Induction, confirmation and choice

Rushworth lab meeting, 28/5/2012

Questions

1/ humans and other primates can exert task-level control over behaviour, i.e. use rules
But how do we create new task sets?

2/ when learning propositional information, humans are subject to **inductive biases**

among the best-known of these is the **confirmation bias**, whereby agents tend to overweight prior hypotheses in decision-making

Questions

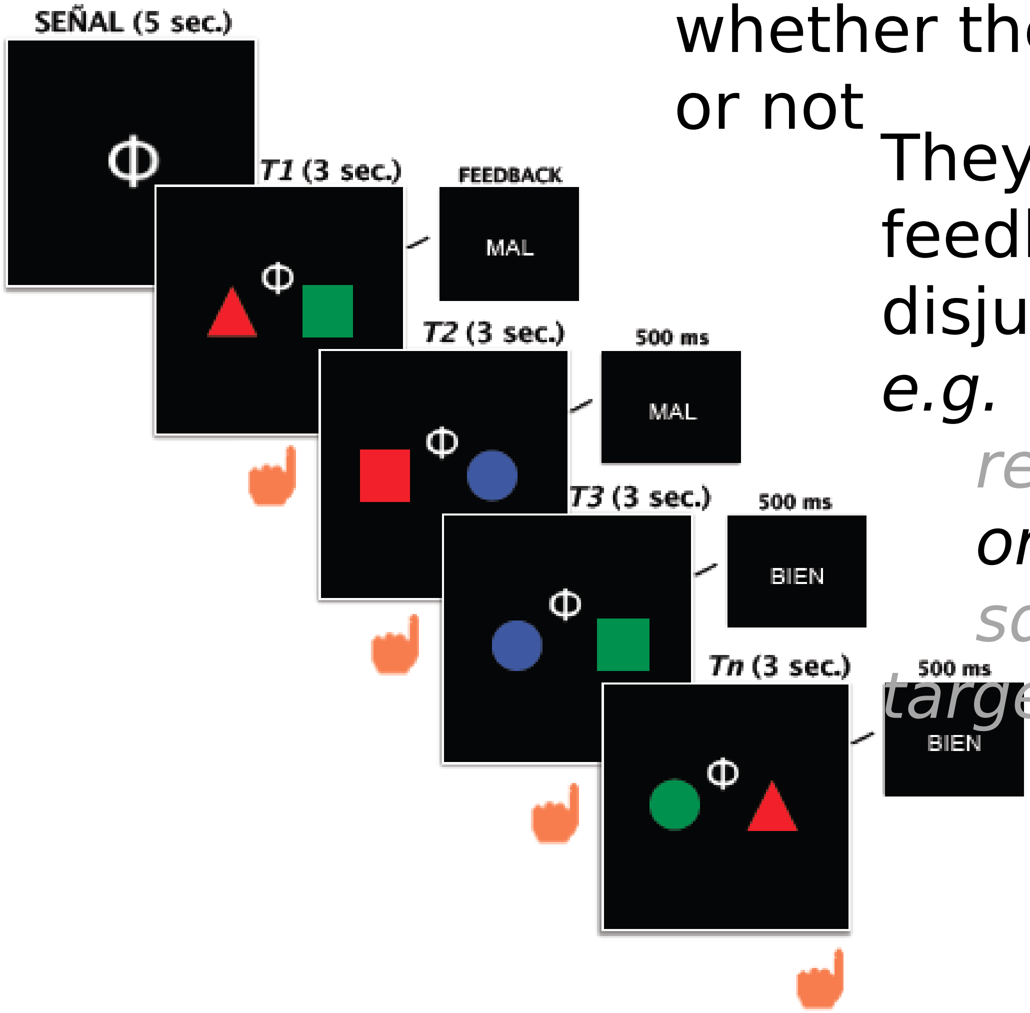
The confirmation bias is often thought of as a bias to **seek** confirmatory evidence, e.g. Wason card selection task

However, humans may also
    fail to learn from disconfirmatory evidence
    fail to use this information to rule out incorrect hypotheses

Task

Participants viewed coloured shapes left and right of the screen in blocks of 16 trials

On each trial they had to decide whether the stimulus was a target or not



They received deterministic feedback according to a disjunctive rule

e.g.

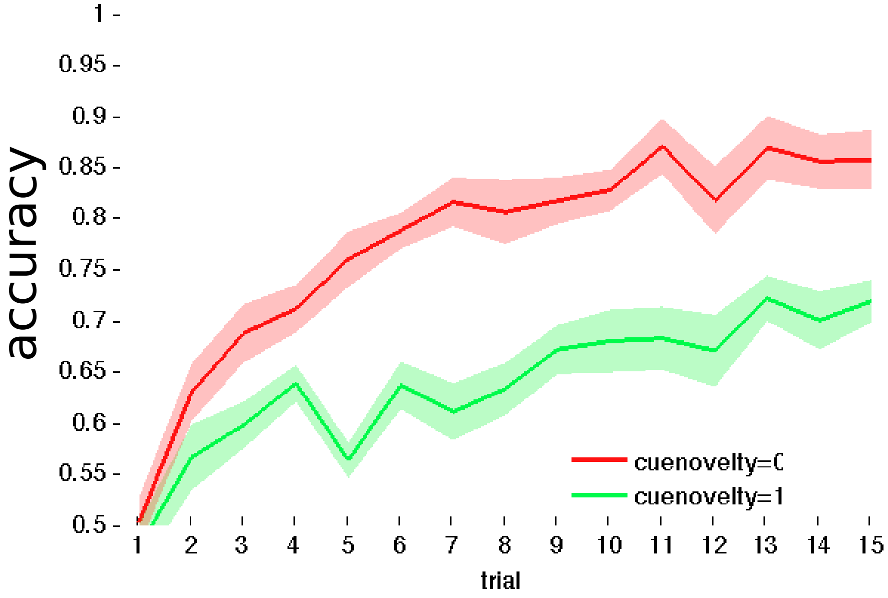
red left | blue right = target

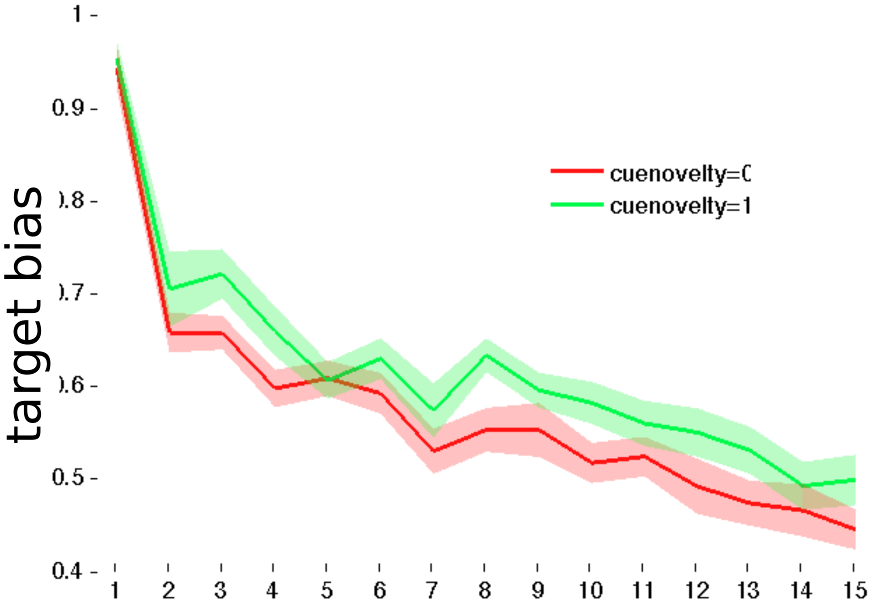
or

square left | green right = target

Symbols instructed the relevant dimension (e.g. shape left/colour right) but not the precise

Behaviour





Participants learned across the block

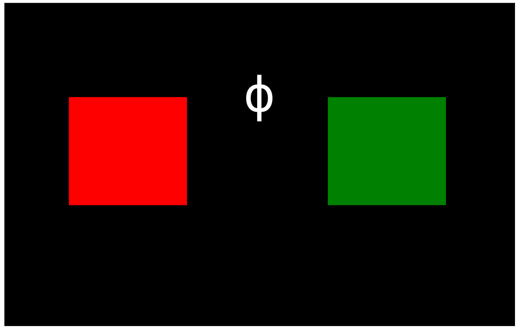
They learned better when the cue was informative

Participants' tendency to respond 'target' began high and declined with time

Models and mechanisms…

ets consider the simple case where you know the dimensions, eg colour-colour
/e define prior belief in any sub-rule as α

	RED	GREEN	BLUE
LEFT	α	α	α
RIGHT	α	α	α



‘target’

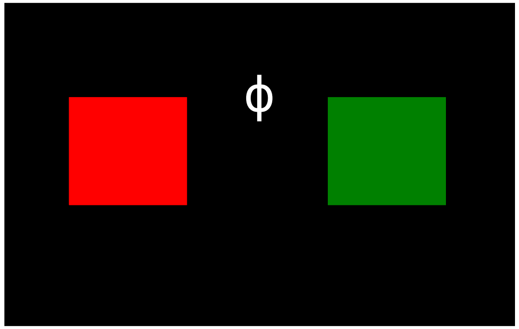
incorrect

	RED	GREEN	BLUE
LEFT	-Inf	α	α
RIGHT	α	-Inf	α

Models and mechanisms…

ets consider the simple case where you know the dimensions, eg colour-colour
/e define prior belief in any sub-rule as α

	RED	GREEN	BLUE
LEFT	α	α	α
RIGHT	α	α	α



‘target’

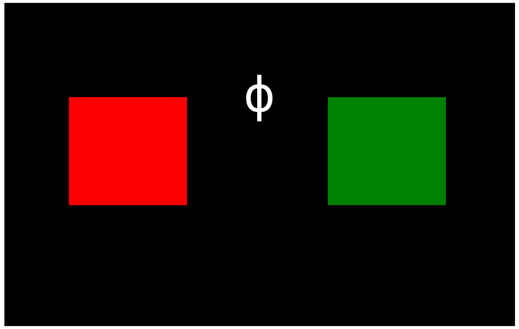
correct

	RED	GREEN	BLUE
LEFT	α	α	α
RIGHT	α	α	α

Models and mechanisms…

ets consider the simple case where you know the dimensions, eg colour-colour
/e define prior belief in any sub-rule as α
nd the increased belief from confirmation as δ

	RED	GREEN	BLUE
LEFT	α	α	α
RIGHT	α	α	α



‘target’ ☝

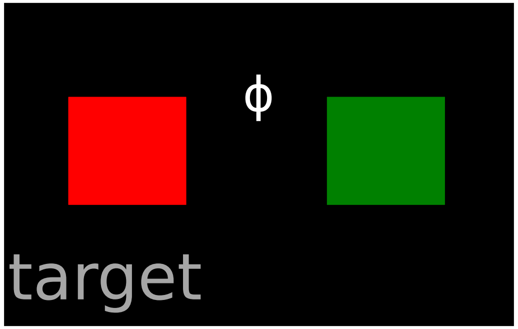
correct 🔊

	RED	GREEN	BLUE
LEFT	α+δ	α	α
RIGHT	α	α+δ	α

3 models

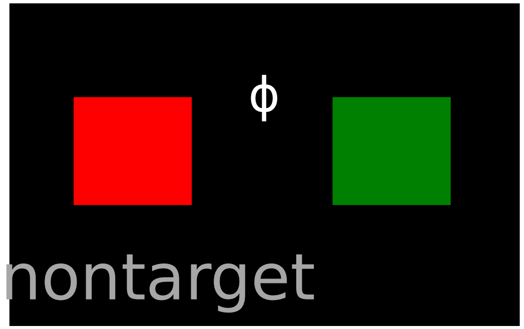
confirmation model

	RED	GREEN	BLUE
LEFT	+1	0	0
RIGHT	0	+1	0



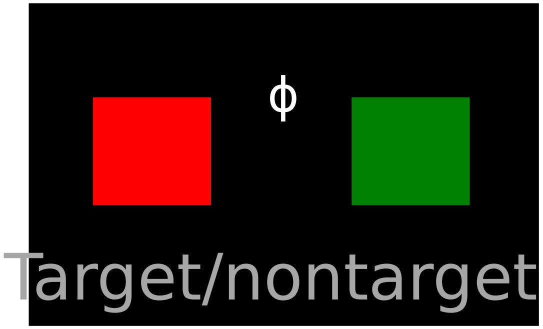
disonfirmation model

	RED	GREEN	BLUE
LEFT	-1	0	0
RIGHT	0	-1	0

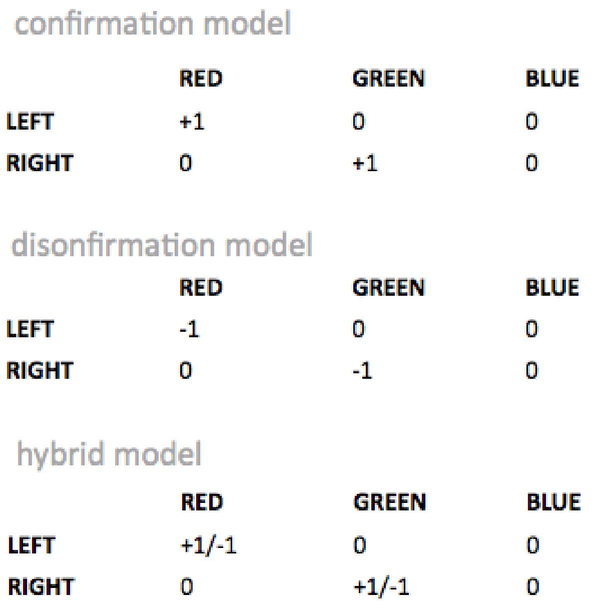


hybrid model

	RED	GREEN	BLUE
LEFT	+1/-1	0	0
RIGHT	0	+1/-1	0



Further assumptions



confirmation model

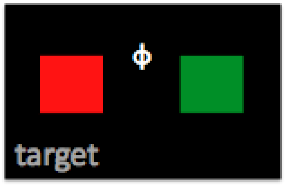
	RED	GREEN	BLUE
LEFT	+1	0	0
RIGHT	0	+1	0

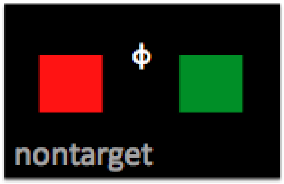
disonfirmation model

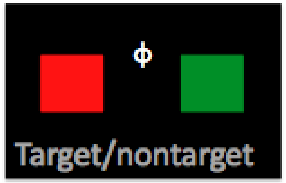
	RED	GREEN	BLUE
LEFT	-1	0	0
RIGHT	0	-1	0

hybrid model

	RED	GREEN	BLUE
LEFT	+1/-1	0	0
RIGHT	0	+1/-1	0

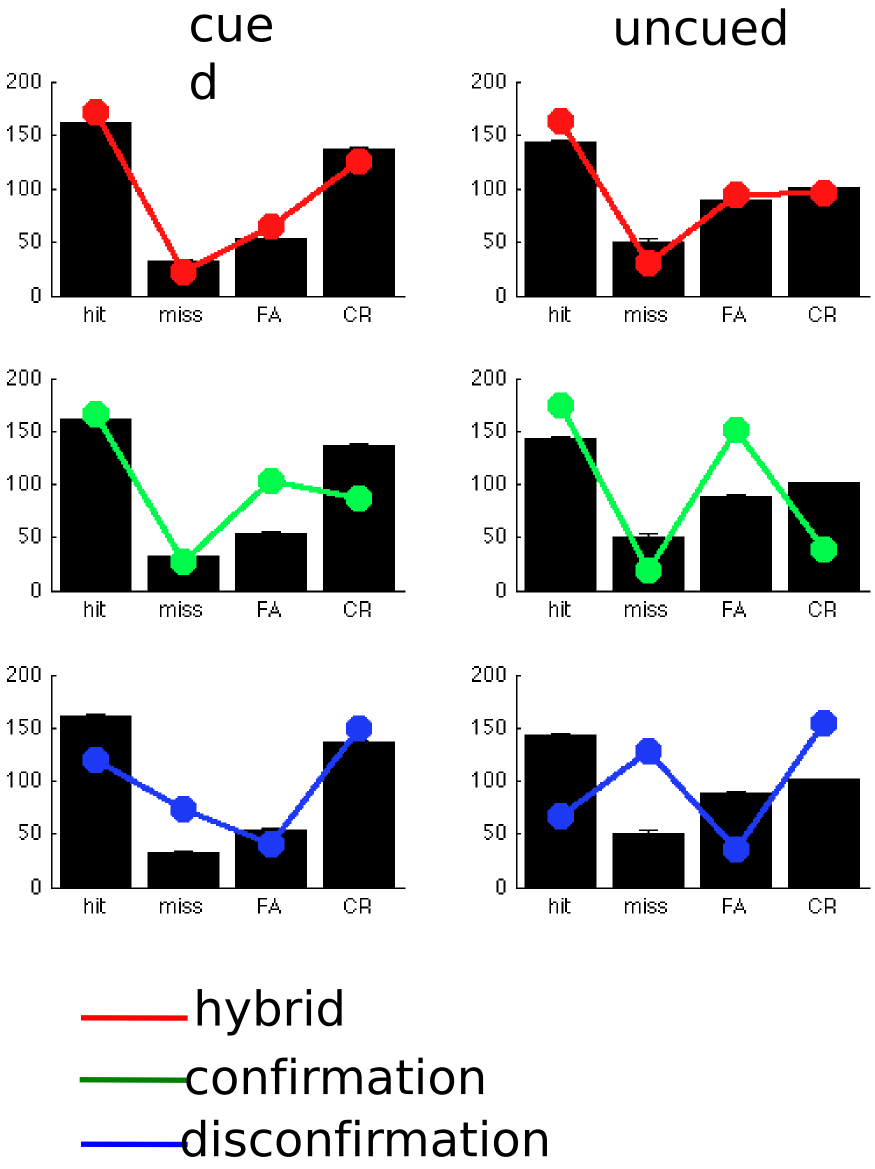






1/ if you have no evidence either way, then assume target
2/ no free parameters – adding leak/bias does not improve fits
3/ hardmax choice rule on expected value:
    confirmation/disconfirmation: defined as abs MAX of two values

Model fits



cue
d

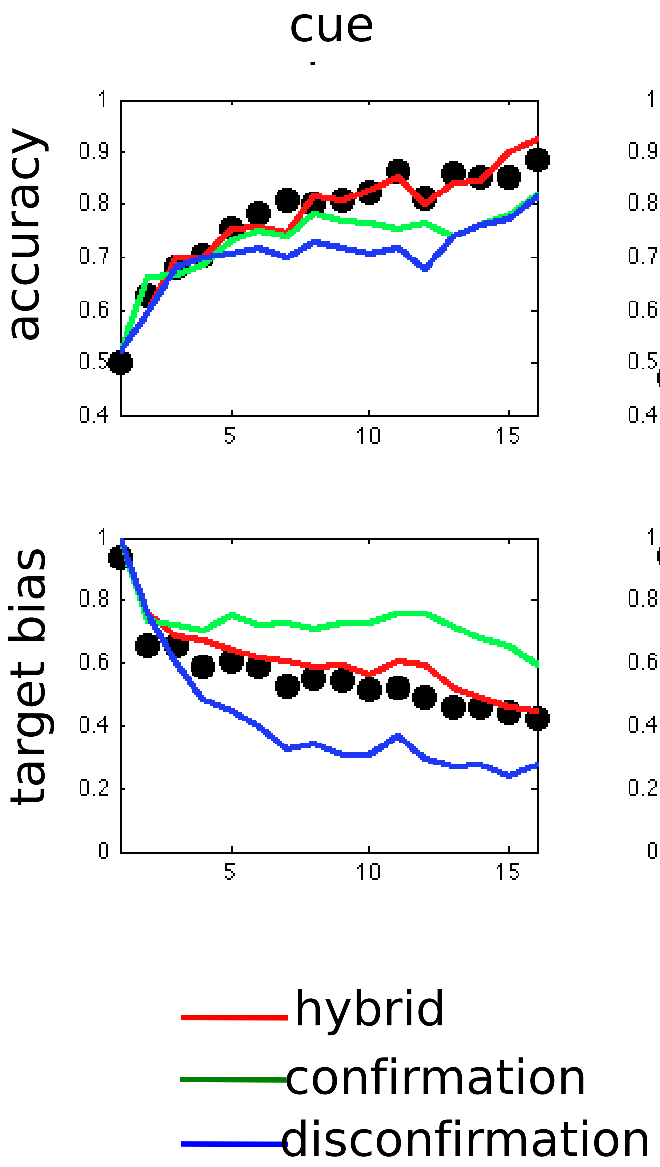
uncued

hybrid

confirmation

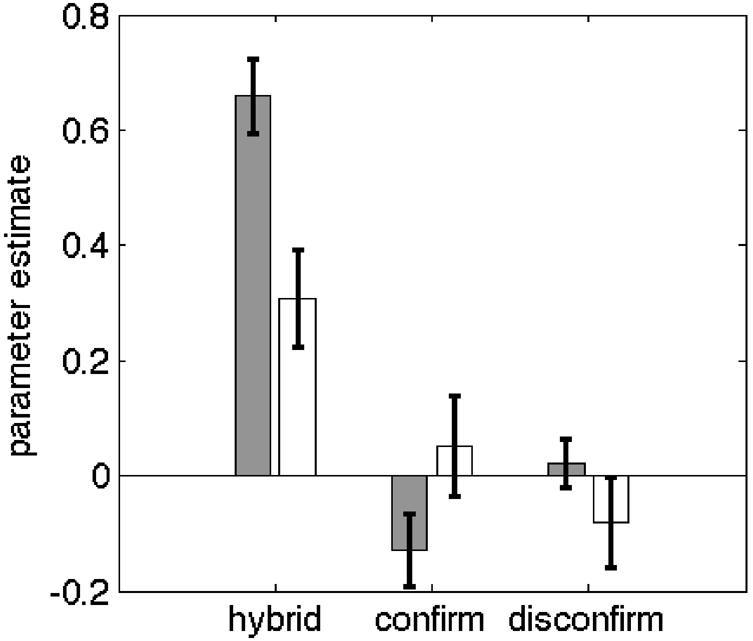
disconfirmation

Model fits

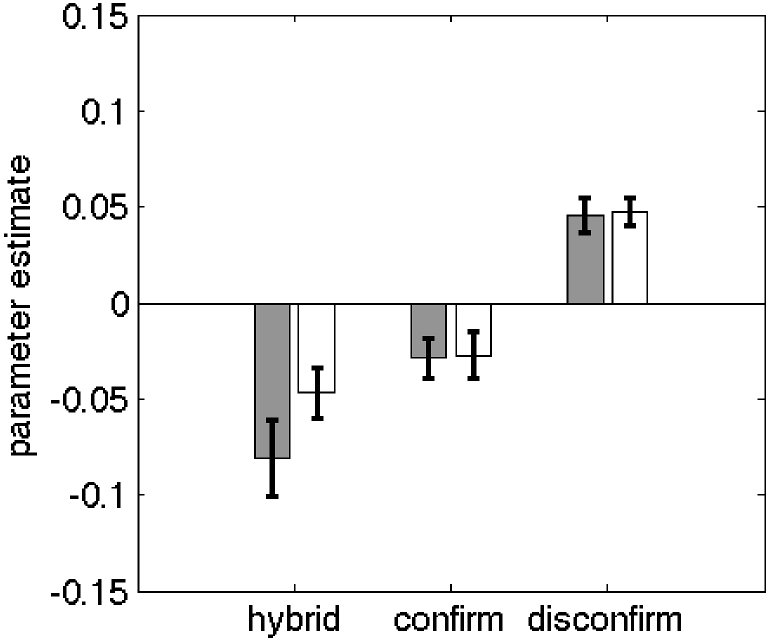

cue
uncued
accuracy
target bias
1
0.9
0.8
0.7
0.6
0.5
0.4
5
10
15
1
0.8
0.6
0.4
0.2
0
hybrid
confirmation
disconfirmation

3 models

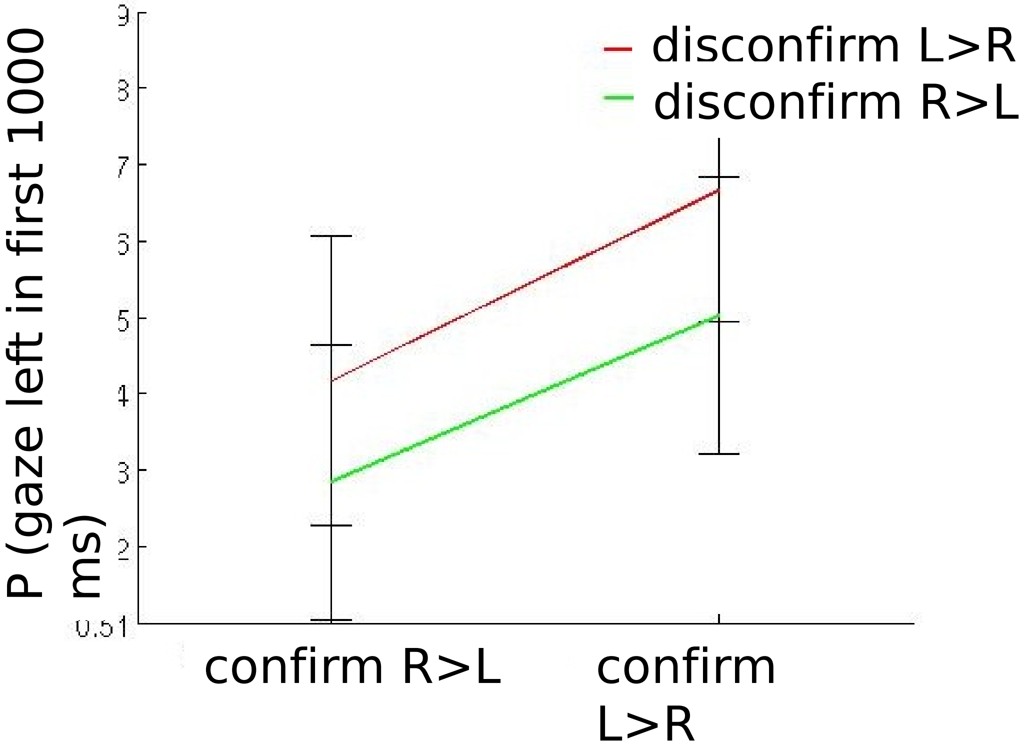
predicting choice



predicting RT



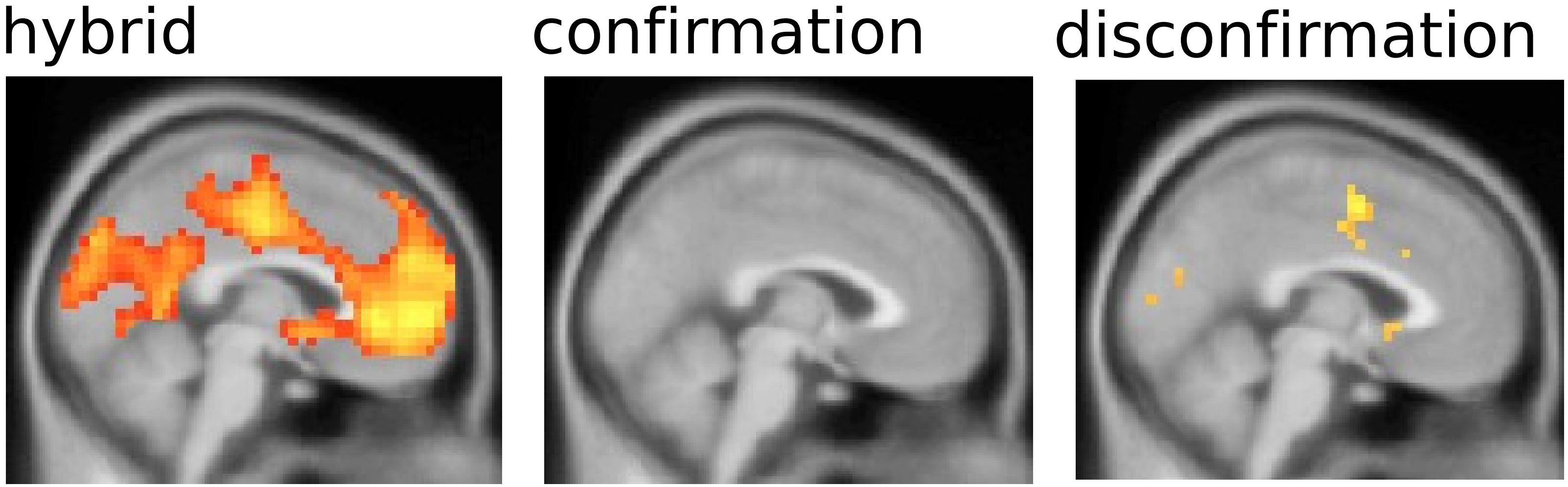
Sampling bias? Eyetracking data..



Eye tracking….is participants’ gaze predicted by the strength of confirmatory or disconfirmatory evidence?

Seems to be both (both main effects significant)

3 models: neural data

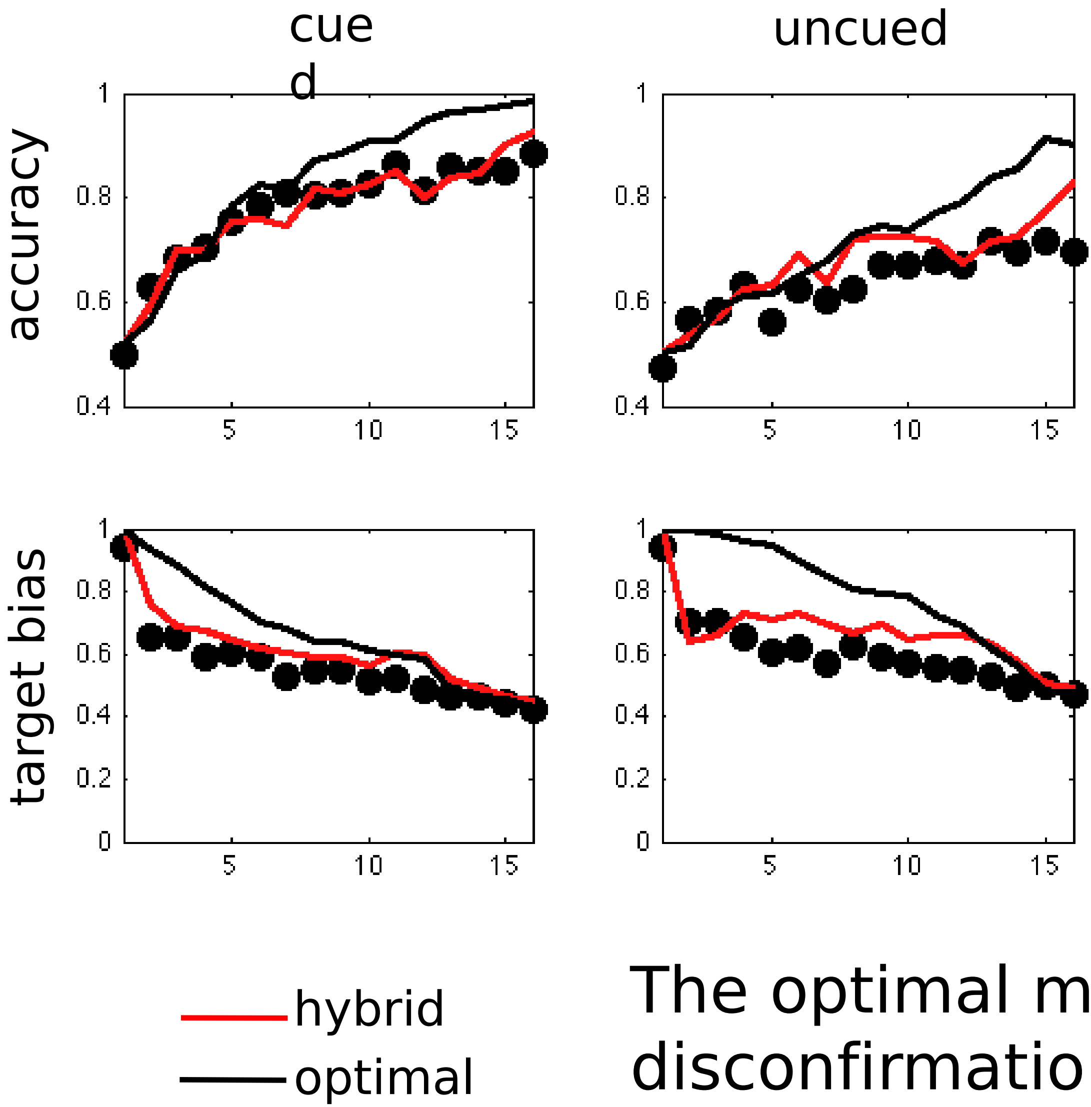


elates of expected value under the three models

Confirms behavioural data in strongly supporting hybrid model

OK so let’s do something more interesting…

Hybrid model is suboptimal



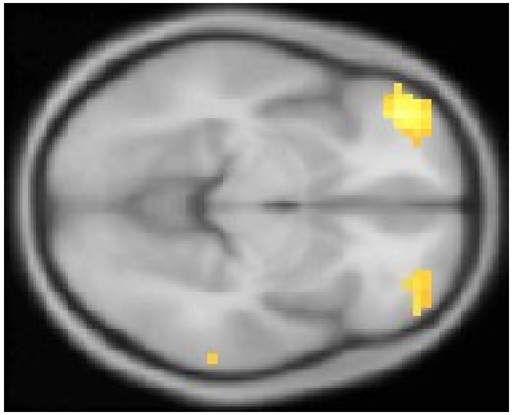
The optimal model is like the disconfirmation model, except that options are ruled out definitively

Confirmation and disconfirmation: choice

Both evidence for confirmation and disconfirmation make a contribution to expected value, so we can look for independent correlates of v[max(confirm)] and v[max(disconfirm)] at the time of choice

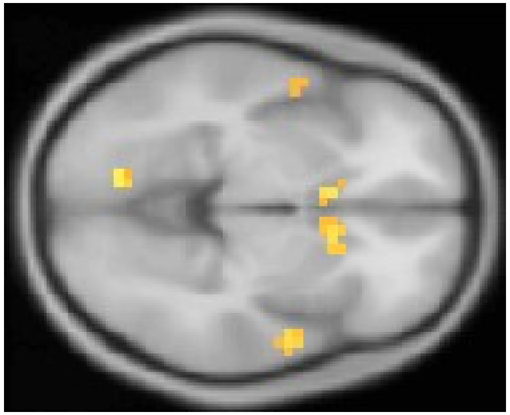
value of confirmation:

lateral OFC



value of disconfirmation

striatum

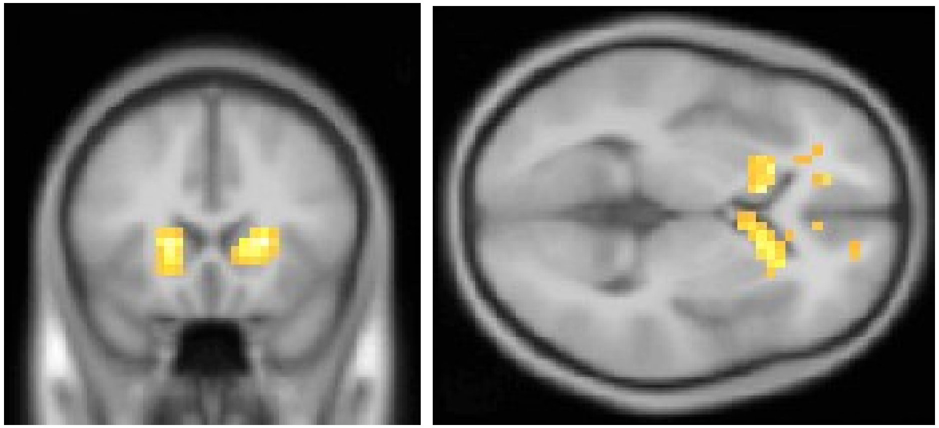


< 0.001, entered alongside overall expected value signal, feedback, target, etc

Confirmation and disconfirmation: learning

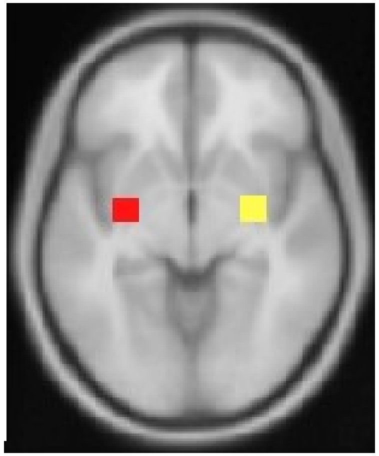
We can also measure prediction error signals for confirmatory and disconfirmatory learning

For example, confirmation bias might be due to larger prediction error signals for confirmatory than disconfirmatory learning



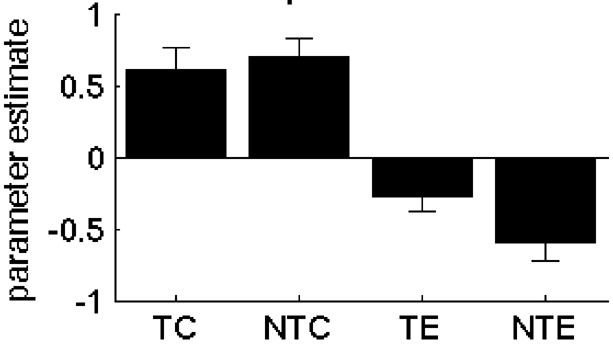
nteraction between prediction error and target/nontarget, p < 0.0001

Hybrid model
ROIs in the basal ganglia - putamen

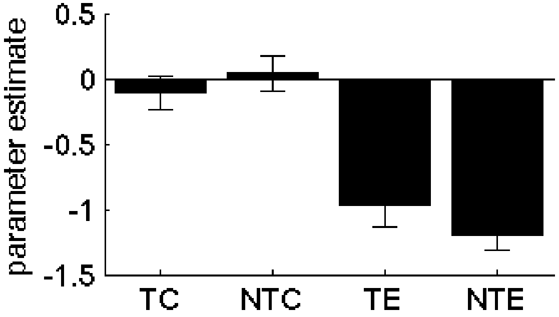


Response to feedback (main effect)

R putamen

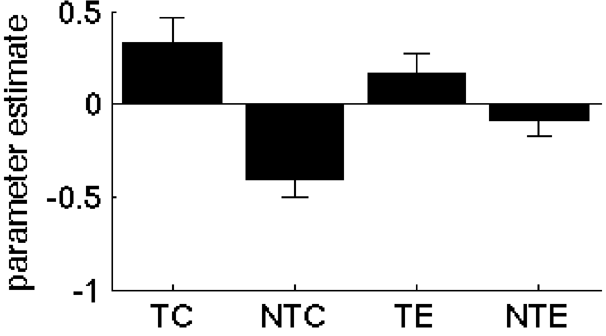


L putamen

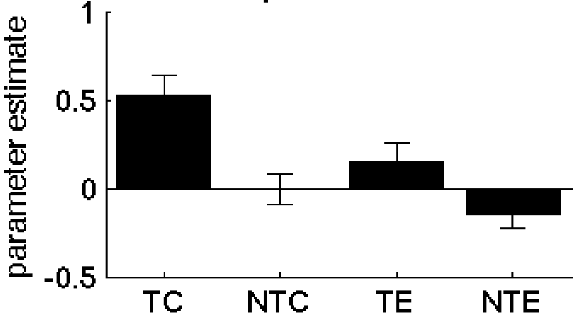


Response to prediction error

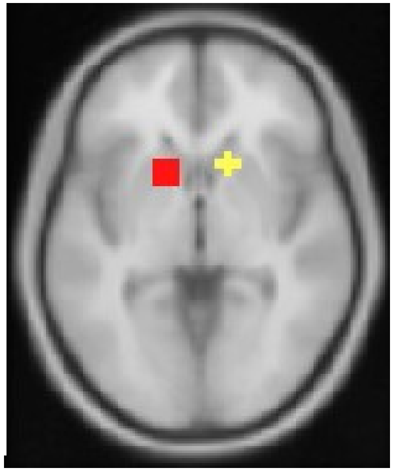
R putamen



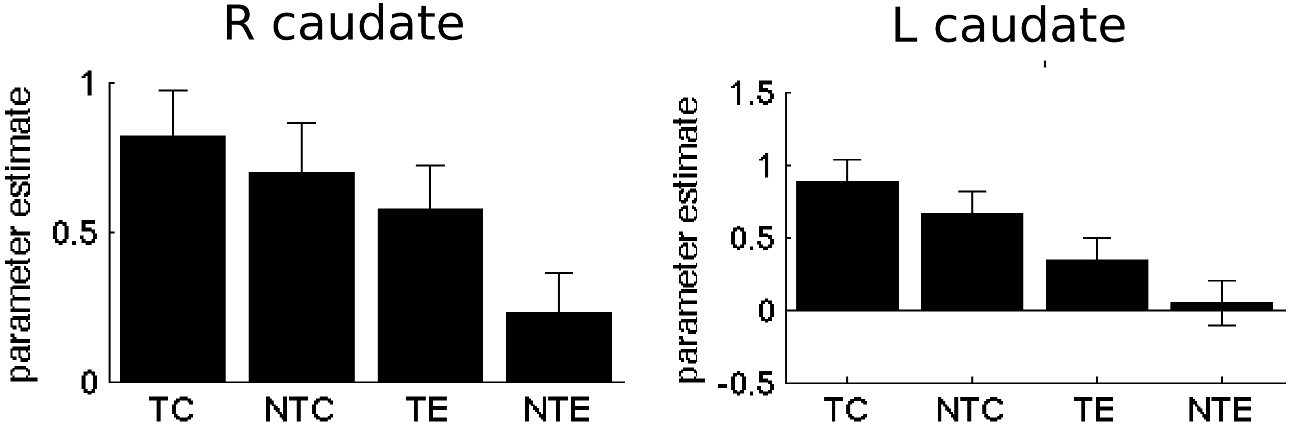
L putamen



Hybrid model
ROIs in the basal ganglia - caudate



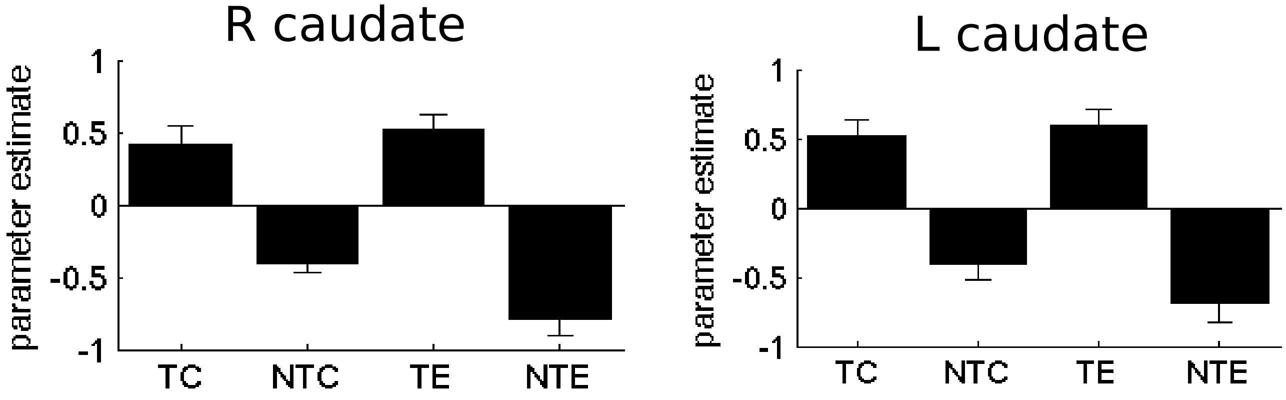
Response to feedback (main effect)



R caudate

L caudate

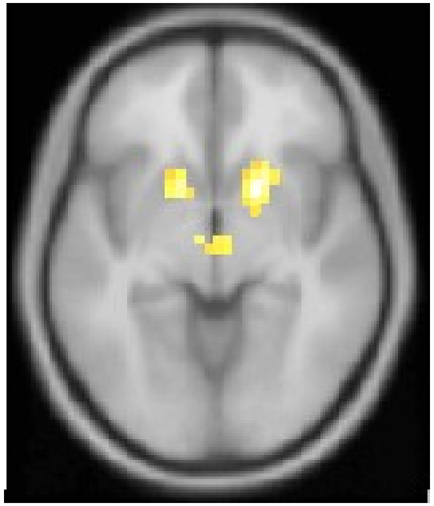
Response to prediction error

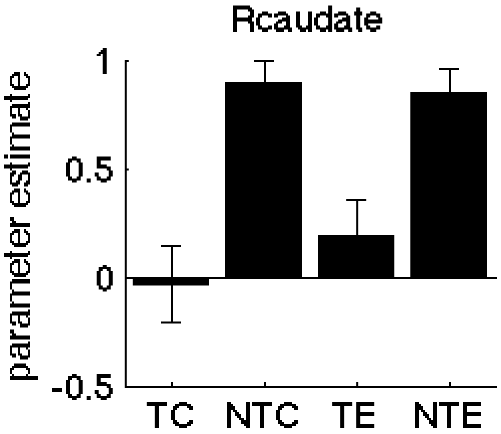


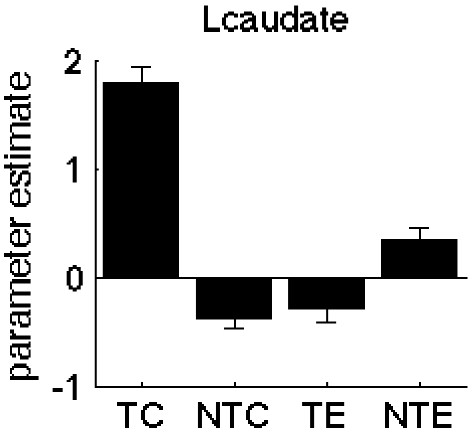
R caudate

L caudate

Hybrid model
Whole brain analysis







action between prediction error, feedback (correct/error) and et/nontarget, p < 0.0001