1 Higher Level Cognition: What's Missing

- Planning, problem solving, reasoning, complex decision-making
- What do all of these have in common?
- Top-down control of behavior: Instead of reacting in a bottom-up fashion to stimuli, behavior is driven (controlled) by an *actively maintained representation* of what we are supposed to be doing...
- Allows us to behave in contextually appropriate fashion instead of just giving the strongest, most dominant response
- Also gives us the ability to link events across time points, and to carry out behaviors that are extended across time...

2 Higher Level Cognition: What We Know

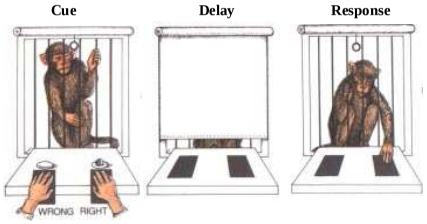
Frontal damage impairs planning, reasoning, decision-making, self-initiated actions, self-awareness, social interaction...

3 The Range of Frontal Functions

Activation-based working memory

4 Activation-based Working Memory

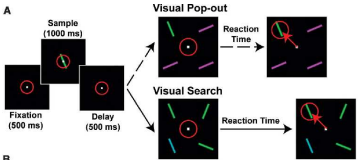
Monkey electrophysiology



The prefrontal cortex (PFC) can maintain information over time as firing of neurons (activation-based memory).

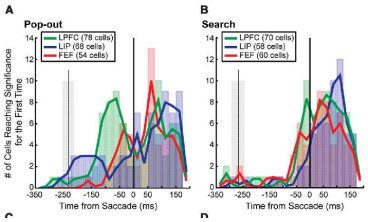
5 Top-down vs bottom-up PFC

Buschman & Miller, 2007, *Science*



6 Top-down vs bottom-up PFC

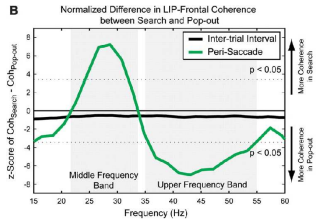
Buschman & Miller, 2007, *Science*



- parietal act for target location precedes pfc act for pop-out
- pfc act precedes parietal for search

7 Top-down vs bottom-up PFC

Buschman & Miller, 2007, *Science*



- greater low freq pfc-parietal synchronization for top-down
- greater high freq synchrony for bottom-up

8 The Range of Frontal Functions

Activation-based working memory

Inhibition Stroop: Difficulty inhibiting prepotent response.

9 The Stroop Task

Pink

Yellow

Green

Red

10 The Range of Frontal Functions

Activation-based working memory Monkey electrophysiology.

Inhibition Stroop: Difficulty inhibiting prepotent response.

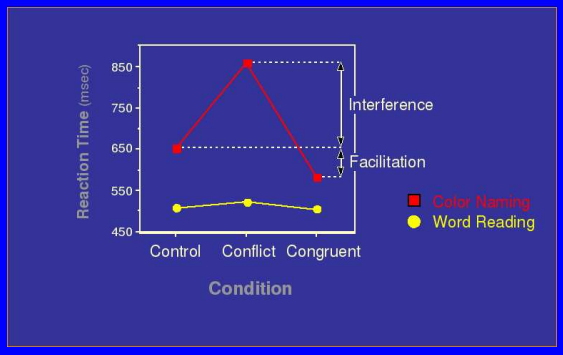
Flexibility Continue with same response after task changes, *perseveration*.

Fluency Difficulty generating variety of responses.

Executive control Probs w/ goal-directed planning, coordinating. (shopping)

Monitoring/evaluation e.g., Error-monitoring.

11 The Stroop Effect



12 Stroop Effect: *GREEN*

Possible explanation: **differential pathway strength**

- two pathways: word reading and color naming
- These **compete** to generate response
- Word reading pathway is much stronger than color naming
- When word identity information doesn't match color, it **interferes strongly** with color naming
- Because color pathway is relatively weak, incongruent color info does not interfere with word reading

13

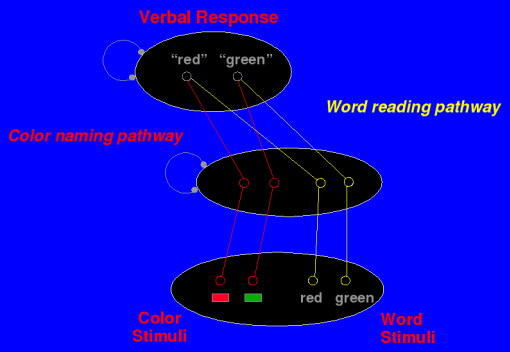
Stroop Effect: *GREEN*

- Puzzle: If the color naming pathway is weaker than word reading, how do we manage to name color of the word "green" above?
- Solution: Prefrontal cortex *actively maintains* a representation of the task that you are supposed to be doing (color naming or word reading)
- This actively maintained task representation **biases processing** in posterior cortex by activating units in appropriate pathway
- e.g.,color naming task rep in PFC sends activation to the units in color naming pathway...

14

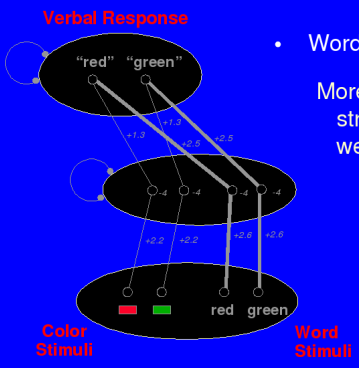
Model of the Stroop Task

(Cohen, Dunbar & McClelland (1990)



15

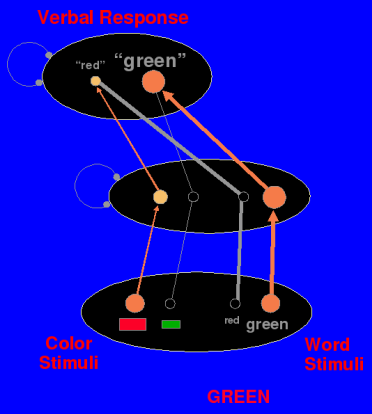
Effects of Training



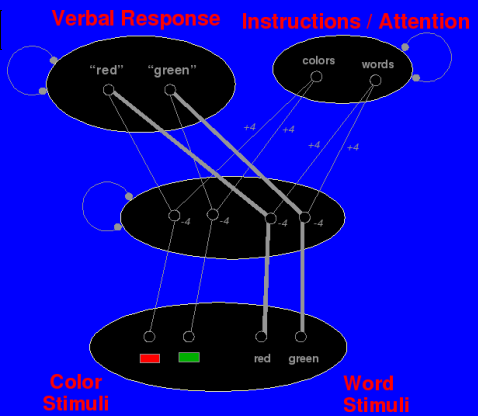
- Word pathway:

More training = stronger weights

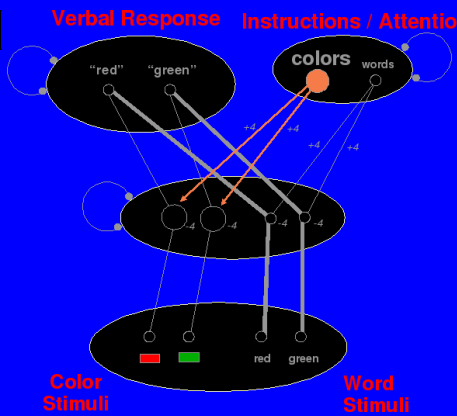
16



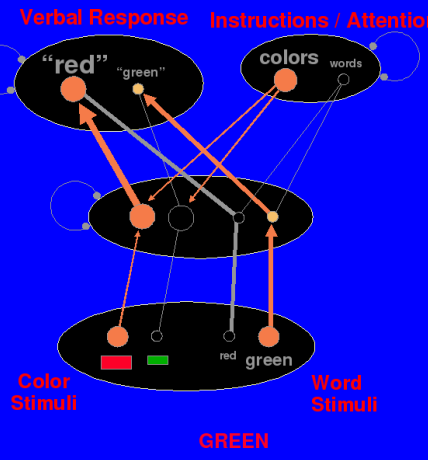
17



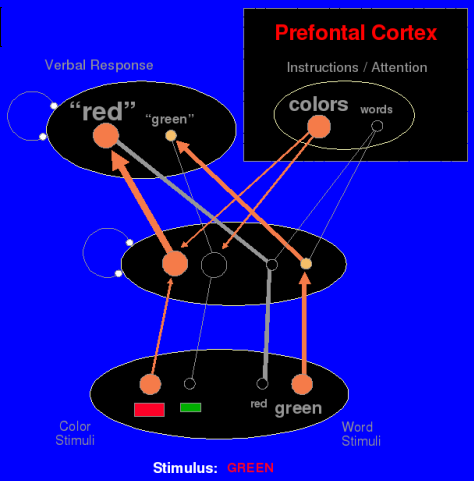
18



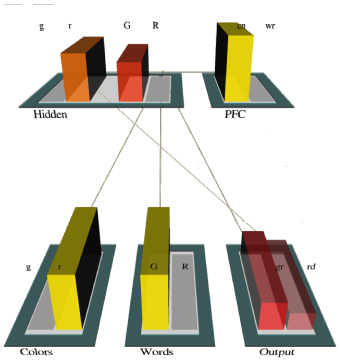
19



20

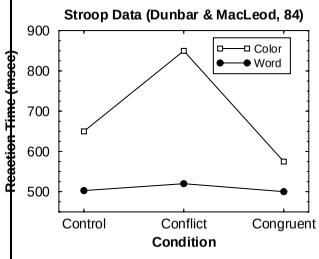


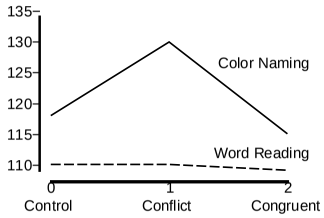
21 Stroop Model



22 The Stroop Task: Model Data

Control: Red. Conflict: Red. Congruent: Red.





23 Pathway Strength vs Processing Speed Theories

- Model predicts key difference between word reading and color naming is **pathway strength** (reading > color naming). This results in:
  - Word reading being faster than color naming
  - Asymmetric interference effects
- Other (verbal) theories posit Stroop effects resulting from a **horse race** rather than direct **competition**

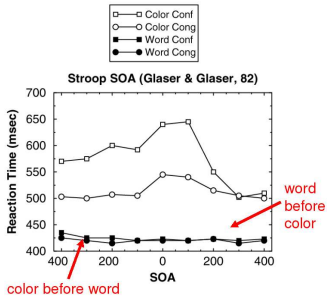
24 Horse Race Theories

- Color does not affect word reading because the word reading process runs to completion before color is processed
- Conversely, word identity **does** affect color naming because word reading process completes before color response is generated
- This theory, stated as such, implies that it should be possible to get color to interfere with word info if the color naming process has a *head start*

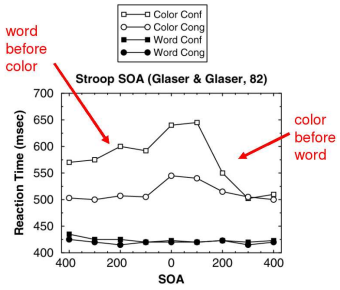
25 Stroop Accounts: Not a Horse Race

⇒ *GREEN*

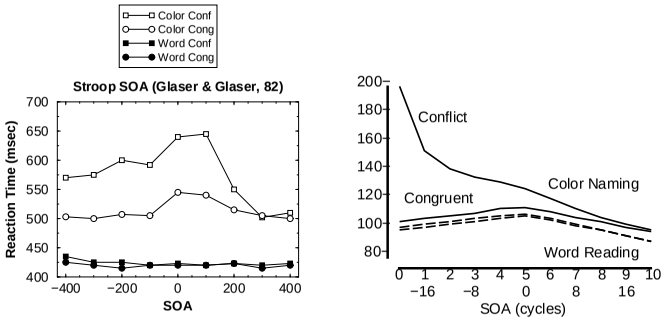
26 Stroop Accounts: Not a Horse Race



27 Stroop Accounts: Not a Horse Race



28 Stroop Accounts: Not a Horse Race



29 Stroop Accounts: Automaticity

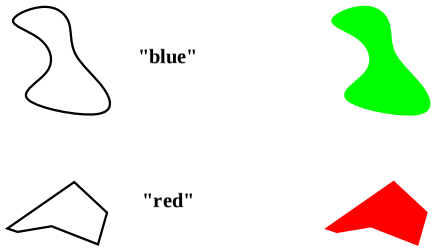
- Early accounts of Stroop focused on automatic vs controlled processing
- According to these theories, word reading is **automatic** and color naming is a **controlled** process.
- Automatic processes don't suffer from interference (they proceed "automatically") but controlled processes do.

30 Stroop Accounts: Automaticity

- Status (whether a process is controlled or automatic) depends on **absolute** pathway strength
- In contrast, model focuses on **relative** pathway strength - stronger pathway interferes with less strong pathway (but not vice-versa)
- *Prediction*: If we could come up with a task that is **even less** well-learned than color naming, we will find that:
- New task will have no effect on color naming
- Color naming will interfere with new task

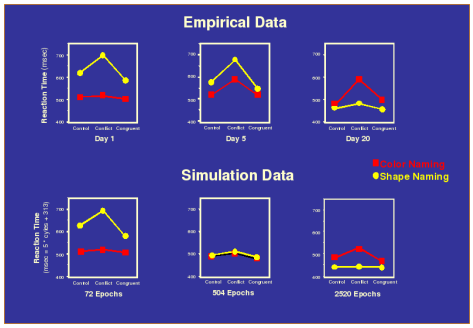
- As new task is practiced repeatedly, effects should reverse

31 Stroop Accounts: Continuum, not a Dichotomy



Shape-naming first like color-naming in standard Stroop, then like-word reading.

32 Stroop Accounts: Continuum, not a Dichotomy



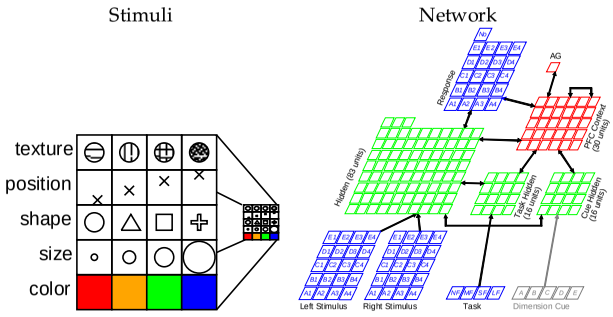
MacLeod & Dunbar, 1988

33 But how do PFC units come to represent task rules??

- Stroop model is a nice simple account of PFC function, but it somehow assumes that PFC 'knows' how to maintain a rule for color naming and to magically bias color-naming hidden units
- Interesting question is how these rule-like representations develop in the first place?
- Can PFC *learn* to assign *abstract* rule-like representations that code for *stimulus dimensions* (e.g., color) by experience (with multiple colors)??

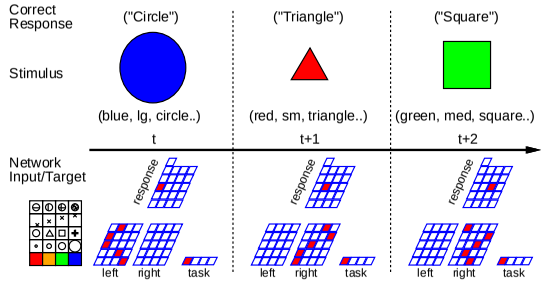
34 PFC Specializations → Rule-Like Abstract Reps

(Rougier, Noelle, Braver, Cohen & O'Reilly, *PNAS*)



35 Developing PFC Reps

(Rougier, Noelle, Braver, Cohen & O'Reilly, *PNAS*)



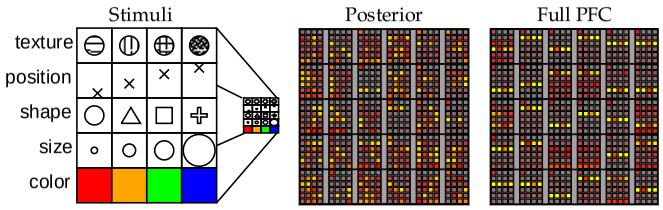
Task 1 = Name the Shape (or color, etc)
Task2: Do two stimuli match along some dimension? (yes/no)
Task3: Which object is larger? etc.

Key: Do repeated trials of same task – continuous attention to shapes, etc

36 PFC Specializations → Rule-Like Abstract Reps

(Rougier, Noelle, Braver, Cohen & O'Reilly, *PNAS*)

Weights from PFC or Hidden to output response units

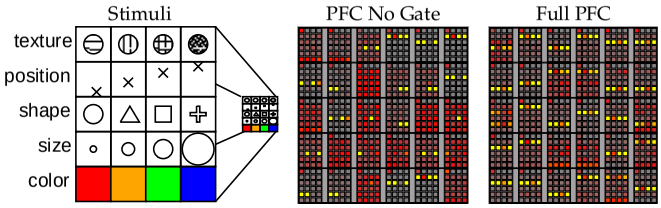


Rule = One stimulus dimension (row) relevant at a time. (e.g., card-sorting tasks)

Abstraction derives from sustained maintenance over trials! Posterior net 'memorizes' specific combinations of features/responses for each task, doesn't develop systematic representations

37 PFC Specializations → Rule-Like Abstract Reps

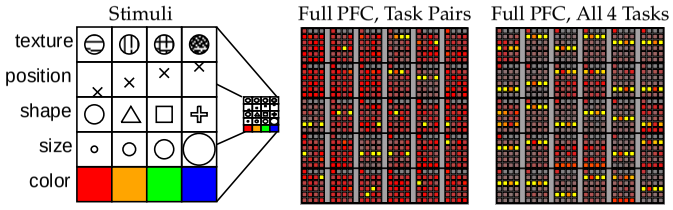
(Rougier, Noelle, Braver, Cohen & O'Reilly, *PNAS*)



Adaptive gating is key:
within block of trials feature changes but gating mech learns to maintain constant PFC rep (until rule switches, performance goes down → update)

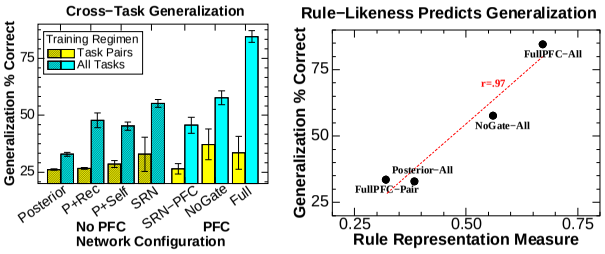
38 PFC Specializations → Rule-Like Abstract Reps

(Rougier, Noelle, Braver, Cohen & O'Reilly, *PNAS*)



As is breadth of experience (same stimuli across different tasks) (increasing pressure to use same pfc reps across tasks → systematicity); with small # tasks can get by with memorizing)

39 Rule-Like Abstract Reps → Generalization

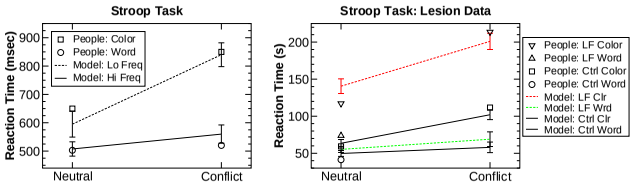


Abstraction → better generalization across tasks
(accuracy on stims not seen in particular task).

Interaction of nature (PFC mechanisms) and nurture (breadth of experience).

40 Stroop Performance

(Rougier et al, *PNAS*)



Same network & parameters: PFC control representations developed entirely through learning from random initial weights!

LF = left frontal (DLPFC) lesions in people and model (posttraining, 30% damage)

41 The Range of Frontal Functions

Activation-based working memory Monkey electrophysiology.

Inhibition Stroop: Difficulty inhibiting prepotent response.

Flexibility Continue with same response after task changes, *perseveration*.

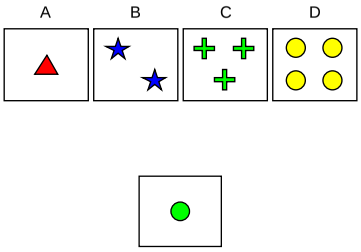
Fluency Difficulty generating variety of responses.

Executive control Probs w/ goal-directed planning, coordinating.

Monitoring/evaluation e.g., Error-monitoring.

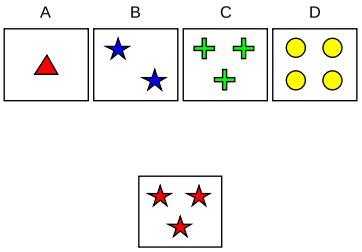
42 Dynamic Categorization Tasks

Wisconsin Card Sort



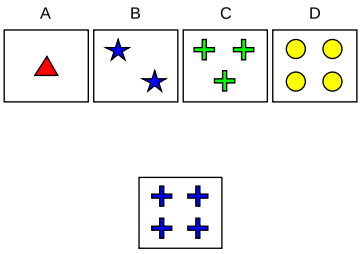
43 Dynamic Categorization Tasks

Wisconsin Card Sort



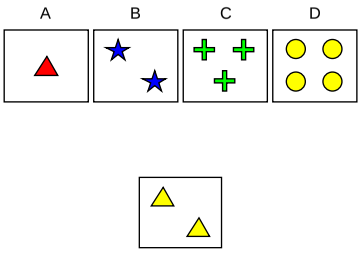
44 Dynamic Categorization Tasks

Wisconsin Card Sort



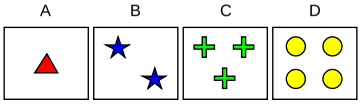
45 Dynamic Categorization Tasks

Wisconsin Card Sort



46 Dynamic Categorization Tasks

Wisconsin Card Sort



Experimental task (like Stroop), but captures some essential aspects of higher level cognition.

Frontal patients perseverate with the first rule.
→ weight-based tendencies build up when categorizing according to first rule, and you need to *actively maintain* the new rule to counteract these weight-based tendencies

47 "Frontal Tasks"

- Stroop: Ability to override prepotent response (word reading) in favor of currently relevant task (color naming) – requires top-down control.
- Activation based directing of attention.
- "Prefrontal control" not just for overriding long term associations like word reading, but also for the ability to quickly change attention in an online fashion in response to changing task demands: UPDATING.

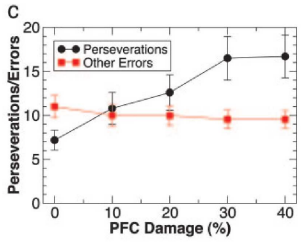
48 Revenge of the Donuts...

49 Revenge of the Donuts...

Two strategies for solving donut categorization task:

- Adjust weights to different donut types
- Actively maintain a representation of your current strategy; deactivate this rep and activate another if you get negative feedback
- Active maintenance does not strongly benefit initial learning of the rule
- However it does greatly facilitate performance when the rule switches

50 WCST in PFC model (Rougier et al)



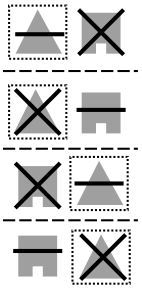
51 Card Sorting Tasks

- Relevant to everyday life, or just to this peculiar task?
- Good measure of online thinking & problem solving: The ability to flexibly consider different possibilities to guide thinking and behavior.

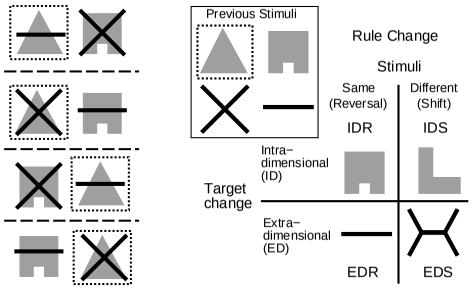
52 Card Sorting Tasks

- In what situations do we need to to consider/represent different rules in mind and have the ability to flexibly update/maintain them until one works well?
- Right now! Thinking. I'm asking you a question, you consider an alternative (e.g., "Never: card sorting tasks are dumb").
- You then evaluate the quality of what you're holding in mind: does it make sense, is it likely to produce a good outcome?
- If yes, maintain info further processing; if not, update.
- Science: hypothesis formulation from experimental data.

53 Dynamic Categorization Tasks: ID/ED task

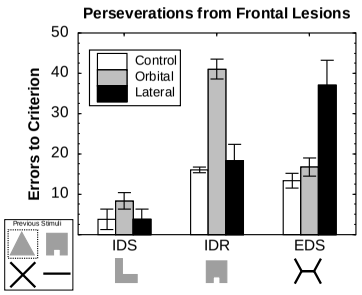


54 Dynamic Categorization Tasks: ID/ED task



55 ID/ED and Frontal Damage

(Dias, Robbins & Roberts (1997), *J Neurosci*)



Original interpretation: Orbital = affective inhibition,
Lateral = attentional selection.

56 Alternative Account

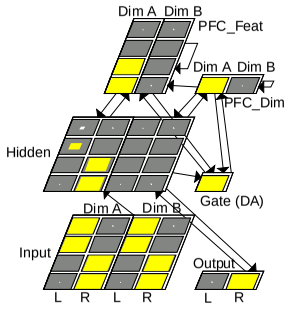
(O'Reilly, Noelle, Braver & Cohen (2002), *Cerebral Cortex*)

Orbital PFC represents detailed features.
Lateral PFC represents abstract dimensions.

Activation-based PFC processing facilitates rule switch:
Orbital = switch to new features (IDR).
Lateral = switch to new dimension (EDS).

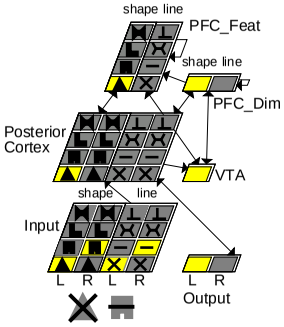
Perseverations = weight-based processing in absence of PFC.

57 ID/ED Model

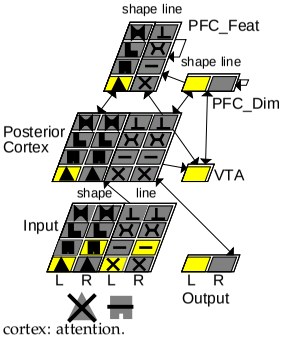


Two dimensions, A and B (shapes & lines)
On each trial, four stims are presented:
Dim A left, Dim A right, Dim B left, Dim B right...

58 ID/ED Model



59 ID/ED Model



Activation limited in cortex: attention.
PFC provides top-down bias, with DA/updating unit.
PFC_Feat = stim features indep of location
PFC_Dim = abstract dimensions

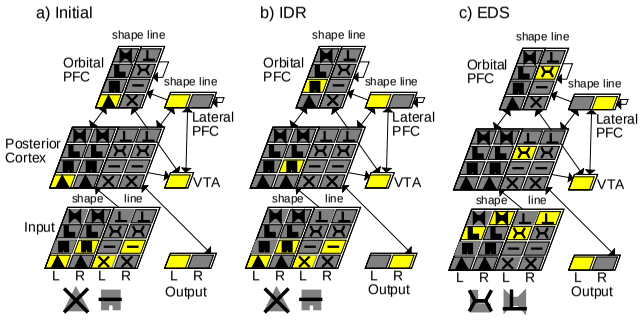
60 PFC updating based on unexpected rewards and errors

- When there is an increase in DA activity (e.g. the model got the answer right but wasn't expecting a reward):
  - hidden unit activity is gated into PFC
  - connections from hidden units to DA are increased
  - PFC serves to *amplify the influence of hidden units associated with correct responding*
- When there is a decrease in DA activity (the model was expecting a reward but gave the wrong response):
  - PFC activity is wiped clean
  - connections from hidden units to DA are decreased
- Also, there is some "gating noise": trial and error search

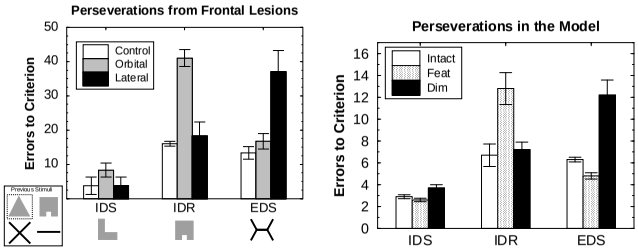
61 Similarities/Differences with Store-Ignore-Recall

- With S-I-R, the model had to gate the "Store" stimulus into PFC (and carry it forward in time) in order to respond correctly; S-I-R can only be solved with the help of active maintenance (working memory)
- The ID/ED task **can** be solved without active maintenance; but PFC can **help** by focusing the model's attention on useful parts of the input but it isn't **necessary**..

62 IDR, EDS in the Model



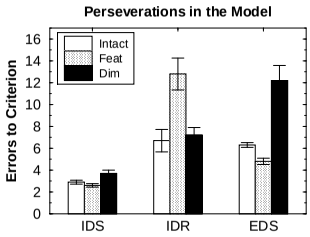
63 Model Data



64 Explanation of Lesion Data: IDS

- Intradimensional shift (IDS): different stimuli pre and post-shift; the relevant dimension (A) stays the same
- No effect of PFC lesions
- PFC is unnecessary because there are no strong, inappropriate tendencies to overcome (new stimuli)

65 Model Data



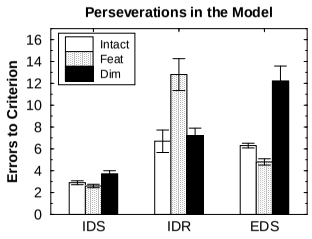
66 Explanation of Lesion Data: IDR

- Intradimensional reversal (IDR): same stimuli pre and post-shift; initially A1 = target; after the shift A2 = target
- Performance is impaired after PFC_Feat lesions but not PFC_Dim lesions
- It's clear why PFC_Dim is *not* important here: It involves a shift of attention *within* a dimension, not across dimensions..
- How does PFC_Feat help performance?

67 Explanation of Lesion Data: IDR

- Before shift, some hidden units learn to generate the A1 response
- After shift, these hidden units point to the *wrong* response
- PFC helps the model focus on other hidden units, which can then be associated with the new response
- This way the model avoids having to fully *unlearn* the association between the original hidden units and A1 response

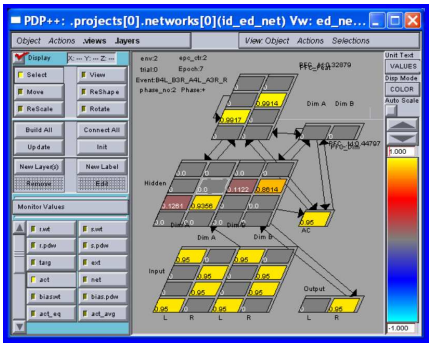
68 Model Data



69 Explanation of Lesion Data: EDS

- Extradimensional shift (EDS): different stimuli pre- and post-shift; initially A1 = target; after the shift, B3 = target
- Performance is impaired after PFC_Dim lesions but not PFC_Feat lesions
- It's clear why PFC_Dim is important: It helps focus attention on the newly relevant dimension
- Why can't PFC_Feat serve the same function? should be able to bias new reps as before...
- *Without PFC_Dim, PFC_Feat has no sense of what constitutes a "dimension", just updates to random new pattern of features from both A and B dimensions...*

70 After EDS: PFC_Dim lesion



71 Advantages of ID/ED model

- PFC reps are not clamped as in Stroop – updated in response to changing task demands.
- Nice fit and explanation of complex monkey data.
- Shows how working memory and cognitive control may be two sides of the same coin: activation-based memory is not just memory but also biases activity elsewhere in the brain.
- Shows that ID/ED data can be explained in terms of lateral and orbital PFC carrying out the same *function* (biasing competition in posterior cortex), applied to different kinds of *content* (features vs dimensions)
- Also provides evidence for a hierarchy of PFC representations

72 Limitations of ID/ED model

- Reps not clamped, but still not learned – one to one connectivity from HL.
- Distinction between OFC = features, DLPFC = dimensions may be too convenient: observed dissociation; not much evidence of OFC-features (see Frank & Claus, 2006).
- Newer models address the issue of how PFC representations can develop in childhood and lead to higher level *abstraction* and generalization to new tasks (Rougier et al, 2005, *PNAS*)

73 Limitations of ID/ED model

- Doesn't distinguish b/w updating and maintenance systems.
- Goal/Subgoal requires *selective* updating with concurrent maintenance of task relevant info.
- Updating system thought to involve the BG and DA, damaged in PD,SZ and lead to "frontal-like" impairments in Stroop, WCST, etc.
- Newer BG models address these issues in more complex tasks (eg. O'Reilly & Frank, 2006)

74 Goal/Subgoal Hierarchical Structure

1. Open fridge.
2. Get food items.
3. Close fridge.
4. Get bread from cupboard

Update these subgoals to guide actions, but to guide the ordering of subgoals themselves, need to maintain overall goal of task (Make sandwich)

75 A Unified Activation-based Account

Central frontal mechanisms:

Activation-based working memory Frontal neurons maintain actively over delays.

Monitoring/evaluation e.g., Error-monitoring, critical for dopaminergic modulation.

76 A Unified Activation-based Account

Inhibition Need to maintain top-down activation for weaker task.

Flexibility Dynamics of activation-based more rapid than weight-based.

Fluency Only problem w/ novel categories of responses — need top-down support to overcome prepotent categories.

Executive control Maintain & update plans / goals over time, avoid distraction.

77 Higher Level Cognition: What's Missing

- Planning
- Reasoning
- Decision-making
- Emotion
- Consciousness, sense of self
- Free will
- Social interaction

Bias & Binding in the PFC and Hippocampus: