Computational Psychiatry Rising
(on the shoulder of reinforcement learning and decision making models)

What are the big problems that neuroscience could solve?

• mood disorder (Depression, Bipolar ..) : ~ 10% of the population (at some point in life) in US
• anxiety disorder (Panic, OCD, PTSD): ~18% of the population
• addiction: alcohol ~ 10% of the population (at some point in life)
• eating disorder (Anorexia, Bulimia): ~ 4 %
• ADHD: ~4 % (adults)

[NIMH]
• drugs often work poorly
• precise mechanisms of action unknown
• computational neuroscience very poorly represented in psychiatry in the past (often not at all)
  -- partly due to nomenclature of psychiatric diseases based on qualitative concepts, incompletely tied to neuroscientific foundations

New hopes

• but this is changing.

• a new approach: seek a firmer foundation of the science of decision making

• pioneers : P. Dayan, Q. Huys, T. Braver., J. Cohen, M. Frank, S. Kapur, R. Montague, D. Pizzagali, K. Stephan, D. Steele, J. Williams, D. Redish and others ...

• “hope of a specific and quantitative anatomy of normal and abnormal function along with the prospect of rigorous tests for each underlying defect”.

The 4 Main Neuromodulators

dopamine

5HT

acetylcholine

norepinephrine
The 4 Main Neuromodulators: critically involved in Major Psychiatric Disease

- **Dopamine (DA)** involved in Parkinson’s, Schizophrenia, Addiction,

- **Serotonin (5HT)** involved in Depression, OCD, Eating disorders

- **Acetylcholine (ACh)** involved in Alzheimer’s Disease

- **Norepinephrine (NA)** involved in ADHD, Depression

Yet How Neuromodulation influences Neural Activity is very poorly understood.

Addiction

A chronically relapsing disorder that is characterised by:

(i) compulsive drug seeking and taking
(ii) inability to limit the intake of drugs
(iii) emergence of a withdrawal syndrome during cessation of drug taking

Goal of neuroscience: understand the cellular & molecular mechanisms that mediate transition between occasional controlled drug use and loss of behavioural control over drug seeking and taking

a promising field for modeling, building on models of decision making and reinforcement learning.

Drug Addiction as abnormal decision making

Yet How Neuromodulation influences Neural Activity is very poorly understood.

Systems involved: the reward system

- mesolimbic dopaminergic system - increase of dopamine release

- mesolimbic DA system: originates in the ventral tegmental area (VTA) of the midbrain, and projects to the nucleus accumbens (NA - ventral striatum). The amygdala (A), hippocampus (HC) and medial prefrontal cortex (PFC) send excitatory projections to the nucleus accumbens.

- drug seeking behaviour induced by glutamatergic projections from the prefrontal cortex to the NAc.
Why making a maladaptive choice over and over again? Theories of addiction

- In the past 30 years, lots of theories
- e.g.
  - compulsion zone: self administration is automatically induced when brain cocaine levels within a specific range.
  - set point model (or allostasis): goal = adjust sensitivity of brain reward system to set level, by increasing tonic dopamine
  - opponent process theory: drug addiction = result of emotional pairing between pleasure and symptoms of withdrawal. Motivation is first related to pleasure, and then to relief from withdrawal.
  - impulsivity.
- recently, addiction as a vulnerability in the decision process -- inspiration from reinforcement learning

Phasic dopamine signals prediction error

- the “largest success of computational neuroscience” [Niv]
  - Monkeys underwent simple instrumental or pavlovian conditioning
  - disappearance of dopaminergic response at reward delivery after learning, in VTA and SN.
  - if reward is not presented, response depression below basal firing at expected time of reward.

TD learning -- 101

- world is made of states, actions and rewards:
  • actions are selected so as to maximize future rewards.
  • states are associated with value functions defined as expected future reward
    \[ V(s) = \sum_{s'} P(s'|s) V(s') + \gamma R(s) \] (1)
  • Goal of TD learning: correctly learn the values. To do this, iteratively use the difference between expected and observed change in value -- the prediction error:
  \[ \delta = \gamma R(s) + V(s') - V(s) \] (2)
  • Value is then updated using:
    \[ V(s) \leftarrow V(s) + \eta \delta \]
  • Once the value correctly predicts the reward, learning stops.
  • a powerful learning algorithm in machine learning

Redish’s (Science, 2004) model

- Addiction as a Computational Process Gone Awry
  - A. David Redish
  - Addiction drugs have been hypothesized to access the same mesolimbic-dopaminergic reward learning systems. These neural learning systems can be studied through temporal difference reinforcement learning (TDRL), which requires a reward signal that has been hypothesized to be carried by dopamine. TDRL learns to predict reward by showing that reward-predicted signal is zero. By adding a transsial dopaminergic dopamine increase to a TDRL model, a computational model of addiction is constructed that connects actions leading to drug reward. The model predicts an exploration for reward that is too fast and too great, which provides a theoretic viewpoint with which to address other aspects.
  - cocaine and other drugs produce a transient increase in dopamine
  - idea: this dopamine surge induce an increase in prediction error \( \delta \) that can’t be compensated by changes in values.
  \[ \delta = \max \{ R(S) + V(S) \} \]
  \[ \delta = V(S) + D(S), D(S) \]

where \( D(S) \) indicates a dopamine surge occurring on entry into \( S \).

Consequence: values of states leading to the drug increase without bound.
Redish’s (2004) model

- Drug is hijacking the learning pathways, creating a prediction error where there should be none.

Gutkin, Dehaene & Changeux (PNAS, 2006) model of nicotine addiction

- A circuit model, 3 time scales
- Nicotine, through action on nAChRs in VTA, evokes phasic DA signal and changes the gain of DA signaling; potentiates DA transmission.
- The phasic DA instructs learning of action selection. Tonic DA gates this process.
- Slow onset opponent process decrease tonic DA neurotransmission to the point that extinction learning and response unlearning is impaired: routinized/ rigid behavior.

Redish’s (2004) model: predictions

- With repeated experience, drug choice become 1) less sensitive to alternative non-drug reinforcers [some evidence]; 2) more inelastic to costs [confirmed]

Models of Addiction: Conclusions

- Redish’s model, extensions and RL framework
  --> a new generation of models and model-driven experiments.

Lots of remaining challenges:
- Addiction to ordinary rewards such as fatty foods, which unlike cocaine produce a dopamine signal that can be accommodated
- Addiction to non-stimulant substances which depend less on mesolimbic dopamine (e.g., alcohol)
- Describing withdrawal symptoms -- opponent mechanisms
- Why do people want to get sober?
- Vulnerability: only a minority of people become addicted -- while other people can enjoy casual use, why?
Serotonin, Inhibition and Negative Mood
P. Dayan & Q. Huys (2008)

- Prediction of a sufficiently distant threat leads to inhibition, in the form of inhibition.
- Idea: inhibition is directly associated with aversive predictions.
- Prediction of a sufficiently distant threat leads to inhibition, in the form of withdrawal and disengagement (as in conditioned suppression).
- Serotonin - 5-HT:
  - role in normal and abnormal function still mysterious
  - involved in prediction of aversive events (opponent of dopamine which would be related to prediction of reward)
  - involved in behavioral inhibition
  - involved in models of depression and anxiety:
    i) depleting 5-HT by dietary depletion of precursor tryptophan can re-instate depression
    ii) selective serotonin re-uptake inhibitors (SSRIs) = antidepressant
    iii) but constitutive decreases in efficiency of 5HT re-uptake is a risk factor for depression.

- Model:
  - a model of trains of thoughts
  - belief = state
  - thought = change of belief = action
  - thoughts gain value through their connections with a group of terminal states O+/O- that are assigned + or - affective values

- 5-HT terminates trains of thought that have a negative value
**Model**

- O+ and O- (each with 100 elements) are associated with value \( r(s) \)
- I+ and I- (400 elements) are internal states
- sparse connections between states

- A fixed policy \( \pi^0 \) defined the transition probabilities from one state to the next.
- Internal states will acquire value through (TD reinforcement) learning.

**Idea**

- 5-HT terminates trains of thought that have a negative value
- Probability of continuing a train of thought depends on \( V(s) \)
  \[
  p_{5HT}(s) = \min(1, \exp(\alpha_{5HT} V(s)))
  \]
- When thoughts are terminated, they stop and restart randomly in I+ or I-.
- Consequence: the more the 5HT the less the 'negative' states are explored -- sampling bias

**Values after learning**

- \( \alpha_{5HT} = 0 \)
- \( \alpha_{5HT} = 20 \)

- 5HT is favorable - enhanced average rewards
- but values are overly optimistic and errors for aversive chains (overvalued)

**Serotonin (via Tryptophan) depletion after learning**

- after learning, switching \( \alpha_{5HT} = 20 \) to \(<20\)
- suddenly more negative states become explored
- \( \rightarrow \) more negative average affective outcome
- surprises (errors) associated with transitions that were previously inhibited

---

![Diagram showing transition probabilities and values after learning](image)

![Graph showing probability of continuing a train of thought](image)

![Graphs showing values and terminal bias after learning](image)
**Decision-making Priors**

**Optimism: a prior on the likelihood of future reward?**

- "Optimism: the extent to which people hold generalised favourable expectancies for the future"

- the LOT-R questionnaire.

1. In uncertain times, I usually expect the best
2. It's easy for me to relax
3. If something can go wrong for me it will
4. I'm always optimistic about my future
5. I enjoy my friends a lot
6. It's important for me to keep busy
7. I hardly ever expect things to go my way
8. I don't get upset too easily

**Discussion**

- 5HT is favorable - enhanced average rewards
- but values are overly optimistic and errors for aversive chains
- consistent with the fact that 5-HT suppression leads to impulsivity (choosing states that would not be selected otherwise)
- consistent with the idea that 5-HT is related to prediction of aversive outcomes
- consistent with the fact that 5-HT depletion after learning leads to depressive symptoms.
- predictions: 5-HT levels during learning would control the extent to which negative states are explored / learned.
- dopamine and serotonin: mutual opposition model. serotonin proposed to report negative prediction errors

**Discussion**

- 5HT is favorable - enhanced average rewards
- but values are overly optimistic and errors for aversive chains

**Decision-making Priors**

**Optimism: a prior on the likelihood of future reward?**

- "Optimism: the extent to which people hold generalised favourable expectancies for the future"

- the LOT-R questionnaire.

1. In uncertain times, I usually expect the best
2. It's easy for me to relax
3. If something can go wrong for me it will
4. I'm always optimistic about my future
5. I enjoy my friends a lot
6. It's important for me to keep busy
7. I hardly ever expect things to go my way
8. I don't get upset too easily

- decision making in rats using adapted version of Iowa Gambling task.
- large inter-individual differences
- poor decision making results from hypersensitivity to reward and higher risk taking
- TD modeling
Questions

- Are people usually biased in estimating probability of future reward?
- Is this bias correlated with the LOT-R score?
- Can this bias be described as a Bayesian Prior?

[Stankevicius, Kalra, Huys, Seriès, Plos Comp Biol. in press]

Bayesian Model

- Assume subjects are Bayesian Optimal. Based on observed data D and their prior belief, they form the posterior p(c|D).

\[
p(c|D) = \frac{p(D|c)p(c)}{p(D)}
\]

- Parametrisation of prior p(c): Beta distribution, parameters α and β.

- Subjects form estimate of c using the mean of the posterior

\[
\hat{c} = \int_0^1 c p(c|D) dc, \quad \hat{c} = \frac{\hat{\alpha} + c}{\hat{\alpha} + \hat{\beta}}
\]

- Subjects make decision based on comparing c with b = probability of reward of square.

\[
p(\text{choose fractal}) = \frac{\exp(\gamma\hat{c})}{\exp(\gamma\hat{c}) + \exp(\gamma b)}
\]

- Each subject is described by 3 parameters (α, β, γ).

Optimists overestimate probability of future reward

- 51 subjects
- People show significant bias in estimation of probability of future reward.
- ML estimation --> estimation of the prior for each participant.
- LOT-R scores correlate with mean of the prior (r=0.438, p<0.001).
The idea of a continuum between health and disease

- decision making in rats
- There is a variability of performance comparable to that in humans -- extreme behaviour could correspond to disorder
- no need for dedicated animal model
- electrophysiology

**Dimensional Analysis of ADHD Subtypes in Rats**

Candice Blondeau and Françoise Della-Hagedorn

**Background:** Attention-deficit/hyperactivity disorder (ADHD) is a heterogeneous disorder that is classified into three subtypes in which the main symptoms, inattention, hyperactivity, and impulsivity, are expressed with various degrees of severity. The nature of the biological underpinnings explaining such subtypes (common or distinct) is unknown, and animal models encompassing different subtypes are needed.

**Methods:** A cluster analysis segregated subgroups of rats on the basis of similarities in both impulsivity and attentional scores in the five-choice serial reaction time task. These subgroups were characterized behaviorally and were compared for several aspects of spontaneous attention sustaining each subtype (common or distinct) is unknown, and animal models encompassing different subtypes are needed. A cluster analysis segregated subgroups of rats on the basis of similarities in both impulsivity and attentional scores in the five-choice serial reaction time task. These subgroups were characterized behaviorally and were compared for several aspects of spontaneous attention sustaining each subtype (common or distinct) is unknown, and animal models encompassing different subtypes are needed.

**Results:** Four distinct subgroups were demonstrated: efficient, middle, inattentive, and inattentive–impulsive. Hyperactivity expressed in the efficient subgroup and on accuracy. Clinical and neuropsychologic indicators of ADHD addressing the main symptoms and reflecting distinct subtypes of the disorder could be of great interest.

**Conclusion:** This new approach is the first to demonstrate behavioral subtypes in rats that parallel those observed in human beings and proposes that common mental disorders can be conceptualized along a continuum of the processes involved. Behavioral traits not encompass the dimensional aspects (differences in degree of similarity) of the processes involved. Behavioral traits not encompass the dimensional aspects (differences in degree of similarity) of the processes involved. Behavioral traits not encompass the dimensional aspects (differences in degree of similarity) of the processes involved. Behavioral traits not encompass the dimensional aspects (differences in degree of similarity) of the processes involved.

**Key Words:** ADHD, attention, cluster analysis, hyperactivity, impulsivity, methylphenidate.

A new approach: seek a firmer foundation of the science of decision making

- new hopes
- “hope of a specific and quantitative anatomy of normal and abnormal function along with the prospect of rigorous tests for each underlying defect”.
- interesting times.