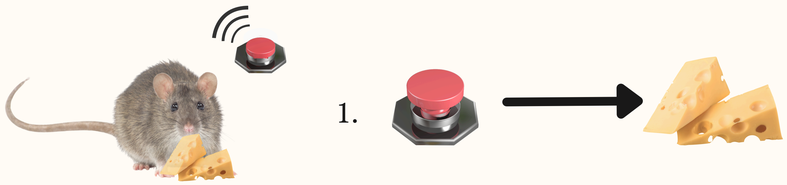
Stacked DAGs for Sequential and Hierarchical Learning

Dominic Le | dominic.le@mail.utoronto.ca | dominicle.net
University of Toronto

Main Question

In extinction, how do the rat's previously learned associations get rewritten?



Background

'Unlearning' [1]

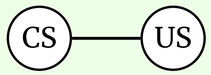
Extinction is unlearning the association from the unconditioned stimuli (US) to the conditioned stimuli (CS) (as in classical conditioning and exposure therapy studies).

Spontaneous recovery is when the extinct US-CS association re-emerges (i.e. relapse).

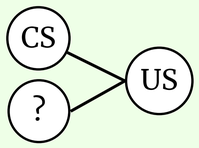
However, there are many ways to induce extinction (e.g., CS-alone, partial, context manipulations, deepened extinction).

Past Modeling Approaches

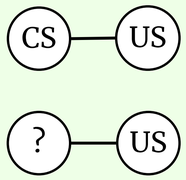
R-W (Rescorla & Wagner, 1972)



Comparator (Stout & Miller, 2007)

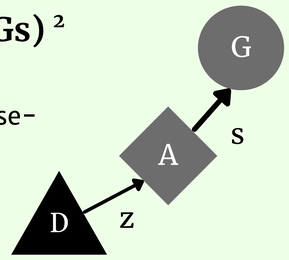


Latent-Cause (Gershman et al., 2015)



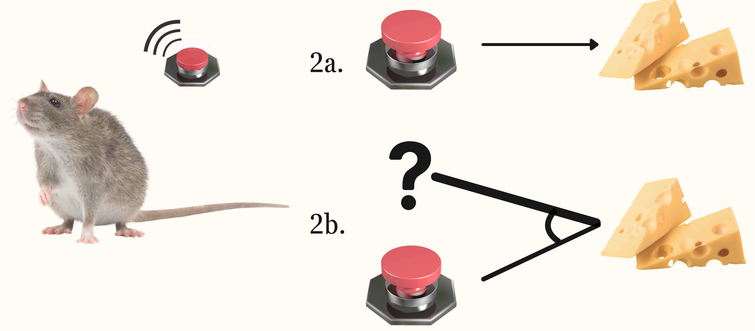
Directed Acyclic Graphs (DAGs) [2]

Effectively represents <u>conditional relationships</u> – essential for many cause-and-effect paradigms (e.g. models of causation, counterfactuals, neural activations)



Stacks

The stack is a layered sequence of DAGs that share the same structure (nodes and edges) where each layer archives changes in parameters (weights).



2-Way Updating

<u>Within</u>: update existing parameters in a stack (e.g., CS-alone, partial reinforcement, reversal).

<u>Between</u>: create a stack with new events/relationships (e.g., context manipulation, deepened extinction).

Discussion

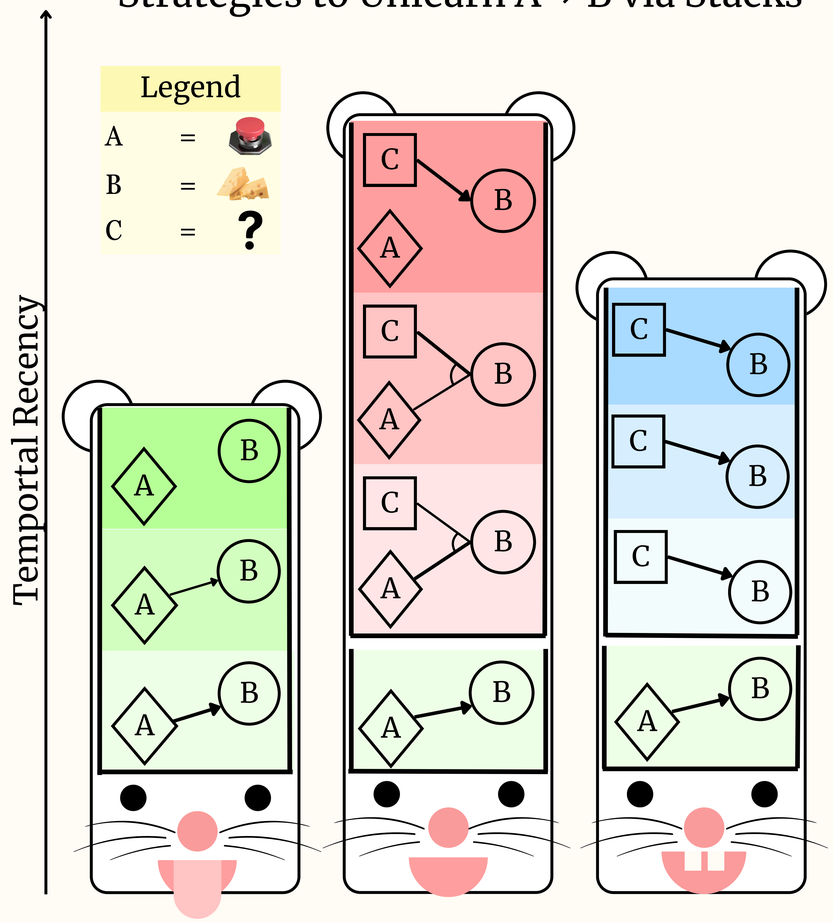
Implications

- Recovery timing depends on update strategy.
  - The competition stack predicts slower recovery than within stack.
- Savings comes from accessing intact layers.
  - The within stack results in quicker reacquisition.
- Context switching requires multiple models.
  - A new physical context re-activates latent-cause stack without new learning.

Future Work

1. Cases for hypothetical associations (e.g., counterfactuals).
2. Logic for nested modeling.
3. Test predictions (e.g., context-switching, access delays).

Strategies to Unlearn A → B via Stacks



Within-Stack: CS-alone | **Between-Stack: Competition** | **Between-Stack: Latent-Cause**

References

1. Gottlieb, D. A. (2012). Pavlovian Conditioning. In Encyclopedia of the Sciences of Learning (pp. 2563–2567). Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-1428-6_1041
2. Danks, D. (2014). Unifying the mind: Cognitive representations as graphical models. the MIT press.

Takeaway

Stacked DAGs may be a flexible framework to model dynamic and iterative structural learning that draws on mechanisms from compatible theories.

Thank you to the Social Sciences and Humanities Research Council (SSHRC), the University of Toronto Faculty of Arts & Science, and the University of Toronto Department of Philosophy for funding this research under the University of Toronto Excellence Award ($)
Special thanks to Sara Aronowitz for her mentorship and supervision (☺)