Brain-inspired replay in artificial neural networks

Current state-of-the-art deep neural networks can solve almost any task they are trained on. But when such a network is trained on a new task, the previously learned task is quickly forgotten. Importantly, this 'catastrophic forgetting' is not due to limited capacity of the network, as the same network could learn both tasks when trained in an interleaved fashion. In the real world, however, training examples are not presented interleaved but typically appear in sequences. A straight-forward solution would be to store the encountered examples from previously learned tasks and revisit them when learning new tasks. Although such 'replay' or 'rehearsal' solves catastrophic forgetting, in the deep learning community replay is typically believed not to be a scalable solution as constantly retraining on all previous problems is very inefficient and the amount of data that would have to be stored becomes unmanageable very quickly.

Yet, in the brain – which clearly has implemented an efficient and scalable algorithm for continual learning – the replay of previous experiences *is* important for stabilizing new memories. Inspired by this, here we revisit the use of replay as a tool for continual learning with artificial neural networks. We find that: (1) fully replaying previously learned problems is not needed, as a handful of replayed examples could be enough; (2) a perfect memory (i.e., storing all encountered examples) is not required, as a low capacity generative model could suffice; and (3) brain-inspired modifications enable generative replay to scale to complicated problems with many tasks (≥ 100) or complex inputs (natural images), resulting in state-of-the-art performance on challenging continual learning benchmarks. Moreover, when incrementally learning new classes (as opposed to new tasks), we find that replay might actually be *necessary*. This last result suggests a specific, so far unappreciated, computational goal for replay in the brain.

RESULTS: Firstly, we find that although most of the recent methods for continual learning with artificial neural networks (ANNs) are very successful for scenarios in which *tasks* must be learned incrementally, only replay-based methods are able to incrementally learn new *classes* (***Fig. 1***). To further strengthen the case for replay as a valuable tool for continual learning with ANNs, we demonstrate that generative replay – which does not rely on storing data, as it trains a generator to generate the data to be replayed – is surprisingly robust and efficient (***Fig. 2***). Despite these promising results on split MNIST, we nevertheless find that scaling up generative replay to more challenging problems with many tasks (≥ 100) or complex inputs (natural images) is not straight-forward (***Fig. 4***, red curves). One possible solution could be to use the recent progress in deep generative models to improve the quality of the generator, but this approach will be inefficient as high-quality generative models can be computationally very costly to train or sample from. Instead, we demonstrate that with several efficient, brain-inspired modifications (***Fig. 3***), we obtain state-of-the-art performance with generative replay on challenging benchmarks (***Fig. 4***, purple curves).

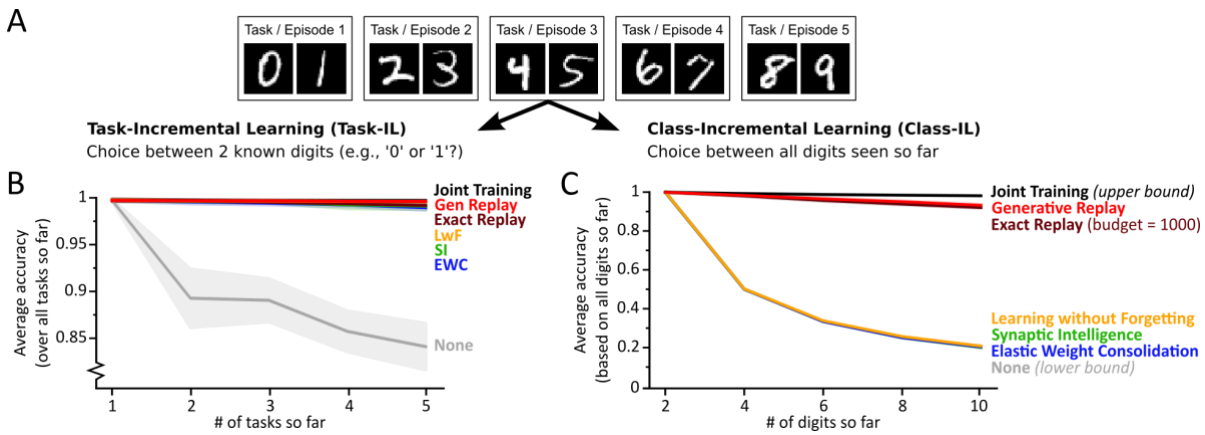

***Fig. 1**: **(A)** Split MNIST performed according to two different scenarios. **(B)** With task-incremental learning, all compared continual learning methods perform very well. **(C)** But with class-incremental learning, only methods using some form of replay are able to prevent catastrophic forgetting.*

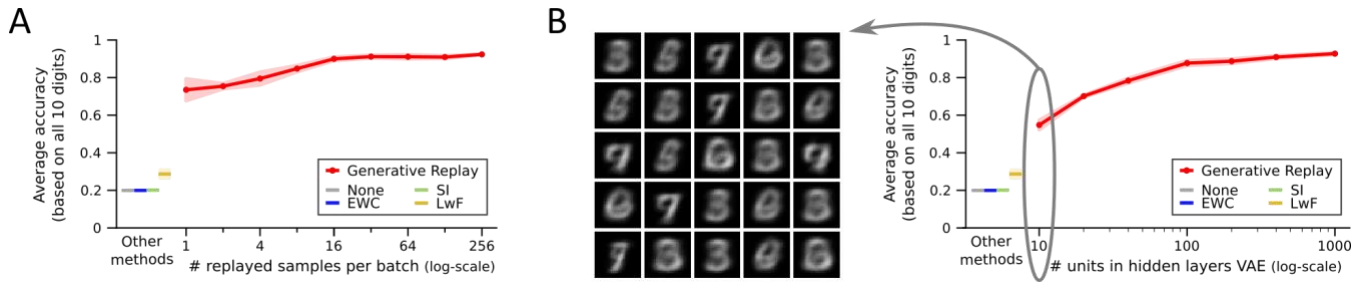

***Fig. 2**: Performance of generative replay on the class-incremental version of Split MNIST as a function of **(A)** the number of replayed samples per batch and **(B)** the number of units in the hidden layers of the variational autoencoder (VAE) used for generating replay. Also shown are random samples from the VAE with 10 hidden units after training on the 4th task (i.e., what is replayed during the final task).*

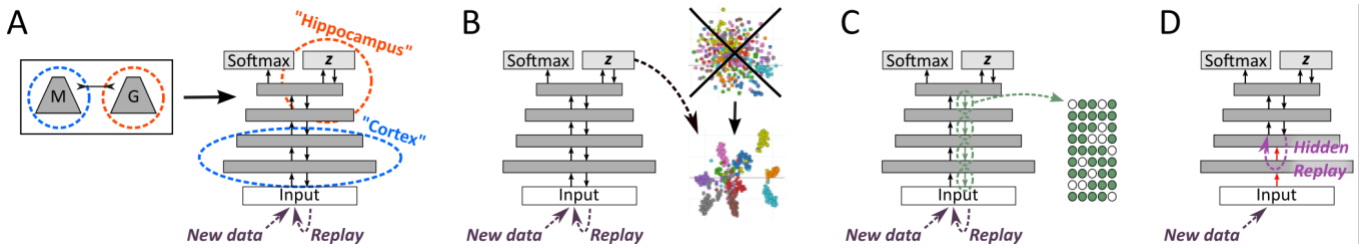

***Fig. 3**: Schematics of our brain-inspired modifications to the standard generative replay framework (Shin et al, 2017; NIPS). **(A)** Replay-through-Feedback. The generator [G] is merged into the main model [M] by equipping it with generative feedback or backward connections, resulting in a VAE with added softmax layer. **(B)** Conditional Replay. To enable the model to generate specific categories, the standard normal prior is replaced by a Gaussian mixture with a separate mode for each category. **(C)** Context gates. For every task or episode, a different subset of neurons in each layer is inhibited during the generative backward pass. **(D)** Internal replay. Instead of representations at the input level (e.g., pixel level), hidden or internal representations are replayed.*

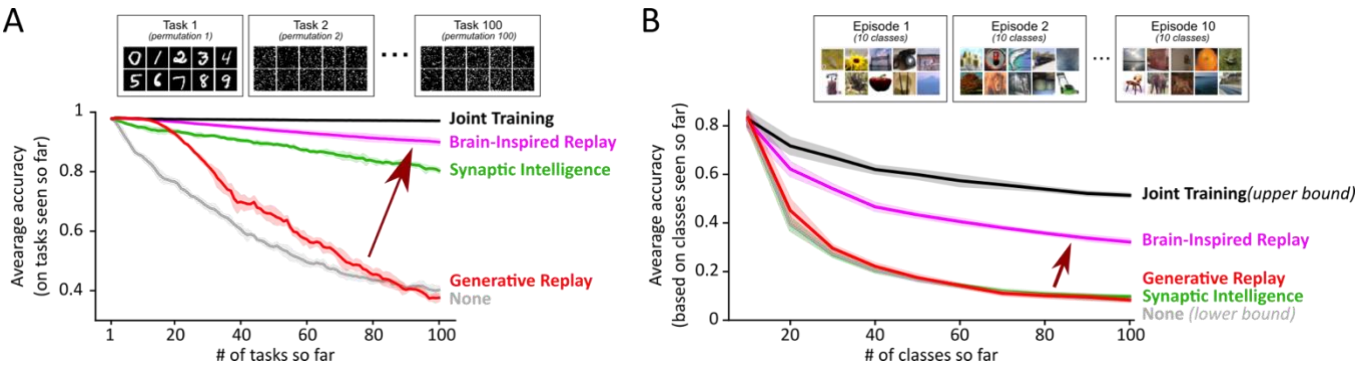

***Fig. 4**: Scaling up generative replay to more challenging problems is not straight-forward, but it can be achieved with our brain-inspired modifications. All of the compared methods use similar-sized networks. **(A)** Permuted MNIST with 100 different permutations. **(B)** Class-incremental learning on CIFAR-100, which is an unsolved benchmark in the continual learning literature; until now, only methods that explicitly store data (e.g., iCaRL) had been able to achieve acceptable performance on it.*

DISCUSSION: Besides providing a demonstration of how insights from neuroscience can make the performance of ANNs more human-like, our work generates new perspectives and hypotheses about the computational role and possible implementations of replay in the brain. In particular, our findings point towards a specific role for replay in *incremental* category learning and they highlight that replaying internal representations could be efficiently implemented by feedback connections using inhibitory gating.