# LEARNING IN AN UNCERTAIN ENVIRONMENT I

Peter Bossaerts

# CONTENT

I. Reinforcement learning… of (reward) risk
   - Link with outlier risk

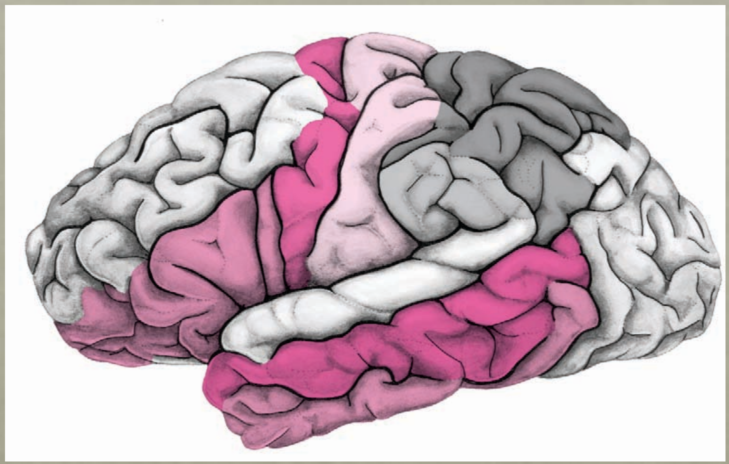
II. Learning of stochastic affectively neutral stimuli
   - Specifically: Bayesian learning

# Messages - Methods

- Why "contrasts" in imaging can be misleading
- Why knowing the biology behind behavior is important
- How "converging evidence" works
- Understanding models beyond the mathematics
- Experiments are not simply about "treatment" vs "control"

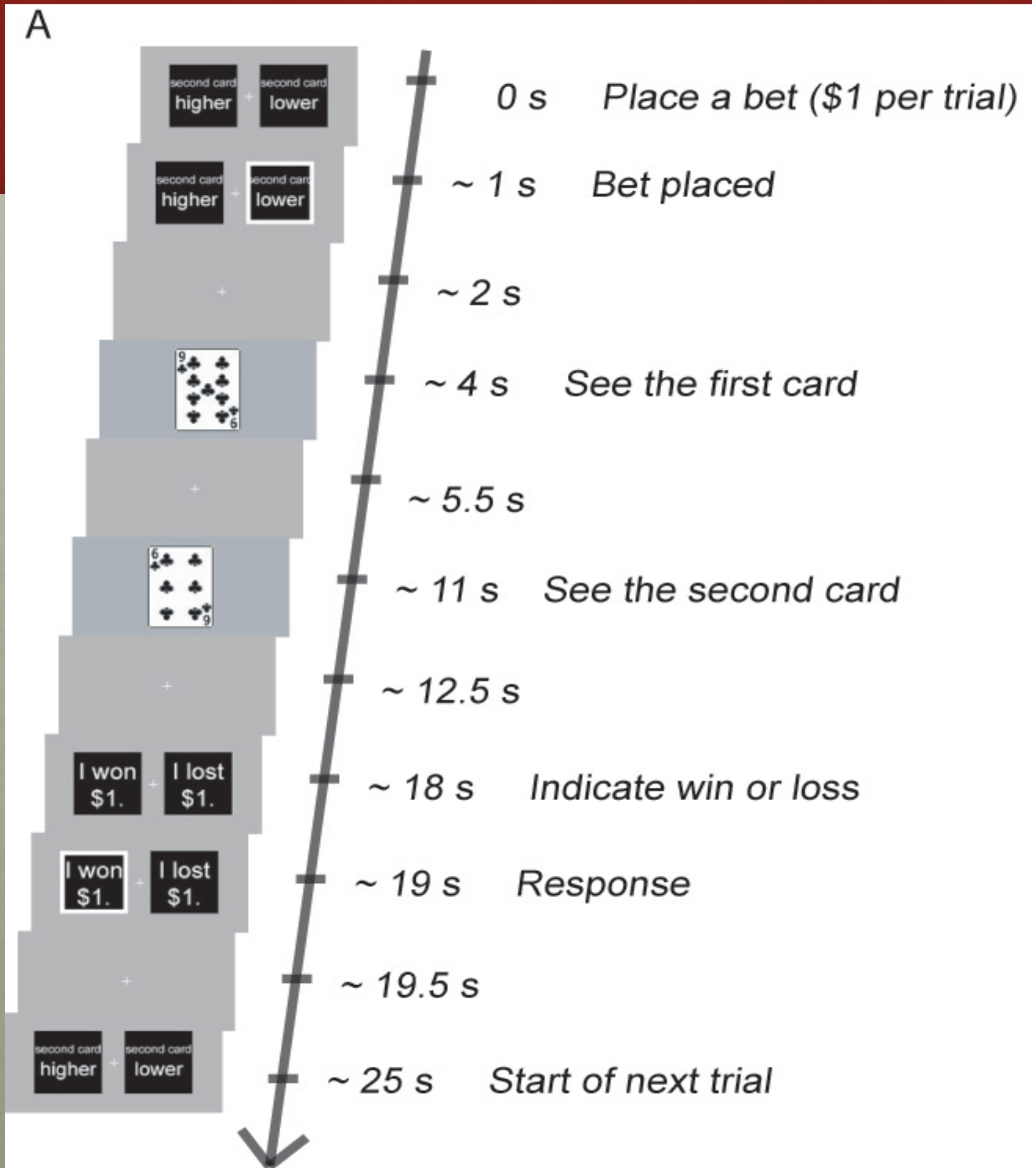
# 2. Reinforcement Learning… of (Reward) Risk

- How does the human brain process risk related to rewards?
- What are the computational algorithms that the brain uses?
- How general are these (algorithms)? Any potential shortcomings that would explain cognitive biases?
  - Or that would explain heterogeneity?



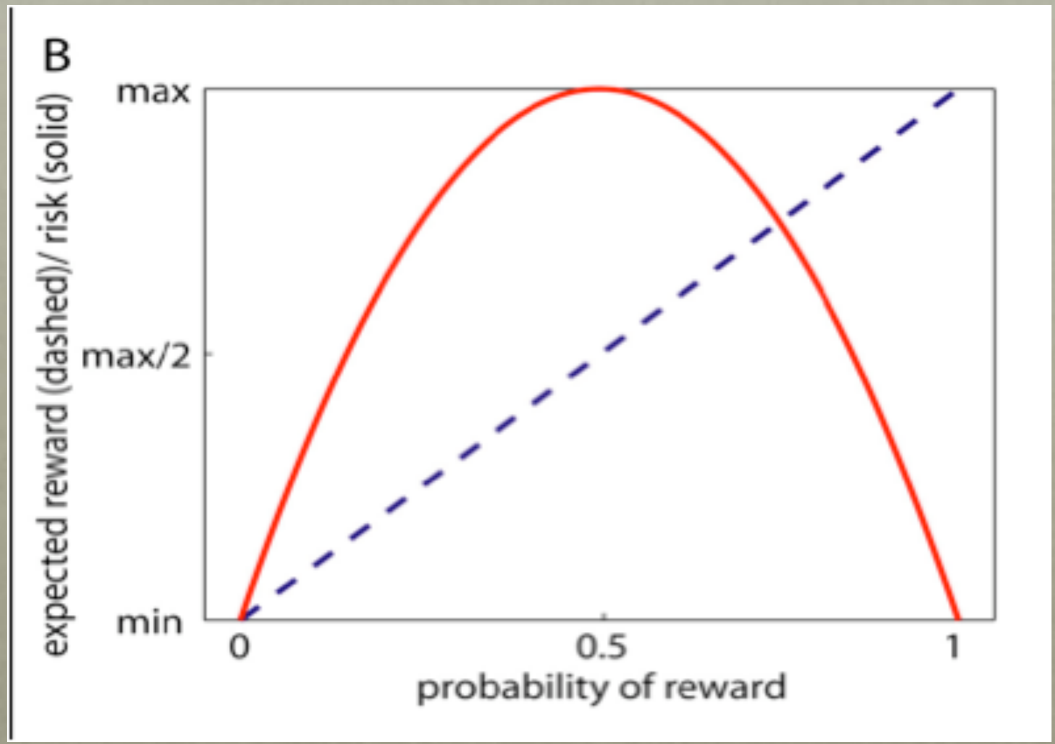
# A SIMPLE GAMBLE

- With each card, update:
  - EXPECTED REWARD
  - RISK



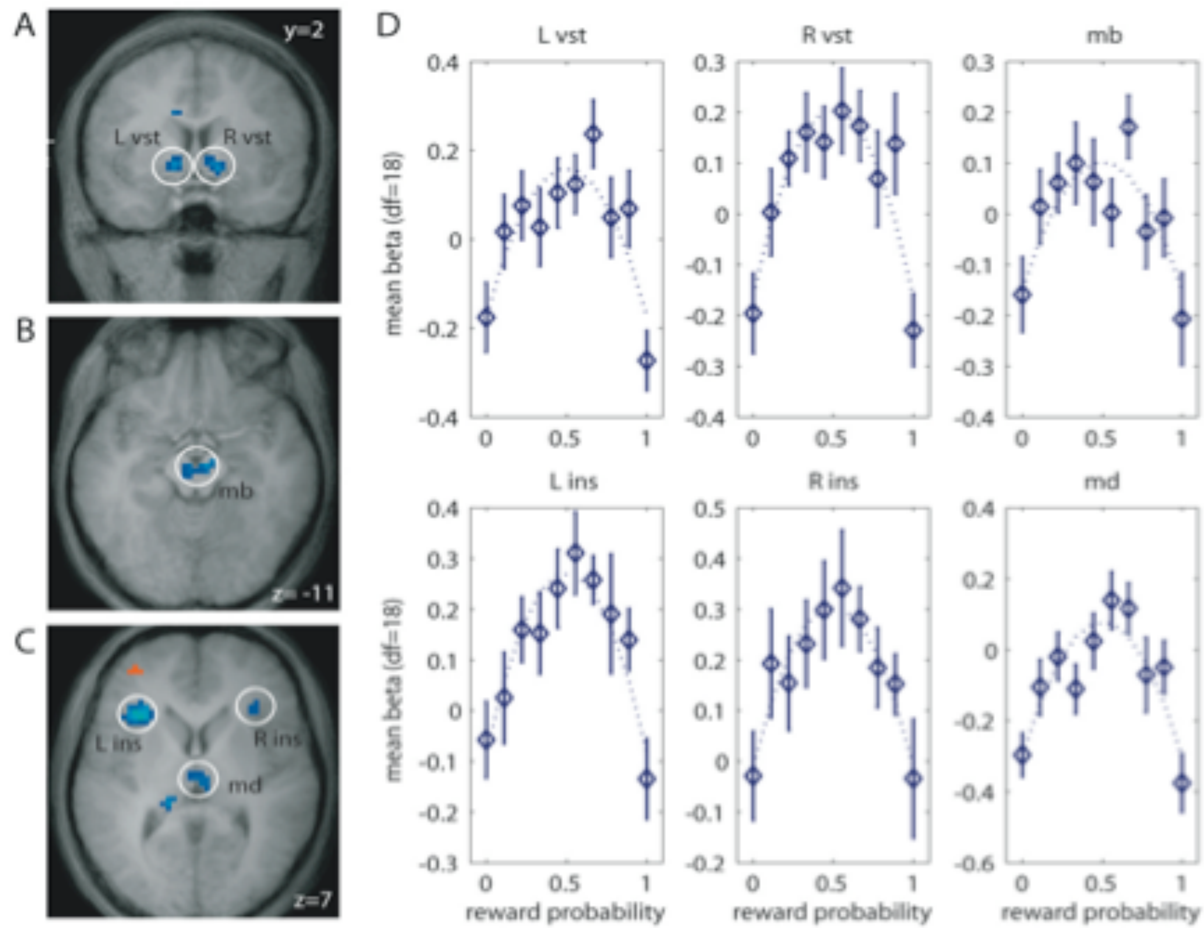
# Risk and Reward

- Expected reward increases linearly
- Risk (reward variance) is quadratic



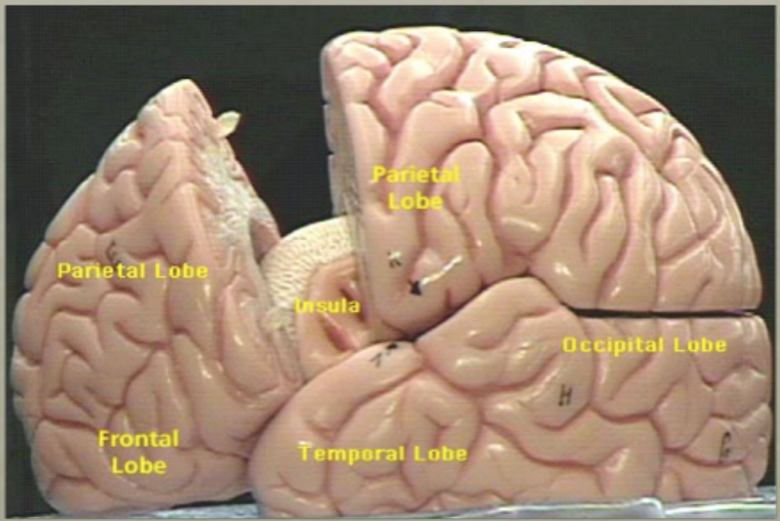
# Neural Activation Correlating with Risk

- fMRI signal per level of reward probability; averaged across subjects
- Activation AFTER seeing first card



# ANTERIOR INSULA

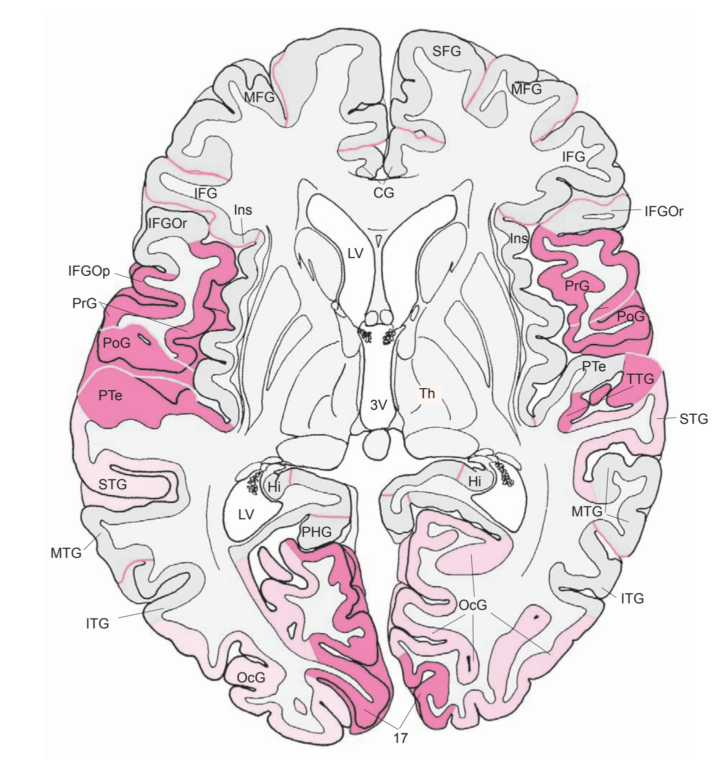
- Known to “convert” emotions into “feelings”
- (Ability to sense heartbeat ~ size)
- Related to self-awareness
- Activates especially in reaction to disgust, pain
- Involved in empathy




**Cool mathematics in a quintessentially emotional part of the human brain!**

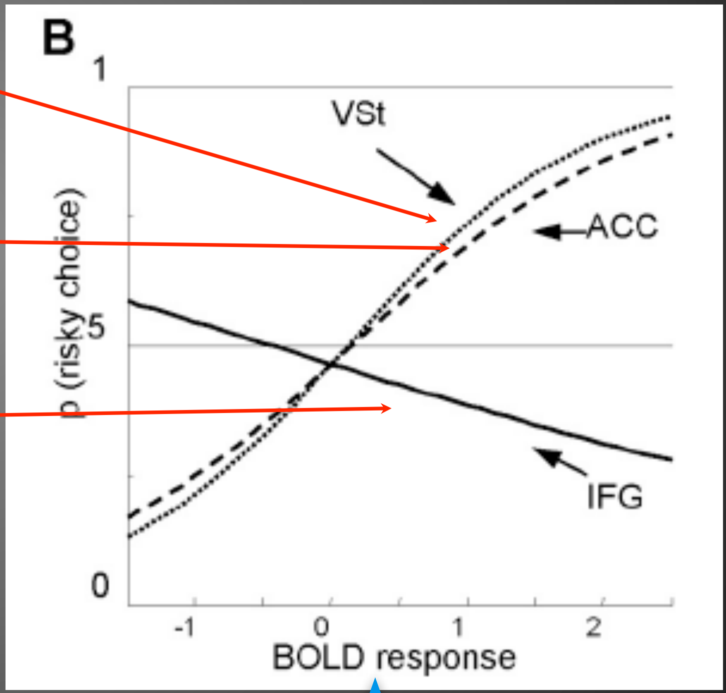
# OTHER REGIONS

- Anterior cingulate cortex
- Inferior frontal gyrus




# PREDICTING CHOICE?

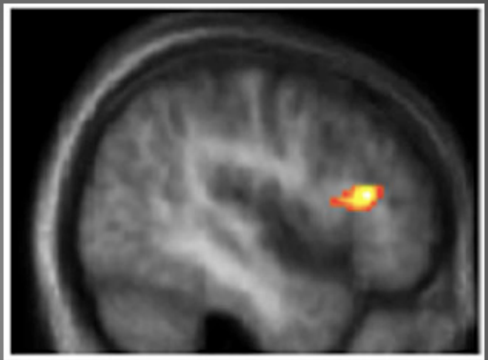
- VSt (ventral striatum): correlates with expected reward
- ACC (anterior cingulate cortex): correlates with (objective) risk
- IFG (inferior frontal gyrus): correlates with (subjective) risk
- *(Christopoulos ea, J Neuroscience 2009)*
- (ACC and IFG have opposite effects: opponent theory in biology)



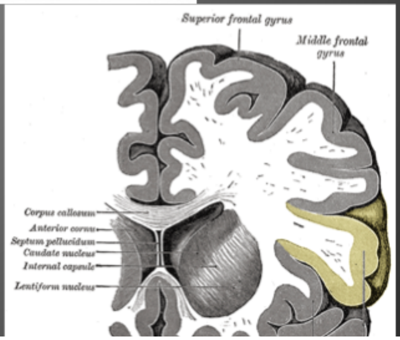

fMRI brain signals predicts risk taking

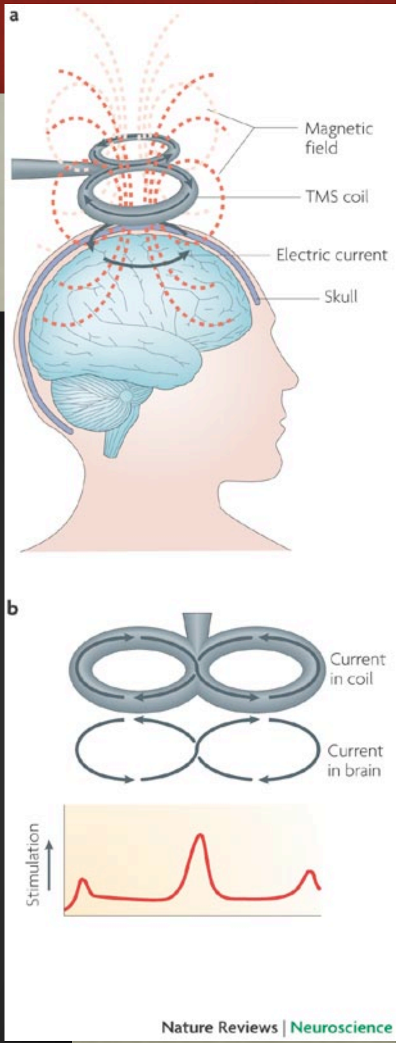
# Causing Temporary "Lesions"

transcranial magnetic stimulation



- Disrupting inferior frontal gyrus leads to reduced risk aversion
- *(Knoch ea, J Neuroscience 2006)*



# Learning Risk

- … based on (reward) reinforcement learning
- Core: *prediction error*
  - Here, it is a <u>risk</u> prediction error
  - Not a reward prediction error

# Risk Prediction Error

- Traditional reward prediction error: reward minus expected reward

$$\delta_t = r_t - \hat{v}_t$$

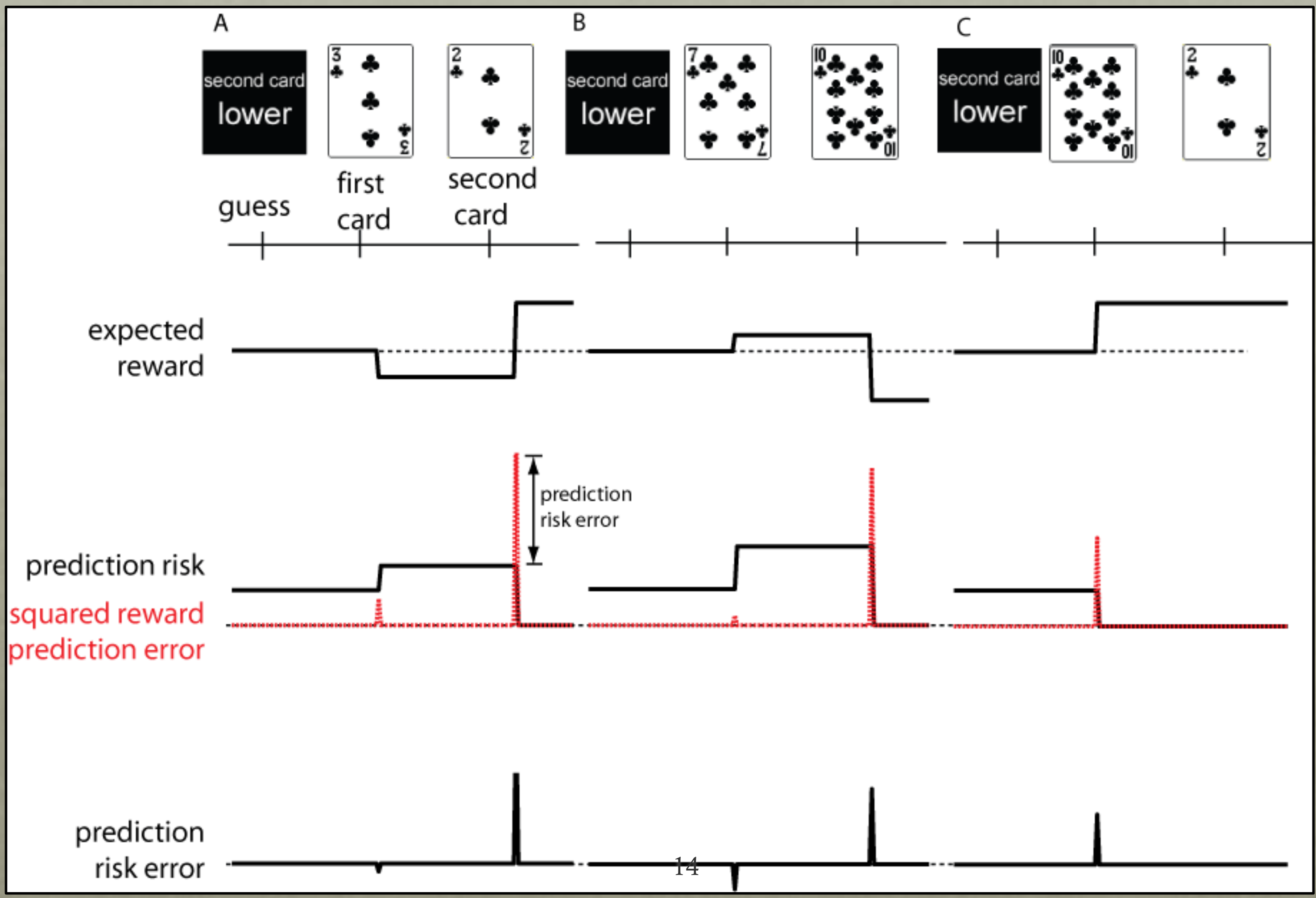
- Risk = expectation of size of reward prediction error

$$\hat{h}_t$$

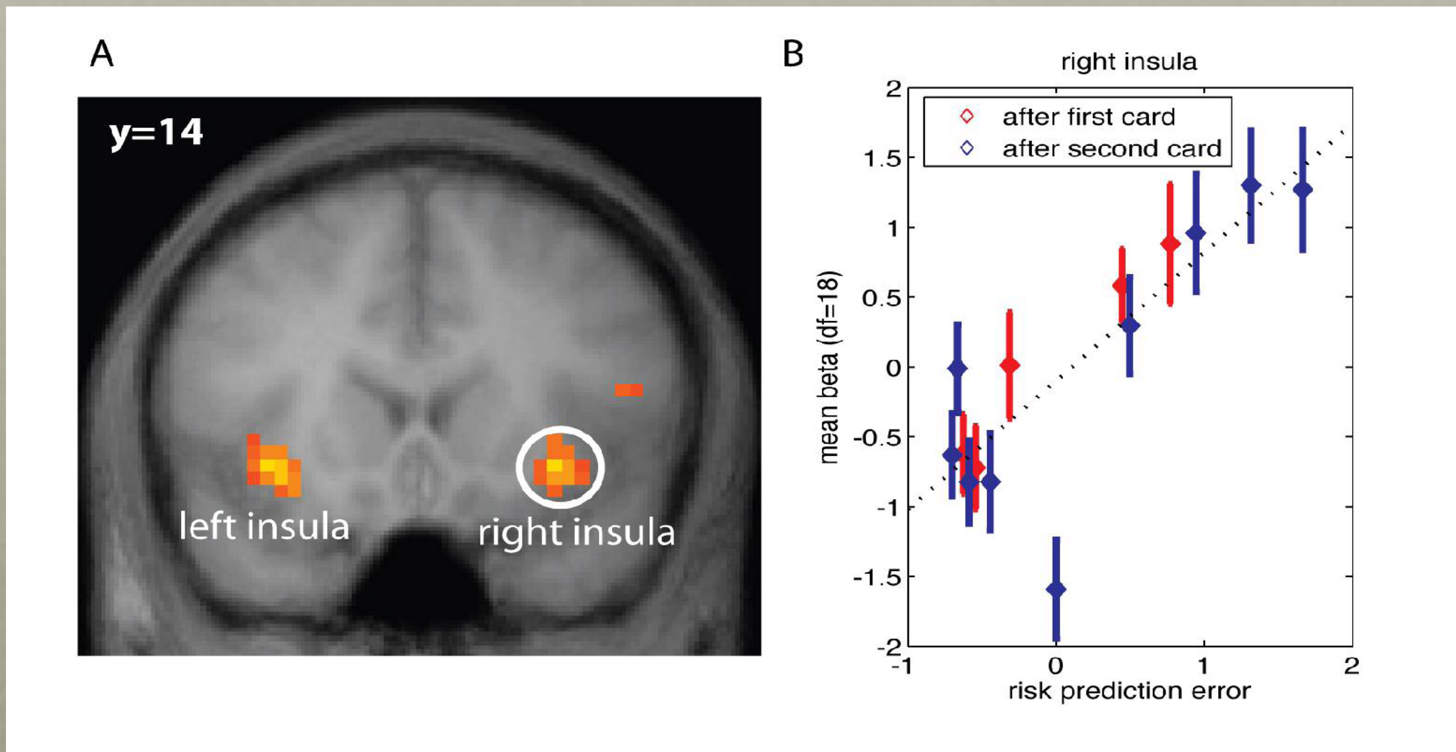
- Risk prediction error

$$\xi_t = \delta_t^2 - \hat{h}_t$$

# SOPHISTICATED…

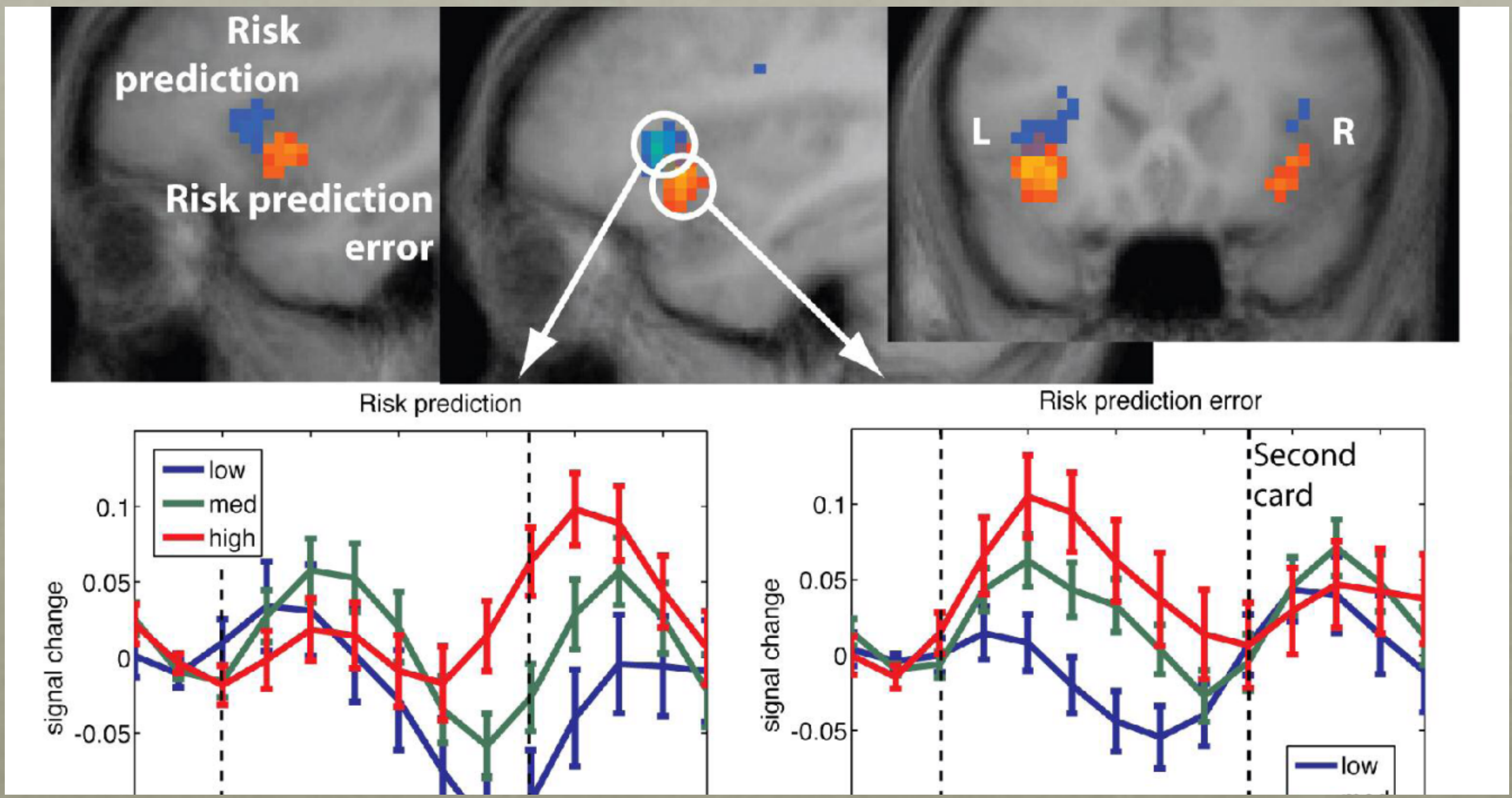


# FMRI EVIDENCE




Risk prediction error after *first* and *second* card. No risk, and hence, zero risk prediction error if first card is 1 or 10

# LOCATION RELATIVE TO RISK SIGNAL

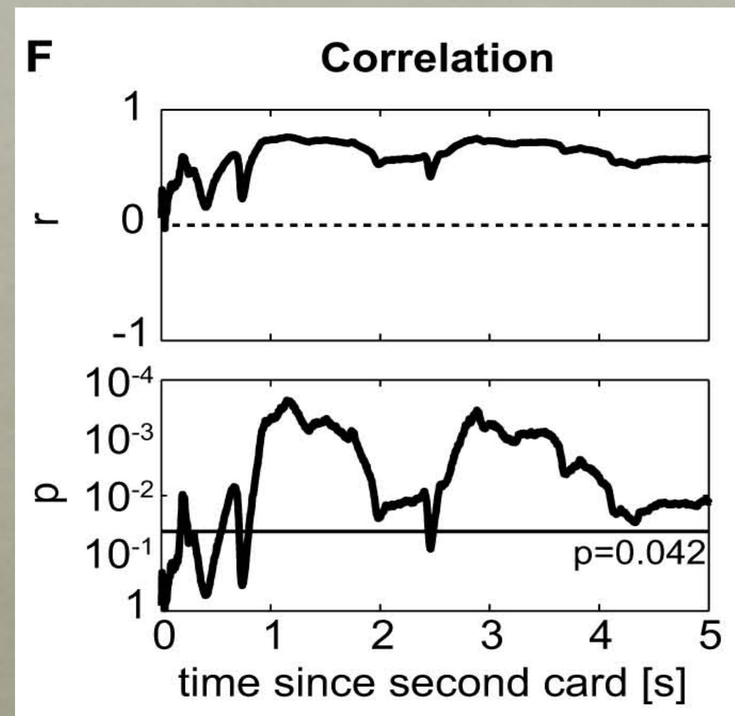


# Psycho-physiological Measurements

- Heartbeat
- Skin conductance
- …
- *PUPIL DILATION*

# PUPIL DILATION

- Pupil dilation
- … correlates with risk prediction error (Preuschoff ea, Frontiers in Neuroscience 2011)
- There is a link between pupil dilation and (phasic) changes in norepinephrine levels
- Could study this pharmacologically (Propranolol? Retalin also blocks dopamine re-uptake)
- … or with fMRI?




# Link with Outliers

- Is size of realized reward prediction error bigger than expected?
- If so, things may have changed…
  - **Regime shifts** (Econometrics)
  - **“Unexpected Uncertainty”** (Yu-Dayan)
  - **“Jumps”** (Payzan)

# Computational Neuroscience (Yu-Dayan)

- Outliers (relative to a gaussian world) signal *regime change*
- Learning rate will go up after regime change is detected, effectively discounting past information faster

$$\begin{aligned} y_t &= x_t + \epsilon_t \\ x_t &= x_{t-1} + \eta_t + \xi_t \end{aligned}$$

- y=observation, x=state, epsilon=observation error (gaussian), eta=gaussian state error, xi=Poisson state shock, i.e., the outlier!

# "ANIMAL MODEL"

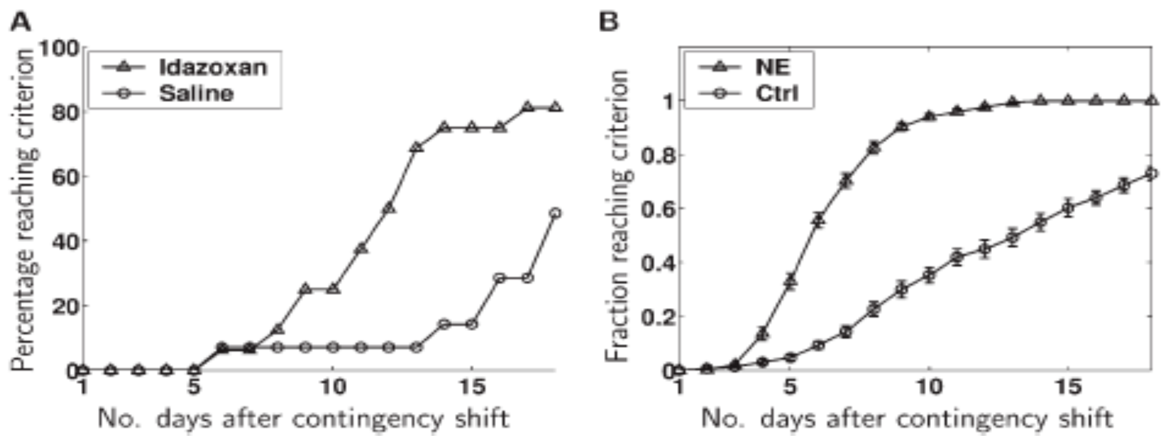
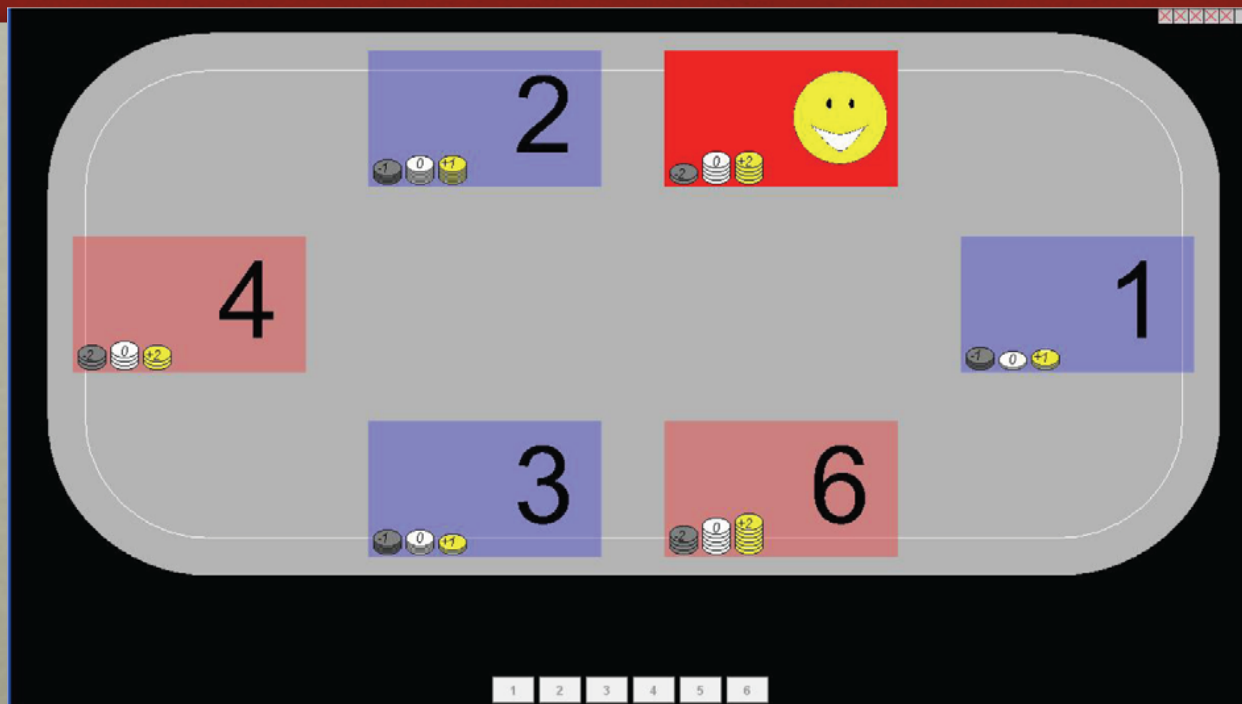



**Figure 4. A Maze Navigation Task and the Effects of Boosting NE**

(A) The cumulative percentage of idazoxan rats reaching criterion (making no more than one error on 2 consecutive days) considerably outpaced that of the saline-control group. Adapted from Devauges and Sara (1990) with permission from Elsevier.
(B) In the model, simulated "rats" with elevated NE levels (10% greater than normal) also learn the strategy shift considerably faster than controls. Data averaged over 20 simulated experiments of 30 model rats each: 15 NE-enhanced, 15 controls. Error bars: standard errors of the mean.

- Enhancing NE levels induces rats to abandon old "hypotheses" and find the newly optimal paths in a navigation task

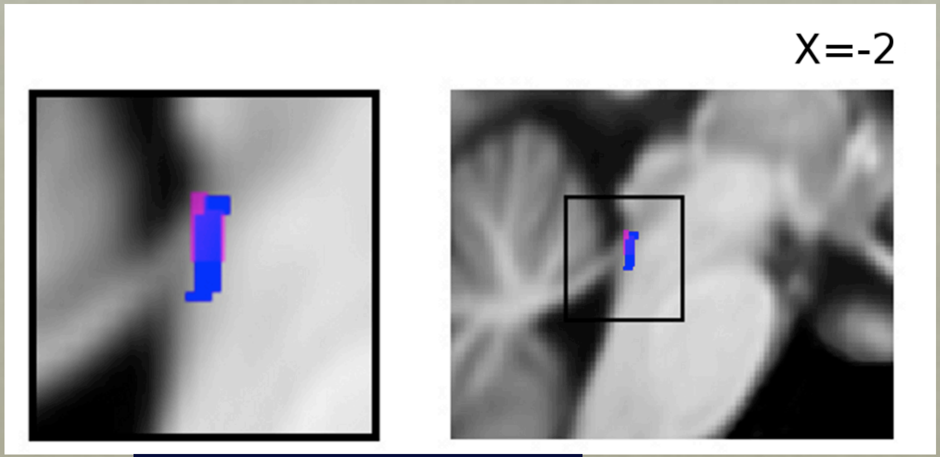
# TASK

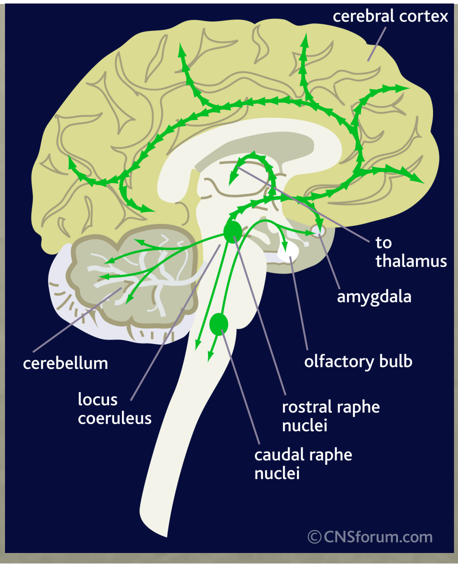



- Let's focus on what happens when reward probabilities JUMP (for a color)

# Norepinephrine Nucleus Reaction to Outliers or "Jumps"

- Locus coeruleus correlates with the probability that a jump has occurred
- Signal CORRECTED for impact on LEARNING RATE (so this is not just a mechanical change in the learning rate due to surprise)




(From Payzan-Le Nestour, ea, *Neuron*, in press)

# DO OUTLIERS ALWAYS SIGNAL A "REGIME SHIFT"?

- NO!
- Think about black swans… (Nassim Taleb?)
- But if the norepinephrine system is built for outliers-that-signal-regime-shifts, it is ill adapted to the other type of outliers!

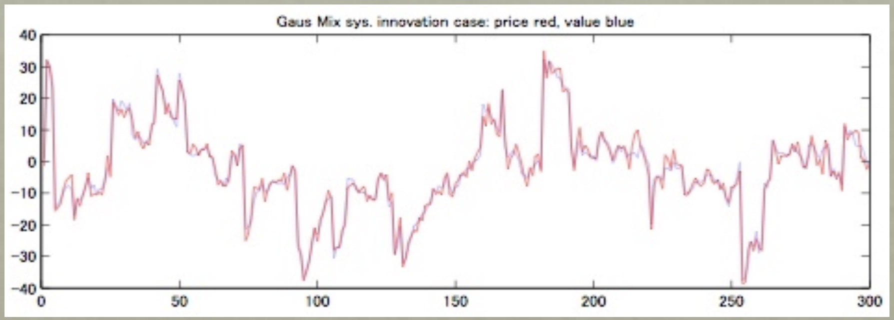
# Allow for a continuum of jumps

$$
\begin{aligned}
y_t &= x_t + \epsilon_t \\
x_t &= x_{t-1} + \eta_t
\end{aligned}
$$

- Here, eta is extreme-value (say, Cauchy – for which variance is not even defined)
- Outliers happen all the time have a permanent effect – signal regime shift: *outliers of the first kind*

# EXAMPLES

- Simulation



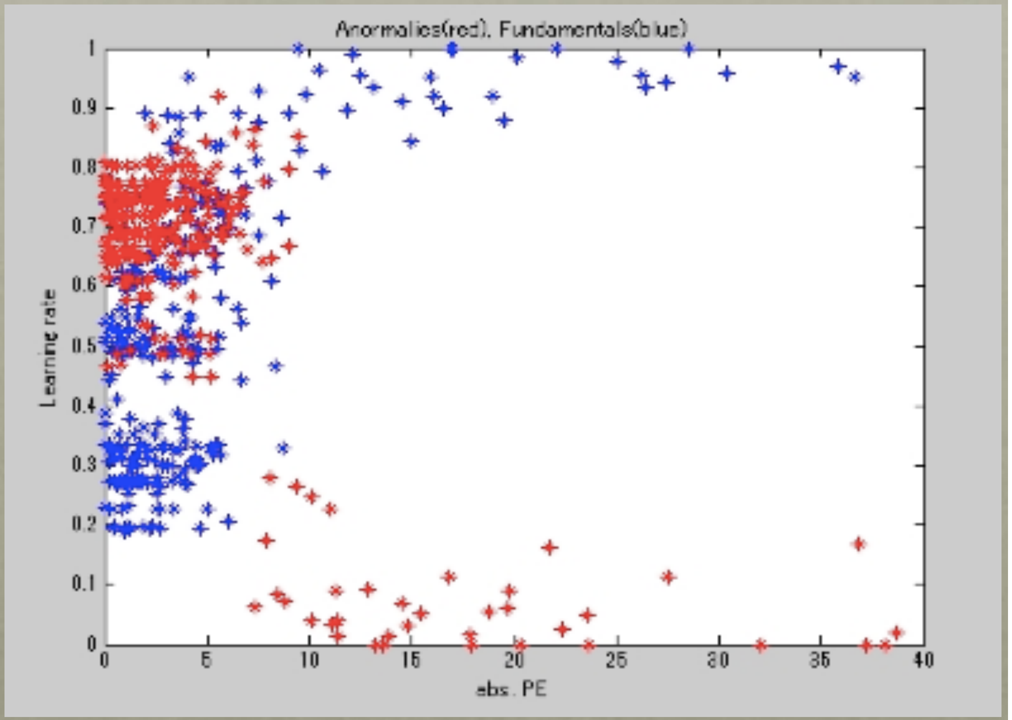

- Empirical Example?

# TASK

- **python game6.py (1,1,21)**

# OPTIMAL POLICY

- *Learning rate increases in prediction error*
- Typically observed: see blue stars
- (Ignore red)



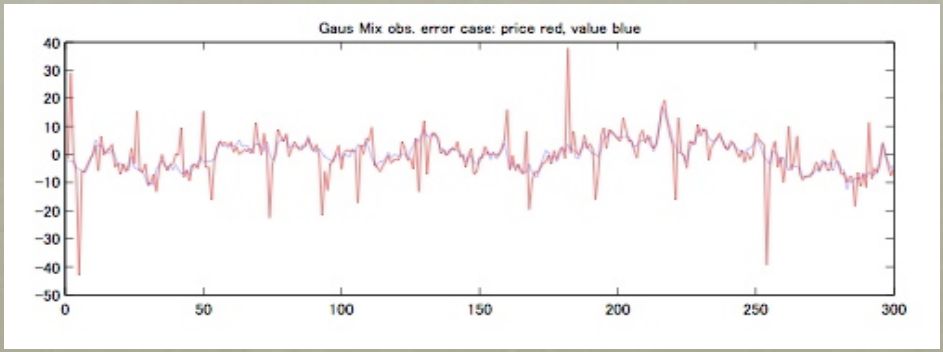
# Outliers of the Second Kind

$$
\begin{aligned}
y_t &= x_t + \epsilon_t \\
x_t &= x_{t-1} + \eta_t
\end{aligned}
$$

- Observation error epsilon is extreme-value, while state transition is standard Gaussian
- Now outliers do NOT signal a "regime change" – only a temporary deviation
- (Stock markets: October 1987; much of 2007 "slow-motion crash;" FX market - see picture)

# EXAMPLES

- Simulation



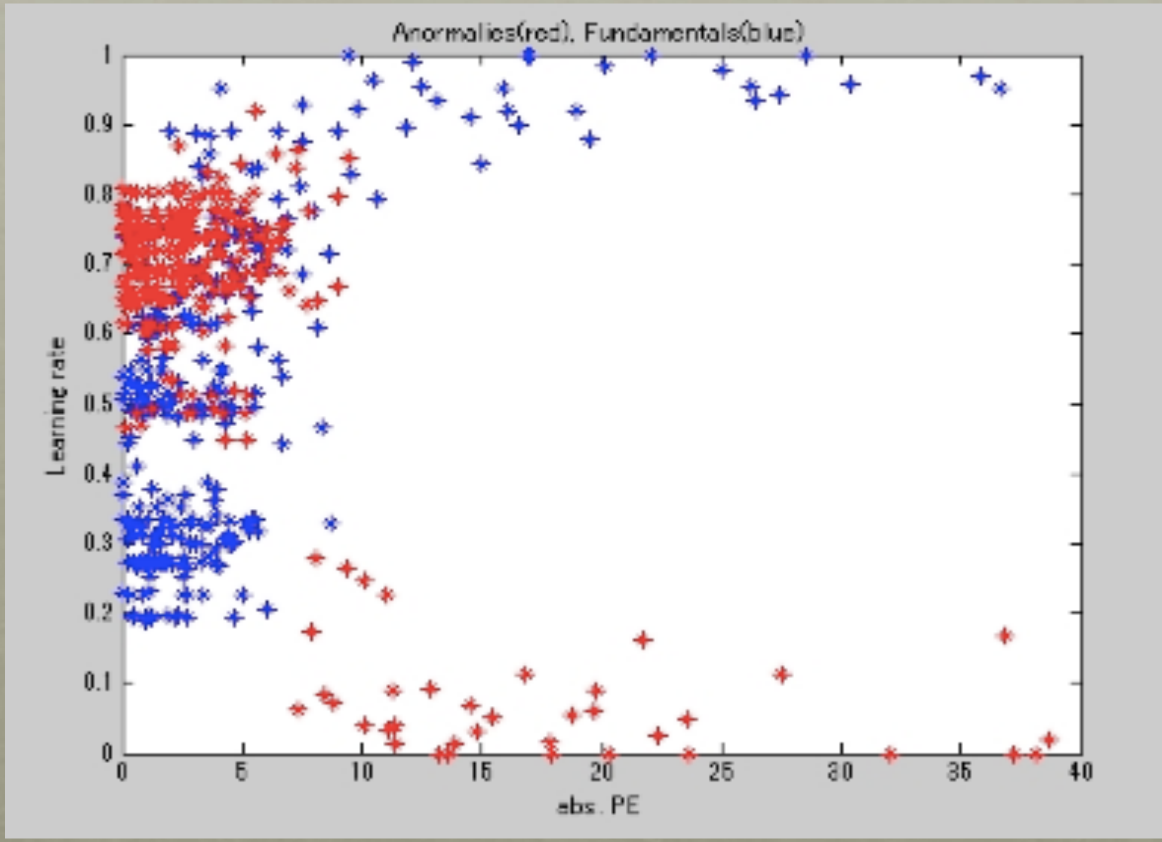

- Empirical Example?

# TASK

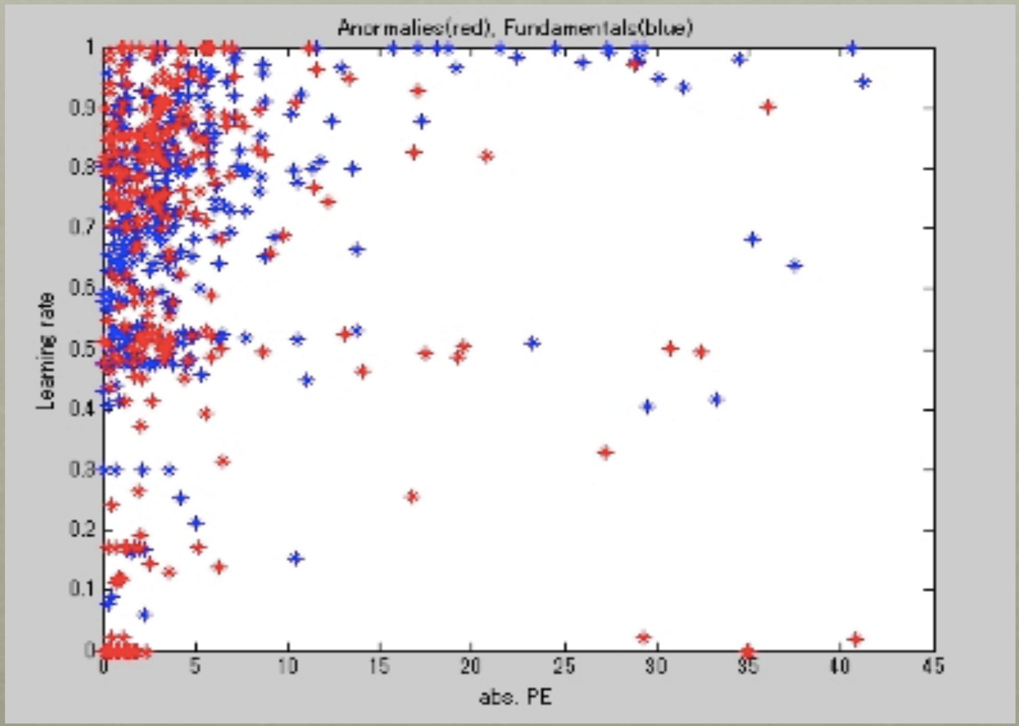
- **python game6.py (1,1,11)**

# This participants got it right… (red crosses)

- <u>Decrease</u> learning rate with prediction error!



# BUT TYPICALLY, SUBJECTS DO NOT GET IT RIGHT




Does the norepinephrine system stand in the way of learning this type of risk?

# Preliminary Lessons

- Why knowing the biology behind behavior is important
  - Biology may be ill adapted (see outliers)
  - Biology arbitrates between multiple models that can hardly/cannot be distinguished behaviorally (see jump risk)
- How "converging evidence" works
  - `fMRI, psychophysiology, animal model, pharmacology, TMS,…
- Understanding models beyond the mathematics
  - Reinforcement learning is about prediction errors
  - Understand what outliers really mean (what does "abnormal" mean?)
- Experiments are not simply about "treatment" vs "control"
  - Change risk prediction error continuously
  - Control for mechanistic changes in learning rate when tracking jump risk