Brain-Inspired Replay in Artificial Neural Networks

Gido M. van de Ven[1,2] & Andreas S. Tolias[1,3]

[1] Center for Neuroscience and Artificial Intelligence, Department of Neuroscience, Baylor College of Medicine, Houston, US
[2] Computational and Biological Learning Lab, Department of Engineering, University of Cambridge, UK
[3] Department of Electrical and Computer Engineering, Rice University, Houston, US

Introduction

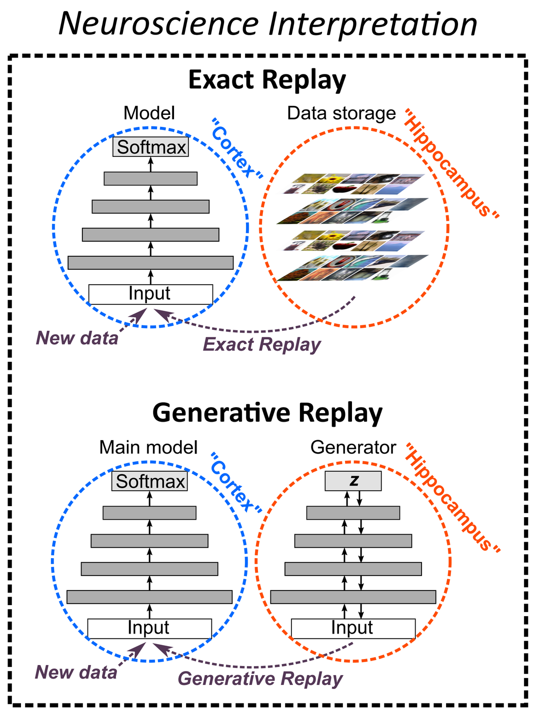
Like our brain, current state-of-the-art deep neural networks can be trained to impressive performance on a wide variety of individual tasks. Unlike our brain, when these networks are trained on a new task, they typically forget any previously learned tasks. A simple solution for such ‘catastrophic forgetting’ is to store examples from previous tasks and revisit them when learning new ones. Despite its effectiveness, such ‘replay’ or ‘rehearsal’ seems not to be a scalable solution as constantly retraining on all previous tasks is very inefficient and the amount of data to be stored becomes unmanageable quickly. Yet, in the brain – which clearly has implemented an efficient and scalable algorithm for continual learning – the replay of previous experiences *is* believed to be important for memory consolidation [1,2]. Inspired by this seeming paradox, here we revisit the use of replay as a tool for continual learning in artificial neural networks (ANNs).

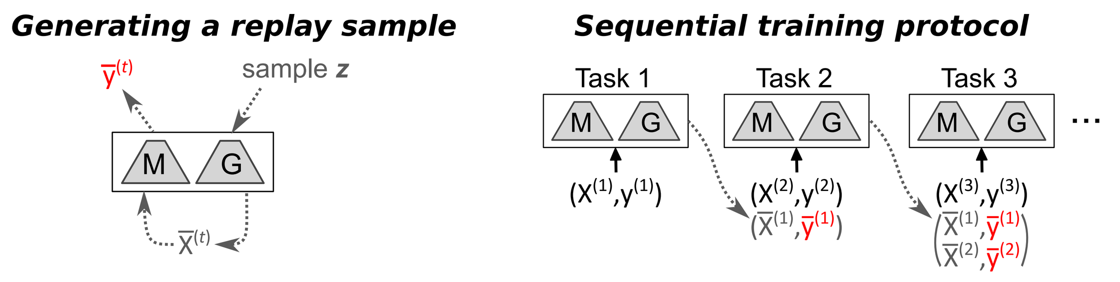
Part I: A Strong Case for Replay

Methods 1: Adding replay to artificial neural networks

Exact Replay (ER): Data from previous tasks is stored and interleaved with the training data of later tasks. However, storing data is not always possible (e.g., privacy / safety concerns, limited storage capacity) and it is unclear how the brain could do this.

Generative Replay (GR): An alternative is to generate the data to be replayed. Besides the main model (**M**) for solving the tasks (e.g., classifier), a separate generator (**G**) is trained [3]. We use a standard variational auto-encoder (VAE). When training on a new task, images generated by the (previous) generator are labelled by the (previous) classifier and interleaved with the current task's training data.





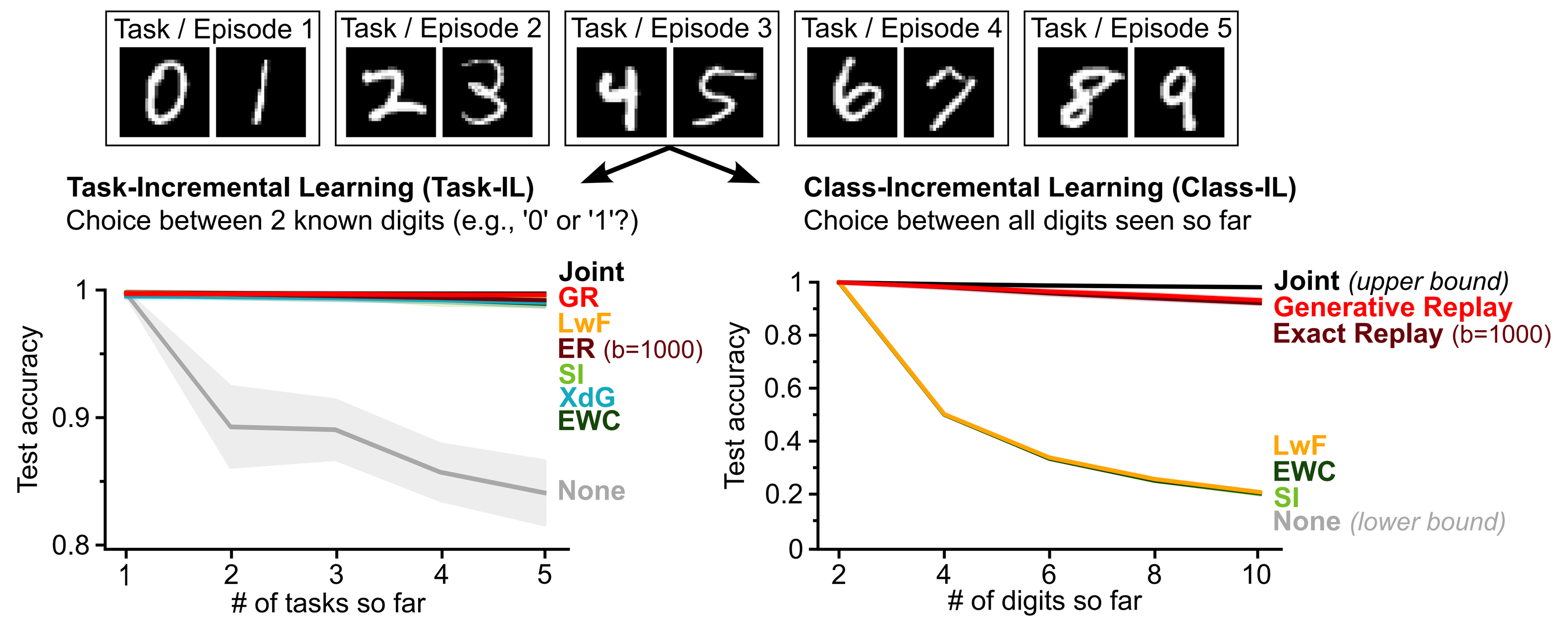
Methods 2: Other continual learning methods

Learning without Forgetting (LwF): Inputs from the current task are replayed, after they have been labelled by the model trained on the previous task [4].

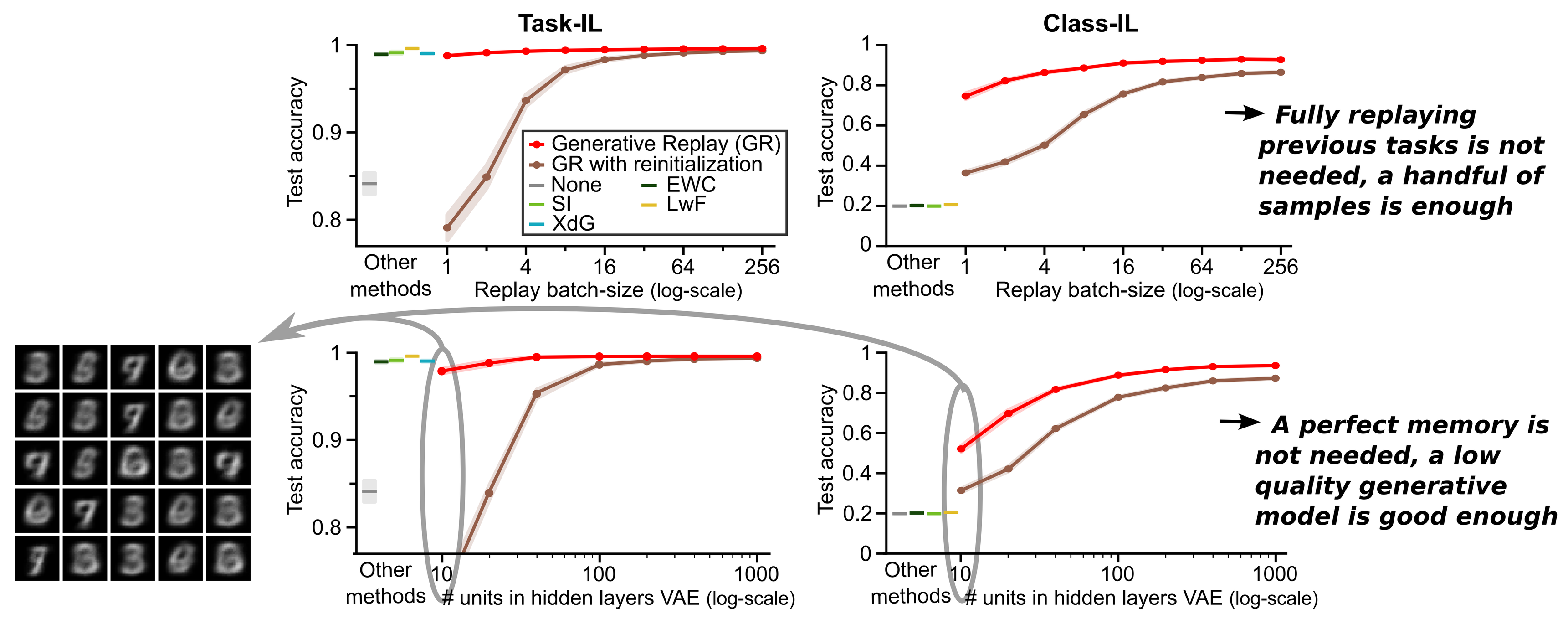
Elastic Weight Consolidation (EWC) / Synaptic Intelligence (SI): These metaplasticity-inspired methods slow down learning for parameters important for previous tasks, by adding a regularization term to the loss that penalizes changes to parameters depending on their estimated importance [5,6].

Context-dependent Gating (XdG): For each task, a different random subset of units is gated (i.e., their activations set to zero) [7]. As this method requires availability of task identity at test time, it cannot always be used.

Results 1: Class-incremental learning *requires* replay

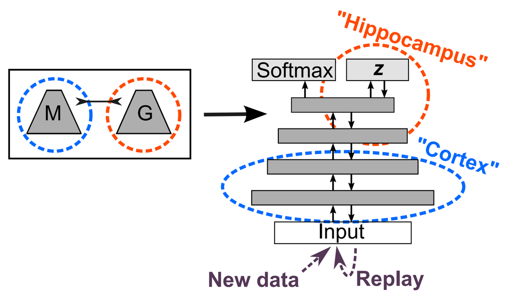


Results 2: Replay can be remarkably efficient and robust



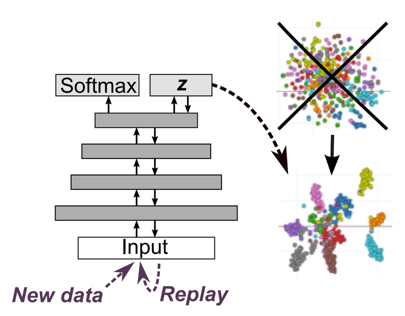
Part II: Brain-Inspired Modifications Enable Generative Replay to Scale

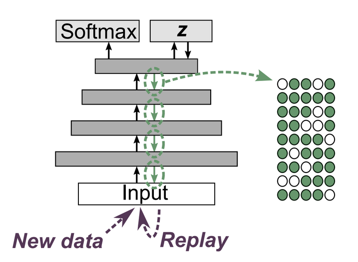
Methods 3: Brain-inspired replay (BI-R)



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.

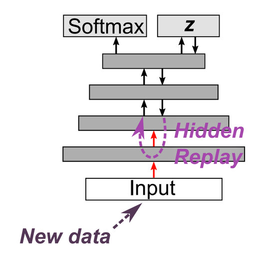
Conditional Replay: To enable the network to generate specific categories, the standard normal prior is replaced by a Gaussian mixture with a separate mode for each category.



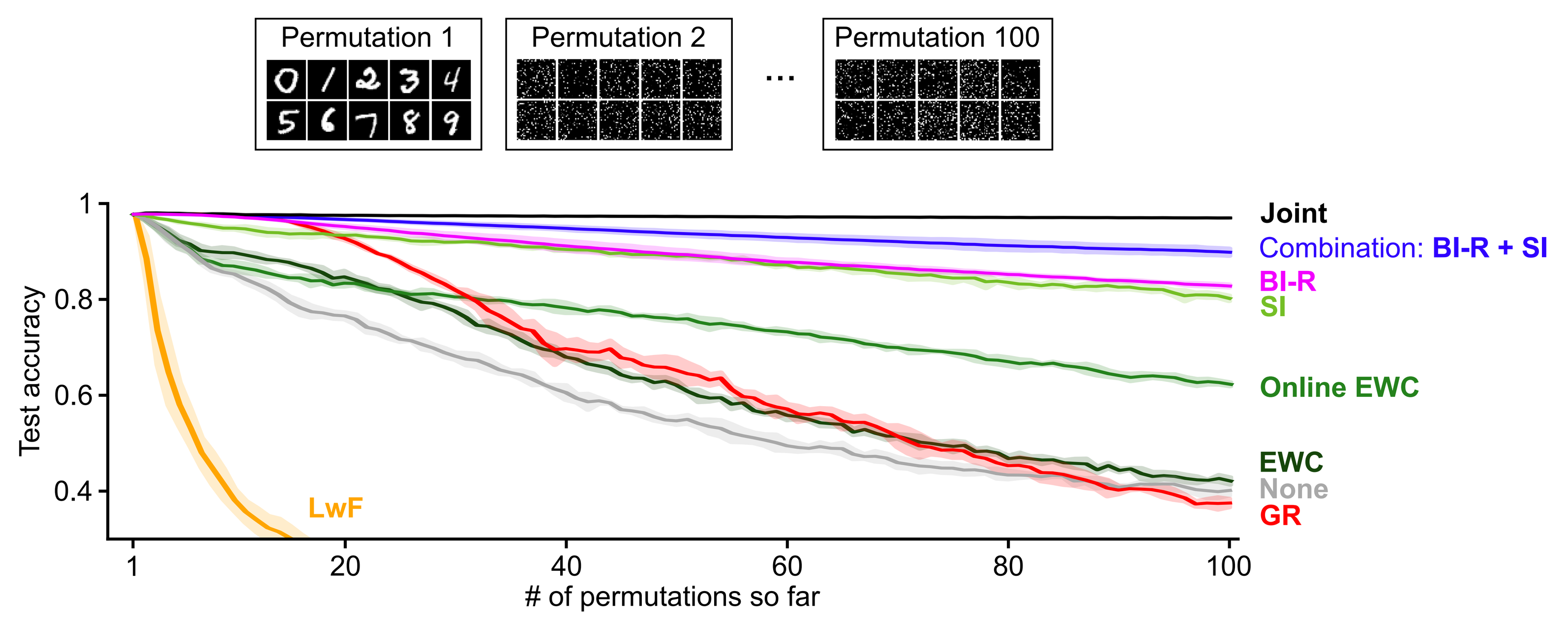


Gating based on Internal Context: For every task or class, a different, randomly selected subset of neurons is inhibited during the generative backward pass.

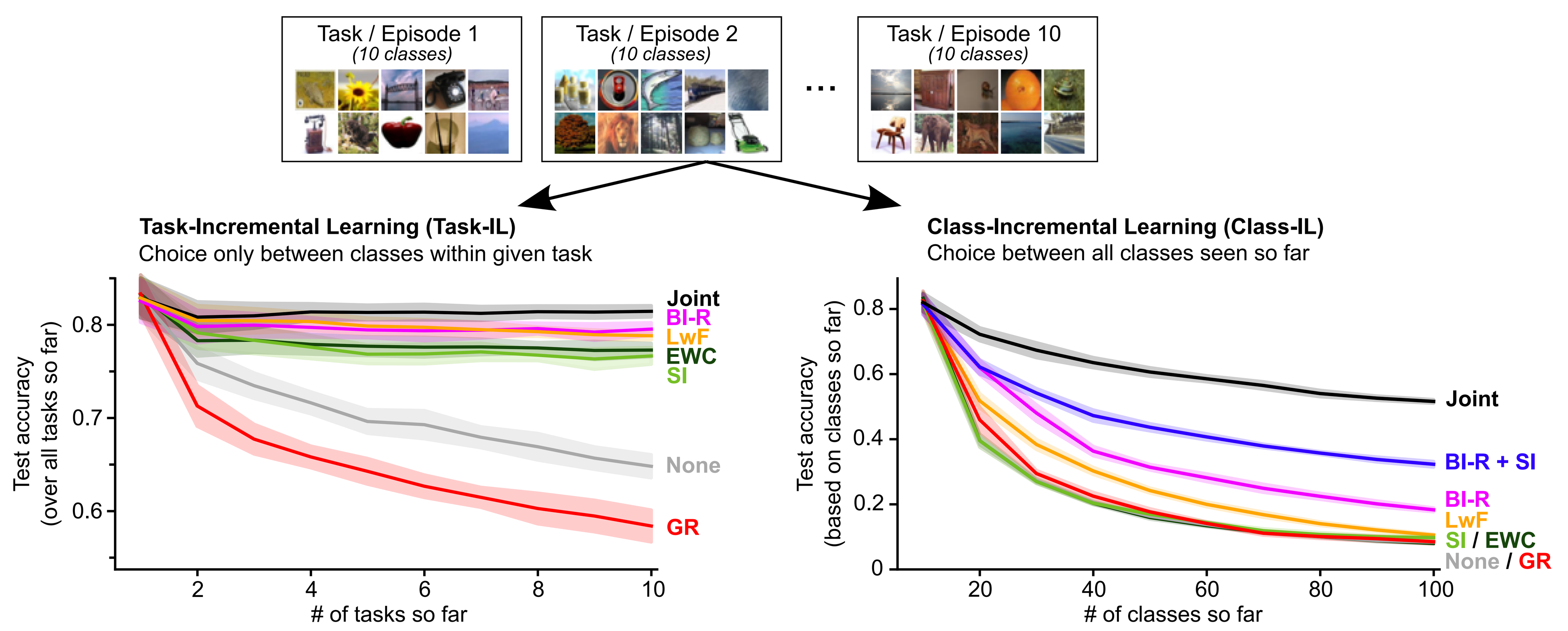
Internal Replay: Instead of representations at the input level (e.g., pixel level), hidden or internal representations are replayed.



Results 3: Scaling up to many tasks



Results 4: Scaling up to complex inputs



Highlights

- Replay might be necessary for *class*-incremental learning
- Generative replay can be remarkably efficient and robust
- Relatively simple, brain-inspired modifications enable generative replay to scale to more challenging problems
- Besides a SOTA solution for an unsolved problem in AI, our work provides a novel model for replay in the brain

References

[1] Wilson & McNaughton (1994) Reactivation of hippocampal ensemble memories during sleep. *Science* **265**: 676-679.
[2] van de Ven *et al.* (2016) Hippocampal offline reactivation consolidates recently formed cell assembly patterns during sharp wave-ripples. *Neuron* **92**: 968-974.
[3] Shin *et al.* (2017) Continual learning with deep generative replay. *NIPS*: 2994-3003.
[4] Li & Hoiem (2017) Learning without forgetting. *IEEE Trans Pattern Anal Mach Intell.*
[5] Kirkpatrick *et al.* (2017) Overcoming catastrophic forgetting in neural networks. *PNAS* **114**: 3521-3526.
[6] Zenke *et al.* (2017) Improved multitask learning through synaptic intelligence. *arXiv*: 1703.04200.
[7] Masse *et al.* (2018) Alleviating catastrophic forgetting using context-dependent gating and synaptic stabilization. *PNAS* **115**: E10467-E10475.

Acknowledgements

We thank Hava Siegelmann, Mengye Ren, Zhe Li, Johannes Oswald, Zhengwen Zeng and Máté Lengyel for comments and/or useful suggestions. This research project has been supported by an IBRO-ISN Research Fellowship, by the Lifelong Learning Machines (L2M) program of the Defence Advanced Research Projects Agency (DARPA) via contract number HR0011-18-2-0025 and by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior/Interior Business Center (DoI/IBC) contract number D16PC00003. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA, IARPA, DoI/IBC, or the U.S. Government.