相关概念：

blocking：

RPE：主要依赖error认知奖励，DA可算RPE生物指示。

{reward-related circuits：DA.

negative reward signals：lateral habenula (LHb)：抑制DA.

强化学习.

① TD learning rule：⇒ Q似乎与striatium活动相关

$Q(S_t, a_t) = E[r(t) + \gamma r(t+1) + \gamma^2 r(t+2) + \ldots | S_t, a_t]$.

期望. 期望[实际最终奖励 + 未来奖励]. state. action.

$\Delta Q(S_t, a_t) = \alpha[r(t) + \gamma Q(S_{t+1}, a_{t+1}) - Q(S_t, a_t)]$.

RPE. ：学习(实际+未来 − 原本预期)

参数：

α：学习率.

γ：延迟满足能力，越高对未来敏感.

β：贪个，越偏好高Q.

强化学习四大核心模块

模块	关键词	解决的问题	神经实现
RPE（Reward Prediction Error）	$\mathrm{RPE} = r + \gamma Q' - Q$	**学习应该发生吗？**（是否低估/高估了奖励）	多巴胺系统（VTA/SNc）
Q值与△Q更新	$Q \leftarrow Q + \alpha \cdot \mathrm{RPE}$	**估值如何修正？**（调整期望行为价值）	纹状体、皮层突触权重调整
Softmax 动作选择模型	$P(a) = \frac{e^{\beta Q(a)}}{\sum e^{\beta Q}}$	**做什么？**（概率性行为选择）	vmPFC、前额叶皮层调控 β
未来奖励贴现模型	$\mathrm{Exp}: \gamma^t \quad / \quad \mathrm{Hyp}: \frac{1}{1+kt}$	**何时行动？**（现在好，还是等待更好）	前额叶-边缘系统、5-HT调控 γ/k

② Exploitation - Exploration.

{ Greedy：best Q.

ε-Greedy：90% best Q，10%随机.

Softmax：β { 低β：随机. (exploration)

高β：Q高 (exploitation)

the matching law：按reward比例做出选择.

③ Future discounting: / hyperbolic discounting.

$\gamma^t \cdot r(t)$ $\frac{r(t)}{1+kt}$.

(针对对近距离敏感而改良)

图式 schema.

模块	关键词	神经 / 行为机制	易考点
图式（Schema）	联想网络、结构知识、贝叶斯先验	海马形成、新皮层迁移、情绪联想	图式作用于记忆巩固 / 快速学习
记忆机制	Tse实验、48h迁移、图式一致性	动物paired-associate学习、人类顺序一致任务	图式能加速皮层整合 / 降低认知负荷 越新越激活海马.
决策建模	证据整合模型（DDM）、LIP区反应	累积-跨阈决策机制	LIP神经元编码证据方向；行为建模可预测脑区活动
行为经济学	期望效用、悖论、前景理论	概率权重、损失厌恶、参考点偏好	阿莱悖论解释 / 框架效应实验 / 凹形效用函数

lateral intraparietal (计算脑区)

↳高估小概率，低估大概率

↓

Petersburg paradox：maximize utility.