when applied to these other datasets. At a high level, the key observations remain unchanged across datasets. We see that across all statistical losses the larger models generally perform better. When looking at classification and marginal log-loss, epinets do not offer any particular advantage over baseline ResNets. However, when we look at the *joint* log-loss we can see that epinets offer huge improvements in performance, even when measured against very large ensembles.

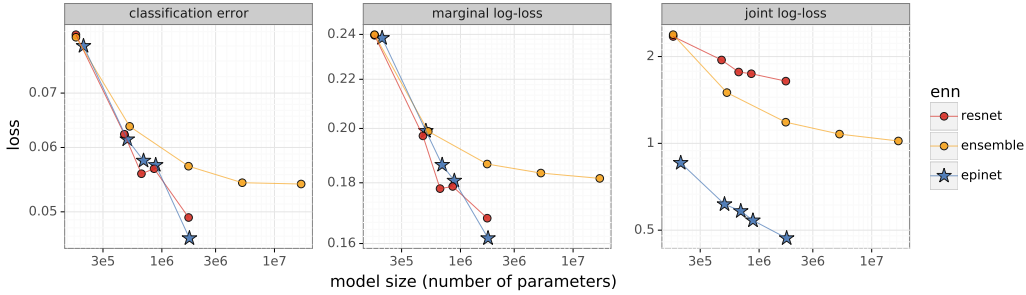


Figure 15: Quality of marginal and joint predictions across models on CIFAR-10.

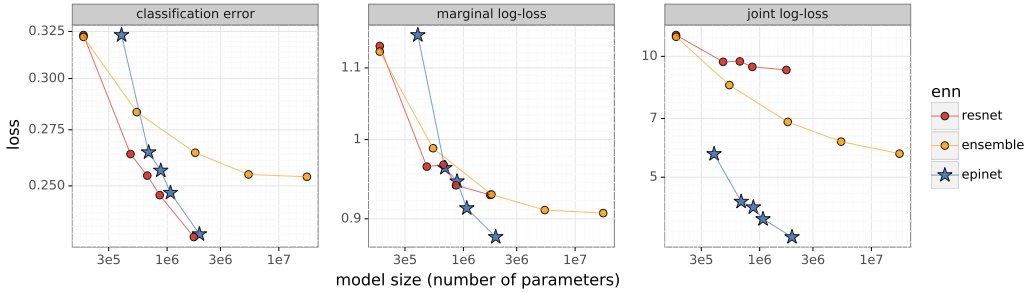


Figure 16: Quality of marginal and joint predictions across models on CIFAR-100.

F.4 Computational cost

This paper highlights a key tension in neural network development: balancing statistical loss against computational cost. Our main results on ImageNet (Figure 2) use ‘memory’ as a proxy for computational cost in large deep neural networks, for which this is often a hardware bottleneck (Kaplan et al., 2020). However, in the case of these models, the results are similar for many alternative measures.

Figure 17 reproduces the results of Figure 2 but with computational cost measured in total floating point operations at inference. This plot includes the total costs of one thousand forwards of the epinet. Even with these additional FLOPs, the overall cost of each epinet is still less than 50% of the total network, and the outperformance of the epinet in terms of joint log-loss is still remarkable. Further, on large modern TPU architectures these epinet operations can often be performed in parallel. This means that, in some cases, these extra FLOPs may require no extra *time* to forward on the device.

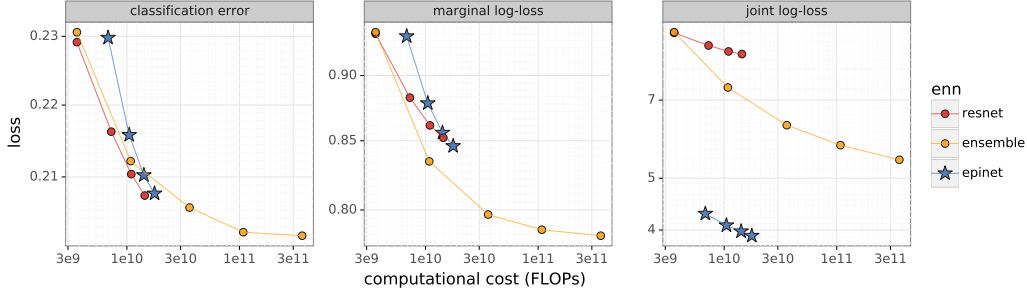


Figure 17: Quality of marginal and joint predictions across models on ImageNet. Reproduces the results of Figure 2 but using inference FLOPs as measure of computation.

The results of Figure 17 hide an extra hyperparameter in epinet development: the number of independent samples of the index z , which we will call M . All of the results presented in this paper focus on the case $M = 1000$, however an agent designer may choose to vary this depending on their tradeoff between statistical loss and computational cost. Figure 18 shows this empirical tradeoff over $M \in \{10, 30, 100, 300, 1000, 3000\}$ for each of the resnet variants. Once again, we see that in terms of *marginal* statistics, there is really no benefit to using an epinet. However, once you look at joint log-loss even a small number of epinet samples improves over the ResNet. Interestingly, these results continue to improve for $M > 1000$ but at a higher computational cost.

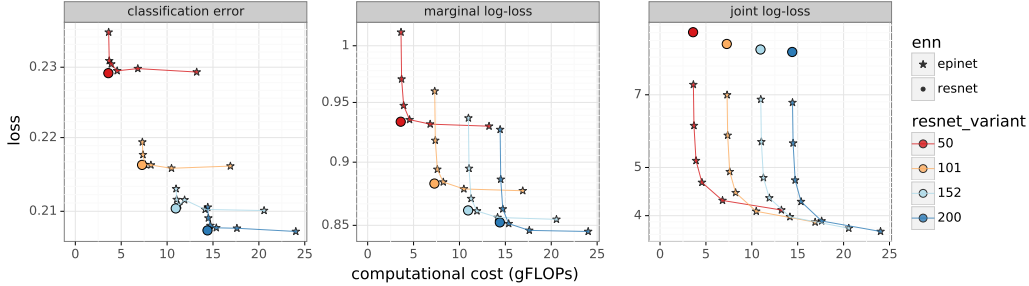


Figure 18: Comparing base ResNet against epinet for differing numbers of sampled indices z .

F.5 Epinet ablations

We run ablation experiments on the epinet agent that builds on the pre-trained ResNet-50 base network. We sweep over various values for the index dimension, the number of hidden layers in the trainable epinet, the prior scale for the matched epinet prior, the prior scale for the ensemble of convolutional networks, $L2$ weight decay, the number of index samples drawn for each gradient step, label smoothing, and temperature re-scaling post-training. We keep all other hyperparameters fixed to the open-sourced default configuration while sweeping over one hyperparameter.

Our results are summarized in Figure 19. We see that a larger index dimension improves the joint loss-loss, but does not necessarily improve the marginal log-loss and classification error. Adding another hidden layer to the trainable epinet makes the marginal and joint loss-loss worse, but we suspect that the performance could be improved with more tuning. The epinet

performance seems sensitive to the prior scales. In the third and fourth rows, we see that the performance degrades quickly when the prior scales become too large. The epinet seems relatively robust to different values of $L2$ weight decay. A large weight decay improves the joint log-loss but slightly worsens the marginal log-loss. The number of index samples and degree of label smoothing do not seem to affect the performance of epinet. Interestingly, we find that using a cold temperature < 1 to re-scale the ENN output logits post-training improves the agent’s performance during evaluation (Wenzel et al., 2020). The improvement is most dramatic in the joint log-loss.

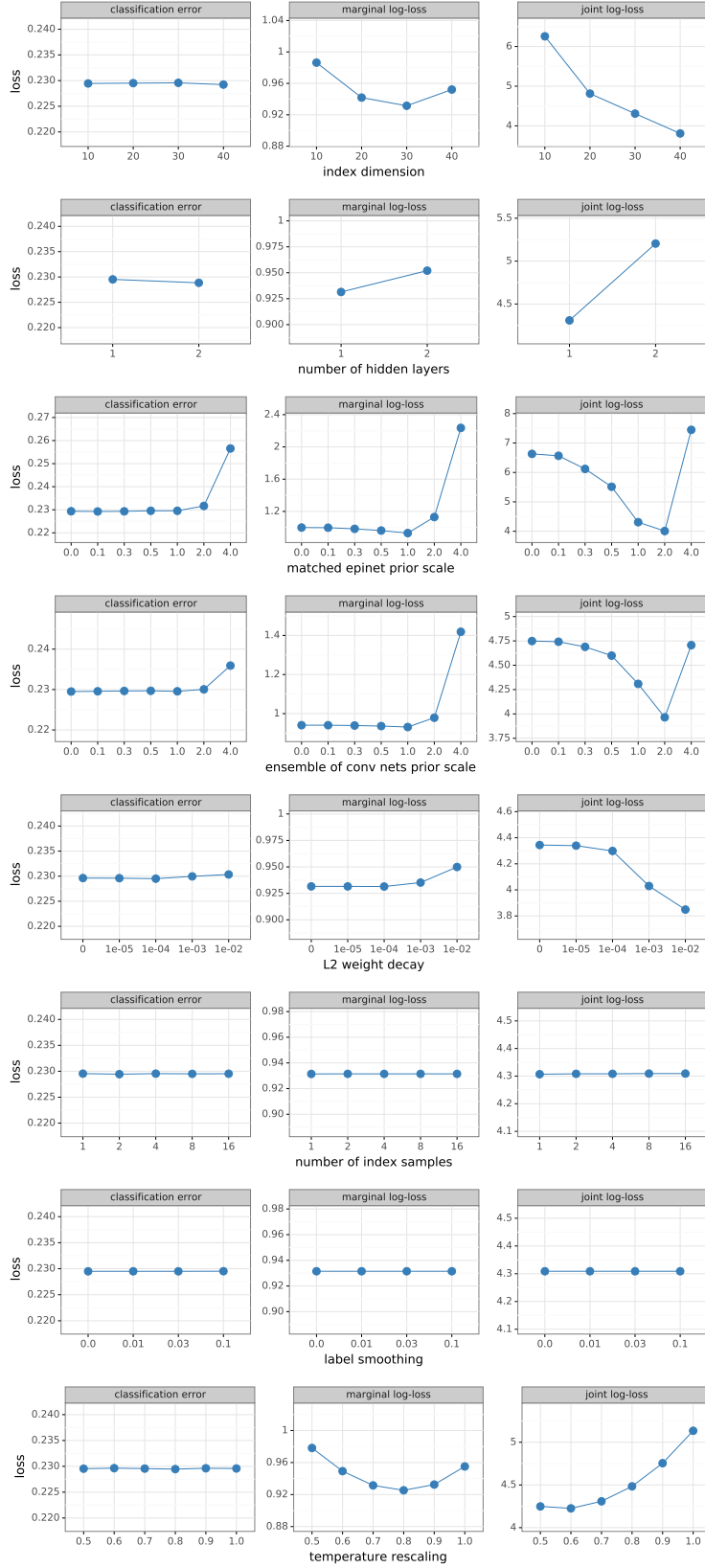


Figure 19: Ablation studies of epinet with ResNet-50 base model on ImageNet.

G Reinforcement Learning

This appendix is designed to provide supplementary details that complement the results of Section 4. We consider the setting of episodic reinforcement learning as outlined by Algorithm 1. At the start of each episode k , the agent samples an index Z_k from the reference distribution P_Z and acts greedily throughout the episode according to the action value function indexed by the sampled index Z_k . At each timestep of an episode, the agent accumulates and adds to the data buffer a tuple in the form of (s, a, r, s') , where s is the state, a is the action taken by the agent at state s , r is the reward, and s' is the observed next state. The agent updates the ENN based on the data in the buffer along the direction of the gradient of the loss computed using a mini-batch of data \tilde{D} and a batch of indices \tilde{Z} sampled i.i.d from P_Z . The gradient is presented in Equation 7. The interaction of an ENN agent with the environment is presented in Algorithm 1. Note that bandit problems are a special case of RL problems whose episode terminates after the first timestep.

Algorithm 1 Episodic RL

Input:	θ_0	initial parameters
	$f_\theta : \mathcal{S} \times \mathcal{Z} \rightarrow \mathbb{R}^{ \mathcal{A} }$	action-value function represented by ENN
	P_Z	index sampler for reference distribution


```

1:  $\theta^{\text{target}} \leftarrow \theta_0$ 
2: initialize buffer
3: for episode  $k = 1, 2, \dots$  do
4:   sample index  $Z_k \sim P_Z$ 
5:    $t \leftarrow 0$ 
6:   observe  $S_0^k$ 
7:   while  $S_t^k$  is not terminal do
8:     apply  $A_t^k \leftarrow \arg \max_a f_\theta(S_t^k, Z_k)_a$ 
9:     observe  $R_t^k, S_{t+1}^k$ 
10:    buffer.add( $(S_t^k, A_t^k, R_t^k, S_{t+1}^k)$ )
11:     $\theta, \theta^{\text{target}} \leftarrow \text{update}(\text{buffer}, \theta, \theta^{\text{target}})$ 
12:     $t \leftarrow t + 1$ 
13:   end while
14: end for
```

We compare `mlp`, `dropout`, `ensemble`, `hypermodel`, `ensemble+`, and `epinet` agents on bsuite and neural bandit problems (Osband et al., 2022a). Similar to (Osband et al., 2022a), we run the enn agents on these problems with their best hyperparameters on the neural testbed for all ENN agents.

G.1 Bandit problem

For our experiments on neural bandit problems, we use a replay buffer of size 10,000. We update the ENN parameters after each observation with one stochastic gradient step computed using a batch of 128 observations from the replay buffer and a batch of i.i.d index samples from P_Z . The gradient step is similar to Equation 7 but with $\gamma = 0$. For computing the gradient, `epinet` agent used a batch of 5 index samples and other agents use the respective default values specified in https://github.com/deepmind/neural_testbed. We use Adam optimizer (Kingma and Ba, 2015) with a learning rate of 0.001 for updating the ENN parameters based on the gradient.

For `epinet`, we make use of the same architecture which is described in Appendix E.3. We use a standard Gaussian with dimension $D_z = 5$ as the reference distribution and scaling factor $\alpha = 0.3$ for the ensemble prior. All of the experiment code is available at `neural_testbed/bandit/`. Figure 20 shows the total regret of the agents after 50,000 timesteps against their computational cost. Once again, we observe that `epinet` performs as well as the best agent with only a fraction of compute even on bandit problems.

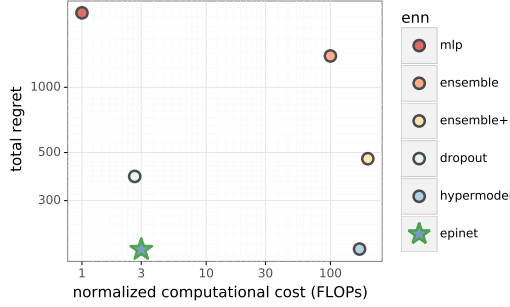


Figure 20: Bandit compute

G.2 Behaviour suite for reinforcement learning

For our experiments on **bsuite** we follow the modified DQN as outlined in Section 4.1. Just as in the bandit experiments, we use the best hyperparameters as selected on the neural testbed for all agents except **epinet**. All ENN agents make use of a replay buffer of size 10,000 and update the ENN parameters after each interaction with the environment. Each update consists of taking a step in the direction of the gradient of the loss function, Equation (7), using a batch of 128 observations from the replay buffer and a batch of 20 i.i.d index samples from the reference distribution. We make use of discount factor $\gamma = 0.99$ for all ENN agents in our experiments.

For **epinet** we use a similar architecture to Appendix E.3 but only a single-hidden layer **epinet** with 50 hidden units along with a 2-hidden layer MLP base model, 2-dimensional normal Gaussian distribution as the reference distribution, and a scaling factor $\alpha = 1$ for the ensemble prior.

We use a single set of hyperparameters for all the **bsuite** environments. However, different **bsuite** environments have different maximum possible rewards, and a single value of prior scale might not suffice for all the environments. To overcome this, we first run a uniform random action policy, which samples actions with equal probability from the set of possible actions, for 100 time steps. We use this data to scale the output of the prior value functions to have a mean 0 and variance 1 for all the agents which use prior functions (**hypermodel**, **ensemble+**, and **epinet**). A detailed breakdown of performance of different agents across different environments is provided in Appendix H.

H bsuite report: Epistemic Neural Networks

The *Behaviour Suite for Core Reinforcement Learning* (?), or **bsuite** for short, is a collection of carefully-designed experiments that investigate core capabilities of a reinforcement learning (RL) agent. The aim of the **bsuite** project is to collect clear, informative and scalable problems that capture key issues in the design of efficient and general learning algorithms and study agent behaviour through their performance on these shared benchmarks. We test agents which use ENNs to represent uncertainty in action-value functions. Please refer to Appendix G for training details.

H.1 Agent definition

In these experiments we use the DQN variants defined in `enn_acme/experiments/bsuite`. These agents differ principally in terms of their ENN definition, which are taken directly from the `neural_testbed/agents/factories` as tuned on the Neural Testbed. We provide a brief summary of the ENNs used by agents:

- **mlp**: A ‘classic’ DQN network with 2-layer MLP.
- **ensemble**: An ensemble of DQN networks which only differ in initialization.
- **dropout**: An MLP with dropout used as ENN (Gal and Ghahramani, 2016).
- **hypermodel**: A linear hypermodel (Dwaracherla et al., 2020).
- **ensemble+**: An ensemble of DQN networks with additive prior (Osband et al., 2016a, 2018).
- **epinet**: The epinet architecture outlined in Section G.2.

H.2 Summary scores

Each **bsuite** experiment outputs a summary score in $[0,1]$. We aggregate these scores by according to key experiment type, according to the standard analysis notebook.

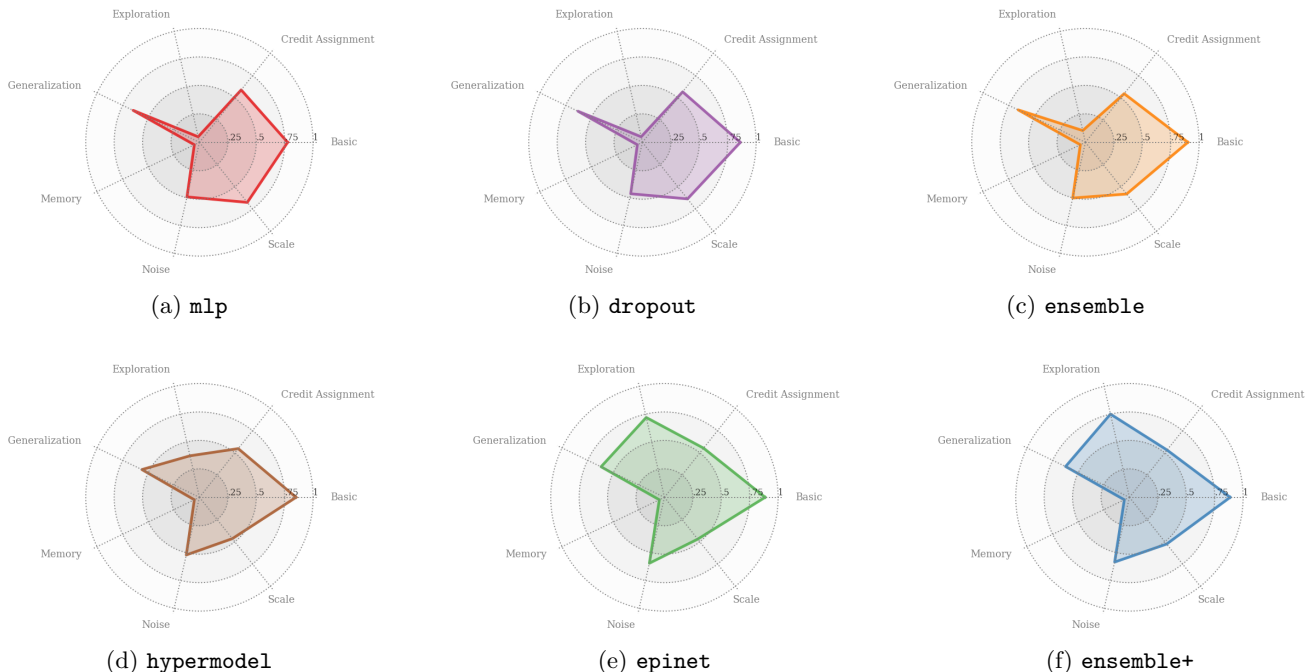


Figure 21: Radar plots give a snapshot of agent capabilities.

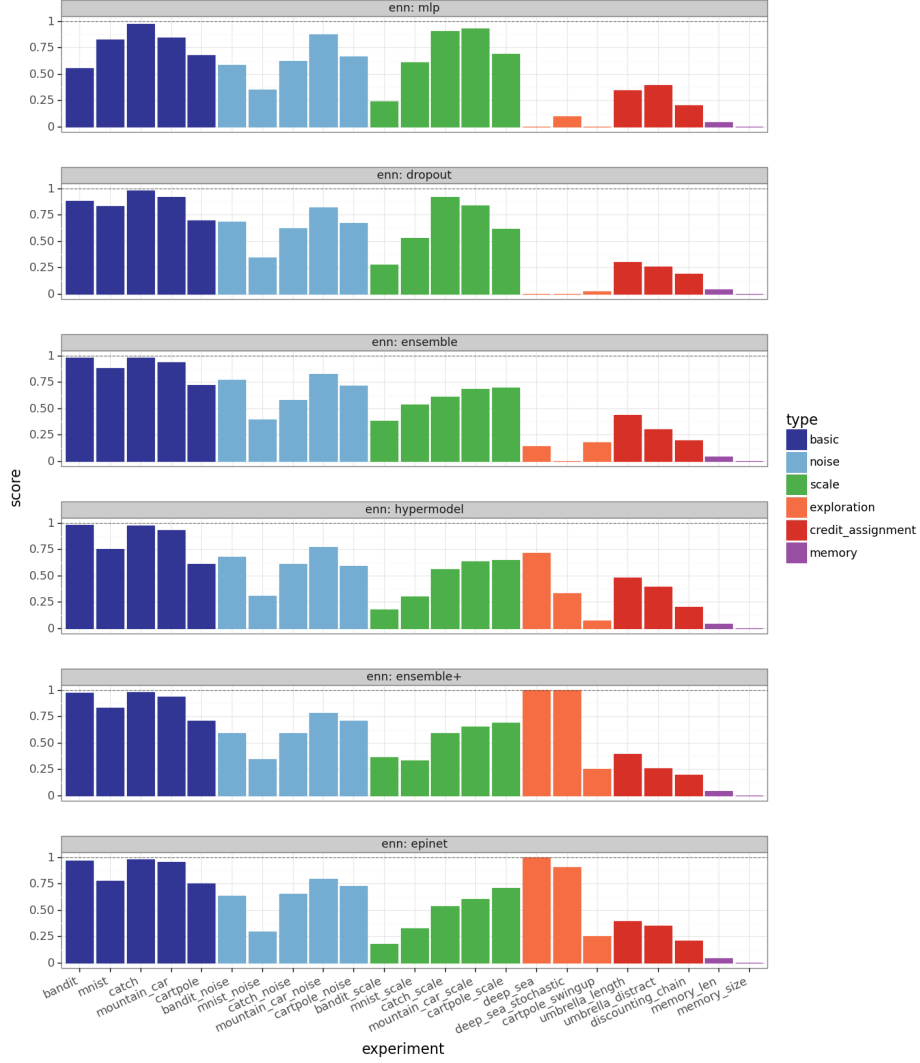


Figure 22: Summary score for each **bsuite** experiment.

H.3 Results commentary

- **mlp** performs well on basic tasks, and quite well on credit assignment, generalization, noise and scale. However, DQN performs extremely poorly across memory and exploration tasks. Our results match the high-level performance of the **bsuite/baselines**.
- **ensemble** performs similar to **mlp** agent. The additional diversity provided by random initialization in ensemble particles is insufficient to drive significantly different behaviour.
- **dropout** performs very similar to **mlp** agent. Different dropout masks are not sufficient to drive significantly different behaviour on **bsuite**.
- **hypermodel** performs better than **mlp**, **ensemble**, and **dropout** agents on exploration tasks, but the performance does not scale to the most challenging tasks in **bsuite**.
- **ensemble+** also known as Bootstrapped DQN (Osband et al., 2016a, 2018). Mostly performs similar to **ensemble** agent, except for exploration where it greatly outperforms **mlp**, **ensemble**, and **dropout** agents. The addition of prior functions is crucial to this performance.
- **epinet** performs similar to **ensemble+** agent, but with much lower compute. We do see some evidence that, compared to other approaches **epinet** agent is less robust to problem *scale*. This matches our observation in supervised learning that **epinet** performance is somewhat sensitive to the chosen scaling of the prior networks σ^P .

None of the agents we consider have a mechanism for memory as they use feed-forward networks. We could incorporate memory by considering modifications to the agents, but we don't explore that here.