# 基底神经节建模的经典的计算方法

Classical Computational Approaches to Modeling the Basal Ganglia

Ahmed A. Moustafa and V. Srinivasa Chakravarthy

© Springer Nature Singapore Pte Ltd. 2018

Computational Neuroscience Models of the Basal Ganglia, Cognitive Science and Technology,

https://doi.org/10.1007/978-981-10-8494-2_2

（Song Jian，translate）

**摘要：**目前已有几种模拟 BG 结构和功能的建模方法。在这一章中，我们讨论了主要的建模框架，这些框架被提议来模拟 BG 的许多功能。许多这样的建模研究都是 BG 建模领域的经典方法，它们反复模拟许多 BG 功能。简言之，我们将讨论以下模型方法：降维模型、行动部分选择模型、Go/NoGo 模型、基底节强化学习（RL）模型和 Actor-Critic 模型。重要的是，本章主要对模拟 BG 结构和功能的主要架构的概述。此外，我们在下面的章节中讨论了许多其他模型，例如步态模型、伸展模型和其他模型。

下面我们将回顾以下 BG 模型：降维模型、行动部分选择模型、Go/NoGo 模型、基底节强化学习（RL）模型和 Actor-Critic 模型。

## 4.1 降维模型

对 BG 功能和解剖的研究大多着眼于理解其在动作选择中的作用。这种方法从解释各种解剖区域、路径和连接框架图开始（Albin、Young 和 Penney，1989；Gurney、Prescott 和 Redgrave，2001a）。Bergman 和他的同事们不同于这条主线，他们使用计算方法研究了 BG 在大脑皮层信息降维中的作用，这些计算方法在执行行为任务时（Bar-Gad、Havazelet-Heimer、Goldberg、Ruppin 和 Bergman，2000；Bar-Gad、Morris 和 Bergman，2003），进一步通过灵长类 BG 的神经活动进行了验证。降维的概念来自于观察到大量的皮质神经元投射到输入端，即纹状体，这在物种中是一致的。据报道，大约有 $17\times10^6$ 个皮质神经元投射到 $17\times10^6$，使会聚率达到 10，这一比例在灵长类（571）和人类（347）中甚至更高（Bar-Gad 等人，2003）。这种趋势在纹状体和苍白球神经元之间可以进一步观察到。然后 GPi 和丘脑投射回皮质，类似于分化。通过学习来保存、更新信息并将其重新发送回大脑皮层，而不会造成任何损失，这一点非常重要。

为了研究这一方面，Bergman 和他的同事们提出了强化-驱动降维（reinforcement-driven dimensionality reduction，RDDR）模型，该模型使用多巴胺能增强信号来调节层之间的学习（Bar-Gad 等人，2000）。基本的 RDDR 模型包括多层前馈，代表皮质、纹状体和 GPi。前馈权值采用赫布（Hebbian）学习法更新，层间横向连接采用反赫比（anti-Hebbian）学习法。输入层比输出层（低维）有更多的神经元（高维）来模拟皮质 BG 的解剖结构。输出神经元的活动是从第一层接收到的输入和其自身层内的横向输入的函数。作为奖励的增强信号调节前馈权重的学习（Bar-Gad 等人，2000）。

利用这种网络结构，他们研究了信息编码过程中每一层的活动，并最终计算了重建误差，即原始元素和重建元素（从输出层）之间的平均平方差整体输入模式（Bar-Gad 等人，2000）。与输入层相关相比，该模型预测输出神经元的活动无/低相关。这个模型的预测与灵长类苍白球的神经记录一致，尽管大脑皮层输入的相关性较高，但这些记录显示出不规则的活动。输出层活动的不规则性和低相关性是由层内横向连接的强度造成的。由于多巴胺能信号的增强在模型的前馈层中调节了赫布学习，作者认为这种强度

导致了一个网络，它不仅编码输入空间的最大变异性，而且编码奖励扭曲空间的变异性。

最后，他们提出了一个高级 RDDR 模型，它克服了基本模型的缺点（Bar-Gad 等人，2003）。第一个也是最重要的约束是限制权重为正或负，这在基本模型中是不存在的。第二，基本模型中的单个神经元是线性的，但由于大多数 BG 神经元的固定率呈非线性，因此建议将单个单元改为 S 型非线性单元。此外，他们计划将模型包含并扩展为多回路、稀疏连接的系统，这在生物学上更为现实。

## 4.2 动作选择模型

生物体对来自世界的不同感官刺激有不同的反应。然而，当同时接收到两个感官刺激时，通常不可能表达两个相应的动作，因为这两个动作可能是不相容的，例如，面对捕食者时是选择逃走还是战斗（两者不可兼得）。因此，虽然来自外部世界的感觉流可能是同时的，但生物体对这些刺激的反应必须首先经过一个特定的仲裁过程，在给定的环境中选择最理想的行动。Gurney 及其同事（Bogacz 和 Gurney，2007；Gurney 等人，2001a；Gurney，Prescott 和 Redgrave，2001b；Humphries，Stewart 和 Gurney，2006）通过分析 BG 在感觉运动皮质通路上的解剖位置，提出了 BG 执行某种动作选择。随着这一思路的发展，BG 系统被认为是解决行动选择问题的脊椎动物解决方案（Redgrave、Prescott 和 Gurney，1999）。一些研究表明基底神经节在动作选择中起着关键作用（Seo、Lee 和 Averbeck，2012）。

为了解释 BG 结构如何实现假定的动作选择功能，Redgrave 等人（1999）提出纹状体计算出在特定时刻影响运动皮层的多种作用替代物的某种显著性。具有最高显著性的行为赢得竞争，并有最高的表达机会。GABA 能介质的多棘神经元（MSNs）之间的局部抑制被认为是在纹状体水平上为多种行为表现之间的竞争提供了细胞水平的机制。此外，从纹状体到 GPi 的聚焦抑制投射和从 STN 到 GPi 的弥漫性兴奋投射被认为在 BG 的输出核水平上产生前馈“偏离中心/围绕”输入/输出响应，即 GPi。其他人也提出了类似的建议（Mink，1996）。这些想法在类-BG 的控制架构驱动的机器人系统中得到了进一步的证实（Prescott，2002）。

纹状体内 MSNs 的横向抑制网络可以实现行为间的竞争，这一观点也被其他研究者所探索。例如，Wickens（1997）研究了精确的拓扑和其他连接参数，使纹状体网络能够执行预期的行动竞争功能。这个网络模型的另一个吸引人的特征是观察到在类似帕金森病的多巴胺减少情况下的强度变化。当多巴胺水平降低到“正常”以下时，网络动态行为从竞争转向协同激活，从而选择多种不相容的动作，这种情况自然可以解释帕金森氏运动障碍，如僵硬。

Humphries 等人（2006）基于 90 年代以来 Gurney 等人开发的 BG 动作选择理论。开发了一个详细的 BG 系统的 spiking 神经元模型，包括所有的主要核。该模型显示了动作选择的作用方式，正如前面在简单的速率编码模型中直观地提出和演示的那样。该模型的另一个显著特征是 STN-GPe 系统的同步振荡，该系统可以通过多巴胺进行调制。在该模型中，在正常或高多巴胺条件下，STN 和 GPe 被解耦，并呈现出非同步振荡；在降低多巴胺条件下，两个核是动态耦合的，显示出与帕金森氏条件下病理同步振荡。

在一项有趣的研究中，Amos（2000）提出了一个模拟威斯康星卡片分类测试（Wisconsin Card Sorting Test，WCST）性能的模型。该模型结合了前额皮质（PFC）和基底神经节之间的相互作用。该模型提供了一个计算有效性，即 PFC 疾病患者在执行

WCST 时如何表现出持续性错误和基底神经节疾病患者如何表现出随机错误，在 WCST 中，智能体人员学习根据特定规则对卡片进行分类。此外，在进行了一些正确的测试之后，智能体必须重新学习，根据不同的规则对卡片进行分类。该模型包括闭合和开放的皮质基底神经节环。PFC 观察到分类规则的主动维持，而基底神经节则选择运动反应。纹状体辅助整合皮层编码的信息，并将皮层活动映射到运动反应中。基底神经节接收来自 PFC 和感觉联想皮层的输入（编码代表目标和输入卡）。基底神经节输入到 PFC 的功能是提供反馈，表明所做的响应是正确的（不是模型化的）。如果反应不正确，PFC 会根据基底神经节的反馈改变其排序规则。请注意，这一假设不同于向 PFC 输入基底神经节的假设，因为在前者中，在 WM 中的信息维持不依赖于基底神经节的完整性。在 Amos 模型中，假设投射到纹状体或 PFC 上的 DA 会增加信噪比（Cohen 和 Servan-Schreiber，1992）。也就是说，投射到纹状体或前额叶神经元上的 DA 降低了噪声的影响，从而增加了神经对刺激的反应。大脑区域的 DA 消耗通过减少代表该区域的 S 型单位的增益来建模（另见 Cohen 和 Servan-Schreiber，1992）。在这个模型中，通过减少代表该区域的神经元的输出来模拟大脑区域的损伤。

仿真结果表明，PFC 中 DA 的降低与持续反应的发生有关，纹状体中 DA 的降低与随机误差的发生有关。额叶功能障碍与持续性反应的发生有关，因为只有感觉联系区向纹状体投射信息。由于排序规则的表示不保持在额叶皮质，所以在整个实验中只选择了与纹状体单元相关的运动响应，而纹状体单元的激活率最高（我假设在整个实验中噪声是固定的；请注意，模型中没有学习）。这个模型有一些局限性。它没有受过执行任务的训练。此外，基底神经节间接通路不纳入模型。而且，该模型不能解释对基底神经节的损伤导致 WM 任务（包括 WCST）中出现持续反应的结论。

基于先前的研究，一些最新的模型还表明，多巴胺投射到纹状体在动作部分起着关键作用，而多巴胺投射到前额叶皮层则是注意力学习的关键，也就是说，学习在纠正反馈的基础上关注环境中的关键信息（Moustafa 和 Gluck，2011a，2011b；Moustafa, Herzallah 和 Gluck，2014）。

基于 BG 函数的作用选择假设，近年来已有多个模型建立。例如，Gurney 及其同事最近的一项扩展表明，纹状体中的神经肽物质 P（SP）和脑啡肽在动作选择和顺序处理中发挥了作用（Buxton、Bracci、Overton 和 Gurney，2017）。在另一项研究中，同一组人模拟了行动选择与γ振荡和β振荡之间的关系（Blenkinsop、Anderson 和 Gurney，2017）。

# 4.3 Go/NoGo 模型

Frank 提供了一个神经计算模型，说明了 BG、丘脑和皮层与 DA 和其他神经递质如何基于奖励机制、运动和认知学习任务相互作用。Frank 模拟了 BG 所支持的不同运动和认知任务的表现，包括 WM 和决策。Frank 的框架表明，BG 调制整合运动和认知行为，这些行为是在运动和前额皮质编码（Frank、Loughry、O'Reilly 和 Houk，2005；O'Reilly 和 Frank，2006）。在运动域中，这些模型假设到运动前皮层的 BG 输出负责动作选择。同样，在认知领域，BG 调节编码在前额叶皮质的表达（Frank 等人，2001；Middleton 和 Strick，2000，2002；O'Reilly 和 Frank，2006）。

最重要的是，Frank 表明 DA 有益于 BG 中运动、认知和基于奖励机制的学习和表现，这得到先前研究的支持（Delgado、Miller、Inati 和 Phelps，2005；Schultz，1998；Schultz、Dayan 和 Montague，1997；Shohamy、Myers、Geghman、Sage 和 Gluck，2006）。

DA 暴增和骤降有助于“学习”通过改变 BG 直接（Go）和间接（NoGo）途径中的突触可塑性来选择最适应性的反应，并避免最不适应性的反应。随后的实验支持了对模型的核心预测，即停药的 PD 患者在从正强化到负强化的学习上受损，而停药的相同患者则表现出相反的学习偏差模式（Frank、Seeberger 和 O'Reilly，2004）。在 Frank 的模型中，模拟的非药物 PD 状态显示了增强的负性，而不是正性，强化学习和模拟的 DA 药物可以逆转这种偏差（Frank 等人，2004）。Frank 还证实了其模型在随后对 ADHD 患者（Frank、Santamaria、O'Reilly 和 Willcutt）以及服用多巴胺激动剂和拮抗剂的正常健康受试者的实验研究中的预测（Frank 和 O'Reilly，2006）。

这些模型的最新扩展探讨了 STN 在决策中的特殊作用。特别是，Frank 的模型表明，STN 提供了一个关于运动和认知行为的动态全局 NoGo 信号。人类受试者的 fMRI 研究报告了这一假设的实验性证据（Aron 和 Poldrack，2006）。模拟结果还表明，STN 在强冲突决策中扮演着重要的角色，即两种备选方案一样好或一样坏的决策（Hershey 等人，2004）。在 Frank 的模型中，通过激活 STN，在运动神经前部和扣带区（Braver 等人，2001；Frank，2005）表现出多种竞争反应，对于减缓反应和防止在不同的决策过程中出现过早反应至关重要。在这些模型中，模拟的 STN 损伤导致选择的模式出现过早反应，这与 STN 损伤大鼠这种行为的证据一致（Baunez 和 Robbins，1997）。Frank 的模型表明，通过以不自然的高频率刺激 STN，DBS 有效地消除了 STN 的动态功能（类似于病变），因此消除了全局 NoGo 信号（Benazzouz 和 Hallett，2000；Limousin 等人，1997；Meissner 等人，2005）。具体地说，当响应冲突较高时，STN 更为活跃（即，两种响应都具有相似的增强值）。这种活动的增加减缓了反应（丘脑活动逐渐增加证明），并防止模型在强冲突条件下仓促做出决定。完整的网络和具有 STN-DBS 的网络都能在训练阶段成功地选择适当的响应。然而，在测试阶段，STN-DBS 在强冲突条件下的破坏选择，在这样的条件下，刺激具有可比较的增强值（80%比 70%）。

此外，有时报告的 DBS 副作用是情绪亢进、不受控制的笑（Czernecki 等人，2002；Funkiewiez 等人，2003；Krack 等人，2001）或注意力分散（Saint Cyr、Trepanier、Kumar、Lozano 和 Lang，2000），可能是由于因目前扩散到边缘或联想性 STN（见 Karachi 等人，2005；Krack 等人，2001）导致的 NoGo 活动减少所致。

Frank 的框架表明，注意力分散是一种更高层次的认知反应，可能是由于信息过度进入前额叶皮质而导致的（Frank 等人，2001）。这一假设是通过一个被称为 AX-CPT 的认知任务来检验的。AX-CPT 是一项工作记忆任务，在该任务中，受试者受到连续字母刺激（A、B、X、Y；用红色打印），并被指示按两个键中的一个来完成每个字母的呈现（Cohen、Barch、Carter 和 Servan Schreiber，1999；Servan Schreiber、Cohen 和 Steingard，1996）。当 A 后接 X（AX‘目标’试验）时，要求受试者按键盘右侧的键（“m”），否则按左键（“z”）（AY、BX 和 BY 试验）。简言之，Frank 的建模框架表明，前额皮质在 WM 中对信息的主动维持起着关键作用，而 BG 在何时和何时不更新信息到 WM 中起着关键的调节作用，这一功能变得更加重要。

Beiser 和 Houk（1998）提出了一个门控模型，该模型概念上类似于 Frank 等人（2001）提出的模型的。在这两个模型中，输入刺激在前额叶皮层被短暂地表现，进入 WM 的信息的门控由 PFC 基底神经节分离环路控制。Beiser 和 Houk 模型模拟延迟序列学习任务中的性能。在这项任务中，模型按一定顺序给出了一系列照明键。经过一段时间的延迟后，模拟对象应该按与呈现顺序相同的键。Beiser 和 Houk 模型假设照明键序列（以时间顺序呈现）在 PFC 中以空间表示。该模型模拟进入 WM 的刺激的门（称为编码问

题）。该模型表明，不同的照明键序列在 PFC 中具有不同的空间表征（即活动模式）。该模型假设输入刺激短暂激活 PFC 神经元，进而激活尾状核。然后，丘脑的去抑制导致前皮质-下丘脑环路刺激的维持。这个模型有一些局限性。Beiser 和 Houk 模型并没有模拟模型如何按它们出现的相同顺序按下按键（称为解码问题）。这个模型的另一个局限性是它没有被训练来执行任务。此外，基底神经节的间接通路也不包含在模型中。

基于基底神经节的 Go/NoGo 模型，最近的一个模型模拟了基底神经节与脊髓相互作用在行为选择中的作用（Kim 等人，2017）。与先前的模型不同，该模型模拟了动态环境中的手臂伸展，包括在多个动作之间进行选择。另一个模型也建立在先前的 Go/NoGo 模型之上，但进一步纳入了一个二项 Hebb 规则来训练纹状体中的突触（Baston 和 Ursino，2015）。

# 4.4 基底节 RL 模型

最常用的模拟 BG 函数的模型是 RL 模型。

强化学习（RL）是一种无监督的机器学习方法，它与大脑的功能机制有很大的相似性。在 RL 中，“$t$”时处于状态（$s_t$）的智能体（Agent）（例如，BG）执行动作（$a_t$），并从环境中获得奖励（$r_t$）。智能体（Agent）的目的是通过选择一个最优的策略来最大化奖励。RL 源于心理学中的操作性条件作用理论，它描述了一种基于行动结果的智能体学习刺激-反应（S-R）关系的方式：与奖励结果相关的 S-R 对得到加强，而那些导致惩罚的对得到减弱。通常，由于一个行为与其结果之间存在延迟，RL 理论提出了一种替代奖励，即价值，供 actor 在选择一个潜在的奖励行为时使用（Sutton 和 Barto，1998）。Schultz 在猕猴身上记录多巴胺活性的开创性实验表明，DA 编码用于奖励预测错误。这个错误项类似于 RL 中的时间差错误（‘$\delta$’）。实验数据表明，BG 以多巴胺能输入纹状体的形式接收奖励相关信息（Chakravarthy、Joseph 和 Bapi，2010；Niv，2009）。多巴胺也能引起皮质纹状体可塑性变化（Reynolds 和 Wickens，2002），从而调节皮质纹状体突触的类似于 Hebb 塑性（Surmeier、Ding、Day、Wang 和 Shen，2007）。基于这些观察，大量的建模文献围绕着这样一个概念发展起来：BG 使用来自 DA 神经元的奖励相关信息来执行各种认知功能，如决策和序列生成（Chakravarthy 等人，2010；Niv，2009）。之前讨论过的许多模型只关注基底神经节的动作选择策略（Bar Gad 等人，2003；Gurney 等人，2001a；Humphries、Khamassi 和 Gurney，2012）。多巴胺的基本增强特性在这类模型中没有得到充分利用。大量证据表明多巴胺活性与突触的长期增强和抑制有关（Schultz，1998；Wickens、Horvitz、Costa 和 Killcross，2007；Wise，2004；Wise 和 Rompre，1989）。也许，突触学习的三因素法则将多巴胺作为一个重要因素，以及突触前和突触后控制突触强度和学习的信息（Wickens 和 Kotter，1995）。此外，Schultz 及其同事的经典实验和 Houk 及其同事的模型（Houk、Adams 和 Barto，1995；Schultz，1998）表明，多巴胺不仅具有奖励方面，而且还具有奖励预测特性。详细的实验证明，一个称为奖赏预测误差的数学量（Sutton 和 Barto，1998）与多巴胺能神经元的信号很匹配。

实验证据从 Schultz 及其同事（Schultz，1998）提出的经典证据开始，该证据表明多巴胺能神经元在奖励“r”时会增加其激活。如果我们将他们的活动表示为变量“$\delta$”，则

$$\delta \propto r$$

引入学习和决策奖励的概念，就需要引入强化学习的概念（Sutton 和 Barto，1998）。强化学习是机器学习的一个分支，智能体通过抽样奖励他/她在某个状态下的行为，更新有关环境的信息并做出有效的决策。目标是在一个状态下最大化获得的奖励。

更具体地说，多巴胺不仅仅是对奖励的反应，而且是对预测的奖励。奖励预测由强化学习中名为“值”的函数跟踪。定义为

$$Q(t)=\sum_{i=t+1}^{\infty} r_i$$

上述公式可以通过使用一个系数γ来计算未来奖励的权重来即兴创作，系数γ满足：

$$Q(t)=r_{t+1}+\gamma r_{t+2}+\gamma^2 r_{t+3}+\cdots+\gamma^{n-1} r_{t+n}$$

血清胺神经元（serotonergic neurons）被认为与权重系数γ相关（Tanaka 等人，2007）。更新“值”函数的方法是：

$$Q(t+1)=Q(t)+\eta_Q \delta_t$$

该算法可以定义为将“状态-动作-奖励-下一个，状态-下一个，动作”（SARSA）解释为序列，Q 函数可以写为 Q（$s_t$，$a_t$），其中“$s_t$”是时间“$t$”的状态，“$a_t$”是时间“$t$”的动作，以及“$\eta_Q$”是动作“值函数”的学习率（$0<\eta_Q<1$）。对于直接报酬问题（$\gamma=0$），DA 的时间差（TD）误差测量由下式中的$\delta_t$确定：

$$\delta_t = r_t - Q(s_t, a_t)$$

这称为奖励预测误差，数量与多巴胺能效应密切匹配，而不仅仅是奖励或奖励预测值。它还代表了 Rescorla-Wagner 的（RW）规则，它带来了无条件刺激（奖励，US）和条件刺激（状态，CS）的关联。在非零贴现系数的情况下，时间预测（TD）定义为：

$$\delta_t = r_t + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)$$

纹状体神经元被用于计算并跟踪价值函数和奖励预测相关数量（Balleine、Delgado 和 Hikosaka，2007；Delgado，2007；O’Doherty 等人，2004；Samejima、Ueda、Doya 和 Kimura，2005）。在 RL 语言中，这近似于 Critic 模块的功能（Joel、Niv 和 Ruppin，2002）。Critic 的对应模块被称为 Actor 模块。此模块使用 Critic 计算的状态和操作的评估来执行选项选择。行动选择可以是探索性的或开发性的（Sutton 和 Barto，1998）。决定一个选择的随机性数量的函数称为策略（policy），π。RL 中一些著名的策略包括 epsilon-greedy，其中在任何时候，对于一个状态，随机行为都是以概率（ε）执行的；soft-max，其中有一个温度参数（$\beta$），控制勘探。$\beta$值越低，差异值函数越小，$Q$ 值越大，对$\beta$值的探索越多。

$$\pi_{a1}=\frac{e^{-\beta Q_{a1}}}{e^{-\beta Q_{a1}}+e^{-\beta Q_{a2}}}$$

这里，$\pi_{a1}$ 表示在时间 $t$ 时，从一个状态 $s$，选择动作 $a1$ 的概率。基底神经节中的策略等价物构成了由直接和间接路径执行的动态。

***Actor-Critic 模型***

另一方面，Actor-Critic 模型专注于模拟（感知和记忆引导）运动行为。这些 BG 模型模拟了运动学习任务中的行为，例如操作性条件（instrumental conditioning）（例如，Houk，1995a）、S-R（例如，Berns 和 Sejnowski，1996；Khamassi，Girard，Berthoz 和 Guillot，2004；Suri，Bargas 和 Arbib，2001）、顺序学习（Suri 和 Schultz，1998）和延迟响应任务（Suri 和 Schultz，1999）

Actor-Critic 模型将运动学习与运动任务中的奖励预测学习分离开来（Houk，1995b；O’Doherty 等人，2004）。例如，在操作性条件任务中，动物被触发（例如，通过光的照射或听觉信号的呈现）去做出特定的运动响应以获得奖励。Actor-Critic 模型假定，在这些任务中，动物学习（a）触发刺激预测奖励的发生（奖励预测）和（b）如何做出奖

励后的运动反应。

在 Actor-Critic 架构（图 4.1）中，Critic 负责学习如何预测奖励，而 Actor 负责根据 Critic 的指示调整动作（Barto，1995，2003）。Critic 向参与者发送一个学习信号（也称为增强或预测错误信号），通知 Actor 它所做的运动响应是否有奖励的结果。一个正信号通知 Actor 增加其刚采取的行动（即加强）的可能性，而一个负信号则通知 Actor 不要做出其刚采取的行动响应。另一方面，Critic 没有收到 Actor 的信号（图 4.1）。然而，它被告知 Actor 的动作反应是否有奖励的后果。

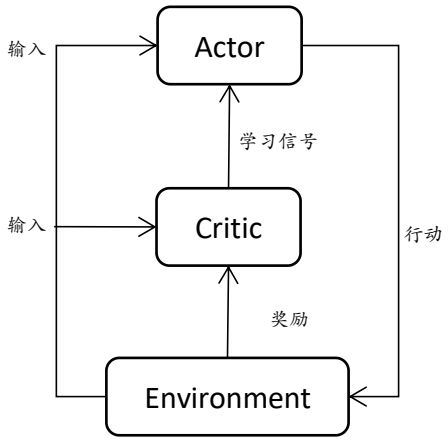


图 4.1：Actor-Critic 架构。Actor 和 Critic 都从环境（包括实验者）中接受输入。Actor 负责做出动作反应。如果它做出正确的电机响应，模型就会得到奖励。Critic 负责奖励预测学习。Critic 从环境中获得奖励，并向 Actor 发送学习信号，告知其所做的行为是否有回报的结果（Barto，2003 年）。

***Houk 及其同事模型***

Houk 及其同事（Houk，1995a；Houk 等人，1995）提出了第一个模型，表明基底神经节在结构和功能上类似于一个 Actor-Critic 架构（Joel 等人，2002）。他们认为 Actor-Critic 架构的基本结构可以映射到纹状体上。他们认为这些基质体（Matrisome）（及其传出的目标）在功能上等同于 Critic。这是基于这样一个事实：（a）纹状小体与 SNc（一个有益于基于奖励的学习脑区）相互连接，（b）皮质-纹状小体通路的突触修饰依赖于 DA 的，可能有益于基于奖励的学习。Houk 等人同时也表明纹状小体（及其传出目标）在功能上等同于 Actor。这是基于这样一个事实：基质体（通过 GPi 和丘脑）向运动皮层发送投射。

Houk 等人（1995）提出了一种概念模型（即没有模拟研究），用于模拟操作性条件任务中的行为。在这项任务中，猴子学会按杠杆以获得奖励。刺激（CS）触发猴子做出运动反应。该模型假设学习预测 US 是由皮质、纹状小体和 SNc 决定的。该模型还假设学习做出运动反应（按下控制杆）是由基质体、皮层、GPi 和丘脑所决定的。首先，Houk 等人（1995）认为基质体-SNc 路径计算 TD 误差。抑制通路，兴奋通路，并且下丘脑外侧输入到 SNc，分别计算了 *P(t-1),P(t),R(t)*。

Houk 等人模型有一些局限性。一个局限性是，该模型不能解释皮质连接在学习操作性条件任务中的作用（尽管它被描述）。同样，该模型不能解释基底神经节间接通路在学习执行运动任务中的作用。此外，该模型是一个概念模型，而不是模拟模型。Houk 等人的假设是，纹状小体在功能上等同于 Critic，而基质在功能上等同于 Actor，这些假设被纳入模型中，模拟由基底神经节所提供的功能，例如 S-R 学习（Khamasi 等人，2004；Suri 等人，2001），空间延迟响应任务（Suri（Schultz，1999）和顺序学习任务（Suri Schultz，1998；Tian，Arnold，Sejnowski，Jabri，2003）。

### *Suri 及其同事的模型*

Suri 和 Schultz（1998）提出了一个 Actor-Critic 模型，该模型模拟顺序学习任务中的行为。在这项任务中，模型被训练成将不同刺激（A、B、C、D、E、F 和 G）的存在与做出不同运动反应（Q、R、S、T、U、V 和 W）相关联。换言之，在每次试验中，模型学习将 A 的表示与作出响应 Q 联系起来，B 的表示与作出响应 R 联系起来，等等。类似于 Houk 等人模型中，该模型假设了纹状小体有益于奖励预测学习，而基质体有利于行为学习。每个隔室使用一个单层网络进行模拟。与基质体和纹状小体的皮层连接是完全连接的。该模型假设在基质体中实现了动作选择（本研究中的选择被解释为从若干潜在动作中选择一个动作）。一个 WTA 网络，大概模拟基质体单位之间的横向抑制，选择活性最高的单位。每个基质体单位对应一个行动。

Suri 和 Schultz（1998）模型采用 TD 算法进行训练。据我所知，当动物执行序列学习任务时，没有来自 DA 神经元的记录研究。然而，该模型假定 DA 阶段信号在到达最早的 CS 之前，会移动到 CS 的时间。这一假设基于这样一个事实：在 DRT（Schultz、Apicella 和 Ljungberg，1993）和学习操作性条件任务（Schultz 等人，1997）中，DA 阶段响应会随着时间而向后移动。

他们还研究了使用无条件强化信号训练模型的效果，这意味着 DA 信号不会转移到条件刺激的时间，并且无论是否预测总是与奖励相关——因此得名。仿真结果表明，DA 信号的非移位与学习障碍有关，这提供了 DA 阶段信号的移位与执行任务的强化学习有关的证据。Tian 等人（2003）提出一个 Actor-Critic 模型，在概念上类似于 Suri 和 Schultz（1998）模型。Tian 等人模型模拟序列学习任务，并应用于机器人领域。该模型的一个局限性是，它假定只有在刺激出现时，模型才会采取行动。这是不可能的，因为动物研究表明，在学习过程中，动物有时会在触发呈现之前或之后做出运动反应。Suri 和 Schultz（1999）提出了另一个 Actor-Critic 模型，该模型模拟了空间 DRT 中的行为（Schultz 等人，1993）。该模型的假设与 Suri 和 Schultz（1998）模型的假设相似。该模型模拟记忆引导的运动响应。该模型假设记忆引导的运动反应被 PFC 基底神经节通路所阻断。它还假设学习做出记忆引导的运动反应是由 DA 投射到纹状体的。Suri 和 Schultz（1999）还模拟了 PD 患者的保护反应的发生。他们假设纹状体 DA 在 PD 中的减少导致 DA 阶段信号不向条件刺激时间移动。他们用无条件强化信号训练他们的模型。他们的模型表明，DA 阶段信号的不适当的时间转移是导致局部放电持续反应的原因。仿真结果还表明，学习后不给出预期的奖励会导致行为消亡。Suri 和 Schultz（1999）模型有一些局限性。该模型没有模拟输入信息如何被选入 WM。此外，该模型没有解决分心器的表示。

Suri 等人（2001）提出另一个 Actor-Critic 模型，模拟 S-R 任务中的行为，即 T-maze 任务。在这个任务中，老鼠学会了将右转或左转分别与面对绿色或红色刺激联系起来。如果老鼠右转并达到绿色刺激，就会得到奖励。如果老鼠左转，就得不到奖励。该模型的假设与 Suri 和 Schultz（1998、1999）的假设相似，但在生理学上更为详细。

在概念上类似于 Suri 和 Schultz（1998）模型，Baldassarre 及其同事提出了 Actor-Critic 模型，模拟觅食任务中的表现（Baldassarre，2002；Baldassarre 和 Parisi，2000）。在这项任务中，一名智能体搜索放在二维板中的许多食物颗粒（Baldassarre 和 Parisi（2000）研究中模拟的任务与 Baldassarre（2000）研究中使用的任务略有不同，前者使用 10 个奖励对象，而后者仅使用 3 个奖励对象）。该模型假定基底神经节对 S-R 学习有利。采用 TD 算法进行训练。WTA 网络选择具有最高激活率的动作。利用具有噪声阈值的 S 状体单元来模拟 Actor。该模型使用一个称为匹配器的单元来计算模型的动机，例如在

饥饿和食物存在时移动。匹配器的基本生物学机制尚未明确。

模拟研究表明，Actor-Critic 体系结构足以模拟相对复杂的任务。在这个任务中，模型学习在获得奖励之前做出许多正确的运动响应（即朝食物的方向移动）。结果表明，该模型学习以大约 30 个步骤（随机约 100 个步骤）找到食物。这些模型的一个局限性是，一个接受了 200000 次试验的培训，另一个接受了 10000 次试验来执行任务。

***Berns 和 Sejnowski（1996）模型***

Berns 和 Sejnowski（1996）提出了一个 Actor-Critic 模型，该模型假定基底神经节对动作选择有利。类似于 Houk 等人（1995）模型，在该模型中，纹状小体对任务的基于奖励学习有利，而基质体对产生行动过程有利。神经元用 S 状单位模拟。模型采用 TD 算法进行训练。该模型假设投射到纹状体的 VTA 和 SNc 对 TD 学习有利；然而，这并没有给模型增加任何计算能力，而 Suri 和 Schultz 的模型只使用 SNc。

该模型中动作选择的神经基质不同于 Suri 及其同事（见上文）。Suri 和他的同事们假设了基质神经元（通过 WTA 网络模拟）有益于行为选择。然而，在这个模型中，行动选择是在 GPi 中实现的。Berns 和 Sejnowski 模型假设输入到 GPi 的纹状体和 STN 对动作选择有益。Berns 和 Sejnowski 模型使用术语 winner-lose-all 而不是 winner-take-all，因为获胜的 GPi 单元被抑制，而不是被激活。在这个模型中，纹状体稀疏地与 GPi 相连，而 STN 则向 GPi 发送兴奋的扩散投射。该模型假定每个 GPi 单元都会进行不同的动作响应。在这个模型中，STN 防止除了获胜的 GPi 单元之外的所有单元被抑制。模拟结果发现并实际预测，STN 的损坏导致无法停止选定的操作。他们将坚持不懈与“DA 减少”相关联。然而，还不知道如何模拟“DA 减少”。或者他们塑造了什么样的毅力。

这个模型的一个限制是它不包含在模型中。该模型的另一个局限性是，它假设相同的纹状体神经元向 GPe（间接途径）和 GPi/SNr 发送投射。Wilson（2004）指出，这在生物学上并不可信。

O'Reilly（2003）提出了一个与 Berns 和 Sejnowski（1996）非常相似的模型。该模型模拟 1-2-AX 任务中的性能。此任务是 AX-CPT 任务（如上所述）的扩展，在该任务中，智能体（人工主体或计算机模型）学习在工作内存中维护两个项目。该模型结合了基底神经节和 PFC 之间的相互作用，采用时间差分和监督学习（Leabra）算法相结合的方法对模型进行训练。该模型假设基底神经节辅助行为选择。这个模型没有模拟信息进入 WM 的门控，尽管同一作者早期的工作假设基底神经节辅助信息进入 WM。在这项工作之后，其中一种模型将基底神经节在运动节和信息门控两方面的作用纳入 WM（Moustafa 和 Maida，2007）。

最近的一些模型采用了强化学习方法来模拟基底神经节的各种功能。例如，Shivkumar、Muralidharan 和 Chakravarthy（2017）提出了一个模型来解决基于背景事件的 RL 过程。该模型假定纹状小体通过选择最理想的动作来控制基质体的功能。该模型做了几个预测，可以在未来的实验研究中进行测试。Stocco（2017）还扩展了基底神经节的 RL 和动作选择功能，以模拟更复杂的决策任务，这些任务无法使用更简单的 RL 模型进行模拟。重要的是，BG 的大多数 RL 模型都侧重于行动学习，但往往忽略了时间的表示。因此，最近的一项研究将时间敏感的行动选择机制纳入其中。该模型解释了时间间隔过程，即对持续时间的感知（Gershman、Moustafa 和 Ludwig，2014）。有关 RL 模型中时间表示的类似工作，请参见 Moustafa、Cohen、Sherman 和 Frank（2008）。

# 4.5 结论

本文所描述的建模方法在模拟某些 BG 函数方面显示了一些前景。但是，它们有很多局限性，因为 BG 支持的许多功能不能被这些模型模拟。在下一章中，我们讨论了一种新的建模方法，可以用来模拟大多数（如果不是全部）BG 函数。此外，在接下来的几章中，我们将讨论基于这里介绍的神经架构的其他几个 BG 模型。

# 参考文献


Albin, R. L., Young, A. B., & Penney, J. B. (1989). The functional anatomy of basal ganglia disorders. Trends in Neurosciences, 12(10), 366–375.

Amos, A. (2000). A computational model of information processing in the frontal cortex and basal ganglia. Journal of Cognitive Neuroscience, 12(3), 505–519.

Aron, A.R.,&Poldrack,R.A. (2006). Cortical and subcortical contributions to Stop signal response inhibition: Role of the subthalamic nucleus. Journal of Neuroscience, 26(9), 2424– 2433.

Baldassarre, G. (2002). A modular neural-network model of the basal ganglia's role in learning and selecting motor behaviours. Journal of Cognitive Systems Research, 3, 5–13.

Baldassarre, G., & Parisi, D. (2000). Classical and instrumental conditioning: From laboratory phenomena to integrated mechanisms for adaptation.

Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. Journal of Neuroscience, 27(31), 8161–8165. https://doi.org/10.1523/ JNEUROSCI.1554-07.2007.

Bar-Gad,I.,Havazelet-Heimer,G.,Goldberg,J.A.,Ruppin,E.,&Bergman,H.(2000).Reinforcement-driven dimensionality reduction-a model for information processing in the basal ganglia. Journal of Basic and Clinical Physiology and Pharmacology, 11(4), 305–320.

Bar-Gad, I., Morris, G., & Bergman, H. (2003). Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. Progress in Neurobiology, 71(6), 439–473. https://doi.org/10.1016/j.pneurobio.2003.12.001.

Barto, A. G. (1995). Adaptive critics and the basal ganglia. In J. C. Houk, J. L. Davis & D. G. Beiser (Eds.), Models of information processing in the basal ganglia (pp. xii, 382p). Cambridge, MA: MIT Press.

Barto, A. G. (2003). Reinforcement learning. In M. A. Arbib (Ed.), The handbook of brain theory and neural networks (pp. 963–968). Cambridge, MA: MIT Press.

Baston, C., & Ursino, M. (2015). A biologically inspired computational model of basal ganglia in action selection. Computational Intelligence and Neuroscience, 2015, 187417. https://doi.org/ 10.1155/2015/187417.

Baunez,C.,&Robbins,T.W. (1997). Bilateral lesions of the subthalamic nucleus induce multiple deficits in an attentional task in rats. European Journal of Neuroscience, 9(10), 2086–2099.

Beiser, D. G., & Houk, J. C. (1998). Model of cortical-basal ganglionic processing: Encoding the serial order of sensory events. Journal of Neurophysiology, 79(6), 3168–3188. https://doi.org/ 10.1152/jn.1998.79.6.3168.

Benazzouz, A., & Hallett, M. (2000). Mechanism of action of deep brain stimulation. Neurology, 55(12 Suppl 6), S13–S16.

Berns, G. S., & Sejnowski, T. J. (1996). How the basal ganglia make decisions. In A. Damasio, H. Damasio, & Y. Christen (Eds.), The neurobiology of decision making. Berlin: Springer.

Blenkinsop, A., Anderson, S., & Gurney, K. (2017). Frequency and function in the basal ganglia: The origins of beta and gamma band activity. Journal of Physiology, 595(13), 4525–4548. https://doi.org/10.1113/JP273760.

Bogacz, R., & Gurney, K. (2007). The basal ganglia and cortex implement optimal decision making between alternative actions. Neural Computation, 19(2), 442–477.

Braver, T. S., Barch, D. M., Keys, B. A., Carter, C. S., Cohen, J. D., Kaye, J. A., … Reed, B. R. (2001). Context processing in older adults: Evidence for a theory relating cognitive control to neurobiology in healthy aging. Journal of Experimental Psychology: General, 130(4), 746– 763.

Buxton, D., Bracci, E., Overton, P. G., & Gurney, K. (2017). Striatal neuropeptides enhance selection and rejection of sequential actions. Frontiers in Computational Neuroscience, 11, 62. https://doi.org/10.3389/fncom.2017.00062.

Chakravarthy, V., Joseph, D., & Bapi, R. S. (2010). What do the basal ganglia do? A modeling perspective. Biological Cybernetics, 103(3), 237–253.

Cohen, J. D., Barch, D. M., Carter, C., & Servan-Schreiber, D. (1999). Context-processing deficits in schizophrenia: Converging evidence from three theoretically motivated cognitive tasks. Journal of Abnormal Psychology, 108(1), 120–133.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. Psychological Review, 99(1), 45–77.

Czernecki,V.,Pillon,B.,Houeto,J.L.,Pochon,J.B.,Levy,R.,&Dubois,B.(2002). Motivation, reward, and Parkinson's disease: Influence of dopatherapy. Neuropsychologia, 40(13), 2257– 2267.

Delgado, M. R. (2007). Reward-related responses in the human striatum. Annals of the New York Academy of Sciences, 1104(1), 70–88.

Delgado, M. R., Miller, M. M., Inati, S., & Phelps, E. A. (2005). An fMRI study of reward-related probability learning. Neuroimage, 24(3), 862–873.

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. Journal of Cognitive Neuroscience, 17(1), 51–72.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. Cognitive, Affective, & Behavioral Neuroscience, 1(2), 137–160.

Frank, M. J., & O'Reilly, R. C. (2006). A mechanistic account of striatal dopamine function in human cognition: Psychopharmacological studies with cabergoline and haloperidol. Behavioral Neuroscience, 120(3), 497–517.

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in parkinsonism. Science, 306(5703), 1940–1943.

Funkiewiez, A., Ardouin, C., Krack, P., Fraix, V., Van Blercom, N., Xie, J., … Pollak, P. (2003). Acute psychotropic effects of bilateral subthalamic nucleus stimulation and levodopa in Parkinson's disease. Movement Disorders, 18(5), 524–530.

Gershman, S. J., Moustafa, A. A., & Ludvig, E. A. (2014). Time representation in reinforcement learning models of the basal ganglia. Frontiers in Computational Neuroscience, 7, 194. https:// doi.org/10.3389/fncom.2013.00194.

Gurney, K., Prescott, T. J., & Redgrave, P. (2001a). A computational model of action selection in the basal ganglia. I. A new functional anatomy. Biological Cybernetics, 84(6), 401–410.

Gurney,K.,Prescott,T.J.,& Redgrave,P. (2001b).A computational model of action selection in the basal ganglia.II.Analysis and simulation of behaviour. Biological Cybernetics,84(6),411–423.

Hershey, T., Revilla, F. J., Wernle, A., Gibson, P. S., Dowling, J. L., & Perlmutter, J. S. (2004). Stimulation of STN impairs aspects of cognitive control in PD. Neurology, 62(7), 1110–1114.

Houk, J. C. (1995a). Information processing in modular circuits linking basal ganglia and cerebral Cortex. In J. C.

Houk, J. L. Davis & D. G. Beiser (Eds.), Models of information processing in the basal ganglia (pp. xii, 382p). Cambridge, MA: MIT Press. Houk, J. C. (1995b). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In J. C.

Houk, J. L. Davis & D. G. Beiser (Eds.), Models of information processing in the basal ganglia (pp. xii, 382p). Cambridge, MA: MIT Press. Houk, J. C. (2005). Agents of the mind. Biological Cybernetics, 92(6), 427–437.

Houk, J. C., Adams, J. L., & Barto, A. G. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. Models of Information Processing in the Basal Ganglia, 249–270.

Humphries, M. D., Khamassi, M., & Gurney, K. (2012). Dopaminergic control of the exploration-exploitation trade-off via the basal ganglia. Frontiers in Neuroscience, 6.

Humphries, M. D., Stewart, R. D., & Gurney, K. N. (2006). A physiologically plausible model of action selection and oscillatory activity in the basal ganglia. The Journal of Neuroscience, 26 (50), 12921–12942.

Joel, D., Niv, Y., & Ruppin, E. (2002). Actor–critic models of the basal ganglia: New anatomical and computational perspectives. Neural Networks, 15(4), 535–547.

Karachi, C., Yelnik, J., Tande, D., Tremblay, L., Hirsch, E. C., & Francois, C. (2005). The pallidosubthalamic projection: An anatomical substrate for nonmotor functions of the subthalamic nucleus in primates. Movement Disorders, 20(2), 172–180.

Khamassi, M., Girard, B., Berthoz, A., & Guillot, A. (2004). Comparing three Critic models of reinforcement learning in the basal ganglia connected to a detailed actor part in a S-R task. Paper presented at the Proceedings of the Eighth International Conference on Intelligent Autonomous Systems IAS-8, Amsterdam, The Netherlands.

Kim, T., Hamade, K. C., Todorov, D., Barnett, W. H., Capps, R. A., Latash, E. M., … Molkov, Y. I. (2017). Reward based motor adaptation mediated by basal ganglia. Frontiers in Computational Neuroscience, 11, 19. https://doi.org/10.3389/fncom.2017.00019.

Krack, P., Kumar, R., Ardouin, C., Dowsey, P. L., McVicker, J. M., Benabid, A. L., & Pollak, P. (2001). Mirthful laughter induced by subthalamic nucleus stimulation. Movement Disorders, 16(5), 867–875.

Limousin, P., Greene, J., Pollak, P., Rothwell, J., Benabid, A. L., & Frackowiak, R. (1997). Changes in cerebral activity pattern due to subthalamic nucleus or internal pallidum stimulation in Parkinson's disease. Annals of Neurology, 42(3), 283–291.

Meissner, W., Leblois, A., Hansel, D., Bioulac, B., Gross, C. E., Benazzouz, A., & Boraud, T. (2005). Subthalamic high frequency stimulation resets subthalamic firing and reduces abnormal oscillations. Brain, 128(10), 2372–2382.

Middleton, F. A., & Strick, P. L. (2000). Basal ganglia output and cognition: Evidence from anatomical, behavioral, and clinical studies. Brain and Cognition, 42(2), 183–200.

Middleton, F. A., & Strick, P. L. (2002). Basal-ganglia 'projections' to the prefrontal cortex of the primate. Cerebral Cortex, 12(9), 926–935.

Mink, J. W. (1996). The basal ganglia: Focused selection and inhibition of competing motor programs. Progress in Neurobiology, 50(4), 381.

Moustafa, A. A., & Gluck, M. A. (2011a). A neurocomputational model of dopamine and prefrontal-striatal interactions during multicue category learning by Parkinson patients. Journal of Cognitive Neuroscience, 23(1), 151–167. https://doi.org/10.1162/jocn.2010.21420.

Moustafa, A. A., & Gluck, M. A. (2011b). Computational cognitive models of prefrontal-striatal-hippocampal interactions in Parkinson's disease and schizophrenia. Neural Netw, 24(6), 575–591. https://doi.org/10.1016/j.neunet.2011.02.006.

Moustafa, A. A., & Maida, A. S. (2007). Using TD learning to simulate working memory performance in a model of the prefrontal cortex and basal ganglia. Cognitive Systems Research, 8, 262–281.

Moustafa, A. A., Cohen, M. X., Sherman, S. J., & Frank, M. J. (2008). A role for dopamine in temporal decision making and reward maximization in parkinsonism. Journal of Neuroscience, 28(47), 12294–12304. https://doi.org/10.1523/JNEUROSCI.3116-08.2008.

Moustafa, A. A., Herzallah, M. M., & Gluck, M. A. (2014). A model of reversal learning and working memory in medicated and unmedicated patients with Parkinson's disease. Journal of Mathematical Psychology, 59, 120–131.

Niv, Y. (2009). Reinforcement learning in the brain. Journal of Mathematical Psychology, 53(3), 139–154.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science, 304 (5669), 452–454.

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. Neural Computation, 18(2), 283–328.

O'Reilly, R. C. (2003). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. ICS Technical Report, (pp. 1–23).

Prescott, T. J. (2002). Basal ganglia. In M. A. Arbib (Ed.), The handbook of brain theory and neural networks (pp. xvii, 1290p). Cambridge, MA: MIT Press.

Redgrave, P., Prescott, T. J., & Gurney, K. (1999). The basal ganglia: A vertebrate solution to the selection problem? Neuroscience, 89(4), 1009–1023.

Reynolds, J. N. J., & Wickens, J. R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. Neural Networks, 15(4), 507–521.

Saint-Cyr, J. A., Trepanier, L. L., Kumar, R., Lozano, A. M., & Lang, A. E. (2000). Neuropsychological consequences of chronic bilateral stimulation of the subthalamic nucleus in Parkinson's disease. Brain, 123(10), 2091–2108.

Samejima, K., Ueda, Y., Doya, K., & Kimura, M. (2005). Representation of action-specific reward values in the striatum. Science, 310(5752), 1337–1340.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. Journal of Neurophysiology, 80(1), 1–27.

Schultz, W., Apicella, P., & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. Journal of Neuroscience, 13(3), 900–913.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. Science, 275(5306), 1593–1599.

Seo, M., Lee, E., & Averbeck, B. B. (2012). Action selection and action value in frontal-striatal circuits. Neuron, 74(5), 947–960. https://doi.org/10.1016/j.neuron.2012.03.037.

Servan-Schreiber, D., Cohen, J. D., & Steingard, S. (1996). Schizophrenic deficits in the processing of context. A test of a theoretical model. Archives of General Psychiatry, 53(12), 1105–1112.

Shivkumar, S., Muralidharan, V., & Chakravarthy, V. S. (2017). A biologically plausible architecture of the striatum to solve context-dependent reinforcement learning tasks. Frontiers in Neural Circuits, 11(45). https://doi.org/10.3389/fncir.2017.00045.

Shohamy, D., Myers, C. E., Geghman, K. D., Sage, J., & Gluck, M. A. (2006). L-dopa impairs learning, but spares generalization, Parkinson's disease. Neuropsychologia, 44(5), 774–784.

Stocco, A. (2017). A biologically plausible action selection system for cognitive architectures: Implications of basal ganglia anatomy for learning and decision-making models. Cognitive Science https://doi.org/10.1111/cogs.12506.

Suri, R. E., Bargas, J., & Arbib, M. A. (2001). Modeling functions of striatal dopamine modulation in learning and planning. Neuroscience, 103(1), 65–85.

Suri, R. E., & Schultz, W. (1998). Learning of sequential movements by neural network model with dopamine-like reinforcement signal. Experimental Brain Research, 121(3), 350–354.

Suri, R. E., & Schultz, W. (1999). A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. Neuroscience, 91(3), 871–890.

Surmeier, D. J., Ding, J., Day, M., Wang, Z., & Shen, W. (2007). D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. Trends in Neurosciences, 30(5), 228–235.

Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: An introduction (Vol. 1). Cambridge: Cambridge University Press.

Tanaka, S. C., Schweighofer, N., Asahi, S., Shishida, K., Okamoto, Y., Yamawaki, S., & Doya, K. (2007). Serotonin differentially regulates short- and long-term prediction of rewards in the ventral and dorsal striatum. PLoS One, 2(12), e1333. https://doi.org/10.1371/journal.pone. 0001333.

Tian, L., Arnold, M., Sejnowski, T., & Jabri, M. (2003). A biologically inspired computational model of the block copying task. Paper presented at the Proceedings of the third international workshop on Epigenetic robotics, Lund University Cognitive Studies. Wickens, J., & Kötter, R. (1995). Cellular models of reinforcement.

Wickens, J. R. (1997). Basal Ganglia: Structure and computations [Invited Review]. Network: Computation in Neural Systems, 8, R77–R109. Wickens, J. R., Horvitz, J. C., Costa, R. M., & Killcross, S. (2007). Dopaminergic mechanisms in actions and habits. Journal of Neuroscience, 27(31), 8181–8183.

Wilson, C. J. (2004). Basal ganglia. In G. M. Shepherd (Ed.), The synaptic organization of the 136 brain (pp. 361–413). New York: Oxford University Press.

Wise, R. A. (2004). Dopamine, learning and motivation. Nature Reviews Neuroscience, 5(6), 483– 494. Wise, R. A., & Rompre, P.-P. (1989). Brain dopamine and reward. Annual Review of Psychology, 40(1), 191–225.