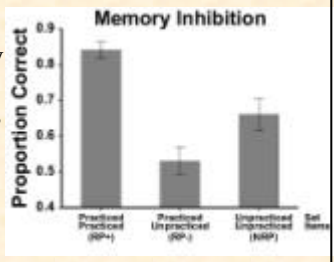
Modeling retrieval induced forgetting

Is it just a case of forming representations?

Retrieval Induced Forgetting

- Recall of some items from memory can reduce ability to recall other items
- Reduction is for related items only



Modelling retrieval induced forgetting

- Build model
  - Define general properties
  - Select model
- Test model
  - Does it show retrieval induced forgetting?
  - Can the model predict?

Building the model

- Caveats
  - Incorporation of psychological processes
    - Can fit the data - but does not imply the processes are in operation
  - High degrees of freedom
    - Over-fitting of data -> poor generalisation
- Get arounds
  - Selection done in broad terms without recourse to specific psychological processes
  - Experimental testing of predictions
    - Parameters should relate to psychological processes

Selection of the model

- Cortical damage necessary for retrograde amnesia
  - Neural representation in cortex relate to LTM
- Inhibitory mechanisms
  - formation of neural representations
  - cognitive interactions (e.g RIF)
- Could formation of neural representations (memories) also explain retrieval induced forgetting?

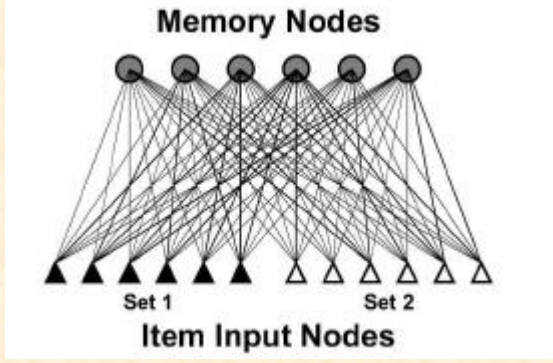
Properties of model

- Fully connected single layer unsupervised competitive model
  - learns from both past and present activity
  - competition (inhibition) between output nodes
  - loads output nodes with equal share of the input variance (Foldiak 1990)
  - forms semantic and episodic like representations (Foldiak 1990; Oram & Foldiak 1996)

Details of model

- 2 sets of 10 input nodes (items in RIF)
- 2 input nodes to indicate "set"
- 10 output nodes (memory / representation)
- Competition as winner takes all
- Initially random weights
- Learning rule: Modified Hebbian learning
  - Uses "trace" activity
    - $Trace_{(time)} = (1-\delta)Trace_{(time-1)} + \delta Act_{(time)}$
  - Weights bounded
    - $\Delta Wt_{[i,j]} = \alpha(Act_i - Wt_{[i,j]})Trace_j$

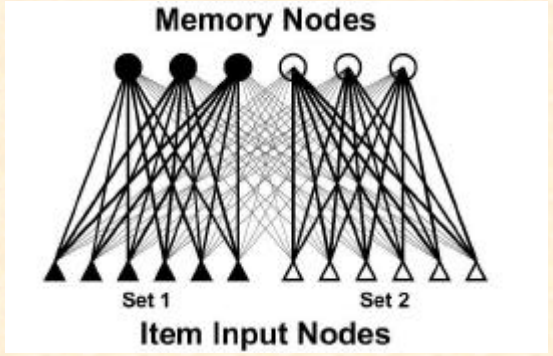
The model - Before learning



Running the model - Learning phase

- Initial training
  - Activate one set node
    - Activate item node associated with set then reset
    - Activate another associated item node then reset
  - Activate second set node
    - Activate item node associated with set then reset
    - Activate another associated item node then reset

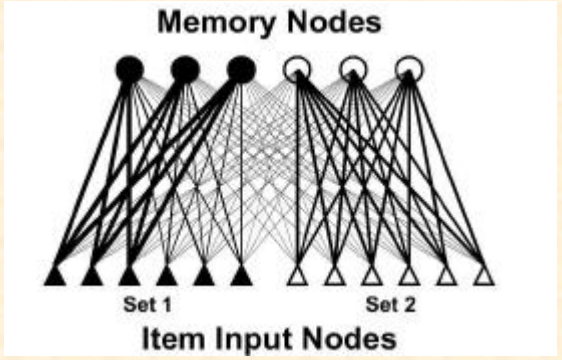
The model - After learning



Running the model - Retrieval Practice Phase

- Practice phase
  - Activate the "practice" set node
    - Partial activation of associated item node, reset
    - Activate another associated item node then reset
  - Restrict to the first half of items associated with the practice set node

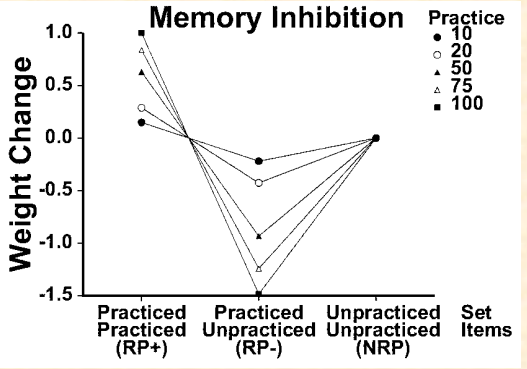
The model - After practice



Quantifying the model effects

- Use strength of representation
  - Activate set node (test phase) ⇒ activation of memory nodes ⇒ activation of item nodes
  - Take average of activation of item nodes as measure of recall performance

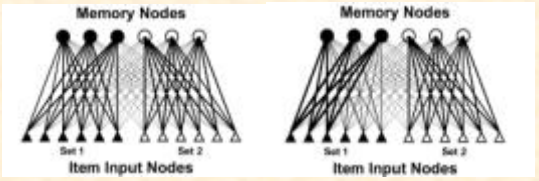
Model of retrieval induced forgetting



Why does the model "work"?

- Basic mechanism
  - Links with both end nodes active are strengthened
  - Links with one node inactive are weakened
  - Training (phase 1) establishes initial representation
  - Retrieval practice modifies that representation

After initial training **After retrieval practice**



Predicting from the model

- Simple model fits the data
  - Selection in broad terms without recourse to specific psychological processes
    - Suggests that RIF not a specific process?
  - High degrees of freedom
    - Over-fitting of data ⇒ poor generalisation?
- Need to test the model
  - Do parameters in the model relate to meaningful psychological concepts?
  - Does the model generalize appropriately?

Empirical testing of the model

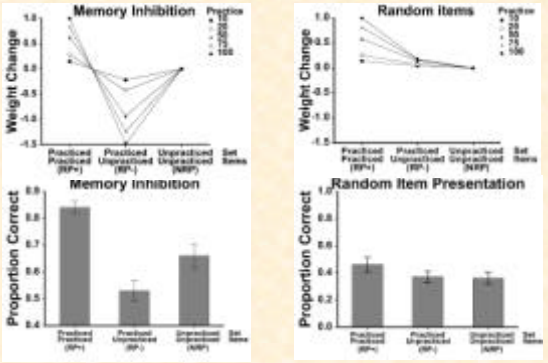
- One test of "psychological process - model parameter"
  - Effect of reducing temporal coherence
    - Setting the trace activity to zero
- Two tests of generalisation
  - What if items in different sets are related to each other?
    - E.g. Addidas trainers & Reebok trainers
  - What if items are related by pre-existing sets?
    - E.g. Peter & Paul are both male

Testing temporal coherence: Experimental details

- Condition 1:
  - Model: Trace activity
  - Expt: Items in order
- Condition 2
  - Model: No trace
  - Expt: Random order
- Recall test:
  - What did the man shout?
  - What colour was the jacket?

- Bank robbery (10 items)
  - The man took a small pistol from his jacket
  - He shouted "Lie down"
  - He made the teller empty the money into a sports bag
  - The police appeared as he was leaving by the back door
- Theft (10 items)
  - Two ladies were talking in the shop
  - The lady in green ran towards the shop exit
  - The other lady realised her wallet was missing

Testing temporal coherence

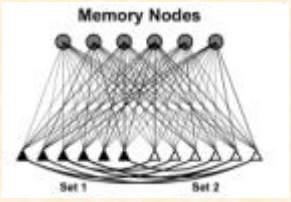


Testing effect of item relatedness and pre-existing memories

- Learning phase
  - each item/set pair presented for 5 seconds
- Practice phase:
  - 1/2 items from one set as stem completion task
- Distracter task
  - name as many capital cities as you can
- Surprise recall task

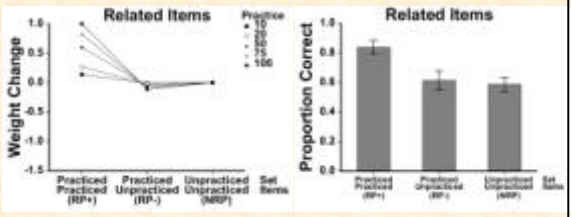
Testing with related items

- Modelling method
  - Same basic model structure, learning etc
  - Add connections between some items in different sets



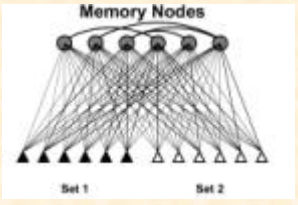
- Experimental detail
  - Use items which have connections between them
    - *Bill_Nike trainers, Bill_Slim build…, John_Adidas trainers, John_Medium build…*
  - *trainers*, *build* etc form links between the items
  - links are made explicit by the items

Testing with related items: Results



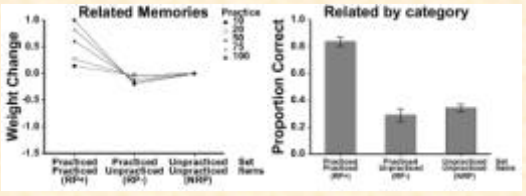
Testing with pre-existing sets

- Modelling method
  - Same basic model structure, learning etc
  - Add connections between memory nodes
    - Conceptually via another set node



- Experimental detail
  - Use items which have connections between them
    - learn representations of employees that were confounded by gender
    - *Tesco_Paul, Tesco_Ann…, Asda_Peter, Asda_Mary…*
  - Links between the items by existing set ("gender")
  - No explicit link

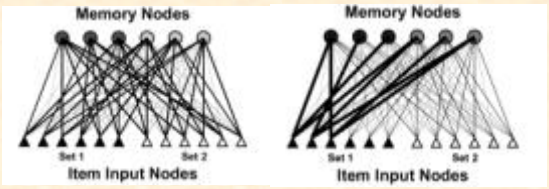
Testing with pre-existing sets: Results



Why does the model predict?

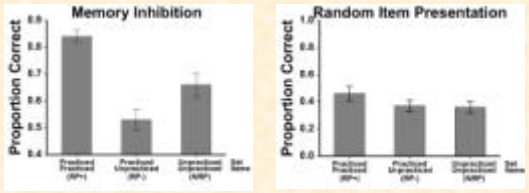
- If initial representation does not separate sets
- Retrieval practice modifies the representation
  - Effects are on items associated with both sets

After initial training **After retrieval practice**



Further observations

- When no retrieval induced forgetting observed
  - Discrimination between sets reduced
    - Predict reduced absolute levels of performance



Interim Summary

- A simple model of memory formation shows retrieval induced forgetting
  - Inhibitory processes can explain RIF
- Parameters of model relate to meaningful psychological concepts
  - The trace activity relates to coherence of items
- Predicts when RIF will and will not occur
  - Depends on the nature of the initial representation that is formed
  - Overlap in initial representation disrupts RIF

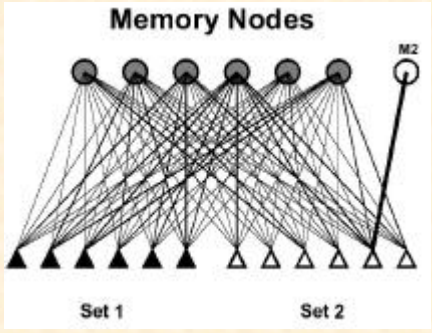
Independent probe technique

- Run a retrieval induced forgetting experiment
  - Learning Phase
    - Red_Bus, Red_Raspberry, Red_Shoes, Red_Head... House_Window, House_Garden, House_Kitchen...
  - Retrieval Practice
    - Red_B..., Red_S..., Red_H...
- Use independent probe for assessing recall
  - Name as many fruit as you can
- Still get retrieval induced forgetting
  - Subjects less likely to recall raspberry

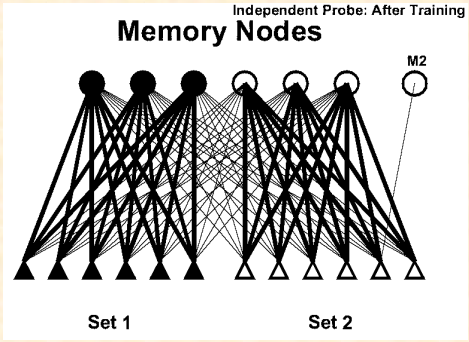
Modeling independent probe effects

- Include an extra “memory” node (M2)
  - Representing fruit-exemplar memory
- Set up a link to one of the input items
  - The “Tomato” node
- Run the standard program sequence

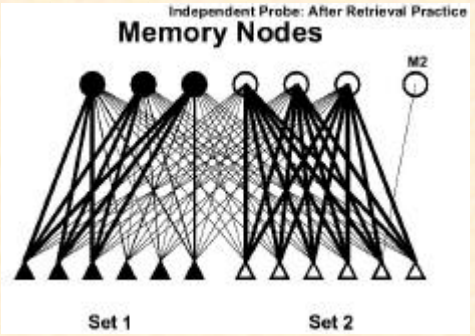
Independent probe effects



Independent probe effects



Independent probe effects



Explaining independent probe effects

- Model results
  - Standard retrieval induced forgetting occurs **AND**
  - During initial learning
    - Item is active, M2 weakly active ⇒ connection weakened
    - Other inputs to M2 inactive ⇒ no change in connection
  - During test phase M2 does not activate I2
    - Get effect of independent probe
- The effect is due to weakening of the item-independent probe during the initial learning phase

Testing independent probe effects

- Model results
  - The effect is due to weakening of the item-independent probe during the initial learning phase
- Utilise the decay of retrieval induced forgetting
  - No delay
    - RIF and independent probe effects
  - Delay between learning and practice/recall
    - RIF effect but no independent probe effect
  - Delay between learning/practice and recall
    - No RIF and no independent probe effect

Testing independent probe effects

- Group 1 - standard RIF and independent probe
- Group 2
  - 1 day delay between learning phase and practice/recall phases
  - Expect to see RIF effect (practice induces RIF)
  - Expect no independent probe effect (dissipates after learning)
- Group 3
  - 1 day delay between learning/practice phases and recall
  - Expect no RIF effect (dissipates after practice)
  - Expect no independent probe effect (dissipates after learning)

Testing independent probe effects

- Results
  - Ask me when the experiment has been run

Question

- Is retrieval induced forgetting a process related to memory but independent of memory formation?
- OR does retrieval induced forgetting offer insight into basic properties of memory formation and updating?
  - The modelling suggests this