Bernstein Conference – Oct 1st 2020

Generative replay in deep neural networks as a model for reactivation in the brain

Gido M. van de Ven[1,2*], Hava T. Siegelmann[3], Andreas S. Tolias[1,4]

[1] Center for Neuroscience and Artificial Intelligence, Department of Neuroscience, Baylor College of Medicine, Houston, TX 77030, USA
[2] Computational and Biological Learning Lab, Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, UK
[3] College of Computer and Information Sciences, University of Massachusetts Amherst, Amherst, MA 01003, USA
[4] Department of Electrical and Computer Engineering, Rice University, Houston, TX 77251, USA
* email: ven@bcm.edu

Acknowledgements: 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 HR0011-18-2-0025 and by the Intelligence Advanced Research Projects Activity (IARPA) via contract D16PC00003.

Replay to protect memories against forgetting

- The reactivation of neuronal activity patterns representing previous experiences is thought to be an important brain mechanism for stabilizing new memories [Wilson & McNaughton, 1994 *Science*; Rasch & Born, 2007 *Curr Opin Neurobiol*]

- A neural network 'catastrophically' forgets previously learned tasks when presented with a new one, but could learn all tasks when trained in an interleaved fashion [McCloskey & Cohen, 1989 *Psych Learn Motiv*; Ratcliff, 1990 *Psych Rev*]

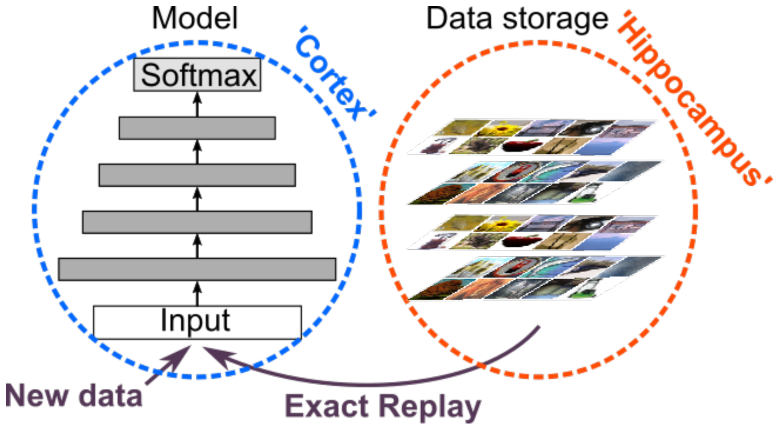
  → *Complementary Learning Systems*: memories are initially stored in the hippocampus, from where they are replayed to the cortex to enable interleaved learning [McClelland et al., 1995 *Psych Rev*]

- Replaying stored data has become important tool in deep learning

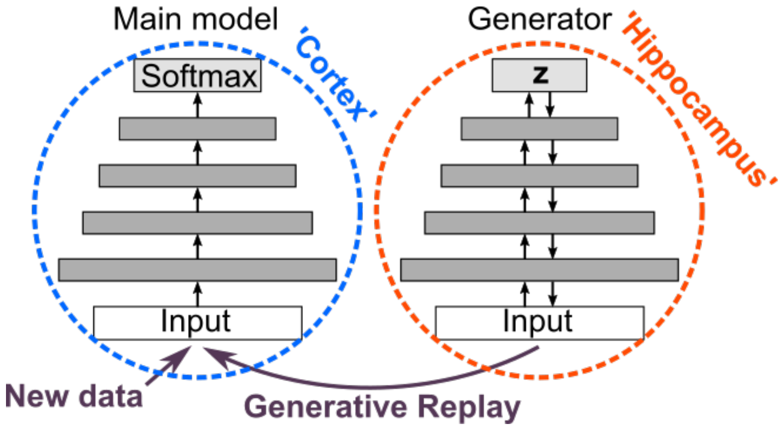
[Rolnick et al., 2019 *NeurIPS*]

Storing data is undesirable, alternative is to generate the data to be replayed

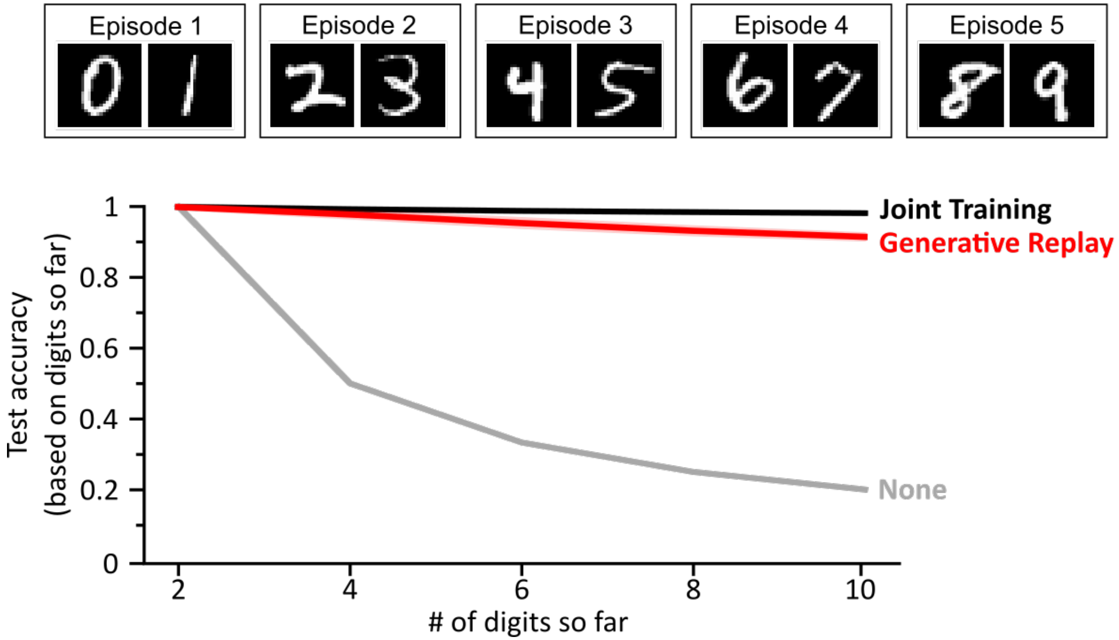
- Store data and interleave – *"exact"* or *"experience replay"*
  - *Initial argument for role of replay in memory consolidation* [McClelland et al., 1995 *Psych Rev*]
  - *But unclear how the brain could do directly store data*
  - Not always possible (e.g., privacy concerns, limited storage)

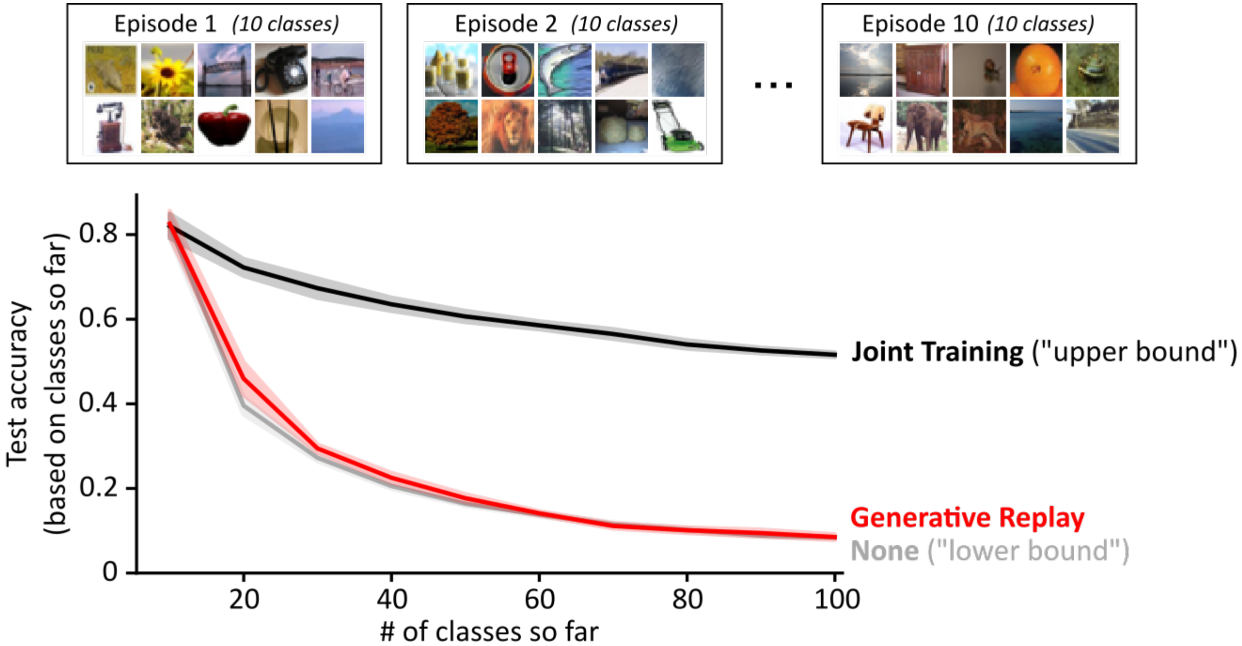


- Use a generative model – *"generative replay"*
  - *More realistic from neuroscience point of view*
  - *Views hippocampus as a generative neural network and replay as a generative process; see also* [Liu et al., 2018 *Neuron*; Liu et al., 2019 *Cell*]
  - Learning a generative model as a more scalable, privacy-preserving way of remembering previous seen data



Generative replay works well on toy problems, but breaks down with more complex inputs

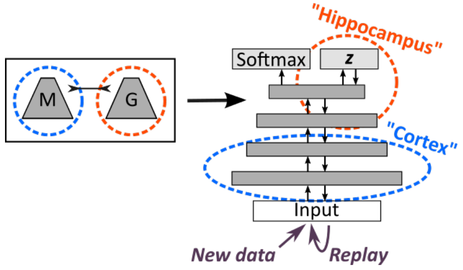




- The inability of replay to scale to realistic problems in a biologically plausible way (i.e., without storing data) raises doubt about how replay could underlie memory consolidation in the brain

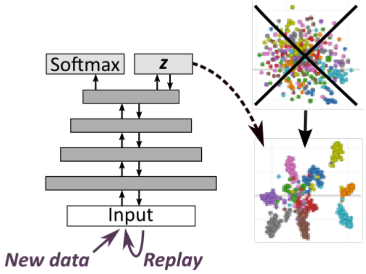
[van de Ven et al., 2020 *Nat Commun*; see also Lesort et al., 2019 *IJCNN*; Aljundi et al., 2019 *NeurIPS*]

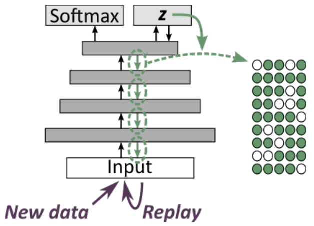
With brain-inspired modifications, generative replay can scale to challenging problems



Replay-through-Feedback: Merge generator into main model; replay is now generated by the feedback / backward connections

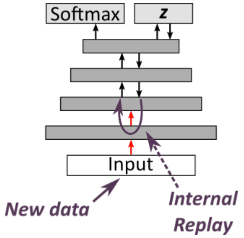
Conditional Replay: Enable model to generate specific classes, by replacing standard normal prior by Gaussian mixture with separate mode per class

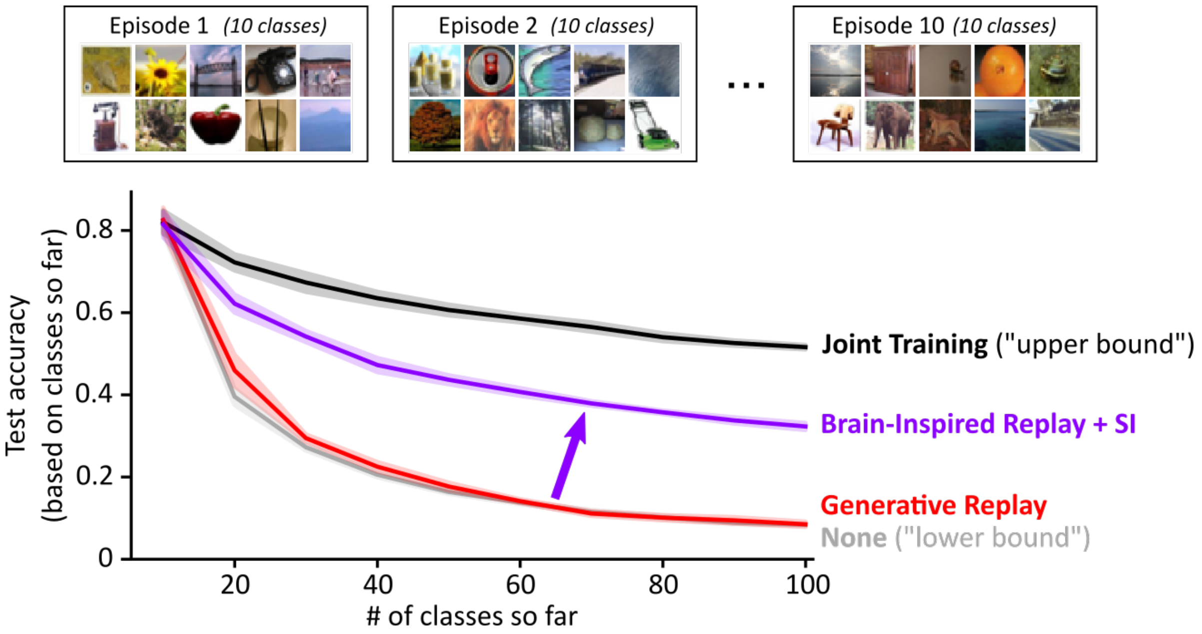




Gating based on Internal Context: For each class, inhibit (or gate) a different subset of neurons during generative backward pass

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





For details:

van de Ven GM, Siegelmann HT, Tolias AS (2020) Brain-inspired replay for continual learning with artificial neural networks. *Nature Communications*, **11**: 4069.