Oscillations Drive Learning in Retrieval-Induced Forgetting

Ehren Newman & Kenneth Norman

Department of Psychology and
Center for the Study of Brain, Mind, and Behavior
Princeton University

Princeton Computational Memory Lab

194.13

Abstract

Retrieval-induced forgetting (RIF) refers to the finding that retrieving a memory hurts subsequent recall of competing memories (see Levy & Anderson, 2002, for a review). In the work presented here, we use neural network simulations to explore how the brain gives rise to RIF. In our prior work on this topic (Newman & Norman, 2003), we showed that a learning rule (suggested by O'Reilly and McClelland) that compares a state of low inhibition to a state of normal inhibition is able to identify and punish competitors in the RIF paradigm. We now present a second-generation learning rule that, unlike its predecessor, is able to simultaneously store new information in the network and punish competitors. The new rule relies on a continuous, "theta-like" oscillation in the strength of inhibition. As before, moving between normal and lower-than-normal inhibition allows the network to identify and punish competitors. Moving between normal and higher-than-normal inhibition serves the complementary function of identifying and strengthening weak parts of the target representation. We show that this new, oscillation-based learning rule is capable of training a large number of heavily overlapping patterns into a network (so that stored patterns can be retrieved given partial cues), and it also can account for detailed patterns of RIF data. We discuss the relationship between our new rule and cortical theta. We also discuss how this view of RIF (which emphasizes basic cortical learning mechanisms) relates to Anderson's account of RIF, which focuses on the role of prefrontal cortex in modulating competition.

RIF Task (from Levy & Anderson, 2002)

Study	Practice	Test
Fruit - Pear Fruit - Apple Red - Apple Red - Rose	Fruit - Pe___ (partial practice) or Fruit - Pear (full practice)	Fruit - P____ Fruit - A____ Red - A____ Red - R____

Recall after practice, relative to baseline

Test Item	After Partial Practice (Fruit - Pe___)	After Full Practice (Fruit - Pear)
Fruit - P(ear)	BETTER	BETTER
Fruit - A(pple)	WORSE	SAME
Red - A(pple)	WORSE	SAME
Red - R(ose)	SAME	SAME

In other words, if given a partial practice -
- Recall of the practiced item improves (Fruit-Pear)
- Recall of competitors gets worse (Fruit-Apple), in a *cue-independent* fashion (Red-Apple)

and if given a full practice -
- Recall of the practiced item improves (Fruit-Pear)
- Other items are unaffected (Red-Rose)

Background

Conflict Resolution:
Anderson has emphasized role of prefrontal cortex (PFC) in resolving competition
- PFC works to inhibit competitors
- Indirect vs. direct suppression not specified
- Mechanism of lasting effects not specified

Our Approach:
Identify basic learning mechanisms that can account for lasting RIF
- Newman & Norman (2003) used a learning algorithm suggested by O'Reilly & McClelland
- Basic idea: Identify competitors by reducing inhibition
  1. Present the input pattern twice

     1st time – Low inhibition
     (allows both the target and competitors to become active)

     2nd time – Normal inhibition
     (only allows the best-fitting pattern to become active)

  2. Record final pattern of activity each time
     Units that "pop up" when inhibition is reduced are competitors.
     Punish these units by making them less excitable.

PROBLEMS:
1. Requires multiple presentations of stimuli
2. Requires mechanism to take and compare snapshots of activity
3. NOT ABLE TO STORE NEW INFORMATION

Oscillation based learning rule (Norman, Newman, & Polyn, in preparation)

Inspired by cortical oscillations such as theta
- There when you need it - Theta is gated by stimuli presentation (Raghavachari et al, 2001)
- Orients to simuli - Theta phase is reset by stimulus onset (Rizzuto et al, 2003)
- Plasticity varies with it - Sign of plasticity depends on phase of theta (Heurta & Lisman, 1996)

NO NEED FOR MULTIPLE PRESENTATIONS
NO NEED FOR SNAPSHOTS
ABLE TO STORE NEW INFORMATION

Extract structure of stored information with oscillations

Low inhibition to punish competitors

Oscillate between NORMAL - LOW - NORMAL inhibition (N-L-N)

Low inhibition = Less constraint on network activity
The network has more space to represent competitors (as well as the target)

LOWERING inhibition lets the network identify competitors

High inhibition to strengthen target

Oscillate between NORMAL - HIGH - NORMAL inhibition (N-H-N)

High inhibition = More constraint on network activity
Stress-test of target: Poorly supported units turn off, well-supported units remain active

RAISING inhibition lets the network identify weak parts of the target

Learn based on changes in activity

Changing activity during N-L-N = competitors popping up

How to map changing activity to learning signal:
- **Inhibition decreases:** Competitors become active
  Therefore - increases in activity should trigger weakening
- **Inhibition returns to normal:** Competitors back off
  Therefore - decreases in activity should trigger weakening

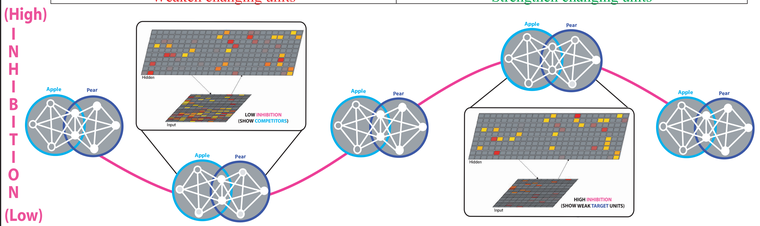
Changing activity during N-H-N = target dropping out

How to map changing activity to learning signal:
- **Inhibition increases:** Weak target units turn off
  Therefore - decreases in activity should trigger strengthening
- **Inhibition returns to normal:** Target turns back on
  Therefore - increases in activity should trigger strengthening

INHIBITION APPLIED TO INPUT LAYER (CONSTRAINT)

Normal	Low	Normal	High	Normal
Target on	Target on	Target struggles to stay on	Target comes back on	
Competitor allowed on	Competitor forced off	Competitor off	Competitor off	
Weaken changing units		Strengthen changing units		

(High) INHIBITION (Low)



Implementation

Only oscillate inhibition in input layer
Input:
1. Calculate baseline inhibition to allow K active units
2. Add an oscillating component to this value

Hidden:
1. Calculate baseline inhibition to allow K active units
2. No oscillating component

Allow one full oscillation each trial

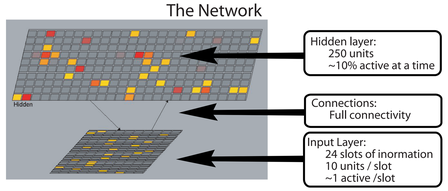
Calculate weight changes at every time step (but do not apply them)

Apply summed weight changes at the end of each trial

Update weights based on phase of theta and change in activation
Phase:
- **Normal to Low**
  Change = -lrate * (Sending_Act * Receiving_deltaAct)
  (Increases in activation will cause negative change)
- **Low to Normal**
  Change = lrate * (Sending_Act * Receiving_deltaAct)
  (Decreases in activation will cause negative change)
- **Normal to High**
  Change = -lrate * (Sending_Act * Receiving_deltaAct)
  (Decreases in activation will cause positive change)
- **High to Normal**
  Change = lrate * (Sending_Act * Receiving_deltaAct)
  (Increases in activation will cause positive change)

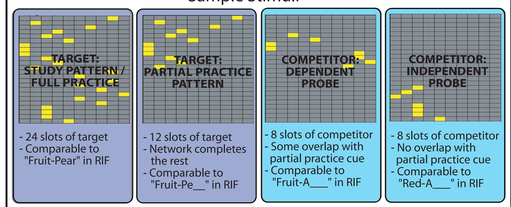
Materials and Procedure

The Network



1. Generate four patterns
A target pattern (presented at study and practice)
A competitor pattern (50% similar to target pattern, presented at study but not practice)
and two controls (50% similar to each other, presented at study but not practice)

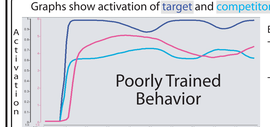
Sample Stimuli



2. Train the network on these patterns
Present the network with the complete patterns
Update weights after each pattern
3. Pretest the network's ability to pattern complete on all patterns
Present 33% of the pattern as cue
4. Allow network to practice target pattern
In case of partial practice:
Network completes 50% of the pattern
In case of full practice
Present the full pattern just like in training
5. Test the network's ability pattern complete on all patterns again
Compare to pretest performance to calculate practice effect

Network behavior during training

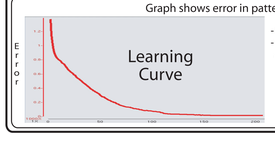
Graphs show activation of target and competitor, and overall level of inhibition on one trial



Early in training:
- High inhibition
  Target units decrease in activity
  Competing units show a slight increase in activity
- Low inhibition
  Competing units increase in activity
  Target units show a slight decrease in activity

Late in training:
- High inhibition
  Target units strong enough to stay on
  Activity of competing units does not change
- Low inhibition
  Competing units too weak to come on
  Activity of target units does not change

Graph shows error in pattern completion

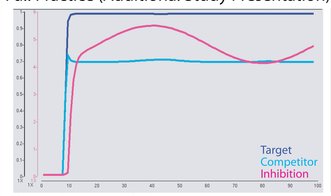


- RIF sims only used 4 patterns at training
- We also ran a more stringent test of the network's ability to learn patterns (results pictured at left):
  - 50 patterns at training
  - 32.5% overlap between patterns
  - Pattern completion tested with 50% partial cues

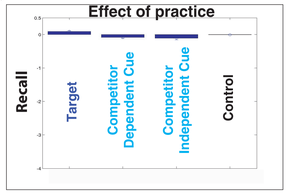
Data

Retrieval Induced Forgetting

Full Practice (Additional Study Presentation)

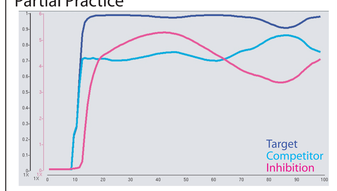


- Cue matches target perfectly
- Just like extra training
- Inhibitory oscillation does not affect network activity

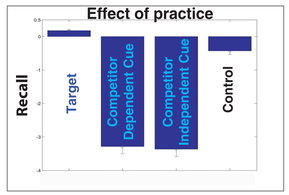


- Some facilitation of target item (already at ceiling)
- Very little effect on competitors
- Very little effect on control items

Partial Practice



- Match between cue and target (relative to match between cue and competitor) is less good here
- Competitor turns on during the low inhibition phase
- Change in activity leads to weakening of competitor



- Some facilitation of target item (already at ceiling)
- Large decrease in recall of competitor
  - This occurs regardless of cue (cue independent)
- More effect on controls than full practice, still very little
  - Also true of Anderson's data

Summary

- Local mechanism (no PFC) can account for basic Retrieval-Induced Forgetting results, including cue-independent effects of partial practice.

- Oscillation of inhibition reveals identity of competitors vs targets

- Weight changes based on oscillation phase and activity change can store new information and reduce competition

Discussion

- Raghavachari et al (2001) and Rizzuto et al.(2003) suggest that cortical oscillations play a role in learning. This model explicitly depicts the role that oscillations can play as an organizing force.

- Although our model (as presented here) does not include PFC, we think PFC plays a critical role in biasing competition when the correct response is not the dominant response (e.g., in the "think-no think" paradigm; Anderson & Green, 2001). Future modeling work will directly address PFC contributions.

- Our model shows how generic theta-like oscillations can help train cortical attractor networks. In future research, we will explore how our model relates to other, more biologically detailed models of how theta modulates learning, in the hippocampus and elsewhere (e.g., Hasselmo, Bodelon, & Wyble, 2002).

References

Anderson, M.C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, 410, 366-369.

Hasselmo, M.E., Bodelon, C., Wyble, B.R. (2002). A proposed function for hippocampal theta rhythm: separate phase of encoding and retrieval enhance reversal of prior learning. *Neural Computation*, 14, 793-817.

Huerta, P.T. & Lisman, J.E. (1996). Synaptic plasticity during the cholinergic theta-frequency oscillation in vitro. *Hippocampus*, 6(1), 58-61.

Levy, B.J. & Anderson, M.C. (2002). Inhibitory processes and the control of memory retrieval. *TRENDS in Cognitive Sciences*, 6(7) , 299-305.

Newman, E.L. & Norman, K.A. (2003, March). A neural network model of retrieval-induced forgetting. Poster at Cognitive Neuroscience Society Annual Meeting, New York, NY.

Norman, K.A., Newman, E.L., & Polyn, S.M. (in preparation). How theta oscillations can train neural networks and punish competitors.

Raghavachari, S., Kahana, M.J., Rizzuto D.S., Caplan, J.B., Kirschen, M.P., Bourgeois, B., Madsen, J.R., and Lisman, J.E. (2001). Gating of human theta oscillations by a working memory task. *J Neurosci*, 21(9), 3175-3183.

Rizzuto, D.S., Madsen, J.R., Bromfield, E.B., Schulze-Bonhage, A., Seelig, D., Aschenbrenner-Scheibe, R., and Kahana, M.J. (2003). Reset of human neocortical oscillations during a working memory task. *Proc Natl Acad Sci U S A*, 100(13), 7931-7936.

ELN was supported by an NIH Training Grant in Quantitative Neuroscience (MH65214) awarded to Princeton University

For a downloadable version, please visit: http://compmem.princeton.edu/publications.html