Malignant Evaluation: Neuromodulation, Depression and Reinforcement Learning

Quentin JM Huys and Peter Dayan

Gatsby Computational Neuroscience Unit, University College London, UK
{qhuys,dayan}@gatsby.ucl.ac.uk www.gatsby.ucl.ac.uk/~qhuys

Overview

Neuromodulators :: RL
    Dopamine – rewards and actions
    Serotonin – aversion, stress, inactivity
Noradrenaline – unexpected uncertainty
    Acetylcholine – expected uncertainty

=> NMs control plasticity

Neuromodulators :: Psychiatry
    Schizophrenia – DA, 5HT
    ADHD – DA, NA
    Parkinson's – DA
    Alzheimer's – Ach
    Depression – 5HT

=> what controls the controllers?
Meta-meta-plasticity
[cf metaplasticity Doya 2002]

Psychiatry :: RL ??

- Analyse the plasticity of NMs themselves
- Set NMs as a function of reward / punishment statistics in the environment
- Relate psychiatric dysfunction to normal function via normative roles of NMs

Depression

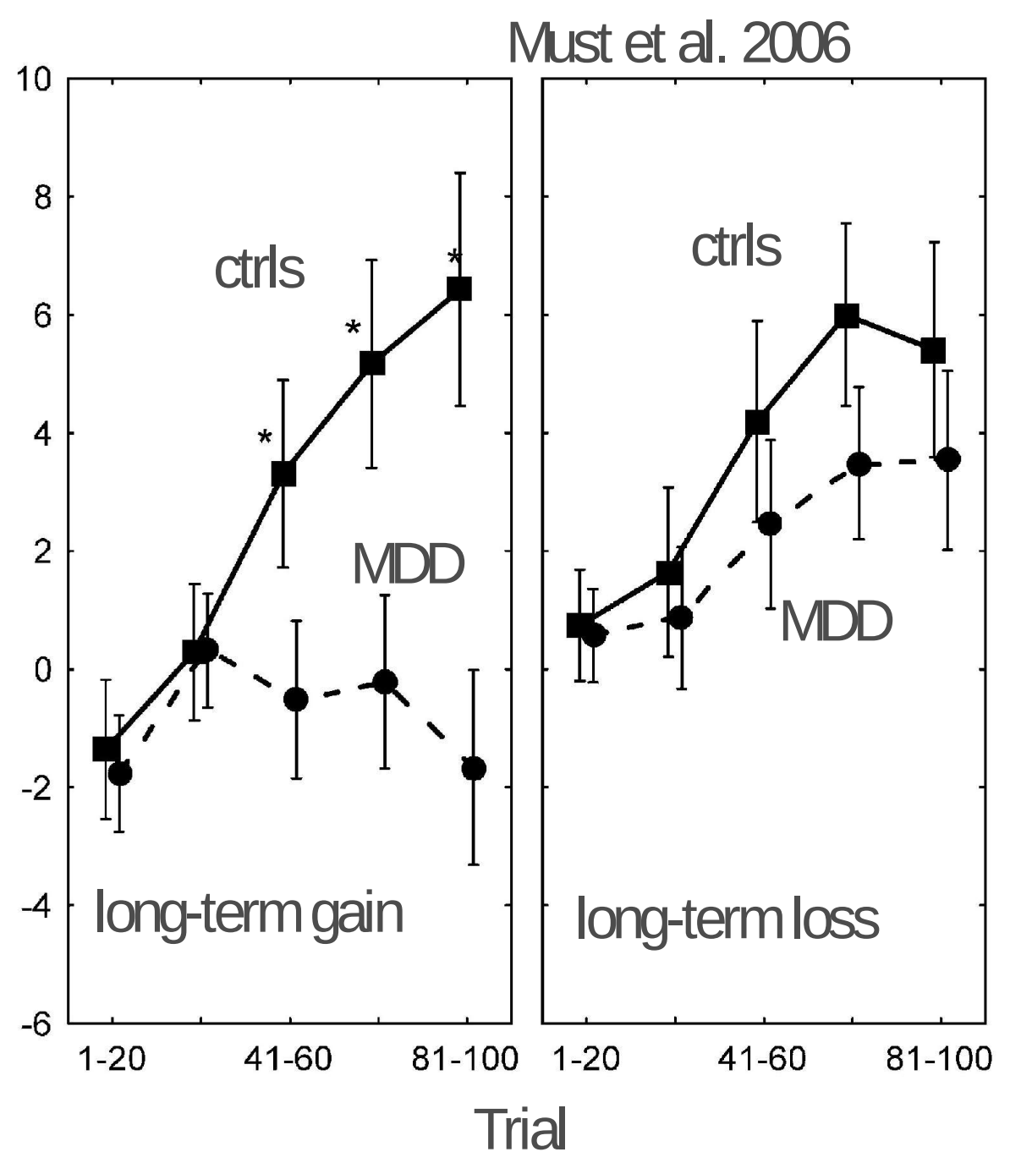
- By 2020: 2nd most important disability worldwide
  - **malfunction in society**
  - suicide, heart disease
- 5-20% lifetime incidence in USA
- Definition:
  - low mood / anhedonia
    -> reward experience
  - hoplessness / worthlessness / helplessness
  - indecisiveness / diminished ability to think
    -> reward usage
  - loss of energy and others
- Aetiology
  - Stress
  - 5HT transporter polymorphism
- Treatment
  - Cognitive behavioural therapy
  - SSRIs (SNRIs, TCA, ECT, stimulants)

Depressed decisions

Depression :: inefficient use of performance evaluations to guide future actions
Serotonin is instrumental

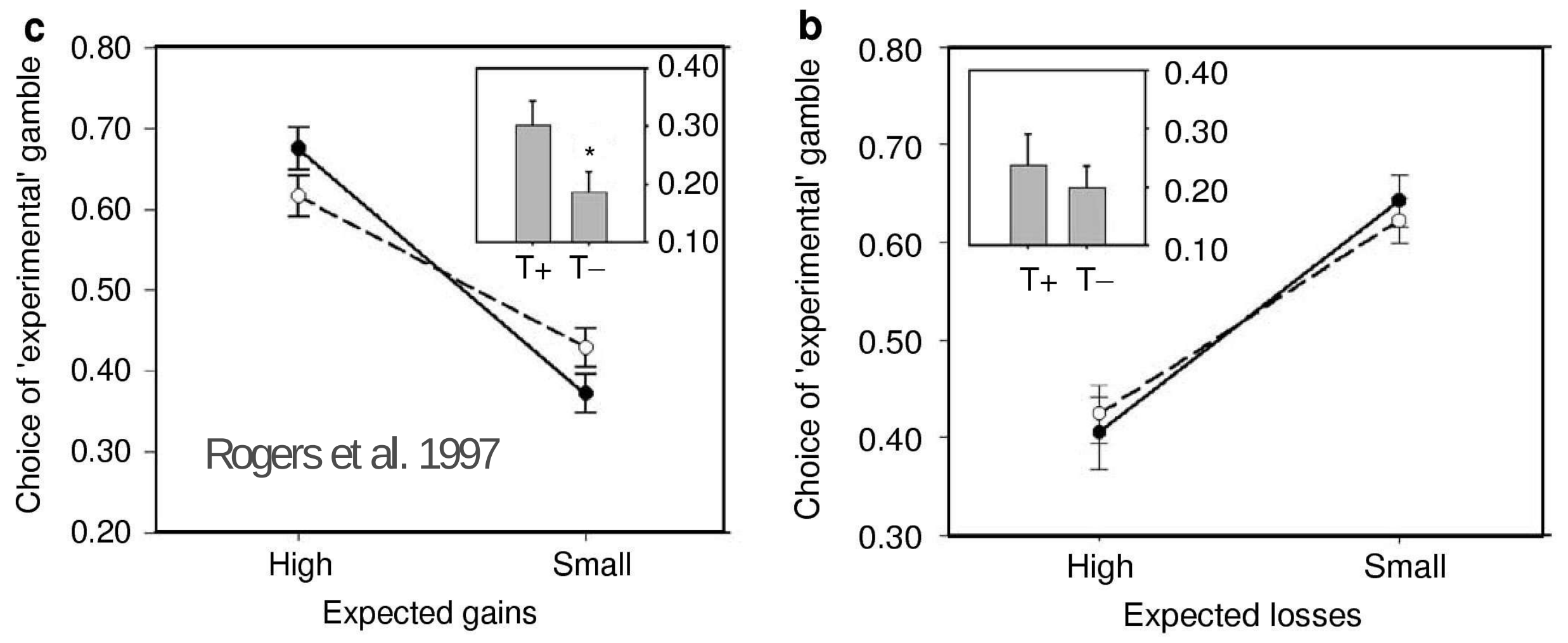
Normal use of negative reinforcement, but suboptimal use of positive reinforcements.

High sensitivity and inefficient use of negative feedback [Elliott et al. 97]

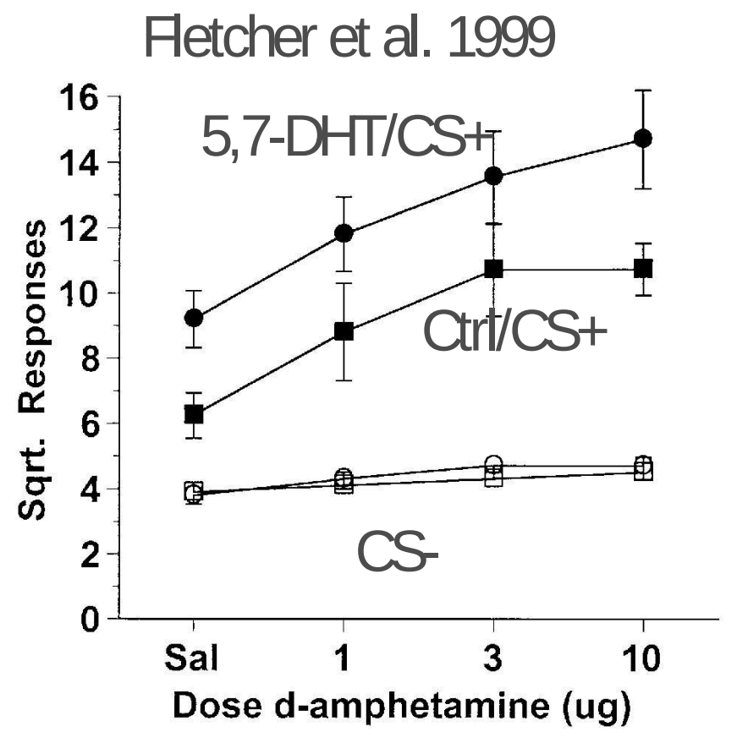

Must et al. 2006

Serotonin

Tryptophan depletion decreases availability of 5HT. This causes depression relapse and qualitatively recovers aspects of decision making in depressed subjects (also mood-congruent biases).



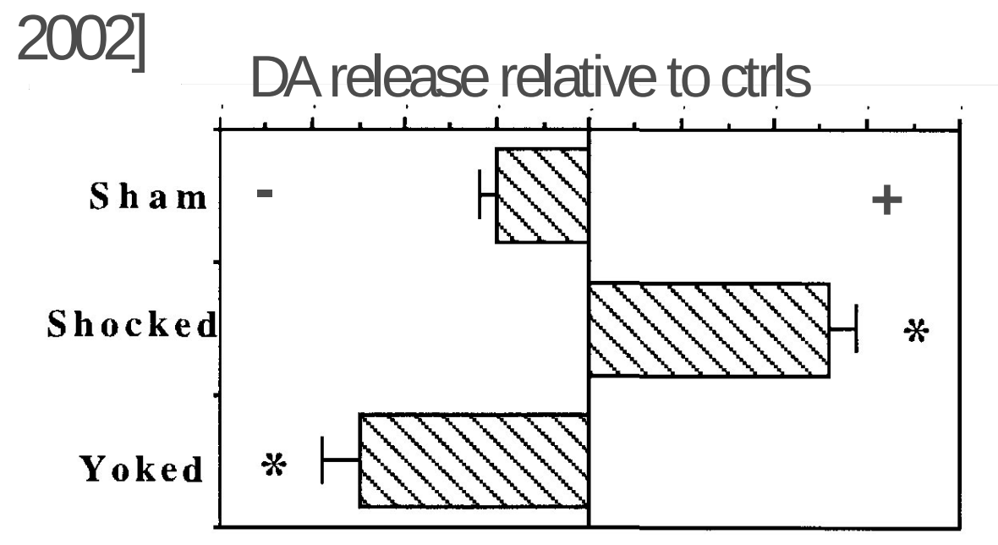
5HT opposes DA


Fletcher et al. 1999

- DA is crucial to learning about rewards
- 5HT mediates response suppression by punishments
- Stress affects DA in NAcc, PFcx
- Conditioned reinforcement potentiated by 5,7-DHT lesions and by amphetamine

=> **5HT opposes DA** [cf. Daw et al. 2002]

D_1 agonist and ADs prevent LH



Cabib and Puglisi-Allegra 1996

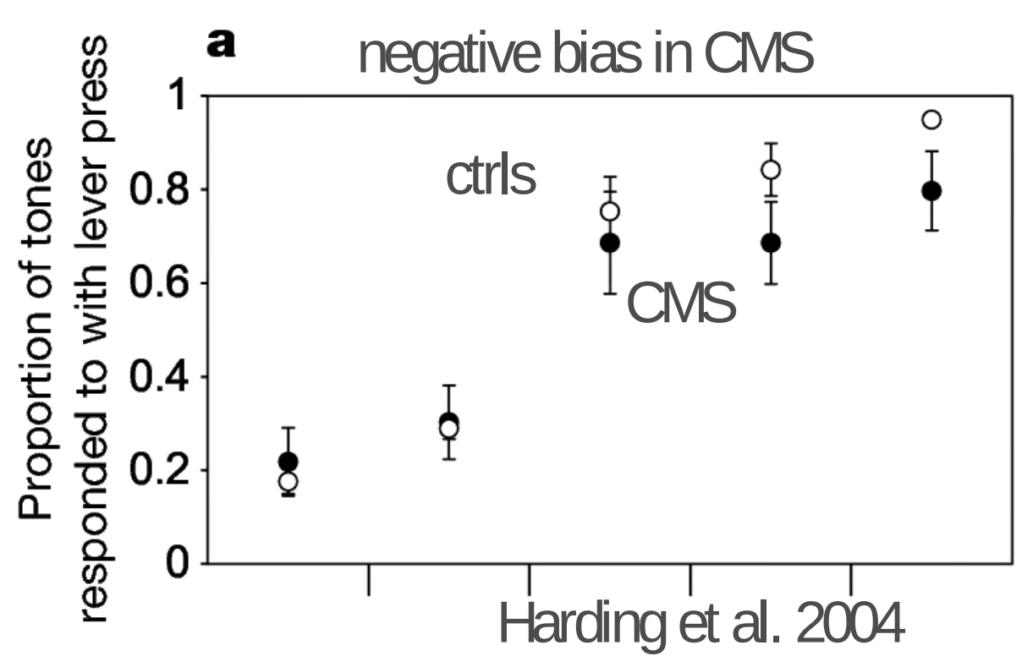
Antidepressants

- Act on 5HT but promote DA
- Increase prefrontal DA
- Increase responding for ICSS
- Potentiate place preference maintained by various rewards
- Chronic imipramine effect on LH is antagonised by D_1 antagonist.

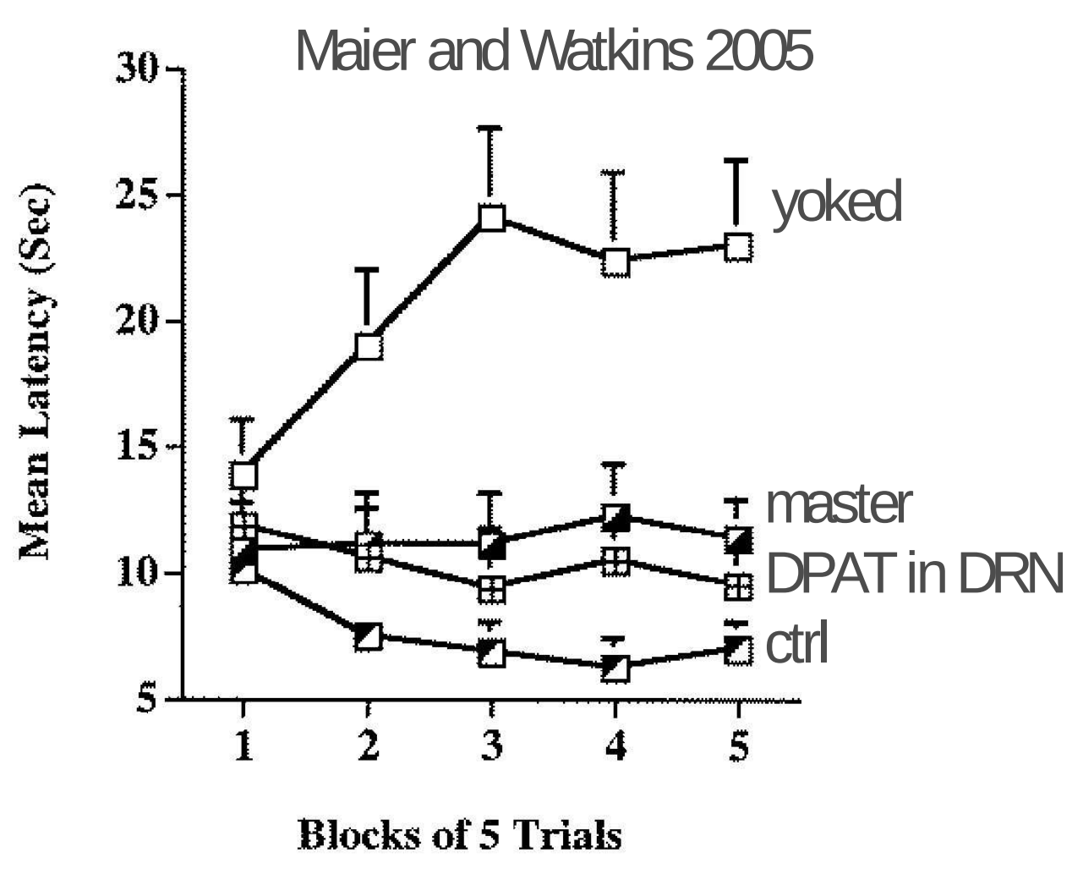
Stress and 5HT

Stress is used as an inducer in most animal models of depression.

Three main stress paradigms
- Learned Helplessness (LH)
  - Uncontrollable but not controllable shock induces depressive state.
- Chronic Mild Stress (CMS)
  - Varied (but not same) mild stressors induce depressive states
- Behavioural despair



Harding et al. 2004


Maier and Watkins 2005

- Behavioural changes in all these models are sensitive to clinically active antidepressants.
- 5HT from dorsal raphe crucial to effect of inescapable shock.
- Previously ineffective mild stressors can be effective after severe stress (life events?)

Stress induces an alteration in the use of rewards. Are different global settings of reward processing optimal under different reward / punishment statistics? This is a RL question.

Coping with stress

LH and CMS together reveal importance of stress size, variability and controllability. Four basic paradigms:

	Paradigm	Paradigm parameters		
		Shock size	**variability**	**Control**
A	LH master	large	low	yes
B	LH yoked	large	low	no
C	CMS constant	small	low	no
D	CMS variable	small	high	no

after Cabib and Puglisi-Allegra 1996

In each of these a different strategy (steady-state action) is optimal. Choose amongst

Lever press, Blunting, Preparation, Nothing

Analyse whether there is regime in which depressive blunting is optimal. Blunting flattens the utility curve, reducing the relevance of **both** rewards and punishments.

LH master rat press lever to terminate shock
LH yoked rat blunting as punishment too large to prepare for it.
CMS constantprepare to minimise impact of specific punishment
CMS varied blunting as punishments unpredictable

Optimal helplessness?

Depressive symptoms as normative stress responses – the case of blunting

In each steady-state scenario (A) LH master (B) LH yoked (C) CMS constant (D) CMS variable, choose least costly action amongst:
- lever press (l) :: cost of action, but may switch off shock
  - cost proportional to shock frequency
- prepare (π) :: for small shocks of known type this is not too costly
  - cost linear in time, varies with shock type
- blunt (b) :: most drastic action – forgo rewards to avoid punishments
  - cost is loss of rewards
- nothing (n) :: just incur shocks as they happen

$$\begin{aligned} C^1(n) &\approx \mathcal{S}^l \\ \rightarrow C^1(l) &= f_s\mathcal{S}^l - f_r\lambda_r\mathcal{R} + \lambda_s c_l \\ C^1(\pi) &\approx \mathcal{S}^l_\pi \\ C^1(b) &\approx \mathcal{S}^l_b \end{aligned}$$

A B

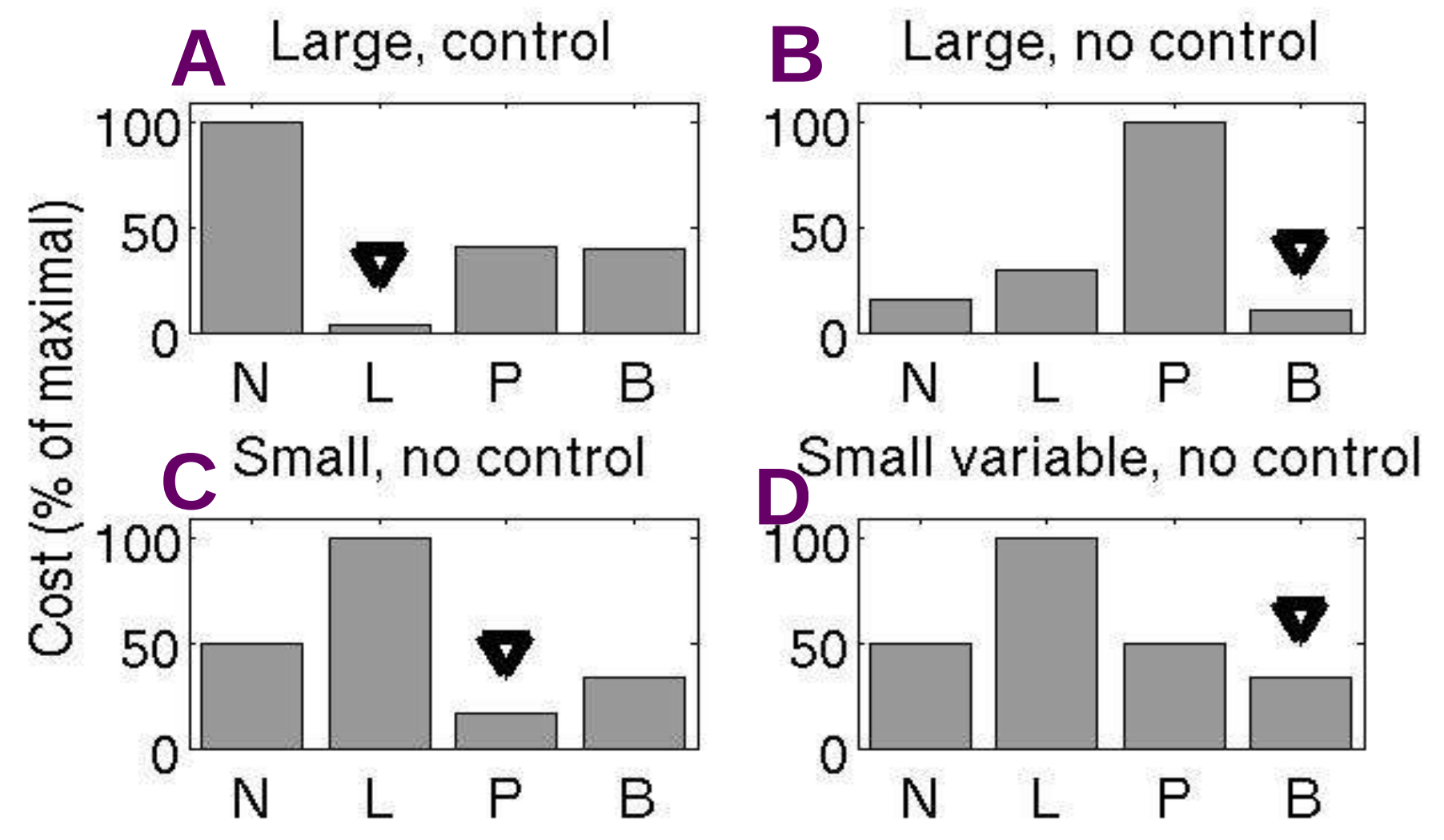
$$\begin{aligned} C^2(n) &= f_s\mathcal{S}^l - f_r\mathcal{R}\lambda_r \\ C^2(l) &= f_s\mathcal{S}^l - f_r\mathcal{R}\lambda_r + \lambda_s c_l \\ C^2(\pi) &= f_s\mathcal{S}^l_\pi - f_r\mathcal{R}\lambda_r + c_\pi(\mathcal{S}^l) \\ \rightarrow C^2(b) &= f_s\mathcal{S}^l_b - f_r\mathcal{R}_b\lambda_r \end{aligned}$$

C D

$$\begin{aligned} C^3(n) &= f_s\mathcal{S}^s - f_r\mathcal{R}\lambda_r \\ C^3(l) &= f_s\mathcal{S}^s - f_r\mathcal{R}\lambda_r + \lambda_s c_l \\ \rightarrow C^3(\pi) &= f_s\mathcal{S}^s_\pi - f_r\mathcal{R}\lambda_r + c_\pi(\mathcal{S}^s) \\ C^3(b) &= f_s\mathcal{S}^s_b - f_r\mathcal{R}_b\lambda_r \end{aligned}$$

$$\begin{aligned} C^4(n) &= f_s\mathcal{S}^s - f_r\mathcal{R}\lambda_r \\ C^4(l) &= f_s\mathcal{S}^s - f_r\mathcal{R}\lambda_r + \lambda_s c_l \\ C^4(\pi) &= f_s\mathcal{S}^s_\pi - f_r\mathcal{R}\lambda_r + c_\pi(n\mathcal{S}^s) \\ \rightarrow C^4(b) &= f_s\mathcal{S}^s_b - f_r\mathcal{R}_b\lambda_r \end{aligned}$$

Just find a set of (12) parameters that satisfies these equalities, plus 55 inequalities (such as $c_\pi(S^1) > c_\pi(S^s)$). $C^1(1)$ appears larger than $C^2(1)$ due to the normalisation only.



Conclusions

Psychiatry :: related to **normative** emotional function [Nesse 00]

Depression ::
    as a (pathologically prolonged) optimal response
    cognition related to pharmacology (CBT vs pharmacology?)

Neuromodulators ::
    opponency
    plasticity related to normative framework
    statistics and controllability

Reinforcement learning ::
    inference of hyperparameters?

Acknowledgements: We thank JD Steele, Barbara Sahakian, Nathaniel Daw and Yael Niv for stimulating discussions. Funding: Gatsby Charitable Foundation, UCL Medical School, BIBA consortium.

References

Cabib and Puglisi-Allegra 1996: Psychopharmacology 128:331-42; **Braver et al.** 99: Biol. Psychiatry 46:312-28; **Doya** 2002: Neural Networks 15:495-506; **Daw et al.** 2002: Neural Networks 15:603-616; **Elliott et al.** 1997: J. Neurol. Neurosurg. Psychiatry 63:74-82; **Fletcher et al.** 1999: Psychopharmacology 147:291-9; **Harding et al.** 2004: Nature 427:312; **Honk et al.** 2003: Neuroreport 14:1993-6; **Maier and Watkins** 2005: Neurosci. Biobehav. Rev. 29:829-41; **Must et al.** 2006: J. Affect. Disord. 90:209-15; **Nesse** 2000: Arch. Gen. Psychiatry 57:14-21; **Rogers et al.** 2003: Neuropsychopharmacology 28:153-62; **Schultz and Dickinson** 2000: Ann. Rev. Neurosci. 23:473-500; **Smith et al.** 1997: Lancet 249:915-9; **Williams and Dayan** 2005: J. Child Adolesc. Psychopharmacol. 15:160-79; **Yu and Dayan** 2005: Neuron 46:681-92