To collect reward or to defend homeostasis?

Boris Gutkin

Group for Neural Theory, ENS, Paris, France
Center for Cognition and Decision Making, Higher School of Economics, Russia
To collect reward or to defend homeostasis?

Needy RL: An integrated hypothalamo-cortico-striatal Circuit collects Rewards and defends homeostasis

Boris Gutkin

Group for Neural Theory, ENS, Paris, France
Center for Cognition and Decision Making, Higher School of Economics, Russia

24 September 2015
To collect reward or to defend homeostasis?

Needy RL: An integrated hypothalamo-cortico-striatal Circuit collects Rewards and defends homeostasis

Boris Gutkin

Group for Neural Theory, ENS, Paris, France
Center for Cognition and Decision Making, Higher School of Economics, Russia

Dr. Mehdi Keramati (UCL)

24 September 2015
To collect reward or to defend homeostasis?

Needy RL: An integrated hypothalamo-cortico-striatal Circuit collects Rewards and defends homeostasis

Boris Gutkin

Group for Neural Theory, ENS, Paris, France
Center for Cognition and Decision Making, Higher School of Economics, Russia

Dr. Mehdi Keramati (UCL)

24 September 2015
Two critical processes:

Reinforcement Learning System ⇔ External World

Homeostatic Regulation System ⇔ Internal World
Two critical processes:

Reinforcement Learning System ⇔ External World

Homeostatic Regulation System ⇔ Internal World
Reinforcement Learning:

Learning Value of Each Choice

\[ Q(s_1, a_1) \]
\[ Q(s_1, a_2) \]
\[ Q(s_1, a_3) \]
Reinforcement Learning:

Learning Value of Each Choice

\[ Q(s_1, a_1) \]
\[ Q(s_1, a_2) \]
\[ Q(s_1, a_3) \]
Reinforcement Learning:

Learning Value of Each Choice

Prediction error: \[ \delta = r_1 + V(s_2) - Q(s_1, a_1) \]

Updating estimates: \[ Q(s_1, a_1) \leftarrow Q(s_1, a_1) + \alpha \delta \]
Reinforcement Learning:

Learning Value of Each Choice

Prediction error: \[ \delta = r_1 + V(s_2) - Q(s_1, a_1) \]

Updating estimates: \[ Q(s_1, a_1) \leftarrow Q(s_1, a_1) + \alpha \delta \]
Reinforcement Learning, maximizes reward.

\[ \delta = r_1 + V(s_2) - Q(s_1, a_1) \]

\[ Q(s_1, a_1) \leftarrow Q(s_1, a_1) + \alpha \delta \]
Reinforcement Learning, maximizes reward.

Prediction error: \[ \delta = r_1 + V(s_2) - Q(s_1, a_1) \]

Updating estimates: \[ Q(s_1, a_1) \leftarrow Q(s_1, a_1) + \alpha \cdot \delta \]
Reinforcement Learning, maximizes reward.

Open issue: What is reward? What is the role of the internal state?

Prediction error: \[ \delta = r_1 + V(s_2) - Q(s_1, a_1) \]

Updating estimates: \[ Q(s_1, a_1) \leftarrow Q(s_1, a_1) + \alpha \delta \]
Two critical processes:

Reinforcement Learning System ↔ External World

Homeostatic Regulation System ↔ Internal World
Homeostatic Regulation:

Two facts:

- Animal’s behavior is variable in a fixed external environment.
- Some physiological variables have stable values, despite changes in their input and output.
  
  Blood glucose, body temperature, body weight, blood volume, plasma sodium, plasma calcium, ...
Homeostatic Regulation:

Two facts:

- Animal’s behavior is variable in a fixed external environment.
- Some physiological variables have stable values, despite changes in their input and output.
  
  Blood glucose, body temperature, body weight, blood volume, plasma sodium, plasma calcium, ...
Homeostatic Regulation:

Two facts:

- Animal’s behavior is variable in a fixed external environment.
- Some physiological variables have stable values, despite changes in their input and output.

Blood glucose, body temperature, body weight, blood volume, plasma sodium, plasma calcium, ...
Homeostatic Regulation: minimizes deviation.

Two facts:

- Animal’s behavior is variable in a fixed external environment.
- Some physiological variables have stable values, despite changes in their input and output.

  Blood glucose, body temperature, body weight, blood volume, plasma sodium, plasma calcium, ...
Homeostatic Regulation: minimizes deviation.

Goals:

- ✔ What rewards to seek: food vs. water
- ✗ How to navigate in the complex environment to achieve the rewards

Blood glucose, body temperature, body weight, blood volume, plasma sodium, plasma calcium, ...
Homeostatic Regulation: minimizes deviation.

Open issue: How should the animal translate its physiological deficits into appropriate instrumental behaviors.

Goals:

☑️ What rewards to seek: food vs. water

☒ How to navigate in the complex environment to achieve the rewards
<table>
<thead>
<tr>
<th>Feature/Theory</th>
<th>Homeostatic Regulation</th>
<th>Reinforcement Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioral explanation</td>
<td>Motivation</td>
<td>Learning</td>
</tr>
<tr>
<td>Neural substrates</td>
<td>Hypothalamic nuclei</td>
<td>Cortico-basal ganglia and</td>
</tr>
<tr>
<td></td>
<td>Internal states</td>
<td>dopaminergic system</td>
</tr>
<tr>
<td>Behavioral adaptation to</td>
<td>Deviation minimization</td>
<td>External cues</td>
</tr>
<tr>
<td>hypothetical assumption</td>
<td></td>
<td>Reward maximization</td>
</tr>
</tbody>
</table>
### Open Issues with RL models:
- Internal state is not taken into account.

### Open Issues with negative-feedback models:
- The complexity and dynamic structure of the external world is ignored.

<table>
<thead>
<tr>
<th>Feature/Theory</th>
<th>Homeostatic Regulation</th>
<th>Reinforcement Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioral explanation</td>
<td>Motivation</td>
<td>Learning</td>
</tr>
<tr>
<td>Neural substrates</td>
<td>Hypothalamic nuclei</td>
<td>Cortico-basal ganglia and dopaminergic system</td>
</tr>
<tr>
<td>Behavioral adaptation to</td>
<td>Internal states</td>
<td>External cues</td>
</tr>
<tr>
<td>hypothetical assumption</td>
<td>Deviation minimization</td>
<td>Reward maximization</td>
</tr>
</tbody>
</table>
• Open Issues with RL models:
  – Internal state is not taken into account.

• Open Issues with negative-feedback models:
  – The complexity and dynamic structure of the external world is ignored.
Reward <-> Homeostasis Interact

Trends in Neuroscience
Palmiter, 2007
Reward <-> Homeostasis Interact

Trends in Neuroscience
Palmiter, 2007
Reward <-> Homeostasis Interact

Trends in Neuroscience
Palmiter, 2007
Reward <-> Homeostasis Interact

\[ d(H_t) = \sqrt{\sum_{i=1}^{N} |h_i^* - h_{i,t}|^n} \]

Trends in Neuroscience
Palmiter, 2007
Reward <-> Homeostasis Interact

**Reward:**
\[ r(H_t, K_t) = d(H_t) - d(H_{t+1}) = d(H_t) - d(H_t + K_t) \]

**Drive:**
\[ d(H_t) = \left( \sum_{i=1}^{N} |h_i^* - h_{i,t}| \right)^n \]

---

Trends in Neuroscience
Palmiter, 2007
Reward <-> Homeostasis Interact

**Prediction error:** \[ \delta = r_t + Q(s_t, a_t) - V(s_{t+1}) \]

**Reward:** \[ r(H_t, K_t) = d(H_t) - d(H_{t+1}) \]
\[ = d(H_t) - d(H_t + K_t) \]

**Drive:** \[ d(H_t) = \sqrt{\sum_{i=1}^{N} |h_i^* - h_{i,t}|^n} \]

Trends in Neuroscience
Palmiter, 2007
Updating values: \[ Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \cdot \delta \]

Prediction error: \[ \delta = r_t + Q(s_t, a_t) - V(s_{t+1}) \]

Reward: \[ r(H_t, K_t) = d(H_t) - d(H_{t+1}) = d(H_t) - d(H_t + K_t) \]

Drive: \[ d(H_t) = \sum_{i=1}^{N} |h_i^* - h_{i,t}|^n \]

Reward <-> Homeostasis Interact

Trends in Neuroscience
Palmiter, 2007
Homeostatic Reinforcement Learning:

Reward <-> Homeostasis Interact

Updating values:
$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \cdot \delta$$

Prediction error:
$$\delta = r_t + Q(s_t, a_t) - V(s_{t+1})$$

Reward:
$$r(H_t, K_t) = d(H_t) - d(H_{t+1}) = d(H_t) - d(H_t + K_t)$$

Drive:
$$d(H_t) = \sum_{i=1}^{N} |h_i^* - h_{i,t}|^n$$
Homeostatic Reinforcement Learning:

- **$K_i$** is the homeostatic outcome of an action
- **$H_i$** is the homeostatic deviation
- **$h_i$** is location in the homeostatic space

Updating values: $Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \cdot \delta$

Prediction error: $\delta = r_t + Q(s_t, a_t) - V(s_{t+1})$

Reward: $r(H_t, K_t) = d(H_t) - d(H_{t+1})$

Drive: $d(H_t) = \sqrt{\sum_{i=1}^{N} |h_i^* - h_{i,t}|^n}$
Homeostatic Reinforcement Learning:

Primary Reward Definition: reduction in drive (an action is rewarding if its outcome reduces homeostatic deviation)

Value Modulation by internal state:

$$Q_1(s,a) = \frac{d(H_1)}{Q_1(H_0)}Q_0(s,a)$$

Updating values:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \cdot \delta$$

Prediction error:

$$\delta = r_t + Q(s_t, a_t) - V(s_{t+1})$$

Reward:

$$r(H_t, K_t) = d(H_t) - d(H_{t+1})$$

$$= d(H_t) - d(H_t + K_t)$$

Drive:

$$d(H_t) = \sqrt{\sum_{i=1}^{N} |h_i^* - h_{i,t}|^n}$$
Normative Theory:

Defending Physiological Stability is RATIONAL!

Obtaining reward $\approx$ Defending homeostasis

Any algorithm that maximizes sum of discounted rewards (SDR), also minimizes the sum of discounted drives (SDD), and vice versa.
Normative Theory:

Defending Physiological Stability is RATIONAL!
Obtaining reward $\approx$ Defending homeostasis

Any algorithm that maximizes sum of discounted rewards (SDR), also
minimizes the sum of discounted drives (SDD), and vice versa.

$\mathcal{P}(H_0)$ is the set of all pathways start at $H_0$ and end up at $H^*$.

$$SDD_p(H_0) = \sum_{t=0}^{n-1} \gamma^t d(H_{t+1})$$

$$SDR_p(H_0) = \sum_{t=0}^{n-1} \gamma^t r_t = \sum_{t=0}^{n-1} \gamma^t (d(H_t) - d(H_{t+1}))$$

If $\gamma < 1$, \( \arg\min_{p \in \mathcal{P}(H_0)} SDD_p(H_0) = \arg\max_{p \in \mathcal{P}(H_0)} SDR_p(H_0) \)
Normative Theory:
Defending Physiological Stability is RATIONAL!

\( \mathcal{P}(H_0) \) is the set of all pathways start at \( H_0 \) and end up at \( H^* \).

\[
SDD_p(H_0) = \sum_{t=0}^{n-1} \gamma^t d(H_{t+1}) \\
SDR_p(H_0) = \sum_{t=0}^{n-1} \gamma^t r_t = \sum_{t=0}^{n-1} \gamma^t (d(H_t) - d(H_{t+1}))
\]

**Proposition:**

\[ \text{if } \gamma < 1, \quad \arg\min_{p \in \mathcal{P}(H_0)} SDD_p(H_0) = \arg\max_{p \in \mathcal{P}(H_0)} SDR_p(H_0) \]

**Proof (sketch):**

\[
SDD_{p_i}(H_0) = d_{i,1} + \gamma d_{i,2} + \gamma^2 d_{i,3} + \ldots + \gamma^{n-2} d_{i,n-1} + \gamma^{n-1} d^* \\
SDR_{p_i}(H_0) = r_{i,0} + \gamma r_{i,1} + \gamma^2 r_{i,2} + \ldots + \gamma^{n-1} r_{i,n-1}
\]
Normative Theory:

Defending Physiological Stability is RATIONAL!

\[ \mathcal{P}(H_0) \] is the set of all pathways start at \( H_0 \) and end up at \( H^* \).

\[ SDD_p(H_0) = \sum_{t=0}^{n-1} \gamma^t d(H_{t+1}) \quad SDR_p(H_0) = \sum_{t=0}^{n-1} \gamma^t r_t = \sum_{t=0}^{n-1} \gamma^t (d(H_t) - d(H_{t+1})) \]

Proposition:

If \( \gamma < 1 \),

\[ \arg\min_{p \in \mathcal{P}(H_0)} SDD_p(H_0) = \arg\max_{p \in \mathcal{P}(H_0)} SDR_p(H_0) \]

Proof (sketch):

\[ SDD_{p_i}(H_0) = d_{i,1} + \gamma d_{i,2} + \gamma^2 d_{i,3} + \ldots + \gamma^{n-2} d_{i,n-1} + \gamma^{n-1} d^* \]

\[ SDR_{p_i}(H_0) = r_{i,0} + \gamma r_{i,1} + \gamma^2 r_{i,2} + \ldots + \gamma^{n-1} r_{i,n-1} \]

\[ = (d_0 - d_{i,1}) + \gamma (d_{i,1} - d_{i,2}) + \gamma^2 (d_{i,2} - d_{i,3}) + \ldots + \gamma^{n-1} (d_{i,n-1} - d^*) \]
Normative Theory:

Defending Physiological Stability is RATIONAL!

\[ \mathcal{P}(H_0) \text{ is the set of all pathways start at } H_o \text{ and end up at } H^*. \]

\[ SDD_p(H_0) = \sum_{t=0}^{n-1} \gamma^t d(H_{t+1}) \quad \text{ and } \quad SDR_p(H_0) = \sum_{t=0}^{n-1} \gamma^t r_t = \sum_{t=0}^{n-1} \gamma^t (d(H_t) - d(H_{t+1})) \]

**Proposition:**

if \( \gamma < 1 \),

\[ \arg \min_{p \in \mathcal{P}(H_0)} SDD_p(H_0) = \arg \max_{p \in \mathcal{P}(H_0)} SDR_p(H_0) \]

**Proof (sketch):**

\[ SDD_{p_i}(H_0) = d_{i,1} + \gamma d_{i,2} + \gamma^2 d_{i,3} + \ldots + \gamma^{n-2} d_{i,n-1} + \gamma^{n-1} d^* \]

\[ SDR_{p_i}(H_0) = r_{i,0} + \gamma r_{i,1} + \gamma^2 r_{i,2} + \ldots + \gamma^{n-1} r_{i,n-1} \]

\[ = (d_0 - d_{i,1}) + \gamma(d_{i,1} - d_{i,2}) + \gamma^2(d_{i,2} - d_{i,3}) + \ldots + \gamma^{n-1}(d_{i,n-1} - d^*) \]

\[ = d_0 + (\gamma - 1)(d_{i,1} + \gamma d_{i,2} + \gamma^2 d_{i,3} + \ldots + \gamma^{n-2} d_{i,n-1}) \]

\[ = d_0 + (\gamma - 1) SDD_{p_i}(H_0) \]
Normative Theory:

**Defending Physiological Stability is RATIONAL!**

\[ P(H_0) \text{ is the set of all pathways start at } H_o \text{ and end up at } H^*. \]

\[
SDD_p(H_0) = \sum_{t=0}^{n-1} \gamma^t d(H_{t+1}) \\
SDR_p(H_0) = \sum_{t=0}^{n-1} \gamma^t r_t = \sum_{t=0}^{n-1} \gamma^t (d(H_t) - d(H_{t+1}))
\]

Proposition:

if \( \gamma < 1 \)

\[
\arg\min_{p \in P(H_0)} SDD_p(H_0) = \arg\max_{p \in P(H_0)} SDR_p(H_0)
\]

Normative Discounting: **Discounting is key to Physiological Stability thru reward maximization**
Anticipatory Responding:

Animals make anticipatory responses to preclude perturbations in regulated variables, even before any physiological depletion (negative feedback) is detectable:

- Anticipatory eating, drinking
- Salivation in response to food-associated stimuli
- Insulin secretion prior to meal initiation
- Anticipatory shivering
Anticipatory Responding:

Animals make anticipatory responses to preclude perturbations in regulated variables, even before any physiological depletion (negative feedback) is detectable:

- Anticipatory eating, drinking
- Salivation in response to food-associated stimuli
- Insulin secretion prior to meal initiation
- Anticipatory shivering
Anticipatory Shivering:

$P = 0.9$

$P = 0.1$
Anticipatory Shivering:

- Shivering
- No-shivering

Trial (#visiting the normal state)

Trial (#visiting the cold state)
Anticipatory Shivering:

\[ p = 0.9 \]

\[ p = 0.1 \]
Anticipatory Shivering:

- Shivering
- No-shivering

Choice probability in the normal state

Choice probability in the cued state

Choice probability in the cold state

Internal temperature (arbitrary unit)

Trial (#visiting the normal state)

Trial (#visiting the cued state)

Trial (#visiting the cold state)

Trial (#visiting any state)
Anticipatory Shivering:

\[ p = 0.9 \]

\[ p = 0.1 \]
Alcohol Tolerance is Learned:
Alcohol Tolerance is Learned:

Alcohol Tolerance is Learned:

Alcohol Tolerance is Learned:

Alcohol Tolerance is Learned:

Alcohol Tolerance is Learned:

Alcohol Tolerance is Learned:

![Drosophila melanogaster](image)

- Background
- Reinforcement Learning
- Homeostatic Regulation
- Homeostatic Reinforcement Learning
- Anticipatory Responding
- Alcohol Tolerance
- Further explanations
- Future Directions
- Conclusion

Alcohol Tolerance is Learned:

J. G. Mansfield, C. L. Cunningham, "Journal of Comparative and Physiological Psychology (1980)."
Anticipatory Responding:

Response is conditioned
Explains alcohol tolerance

Predictions:
Tolerance response preceding deviation (e.g. alcohol injections)
Animals are capable of learning not only Pavlovian, but also instrumental anticipatory responding

(As opposed to the predictive homeostasis theory [Sterling]).
Further explanations by the model:

- Reward value increases with the dose of outcome
- Excitatory effect of deprivation level on the rewarding value of outcome
- Inhibitory effect of the irrelevant drive
- Risk aversion
For all $m > n > 2$, the rewarding Value of $K_t = (0, 0, ..., k_{j,t}, .., 0)$:

**Reward value increases with the dose of outcome**

$$\frac{dr(H_t, K_t)}{dk_{j,t}} > 0 : \text{ for } k_{j,t} > 0$$

- Higher breakpoint in progressive ratio schedules, as the outcome gets bigger.
Higher breakpoint in progressive ratio schedules, as deprivation level increases.
For all, the rewarding Value of $K_t = (0, 0, \ldots, k_{j,t}, \ldots, 0)$:

Inhibitory effect of the irrelevant drive

$$\frac{dr(H_t, K_t)}{d|h_i^* - h_{i,t}|} < 0 \quad : \quad \text{for all } i \neq j \text{ and } k_{j,t} > 0$$
Thirst impairs Pavlovian responses for food, as well as instrumental responses for food during both acquisition and extinction.

Reciprocally, food deprivation suppresses Pavlovian and instrumental water-related responses.

Increased calcium appetite reduces appetite for phosphorus.

Increased level of hunger inhibits sexual behavior.
For all \( m > n > 2 \), the rewarding Value of \( K_t = (0, 0, \ldots, k_{j,t}, \ldots, 0) \):

\[
d(H_t) = \sqrt[n]{\sum_{i=1}^{N} |h_i^* - h_{i,t}|^n}
\]

\[
r(H_t, K_t) = d(H_t) - d(H_{t+1})
\]

Risk Aversion

\[
\frac{d^2 r(H_t, K_t)}{dk_{j,t}^2} < 0 : \text{ for } k_{j,t} > 0
\]
For all \( m > n > 2 \), the rewarding Value of \( K_t = (0, 0, \ldots, k_{j,t}, \ldots, 0) \):

**Risk Aversion**

\[
\frac{d^2 r(H_t, K_t)}{d k_{j,t}^2} < 0 : \text{ for } k_{j,t} > 0
\]

\[
U = r((x_t, y_t), (k_X, 0))
\]

\[
\frac{1}{2} r(10) + \frac{1}{2} r(30)
\]

- **certain choice**: 100%
- **risky choice**: 50%
Risk aversion

2 units of energy with $p=1$

8 units of energy with $p=0.25$

2 units of energy

Small-outcome state
Big-outcome state

Place preference

Internal state

Trial

Background □ Reinforcement Learning □ Homeostatic Regulation □ Homeostatic Reinforcement Learning □ Anticipatory Responding □ Alcohol Tolerance □ Further explanations □ Future Directions □ Conclusion
Experimental Paradigms on feeding behavior:

**Food-seeking:**
The motivational state is tried to be kept fixed
(or there are only two motivational states).

Question: Behavioral adaptation to external cues

Analytical framework: Reinforcement Learning Models

**Food-taking:**
Under ad-libitum access to food.
No learning is involved.

Question: Behavioral adaptation to the internal state. Meal patterns.

Analytical framework: Negative feedback Models

**Our framework:**
Allows for analyzing experiments where internal and external adaptations are simultaneously involved.
Behavioral control of the internal state:

1 unit of energy → 4 units of energy

2 units of energy

\[a \rightarrow s \rightarrow a' \rightarrow s \rightarrow a\]

Graphs showing place preference and internal state over trials.
Human version:

B eing tested by Oliver Hulme, Kopenhagen
Incorporating Sensory properties of food:

\[ r(H_t, K_t) = d(H_t) - d(H_{t+1}) = d(H_t) - d(H_t + K_t) \]
Incorporating Sensory properties of food:

\[
r(H_t, K_t) = d(H_t) - d(H_{t+1}) = d(H_t) - d(H_t + K_t)
\]

Sensory properties of food outcome give an estimate, \( \hat{k}_X \), of its true nutritional content, \( k_X \).
Incorporating Sensory properties of food:

$$r(H_t, K_t) = d(H_t) - d(H_{t+1})$$

$$= D(H_t) - D(H_t + \hat{K}_t)$$

Sensory properties of food outcome gives an estimate, $\hat{k}_X$, of its true nutritional content, $k_X$. 

Trends in Neuroscience
Palmiter, 2007
Incorporating Sensory properties of food:

\[ r(H_t, K_t) = d(H_t) - d(H_{t+1}) \]
\[ = D(H_t) - D(H_t + \hat{K}_t) \]
\[ H_{t+1} = H_t + K_t. \]

Resolves problems with the basic model:
- Dopamine neurons show **instantaneous** burst activity in response to unexpected food rewards.
- Intragastric intubation and intravenous injection of food is not rewarding.
- Palatable foods have reinforcing effect, even when they do not have any nutritional value.
**Assumption:**

Highly-palatable foods escape homeostatic constraints as a result of the inability of internal satiety signals in blocking the opioid-based stimulation of DA neurons.

\[ r(H_t, K_t) = d(H_t) - d(H_t + K_t) + T \]

- \( T > 0 \)
- \( T = 0 \)
Assumption:
Highly-palatable foods escape homeostatic constraints as a result of the inability of internal satiety signals in blocking the opioid-based stimulation of DA neurons.

\[ r(H_t, K_t) = d(H_t) - d(H_t + K_t) + T \]
Some future directions:

- **Allsotatsis: adaptive homeostatic setpoint**
  - Reward system modulating the homeostatic system.
  - Explaining obesity
  - Explaining drug-addiction
- **Non-symmetric drive functions**
  - Unequal impact of need modalities
  - Line-attractors in drive function
- **Physiological/Social interactions**
  - Agent modelling
**Conclusion:**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Homeostatic Regulation</th>
<th>Reinforcement Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioral explanation</td>
<td>Motivation</td>
<td>Learning</td>
</tr>
<tr>
<td>Neural substrates</td>
<td>Hypothalamic nuclei</td>
<td>Cortico-basal ganglia and dopaminergic system</td>
</tr>
<tr>
<td>Behavioral adaptation to hypothetical assumption</td>
<td>Internal states</td>
<td>External cues</td>
</tr>
<tr>
<td>hypothetical assumption</td>
<td>Deviation minimization</td>
<td>Reward maximization</td>
</tr>
</tbody>
</table>

**Homeostatic Reinforcement Learning**
... you can't always get what you want
But if you try sometime,
you just might find
You get what you need

The Rolling Stones
## Conclusion:

<table>
<thead>
<tr>
<th>Feature\Theory</th>
<th>Homeostatic Regulation</th>
<th>Reinforcement Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioral explanation</td>
<td>Motivation</td>
<td>Learning</td>
</tr>
<tr>
<td>Neural substrates</td>
<td>Hypothalamic nuclei</td>
<td>Cortico-basal ganglia and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dopaminergic system</td>
</tr>
<tr>
<td>Behavioral adaptation to</td>
<td>Internal states</td>
<td>External cues</td>
</tr>
<tr>
<td>hypothetical assumption</td>
<td>Deviation minimization</td>
<td>Reward maximization</td>
</tr>
</tbody>
</table>

**Homeostatic Reinforcement Learning**
\[ d(H_t) = \sqrt[m]{\sum_{i=1}^{N} |h_i^* - h_{i,t}|^n} \]
\[ r(H_t, K_t) = d(H_t) - d(H_{t+1}) \]

For all \( m > n > 2 \), the rewarding value of \( K_t = (0, 0, \ldots, k_{j,t}, \ldots, 0) \):

**Reward value increases with the dose of outcome**

\[ \frac{dr(H_t, K_t)}{dk_{j,t}} > 0 \quad : \quad \text{for } k_{j,t} > 0 \]

- Higher breakpoint in progressive ratio schedules, as the outcome gets bigger.
\[ d(H_t) = \sqrt[\í]{\sum_{i=1}^{N} |h^*_i - h_{i,t}|^n} \]

\[ r(H_t, K_t) = d(H_t) - d(H_{t+1}) \]

For all \( m > n > 2 \), the rewarding Value of \( K_t = (0, 0, .., k_{j,t}, .., 0) \):

**Excitatory effect of deprivation level**

\[ \frac{dr(H_t, K_t)}{d|h^*_j - h_{j,t}|} > 0 : \text{ for } k_{j,t} > 0 \]

- Higher breakpoint in progressive ratio schedules, as deprivation level increases.
For all $m > n > 2$, the rewarding value of $K_t = (0, 0, \ldots, k_{j,t}, \ldots, 0)$:

\[
d(H_t) = \sqrt[n]{\sum_{i=1}^{N} |h_i^* - h_{i,t}|^n}
\]
\[
r(H_t, K_t) = d(H_t) - d(H_{t+1})
\]

Inhibitory effect of the irrelevant drive

\[
\frac{dr(H_t, K_t)}{d|h_i^* - h_{i,t}|} < 0 : \text{ for all } i \neq j \text{ and } k_{j,t} > 0
\]
Thirst impairs Pavlovian responses for food, as well as instrumental responses for food during both acquisition and extinction.

Reciprocally, food deprivation suppresses Pavlovian and instrumental water-related responses.

Increased calcium appetite reduces appetite for phosphorus.

Increased level of hunger inhibits sexual behavior.

...
For all \( m > n > 2 \), the rewarding

Value of \( K_t = (0, 0, \ldots, k_{j,t}, \ldots, 0) \):

**Risk Aversion**

\[
\frac{d^2 r(H_t, K_t)}{dk_{j,t}^2} < 0 : \text{ for } k_{j,t} > 0
\]
For all $m > n > 2$, the rewarding Value of $K_t = (0, 0, \ldots, k_{j,t}, \ldots, 0)$:

**Risk Aversion**

$$\frac{d^2 r(H_t, K_t)}{dk_{j,t}^2} < 0 : \text{ for } k_{j,t} > 0$$

$$U = r((x_t, y_t), (k_X, 0))$$

$\frac{1}{2} r(10) + \frac{1}{2} r(30)$

$r(30)$

$r(20)$

$r(10)$

$100\%$  $50\%$  $50\%$
Risk aversion

2 unit of Energy with \( p=1 \)

8 units of Energy with \( p=0.25 \)

2 units of energy

![Diagram showing risk aversion with two outcomes and preference changes over trials.](image)

![Graph showing place preference and internal state changes over trials.](image)
Experimental Paradigms on feeding behavior:

**Food-seeking:**
The motivational state is tried to be kept fixed
(or there are only two motivational states).

Question: Behavioral adaptation to external cues

Analytical framework: Reinforcement Learning Models

**Food-taking:**
Under ad-libitum access to food.
No learning is involved.

Question: Behavioral adaptation to the internal state. Meal patterns.

Analytical framework: Negative feedback Models

**Our framework:**
Allows for analyzing experiments where internal and external adaptations are simultaneously involved.
Behavioral control of the internal state:

1 unit of energy 4 units of energy

2 units of energy

Internal state

Place preference

Small-outcome state Big-outcome state

Trial

Internal state

0 100 200 300

0 0.2 0.4 0.6 0.8 1.0

0 100 200 300

-60 -40 -20 0 20
Human version:

- Choice
  - Left
    - 80%: Small outcome
    - 20%: Big outcome
  - Right
    - 80%: Big outcome
    - 20%: Small outcome

- Internal state

- Action probability
  - Left: Red
  - Right: Blue

- Internal state over time
  - Time: 0 to 300
  - Internal state: -60 to 20

- Background
- Reward as Drive Reduction
- Regulation + Learning
- Anticipatory Responses
- Behavioral Plausibility
- Energy Regulation
- Overeating
- Orexin as a Candidate
- Conclusion
Food vs. Money:

- **Choice**
  - Left (Food)
  - Right (Money)

- **Internal State Action Probability**

- **Graphs**
  - Action probability over time for big money trials (Left and Right)
  - Internal state over time
Assumption:

Sensory properties of a food outcome gives an estimate, $\hat{k}_X$, of its true nutritional content, $k_X$. 

$$r(H_t, K_t) = d(H_t) - d(H_{t+1}) = d(H_t) - d(H_t + K_t)$$

Mechanism:

- Endogenous opioids that signal palatability, and stimulate DA neurons through inhibiting GABAergic inhibitory signaling onto the DA cells.

Resolves problems with the basic model:

- Dopamine neurons show *instantaneous* burst activity in response to unexpected food rewards.
- Intragastric intubation and intravenous injection of food is not rewarding.
- Palatable foods have reinforcing effect, even when they do not have any nutritional value.
Assumption:

Highly-palatable foods escape homeostatic constraints as a result of the inability of internal satiety signals in blocking the opioid-based stimulation of DA neurons.

\[ r(H_t, K_t) = d(H_t) - d(H_t + K_t) + T \]
Overeating:

\[ r(H_t, K_t) = d(H_t) - d(H_t + K_t) + T \]

- Background
- Reward as Drive Reduction
- Regulation + Learning
- Anticipatory Responses
- Behavioral Plausibility
- Energy Regulation
- Overeating
- Orexin as a Candidate
- Conclusion
Overeating:

- Background
- Reward as Drive Reduction
- Regulation + Learning
- Anticipatory Responses
- Behavioral Plausibility
- Energy Regulation
- Overeating
- Orexin as a Candidate
- Conclusion

Graphs showing choice probabilities and internal state over trials.
• Orexin neurons:
  – Orexin neurons are responsive to peripheral metabolic signals (Leptin, insulin, glucose, etc), as well as deprivation level.
  – Orexin agonist in VTA induces feeding behavior, and orexin antagonist reduces food intake.
  – Orexin modulates firing activity of dopamine neurons.
Food vs. Money:

- **Background**
- **Reward as Drive Reduction**
- **Regulation + Learning**
- **Anticipatory Responses**
- **Behavioral Plausibility**
- **Energy Regulation**
- **Overeating**
- **Future Directions**
- **Conclusion**
Anticipatory Shivering:

\[ SDR = r(x^*, k_x) + \gamma \cdot r(x^* + k_x, -l_x) + \gamma^2 \cdot r(x^* + k_x - l_x, l_x - k_x) \]

\[ = [D(x^*) - D(x^* + k_x)] + \gamma \cdot [D(x^* + k_x) - D(x^* + k_x - l_x)] + \gamma^2 \cdot [D(x^* + k_x - l_x) - D(x^*)] \]

\[ = 0 - \frac{m}{\sqrt{|k_x|^n}} + \gamma \left[ \frac{m}{\sqrt{|k_x|^n}} - \frac{m}{\sqrt{|l_x - k_x|^n}} \right] + \gamma^2 \left[ \frac{m}{\sqrt{|l_x - k_x|^n}} - 0 \right] \]

\[ \frac{d(SDR)}{dl_x} = 0 \Rightarrow (\gamma - 1). (k_x)^{n-1} = \gamma(\gamma - 1). (l_x - k_x)^{n-1} \]

If \( \gamma \neq 1 \), then:

\[ k_x = l_x \cdot \frac{1}{1 + \gamma^{m-n}} \]
Extinction Burst is Learned:

![Graph showing Extinction Burst](image-url)
Extinction Burst is Learned:

![Graph showing lever press behavior over time](image1)

- **Lever Press / 20 min**
- **Time (hour)**

![Graph showing outcome expectancy and internal state over time](image2)

- **Outcome Expectancy**
- **Internal State**
- **Time (hour)**