Moderate Memory Activation Leads to Forgetting in the Think-No Think Paradigm

PRINCETON Neuroscience INSTITUTE

Greg J. Detre[1,2], Annamalai Natarajan[1], and Kenneth A. Norman[1]

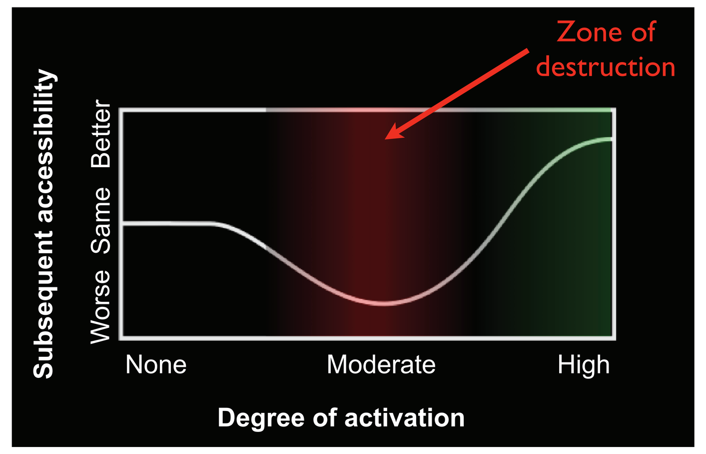
[1] Department of Psychology and Princeton Neuroscience Institute, Princeton University [2] Memrise, Inc. (www.memrise.com)

Princeton Computational Memory Lab

396.27

Introduction

Hypothesis: the relationship between memory activation and learning is **nonmonotonic**



If a memory is strongly retrieved, it gets strengthened

If a memory does not activate, nothing happens

If a memory activates to a moderate degree, it gets weakened

This nonmonotonic relationship is predicted by computational models of learning (e.g., Bienenstock, Cooper, & Munro, 1982; Norman, Newman, Detre, & Polyn, 2006)

The nonmonotonic pattern has been found at the synaptic level (post-synaptic potential: Artola et al., 1990; post-synaptic Ca2+ concentration: Hansel et al., 1996)

We want to see if this pattern occurs at the level of **memory representations**

Prior Work: Newman & Norman (2010)

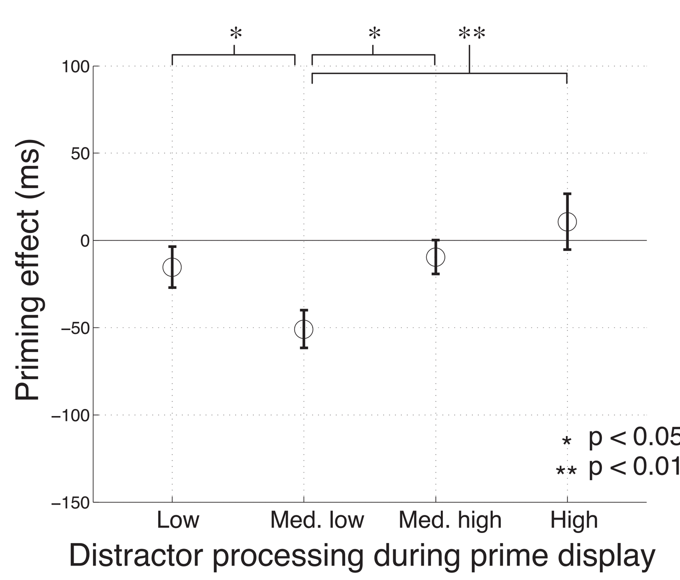
Newman & Norman set out to test the nonmonotonic plasticity hypothesis using a negative priming paradigm

Negative priming effect: Ignoring a distracting stimulus makes you slower to respond to it later (Tipper, 1985)

According to the nonmonotonic plasticity hypothesis, moderate activation of the distractor should weaken the distractor, leading to negative priming

Approach: Use pattern classifiers to track activity of the distractor. Relate this neural measure of distractor activity to priming effects.

Results (shown at right) fit with the nonmonotonic plasticity hypothesis.



Background: The Think/No Think Paradigm

Anderson & Green (2001) had subjects study novel word pairs like elephant-wrench

Think-no think phase:

For pairs assigned to the think condition, subjects are given the first word of the pair & told to recall the studied associate: ELEPHANT -

For pairs assigned to the no think condition, subjects are given the first word of the pair & told **not** to think of the studied associate: ELEPHANT -

Final memory test: ELEPHANT -

The no-think procedure leads to impaired recall, relative to baseline... but not always (e.g., Bulevich et al., 2006)

Puzzle: What factors account for variability in the size of the forgetting effect?

Hypothesis/Approach

Key prediction: Forgetting of no-think items will happen when the associate activates to a moderate degree

Failures to see forgetting of no-think items can be explained in terms of the associate activating too much or not enough

Strategy for testing this prediction: Use fMRI pattern classifiers to read out retrieval of the associate during no-think trials

Use this covert neural measure (taken during no-think trials) to predict recall of the item on the final memory test

Paradigm and Experimental Logic

To facilitate our classifier analyses, we had subjects learn word-picture pairs instead of word-word pairs. Pictures were drawn from four categories: Face, Scene, Car, Shoe

Study Phase (day 1, outside of scanner): Learn-to-criterion procedure; each pair was trained until subjects correctly recalled it once

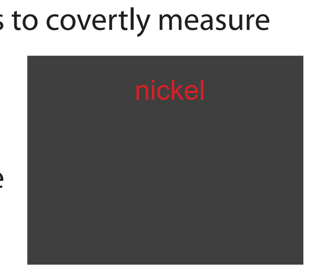
Think-No Think Phase (day 2, in scanner): Think items were shown in green and no-think items were shown in red. Think items were presented 6x. No-think items were presented 12x. All items shown in the Think-No Think phase were associated with either Faces or Scenes (never Shoes or Cars).

Functional Localizer Phase (day 2, in scanner): Subjects viewed pictures blocked by category and performed a one-back task. Data from this phase were used to train the category-specific classifiers

Final memory test (day 2, outside of scanner)

Analysis Plan: We trained pattern classifiers to track activation relating to the four categories (Face, Scene, Car, Shoe). We used these classifiers to covertly measure recall of associated items during the think-no think phase.

For example: Since nickel was paired with a face at study, we would use the face classifier on this trial to measure the extent to which the face associate was coming to mind



Prediction for this trial: Moderate levels of face activity should be associated with forgetting, higher levels of face activity should be associated with improved memory

Alternative analysis: Instead of using classifier outputs, we used time-courses from face-selective ROIs (FFA) and scene-selective ROIs (PPA, retrosplenial) to index face and scene recall.

fMRI Analysis Strategy

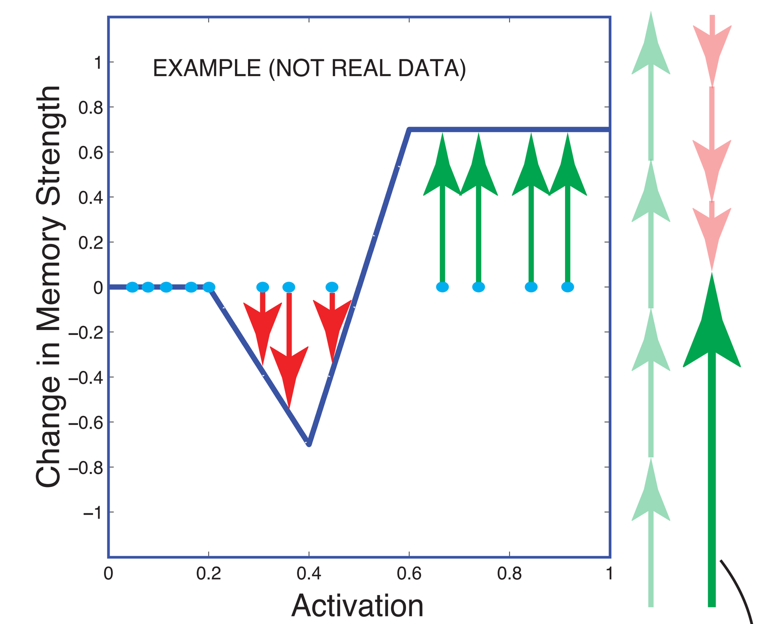
Our goal here was to use neural activity elicited by cues during the no-think phase to predict subsequent memory for the associated item

Each no-think cue (e.g., nickel) appeared 12 times during the no-think phase

Each of those 12 no-think trials was a separate learning opportunity

For each pair assigned to the no-think condition, we want to use neural data from that pair's 12 no-think trials to predict that pair's recall on the final test

If we assume a particular shape for the plasticity curve, we can **PREDICT** (for a particular pair) the effect of a no-think trials on subsequent memory for that pair



The 12 blue dots correspond to specific no-think trials for a particular pair. To evaluate the effect of a no-think (NT) trial, evaluate the plasticity curve at that x-value. We assume that the effects of no-think trials sum in a linear fashion.

Next, we can **EVALUATE** how well these predictions fit actual recall data.

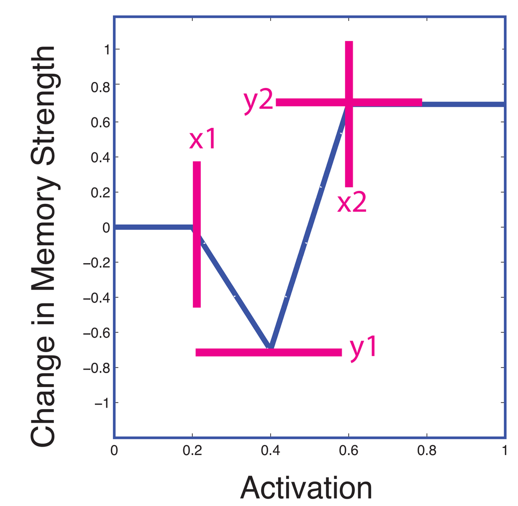
For this example, the correlation between actual and predicted recall = .3

Predicted Effect of NT	Final Test Recall
0.36	1
0.22	1
-0.15	0
0.25	0
-0.94	0
0.11	1
-0.45	0
0.84	0

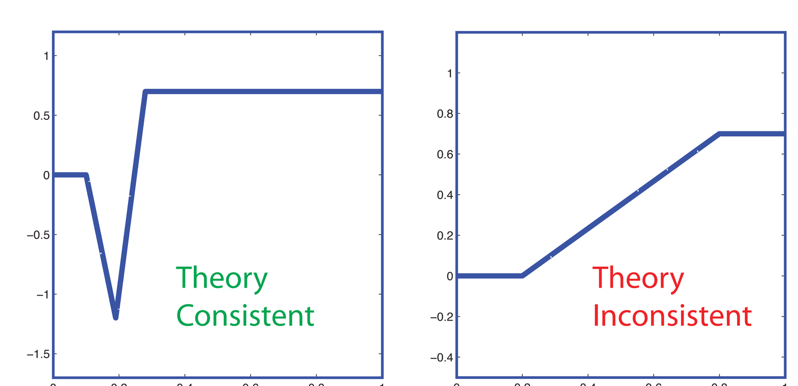
Analysis plan: **SEARCH** over the space of possible plasticity curves to find the curve that does the best job of predicting behavior

To do this, we parameterized the plasticity curves using four values (x1, x2, y1, y2).

Some of the resulting curves are theory consistent ($y1 < 0 < y2$) and some are not



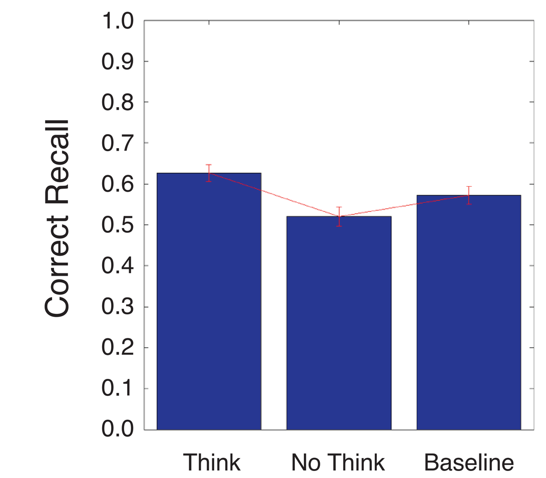
Examples of curves that were considered



For each curve, we get a correlation score saying how well that curve predicts final recall performance for no-think trials: Goal: Find the curve with the best correlation score.

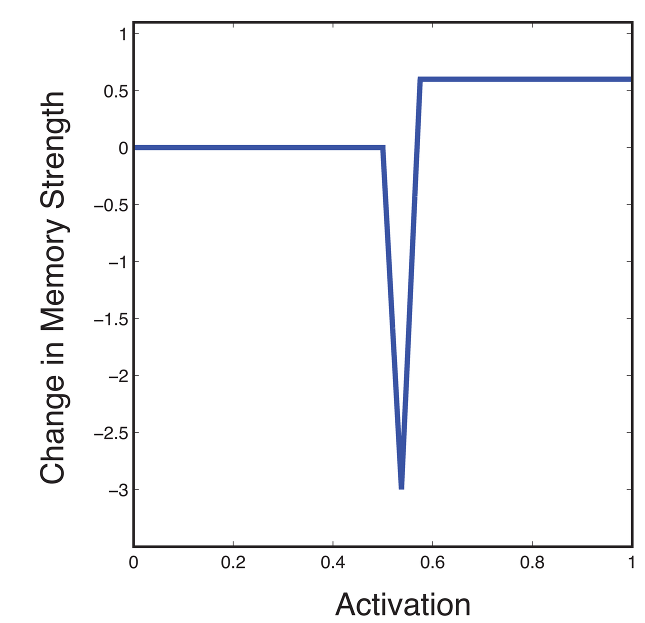
Behavioral Results From the Final Memory Test

Think > Baseline > No Think but Think vs. Baseline and No Think vs. Baseline differences were not significant



Preliminary fMRI Curve Fitting Results

Best-fitting curve:



The best-fitting curve is **theory-consistent** (it has a dip and then increases)

How well does the curve predict the data?

For this curve, the correlation between predicted and actual recall (across trials) for no-think items was **.27**. The probability of getting a correlation this large due to chance (assessed using a nonparametric permutation test) was $p < .001$.

Current and Future Work

The above analysis tells us which curve fits best

What we really want to know is the **overall probability** (based on the data) that the curve is theory-consistent

To answer this question, we are currently running a hierarchical Bayesian analysis (Gelman & Hill, 2007)

Instead of picking out a single best-fit curve, hierarchical Bayes yields a posterior probability distribution over plasticity curves. To compute the posterior probability that the curve is theory-consistent:

- Sample from the posterior distribution
- Compute the proportion of sampled curves that are theory-consistent

Hierarchical Bayes also allows us to model inter-subject variability

Our analyses up to this point have assumed that all subjects have the exact same plasticity curve. This one-size-fits-all approach may be impeding our ability to fit the model.

By contrast, hierarchical Bayes fits curve parameters at both the group level and the subject level. **This approach allows us to model commonalities in subjects' plasticity curves while still allowing for between-subject differences in curve shape.**

Conclusions

Preliminary results from the curve-fitting analysis support the nonmonotonic plasticity hypothesis:

Moderate levels of memory activation during the no-think phase were associated with forgetting.

Higher levels of memory activation during the no-think phase were associated with improved memory

But: **This is still a work in progress.** More analyses are needed to assess the robustness of these findings.

References

Anderson, M. C. & Green, C. (2001). Suppressing unwanted memories by executive control. Nature, 410, 366-9.

Artola, A., Bröcher, S. & Singer, W. (1990). Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. Nature, 347, 69–72.

Bienenstock, E. L., Cooper, L. N. & Munro, P. W. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. Journal of Neuroscience, 2, 32–48.

Bulevich, J. B., Roediger, H. L., Balota, D. A., & Butler, A. C. (2006). Failures to find suppression of episodic memories in the think/no-think paradigm. Memory & Cognition, 34, 1569-77.

Gelman, A. & Hill, J. (2007). Data Analysis Using Regression and Multilevel/Hierarchical Models. New York: Cambridge University Press.

Gotts, S. J. & Plaut, D. C. (2005). Neural mechanisms underlying positive and negative repetition priming. Poster presented at the Annual Meeting of the Cognitive Neuroscience Society.

Hansel, C., Artola, A. & Singer, W. (1996). Different threshold levels of postsynaptic [Ca2+]i have to be reached to induce LTP and LTD in neocortical pyramidal cells. Journal of Physiology Paris, 90, 317–9.

Newman, E. L. & Norman, K. A. (2010). Moderate excitation leads to weakening of perceptual representations. Cerebral Cortex, 20, 2760-70.

Norman, K. A., Newman, E. L., Detre, G. J. & Polyn, S. M. (2006). How inhibitory oscillations can train neural networks and punish competitors. Neural Computation, 18, 1577–610.

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. The Quarterly Journal of Experimental Psychology, 37A, 571-590.

Funding for this research was provided by NIMH grant R01 MH069456 to KAN

This poster can be downloaded from http://compmem.princeton.edu