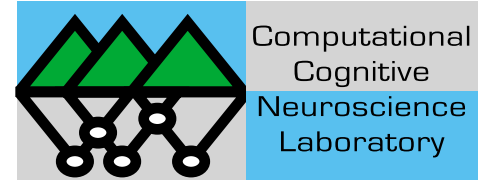# Differential Effects of List Strength on Recall and Familiarity

Kenneth A. Norman (norman@psych.colorado.edu)
Dept. of Psychology, University of Colorado at Boulder

poster presented at the 2000 meeting of the Psychonomic Society, New Orleans, LA

Computational
Cognitive
Neuroscience
Laboratory

Introduction

The effect of *list strength* on recognition memory has been the subject of intense scrutiny over the past decade:

Is there a COST associated with memory strengthening, whereby strengthening memory for some list items impairs recognition of other (non-strengthened) list items?

Compare:

- study Apple, Robot, vs.
- study Apple, Robot, Robot, Robot

=> Does strengthening your
     memory for "Robot"
     hurt recognition of "Apple"?

Several studies, starting with Ratcliff et al. (1990) have failed to find a list strength effect *(LSE)* on recognition.

Math models of recognition have undergone substantial revision to accommodate this (null) finding.

However, the range of conditions under which the LSE has been studied is still limited -- just because the LSE is sometimes null does not mean it will always be null!

Overview

In the *theory* section of the poster (outlined in GREEN), I present a new, biologically-based dual-process neural network model of recognition memory. (Norman & O'Reilly, in prep.)

The model predicts that an LSE should be present for one process that contributes to recognition (hippocampally-driven *recall*) but not for the other (neocortically-driven *familiarity*).

=> This prediction implies that an LSE on recognition should be present whenever recall is making a substantial contribution to recognition (relative to familiarity).

In the *data* part of the poster (outlined in BLUE), I present several new list strength experiments that test this hypothesis.

Modeling the Neural Basis of Recognition

Our model focuses on the contributions of two brain structures to recognition memory:

- the hippocampus
- medial temporal neocortex *(MTLC)*,
  which serves as the interface between
  the hippocampus and the rest of neocortex
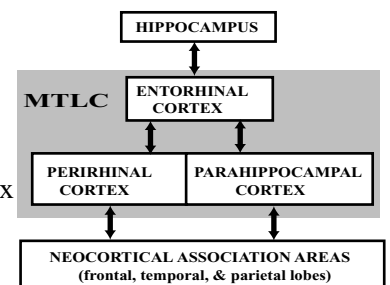
Lesion evidence indicates that:
- the hippocampus is essential for **recall**
- when the hippocampus is lesioned, MTLC can support
  some degree of recognition based on nonspecific feelings
  of **familiarity** (for a review, see Aggleton & Brown, 1999)

Norman & O'Reilly have constructed neural network models of hippocampus and MTLC -- the models can be used to simulate these structures' (respective) contributions to recognition memory.

The models incorporate several key principles of neural computation, including -- but not limited to -- Hebbian LTP/LTD (long-term potentiation/depression) and inhibitory competition between neurons (O'Reilly & Munakata, 2000).
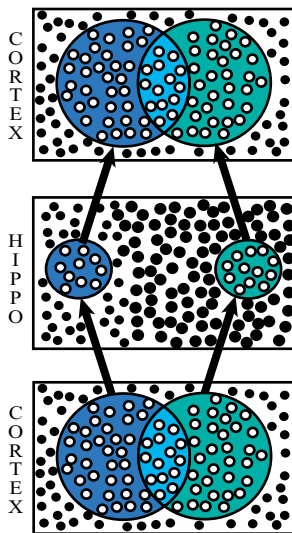
Hippocampal Recall

The hippocampal network rapidly memorizes patterns of cortical activity in a manner that supports subsequent recall, based on partial cues.

The hippocampus links input patterns to clusters of units in region CA3, which are linked back to a copy of the input (via region CA1).

Hippocampal processing can be understood in terms of the interplay between *pattern separation* and *pattern completion*.
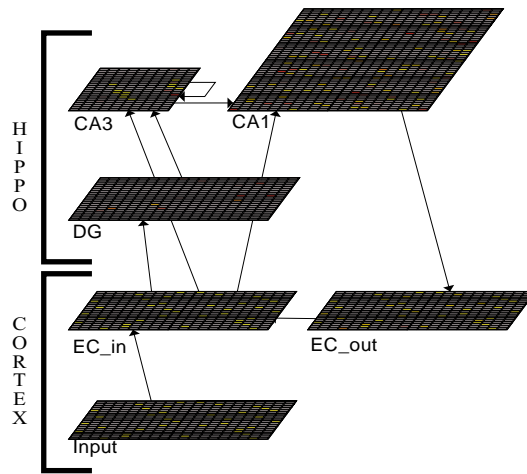
To minimize interference between traces, the hippocampus is biased to assign relatively non-overlapping *(pattern-separated)* representations to stimuli.

However, if an input pattern overlaps strongly with a previously studied input, the second pattern will activate the CA3 representation of the first pattern *(pattern completion)*, and recall occurs.
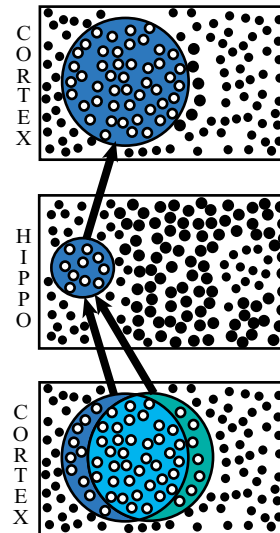
To apply the hippocampal model to recognition, we examine the extent to which the test cue is recalled:

Recall measure = (# of recalled features that *match* the test cue)
   - (# of recalled features that *mismatch* the test cue)

False recall is very rare in the hippocampus because of the pattern separation/completion dynamic...

=> lures have to be very similar to studied items to trigger pattern completion
=> when pattern completion occurs, features of the ***studied item*** are recalled, and the lure can be rejected based on mismatch between the test cue and recalled features

Neocortical Familiarity

Neocortex learns gradually, integrating across events to arrive at a representation of what is *generally* true in the environment. (McClelland et al., 1995)

Neocortex assigns similar representations to similar stimuli => this allows neocortex to represent what different events have in common

MTLC

Input
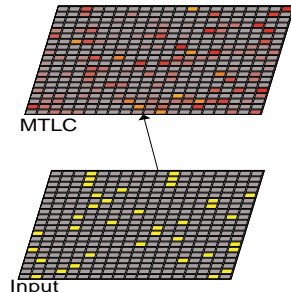
How can neocortex contribute to recognition after a single study exposure, if it is supposed to learn slowly (integrating over events)?

- when an item is studied, Hebbian learning tunes a small number of units to respond strongly to the item's features; these units inhibit units that are less strongly active

Thus, as items become more familiar, representations become *sharper*:

- unfamiliar stimuli *weakly* activate a *large* number of units
- familiar stimuli *strongly* activate a relatively *small* number of units

unfamiliar   **familiar**
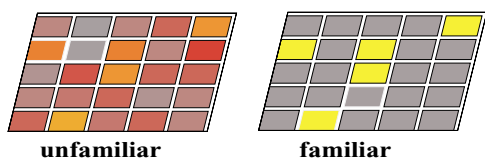
To index sharpness -- and thus familiarity -- we compute the following measure:
average activity of the K most active units (K is a model parameter)

Simulations show that the neocortical familiarity measure increases as a function of "global match"; the familiarity of nonstudied items increases smoothly as a function of their similarity to studied items.

List Strength Simulations

PARADIGM used in simulations and subsequent experiments:

Subjects study **target items** (which are later tested) and **interference items** (which are not tested).

compare two conditions:

=> in the **weak interference** condition, interference items are studied **once**
=> in the **strong interference** condition, interference items are studied **multiple times**

Subjects are given a recognition test consisting of target items and nonstudied lure items.

=> if recognition is better in the weak interference condition, this constitutes a list strength effect!

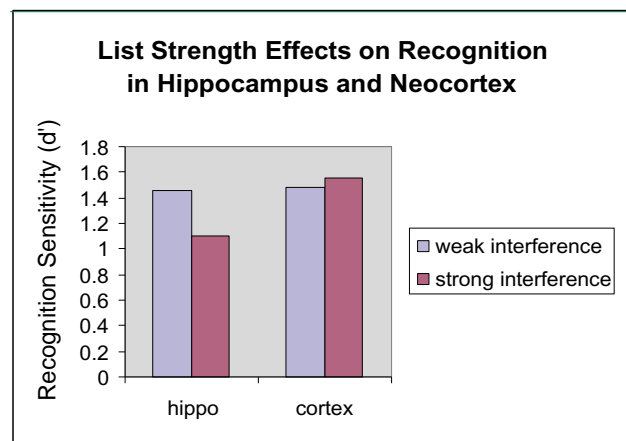**SIMULATION:**

parameters:

-10 studied items, 10 interference items, 1X vs. 2X strengthening, 20% overlap between input patterns

result:

- There was a list strength effect for hippocampal recall but not for neocortical familiarity!

List Strength Effects on Recognition in Hippocampus and Neocortex

Recognition Sensitivity (d')

weak interference
strong interference

hippo   cortex

Explaining the Recall LSE

As a general principle, interference occurs in neural network models whenever items have overlapping representations (i.e., they activate the same hidden units)

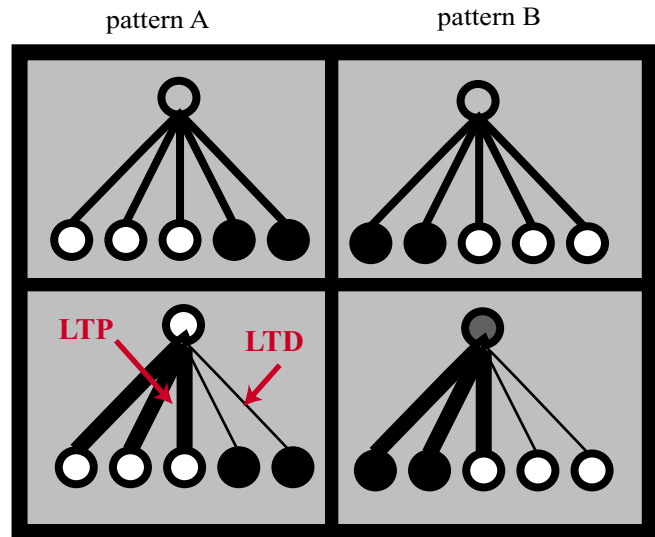This diagram illustrates interference between memory traces.

When pattern A is studied:

Weights to features *shared* by patterns A and B *increase* (due to Hebbian LTP)

However, weights to unique, *discriminative* features of pattern B *decrease* (due to Hebbian LTD)

pattern A   pattern B

BEFORE studying pattern A

LTP   LTD

AFTER studying pattern A

This latter factor (LTD) hurts memory for pattern B!

Interference occurs in the hippocampus because:

- even though there is *less* overlap between representations of list items in the hippocampus (vs. neocortex), there is typically *some* overlap in the hippocampus
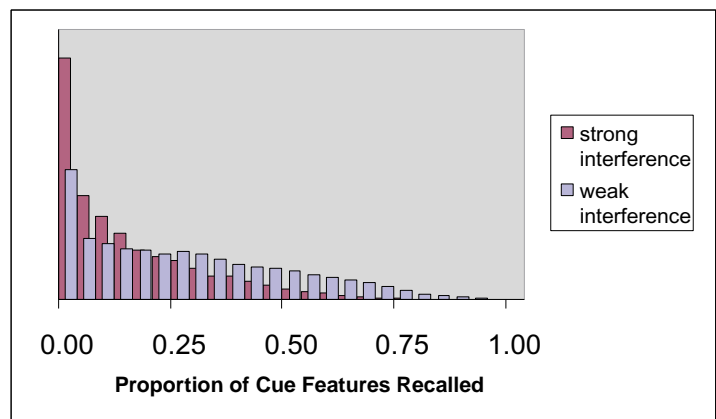
=> every time the hippocampal representations of two studied items overlap, the representation of the first item is weakened (due to Hebbian LTD)

=> the only time we would expect NO interference in the hippocampus is when cortical memory traces are very distinctive; in this case, there might not be enough hippocampal overlap to result in noticeable interference

This histogram shows the distribution of recall scores for studied items, as a function of interference

Interference pushes the studied recall distribution towards zero.

As mentioned earlier, false recall of lure items is rare. In this simulation, **recall of lure items was at floor in both conditions: > 97% of lure items triggered zero recall...**

- strong interference
- weak interference

0.00   0.25   0.50   0.75   1.00

Proportion of Cue Features Recalled

Since interference shifts the studied recall distribution to the left, and the lure recall distribution is unaffected by interference (due to floor effects), the net effect is a decrease in d'.

Explaining the (Null) Familiarity LSE

Recognition discrimination in neocortex depends on the fact that -- because of learning -- the network is *more sensitive to* (i.e., has stronger weights to) *the discriminative features of studied items than to the discriminative features of lures.* Note that sensitivity to features *shared* by studied items and lures (e.g., context; prototypical stimulus features) does not benefit recognition performance. Put another way: Recognition is a function of the **gap** in the network's sensitivity to discriminative features of studied items vs. lures.

This graph illustrates how interference affects the network's sensitivity to *discriminative* and non-discriminative (*shared*) features of studied items and lures. Interference has two primary effects:

- Overall, network becomes *less* sensitive to *discriminative* features of individual items, and *more* sensitive to non- discriminative, *shared* features => this tends to **hurt** recognition performance

- However, there is also an interaction, whereby ***the effects of interference are stronger for lure items than studied items***: Sensitivity to *lure items'* features decreases more rapidly than sensitivity to *studied items'* features => the *gap* in sensitivity between studied items and lures increases, and this tends to **boost** recognition performance.

- The second effect (an increase in the studied-lure "sensitivity gap") initially outweighs the first (an overall decrease in sensitivity to discriminative features), causing a slight increase in d' => However, the overall decrease in sensitivity to discriminative features eventually takes its toll, and d' starts to decrease
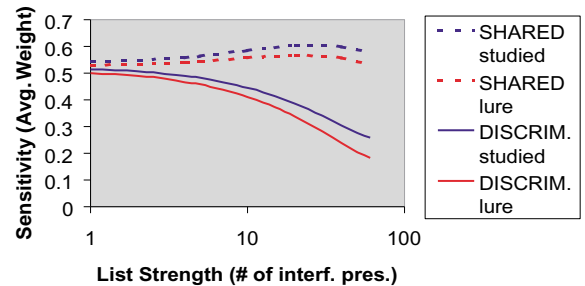
The fact that interference affects studied items less than lures can be explained in terms of **differentiation** (Shiffrin et al., 1990; McClelland & Chappell, 1998) -- strengthening an item's representation makes it more *selective*, such that it is *more* strongly activated by the item itself, but *less* strongly activated by other, interfering items

Differentiation is a simple consequence of Hebbian learning! Referring back to the 6-neuron diagram (on the previous page), studying pattern A has two consequences:
- it *strengthens* weights to the features of pattern A (Hebbian LTP), but
- it *weakens* weights to the (discriminative) features of pattern B (Hebbian LTD)
... as a result, the receiving unit fires *more* to pattern A but *less* to pattern B.

=> effectively, studying an item *pulls its representation away* from other (dissimilar) items' representations in feature space. Because studying an item ***decreases representational overlap*** with other items, studied items' representations suffer ***less interference*** than nonstudied items' representations!

This graph shows how list strength affects studied and lure familiarity. Initially, the familiarity gap between studied items and lures *increases* because of differentiation (leading to an increase in d').
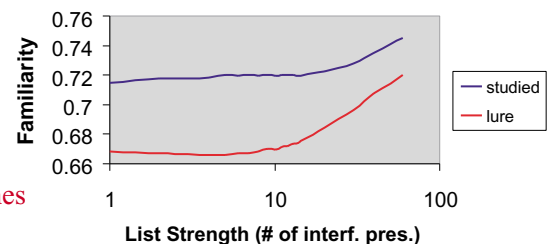
Then, the gap decreases (leading to a decrease in d') as the network becomes increasingly sensitive to non-discriminative prototype/context features (at the expense of representing discriminative features).

BOUNDARY CONDITIONS:
The differentiation dynamic only holds when items do not overlap too strongly. When items are very similar to one another, studying an item makes its representation *more* (not less) similar to other items' representations => in this case, studied items suffer *more* interference than lures and d' decreases monotonically as a function of interference.

Asymptotically (with enough interference), the network will always degrade to the point where it only represents what items have in *common* (e.g., context info), but not what makes items *distinct* from one another -- at this point, d' = 0. The amount of overlap between items determines *how quickly* the network arrives at this degenerate state (more overlap => faster degeneration), and what *trajectory* the d' scores follow (i.e., do they increase, then decrease, or do they decrease monotonically).

Testing the Model's Predictions

To recap: the model predicts that list strength...

- *should* impair recognition based on the hippocampal *recall* process

- should *not* impair recognition based on the cortical *familiarity* process

... assuming that overlap between items is not extremely high (in which case both processes should show an LSE), or extremely low (in which case neither process should show an LSE)

=> this implies that an LSE on recognition should be present whenever recall is making a substantial contribution (relative to familiarity)

Existing data:

- some studies have found an LSE on cued recall (e.g., Kahana, submitted), but...

- all studies that have looked for an LSE on recognition have failed to find one!

maybe recall was not making enough of a contribution...

Goals for experiments:

- to show that an LSE on recognition emerges when recall is contributing (exp. 1)

- to obtain direct evidence that list strength affects recall but not familiarity (exp. 2)

Expt. #1: Filter by Confidence Ratings
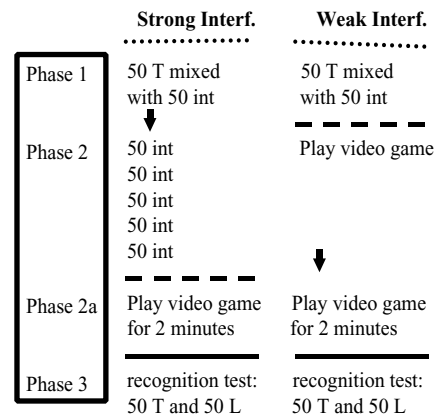
Several studies (e.g., Yonelinas et al., 1996; Yonelinas, in press) have found that recall is associated with high confidence recognition responses, but familiarity is associated with a wide range of confidence responses

=> this implies that we can (at least partially) isolate the contribution of recall by focusing on high-confidence responses

IF: high-confidence "old" responses are primarily driven by recall
AND: there is an LSE on recall
THEN: *if we restrict the analysis to high confidence "old" responses, we should see a list strength effect on recognition accuracy*

METHOD:

Concrete noun stimuli:
50 Targets (T)
50 Lures (L)
50 Interference items (int)

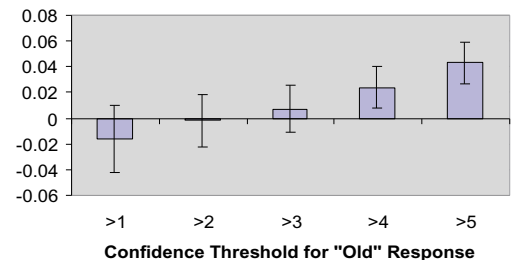|	Strong Interf.	Weak Interf.
Phase 1	50 T mixed with 50 int	50 T mixed with 50 int
Phase 2	50 int / 50 int / 50 int / 50 int / 50 int	Play video game
Phase 2a	Play video game for 2 minutes	Play video game for 2 minutes
Phase 3	recognition test: 50 T and 50 L	recognition test: 50 T and 50 L

- Subjects rated recognition confidence from 1 (sure new) to 6 (sure old); recognition accuracy was computed using different confidence thresholds for accepting an item as "old", e.g., (conf. > 3) = "old"; (conf. <= 3) = "new"

- The encoding task ("would this item fit in a small box"; 1.15 sec per word) was designed to yield memory traces rich enough to support some recall, but not so distinctive as to yield ceiling effects on recall

RESULTS:

As predicted, a significant LSE on recognition accuracy emerged when accuracy (A') was computed using a high confidence threshold (conf. > 4 or 5) for acceping an item as "old".

LSE as a Function of Criterion Placement (in A' units)

ROC data indicate that recognition accuracy was significantly higher in the Weak Interference condition, Ag = .914 in the Weak Interference condition and Ag =.881 in the Strong Interference condition (Ag is an assumption-free estimate of the area under the ROC, Macmillan & Creelman, 1991).
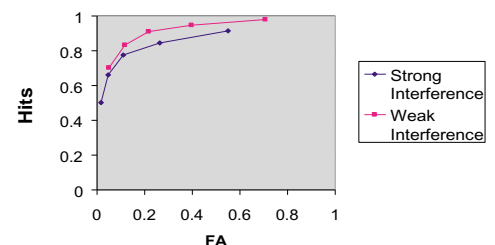
Experiment #2: Self-Report Measures of Recall & Familiarity

PARADIGM:
Whenever a subject thinks that an item was studied, ask them if they *remember* specific details from when they studied the item, or it just seems *familiar*.

Jacoby et al. (1997) showed that, *if* you assume recall and familiarity are independent, it is possible to use subjective-report data to quantiatively estimate:

1) P(R): the probability of recalling a studied item, and
2) Fd': the extent to which familiarity discriminates between studied items and lures
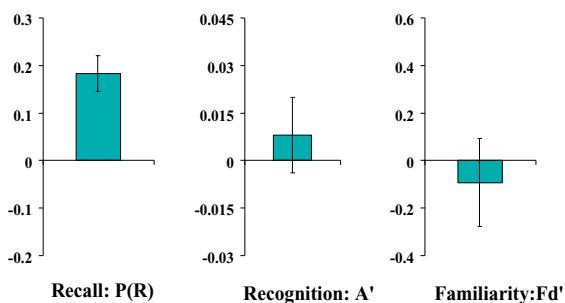
PREDICTIONS:
List strength should affect the derived measure of recall, P(R), but list strength should *not* affect the derived measure of familiarity, Fd'

Methodologically, the paradigm was very similar to the paradigm used in Expt. 1, except "remember/familiar" responses were obtained instead of confidence ratings.

RESULTS:

List Strength Effect

Recall: P(R)   Recognition: A'   Familiarity: Fd'

As predicted, the LSE on Recall was highly significant, and the LSE on Familiarity was not significant; consistent with other list strength experiments, the overall LSE on Recognition was not significant.

Future Directions: Using Related Lures to Boost the Contribution of Recall...

The model predicts that our ability to discriminate between studied items and related lures depends critically on the hippocampal recall process.

=> the cortical familiarity model performs poorly on tests with related lures, because these lures trigger strong feelings of familiarity, leading to false recognition (and low d' scores); the hippocampal model performs better, because of its ability to assign distinct *(pattern-separated)* representations to similar items

IF recognition tests with related lures load heavily on the recall process, THEN increasing list strength should hurt performance on recognition tests with related lures (moreso than on tests with unrelated lures).
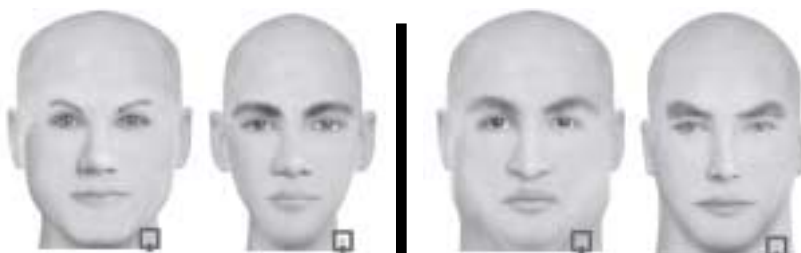
Face morphing can be used to create face pairs that are related to one another. An experiment using related vs. unrelated face pairs is in progress now...

STUDY:

then TEST:

studied vs. related lure   or   studied vs. unrelated lure

List strength should impair studied vs. related judgments more than studied vs. unrelated judgments!

Summary and Conclusions

I presented a new, biologically-based dual-process network model of recognition memory, which predicts that a list strength effect should be obtained for hippocampal **recall** but not neocortical **familiarity** (assuming that between-trace overlap is neither extremely high nor extremely low). This implies that an LSE on recognition should be present whenever recall is making a large contribution to recognition performance (relative to familiarity). In two separate experiments, I obtained evidence consistent with this prediction.

=> Yonelinas et al. (1996) and others have found that recall is associated with high-confidence recognition responses. In Experiment 1, I found a significant recognition LSE when I computed recognition accuracy based on high confidence "old" responses. Lowering the confidence threshold (which allows for a greater contribution of familiarity) eliminated the recognition LSE. The area under the ROC was also significantly affected by list strength in this experiment. These findings demonstrate that (contrary to published findings) **recognition sensitivity sometimes is affected by list strength**.

=> In Experiment 2, I found that measures of recall- and familiarity-based discrimination derived from self-report data (using the IRK procedure; Jacoby et al., 1997) were differentially affected by list strength. There was a significant LSE for the recall measure but not the familiarity measure.

THEORETICAL IMPLICATIONS:
Our hippocampal and neocortical models -- like all neural network models where there is overlap between traces -- suffer from interference at **storage**; new learning experiences degrade the memory traces of other stored items.

- This differs from other recognition models (e.g., REM; Shiffrin & Steyvers, 1997), which posit that memory traces are stored separately, and that interference arises from spurious trace activation at **retrieval**.

Murnane & Shiffrin (1991) questioned whether models that posit interference at storage could account for the null LSE on recognition. Our cortical-model simulations show that biologically realistic neural network models (with overlapping representations) **are** capable of accommodating the null list strength finding

Although there are some similarities between our account of the null LSE on familiarity, and REM's account (they both rely on the differentiation principle, albeit in different ways), the accounts are different enough that it should be possible to tease them apart empirically. In particular, our model's prediction that interference causes cortical traces to degrade (in a very predictable way -- cortical representations become more sensitive to shared features, and less sensitive to discriminative features) seems very distinct from the predictions of models like REM (which posit, for some parameter settings, that strengthening interfering items can actually *remove* interference).

LIST LENGTH:
Although extant studies have failed to obtain an effect of list **strength** on recognition, several studies have found that increasing list **length** -- adding new items to the study list -- impairs recognition (e.g., Ohrt & Gronlund, 1999). Our model can accommodate the length/strength dissociation given the added postulate that the first presentation of an item leads to substantially more weight change than subsequent presentations of an item.

If this is true, presenting several items for the first time will lead to more weight change than repeating already-studied items, and therefore is more likely to push the network into the degenerate range where shared features are represented at the expense of discriminative features (causing a decrease in d'). There is good neurobiological evidence for this claim! The effect of a stimulus on synaptic weights (LTP) is transiently quite large relative to its stable long-term value (e.g., Malenka & Nicoll, 1993). All studied stimuli incur this large, transient weight change. Stimulus repetitions result in adjustments to the stable long-term weight value that are small relative to the transient effect (and thus less likely to cause interference).

PREDICTION: If list length effects on familiarity are caused by the large, transient component of LTP, it should be possible to eliminate the list length effect on familiarity by waiting for the transient component to decay to baseline (on the order of 10-20 minutes).
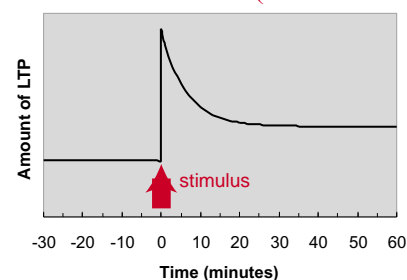
References

Aggleton, J.P., & Brown, M.W. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behavioral & Brain Sciences, 22*, 425-489.

Jacoby, L.L., Yonelinas, A.P., & Jennings, J.M. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific Approaches to Consciousness* (pp. 13-47). Mahwah, NJ: Erlbaum.

Kahana, M. J. (submitted). A factorial analysis of the recogntion-recall relation in four distributed memory models.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge Univ. Press.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105*, 724-760.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419-57.

Malenka, R. C., & Nicoll, R. A. (1993). NMDA-receptor-dependent synaptic plasticity: multiple forms and mechanisms. *Trends in the Neurosciences, 16*, 521-527.

Murnane, K., & Shiffrin, R. M. (1991). Interference and the representation of events in memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 17*, 855-874.

Norman, K. A. (submitted). Differential effects of list strength on recollection and familiarity.

Norman, K. A., & O'Reilly, R. C. (in prep.). A complementary learning systems model of recognition memory.

Ohrt, D.D., & Gronlund, S. D. (1999). List-length effect and continuous memory: Confounds and solutions. In C. Izawa (Ed.), *On Human Memory: Evolution, Progress, & Reflections on the 30th Anniversary of the Atkinson-Shiffrin Model*. Mahwah, NJ: Erlbaum.

O'Reilly, R.C., & Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. Cambridge, MA: MIT Press.

Ratcliff, R., Clark, S.E., & Shiffrin, R.M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16*, 163-178.

Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16*, 179-195.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM -- retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*, 145-166.

Yonelinas, A. P. (in press). Consciousness, control, and confidence: The three Cs of recognition memory. *Journal of Experimental Psychology: General*.

Yonelinas, A.P., Dobbins, I., Szymanski, M.D., Dhaliwal, H.S., & King, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness and Cognition, 5*, 418-441.