university of groningen

Computational principles of continual learning

Gido van de Ven

E-mail: g.m.van.de.ven@rug.nl

Website: https://gmvandeven.github.io

29 June 2026

What is continual learning?

- Learn “Task A”, then learn “Task B”, after which we care about the performance on ***both*** tasks

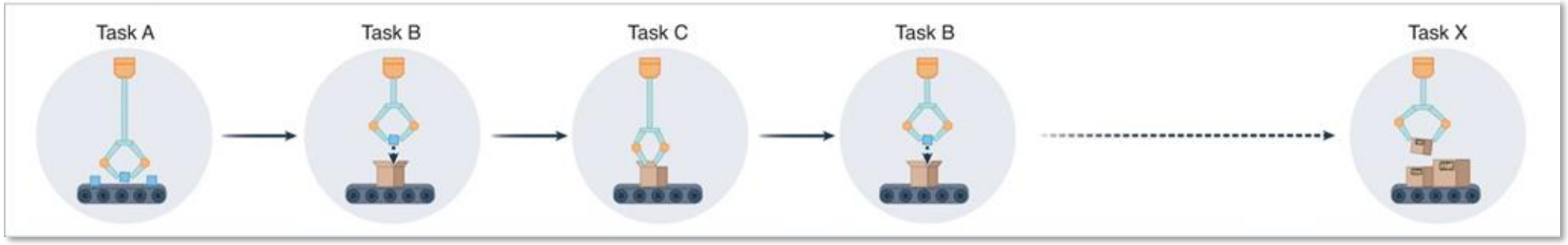


Figure source: Kudithipudi et al. (2022, Nature Machine Intelligence)

What is continual learning?

- Learn “Task A”, then learn “Task B”, after which we care about the performance on ***both*** tasks

- The problem of *incrementally* learning from a *non-stationary* data stream
  - Non-stationary: distribution of data from which is learned changes over time
  - Incrementally: new learning should not overwrite what was learned before, but knowledge should be accumulated

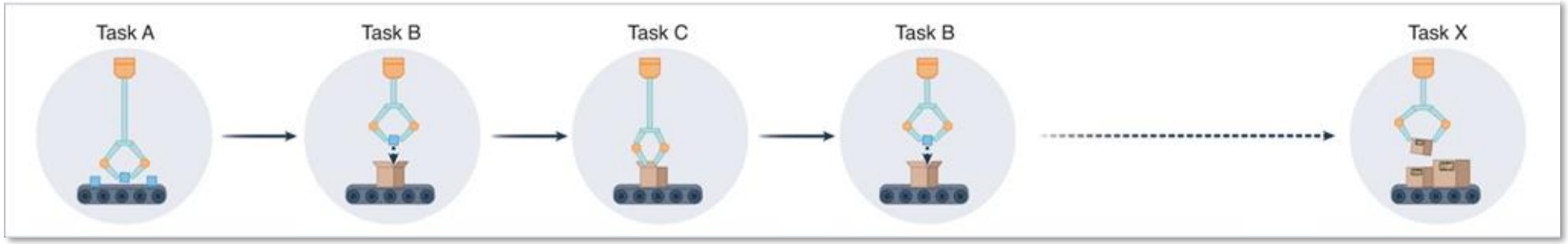


Figure source: Kudithipudi et al. (2022, Nature Machine Intelligence)

What is continual learning?

- Learn “Task A”, then learn “Task B”, after which we care about the performance on ***both*** tasks

- The problem of *incrementally* learning from a *non-stationary* data stream
  - Non-stationary: distribution of data from which is learned changes over time
  - Incrementally: new learning should not overwrite what was learned before, but knowledge should be accumulated

Continual learning is a **key aspect of intelligence**, but **very challenging** for current AI algorithms

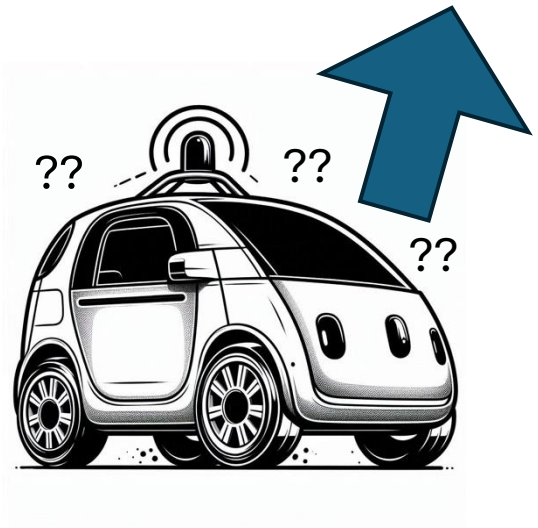
Motivating example: self-driving cars



Cartoon created with generative AI

Motivating example: self-driving cars

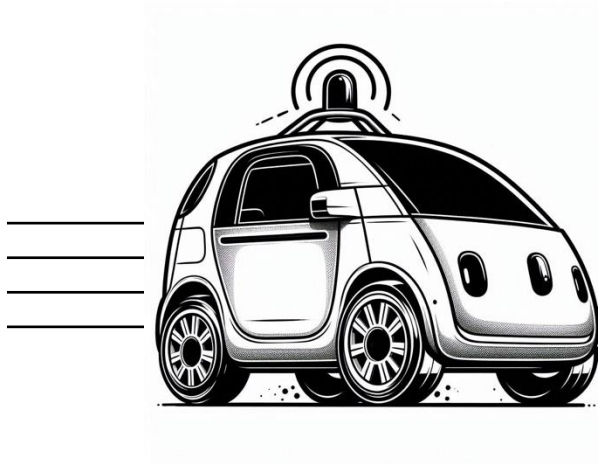
A LOT of training data



Cartoon created with generative AI

Motivating example: self-driving cars

A LOT of training data

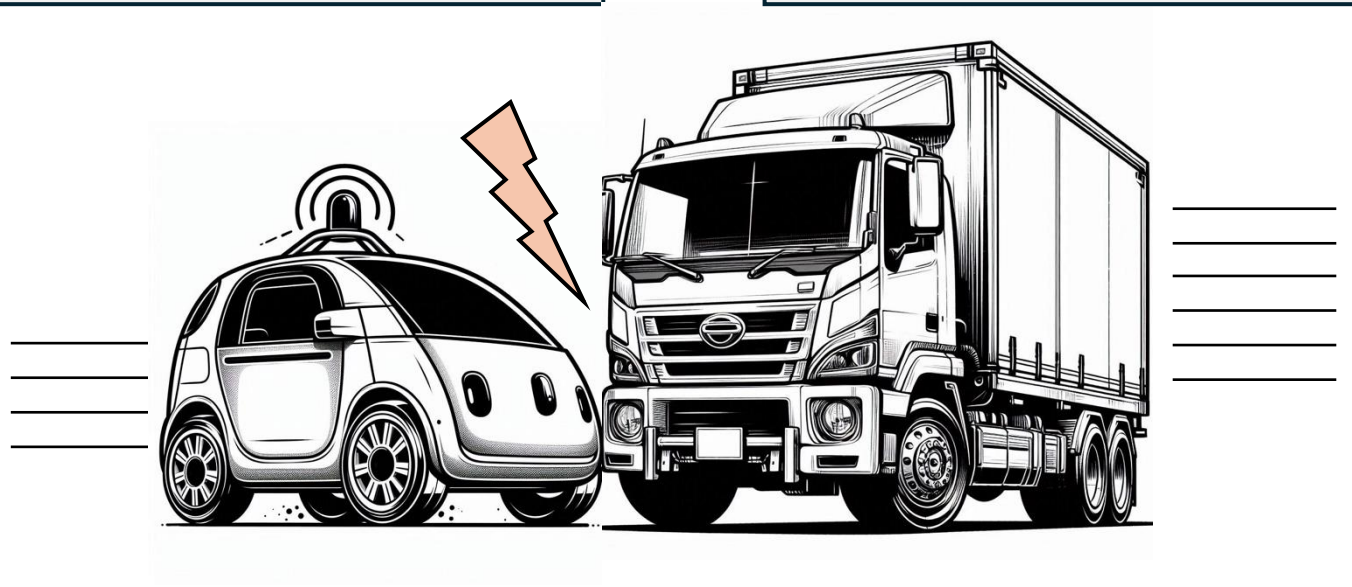


Cartoon created with generative AI

Motivating example: self-driving cars

A LOT of training data

More data of white trucks

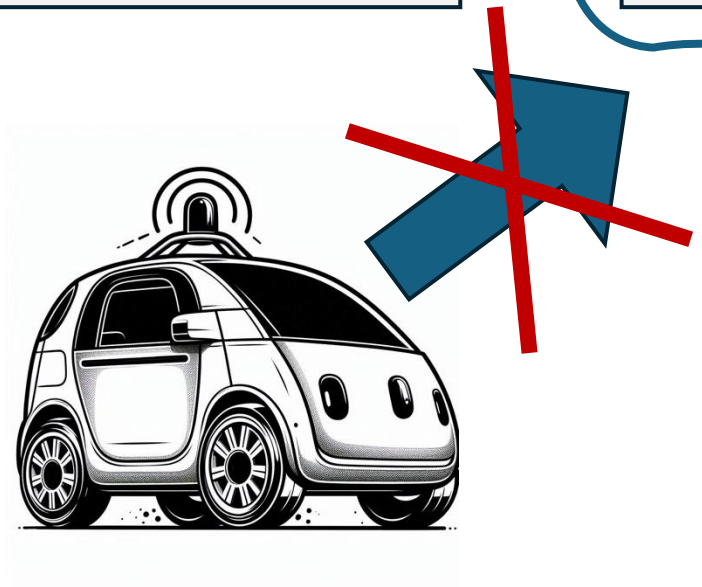


Cartoon created with generative AI

Motivating example: self-driving cars

A LOT of training data

More data of white trucks

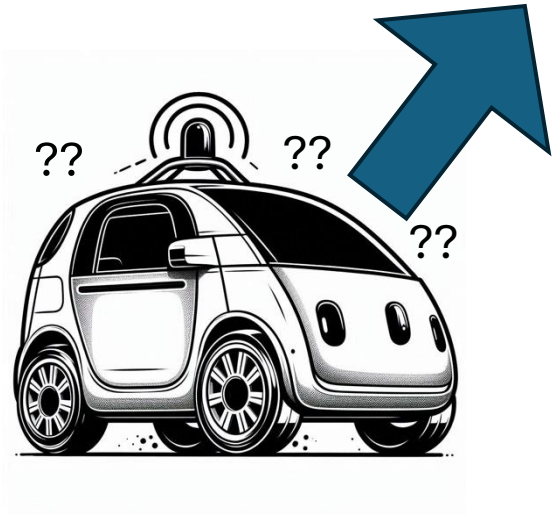


Cartoon created with generative AI

Motivating example: self-driving cars

A LOT of training data

More data of white trucks



→ **Inefficient!**

Cartoon created with generative AI

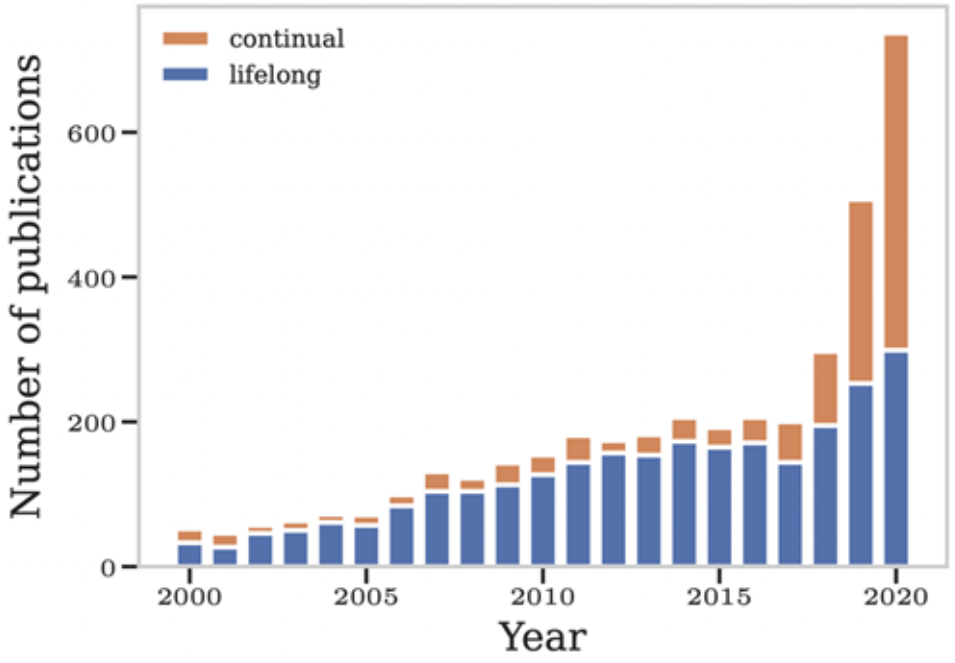
Promising solution: continual learning

- Add capabilities to existing models in much more efficient manners
- Hot topic in deep learning

But practical applications remain rare!

Much less studied from a cognitive science perspective!

Figure source: Mundt et al. (2022, ICLR)



My research goal

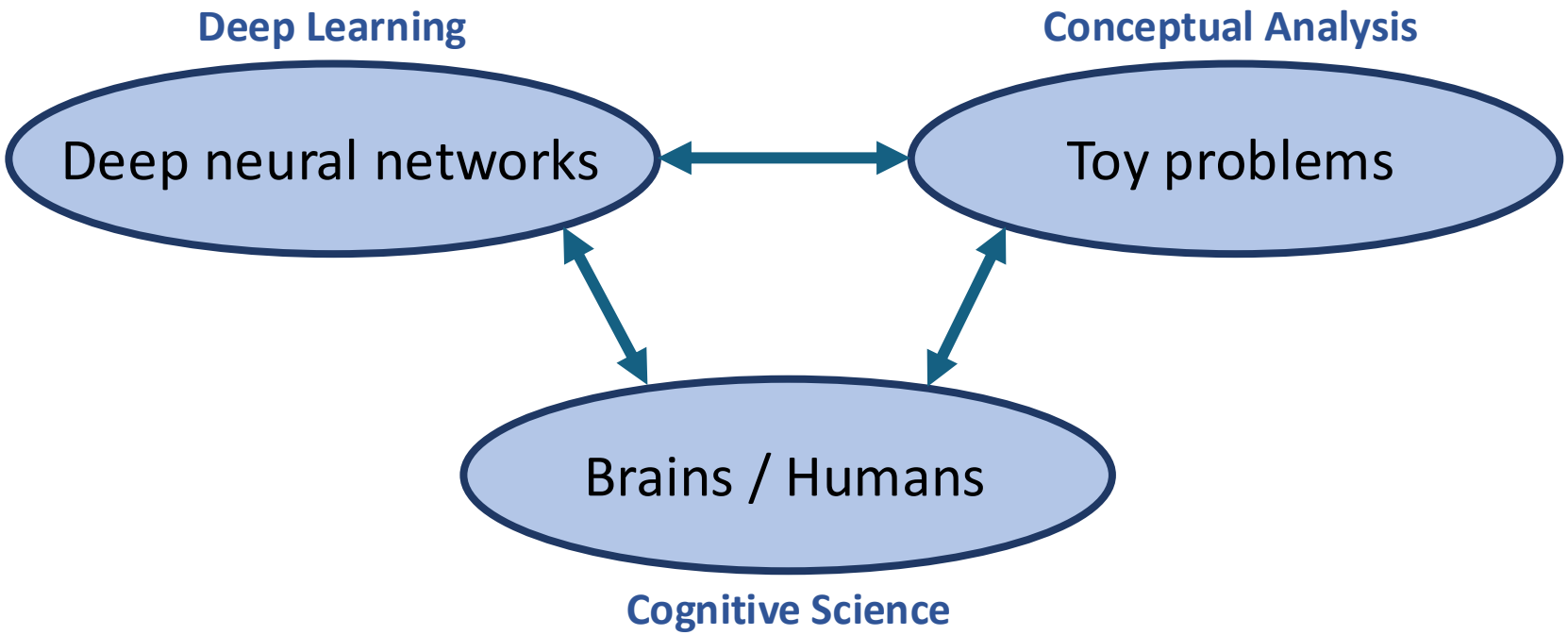
Understand the computational principles underlying continual learning

Important because:

- Helpful for desiging better continual learning algorithms for deep neural networks
- Gain insight into the computational mechanisms that might underlie an important aspect of our cognition

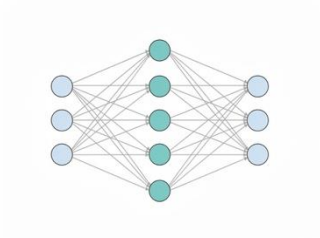
My approach

- Study the computational principles of continual learning in …

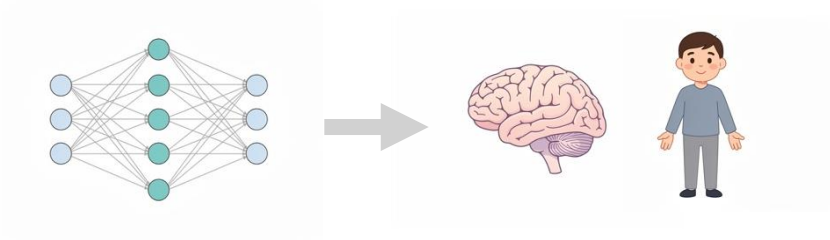


Overview

- Characterizing the continual learning problem
- Replay *[the current approach to continual learning]*
- Stability gap *[fundamental limitation of the current approach]*

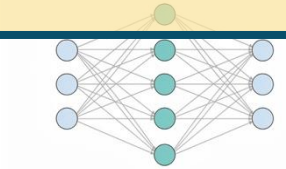


- Transient forgetting in humans?

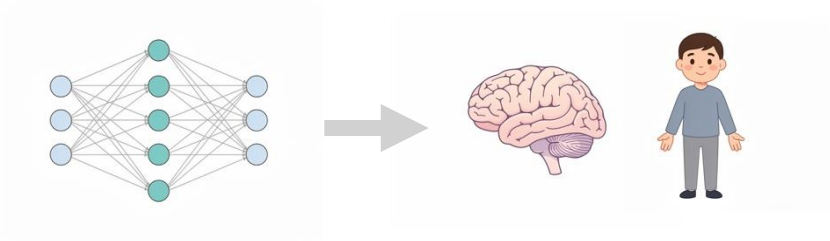


Overview

- **Characterizing the continual learning problem**
- Replay *[the current approach to continual learning]*
- Stability gap *[fundamental limitation of the current approach]*



- Transient forgetting in humans?



Why is continual learning difficult?

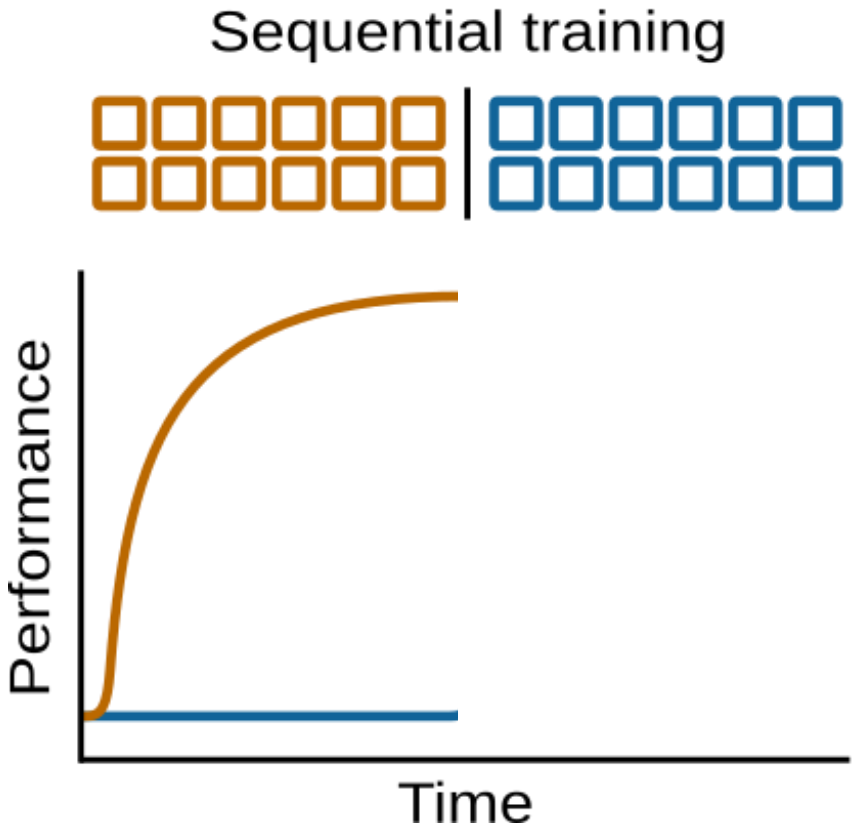


Figure source: **van de Ven**, Soures & Kudithipudi (2025, in: *Learning & Memory: A Comprehensive Reference, 3rd Edition)*

Catastrophic forgetting (McCloskey and Cohen, 1989; Ratcliff, 1990)

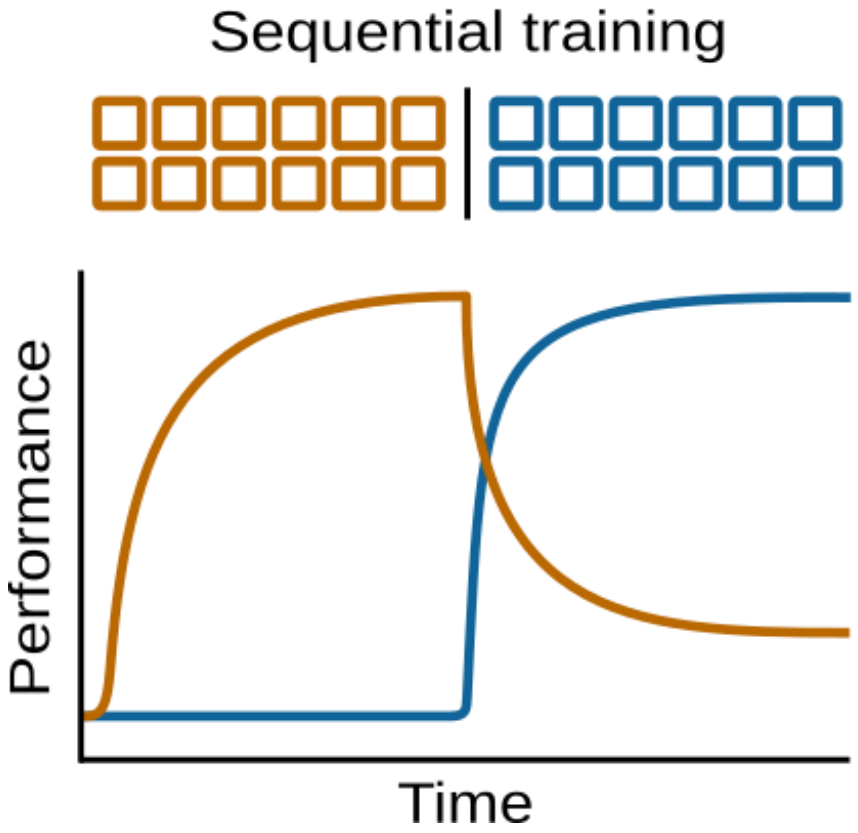


Figure source: **van de Ven**, Soures & Kudithipudi (2025, in: *Learning & Memory: A Comprehensive Reference, 3rd Edition)*

Catastrophic forgetting (McCloskey and Cohen, 1989; Ratcliff, 1990)

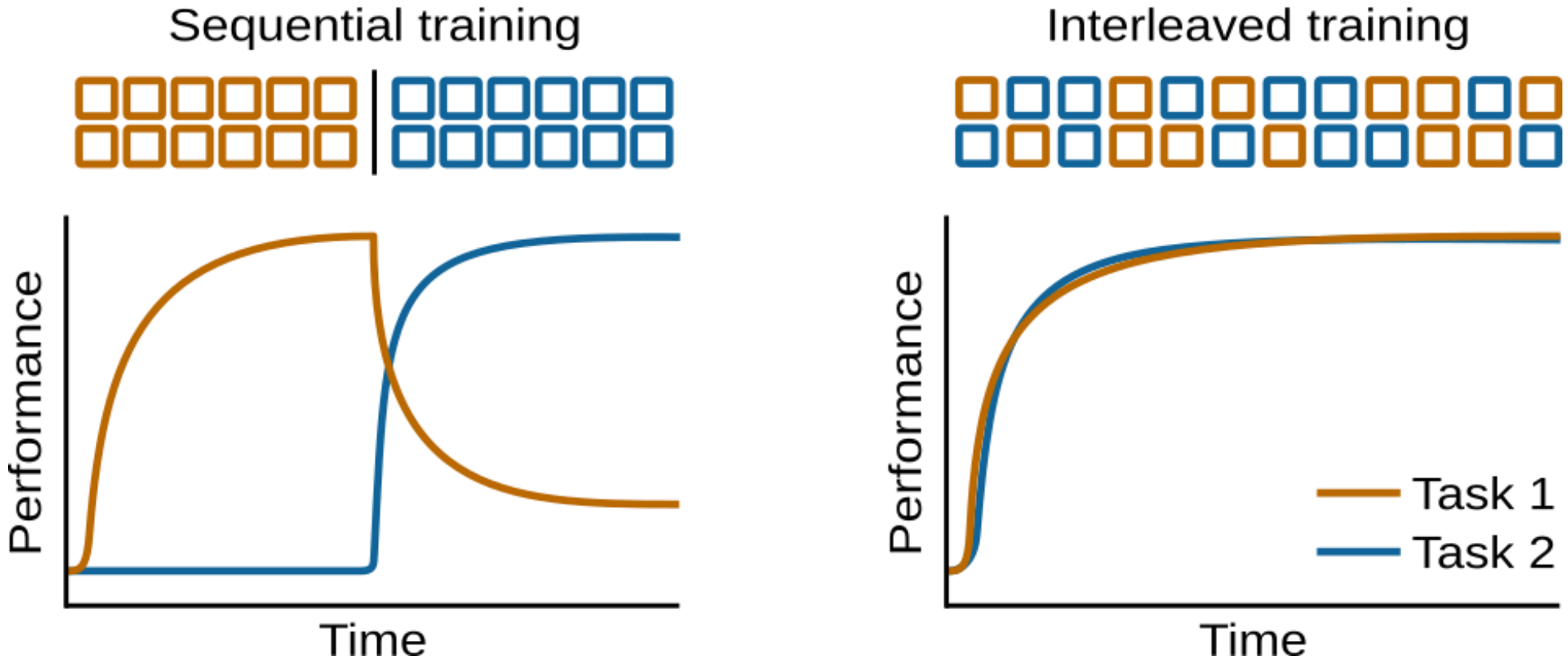


Figure source: **van de Ven**, Soures & Kudithipudi (2025, in: *Learning & Memory: A Comprehensive Reference, 3rd Edition)*

Catastrophic forgetting: expected or surprising?

- We first optimize the model for one task, and then we optimize it for another task. Why would the optimal solution for the second task still be good for the first one?

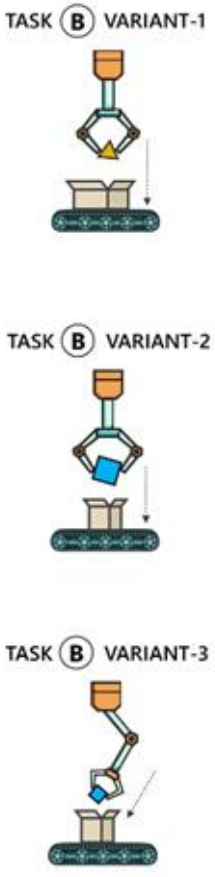
- One reasonable hypothesis:

  ``*If the modifications are all in the direction that helps the pattern that is being stored, there will be a conspiracy effect: The total help for the intended pattern will be the sum of all the small separate modifications. For unrelated patterns, however, there will be very little transfer of effect because some of the modifications will help and some will hinder. Instead of all the small modifications conspiring together, they will mainly cancel out.''* (pp. 81-82; Hinton et al., 1986)

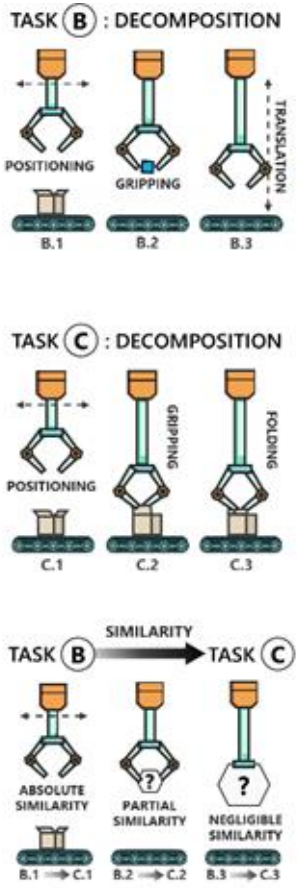
- McCloskey and Cohen (1989) and Ratcliff (1990) were the first to demonstrate catastrophic forgetting

Other features important for continual learning

Rapid adaptation



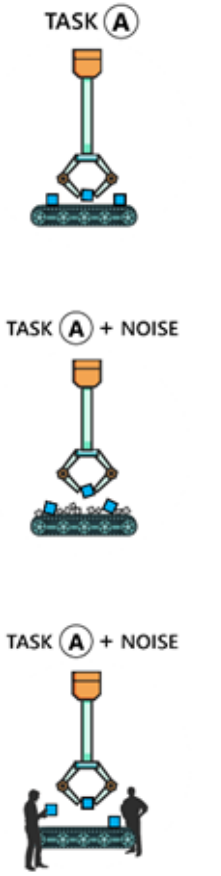
Exploiting task similarity



Task agnosticity



Noise tolerance



Resource efficiency and sustainability

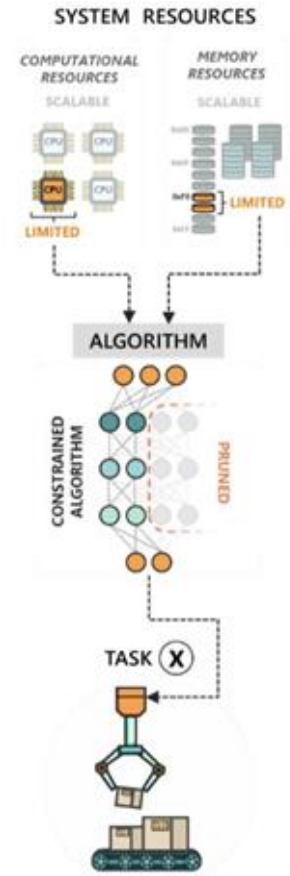


Figure source: Kudithipudi et al. (2022, Nature Machine Intelligence)

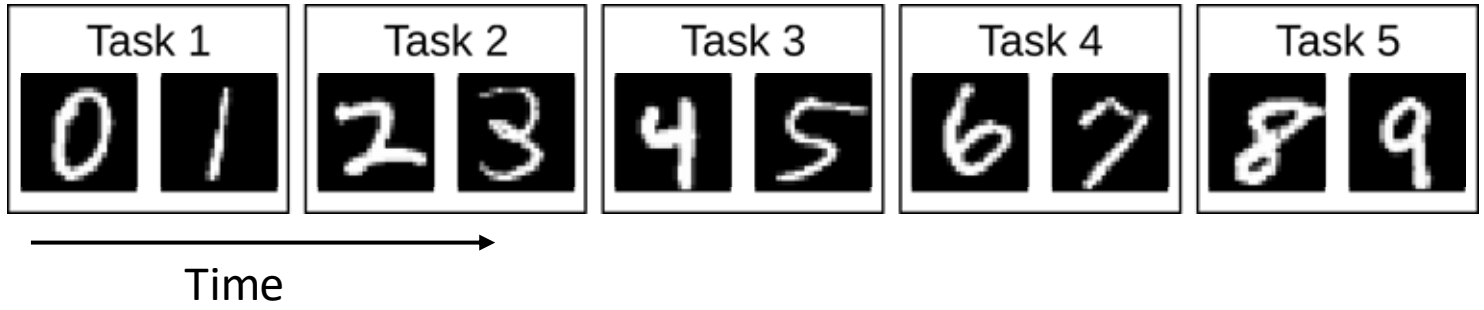
The canonical continual learning example: Split MNIST

- MNIST dataset is split in multiple tasks that are encountered sequentially
- After all tasks have been learned, the model should be good at all tasks

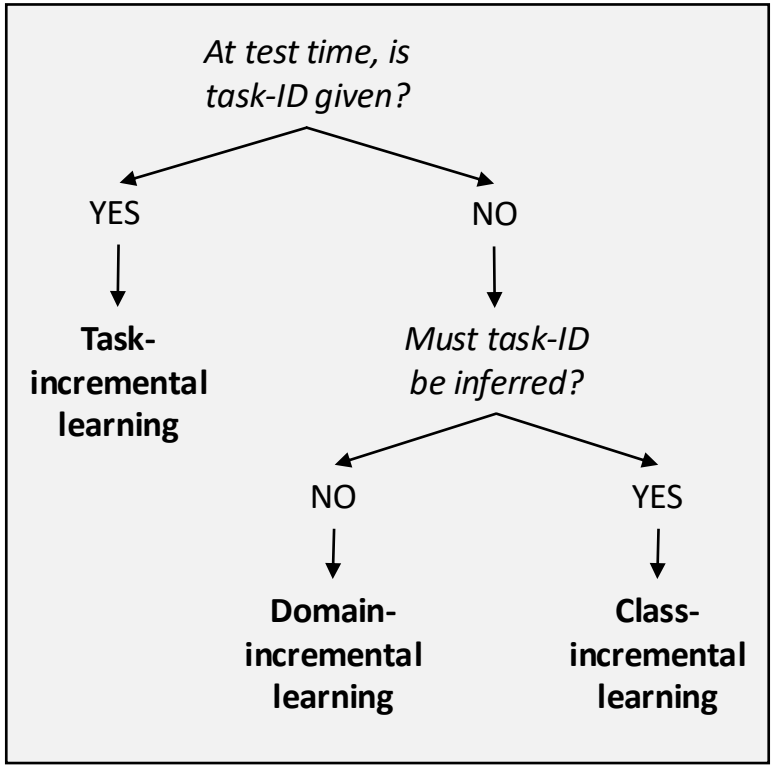


Three continual learning scenarios

Split MNIST:



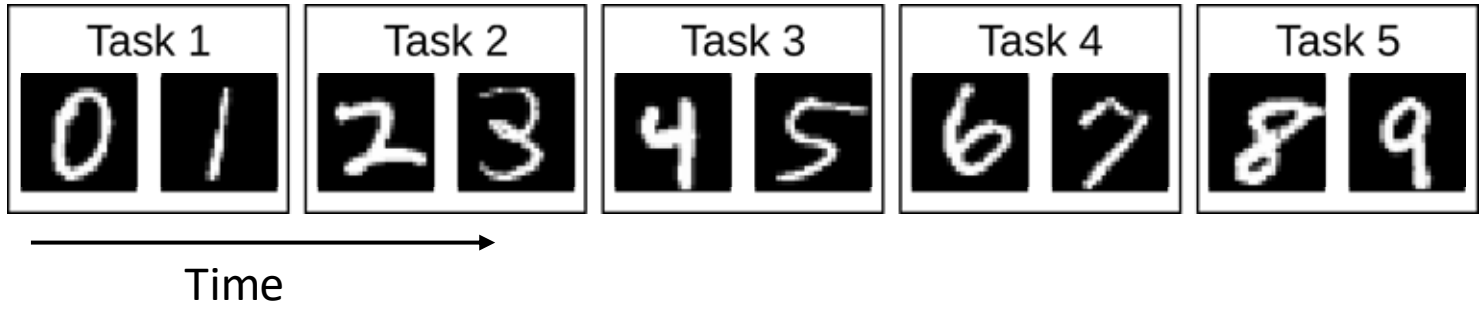
Scenario	*Type of choice*
Task-incremental learning	Choice between two digits of same task (e.g., 0 or 1?)
Domain-incremental learning	Is the digit odd or even?
Class-incremental learning	Choice between all ten digits



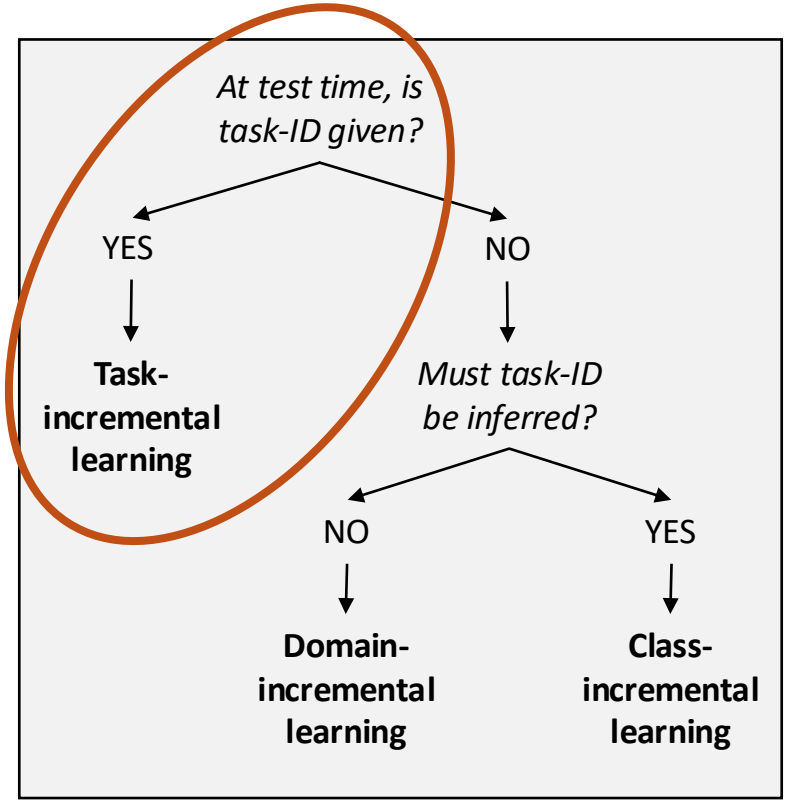
van de Ven & Tolias (2018) *NeurIPS workshop*

Three continual learning scenarios

Split MNIST:



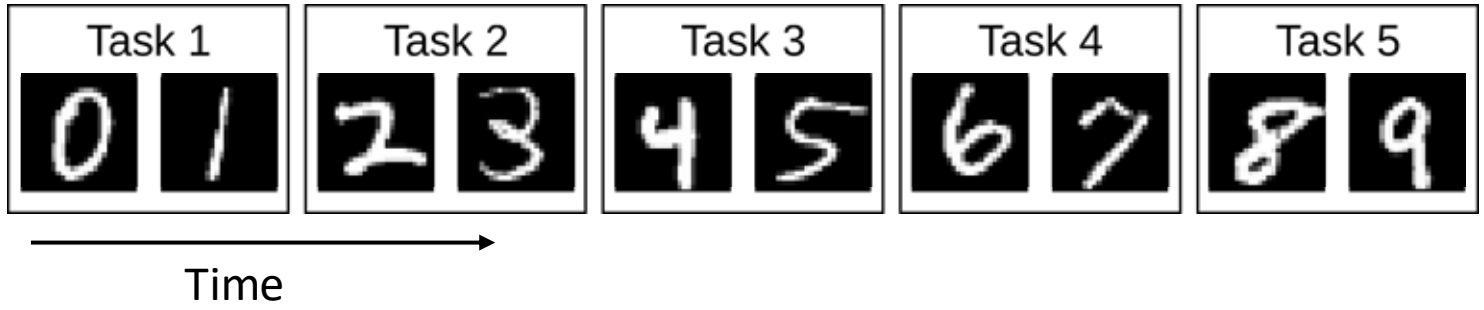
Scenario	*Type of choice*
Task-incremental learning	Choice between two digits of same task (e.g., 0 or 1?)
Domain-incremental learning	Is the digit odd or even?
Class-incremental learning	Choice between all ten digits



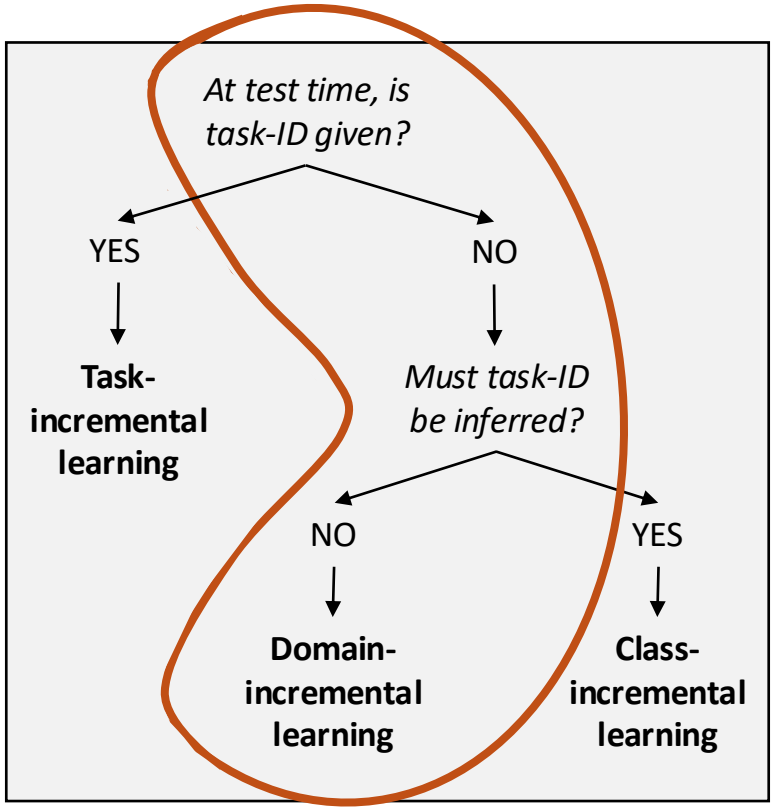
van de Ven & Tolias (2018) *NeurIPS workshop*

Three continual learning scenarios

Split MNIST:



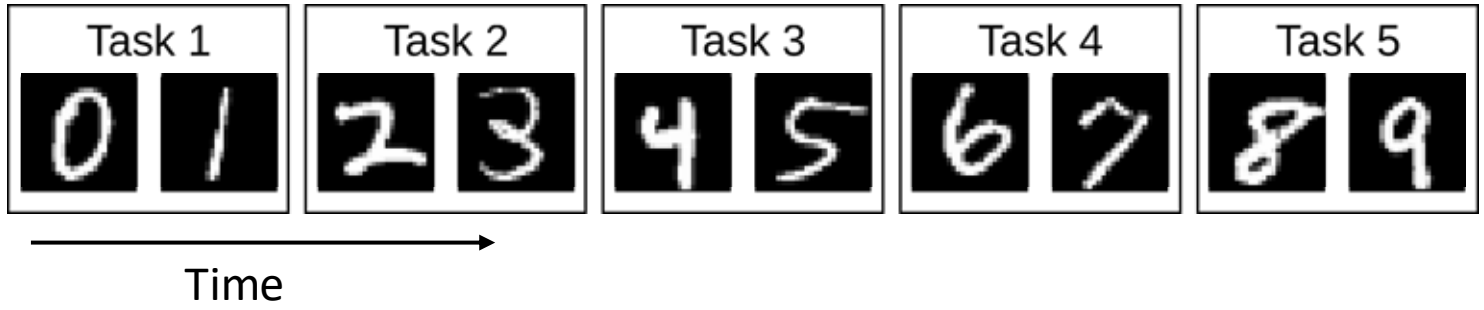
Scenario	*Type of choice*
Task-incremental learning	Choice between two digits of same task (e.g., 0 or 1?)
Domain-incremental learning	Is the digit odd or even?
Class-incremental learning	Choice between all ten digits



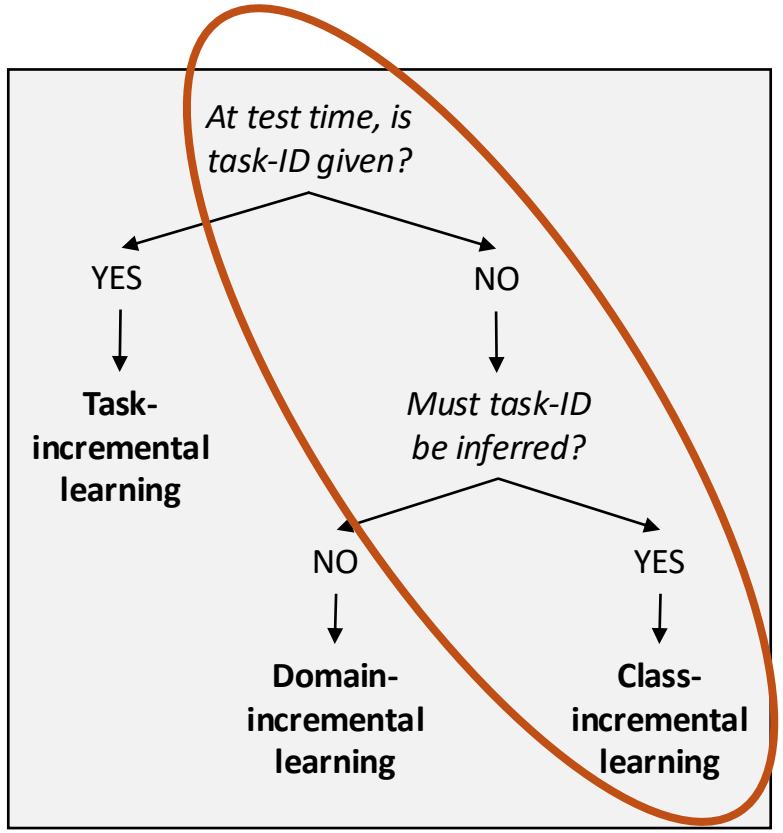
van de Ven & Tolias (2018) *NeurIPS workshop*

Three continual learning scenarios

Split MNIST:



Scenario	*Type of choice*
Task-incremental learning	Choice between two digits of same task (e.g., 0 or 1?)
Domain-incremental learning	Is the digit odd or even?
Class-incremental learning	Choice between all ten digits

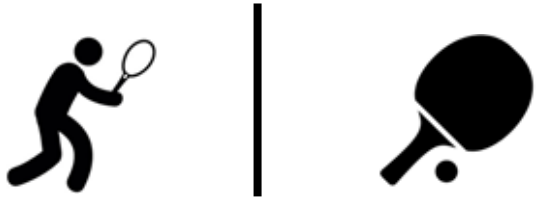


van de Ven & Tolias (2018) *NeurIPS workshop*

Three types of incremental learning

- Task-incremental learning *(Task-IL)*
  - Incrementally learn a set of clearly distinguishable tasks

Important challenge: achieve positive transfer between tasks



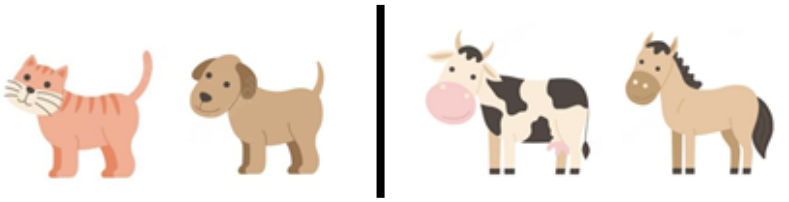
- Domain-incremental learning *(Domain-IL)*
  - Learn the same type of problem in different contexts

Important challenge: alleviate catastrophic forgetting



- Class-incremental learning *(Class-IL)*
  - Incrementally learn a growing number of classes

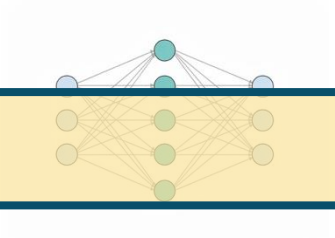
Important challenge: learn to discriminate between objects not observed together

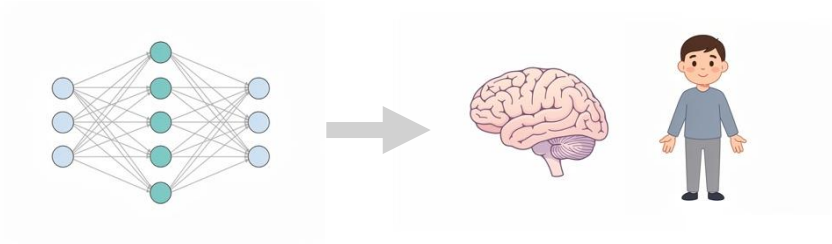


Images designed by Freepik

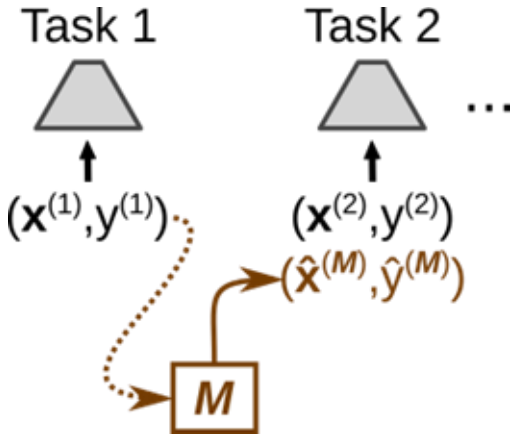
Overview

- Characterizing the continual learning problem
- **Replay** ***[the current approach to continual learning]***
- Stability gap *[fundamental limitation of the current approach]*
- Transient forgetting in humans?





Replay



Cognitive (neuro)science:

- Replay is hypothesized to have an important role in memory consolidation

[e.g., **van de Ven** et al. (2016) *Neuron*]



Deep learning:

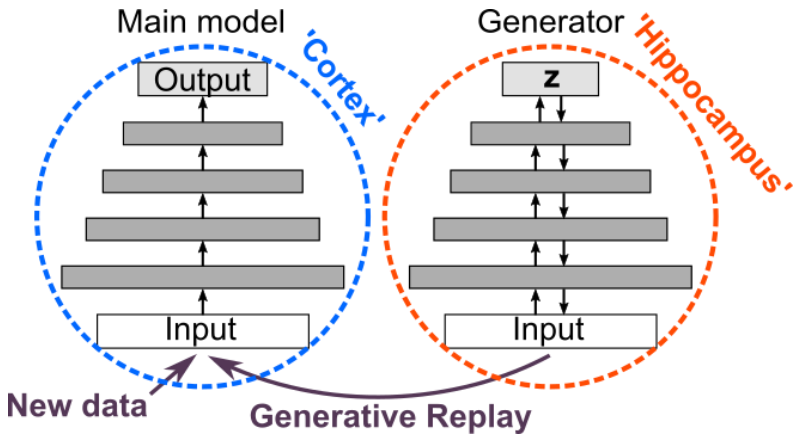
- Interleaved learning prevents catastrophic forgetting

But:

- Does the brain simply store its past experiences?
- Seems very inefficient to constantly retrain on all past data

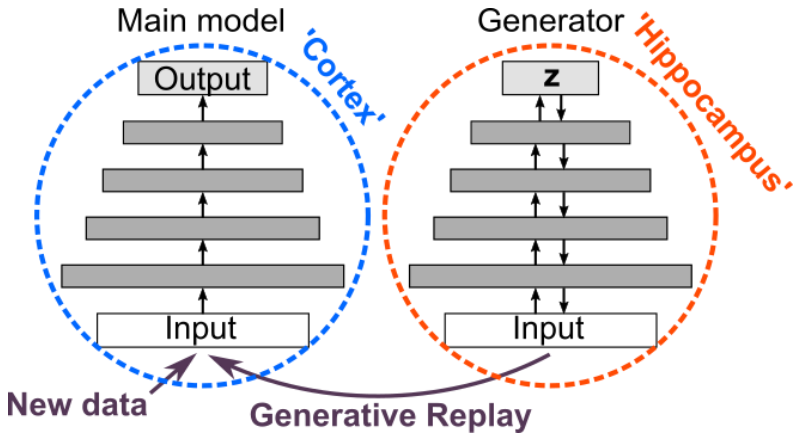
Generative replay

- Generative Replay (Shin et al., 2017; *NeurIPS*)
  - Works great on MNIST, but fails with more complex inputs (e.g. natural images)

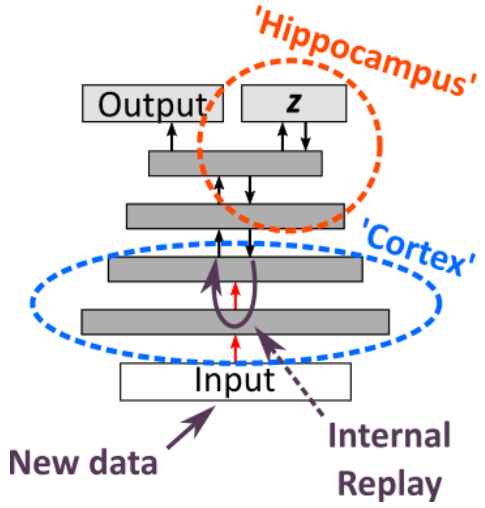


Generative replay → Brain-inspired replay

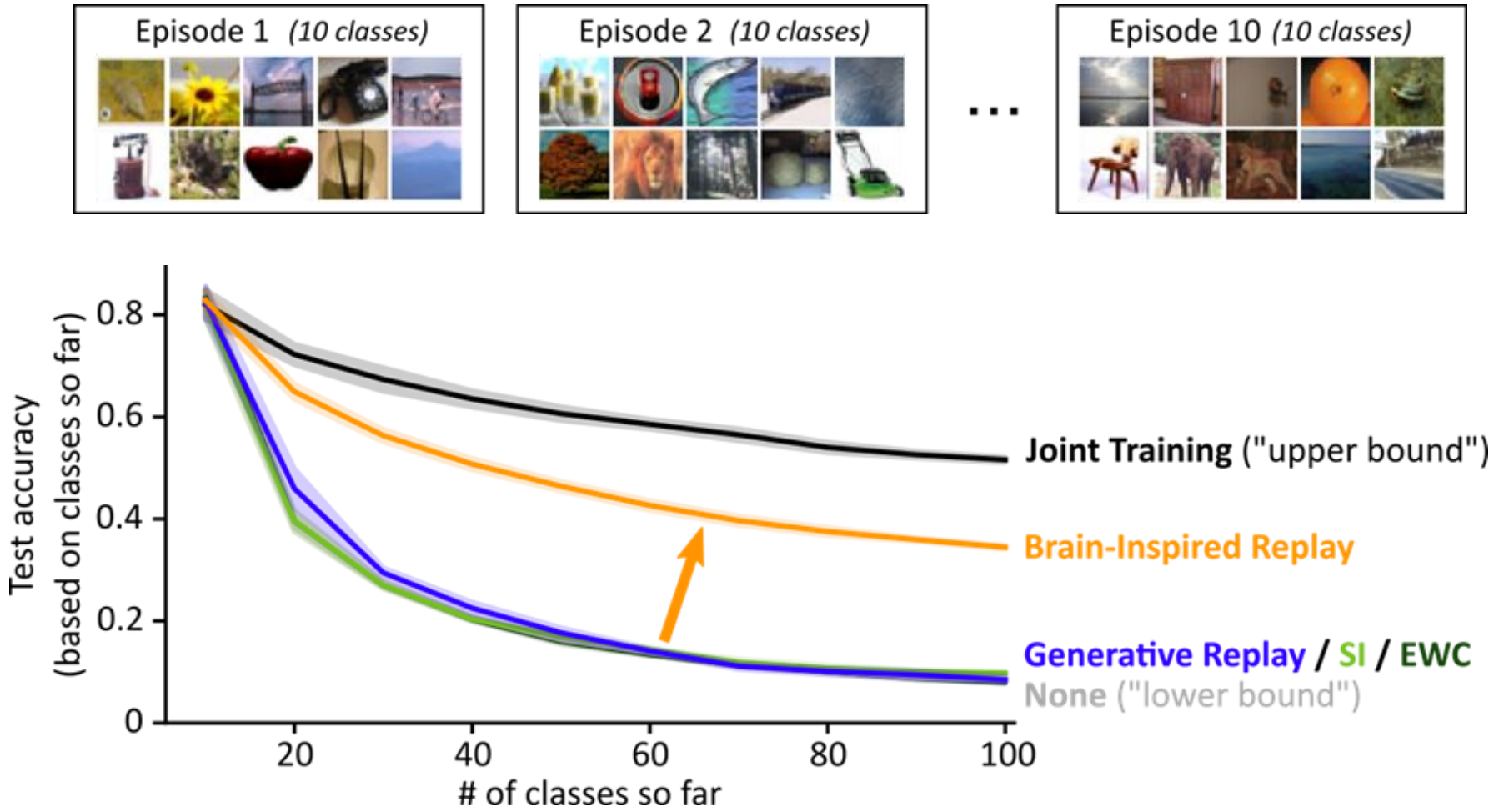
- Generative Replay (Shin et al., 2017; *NeurIPS*)
  - Works great on MNIST, but fails with more complex inputs (e.g. natural images)



- Brain-inspired replay [**van de Ven** et al. (2020) *Nature Communications*]
  - Replay abstract, high-level representations, rather than detailed, input level samples



Generative replay → Brain-inspired replay

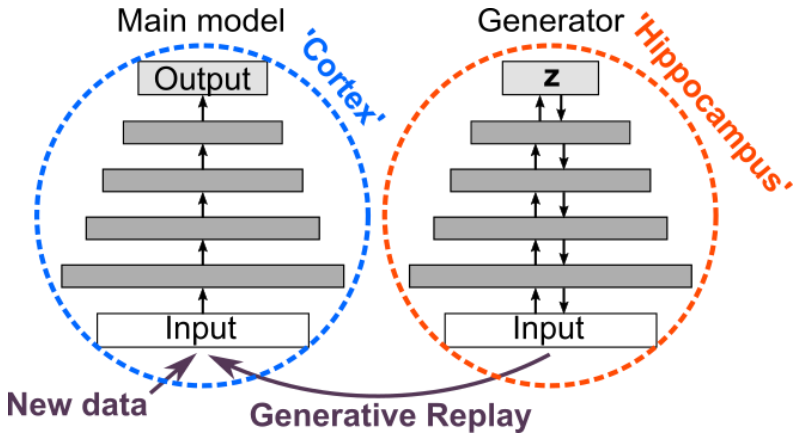


(all methods use pre-trained convolutional layers)

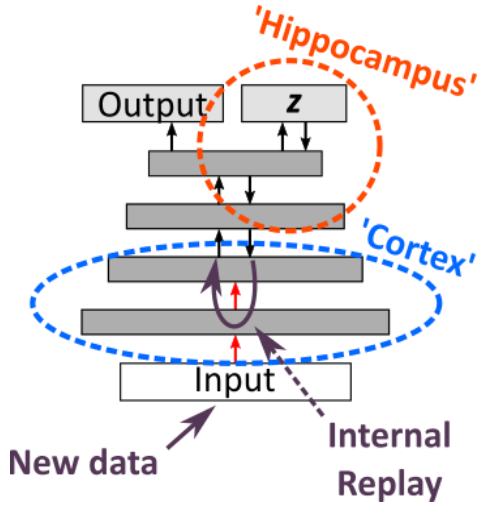
van de Ven et al. (2020) *Nature Communications*

Generative replay → Brain-inspired replay

- Generative Replay (Shin et al., 2017; *NeurIPS*)
  - Works great on MNIST, but fails with more complex inputs (e.g. natural images)
  - Efficiency issue left unaddressed



- Brain-inspired replay [**van de Ven** et al. (2020) *Nature Communications*]
  - Replay abstract, high-level representations, rather than detailed, input level samples
  - Replay can be sparse and low quality: "not forgetting" is easier than "learning"



Efficiency: How much replay is needed?

$$\ell_{\text{total}} = \ell_{\text{new}} + \ell_{\text{replay}}$$

$$N_{\text{new}} = 128$$

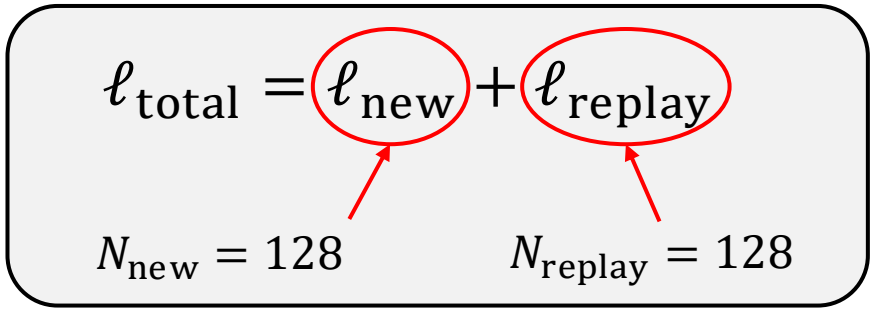
Efficiency: How much replay is needed?

$$\ell_{\text{total}} = \ell_{\text{new}} + \ell_{\text{replay}}$$

$N_{\text{new}} = 128$

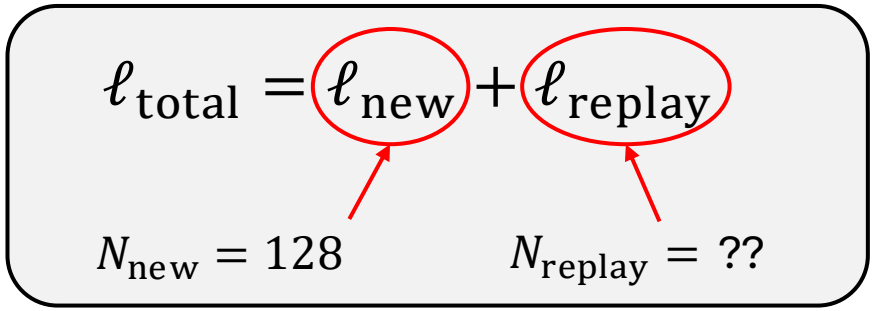
$N_{\text{replay}} = 128 * n_{\text{tasks}}$

Efficiency: How much replay is needed?

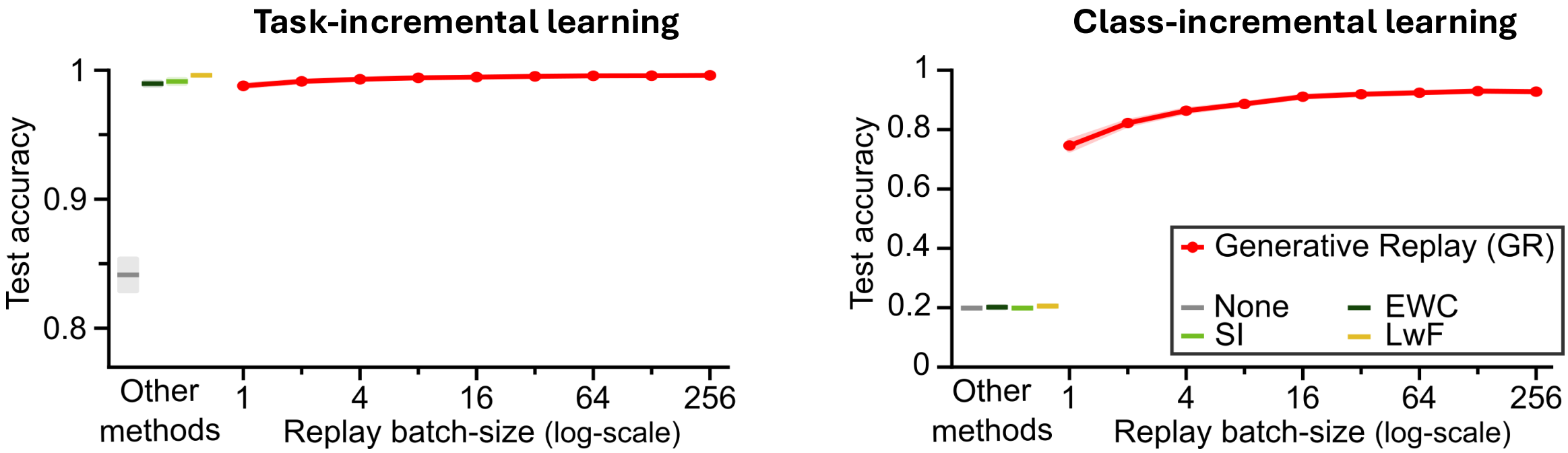


Efficiency: How much replay is needed?

- How far could the number of replayed sampled per batch be reduced?



Efficiency: How much replay is needed?

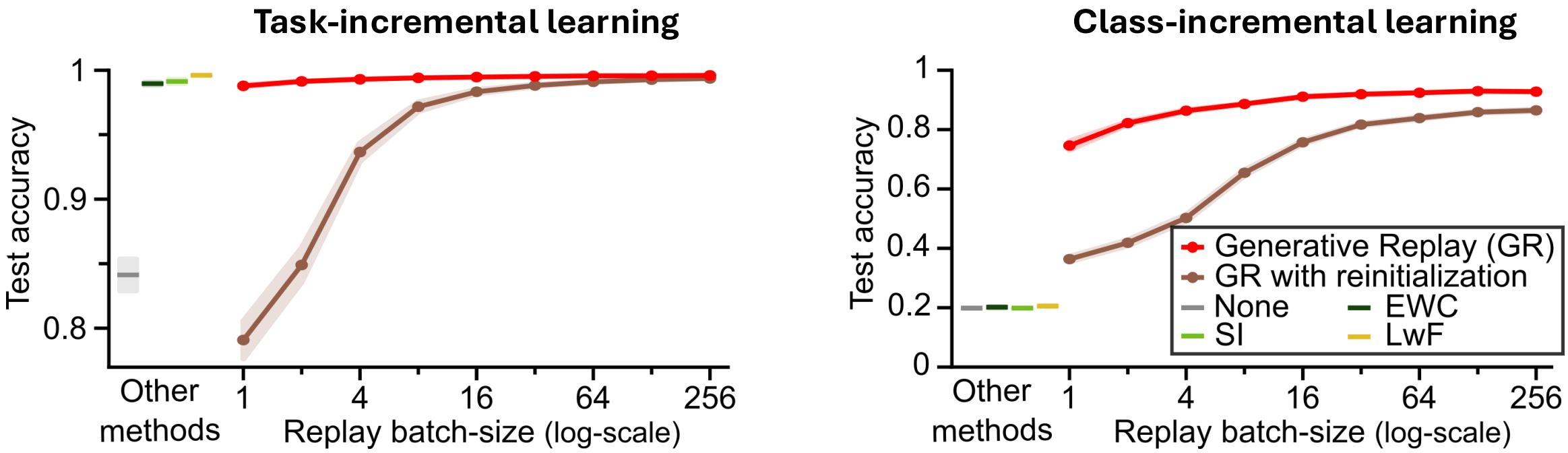


- Fully replaying past tasks is not needed, replaying just a few samples could suffice

(these experiments are on Split MNIST)

van de Ven et al. (2020) *Nature Communications*

Efficiency: How much replay is needed?



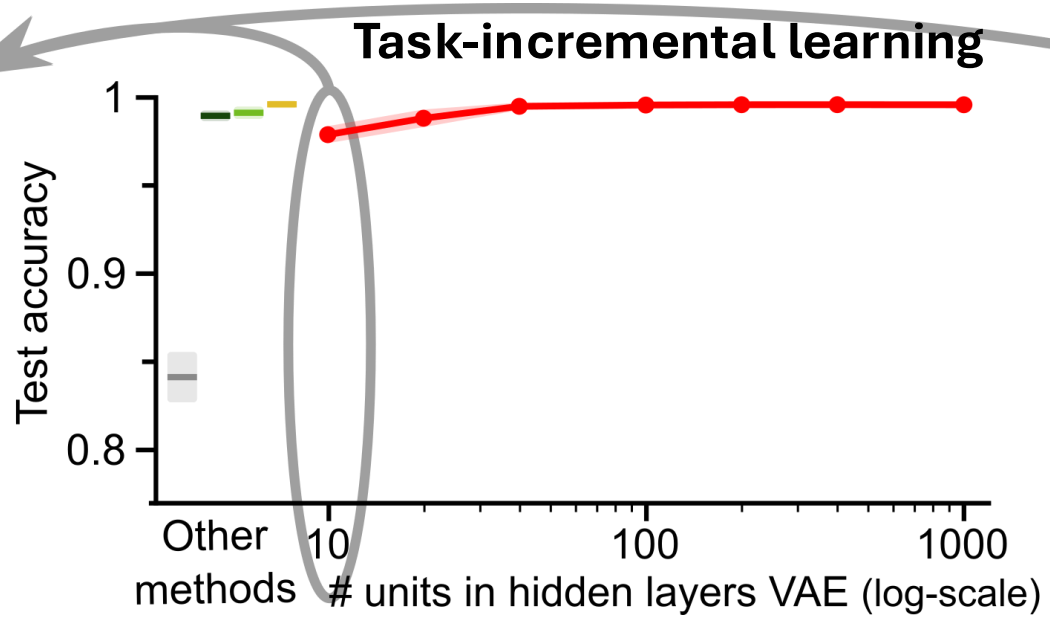
- Fully replaying past tasks is not needed, replaying just a few samples could suffice
- Because “learning something” is harder than “not forgetting it”

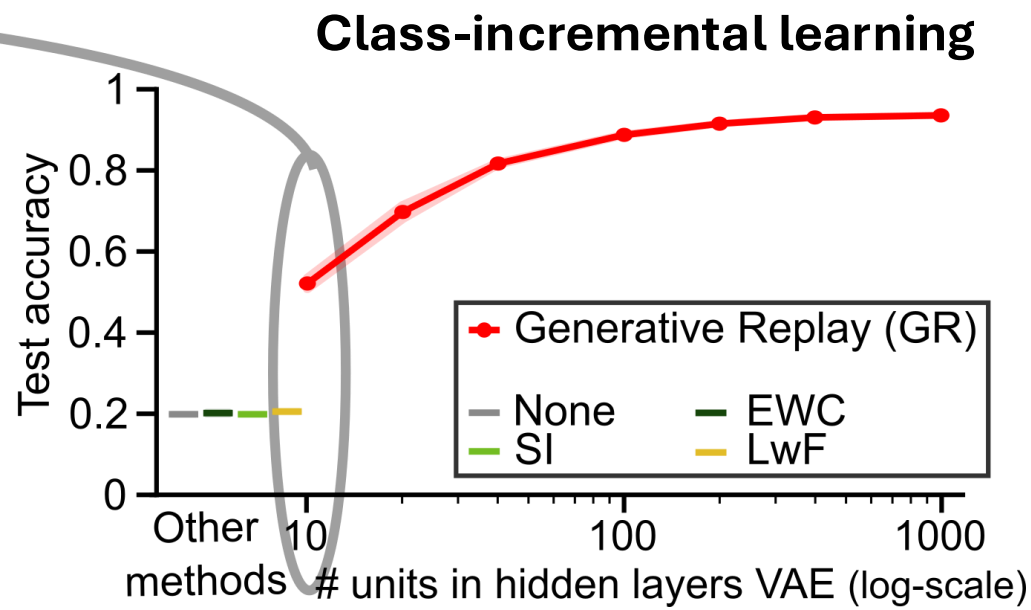
(these experiments are on Split MNIST)

van de Ven et al. (2020) *Nature Communications*

Robustness: How good does the replay need to be?







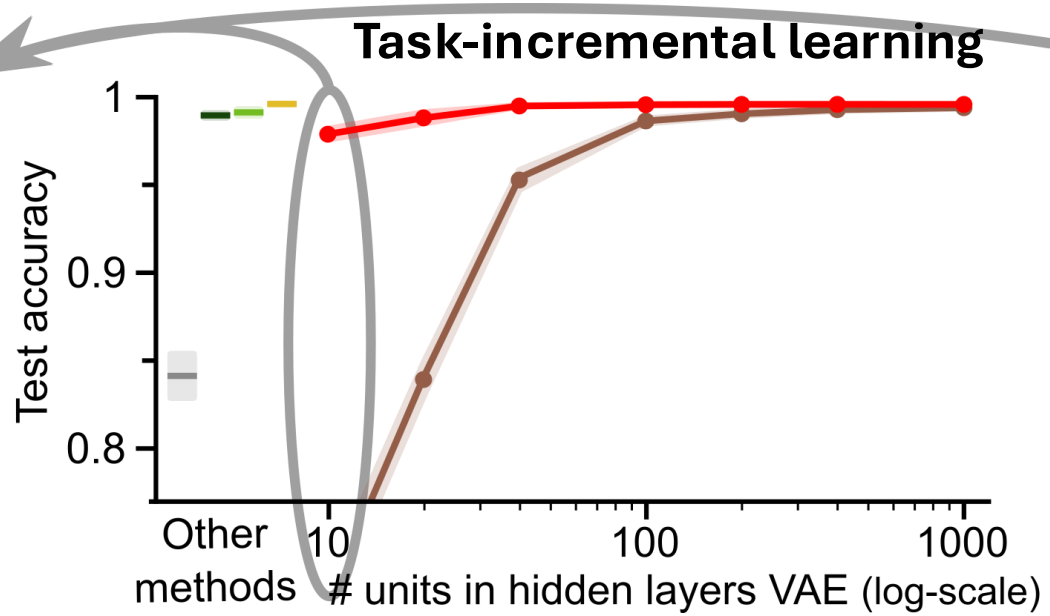
- A perfect memory is not needed, a low-quality generative model could suffice

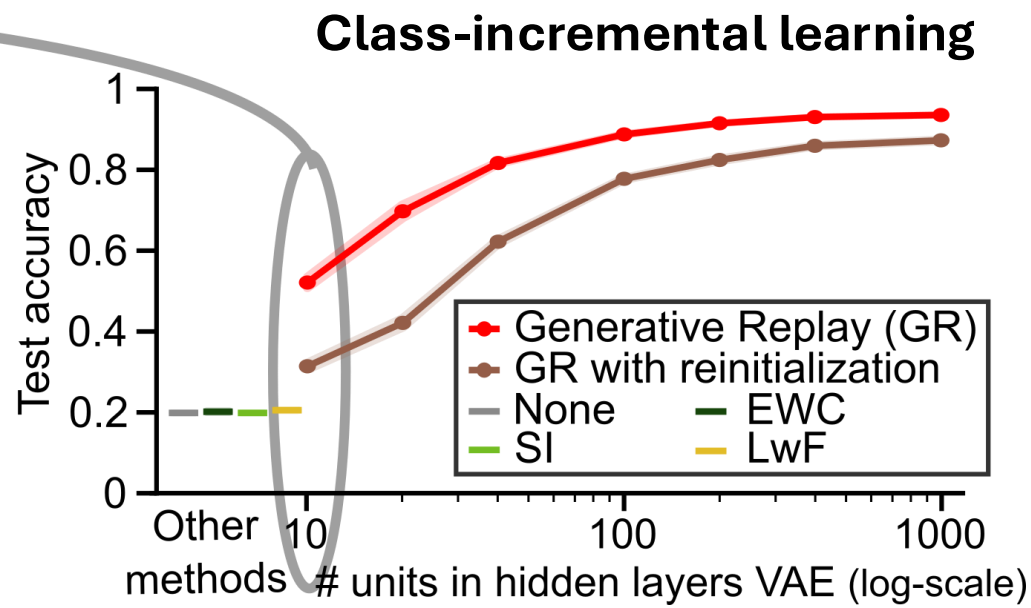
(these experiments are on Split MNIST)

van de Ven et al. (2020) *Nature Communications*

Robustness: How good does the replay need to be?







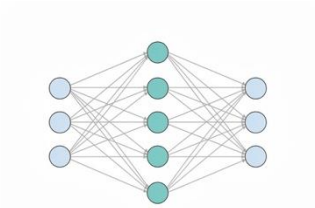
- A perfect memory is not needed, a low-quality generative model could suffice
- Because “learning something” is harder than “not forgetting it”

(these experiments are on Split MNIST)

van de Ven et al. (2020) *Nature Communications*

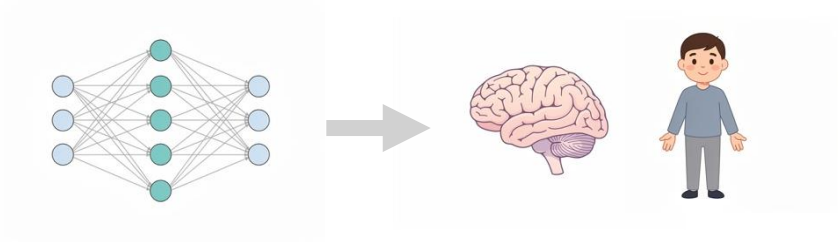
Overview

- Characterizing the continual learning problem
- Replay *[the current approach to continual learning]*



- **Stability gap** ***[fundamental limitation of the current approach]***

- Transient forgetting in humans?

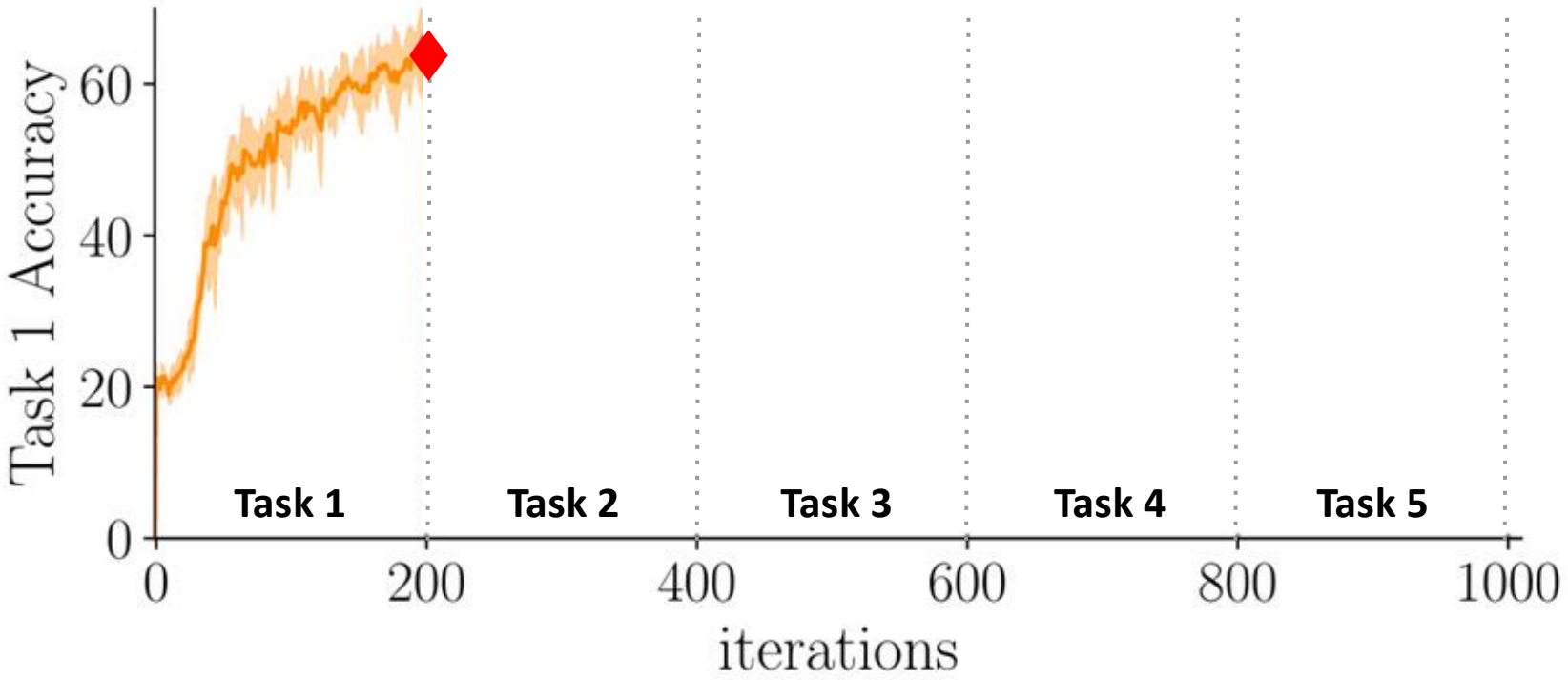


Does replay prevent forgetting?



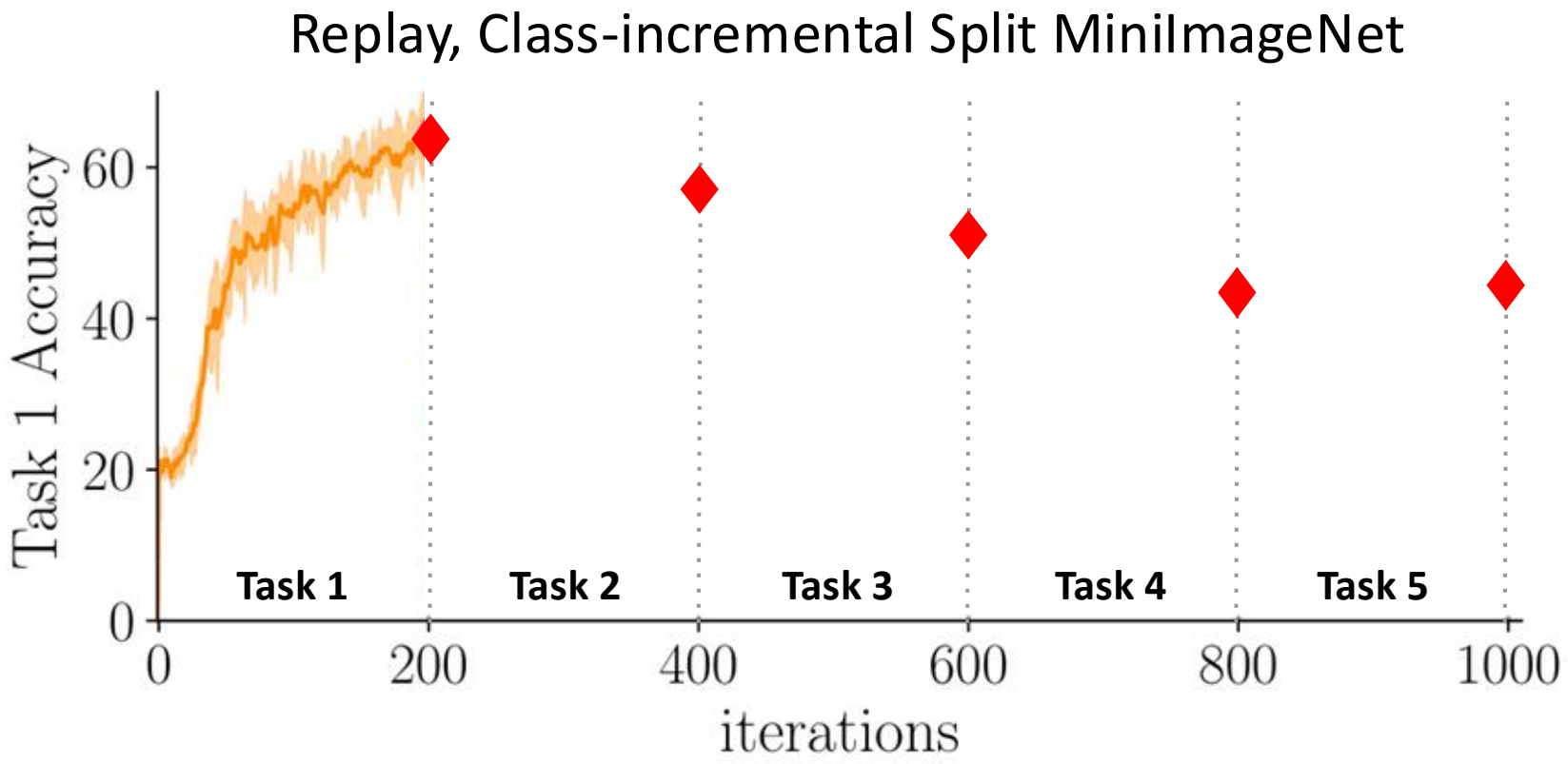
Matthias De Lange

Replay, Class-incremental Split MiniImageNet



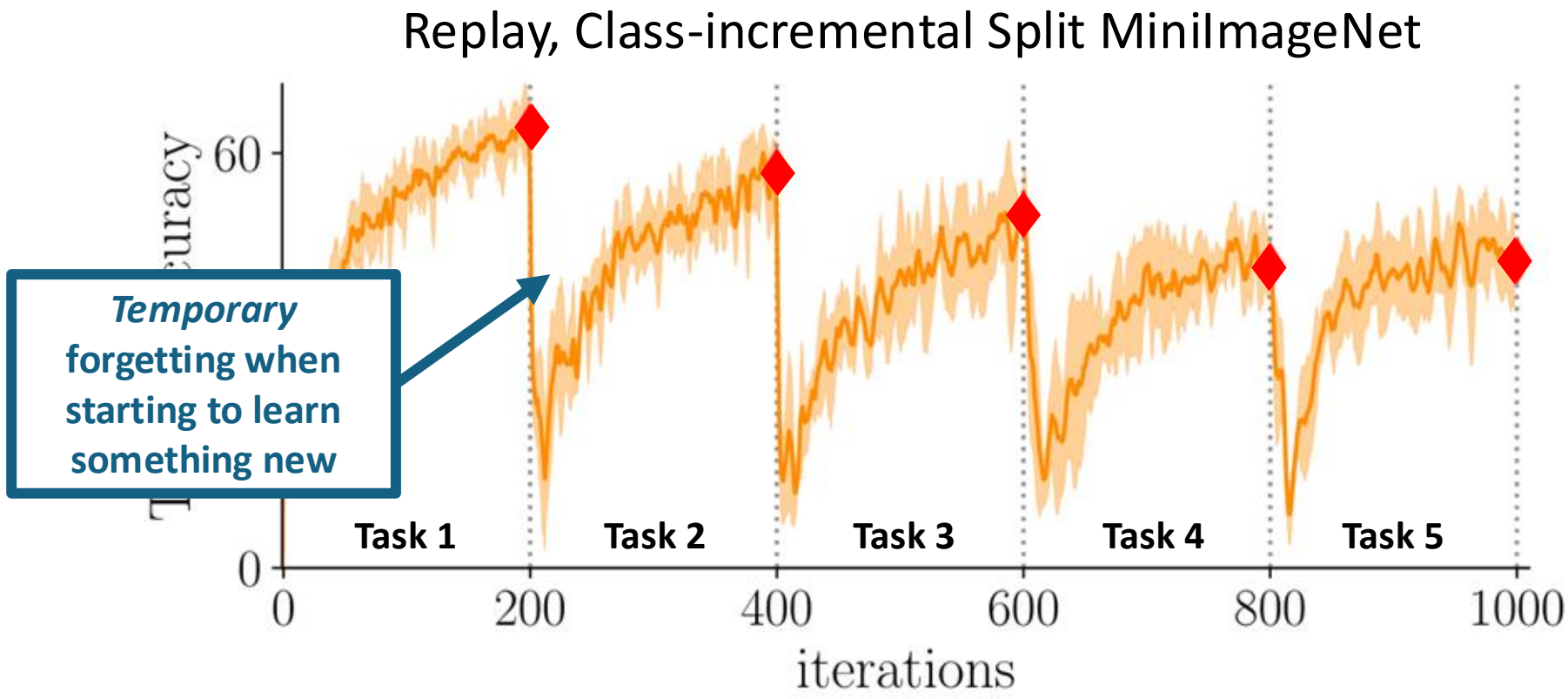
De Lange, **van de Ven** & Tuytelaars (2023) "Continual evaluation for lifelong learning: Identifying the stability gap" *ICLR, spotlight*

Does replay prevent forgetting?



De Lange, **van de Ven** & Tuytelaars (2023) "Continual evaluation for lifelong learning: Identifying the stability gap" *ICLR, spotlight*

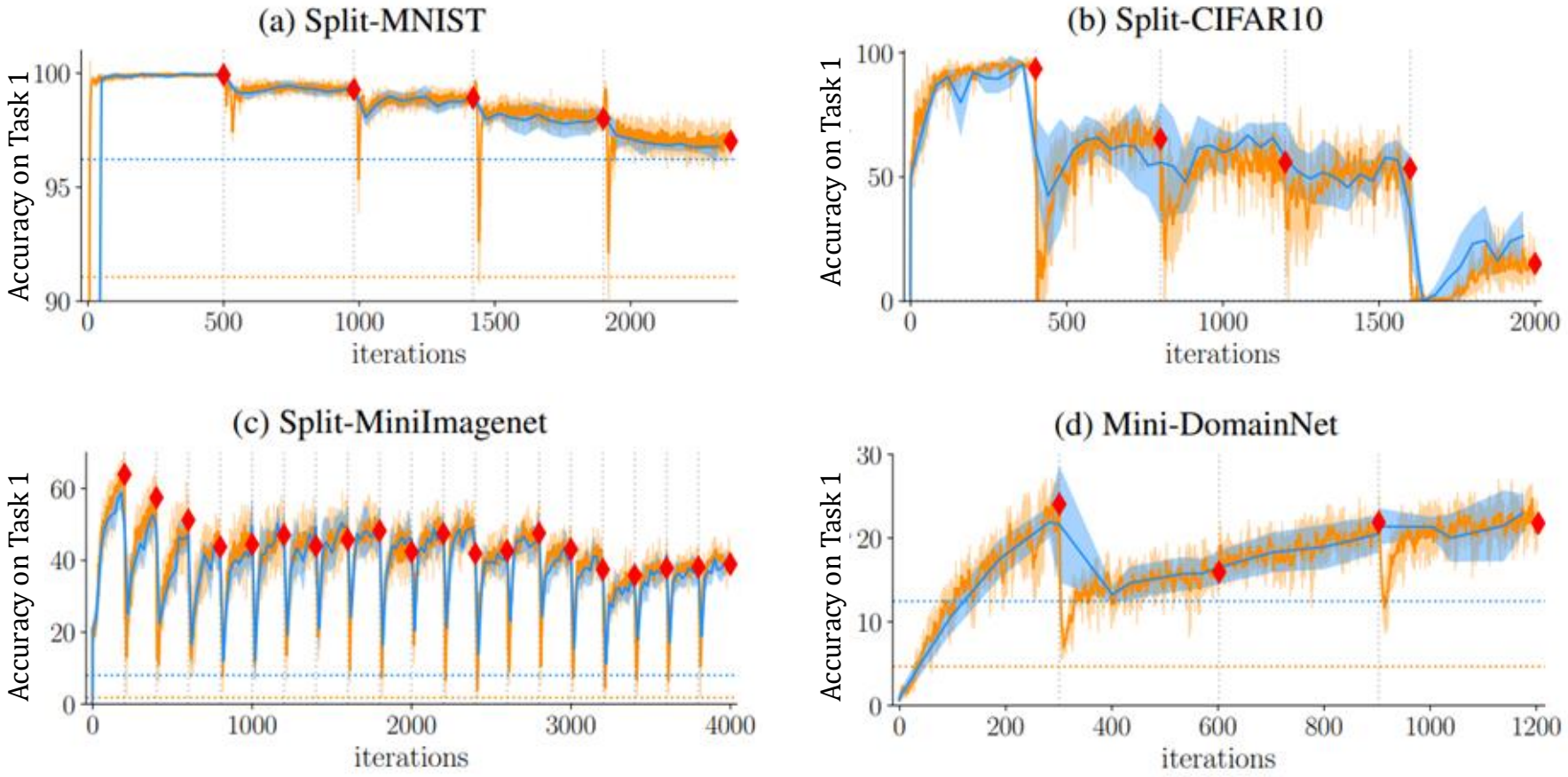
Does replay prevent forgetting?



De Lange, **van de Ven** & Tuytelaars (2023) "Continual evaluation for lifelong learning: Identifying the stability gap" *ICLR, spotlight*

The stability gap is consistently observed

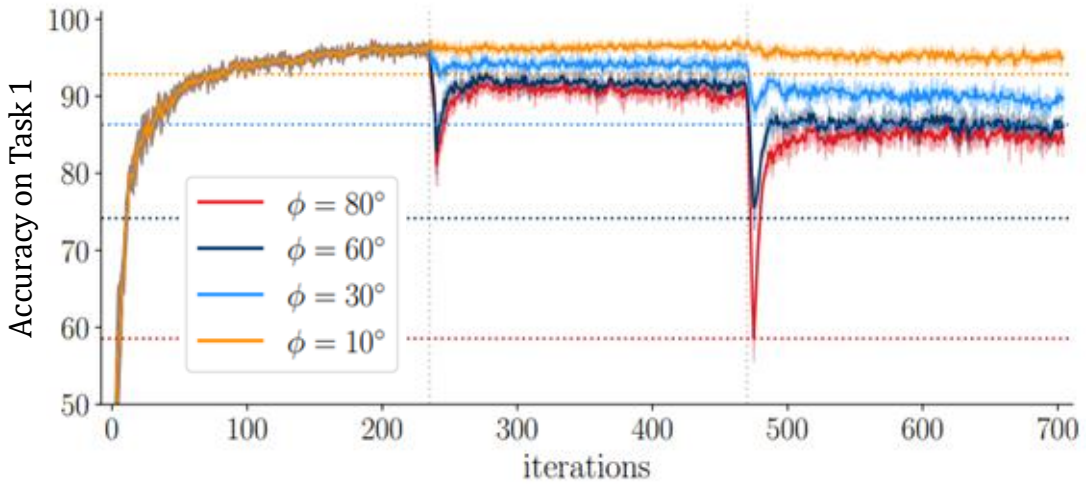
Replay, Class-incremental on …



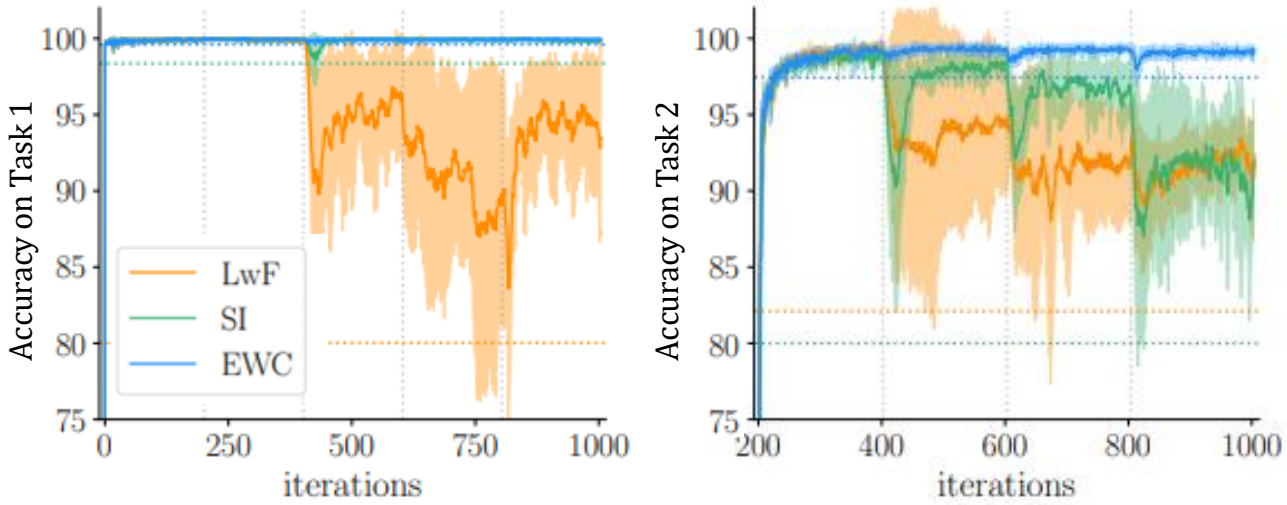
De Lange, **van de Ven** & Tuytelaars (2023) "Continual evaluation for lifelong learning: Identifying the stability gap" *ICLR, spotlight*

… also in other settings or with other methods

Replay, Domain-incremental Rotated MNIST



Regularization, Task-incremental Split MNIST



De Lange, **van de Ven** & Tuytelaars (2023) "Continual evaluation for lifelong learning: Identifying the stability gap" *ICLR, spotlight*

Why should we care?

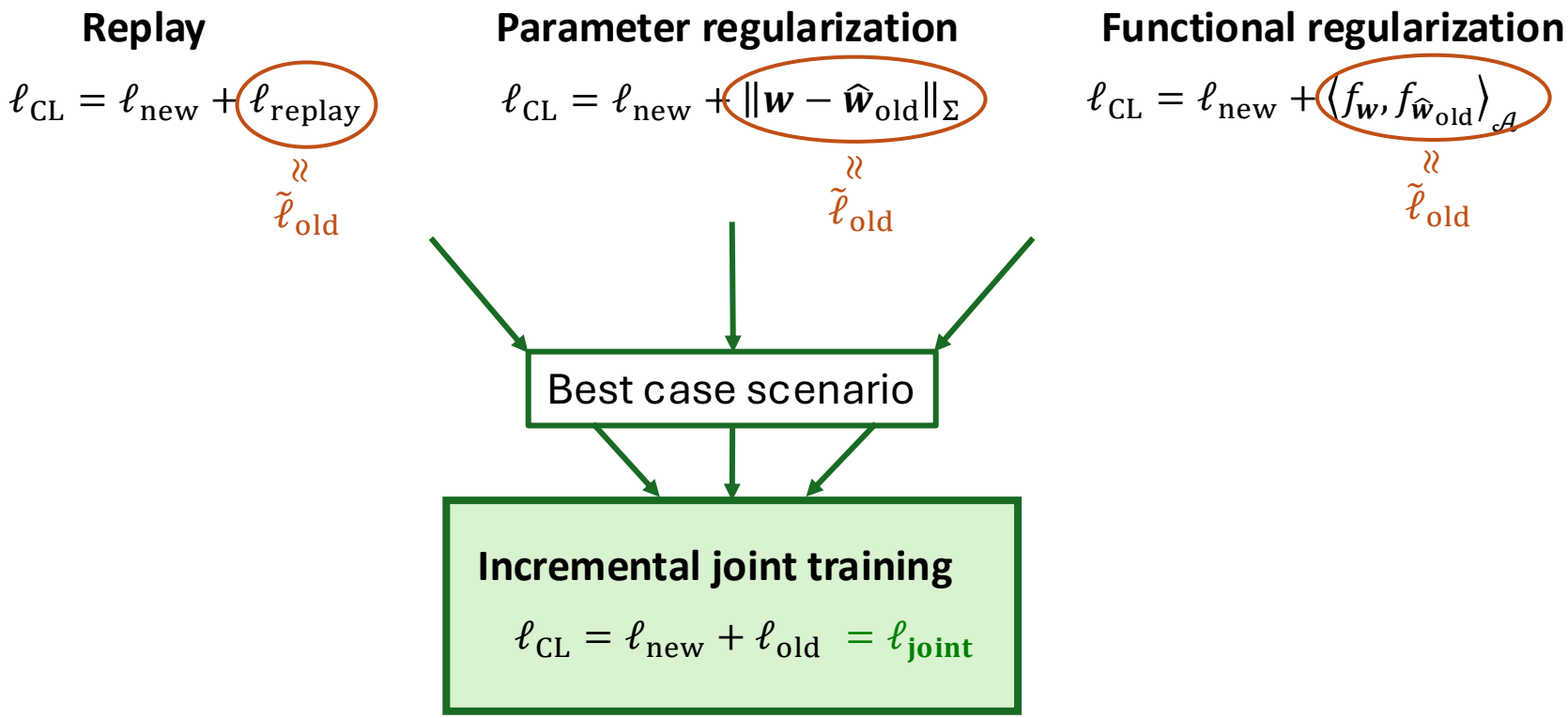
- Scientifically interesting
  - Insight into how replay works
  - Intriguing phenomenon

- Problematic for safety-critical applications
  - Worst-case performance might be important
  - Could be exploited by adversarial agent with control over training stream

- Seems highly inefficient
  - Preventing forgetting seems more efficient than having to re-learn

Why does the stability gap happen?

- Perhaps not enough data was replayed

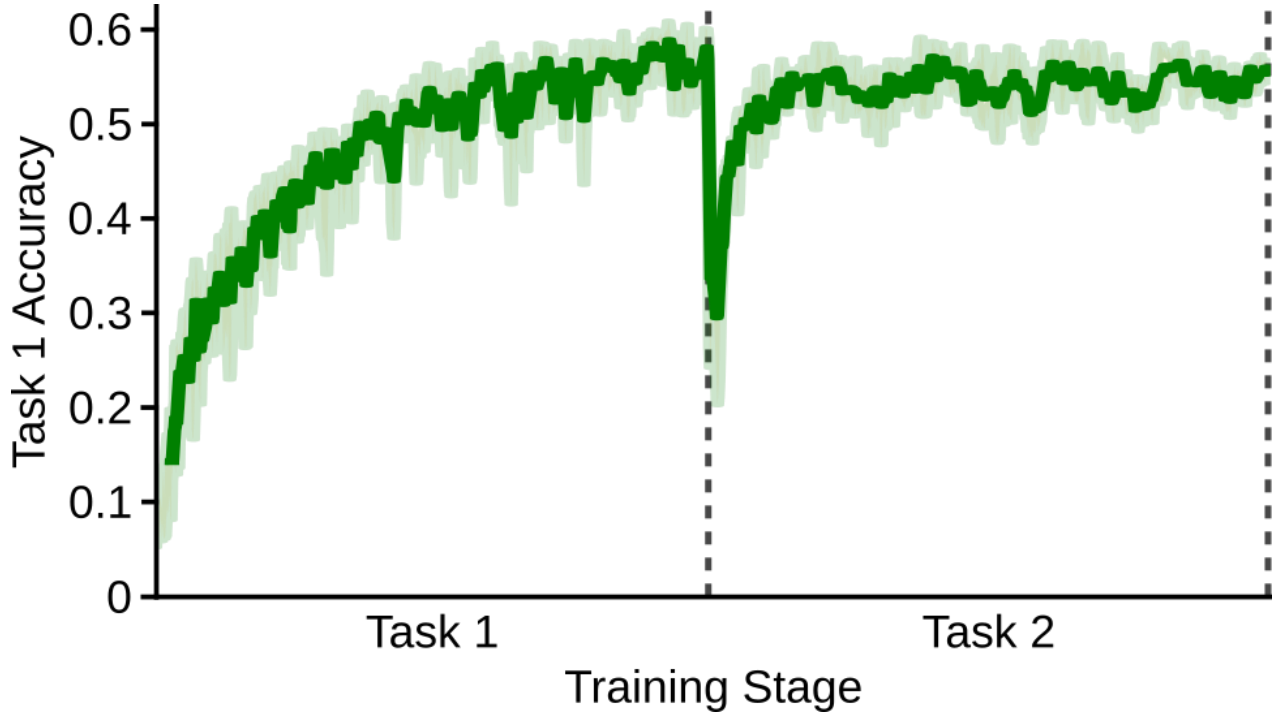


The stability gap occurs even with ‘full’ replay



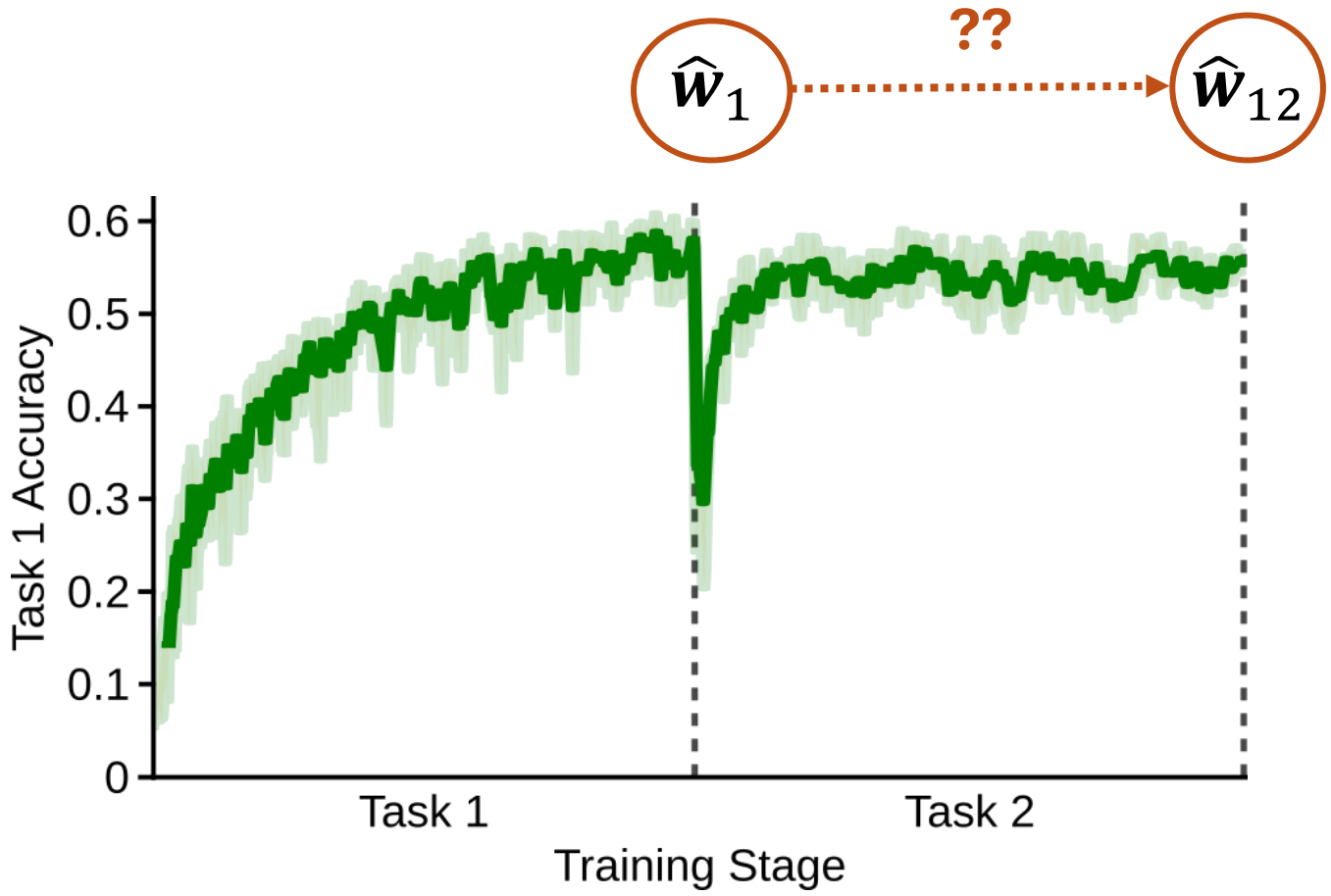
Timm Hess

Incremental joint training – Domain CIFAR-100

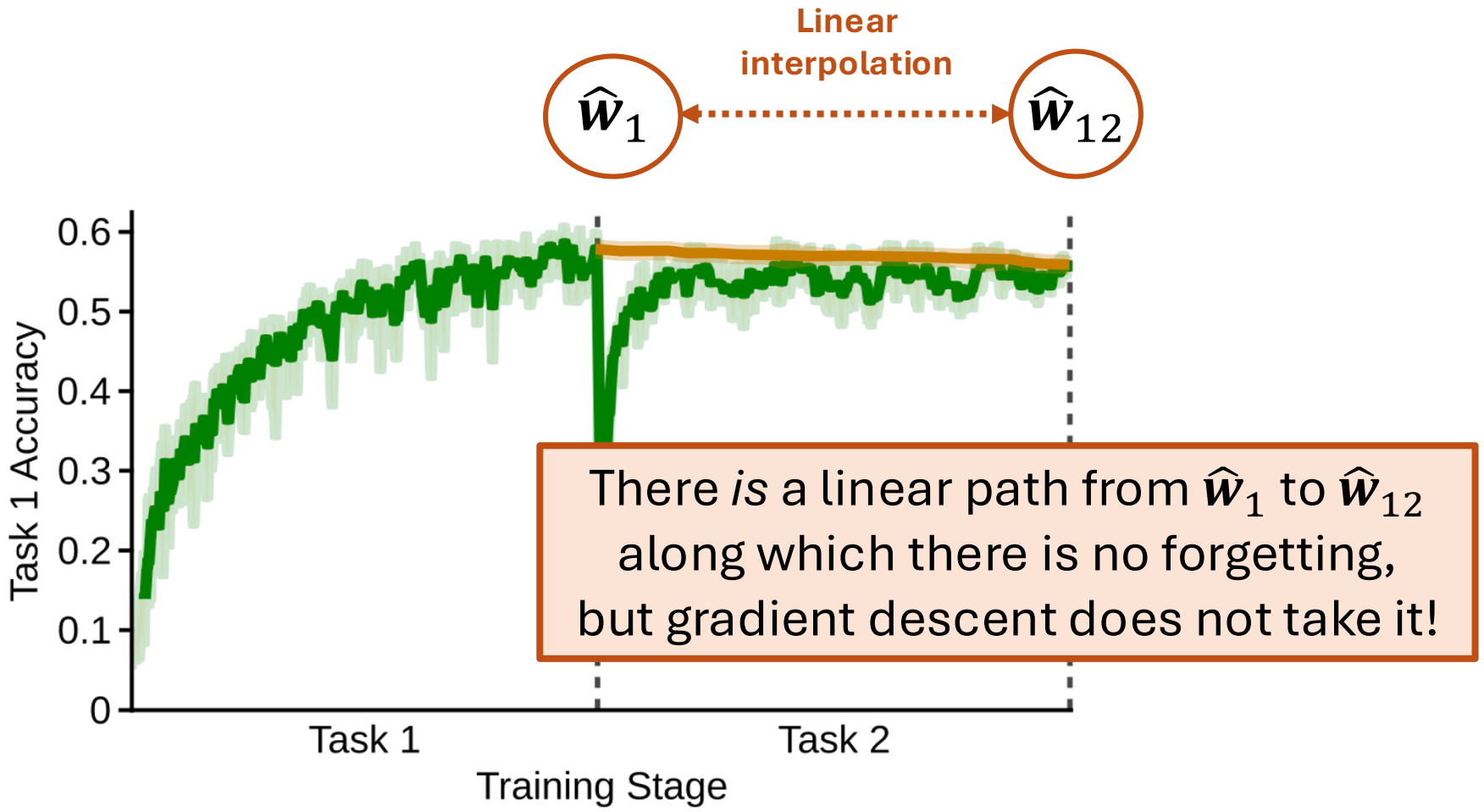


Hess, Tuytelaars & **van de Ven** (2023) “Two Complementary Perspectives to Continual Learning” *Pre-registered Report*

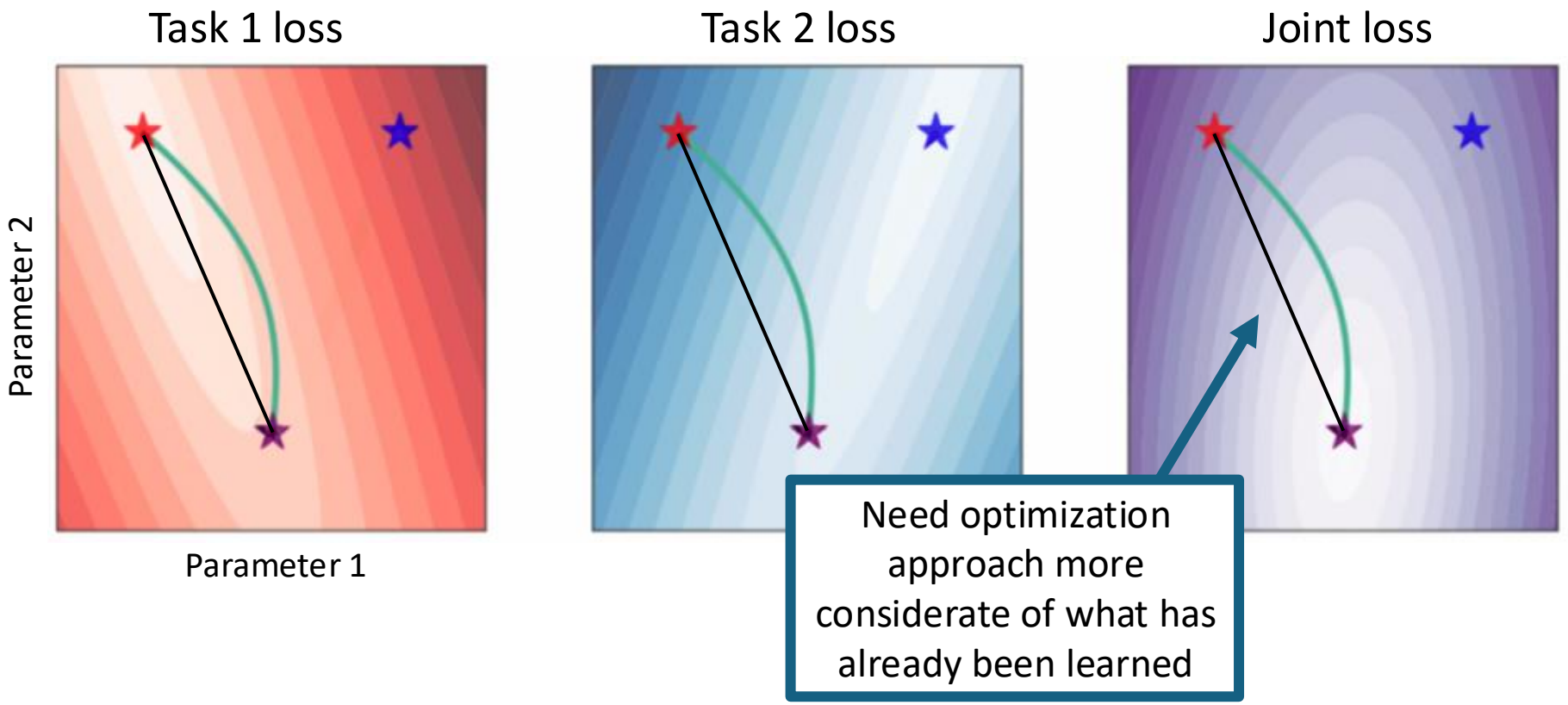
Can the stability gap be avoided at all?



Can the stability gap be avoided at all? → Yes!



Stability gap in a toy example

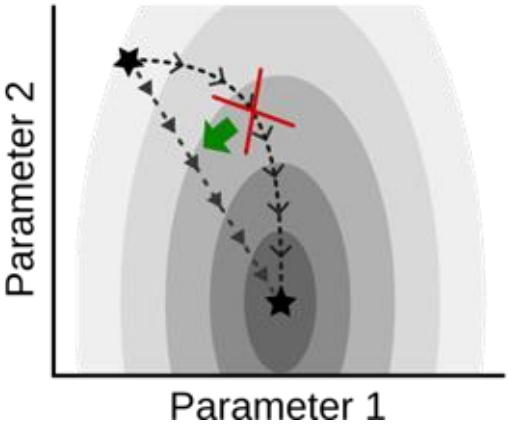


Kao*, Jensen*, **van de Ven** et al. (2021) "Natural continual learning: success is a journey, not (just) a destination" *NeurIPS*

Implications of the stability gap & future directions

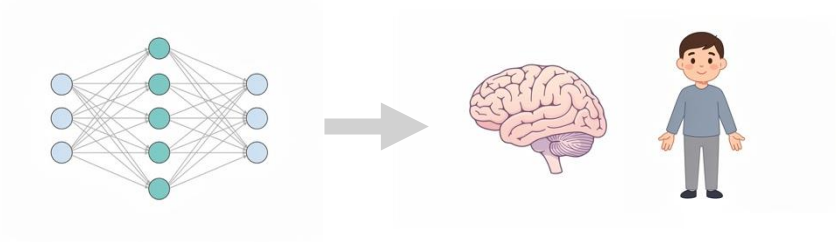
- Highlights ***optimization-based continual learning*** as a promising research direction

  → See Hess et al. (2024, PMLR) for a written-out argument



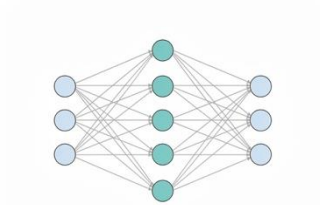
- Do humans suffer from transient forgetting when starting to learn something new?

  → Last part of this presentation!



Overview

- Characterizing the continual learning problem
- Replay *[the current approach to continual learning]*
- Stability gap *[fundamental limitation of the current approach]*



- **Transient forgetting in humans?**



Transient forgetting in humans



Milan Van Maldegem

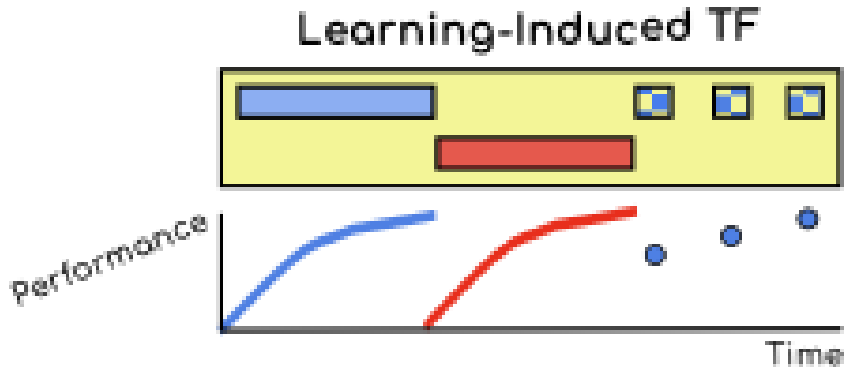
- Do humans suffer from a stability gap-like phenomenon?
  (i.e., transient forgetting when starting to learn something new)

- Survey of cognitive science literature

→ Six types of transient forgetting

Learning-induced transient forgetting

<u>Retroactive interference:</u>
learning new information can impair memory for previously learned information (~catastrophic forgetting in ANNs)



... this interference can recover over time!

Wheeler (1995) - *Journal of Experimental Psychology*

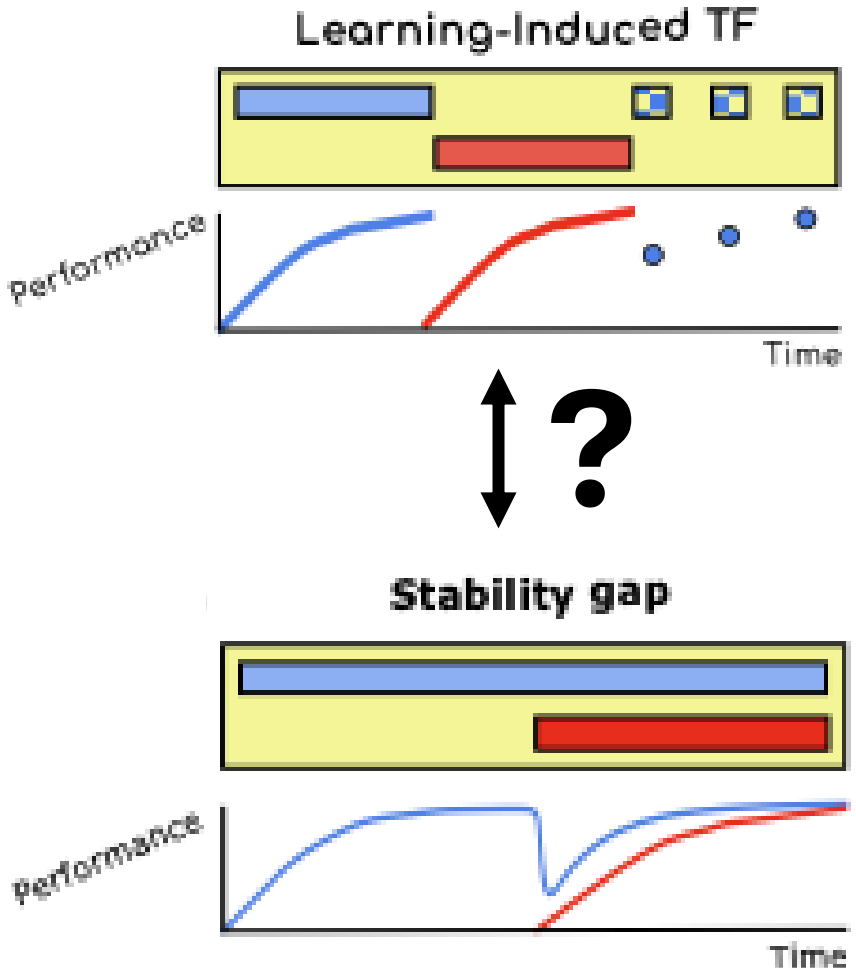
Learning-induced transient forgetting

<u>Retroactive interference:</u>
learning new information can impair memory for previously learned information (~catastrophic forgetting in ANNs)

… this interference can recover over time!

Wheeler (1995) - *Journal of Experimental Psychology*

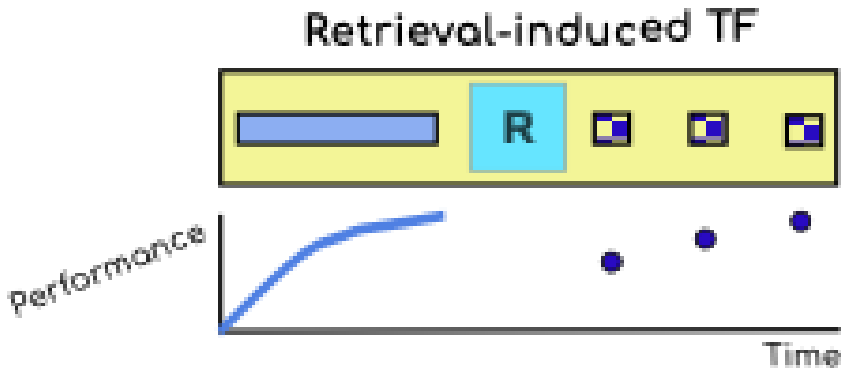
Similar to stability gap?



Retrieval-induced transient forgetting

<u>Retrieval-induced forgetting:</u>
selectively retrieving a subset of studied information can impair subsequent memory to related but unretrieved information

... this interference can recover over time!

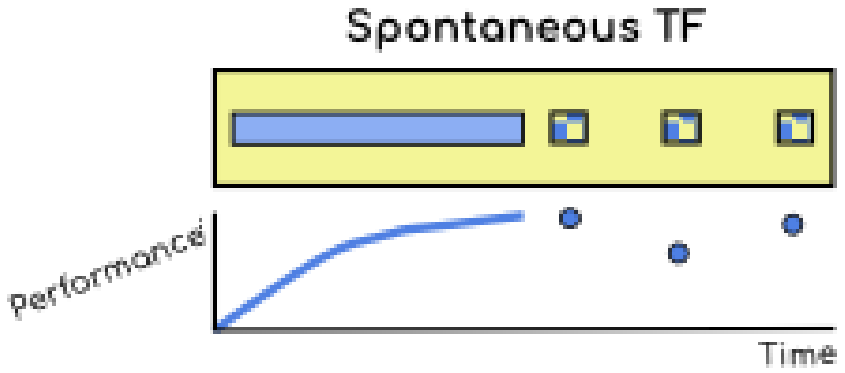


MacLeod & Macrae (2001) - *Psychological Science*

Spontaneous transient forgetting

<u>Transient memory lapses:</u>
drops in memory performance at intermediate intervals after learning

Wright & Ning (2024) - *npj Science of Learning*



<u>Tip-of-the-tongue states:</u>
transient failures in word retrieval despite a strong sense of knowing the word

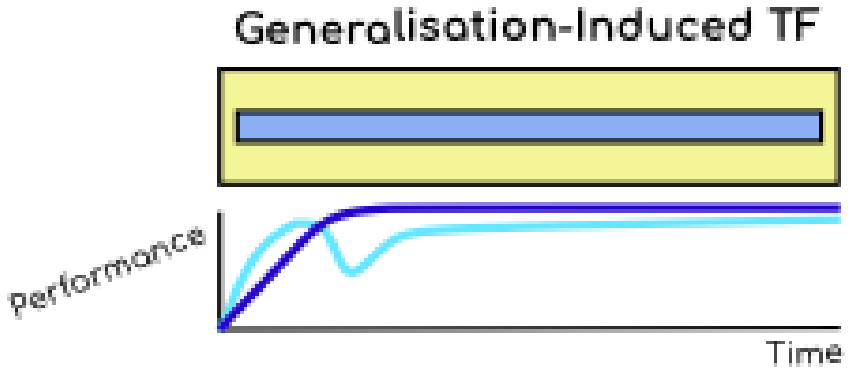
Brown & McNeill (1966) - *Journal of Verbal Learning & Verbal Behavior*

Also: reminiscence (Butler, 1963 *Psychiatry*) and involuntary autobiographical memories (Berntsen, 1996 *Applied Cognitive Psychology*)

Generalisation-induced transient forgetting

<u>U-shaped learning:</u>
Type of learning going through three stages:

1. correct use of exceptions (light blue curve)
2. overgeneralisation of a general rule to these exceptions
3. correct use of both the general rule and exceptions



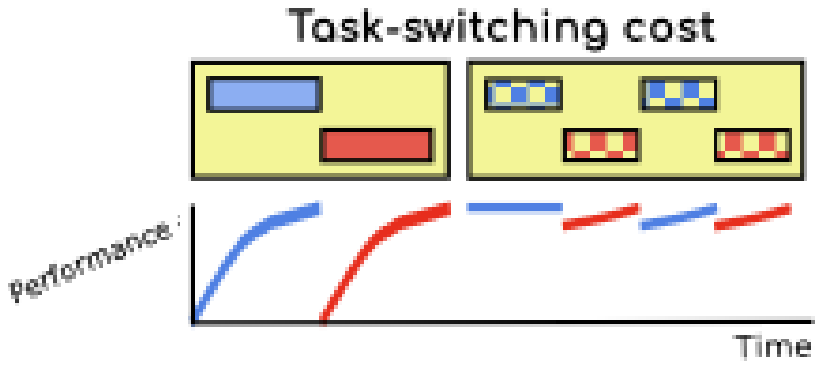
Carlucci & Case (2013) - *Topics in Cognitive Science*

(For a clear example, see acquisition of English past tense: Taatgen & Anderson, 2002 – *Cognition*)

Multitask and task-switching costs

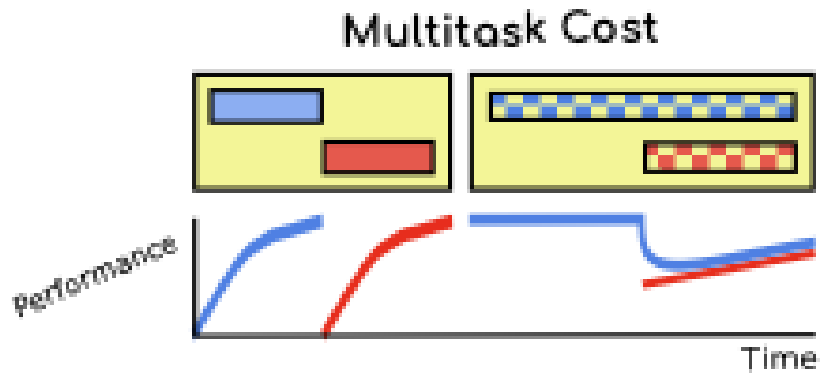
Local task-switching cost:
worse performance on task-switch trials compared to task-repeat trials, followed by performance recovery on subsequent trials

Rogers & Monsell (1995) - *Journal of Experimental Psychology*

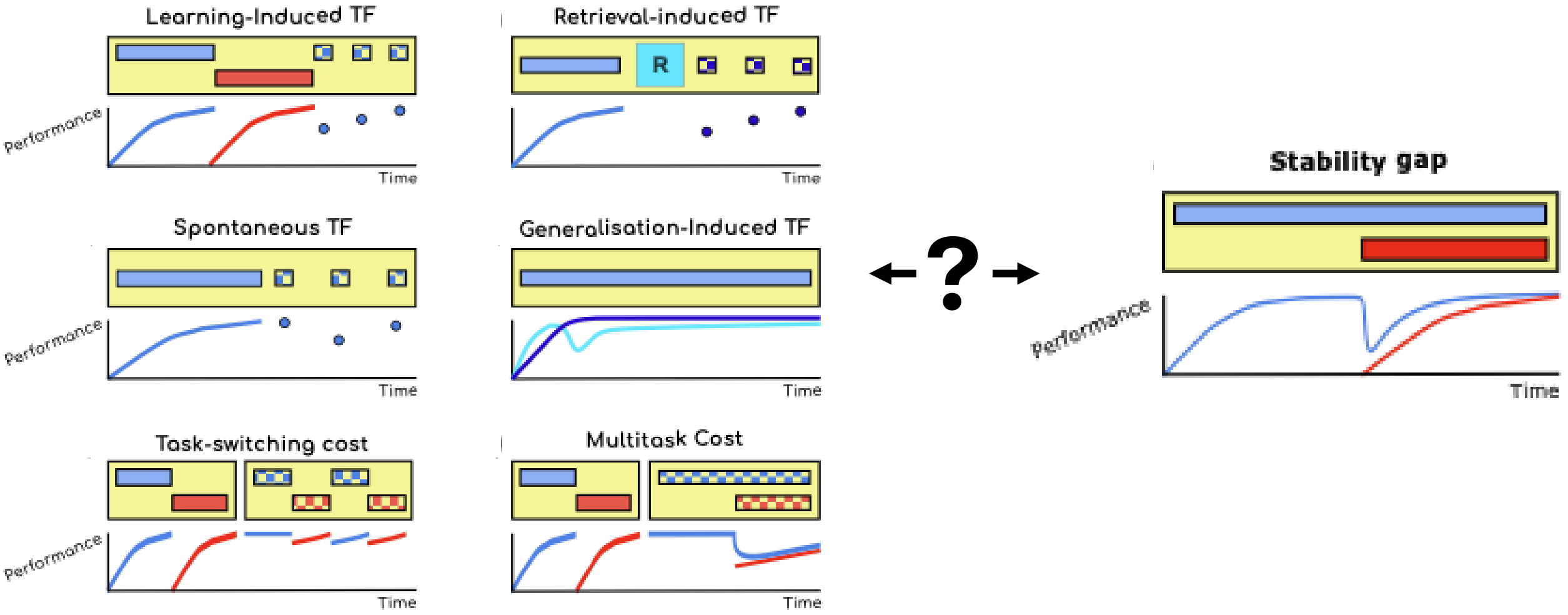


Dual-task interference:
transient reduction in performance on one or both tasks when they are performed simultaneously, compared to when they are performed in isolation

Van Selst et al. (1999) - *Journal of Experimental Psychology*



A stability gap-like phenomenon in humans?

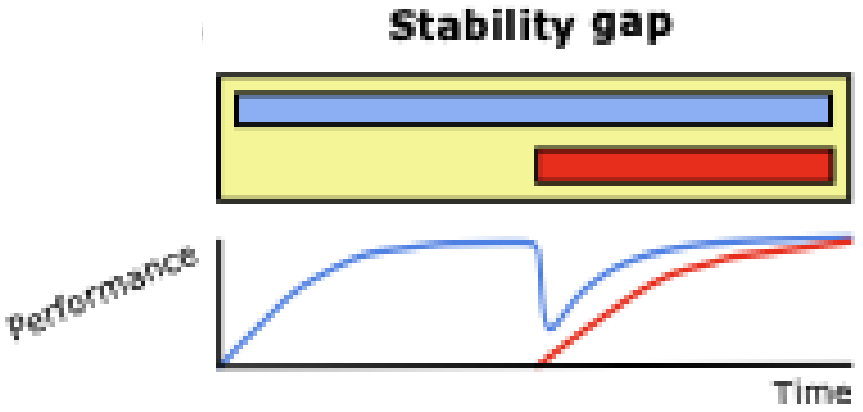


A stability gap-like phenomenon in humans?

Conditions an experiment should satisfy:

- performance on previously learned information should be **assessed continuously** while new learning is taking place
- the transient forgetting must be **linked to learning new information**

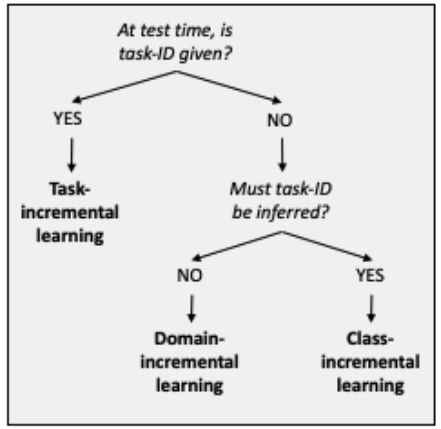
  (rather than passage of time, attentional demands or the *execution* of other tasks)



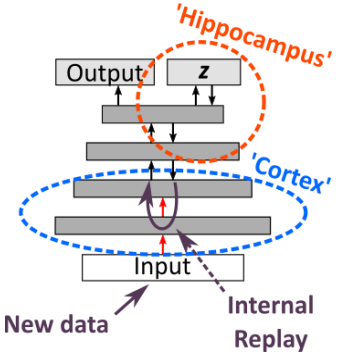
TO BE CONTINUED!

Summary

- Characterizing the continual learning problem:
  → distinguishing three types of incremental learning



- Analyzing and improving the scalability of replay



- Discovering the stability gap:
  → fundamental limitation of the current approach to continual learning
  → raises question whether humans suffer from temporary forgetting when learning something new – **BEING TESTED**

