

# BG 建模概述

## 目录

一、引言.....	3
二、BG 模型的生物学基础.....	5
2.1.1 纹状体.....	6
2.1.2.STN.....	8
2.1.3.GPe.....	9
2.1.4.GPi.....	9
2.1.5.SNr.....	10
2.1.6.SNc.....	10
2.3.1.STN-GPe 环路层作为随机性的来源.....	11
2.3.2.多种前馈途径(Multiple feed-forward pathways).....	12
2.3.2.1“皮质-BG”系统多个平行回路.....	12
2.3.2.2 直接回路和间接回路.....	13
2.3.2.3.超直接回路.....	14
2.3.2.3.1.“皮质-STN-苍白球“超直接”通路.....	14
2.3.2.3.2.“超直接”通路的功能性意义.....	15
2.3.3.反馈路径(Feedback pathways).....	16
2.3.4.完整循环(Completing the loop).....	16
2.7.1.多巴胺能信号(dopaminergic signal).....	21
2.7.2.胆碱能信号(cholinergic signal).....	22
2.7.3.其他强化信号.....	22
三、BG 模型的演变与革命(evolution and revolutions).....	23
3.1.运动准备.....	24
3.2 动作选择(action selection).....	25
3.3.序列生成(sequence generation)与依序处理.....	32
3.4.Go/NoGo 模型.....	35
3.5.单通道模型(Single pathway models).....	37
3.6.多路径模型(multiple pathway models).....	37
四、BG 的计算模型.....	38
4.1.从机器人到膜.....	38
4.1.1 系统层面的模型.....	39
4.1.2.微电路模型.....	42
4.1.3.基于电导的单神经元模型.....	43
4.1.4.模型间的关系.....	44
4.1.5.分析.....	44
4.2.BG 中的动作选择的计算模型：I.一个新的功能解剖.....	44
4.2.1. BG 解剖结构和生理机理.....	45
4.2.2. 基底节功能机制.....	47
4.2.3.选择过程.....	47
4.2.4.神经网络选择.....	48

4.2.5.模型建立.....	50
4.2.6.讨论.....	53
4.3.BG 动作选择的计算模型。II.行为分析与模拟.....	54
4.3.1.量化新的功能架构.....	54
4.3.2.分析结果.....	57
4.3.2.2.1.保序映射.....	58
4.3.2.2.2.尺度缩放.....	59
4.3.2.2.3.竞争力.....	60
4.3.2.2.4.选择性.....	60
4.3.3.模拟结果.....	61
4.3.4.讨论.....	64
4.4.BG 的机器人模型：动作选择与内部处理.....	66
五、强化学习(Reinforcement learning).....	67
5.4.1.“Actor-Critic”体系结构.....	75
5.4.2.BG 强化学习的 Actor-Critic 模型.....	76
5.4.2.1.Houk,Adams 和 Barto(1995).....	76
5.4.2.2.Berns 和 Sejnowski(1996)模型.....	77
5.4.2.3.Suri 和 Schultz(1998、1999).....	78
5.4.2.4.Contreras Vidal 和 Schultz(1999).....	80
5.4.2.5.Brown、Bullock 和 Grossberg(1999).....	80
5.4.2.6.Suri、Bargas 和 Arbib(2001).....	80
5.7.1.DA 神经元的兴奋和抑制源.....	82
5.7.2.DA 信号在强化学习和 Actor 转换中的双重作用.....	83
5.7.3.纹状小体与适应的 Critic.....	85
5.7.4.结论.....	85
六、降维(Dimensionality reduction).....	86
七、BG 基础强化驱动降维模型.....	90
7.1.模型.....	90
7.2.结果.....	91
7.2.1.关联性(Correlation).....	91
7.2.2.横向连通性(Lateral connectivity).....	92
7.2.3.信息编码(Information encoding).....	92
7.2.4.强化信号(Reinforcement signal).....	93
7.2.5.病理学(Pathologies).....	93
八、BG 强化驱动降维模型的建立.....	93
8.1.约束权重(Constrained weights).....	94
8.2.非线性元件(Non-linear elements).....	94
8.3.多层(Multiple layers).....	95
8.4.部分闭环(Partially closed loop).....	96
8.5.稀疏连接(Sparse connectivity).....	97
8.6.多种途径(Multiple pathways).....	98
8.7.增强驱动的降维——BG 奖励偏向表示.....	98
九、在 BG 的同步与探索——尖峰网络模型.....	100
9.1.BG 脉冲神经元模型.....	101

9.2.二元动作选择任务.....	101
9.3. N-臂强盗任务.....	101
9.3.1 行为模型【改编自 Bourdaud 等人(2008)】 .....	102
9.3.2.自动控制机械选择策略.....	103
9.4.对策.....	104
9.4.1.同步.....	104
9.5.使用竞赛模型的动作选择.....	104
9.6.结果.....	105
9.6.1.神经动力学.....	105
9.6.2.决策.....	105
9.7.分析.....	107
十、结论(Conclusion).....	107

## 一、引言

*BG*(Basal ganglia, *BG*)位于大脑皮质(*Cerebral cortex*)的内部,是大脑的深层核,它们接收来自所有大脑皮层(*the pallium*)的显著输入,包括等厚皮质(*Isocortex*)和异厚皮质(*Allocortex*)(Swanson 2000)。一般来说,这些输入形成相对离散的通道,循环回到它们起源的大脑皮层的同一区域,这一特征类似于小脑(*Cerebellum*)和大脑皮层之间的循环(Middleton 和 Strick 2000)。图 1 总结了这种循环结构,为 *BG* 模型的输入提供了重要的研究视角。在深入了解 *BG* 之前,回顾一下解剖学(*Anatomy*)和生理学(*Physiology*),便可得出 *BG* 涉及多种重要功能,如(1)基于奖励的学习 (*reward-based learning*)、(2)探索行为(*exploratory behavior*)、(3)目标导向行为(*Goal-oriented behavior*)、(4)运动准备(*motor preparation*)、(5)工作记忆 (*working memory*)、(6)计时(*timing*)、(7)动作门控(*action gating*)、(8)动作选择(*action selection*)、(9)疲劳(*fatigue*)和(10)冷漠(*apathy*)。其中两个基本功能是动作选择(*action selection*)和强化学习(*Reinforcement learning*)。

同时,运动相关皮质区如何调节 *BG* 输出核的活动是理解 *BG* 运动控制机制的重要问题。皮质-下丘脑(*Subthalamus, STN*)-苍白球(*Globus pallidus*)“超直接”通路(*Hyperdirect pathway*)将运动相关皮质区的强烈兴奋效应传递到苍白球,绕过纹状体(*Striatum*),传导时间比通过纹状体传递的效应短。本综述会着重强调了“超直接”通路的功能意义,并描述了一个动态的 *BG* 功能“中心环绕模型(*Central Surrounding Model*)”,用于控制自主性肢体运动。当一个自主性运动,即将由皮质机制启动时,通过皮质-下丘脑-苍白球“超直接”途径传递的一个必然信号(*corollary signal*)首先会抑制丘脑和大脑皮层的大片区域,这些区域与选定的运动程序和其他竞争程序都有关。然后,通过皮质-纹状体-苍白球“直接”途径的另一个必然信号抑制其目标并仅释放选定的运动程序。最后,第三个可能通过皮质-纹状体-外苍白球-下丘脑-内苍白球“间接”途径的必然信号广泛抑制其靶点。通过这种顺序信息处理,只有选定的电机程序在选定的时间被启动、执行和终止,而其他竞争程序则被取消。

已知 *BG* 回路在各种动作中起着作用,从运动开始到高层次认知过程,都是基于奖励的学习,而动作选择是这些中枢大脑结构的主要作用。在这里,总结了一些用于自主动作选择的 *BG* 回路的生物物理学现实模型。由于感觉运动系统与 *BG* 之间存在多个平行回路(*Parallel circuit*),因此有人提出,这些连接机制可以作为一种选择机制,解决在特定环境下可用且可选择动作之间的竞争问题。对这一假设的一个强有力的检验是,需要一个 *BG* 的计算模型

来在一个自治体中生成综合选择序列，因此，本综述提出了三个生物学上可信的 *BG* 固有处理模型，这些模型的计算前提是 *动作选择*。介绍的第一个模型是通过给定动作在标量值中的选择倾向(*显著性(Saliency)*)进行编码，表明 *动作选择* 可以根据信号选择进行更正(Gurney, Prescott, Redgrave, 2001)。在文中定义了信号选择的一般特性，并研究了这种计算的神经网络。这些网络与 *BG* 解剖的比较导致了 *BG* 结构新的功能分解成“选择”和“控制”路径。前一条路径本身通过 *环绕网络* 中心的 *前馈* 进行选择。控制通路调节选择通路的作用以确保其有效运行，并协同补充其多巴胺能调节。该模型与 *BG* 的主要功能分离形成了“*直接*”和“*间接*”路径。之后介绍了一个嵌入这种模型的机器人架构(Prescott, Montes-González, Gurney 等人, 2006)，并要求它在受动物观察启发的机器人任务中控制 *动作选择*。他们的结果证明，嵌入式模型在多感官和 *动机* 促进条件下，会有有效的 *动作选择*。当面对多个高度显著的替代方案时，机器人还表现出 *动作选择* 崩溃的情况，这与动物在冲突情况下的行为相似。该模型还揭示了最近关于 *行为转换* 和 *测序* 的神经生物学发现。最好介绍一个模型(Yucelgen, Denizdurduran, Metin 等人, 2012)，帮助理解复杂的过程，并提出了一个能够用改进的 *Hodgkin-Huxley* 神经元模型模拟 *直接(Direct)*、*间接(Indirect)* 和 *超直接* 路径的模型。该模型考虑的神经结构多于文献中类似的工作，能够模拟参与 *动作选择* 的神经结构中神经元的活动。所提出的模型具有通用性，因为所获得的模拟结果与之前发表的所考虑的神经结构的神经元活动记录类似。

多巴胺神经元似乎在预测奖赏时编码误差。它们被不可预测的 *奖励* 激活，不受预测的 *奖励* 的影响，并且在忽略预测的 *奖励* 时被抑制。在条件作用之后，他们以类似的方式对 *奖励* 预测刺激做出反应。具有这些特点的多巴胺反应与实现 *时序差分* 学习算法的神经网络模型的预测强化 *教学信号* 十分相似。本综述中给出了一个神经网络模型，该模型使用与多巴胺反应极为相似的 *奖赏预测 误差信号(error signal)* 来学习 *动作序列*。在序列的每一步中都有不同的刺激，需要不同的运动反应，并且 *奖励* 发生在正确执行序列的末尾。类多巴胺的 *预测强化信号* 有效地允许模型学习长序列。相比之下，在无 *条件强化信号(Reinforcement signal)* 下学习需要更长的 *突触 合格性追踪(eligibility trace)* 和生物学上不太可信的持续时间，才能获得令人满意的表现。因此，类多巴胺神经元信号是学习顺序行为的良好 *教学信号(Teaching signal)*。

我们不断面临着在 *探索(Explore)* 新信息的行动和 *运用(Exploit)* 现有知识的行动之间做出选择的两难境地。这种“*探索-运用*”的平衡取决于环境：稳定有利于“*运用*”知识以获得最大收益；波动性有利于“*探索*”新的选择和发现新的结果。在综述中，介绍了一种 *运用* 生物物理学和 *Izhikevich* 神经元的结合建立了一个混合模型，并验证了我们先前关于 *BG* 中 *Go-Explore-NoGo(GEN)* 机制的假设。混合模型由 *Hodgkin-Huxley* 型 STN 模型、GPe 模型、GPi 模型和 *纹状体 spiking* 模型组成。为了捕捉多巴胺(DA)对 *BG* 核动态的影响，通过 DA 调节 STN-GPe 与已知诱发突发性行为的 STN 中 T 型钙电流之间的突触权重。我们比较了混合模型与 *spiking-Izhikevich* 模型和速率编码模型对二元 *动作选择* 任务的结果。混合模型的结果进一步强化了 *GEN* 理论，表明“*探索*”水平取决于 DA 的水平。N 臂强盗任务的结果也表明，通过在 *spiking* 模型中将 *纹状体(D1)* 降低到 GPi 的权重，我们可以提高系统中的“*探索*”水平，从而降低模型获得的平均回报。将 N 臂强盗结果与速率编码和集总 Softmax 模型的结果进行了比较。

*奖励* 或 *惩罚* 产生的强化对于塑造人类和动物行为很重要，有相当多的证据表明，*BG* 的多巴胺(DA)系统是通过强化来调节学习的关键大脑机制(Beninger, 1983; Wise 和 Rompre, 1989; Wickens, 1990)，在工程控制中，强化也作为一种强有力的策略出现(Barto 等人, 1990; Werbos, 1992)。虽然 *强化学习* 有许多优点，但它有一个重要的局限性，称为 *信任分配问题(the credit assignment problem)*，这在学习应用中至关重要(Minsky, 1963; Barto 等

人,1983)。在生物学术语中,这是在正确的时间(时间信用分配)将强化信号传递到正确的突触(空间信用分配),以便有效地指导学习过程的问题。在这里,介绍了一个神经模型,通过检测事件预测随后的强化,解决了时间信贷分配问题。他们根据纹状体的纹状小体隔室的解剖和生理学(Gerfen 等人 1987; Graybiel,1991)以及 DA 神经元的信号特性介绍了该模型。然后,给出了一个概念基础,这是基于 Barto 所描述的适应性 Critics 理论。该模型的初步版本已在之前发布(Houk,1992)。

在本文中,首先描述了穿过 BG 的环路的计算结构的主要特征。然后回顾了具体的动作选择模型。其中一些模型在解剖学和生理学上受到限制,而另一些模型则是抽象的,但仍受 BG 行为功能的驱动。建模的层次也不同,从抽象的、基于人群的(population-based)、联通的(connectionist)、生物物理学的(biophysical)、分子的(molecular)。混合模型结合了不同神经科学水平的显著特征。

为理解 BG 功能打开大门的一个关键观点是, BG 中多巴胺能细胞的活动代表着奖赏信号(reward signaling)(Schultz,1998)。这种概念关联使得强化学习(Reinforcement learning, RL)概念(Sutton 和 Barto,1998)在 BG 研究中的应用成为可能。RL 是机器学习的一个分支,它研究的是在没有明确“老师指导”的情况下,“智能体(Agent)”如何学习对刺激做出最佳反应;“智能体”的学习过程是由环境对“智能体”行为的反应而产生的奖惩信号驱动的。今天,RL 在机器人领域有着出色的应用(DoriGo 和 Colombetti,1994)。计算 RL 实际上是受到“条件反射”作用的启发,条件反射作用是心理学的一个主题,在这个主题中,动物学习对感官刺激产生有益的反应(Sutton 和 Barto,1990)。导致奖励的反应得到加强,而惩罚的反应得到避免。这种对有回报的反应的搜索通常包括反复试验或探索过程。因此,如果假设 BG 辅助强化学习动态情况,则很容易预测其在探索行为中的作用。“奖励”是一个抽象的概念,在我们对 BG 功能提出一个观点之前,必须仔细定义它。虽然食物或果汁奖励是主要的奖励形式,但更抽象的(次要的)奖励形式可能被认为是成功地接近和达到目标状态的方法。大量的实验文献表明 BG 参与了目标导向行为。运动准备的定义通常很模糊,就像大脑在开始执行动作之前所做的那样。虽然 BG 一直被认为参与了运动的准备阶段,但对于 BG 在运动的准备阶段中的作用还没有令人信服的理论。我们将运动准备描述为 BG 和运动皮质相互作用的结果。目标导向行为要求目标状态以适当的工作记忆表征状态保持。越来越多的人认识到多巴胺能投射在维持和更新前额皮质(PFC)工作记忆表征中的作用(Cohen 等人,2002)。RL 的一个重要组成部分是预测未来的回报。预测与刺激-行动对相关的潜在奖励不仅包括预测奖励的存在/不存在,甚至是奖励的大小,还包括预测未来奖励发生的时间。因此,辅助 RL 机制的结构自然适合学习奖励未来事件的时机。确定对给定刺激的最有益的反应包括比较给定背景事件中的许多潜在行为,并选择最佳行为。BG 被称为脊椎动物大脑对动作选择问题的解决结构。因此,大多数 BG 功能至少可以在 RL 框架内进行概念上的调整,尽管在几个领域还需要更详细和更精确的知识。然而, BG 功能疲劳、冷漠等方面似乎并不容易在 RL 的网络中以任何明显的方式被捕获。这些为建模者在未来 BG 研究中提供了一些最困难的挑战。

在这篇文章中,我们回顾了大量关于 BG 功能计算模型的代表性文献。尽管我们努力提供一个全面的文献回顾和总结在神经解剖学这个复杂的模块的当代思维,我们真正的目标是收集许多建模工作链式线索,把它们统一在 RL 的结构体系之下,呈现出一个统一的,综合的 BG 功能。

## 二、BG 模型的生物学基础

BG 参与复杂的行为,需要在认知(Cognition)、动机(Motivation)和运动(Movements)之间

进行协调。这一作用与 *BG* 的解剖位置紧密相连, *BG* 是连接大多数皮质区域和额叶皮质 (*Frontal cortex*) 的神经元回路的中心部分。*BG* 由多个核组成, 每个核内的神经元和不同的核之间有复杂的相互作用。本节描述 *BG* 电路的主要问题, 与 *BG* 的过去和现在的模型有关。关于 *BG* 解剖、生理学和生物化学的更广泛和全面的综述可以参考其他文章 (Parent 和 Hazrati,1995;Gerfen 和 Wilson,1996;Bolam 等人,2000)。本综述的生物学背景主要基于灵长类和人类 *BG* 的组织。然而, 某些数据和见解也来自啮齿动物的工作。

## 2.1. *BG* 核

*BG* 由 7 个广泛连接的皮质下核组成: 尾状核 (*Caudate nucleus*)、壳核 (*Putamen*)、苍白球 (内部 (*Internal Globus Pallidus, GPi*) 和外部 (*External Globus Pallidum, GPe*))、底丘脑核 (下丘脑, *Subthalamic nucleus, STN*) 和黑质 (致密部 (*Substantia nigra pars compacta, SNc*) 和网状部 (*Substantia nigra pars reticulata, SNr*))。纹状体、苍白球、黑质和下丘脑核是 *BG* 的主要组成部分。其中部分在正常的自主运动中起着重要作用 (Gerfen 和 Wilson,1996)。这些成分进一步被分为亚核。纹状体由尾状核、壳核和腹侧纹状体 (*Ventral striatum*) (伏隔核 (*Nucleus accumbens*) 和嗅结节 (*Olfactory tubercle*)) 组成, 纹状体是 *BG* 的输入核。苍白球复合体由外部 (*GPe*) 和内部 (*GPi*) 段以及腹侧苍白球组成。最后, 黑质被分为致密部 (*SNc*) 和网状部 (*SNr*) (图 1), *SNc* 将多巴胺能神经元的轴突投射到 *STR* 上, *STR* 的神经元投射到 *GPi* 和 *GPe* 上。*STN* 与 *GPe* 之间存在兴奋性和抑制性连接。*STN* 神经元投射到 *GPi* 上。*GPi* 核和 *SNr* 核构成 *BG* 的输出核, 后者将氨基酸能投射到丘脑。多巴胺能细胞也存在于腹侧被盖区 (*VTA*), 即三叉神经中脑核 (中脑核)。伏隔核 (*NAc*) 是另一个重要的 *BG* 核, 与嗅结节一起构成腹侧纹状体的一部分。它被认为在奖赏处理、快乐、攻击性和成瘾方面有作用。*NAc* 接收来自前额皮质区域和 *VTA* 的输入。*VTA* 神经元在 *NAc* 上的末端是可卡因和安非他明等成瘾药物的作用区域 (Di Chiara 等人,2004)。背侧纹状体的 *SNc* 多巴胺能投射被认为参与了自主运动和习惯学习, 而腹侧纹状体的 *VTA* 投射有助于奖励和动机 (Keath 等人,2007)。多巴胺能投射到 *PFC* (Cohen 等人,2002) 和脊髓 (Heise 和 Kayalioglu,2008)。然而, 它们与脊髓没有直接的输入或输出连接。这些核从大脑皮层接收它们的主要输入, 并将它们的输出发送到脑干, 然后通过丘脑返回到前额叶、运动前皮质和运动皮质。因此, *BG* 的运动功能在很大程度上是由额叶皮层的运动区域介导的。

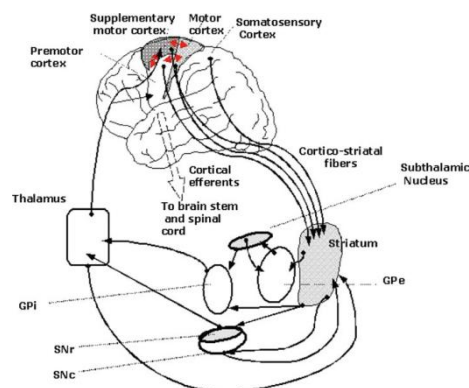


图 1: *BG* 运动功能的解剖学基础

### 2.1.1 纹状体

纹状体是 *BG* 的主要输入结构。分为背侧纹状体和腹侧纹状体。背侧纹状体进一步分为形成尾状核和壳核。腹侧纹状体是纹状体的腹侧延伸, 包括伏隔核、尾状核和壳核的内侧和

腹侧部分，以及嗅结节和前穿质(*Anterior perforated substance*)的纹状体细胞(Gerfen 和 Wilson,1996)。纹状体作为 BG 系统的主要输入团：接受①来自大脑皮层的谷氨酸能兴奋性投射；②来自内侧丘脑的兴奋性输入；③来自黑质致密部的多巴胺能投射；④来自中缝核的 5-羟色胺能(血清素)投射。

它从大脑皮层的大部分区域(Parent 和 Hazrati,1995)和丘脑的多核(McFarland 和 Haber,2000)接收大量和局部的谷氨酸(兴奋性)输入。第三个主要纹状体输入来自中脑(SNc 和 VTA)多巴胺能细胞(Haber 等人,2000)。投射神经元占纹状体神经元的绝大多数，被称为纹状体中的中棘神经元(MSN)(Gerfen 和 Wilson,1996)。这些中等大小(12-20 微米)的细胞有 25-30 个树突分支，被大量的棘覆盖(Difiglia 等人,1976;Wilson 和 Groves,1980)。这些分支从细胞体向各个方向辐射，最终形成 0.3-0.5mm<sup>3</sup>的球形体积(Gerfen 和 Wilson,1996)。细胞内研究表明，这些细胞的膜电位具有奖励(*up-down states*)(Wilson 和 Kawaguchi,1996;Stern 等人,1997;Plenz,2003)。这种双态膜电位转变的发生证据来自于清醒猴子的数值模拟和电生理记录(Kitano 等人,2002)。MSN 的超极化(*Hyperpolarization*)静息膜电位(*Membrane potential*)、较低的输入电阻和双稳态(*Bistability*)行为，MSN 的奖励之间的转换在很大程度上是由外部连接控制的(Wilson,1993;Wickens 和 Wilson,1998)。这导致他们在以零增长率为特征的下降状态与以上升状态为特征的高增长率(20-40 峰/秒)的短爆发之间发生转变(Crutcher 和 DeLong,1984b;Hikosaka 等人,1989;Lee 和 Assad,2003)。MSN 使用氨基丁酸(GABA)(Bolam 等人,1985)，通常被认为是一种抑制性神经递质(*Neurotransmitters*)(Tremblay 和 Filion,1989)，作为主要递质并投射到苍白球的两个部分和 SNr(Parent 和 Hazrati,1995)。纹状体-黑质 MSN 表达 D1 类多巴胺能受体(D1 和 D5)，D1 型介导时，多巴胺能传递主要是兴奋性的，而纹状体苍白球 MSNs 表达 D2 类(细分为 D2、D3 和 D4)，D2 型介导时，主要发挥抑制作用。

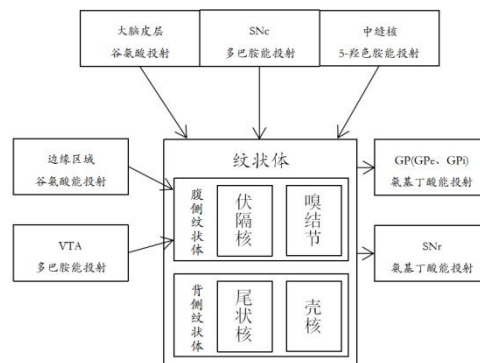


图 2:纹状体结构及其连接。纹状体作为基底神经节系统的主要输入团：接受①来自大脑皮层的谷氨酸能投射；②来自内侧丘脑的兴奋性输入；③来自黑质致密部的多巴胺能投射；④来自中缝核的 5-羟色胺能(血清素)投射。主要的组成神经元有①氨基丁酸能中等多棘神经元(MSN, 12-20 微米),占总数的 95%以上;②无棘胆碱能中间神经元(约 2%) ;③氮能神经元;④帕伐布明神经元;⑤钙素神经元。电生理上，MSNs 的特点是超极化静息膜电位、较低的输入电阻和双稳态行为，纹状体-黑质 MSNs 表达 D1 类多巴胺能受体 (D1 和 D5)，D1 型介导时，多巴胺能传递主要是兴奋性的，而纹状体苍白球 MSNs 表达 D2 类(细分为 D2、D3 和 D4)，D2 型介导时，主要发挥抑制作用。这些细胞有 25-30 个树突分支，被大量的棘覆盖，这些分支从细胞体向各个方向辐射，最终形成 0.3-0.5mm<sup>3</sup>的球形体积无棘神经元，进一步分为快突神经元(FS)、低阈值突神经元(LTS)和胆碱能神经元(TAN)——最大的。

纹状体包含几种类型的中间神经元(Kawaguchi 等人,1995;Wilson,1998;Haber 和 Gdowski,2003)，但是我们只讨论其中两种。棘突 II 型(Aspiny type II)神经元是纹状体细胞中最大的，约占纹状体总数的 1-2%。这些是胆碱能间神经元，可通过其特征、自发放电活动(3-10Hz)来识别(Kimura 等人,1984;Hikosaka 等,1989;Wilson 等人,1990;Aosaki 等人,1995)，因此被称为张力活性神经元(*tonically active neurons, TAN*)。TAN 的自发放电模式大体上与这些细胞内在膜特性相关(Bennett 和 Wilson,1999;Bennett 等人,2000)。这些特性是由突触输入



改变的，这些突触输入对峰值序列的时间调节有影响(Aosaki 等人,1994;Raz 等人,1996)。此外，TAN 只需要相对少量的外源性突触输入直接影响它们的 *spiking* 模式(Bennett 和 Wilson,1999)。细胞外记录显示 TAN 编码与强化或激励行为相关的信息(见第 2.6 节)。另一种类型的纹状体间神经元是 parVAbumin-阳性氨基丁酸能快速尖峰间神经元(FS)(Bolam 等人, 2000)，约占纹状体神经元种群的 3%。这些细胞具有间隙连接，表明它们在连续网络中耦合(Koos 和 Tepper,1999;Bolam 等人,2000)。因此，尽管数量相对较少，但它们对中等多棘投射神经元的抑制可能特别有效。不幸的是，氨基丁酸能中间神经元不能被标准的细胞外记录方法识别，而且在行为过程中有关其调制的数据仍然缺乏。组织化学和追踪研究已经证明了纹状体中与传输器相关的分子和传入末端分布模式的不连续性(Kunzle,1975;Goldman 和 Nauta,1977;Yeterian 和 Vanhoesen,1978;Graybiel 和 Ragsdale,1980)。乙酰胆碱酯酶(AChE)-缺乏纹状体的区域(称为纹状体)被密集染色的“基质”包围(Graybiel 和 Ragsdale,1978)。纹状体中这种分隔组织的重要性仍在争论中(Gerfen,1989)。在任何情况下，纹状体内大多数类型纹状体神经元的树突的形状和范围都受到纹状体室的大小和形状的限制(Gerfen,1985;Malach 和 Graybiel,1986;Penny 等人,1988)。相反，胆碱能中间神经元的树突很容易穿过纹状小体/基质边界。因此，这些中间神经元可能有助于在纹状体间建立联系(Graybiel 等人,1994)。

### 2.1.2.STN

BG 的第二个主要输入结构是 STN，它接收来自额叶皮质和皮质躯体运动区域的谷氨酸能投射(Monakow 等人,1978;Nambu 等人,1996)。与纹状体相比，STN(体积和细胞数)要小得多，灵长类动物的比例为 1:60(Yelnik,2002)，啮齿动物的比例为 1:200(Oorschot,1996)。然而，它在正常和病理条件下的 BG 活动中起主要作用(Carpenter 等人,1950;Kitai 和 Kita,1986;Bergman 等人,1990;Nambu 等人,2002b)。与大多数 BG 结构一样，STN 主要由投射神经元组成，仅包含少数小的中间神经元(Rafols 和 Fox,1976;Yelnik 和 Perperceron,1979;Chang 等人,1983)。投射细胞在运动过程中是活跃的(20-30 峰/秒)和短暂的爆发(Matsumura 等人,1992;Wichmann 等人,1994)。STN 谷氨酸能(兴奋性)投射神经元投射到苍白球的两个部分和 SNr(Parent 和 Hazrati,1993;Smith 等人,1998a)。底丘脑核的作用是刺激 SNr-GPi 复合体，它是间接途径的一部分。底丘脑核从苍白球的外部接收抑制性输入，并向 GPi 发送兴奋性输入。在最近的观点中，纹状体接收直接的兴奋性皮质输入，并投射到输出核，即苍白球内段(GPi)和黑质网状部分(SNr)均参与两个主要投射系统，“直接”和“间接”路径(Albin 等人,1989; Alexander 和 Crutcher,1990)。这种“直接”通路来自含 P 物质的氨基丁酸能纹状体神经元，单合性地投射到 GPi-SNr 上。“间接”途径来源于含脑啡肽的氨基丁酸能纹状体神经元，并通过一系列涉及 GPe 和 STN 的连接，多合性地投射到 GPi-SNr。因此，STN 被认为作为“间接”通路的中继核，占据着关键位置。此外，越来越多的证据表明，STN 可被视为除纹状体外的 BG 的另一输入站(Mink 和 Thach,1993; Kita,1994; Mink,1996; Levy 等人,1997; Nambu 等人,2000b)，因为 STN 接收到直接的皮质投射，特别是来自纹状体的皮质投射(Hartmann-von Monakow 等人,1978; Kitai 和 Deniau,1981; Nambu 等人,1996、1997)。

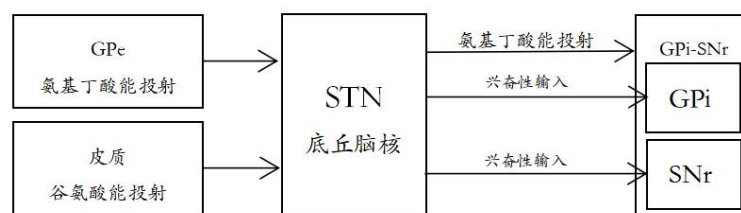




图 3:底丘脑核结构及其连接。它接收来自额叶皮质和皮质躯体运动区域的谷氨酸能投射。STN 谷氨酸能(兴奋性)投射神经元投射到苍白球的部分和 SNr。底丘脑核的作用是刺激 SNr-GPi 复合体,它是间接途径的一部分。底丘脑核从苍白球的外部接收抑制性输入,并向 GPi 发送兴奋性输入。纹状体接收直接的兴奋性皮质输入,并投射到输出核,即苍白球内段(GPi)和黑质网状部分(SNr)均参与两个主要投射系统,“直接”和“间接”路径。这种“直接”通路来自含 P 物质的氨基丁酸能纹状体神经元,单合性地投射到 GPi-SNr 上。“间接”途径来源于含脑啡肽的氨基丁酸能纹状体神经元,并通过一系列涉及 GPe 和 STN 的连接,多合性地投射到 GPi-SNr。因此,STN 被认为作为“间接”通路的中继核,占据着关键位置。

### 2.1.3.GPe

传统上, GPe 被认为是 BG 内的中继站,接收纹状体的输入并投射到 STN(Albin 等人,1989;Alexander 和 Crutcher,1990)。新数据的积累表明,除了传统的连接之外,还有对纹状体的 GPe 和 parVAalbumin-阳性氨基丁酸能中间神经元的 STN 输入和反投射(kita 等人,1999;Bolam 等人,2000)。因此,GPe 与 BG 的输入结构,即 STN(Carpenter 和 Strominger,1967)和纹状体相互连接。此外, GPe 直接投射到 SNr(Sato 等人,2000),其投射与 GPi 神经元的细胞体形成密集的氨基丁酸能突触接触(Hazrati 等人,1990;Shink 和 Smith,1995;Smith 等人,1998a)。因此,在本综述中,我们将强调 GPe(或苍白球(GP))——相当于 GPe 的啮齿动物作为 BG 内在电路的一部分的作用。

GPe 中观察到的绝大多数神经元是大的投射神经元,树突厚而光滑(无脊椎)且长(可达 1000 微米)(Fox 等人,1974;Difiglia 等人,1982;Park 等人,1982;Francois 等人,1984)。细胞内生理学研究还发现,苍白球中有两到三种类型的神经元,具有单一的优势类型(Nakanishi 等人,1990;Nambu 和 Llinas,1997;Cooper 和 Stanford,2000)。细胞外记录的苍白球刺激活动显示,大多数(>85%)的 GPe 神经元有一个高频率(50-70 峰/秒)放电暂停(Delong,1972)。与纹状体一样,苍白球投射神经元的主要传输器是 GABA(Oertel 和 Mugnani,1984)。

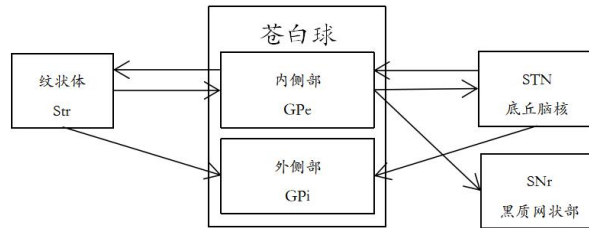


图 4:苍白球结构及其连接。主要包含氨基丁酸能的神经元,因此对其目标具有抑制作用。这些神经元在没有输入的情况下以稳定的高频率放电,纹状体发出的信号使它们暂停或降低放电速度。从而减少苍白球细胞对其目标的强直抑制(去抑制),增加目标的放电率。GPi 是 BG 最常见的输出端口,它接收来自表达 D1 的纹状体 MSNs 和 GPe 的抑制性氨基丁酸能输入,以及来自 STN 的兴奋性谷氨酸输入。GPe 主要接受纹状体的输入,并投射到底丘脑核。

### 2.1.4.GPi

从纹状体接收氨基丁酸能输入,从 STN 接收 GPe 和谷氨酸能输入的 GPi,被称为 BG 的输出站。GPi 通过“直接”和“间接”途径从纹状体接收信号。从 GPi 项目到丘脑中继核(丘脑腹侧核(ventro-lateral thalamic nucleus)、丘脑腹前核(ventral anterior thalamic nucleus)、内侧背核(medio-dorsal nucleus, MD)和丘脑外侧核(lateral habenular nucleus))以及层内(非特异性)丘脑核、中心体和锁骨旁核的氨基丁酸能输出(Kuo 和 Carpenter,1973;Kim 等人,1976;Parent 等人,2001)。GPi 神经元的大多数解剖、生理和生化特征与 GPe 相似。(Fox 等人,1974;Difiglia 等人,1982;Park 等人,1982;Francois 等人,1984;Nakanishi 等人,1990)。然而,与 GPe 神经元不同的是,几乎所有的 GPi 神经元都有高频率(60-80 峰/秒)放电,没有停顿(Delong,1972)。卵圆内核(EP)相当于啮齿动物的 GPi。

GPe 和 GPi 主要包含氨基丁酸能的神经元,因此对其目标具有抑制作用。这两个部分参与不同的神经回路。苍白球神经元运用去抑制原理工作。这些神经元在没有输入的情况下以

稳定的高频率放电, 纹状体发出的信号使它们暂停或降低放电速度。由于苍白球神经元本身对其目标具有抑制作用, 因此纹状体输入苍白球的净作用是减少苍白球细胞对其目标的强直抑制(去抑制), 增加目标的放电率(Firing rate)。

### 2.1.5.SNr

通常认为 SNr 是基于形态学、连接性、生物化学和生理学的 GPi 尾状延伸(Schwyn 和 Fox,1974;Kitai,1981;Yelnik 等人,1987;Francois 等人,1987)。与苍白球一样, SNr 主要由具有长而厚树突的大投射神经元组成, 树突几乎完全被纹状体的突触接触所覆盖(Yelnik 等人,1987;Francois 等人,1987)。然而, 其他证据也指出了 SNr 和 GPi 之间的一些差异。SNr 似乎不是来源于与外部或内部苍白球段相同的发育胚(Merchand 等人,1986)。此外, SNc 腹侧层的多巴胺能细胞具有树突, 可深入 SNr。因此, 多巴胺的树状体延伸(Jaffe 等人,1998)以及 GABA 对 SNc 神经元的影响表明, 与 SNc 多巴胺能神经元的 SNr 关系比 GPi 更直接和复杂(CelaDA 等人,1999)。最后还有对眼睛和注意力转移运动有重要影响的 SNr 上丘投射(Hikosaka 和 Wurtz,1983;Redgrave 等人,1992;Handel 和 Glimcher,1999;Jiang 等人,2003)。然而, 由于其功能的一般相似性, 我们将把 GPi/SNr 视为 BG 的单一输出阶段。

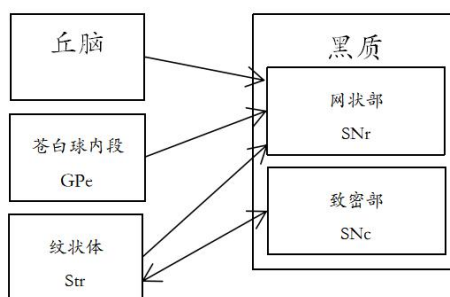


图 5:黑质结构及其连接。黑质是大脑功能的重要参与者, 尤其在眼球运动、动作计划、寻求奖励、学习和成瘾等方面。网状部的氨基丁酸能神经元自发引发动作电位。致密部最突出的功能是运动控制, 主要参与涉及学习的对刺激的反应。近年来发现, 致密部疑似与调节睡眠周期有关。SNr 通过抑制性 GABA 能投射, 从表纹状体 MSNs、GPe、STN 和 SNc 的多巴胺能细胞与上丘、丘脑腹侧部分和 SNc 接收输入。SNr 主要参与了眼动运动。SNc 是黑质-纹状体通路的一部分, 是中脑多巴胺能细胞簇之一。SNc 的多巴胺能投射针对包括纹状体 (背侧)、GPe 和 STN 在内的大脑多个区域, 特别是在 GPe 和 STN 中调节活动模式。

### 2.1.6.SNc

SNc 和其他中脑多巴胺能神经元从许多结构接收输入, 包括纹状体(Heimer 等人,1982;Haber 等人,2000)、STN 和边缘系统(Haber 和 Gdowski,2003)。中脑多巴胺系统分为背侧层和腹侧层细胞(Haber 等人,2000)。背侧层细胞包括 VTA(啮齿类动物的 A10)、红细胞后群(A8)和 SNc 的背侧细胞(A9)。腹侧层细胞包括致密细胞群和延伸至网状部的细胞柱。然而, 在不同区域的 SNc 和 VTA 上记录的 DA 神经元的反应没有显著差异(Schultz,1998), 因此在本综述中, 我们将把中脑 DA 神经元视为一个均匀的神经结构, 向其传出结构报告相同的信息。美国的电生理研究揭示了两种不同的模式: 单尖峰或迸发(Grace 和 Bunney,1984a,b)。在行为灵长类动物中, 多巴胺能细胞倾向于以较低的速率(4-10 峰/秒)自发地释放, 其释放速率的升高和抑制与行为和预测之间的不匹配有关(Schultz,1998)(第 2.6 节)。SNc 和 VTA 的多巴胺能投射终止于纹状体 MSNs 的棘和树突轴(Freund 等人,1984;Smith 等人,1994;Hanley 和 Bolam,1997)。电子显微镜研究表明, 皮质末梢经常出现在尖棘的头部, 多巴胺末梢在附近, 但在树突轴上(Dube 等人,1988;Smith 等人,1994)。纹状体中多巴胺与丘脑末梢的关系尚不清楚, 本文未作讨论。

## 2.2.BG 的输入和输出

BG 接收来自大脑皮层大部分感觉运动区域的输入，包括初级和次级躯体感觉区、初级运动皮层(M1)和各种运动前区，包括补充区、背侧和腹侧运动前区。BG 运动功能的解剖学基础如图 1 所示。皮质部分负责运动，即补充运动区(SMA)、前运动区(PM)、初级运动区(M1)、体感皮质和顶上小叶，向壳核(BG 的输入核)的运动部分作出密集的、有拓扑组织的投射。这个通路的输出，称为 BG 的运动电路，主要返回到 SMA 和 PM 皮质。这些区域相互联系，并与运动皮层相互联系，这三个区域都有直接的下行投射到脑干运动中心和脊髓。

纹状体(STR)是皮层向 BG 输入信号的主要靶点。在 BG 的“皮质-纹状体-苍白球”通路中，纹状体输出预测形成两个截然不同的平行通道。因为它们对靶核，即丘脑的影响，这些投射被称为直接和间接的途径。直接路由纹状体输出神经元对 GPi 神经元的抑制投射形成。纹状体神经元的激活抑制 GPi 中的神经元，GPi 抑制丘脑核。从概念上讲，直接通路可以看作是一个常闭的运动门。这个门是由纹状体活动打开，抑制苍白球的输出，允许出现运动。间接通路是通过抑制 GPe 的抑制投射而形成的，与直接通路中 GPi 神经元的抑制投射作用相反。间接通路的活性倾向于增加 GPi 细胞的活性，因此通过解除丘脑下核的抑制，关闭了“门”。这些通路可能参与调节运动参数(Contreras-Vidal 和 Stelmach,1996)。

除了皮质，丘脑还向纹状体发送大量的反馈投射。丘脑尾端髓板内核、中心核(CM)和束旁核(Pf)是丘脑纹状体投射的重要来源。CM 主要作用于壳核，而 Pf 主要作用于尾状核和 NAc (Smith,2008;Arbib 等人,1998)。

纹状体神经元由两种多巴胺受体组成：D1 和 D2 (Clark 等人,2005)。当 D1 受体被 SNc 的多巴胺信号激活时，直接通路被激活。同样，D2 受体被激活时，间接通路也被激活。此外，纹状体多巴胺的增加使平衡转向直接途径，从而增加整体运动活动。因此，间接途径是正常的活性途径。在运动开始前，多巴胺释放到纹状体激活了直接通路，平衡就被改变了 (Clark 等人,2005)。图 2 显示了 BG 的功能解剖。

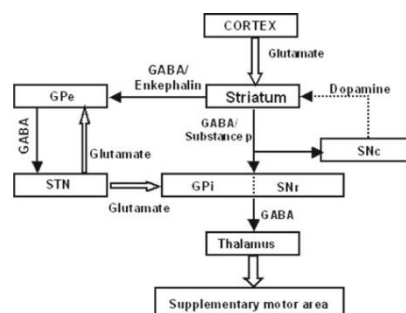


图 6: BG 功能解剖。

### 2.3.BG 通路

神经系统的许多区域的特点是相互联系。一个很好的经验法则是，如果从 A 区到 B 区有连接，那么从 B 区到 A 区也会有解剖连接。这就是大脑皮层 (Abels,1991;Scannell 等人,1995;Sporns 等人,2002) 和丘脑皮质环 (Sherman 和 Guillery,2001;McFarland 和 Haber,2002)。然而，BG 连接的经典图像与这种相互连接规则形成了鲜明的对比。BG 的主要通路形成一个前馈网络，从皮层到 BG 的输入核(纹状体和 STN)，从纹状体和 STN 到 GPe，从这三个结构到 BG 的输出核(GPi 和 STN)呈单向连接。最后，GPi 和 SNr 对丘脑的靶核进行单向投射。前馈结构在丘脑和皮质相互连接的网络终止时发生变化。这些通路最终通过投射到丘脑和额叶皮质，将输入信号传回 BG 的输入核，形成一个部分闭合的回路。

#### 2.3.1.STN-GPe 环路层作为随机性的来源

Plenz 和 Kitai(1999)在对 BG 器官型组织培养的研究基础上提出, STN 和 GPe 结构中都可能出现相关的活性, 并且相关的活性是由这两个结构之间的相互作用而不是由外部源驱动的。Terman 等人(2002)提出 GPe 内连接减弱和纹状体输入增强的组合为同步 STN-GPe 振荡和相关节律性 STN 输出奠定了基础。关于 GPe-STN 和 STN-GPe 投射的空间组织, 以及这两个核是否以“互反”或“失域”的方式相互投射, 目前缺乏精确的解剖学信息。Hurtado 等人(1999)记录了清醒帕金森患者的神经元活动, 并观察了 GPi 内的一些成对记录位点, 它们显示出短暂的同步。Bergman 等(1998)报道了 MPTP 治疗猴子大脑皮层活动的动态同步。最近的实验研究显示, 放电具有显著的低频周期性(4-30Hz), 并显著增加了 GPe 和 STN 神经元之间的相关性(Bergman 等人,1994; Nini 等人,1995; Magnin 等人,2000; Raz 等人,2000; Brown 等人,2001)。对 STN 和 GPe 活性的实验研究表明, 在多巴胺耗尽的情况下(类似于帕金森条件), 这些核的活性虽然没有显著降低放电率, 但神经元之间的相关性却显著增加(Bergman 等人,1994; Brown 等人,2001)。在建模研究中, STN-GPe 回路神经元的相关活动在功能上与帕金森震颤频率(Terman 等人,2002)和帕金森笔迹扭曲(Gangadhar 等人,2008)有关。

STN-GPe 环路在正常 BG 中的复杂活性及其在帕金森条件下的损失被认为具有深刻的功能意义, 并被解释为被解释为 RL 所要求的随机勘探的来源(Sridharan 等人,2006)。Sridharan 等人(2006)的模型描述了一个模拟的莫里斯水池实验, 其中一只虚拟老鼠在可见地标的帮助下探索一个隐藏的平台。当平台(即与之相关的地标集)不在“智能体”的视野中时, 模型的 STN-GPe 部分表现出不相关的活动, 反映了探索性行为。当平台(即与之相关的一组地标)不在智能体视野之内时, 模型的 STN-GPe 部分表现出不相关的活动, 反映了探索性行为。

Doya(2002)从 RL 的角度讨论了各种神经调节因子(多巴胺、血清素、去甲肾上腺素和乙酰胆碱)在脑功能中的可能作用, 并假设: (i)多巴胺是预测奖励和强化行为的全局学习信号, (ii)血清素控制着对奖励的短期和长期预测之间的平衡, (iii)去甲肾上腺素控制着广泛探索和集中执行之间的平衡, (iv)乙酰胆碱控制着记忆存储和更新之间的平衡。特别地, 为了支持去甲肾上腺素在探索中的假设作用, Doya 指出蓝斑(LC)中的去甲肾上腺素能神经元在紧急情况下被激活(Doya,2002)。此外, 我们知道, 在刺激呈现时, LC 神经元的相态反应与反应的高精度相关(Aston-Jones 等人,1994)。其他作者也提出去甲肾上腺素与“探索-运用”问题之间存在联系(Usher 等人,1999)。也有证据表明去甲肾上腺素与苍白球的活动水平有关(Russell 等人,1992)。这种感知与 STN-GPe 环路的观点非常一致, STN-GPe 环路在探索行为中扮演着重要角色(Sridharan 等人,2006)。

一个主要的理论问题涉及探索行为, 以及探索获取新信息和运用现有知识以获得回报之间的权衡。研究动物如何选择在“探索”和“运用”之间保持平衡是一个丰富的研究领域(Montague 和 Berns,2002)。已有研究表明, 多巴胺系统的活动与奖励预测误差无关, 但可能与确保适当的探索有关(Suri 和 Schultz,1999; Kakade 和 Dayan,2002)。多巴胺与强化学习的时间差误差之间的密切关系, 揭示了动物与人工决策系统之间新的、强有力的联系。

### 2.3.2.多种前馈途径(Multiple feed-forward pathways)

#### 2.3.2.1“皮质-BG”系统多个平行回路

早期的研究认为存在一个单独的“皮质-BG”回路, 所有的皮质区域都投射到 BG 上。通过各种解剖学和追踪研究的进一步研究, 我们发现“皮质-BG”系统确实由多个平行的回路组成, 其中皮质区域投射到 BG 的不同区域, 且大多不重叠(Alexander, DeLong 和 Strick, 1986; DeLong, Wichmann, 2010; Nakano, 2000)。平行回路主要分为运动回路(运动和动眼神经)、联想回路(*Associative circuit*)(背外侧和眶额前额皮质)和边缘回路。图 8 所示为各皮层区

域向各环路不同 BG 核的投射。

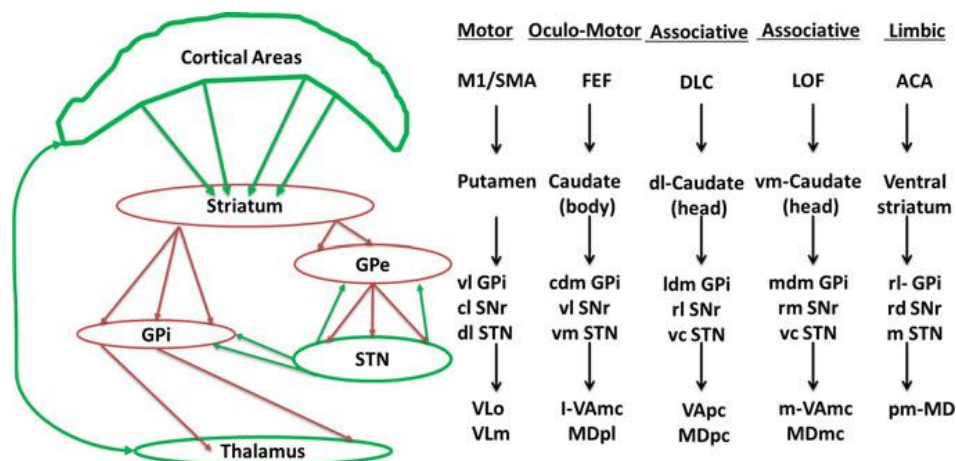


图 7: 此图显示了运动、动眼神经、联想和边缘区域的平行 BG 皮质环, 具体区域为 M1: 初级; SMA: 辅助运动区; FEF: 额叶视野; DLC/DLPFC: 背外侧前额叶皮质; OFC/LOF: 外侧前额叶皮质; ACA: 前扣带区; GPI: 苍白球; SNr: 黑质网状部; STN: 底丘脑核; MDPL: 中隔背侧副板层; MDMC: 中隔背侧厚朴细胞; MDPC: 中隔背侧厚朴细胞; VAMC: 中隔前厚朴细胞; VAPC: 中隔前厚朴细胞; VLM: 中隔腹侧侧部; VLO: 中隔腹侧部 teralis pars oralis; vp: 腹侧苍白球; vs: 腹侧纹状体; cl: 尾侧纹状体; cdm: 尾侧背内侧; dl: 背外侧; l: 外侧; ldm: 外侧背内侧; m: 内侧; mdm: 中隔背内侧; pm: 后内侧; rd: 罗斯特洛背侧; rl: 前外侧; rm: 罗斯特洛内侧; vm: 腹内侧; vl: 腹外侧。

虽然 BG 回路在功能上/解剖学上是分离的, 但这些单独的回路相互作用是很重要的, 以确保学习和信息传递发生在运动、认知和情感领域。最新的解剖学证据确实表明这些闭环之间存在相互作用(Haber 和 Calzavara, 2009)。下面将简要介绍各种用来解释多皮层闭环功能的理论。第一个假说认为, 信息在回路之间的传递是通过树突结构在回路之间的交叉来实现的。第二种是基于较小的 BG 结构的重叠, 它们之间也有侧突。这种结构创造了一个“边缘”, 在“边缘”处神经元对不止一种模态(运动/认知)作出反应(Yelnik, 2002)。第三种是提供方向信息流的复杂的非交互连接。例如, 已经观察到边缘纹状体通过纹状体-黑质通路影响啮齿动物的运动输出, 腹侧纹状体通过多巴胺能系统(Haber, Fudge 和 McFarland, 2000)影响背侧纹状体。最后一个是基于“热点”, 在一个结构内的解剖区域接收来自多个功能区域的输入, 从而实现一体化连接(Haber 和 Calzavara, 2009)。

### 2.3.2.2 直接回路和间接回路

纹状体中主要存在两种类型的 MSNs: 一种含有 P 物质和 GABA, 主要作用于 GPi 和 SNr; 一种含有脑啡肽和氨基丁酸, 主要用于 GPe(Gerfen 等人, 1990)。含有 P 物质高 mRNA 表达水平的 MSNs 也含有多巴胺 D1 受体的 mRNA, 含有脑啡肽的 mRNA 的细胞也含有 D2 受体亚型的高 mRNA 表达水平(Gerfen 等人, 1990; Aubert 等人, 2000)。不同途径内受体亚型的分离及其不同的药理作用和共定位神经肽, 在确定“直接”和“间接”途径的功能框架方面尤为重要(Albin 等人, 1989; Alexander 和 Crutcher, 1990; Gerfen 等人, 1990)(第 3.2 节)。在 20 世纪 80 年代早期, STN 被认为是 BG 电路中一个小的闭合辅助回路的一部分, 包括苍白球(GPe)的外段(Delong 和 GeorGopoulos, 1981)。在最近的观点中, 纹状体接收直接的兴奋性皮质输入, 并投射到输出核, 即苍白球内段(GPi)和黑质网状部分(SNr)均参与两个主要投射系统, “直接”和“间接”路径(Albin 等人, 1989; Alexander 和 Crutcher, 1990)。

直接通路: 皮层 → 纹状体 → SNr - GPi 复合体 → 丘脑 → 皮层 → 肌肉等 → (运动亢进状态)

间接通路: 皮层 → 纹状体 → GPe → 底丘脑核 → SNr - GPi 复合体 → 丘脑 → 皮层 → 肌肉等 → (运动减少状态)



在这个框架中，“直接”路径直接从纹状体通向 GPi，这种“直接”通路来自含 P 物质的氨基酸能纹状体神经元，单合性地投射到 GPi-SNr 上。而在“间接”途径来源于含脑啡肽的氨基酸能纹状体神经元，并通过一系列涉及 GPe 和 STN 的连接，多合性地投射到 GPi-SNr。“间接”路径中，信息流从纹状体流向 GPe，再流向 STN，再从 STN 流向 GPi(图 1a)。因此，STN 被认为作为“间接”通路的中继核，占据着关键位置。然而，在灵长类动物中，肽副递质(peptide co-transmitters)的分离并不完全。相反，GPe 的内侧边界含有 P 物质阳性免疫反应活性，而脑啡肽的免疫反应活性存在于 GPi 的内侧(Haber 和 Elde,1981;Reiner 等人,1999)。此外，最近的研究报告了共同表达多巴胺受体不同亚型的多发性硬化症亚群(Surmeier 等人,1996;Nicola 等人,2000)，这表明直接/间接途径不像以前想象的那样分离。因此，我们将把直接/间接概念的讨论推迟到本综述的高级主题部分(第 7.6 节)。P 物质和脑啡肽的功能作用仍存在疑问(Steiner 和 Gerfen,1998);因此，在本综述中，我们将不研究它们的神经调节作用。

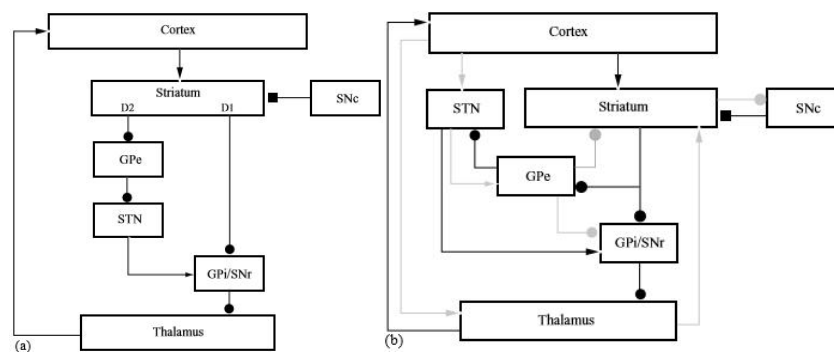


图 8:皮质-BG 网络。BG 不同通路的“盒箭”状网络。(a)早期的 Albin DeLong 网络。(b)最新的网络。早期的网络是黑色的，后来添加的是灰色的。谷氨酸能突触用箭头表示，氨基酸能突触用圆圈表示，多巴胺能突触用正方形表示。

### 2.3.2.3.超直接回路

除了皮质纹状体苍白球通道外，对啮齿类动物(Kita,1992;Ryan 和 Clark,1992)和灵长类动物(Nambu 等人,2000,2002b)的研究，还强调了从皮质到 STN 的“超直接”快速兴奋通道的功能显著性，其对输出核的影响。通常比纹状体“直接”和“间接”路径的作用快。在这篇文章中，是通过 STN 来描述这个新的 BG 回路通路，“皮质-STN-苍白球”超直接”通路(也可参见 Gerfen,2000)，并提出一个动态的“中心环绕模型”，用于控制自主性肢体运动。

超直接路径：皮质  $\xrightarrow{\text{刺激}}$  底丘脑核  $\xrightarrow{\text{刺激}}$  SNr-GPi 复合体  $\xrightarrow{\text{抑制}}$  丘脑  $\xrightarrow{\text{刺激/抑制}}$  皮质  $\xrightarrow{\text{刺激/抑制}}$  肌肉等

#### 2.3.2.3.1.“皮质-STN-苍白球”超直接”通路

STN 接收来自大脑皮层的输入，然后将输出发送到 GPi-SNr。皮质刺激在猴子的苍白球神经元中诱导一种早、短潜伏期的兴奋，随后是抑制和晚期的兴奋。根据以下发现(Nambu 等人,2000b)，早期兴奋被认为来自“皮质-STN-苍白球”通路。同时记录苍白球复合体和 STN 的神经元活动表明，皮质刺激在 STN 神经元中诱导了一种早、短潜伏期的兴奋，高于苍白球神经元。通过记录电极对 STN 的刺激可使苍白球神经元得到原位活化。通过注射肌肉注射(GABAA 受体激动剂)来阻断 STN 中的神经元活动，从而消除皮质刺激引起的苍白球早期和晚期兴奋。此外，通过向 STN 中注入(±)-3-(2-羧基哌嗪-4-基)-丙基-1-磷酸[CPP,N-甲基-D-天冬氨酸(NMDA)受体拮抗剂]来阻断谷氨酸性“皮质-STN”神经传递，抑制了苍白球神经元的早期兴奋，而干扰了 STN 的表达。在 STN 中注射双核谷蛋白(GABA<sub>A</sub> 受体拮抗剂)后，氨基酸能苍白球-STN 神经传递对早期兴奋没有影响。在麻醉大鼠的电生理实验中也获得



了类似的结果,即皮层刺激在苍白球(灵长类的 GPe)和 SNr 中诱发了三个阶段的反应,包括早期激发、抑制和晚期激发(Ryan 和 Clark,1991;Fujimoto 和 Kita,1992;Kita,1992,1994; Maurice 等人,1999),在 STN 永久性损伤(Ryan 和 Clark,1991;Kita,1992,1994)或阻断来自前额皮质的“皮质-STN”神经传递后,早期的兴奋消失(Maurice 等人,1999)。

皮质刺激对苍白球神经元的抑制作用被认为是由皮质-纹状体-苍白球“直接”途径介导的,因为(1)纹状体中的刺激诱发了苍白质神经元的抑制反应,皮质刺激诱发的抑制潜伏期与纹状体刺激诱发的抑制潜伏期之差与皮质纹状体传导时间相当(Yoshida 等人,1993);(2)通过全身注射氨基丁酸能阻滞剂(Kita,1992)或通过阻断皮质-纹状体神经传递(Maurice 等人,1999),抑制性突触后电位(IPSP)或皮质刺激诱发的抑制被消除。另一方面,大脑皮层刺激引起的苍白球神经元迟发性兴奋的起源是不明确的。这可以归因于净兴奋性皮质-纹状体-GPe-STN-GPi“间接”途径,因为在阻断皮质-纹状体或纹状体-GPe 神经传递后,类似的 SNr 神经元的后期兴奋反应显著降低(Maurice 等人,1999)。然而,向 STN 中注射 CPP 也减弱了晚期兴奋,表明“皮质-STN-苍白球”通道参与了苍白球神经元的晚期兴奋(Nambu 等人,2000 b)。此外,还不能排除通过“直接”途径在 IPSPs 后反弹放电的作用,因为纹状体刺激导致切片制备中的苍白球神经元在 IPSPs 后反弹放电(Nambu 和 Llina,1994)。

这些观察结果表明,“皮质-STN-苍白球”通路从运动相关皮质区向苍白球传递强大的兴奋效应,绕过纹状体,传导时间比通过“直接”和“间接”通路传递的效应短(图 2a)。基于这些发现,我们建议将这种“皮质-STN-苍白球”通道称为 BG 的“超直接”通道(Nambu 等人,1996)。

### 2.3.2.3.2.“超直接”通路的功能性意义

最近的解剖学研究表明,与纹状体-苍白球纤维相比,STN-苍白球纤维树状分布更广,终止于更近端的神经元元件(Hazrati 和 Parent,1992a, b)。这表明了 BG 功能的“中心-环绕模型”,提出了对竞争性运动程序的集中选择和抑制(Mink 和 Thach,1993; Mink,1996; Hikosaka 等人,2000)。事实上,分析大脑皮层诱发反应在苍白球中的精确分布表明,具有典型兴奋抑制-兴奋反应的神经元倾向于被仅受兴奋作用的神经元包围,支持苍白球的中心-环绕组织(未发表的观察)。

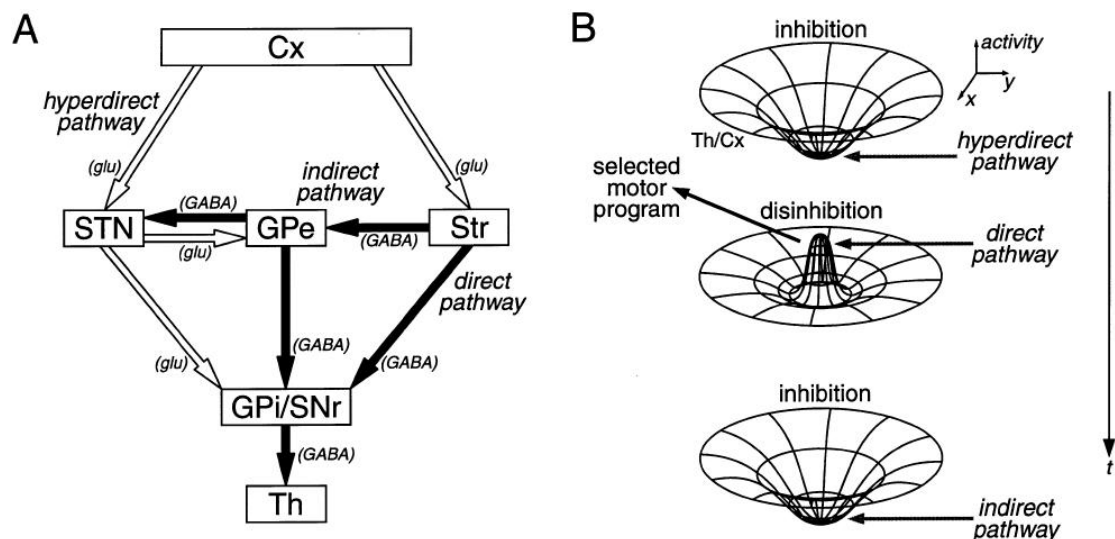


图 9: (A)皮质-STN-GPi/SNr“超直接”通路、皮质-纹状体-GPi/SNr“直接”通路、皮质-纹状体-GPe-STN-GPi/SNr“间接”通路示意图。空心 and 实心箭头分别代表兴奋性谷氨酸能(glu)和抑制性氨基丁酸能(GABA)投射。Cx,大脑皮层;GPe,苍白球的外段;GPi,苍白球内段;SNr,网状黑质;STN,丘脑核;Str,纹状体;Th,丘脑。(B)显示 BG 功能“中心-环绕模型”动态方面的示意图,该模型解释了丘脑和/或皮质(Th/Cx)的活动变化,这些活动变化是由超直接(顶部)、直接(中部)和间接(底部)通路的顺序输入引起的。(由 Nambu 等人修改,2000b)

“皮质-STN-苍白球”超直接”通路对 BG 的输出核产生强烈的兴奋作用，并且大脑皮层的信号传导比“直接”和“间接”通路快(图 2a)。基于这些发现，我们提出了 BG 功能的动态模型，该模型在调节自主性肢体运动的时间域中扩展了“中心环绕模型”(图 2b; Nambu 等人,2000b)。当一个自主性运动即将由皮质机制启动时，一个必然的信号通过“皮质-STN-苍白球”的“超直接”途径同时从运动皮层传输到 GPi，从而导致丘脑和皮质的大片区域受到抑制，这些区域与所选择的运动程序和其他竞争程序都有关(图 2b，顶部)。然后，另一个通过“直接”通路的推论信号被传递到 GPi，以抑制中心区域特定的苍白球神经元群。这样的苍白球神经元解除对目标的抑制，只释放选定的运动程序(图 2b，中间)。最后，可能通过“间接”途径的第三个必然信号(corollary signal)到达 GPi，激活其中的神经元，广泛抑制其目标(图 2b，底部)。通过这种顺序信息处理，只有选定的运动程序在适当的时间启动、执行和终止，而由周围区域的苍白球神经元介导的其他竞争程序则被取消。实际上，阻断 STN 神经元活动(Hamada 和 Hasegawa,1994; Nambu 等人,2000b)或 STN 损伤(Whittier 和 Mettler,1949; Carpenter 等人,1950; Carpenter 和 Carpenter,1951; Hamada 和 DeLong,1992)导致偏瘫。这可以解释为通常被 STN 抑制的电机程序是在随机时间释放的。另一方面，在眼球跳动的调节中，Hikosaka 等人(2000)提出了“间接”途径也可能增强时间对比度的可能性。

应考虑通过“超直接”、“直接”和“间接”路径传输的信号差异。“皮质-STN”神经元和皮质-纹状体神经元属于不同的群体。据报道，“皮质-STN”投射源于锥体束神经元的轴突侧支(Giuffrida 等人,1985)，因此 STN 接收并显示与运动直接相关的活动(Georgopoulos 等人,1983; DeLong 等人,1985; Wichmann 等人,1994)。相比之下，皮质-纹状体神经元传输的信号不同于发送到脊髓-脑干的信号，即皮质-纹状体神经元的活动对行为任务参数的选择性高于皮质-足神经元(Bauswin 等人,1989; Turner 和 DeLong,2000)。纹状体神经元的活动类似于皮质-纹状体神经元(Crutcher 和 DeLong,1984; Crutcher 和 Alexander,1990)。此外，纹状体神经元表现出背景事件依赖性(Kimura,1990; Kimura 等人,1992)或奖赏偶然性(KawaGoe 等人,1998)活动。因此，通过“皮质-STN-苍白球”超直接”途径的信号可以广泛抑制运动程序，然后通过“直接”途径的信号可以根据情况调整选定的运动程序。

### 2.3.3. 反馈路径(Feedback pathways)

任何生物系统都有例外。前馈结构的主要例外是 GPe 的双重作用。传统上，GPe 被认为是间接路径中的中继站，从纹状体接收信息并发送给 STN(图 1a)。然而，新的研究表明，GPe 也接收来自 STN 的输入，并将输出发送到纹状体(图 1b)。纹状体的 parValbumin-阳性 GABA 中间神经元接收来自皮层的强大输入(Lapper 等人,1992; Kita,1993)，也是 GPe 一个重要的反向预测的目标(Bolam 等人,2000)。此外，最近的生理学(Plenz 和 Kitai,1999)和计算学(BeVan 等人,2002; Terman 等人,2002)研究强调了 GPe 和 STN 之间相互联系的作用。总之，GPe 与 BG 的输入核保持高度的双向信息流动(另见第 2.1.3 节)。然而，由于这些预测的生理作用还不清楚，我们将在本综述中进一步讨论它们。

纹状体和中脑多巴胺能神经元之间有一个不同的相互循环。除了纹状体的多巴胺能神经支配(Freund 等人,1984; Smith 等人,1994; Hanley 和 Bolam,1997)(另见第 2.1.6 节)，纹状体大量投射回黑质(Szabo,1980; Parent 等人,1983)。一些研究表明，这些投射主要来自纹状体的纹状小体(Gerfen,1992)。这种相互作用途径是产生增强信号的主要原因(Houk 等人,1995)。

### 2.3.4. 完整循环(Completing the loop)

丘脑长期以来被认为将脊髓和皮质下的信息传递到皮质，形成皮质-BG 回路的最后一个环节(Albin 等人,1989; DeLong,1990)。事实上，BG 功能模型认为丘脑的主要作用是传递 BG

从 GPi/SNr 水平到皮质的信息。主要与此功能相关的丘脑核群是腹前核(ventral anterior,VA)和腹外侧核(ventral lateral,VL)以及丘脑背内侧核(medio-dorsal thalamic nucleus)(Haber 和 Gdowski,2003)。丘脑的组织(网状核、特异性核和非特异性核)及其与皮质的相互联系(Sherman 和 Guillery,2001)超出了本综述的范围。然而,丘脑和皮质之间复杂的相互联系需要一些解释。向丘脑和额叶皮质部位注射双向示踪剂也表明,与前向丘脑-皮质投射相比,皮质-丘脑后向腹前核、腹外侧核和丘脑背内侧核的投射更为广泛。这些发现显示了丘脑-皮质-丘脑相互作用和非相互作用的组成部分(McFarland 和 Haber,2002)。

解剖学和生理学研究表明,丘脑的氨基内核(中央-内侧/束旁核(parafascicular))向纹状体和 STN 有投射(Oreux 等人,2000;Haber 和 Gdowski,2003)。最近对灵长类动物的研究表明,VA 和 VL 核向纹状体的投射非常明显,这表明 BG 环在额叶皮质和纹状体层面都是闭合的(McFarland 和 Haber,2001)。

#### 2.4.沿皮质-纹状体-苍白球通路的神经元数量减少

根据神经元数量,BG 的主轴是从皮质到纹状体,再从纹状体到 GPi 和 SNr 的通路。这条通路的一个显著特征是观察通路每一步神经元数量时观察到的漏斗结构。根据动物和研究方法,这些数字本身差异很大。然而,相同的比例似乎在不同的研究中都存在。就神经元数量而言,相对于纹状体,STN 非常小(另见第 2.1.2 节),因此这里不讨论。

第一次会聚发生在从皮层到纹状体之间。大鼠皮质纹状体神经元数量确定为  $17 \times 10^6$  个,汇聚到  $1.7 \times 10^6$  个纹状体神经元上(Zheng 和 Wilson,2002),其会聚率为 10。第二个更大的神经元数量减少是在纹状体和 GPi 和 SNr 之间,其中因子在  $10^2$ - $10^3$  范围内。对大鼠的研究(Oorschot,1996)显示,因子为  $95[N_{\text{纹状体}}/(N_{\text{GPi}}+N_{\text{SNr}})=2790 \times 10^3/(3.2 \times 10^3+26.3 \times 10^3)]$ 。其他物种的数量需要从对不同核分别进行研究的许多研究中推断出来。重要的是要注意,汇集这些数据可能会导致失真。在猕猴中,数据由 Percheron(1987)等人整合,显示出更高的比率为  $571[12000 \times 10^3/(9 \times 10^3+12 \times 10^3)]$ 。由几个小组进行的人类研究(Thorner 等人,1975;Schroder 等人,1975;Percheron 等人,1987)报告了  $347[11000 \times 10^3/(160 \times 10^3+157 \times 10^3)]$ 。一个悬而未决的问题是接收 GPi/SNr 投射的丘脑神经元的数量,以及与这些丘脑神经元相连的(额叶)皮质神经元的数量。尽管间接证据支持这些阶段神经元数量增加的概念(Arecchi 等人,1997;Francois 等人,2002),因此对于皮质-BG-皮质环的瓶颈结构,我们还没有发现对这些方面的任何定量研究。

#### 2.5.抑制性络脉(Inhibitory collaterals)——解剖学和生理学(anatomy and physiology)

纹状体中多棘神经元的轴突侧支终止于纹状体内,位于两个纹状体上。

中间神经元和其他中等多棘细胞(Bolam 等人,2000)。MSNs 的末端形成对称的突触接触并含有 GABA(Bolam 等人,1985;Kita,1993;Smith 等人,1998b)。这种氨基丁酸能突触被认为是抑制性的,它们与附近的投射神经元的连接理论上表明是侧抑制(Wickens,1993;Wickens 和 Oorschot,2000)。然而,早期的生理学研究未能确定纹状体中这种横向相互作用的特征,并得出结论,它们是弱的或不存在的(Jaeger 等人,1994)。最近的研究表明,在切片和器官型组织培养制备中,使用平均方法(提高 SNr)可以抑制纹状体投射神经元之间的突触电位。(Tunstall 等人,2002;Czubayko 和 Plenz,2002)。在不到三分之一的受试组中发现抑制电位,大多数情况下是单向的,而不是相互的。

苍白球的投射神经元产生厚的、稀疏的棘状、分支不良的树突(Fox 等人,1974;Iwahori 和 Mizuno,1981;Difiglia 等人,1982;Yelnik 等人,1984)。这些树突非常长,有时在其主平面上

形成超过 2 毫米的树突半径，其远端部分通常精心分支以形成复杂的树突末端(Difiglia 等人,1982;Francois 等人,1984;Difiglia 和 Rafols,1988)。这些特殊树突的生理意义尚不清楚，尽管有人认为它们可能有助于相邻苍白球神经元之间的局部突触(Yelnik 等人,1984;Francois 等人,1984;Shink 和 Smith,1995)。许多大的苍白球和 SNr 神经元的另一个特征是，从苍白球树突开始，以不规则间隔出现的线状、珠状、一般不分枝的轴索样突起。它们延伸到中等距离(平均 80 微米)，在某些情况下，似乎与相邻神经元的体细胞或树突接触(Francois 等人,1984;Yelnik 等人,1997)。在多种动物中发现了局部树状化的证据(Kita 和 Kitai,1994;Yelnik 等人,1997;Nambu 和 Llinas,1997;Bevan 等人,1998;Parent 等人,2000;Sato 等人,2000)。Sato 等人(2000)已经表明，三分之一的灵长类苍白球神经元有几个侧枝，在它们的整个树突结构域内，甚至在其之外。与苍白球树突的复杂末端一样(Francois 等人,1984)，这些薄过程在 GPe 内似乎比在 GPi 内更频繁。在 GP 的切片制备中，可以观察到自发的 IPSP(Cooper 和 Stanford,2000)。作者认为这些 IPSP 来源于切片中的活动 GP 细胞。然而，在 40 对记录的配对中，只有 1 对观察到 GP 与 GP 的突触连接，并且它们的自发和谷氨酸诱发的 *spiking* 活性不相关(Stanford,2003)。

对同时记录在行为灵长类动物身上的成对神经元的相互关联研究，可以揭示这些神经元是否接受共同的输入，以及它们是否直接影响彼此的活动(Perkel 等人,1967;Eggermont,1990)。公共输入导致高水平的同步动作电位，导致相互关联功能的双面峰，而直接突触连接导致相互关联功能的单侧峰或低谷。另一方面，扁平的交叉相关图表明神经元之间缺乏直接和间接的相互作用。因此，相互关系研究对于确定 BG 不同可能的功能模式至关重要(Bergman 等人,1998)。实现这一目标的主要方法是使用多个电极记录(Baker 等人,1999)和脉冲分类方法(Lewicki,1998)，这些方法能够区分两个相邻的神经元，它们的电活动被一个电极记录下来。行为猴子纹状体对的记录(Jaeger 等人,1995)和麻醉大鼠纹状体对的记录(Stern 等人,1998)在短时间内没有显著相关性。在正常猴子中，不到 10%的大脑皮层交叉相关图(在不同行为阶段计算)显示出显著的相关性(Nini 等人,1995;Raz 等人,2000;Heimer 等人,2002)。对邻近神经元放电模式的研究表明，它们具有与苍白球内远端神经元相同的特征，即不相关放电(Bar-Gad 等人,2003)。

总之，纹状体存在广泛的侧抑制连接的解剖学证据。在苍白球中也可以发现类似的图像，尽管可能没有那么广泛。然而，细胞内和细胞外的生理学研究显示，这些结构内的功能侧连性较弱且稀疏。这种功能性侧连性的缺乏是我们的 BG 模型的主要推动力，我们将在下面深入讨论。

## 2.6.BG 连接性

核之间的识别路径只提供了理解 BG 内复杂连接所需的部分信息。理解核之间的连接性模式需要更全面的描述，包括许多方面。首先是核内发生的收敛量(*amount of convergence*)(Bergman 等人,1998);即，所有信息是合并在一起的，还是存在分离部分信息的平行路径？第二个方面是组织的稀疏性(*sparseness*)，即每个神经元接收到的总输入量。这两个方面实际上交织在一起，形成了许多其他的连通性特征，例如组织的局部性质、区域之间的差异和共同输入的水平。稀疏度和组织之间关系的计算方面将在后面讨论(第 7.5 节)。

### 2.6.1.平行连接与漏斗状连接(Parallel versus funnel-like connectivity)

关于 BG 的解剖学或信息共享的数量，有两种极端的观点(Perceron 和 Fillion,1991)。第一种观点认为 BG 输出阶段的神经元接收高度收敛的输入，导致信息漏斗化(Bolam 等人,1993;Percheron 等人,1994)，而第二种观点则将 BG 描述为分离的平行电路，平行通路之

间的相互作用最小 (Alexander 等人,1986;Alexander 和 Crutcher,1990;Parent 和 Hazrati,1993;Middleton 和 Strick,2000)。

尽管皮质纹状体投射的范围很大,但主要终末区位于其皮质输入最邻近的位置(Kemp 和 Powell,1970;Selemon 和 Goldman Rakic,1985)。一般来说,皮质纹状体投射以功能性局部方式终止(Takada 等人,1998;Tokuno 等人,1999),大致可分为边缘、认知和运动区域(Parent 和 Hazrati,1995)。因此,皮质纹状体投射沿着纹状体的腹轴到背轴终止。背侧纹状体(壳核)从感觉运动区接收皮质输入,中央纹状体(尾状核)从相关皮质区接收输入,腹侧内侧纹状体从边缘区接收输入。例如,在纹状体的头端(运动的高阶认知控制)和尾端(运动执行)扣带运动的皮质输入之间没有明显的重叠(Takada 等人,2001)。猴子纹状体苍白球投射的顺行标记(Anterograde labeling)也支持分离 BG 通路的一般概念。纹状体-苍白球纤维在平行于苍白球外边缘的多个细长带中排列(Parent,1990)。起源于尾状核(相关纹状体区)或连合后壳核(post-commissural putamen)(感觉运动区)的神经元的纹状体-苍白球轴突在 GPe 和 GPi 水平上彼此保持很好的分离(Parent,1990)。此外,顺行双标记(anterograde double labeling)研究显示,从邻近的纹状体神经元群中出现的纹状体轴突在苍白球层分离(Hazrati 和 Parent,1992)。纹状体-黑质投射被组织成多个分散但分离的神经丛,如纹状体-苍白球投射(Hazrati 和 Parent,1992)。跨神经元病毒转运的解剖学研究也支持 BG 平行的宏观组织的概念。因此,不同的额叶皮质区域最密集地投射到不同的纹状体区域(输入通道),这些皮质区域中的每一个都受到不同的 GPi 区域(输出通道)投射的影响(Hoover 和 Strick,1993;Middleton 和 Strick,2002)。

早期的电生理研究也支持(甚至是一个主要的推动力)BG 分离平行组织(segregated parallel organization)的概念(Alexander 等人,1986)。纹状体(Crutcher 和 DeLong,1984a;Alexander 和 DeLong,1985)和苍白球(Delong 等人,1985)的单个单位和微刺激研究揭示了这些结构的躯体定位组织(somato-topic organization)。此外,缺乏与手臂和腿部运动有明确关系的苍白球单位(Delong 等人,1985;Alexander 和 Crutcher,1990)表明分离(segregation)甚至可能适用于特定的身体部位。MI 和 SMA 的电生理刺激研究(Nambu 等人,2002a)表明,大多数纹状体神经元只对 MI 刺激或 SMA 刺激作出反应,而 MI 和 SMA 反应神经元主要分布在壳核的不同部分。类似的研究也显示了接收不同皮质骨输入的苍白球神经元的分离(Yoshida 等人,1993)。

另一方面,“漏斗”模型的支持者强调这样一个事实:在所有总量示踪剂和病毒研究中,不同区域之间的边界并不清晰,纹状体(Kunzle,1975;Yeterian 和 VanHoesen,1978;Selemon 和 Goldman Rakic,1985;Ramanathan 等人,2002)、GPi(Flaherty 和 Graybiel,1993;Bolam 等人,1993)和 STN(Kolomets 等人,2001)中存在广泛的重叠和交叉。此外,与传入的纹状体轴突成直角的苍白球神经元的宽树突树状排列(Percheron 等人,1984),纹状体甚至苍白球的侧支系统(Parent 等人,2000)对 GP 的发散性 STN 投射,以及最后的纹状体到苍白球的神经元数量减少,表明苍白球层次产生广泛会聚或漏斗状突起。对生理学研究结果的仔细分析表明,即使在上述研究中,图像也不完全是黑色或白色的。上面引用的微刺激研究(Nambu 等人,2002a)报告,大约 20%的记录纹状体神经元同时对心肌梗死和心肌梗死的刺激做出反应。同样,大量的苍白球神经元被不止一个皮质区域(前额叶、运动前区、辅助运动和弓形运动前区以及运动皮质)的刺激所抑制(Yoshida 等人,1993)。通过纹状体的原位刺激,也证明了从远侧纹状体神经元到苍白球中一个焦点区域的输入通道(Kimura 等人,1996)。我们最近对参与复杂行为任务的猴子的研究也揭示了对多种行为事件有反应的神经元的显著比例,包括经典上属于边缘区,认知区和运动区。

## 2.6.2. 稀疏连接(Sparse connectivity)

沿着 BG 主轴的核之间的连接非常稀疏。在灵长类动物  $10^8$  个皮质纹状体神经元中(基于皮质纹状体和纹状体神经元之间的比例为 10(Zheng 和 Wilson,2002)和灵长类动物  $10^7$  个纹状体神经元之间的比例(Percceron 等人,1987)), 单个纹状体神经元接受大约 104 个皮质突触(Wilson 等人,1983;Ingham 等人,1993)。因此, 假设每个突触来源于不同的皮质神经元, 单个纹状体神经元最多从支配纹状体的大约 0.01%皮质神经元接收输入。同样的道理也适用于由大约  $10^7$  个纹状体神经元支配的 GPi(Percceron 等人,1994), 而每个苍白球细胞只有约  $10^4$ - $10^5$  个源自纹状体的突触(Percceron 等人,1994)。因此, 典型的苍白球神经元最多接收 0.1%到 1%的纹状体投射神经元的输入。同一个核内的神经元通过氨基丁酸能络脉的连接也在其范围内受到限制。纹状体神经元约有  $10^3$  个侧支突触(Kawaguchi 等人,1995), 导致与另一个纹状体投射神经元的突触的最大概率为 0.01%。苍白球神经元形成的侧支突触数量较少, 可能约为  $10^2$  个突触(Parent 等人,2000), 核内的连接率为 0.1%。重要的是要记住, 即使核之间和核内的这些低连通性因素也代表着最高的概率估计。神经元对之间的多个突触的出现将进一步降低与其他神经元连接的可能性(Kincaid 等人,1998;Zheng 和 Wilson,2002)。

然而, 不同核的神经元之间以及核内的连接并不均匀。特定的神经元对往往会增加突触的概率, 这取决于分域和其他组织特性。因此, 这些神经元倾向于“附近”, 无论是功能上还是空间上(Wilson,2000)。皮质纹状体轴突倾向于在有限的区域内形成一组突触(Kincaid 等人,1998), 突触之间的距离服从指数分布(Kincaid 等人,1998)。此外, 纹状体的(至少部分)局部定位来源于参与相同功能的神经元趋向于会聚到相同区域的事实(Yeterian 和 Vanhoesen,1978;Selemon 和 Goldman Rakic,1985;Flaherty 和 Graybiel,1991)。因此, 纹状体神经元在同一分域(topographic)间隔内接收来自同一神经元或皮层内相邻神经元的输入的概率显著升高。苍白球也显示出一种局部定位关系(topographic organization)(DeLong 等人,1985), 纹状体轴突的树化倾向于遵循特定模式(Hazrati 和 Parent,1992)。纹状体 MSNs 的树突和局部轴突树呈球形(Wilson,2000), 限制其直径为  $400\mu\text{m}$  的连接范围(Wilson 和 Groves,1980), 从而导致仅限于相邻神经元的横向连接。此外, MSNs 的树突不会穿过斑块/基质室之间的边界(Malach 和 Graybiel,1986)。在苍白球中, 络脉稀疏, 树突树的大小约为  $1000\mu\text{m}$ 。然而, 与纹状体一样, 络脉在空间上有组织, 支配着附近的空间神经元(Kita 和 Kitai,1994;Yelnik 等人,1997;Nambu 和 Lynas,1997;Bevan 等人,1998;Parent 等人 2000;Sato 等人,2000)。耶尔尼克(J. Yelnik, 私人通讯)的一项有趣的观察发现:“与多棘的纹状体神经元相比, 苍白球神经元的数量非常少, 一个相当差的局部树状化就足以将两种结构中相同比例的相邻神经元连接起来。”

总的来说, 纹状体似乎被划分为多个区域(Wilson,2000), 这些区域之间的连通性更高, 区域之间的连通性为零或非常低, 这可能是由神经元间质介导的(Kawaguchi 等人,1995)。苍白球也有一个局部分区(topographic division), 但是它不太清晰(Hamada 等人,1990)。因此, 附近的纹状体和苍白球神经元倾向于共享在其输入结构中代表相同功能的输入。

## 2.7. BG 强化信号(Reinforcement signals in the basal ganglia)

适应性行为要求评估环境刺激对其行为的影响(Robbins 和 Everitt,1996)。作用于纹状体的多巴胺和乙酰胆碱神经元等神经调节剂发出的信号可以进行这种评估。这些物质影响皮质纹状体传导的可塑性(Calabresi 等人,2000;Reynolds 等人,2001), 因此是 BG 系统教师的良好人选。

多巴胺功能的经典理论是 Wise(1982)提出的“RL”。这一假设表明, 多巴胺代表了负责行



为控制的大脑系统的天然增强剂的“享乐品质”。这些理论的强大之处在于，成瘾药物主要通过影响多巴胺能系统来发挥作用。大脑刺激奖励(BSR)实验也支持这一观点。在 BSR 实验中，动物选择接受多巴胺相关回路周围的电刺激，动物甚至对这种刺激比食物或水等自然强化物更喜欢(Daw,2003)。

中脑多巴胺神经元对自然奖励和新奇或意想不到的刺激有反应(Schultz,1998)。然而，多巴胺能神经元对自然奖励的反应只是对其功能的一个近似描述。Schultz 等人(1997)的实验表明多巴胺反应与奖赏本身无关，而是与对未来奖赏的预测有关。多巴胺神经元被比预期更好的奖励事件激活，不受预期更好的事件影响，被比预期更糟的事件抑制。因此，更准确地说，多巴胺信号似乎代表了预测的未来奖励和实际奖励之间的误差(Schultz 等人,1997)。

中脑多巴胺信号代表预测未来奖励和实际奖励之间的误差，这一发现标志着我们对 BG 的理解取得了重要进展。有趣的是，还观察到一个复杂的量，比如预测未来奖励和实际奖励之间的误差，与 RL 术语中称为时序差异(TD)误差的类似量非常相似(Sutton 和 Barto,1998)。这提出了一个实现 RL 机制的可能的场景 BG，这需要 SNc 的多巴胺信号来处理它，并向感觉运动皮层通路发出适当的调节信号，形成 S-R 学习的神经底物(图 1)。

### 2.7.1.多巴胺能信号(dopaminergic signal)

BG 被认为参与了记忆的 S-R 形式的形成，这种记忆形式被称为“操作条件反射(Operant conditioning)”(Squire 和 Zola,1996)。在操作条件反射中，有益的反应得到加强，而无益的反应减弱。操作条件作用也与 RL 密切相关，RL 研究的是个体如何选择自己的行为，从而最大化回报(或最小化错误)(Sutton 和 Barto,1998)。中脑多巴胺神经元在获得新行为中起着关键作用。虽然早期的研究强调多巴胺在奖赏相关(享乐)行为中的作用(Fibiger 和 Phillips,1988;Wise 和 Rompre,1989)，但 Schultz 及其同事的最新开创性研究表明，多巴胺神经元的主要功能是报告动物的预测与现实之间的不匹配(Hollerman 和 Schultz,1998;Schultz,1998;Waelti 等人,2001;Fiorillo 等人,2003)(但见(Redgrave 等人,1999;Horvitz,2000))。纹状体对多巴胺的反应是异质的，可能由多个变量决定，包括多个受体之间的复杂相互作用、纹状体神经元的状态(向上或向下)和多巴胺能神经元的放电模式(强直与突发放电)(Di-Chiara 等人,1994;Arbuthnott 和 Wickens,1996;Cepeda 和 Levine,1998;Wickens 和 Oorschot,2000;Calabresi 等人,2000)。在本综述中，我们将强调最近的研究表明多巴胺在皮质纹状体通路中控制 LTP/LTD(Wickens 等人,1996;Centonze 等人,1999;Kerr 和 Wickens,2001)。

如果我们将 RL 与另一种学习形式(如监督学习)进行比较，就会清楚地认识到将探索作为实现 RL 的过程的必要性。在监督学习中，对刺激物的期望反应对智能体是显式可用的；将智能体的实际响应与期望的响应进行比较，并使用该误差纠正智能体的 S-R 映射。在 RL 的情况下，没有这样明确的期望反应(老师)。智能体必须探索对刺激可能作出反应的空间；从环境中获得奖励的反应得到选择性的加强。在 RL 框架中，这种探索是由一个称为“Explorer”的模块执行的，但是，它可以根据正在面临的问题实现多种实现。

如果不是为了探索“反应或行动”的空间，使用奖励信息学习 S-R 映射本来是很容易的。但是，在自然环境中，奖励通常是在动物完成一系列动作后才发放的，而这些动作对奖励发放的贡献很难确定。根据过去行为与未来奖励的相关性来评估过去行为的需要被称为“时间信用分配问题”。在 RL 框架中，“智能体”通过开发一个称为 Critic 的内部评估模型来解决这个问题，该模型估计了“智能体”当前状态下的总预期未来报酬，即“值函数” $V(t)$ 。

Critic 的评价和 Explorer 的干扰被用来训练“S-R 映射”，这是一个用 RL 术语称为 Actor

的模块。因此, Actor、Critic 和 Explorer 是 RL 机器人的关键执行组件。

建模 BG 的下一步自然是寻找 RL 的 Critic、Explorer、Actor 和其他关键组件的神经类似物, 从而使我们对 BG 的理解更加完整。许多建模工作已经被用于使用 RL 概念描述 BG 功能(Montague 等人,1996; Schultz 等人,1997; Houk 等人,1995; Suri 和 Schultz,1999; Daw 和 Touretzky,2000; Frank,2005)。我们现在描述一些从这一研究中产生的关键思想。

### 2.7.2.胆碱能信号(cholinergic signal)

在醒着的、行为正常的猴子纹状体中, 细胞外的生理记录显示, 自发的 TAN 放电会暂停, 常常伴随着兴奋和奖励(但也令人反感)的事件(Kimura 等人,1984;Crutcher 和 DeLong,1984b;Liles,1985;Apicella 等人,1991;Ravel 等人,2003)。因此, TAN 被认为参与了对具有内在激励意义的刺激的检测(Kimura 等人,1984;Aosaki 等人,1994;Apicella 等人,1997;Ravel 等人,2001)。我们最近对执行概率工具性条件反射(instrumental conditioning)任务的猴子的多巴胺能和 TAN 神经元群体的研究表明, 尽管多巴胺神经元的反应反映了预期和结果之间的不匹配, 但 TAN 的反应与事件的可预测性或奖励价值无关。TAN 的尖峰活动具有高度的时间相关性(Raz 等人,1996;Kimura 等人,2003)。SNc 和 TAN 多巴胺能神经元的不同反应和相关模式表明, 这两个系统并不相互镜像。这些数据符合这样一个作用假设, 即胆碱能信号通知 BG 何时学习, 多巴胺能信号告诉他们如何学习, 以及先前皮质纹状体活动的性质决定了将要学习的内容(第 4 节)。因为乙酰胆碱通过固定纹状体投射神经元的上/下状态来降低它们对皮质输入的反应(Akins 等人,1990), 因此, 对行为显著事件作出反应时, 纹状体神经元的停顿可能有助于确保纹状体神经元及时“获得信息”, 使多巴胺能增强信号加强共同激活的皮质和纹状体神经元之间的突触作用。

### 2.7.3.其他强化信号

纹状体的另一个主要脑干输入是血清素能的(serotonergic), 来自中缝背核(dorsal raphe nuclei)和中缝核(medial raphe nuclei)(Lavoie 和 Parent,1990)。与纹状体中多巴胺能末端的分布不同, 血清素能末端在纹状体中分布不均匀。相反, 它们在腹侧纹状体中更丰富(Lavoie 和 Parent,1990)。关于这些神经调节系统在 BG 功能中的作用的生理学研究很少(Sawyer 等人,1985), 在本综述中我们将不再关注它们。

## 2.8.BG 的病理生理学(Pathophysiology of the basal ganglia)

19 世纪初, 人们首次假设黑质与帕金森病(PD)之间的关系(Brissaud,1895)。这一点后来被证明是在帕金森病人的黑质中进行性细胞丢失证实的(Hassler,1939)。神经递质多巴胺和黑质之间的联系是由于发现 PD 患者多巴胺被耗尽(Ehringer 和 Hornykiewicz,1960)。80 年代早期, 灵长类 1-甲基-4-苯基-1,2,3,6 四氢吡啶(MPTP)帕金森病模型的发展(Burns 等人,1983;Wilms 等人,1999)使我们对帕金森病的病理生理学的理解向前迈进了一大步。MPTP 模型是 BG 疾病的关键模型。此外, 其他动物模型(如 6-OHDA 啮齿动物模型)和人类研究中也包含了这些见解。

几项灵长类动物的研究表明, MPTP 对不同 BG 核的放电率有不同的影响。GPi 和 STN 神经元的平均放电率分别从 ~80 到 ~100 峰/秒(Miller 和 DeLong,1987;Filion 和 Tremblay,1991;Boraud 等人,1998)和从 ~20 到 ~25 峰/秒(Miller 和 DeLong,1987;Bergman 等人,1994)。GPe 神经元的比率从 ~70 降至 ~50 峰/秒(Miller 和 DeLong,1987;Filion 和 Tremblay,1991;Boraud 等人,1998)(但不变的放电率见(Boraud 等人,2002))。在人类 PD 患者中确定了 MPTP 猴子 BG 的异常强直放电率(Merello 等人,1999;Levy 等人,2001), 并通过消融

或灭活 GPi 或 STN 来缓解所有帕金森病症状(Bergman 等人,1990;Benabid 等人,1994)。

此外,在 MPTP 治疗后,灵长类 BG 回路中的放电模式发生了显著改变。在不规则或振荡爆发中放电的神经元百分比有所增加(Miller 和 DeLong,1987;Filion 和 Tremblay,1991;Bergman 等人,1994;Boraud 等人,1998)。在人类 PD 患者的研究报告了细胞的放电在 GPi 和 STN 的震颤频率范围内被调节(Levy 等人,2002)。然而,震颤与神经元振荡之间的相关性充其量是间歇性的,本质上是动态的(Hurtado 等人,1999;Lemstra 等人,1999;Raz 等人,2000)。

在行为正常的动物中,苍白球神经元之间的大多数交叉相关图显示,神经元对在功能上是独立的(Nini 等人,1995;Raz 等人,2000)。然而,对经 MPTP 处理的猴子苍白球的生理学研究表明,交叉相关图会达到峰值并出现振荡(Raz 等人,2001;Heimer 等人,2002)。在主运动皮层(Goldberg 等人,2002)、纹状体 TAN 之间以及 TAN 和苍白球神经元之间,也发现了类似的现象,即在接受 MPTP 治疗后,灵长类动物大脑的同步程度有所提高(Raz 等人,1996、2001)。综上所述,对 MPTP 治疗的灵长类和 PD 患者的 BG 的生理学研究显示,这些结构的放电率、模式和同步性发生了显著变化。通过几种 BG 靶点的失活或脑深部刺激改善大多数帕金森病症状,进一步表明 BG 活动的这些或其他变化(纹状体多巴胺耗尽后)是这些症状的主要原因。

### 三、BG 模型的演变与革命(evolution and revolutions)

虽然早期的研究涉及到 BG 在运动功能中的作用,但后来的工作也认识到它对认知、学习和记忆相关现象的重要性。此外,BG 的损伤与复杂的神经精神认知和行为障碍有关,反映了这些核在额叶不同功能中的广泛作用(Stout 和 Johnson,2005)。关于 BG 的建模文献的一个主要部分涉及 BG 在运动功能和运动障碍中的作用(Chesselet 和 Delfs,2003)。虽然在解剖学、病理学、电生理学和与 BG 相关的影像学研究方面已经取得了巨大的进展,但对这些核在行为控制方面的作用的全面理解仍然是难以实现的。关于 BG 的计算建模文献确实很多,但大多数模型往往只关注 BG 的一个或两个功能。需要一个可以将 BG 的所有功能集成到一个框架中的统一功能模型。从 BG 计算模型综述(Prescott 等人,2002; Houk 等人,1995)可知,似乎大多数 BG 模型都是排他性的,并只实现特定的功能角色。

在过去的 20 年里,健康和疾病中的 BG 的信息处理模型(Beiser 等人,1997;Gillies 和 Arbuthnott,2000)经历了革命性和进化性的变化。这些核的早期模型涉及主要途径,即从皮层依次通过纹状体到苍白球的途径。这些单通路模型主要涉及该环路的功能组织以及沿通路的信息并行处理(Alexander 等人,1986)与收敛(Perceron 等人,1984)的问题。将 BG 内的其他已知通路并入模型,将焦点转移到这些多条通路之间的微妙相互作用上。使用“盒箭”模型对通路及其相互作用的描述导致了对 BG 在健康和疾病中的作用的多方面深入研究(Albin 等人,1989;DeLong,1990)。路径(“盒子”)中元素的数量和复杂性及其相互作用(“箭头”)逐渐包含新的解剖和生理数据(Wichmann 和 DeLong,1996)。“盒箭”模型的这种演变反过来导致了盒子的打开,描述了内部核结构,并在空间和时间域中模拟了核内的相互作用(Hikosaka 等人,1993)。这一代模型使用更加复杂的神经网络,着重于 BG 的动作选择特性,强调了核的计算特性(Wickens,1997)。例如,有一些模型描述了 BG 在“动作门控”中的作用(Contreras-Vidal 和 Stelmach,1995);竞争行动之间的“动作选择”(Redgrave 等人,1999b; ,1995);维持“工作记忆表征”(Houk 等人,1995);“顺序学习”(Bapi 和 Doya,1998; Berns 和 Sejnowski,1998),最重要的是 RL(Montague 等人,1995; Schultz,1998; Barto,1995)。Buhusi 和 Meck(2005)也回顾了 BG 在“计时”方面的作用。

BG 建模者的一个巨大的挑战,即未来要构造由以上提到的模型改编而成的具有独特的、有用的洞察力的一个集成框架,构建一个统一的 BG 功能构想,并显示 BG 核运用奖励信息引起的多巴胺信号和用它来进行多方面的操作。最近的研究扩展了 BG 模型,将信息流包含在从皮层到 BG 再回到皮层的信息循环的更大框架内,从而在顺序处理中形成部分闭合的循环功能(Berns 和 Sejnowski,1998)。

BG 模型的一个重要类别被称为 Actor-Critic 模型或基于 RL 的模型。上文提到,中脑多巴胺能神经元的活动代表了预测奖励和实际奖励之间的差异,这个量在 RL 术语中被称为 TD 误差(Schultz,1998)。Houk、Adams 和 Barto(参见 Houk 等人,1995 年的著作)可能是最早提出类似于 Actor-Critic 学习系统的一些作者。Montague 及其同事(Montague 等人,1996; Schultz 等人,1997)提出了一个计算 RL 模型来解释在(Schultz,1998)中的实验研究中观察到的多巴胺能神经活动模式。一些作者认为纹状体在 BG 中起着 Critic 的作用(Houk 等人,1995; Suri 和 Schultz,1999; Contreras-Vidal 和 Schultz,1999)。纹状体接受大量来自感觉运动皮层区域的输入。特别是,每个运动皮质区域都投射到一个由尾状核和壳核构成的结构上截然不同的区域。因为纹状体也收到从 SNc 的多巴胺能预测,它实现的计算是十分恰当的 *值函数*:正如 TD 误差( $\delta$ )是用于更新 *值函数* 的权重,纹状体多巴胺预测调节“皮质-纹状体”连接的可塑性。

虽然 RL 结构成分在 BG 解剖结构上的映射研究已经取得了一定的成果,但在确定一个重要成分的皮层下底物上仍有一定的困难。虽然在 RL 框架中,“运用”和“探索”是同等重要的过程,但基于 RL 的 BG 观似乎主要集中在 *奖赏信号*——它的化学信使,它的解剖位置,它在学习中的结果等等,而只对探索进行了总结。Daw 等人(2006)的研究表明,前额叶和顶叶位置与探索相对应,但没有发现相应的皮质下位置。因此,尽管有大量的文献应用于 BG 功能的 RL 概念,但没有共识,如果,RL 机制可以被映射——组件到组件——BG 解剖,并且映射关系如何解释 BG 功能的全部功能(Joel 等人,2002; WorGotter 和 Porr,2005)。

### 3.1.运动准备

Kornhuber 和 Deecke(1990)在他们关于自主运动的经典脑电图研究中发现,皮质电位在运动开始前就存在缓慢的负迁移。这种电位被称为“准备电位”(Bereitschafts potential, BP),被认为是运动皮层区域运动前准备的标志。仔细分析 BP 的电流偶极子源,可以发现 SMA 是一个关键因素(Lang 等人,1991)。然而,与运动方向相对应的准备活动在许多其他脑区也被发现,包括 M1 (GeorGopoulos 和 Grillner,1989)、运动前皮质(Kubota 和 Hamada,1978)、PFC (Kubota 和 Funahashi,1982)、顶叶皮质(Crammond 和 Kalaska,1989)和 BG (Alexander,1987)。Churchland 等人(2006)在灵长类动物实验中提出了一个有趣的运动准备功能定义。这个小组假设,准备是一个过程,通过这个过程,运动皮层神经元的活动,随机的和可变的在准备的早期阶段,是逐步推进到一个有限的区域的状态空间,这是特定于给定的运动。来自灵长类动物运动前皮层神经元的活动似乎证实了他们的假设 Churchland 等人(2006)。然而,与其他运动皮质区相比,SMA 在运动准备中起着主导作用。

SMA 作为运动准备信号的主要来源,似乎与其他几个相关运动领域竞争。灵长类动物的单细胞记录显示,与 M1 神经元相比,神经神经元的准备相关的变化更为明显(Tanji,1994)。问题是 SMA 和 M1 哪个先出现?如果能证明 SMA 神经元的准备活动先于 M1 神经元的类似活动,就可以解决这个问题。已有研究表明,具有预备活动的 SMA 神经元可以投射到 M1 (Tanji,1994)。相反,研究还发现,表现出准备活动的 M1 神经元接受来自 SMA 的输入,而不是来自丘脑或顶叶皮质(Aizawa 和 Tanji,1994)。这些研究有力地说明 SMA 在运动准备中的作用。然而,SMA 可能并不完全负责运动准备工作。它的准备作用可能包括皮层下结构之间的相互作用,如 BG,这往往涉及运动定时功能:SMA 和 BG 之间的相互作用被认为

在学习的运动序列中起着至关重要的作用(Cunnington 等人,1995; Bapi 等人,2006)。有人认为 *BG* 的阶段性活动可以作为 *SMA* 的复位信号,清除一次运动后的痕迹,为连续运动做好准备(Georgiou 等人,1993)。

从上面的叙述中,我们了解到运动准备是受 *BG* 和 *SMA* 双向作用的影响。类似的观点也出现在“笔迹生成”的计算神经运动模型中(Gangadhar 等人,2007)。该模型的灵感来自于 Hollerbach(1981)的“手写的振子理论”,根据该理论,“手写”可以看作是两个振荡过程的叠加运动——一个是垂直的,另一个是水平的。笔迹的形状可以通过调制各种参数的振荡,如相位,孤立时刻的振幅。Gangadhar 等人(2007)描述了一个神经振荡器网络,该网络可以根据网络中神经元的振荡活动,以傅里叶式的方式解析笔画速度,从而学习任意的笔画。准确的行程执行不仅取决于网络的“权重”,而且关键取决于振荡器的初始相位。模型中的“*BG*”模块向振荡神经元发送预备信号,并通过将它们从随机的初始阶段驱动到“标准阶段”来初始化它们的相位。这种运动准备的观点与 Churchland 等人(2006)的观点强烈呼应,Churchland 等人观察到背侧运动前神经元(PMd)的活动模式是从随机初始运动开始运动的状态到“最优子空间”,该子空间指的是一组受限制的状态(图 3)。

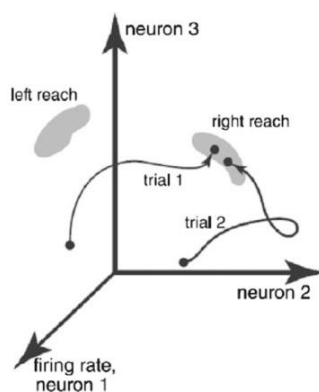


图 10: 显示了运动前活动如何随着时间的推移而稳定下来。活动在高维空间中表示,每个神经元的放电率提供一个轴。活动最初是可变的,接近基线(每次试验一个点)。在目标开始后,活动将(曲线路径)确定为一个子空间状态,该子空间适合所需的运动(阴影区域)(摘自 Churchland 等人,2006 年的研究)。

作为(Gangadhar 等人,2008)中提出的运动准备研究的延伸,我们的小组最近研究了帕金森条件下的准备(未发表的结果)。在模型中,我们发现从运动开始到运动结束的驱动力与阶段的多巴胺爆发有关,而阶段的多巴胺爆发已知发生在运动开始或运动之前。由于多巴胺缺乏症是帕金森病(PD)的一个特征,我们在模型中通过减弱准备神经振荡器的驱动信号来模拟 PD 的情况。驱动信号(缩放因子称为驱动因子,  $\gamma_{df}$ )被削弱,准备花费的时间更长;  $\gamma_{df}$  低于临界值时,准备永远不会终止,动作永远不会启动。这种不能使运动皮层为运动做好准备的情况类似于 PD 患者在自主控制时所面临的困难。特别是对 PD 的冻结现象(Giladi 等人,1992)提供了一个有趣的解释,我们认为这可能是一个失败的运动准备的例子。

### 3.2 动作选择(action selection)

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。*BG* 解决竞争行动之间的联系的建议是基于一个日益增长的共识,即这些结构的一个关键功能是在竞争进入最终共同运动路径的传感器运动系统之间进行仲裁。从这个想法发展而

来的一个计算假设依赖于这样一个前提:传入纹状体的信号编码了进入运动系统的请求的显著性(Redgrave 等人,1999b)。BG 中嵌入的多重选择机制可以解决竞争对手之间的冲突,并在赢家之间提供干净、快速的切换。首先,纹状体神经元的上/下状态可以作为第一遍过滤器来排除弱支持请求。其次,纹状体内的局部抑制可以选择性地增强最显著通道的活性。第三,纹状体的聚焦抑制与 STN 的扩散(发散)激发相结合,可以作为一个前馈的、离中心的/环绕的网络,作为一个整体贯穿 BG(参见 Mink,1996)。最后,输出核内的局部相互抑制可以提高最终的选择。一些研究表明 BG 在动作选择中起着关键作用(Seo、Lee 和 Averbeck,2012)。

BG 的主要输出核(GPi, 黑质网状部分[SNr])表现出对其靶点产生抑制作用的强直活性。由于沿着直接路径的活动会导致这些目标的去抑制,因此有人提出,这允许选择特定的运动程序(Chevalier 和 Deniau,1990; Mink 和 Thach,1993)。聚焦去抑制周围的抑制目标将允许这一过程被微调。在该理论模型中,通过向输出核提供广泛激发的间接途径,通过 STN 可以维持或增强强直性抑制(Mink 和 Thach,1993)。这样一个优雅的“BG 处理”的中心环绕模型为 BG 在动作或运动程序选择中的作用提供了动力。Redgrave 等人(Redgrave, Prescott 和 Gurney,1999b; Gurney, Prescott 和 Redgrave,1999; Redgrave, Prescott 和 Gurney,1999a)对这一观点进行了扩展,以便将 BG 视为一般的资源选择机制。

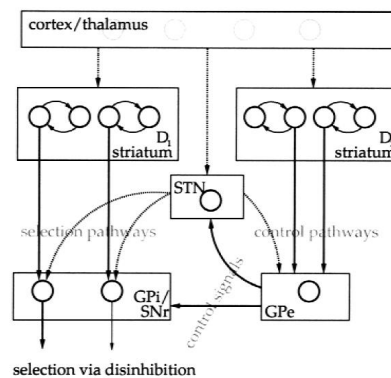


图 11: Gurney 等人模型中连接的示意图。36 条实线表示抑制连接;虚线表示兴奋性连接。

从统筹效率的角度提出了这种集中资源选择系统的观点。中央系统省去了在每个竞争资源之间进行组合大规模连接的需要。它只需要每个资源区域连接到中心机制。从广泛的皮质区域到纹状体的投射当然鼓励了这种观点。图 5 说明了 Gurney 等人(Gurney, Prescott 和 Redgrave,1999)在 BG 电路背景下提出的中心选择模型。纹状体的皮质输入是用“通道”来描述的,其中每个通道指的是竞争的资源或行为,并来源于负责处理它的(皮质)神经基质。假设通道携带显著信息,从而在纹状体水平上,中等多棘神经元之间的竞争由显著性决定。模型分为选择和控制机制。与 STN 信号合作的直接途径介导选择。模型纹状体内的局部竞争解决了局部资源冲突,使单个通道处于活动状态。例如,在细胞体主题表示下,这可能解决公共运动资源的冲突。局部获胜活动通道的活动代表了该通道的显著性。因此,纹状体的产出可以用于全局资源竞争。纹状体到输出核(GPi, SNr)的直接投射抑制了这些核中与竞争通道显著性成比例的通道区。STN 为输出核上的通道区提供扩散均匀的激发。这种来自 STN 的弥漫性兴奋与来自 GPe 的抑制相结合,扩大了输出核的活性,因此,只有具有最高显著性的通道仍然保持足够的抑制,以解除对目标的抑制,从而进行选择。在控制通路中,从 GPe 到 STN 的反馈可以衡量 STN 的活性,使其独立于竞争通道的数量。

“盒箭”模型的缺点促使从系统级模型转向更详细的网络级模型,目的是了解 BG 的正常功能和病理状态。目前, BG 的主要网络层模型是动作选择模型。他们的共同重点是在皮质



-BG- 皮质环内的 BG 的作用，在皮质向 BG 呈现的众多动作中选择一个或多个动作 (Mink,1996)。这些模型在所选动作的性质上有所不同，其定义范围从简单的水平运动动作到高水平的复杂行为方案。根据*选择机制*是核内选择还是核间选择，模型还分为两大类。核内选择是通过核内的横向(或反复)抑制来实现的，而核间选择则是通过不同途径之间的*前馈*竞争来实现的。一些模型集成了两种*选择机制*，以实现增强的选择过程。

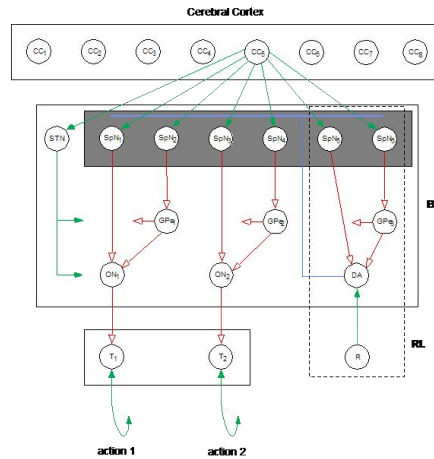


图 12: 一个连接 BG 和大脑皮层的宏观信号处理模块。大脑皮质将不同的兴奋性投射发送到中等多刺神经元的网络(阴影框)和 STN。BG 的输出神经元是苍白球(GPI)和黑质网状(snr)中的细胞。他们投射到的相互兴奋的丘脑皮质环显示为绿色。红色开放箭头表示抑制连接,紫色表示位于黑质致密部(SNC)和腹侧被盖区(VTA)的多巴胺(DA)神经元的神经调节连接。标有 RL 的虚线框表示强化学习电路。

Albin-DeLong 模型(Albin, Young 和 Penney,1989; DeLong,1990)是最早的 BG 现代模型之一,该模型建立的目的是研究帕金森病的运动功能减退的临床表现。根据这个模型, BG 输出神经元的过度活动,由于直接通路上的多棘神经元的活动不足,抑制了丘脑皮质的活动,导致了失运动症。正常受试者的推导模型是,直接通路中的多棘突神经元(bursts of spiny neuron)活动突然抑制其丘脑皮质靶点以启动所需动作,而间接通路中的多棘突神经元放电突然抑制其丘脑皮质靶点以抑制不需要的动作(图 2)。该模型已应用于肢体运动(Mink ,1996)和眼部运动(Hikosaka、Matsumura、Kojima 和 Gardiner ,1993),尽管眼部运动回路略有不同,因为它们由上丘(the superior colliculus)皮质下结构(subcortical structure)控制, (McHaffie、Stanford、Stein、Coizet 和 Redgrave, 2005)。临床模型倾向于强调图 1 中控制行为的 M1 循环,但这已扩展到控制思维的认知循环(Graybiel,1997)。多巴胺的作用包括在模型中,通过将纹状体分裂成具有 D1 或 D2 受体的中等多棘神经元。与 Albin 等人一致(Albin, Young 和 Penny,1989),多巴胺抑制了 D2 介导的途径并促进了 D1 介导的途径。因此,多巴胺的减少增强了控制机制,抑制了*选择机制*。随着模拟中多巴胺水平的降低,多通道选择的可能性也降低了。如果多个通道的选择与两个或多个同时行为的选择和操作相对应,那么模型中减少多巴胺的作用与早期帕金森病的“相关运动丧失”现象相一致。模型中多巴胺的低水平会阻止任何行为或资源的选择。这与 Akinesia 是一致的。随着多巴胺水平的增加,一些异常的可能性也会增加。除了正确选择最突出的通道外,在多巴胺水平中等的模型中,两个高度突出的通道可能会导致彼此无法选择。这导致两个通道都没有被选择。在多巴胺含量高的情况下,可以选择许多具有高显著性的竞争通道。在选择系统中,这些行为可能是不可取的。尽管这些异常现象可能与某些行为条件(例如,可能是精神分裂症)有着松散的联系,但它们作为实验可预测行为的解释仍然不清楚。

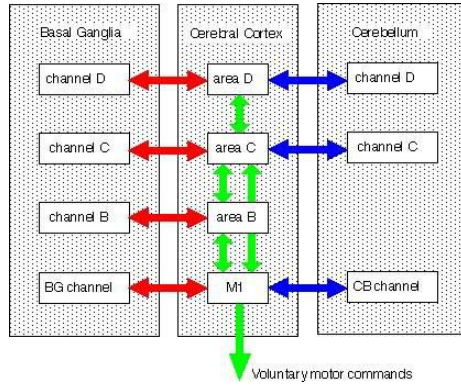


图 13: 连接大脑皮层、基底节和小脑的分布式模块化结构(见(Houk 2005))。绿色双向箭头表示大脑皮层不同区域之间主要的兴奋联系; M1 表示产生自主运动指令的初级运动皮层。红色双向箭头表示大脑皮层任何特定区域和基底节之间的主要抑制联系。蓝色双向箭头表示大脑皮层大部分区域和小脑之间的联系。蓝色是小脑相互作用的双环, 一个环穿过小脑核, 另一个环穿过小脑皮质。一个单独的皮质区域及其穿过基底节和小脑的环路形成了一个强大的计算结构, 被称为分布式处理模块(DPM; (Houk 2005))。DPMs 通过皮质皮质连接相互通信。在人脑中有大约 100 个 dpm, 形成了一个大规模的神经网络。

Wickens(1993)的纹状体理论引用了局部生物物理和全局连接机制的结合——这是一个混合模型。生物物理水平涉及纹状体中 MSN 之间竞争的神经动力学(Wickens, Alexander 和 Miller,1991; Wickens 和 Oorschot,2000)。纹状体内 MSN 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。对 MSN 给予钾离子通道和突触输入, 对其投射大的无棘中间神经元给予胆碱能突触传递(Cholinergic synaptic transmission), 对多棘神经元钾通道有调节作用。无棘中间神经元(图 2 中未明确显示)被给予 DA 输入对它们的神经产生调节作用。DA 和胆碱能活性相互调节竞争的程度。另一个关键特征是将多棘神经元组织成基于细胞解剖学的竞争领域。该模型通过抑制时的波谷预测了多个活动峰, 对于这一点, 作者认为其与肌肉群控制行为的选择有关。这是 Hikosaka, Takikawa 和 KawaGoe(2000)注视控制(gaze control)神经生理模型的一个更复杂的版本, 它解释了如何通过直接路径选择眼球运动的突出目标, 而许多不太突出的替代目标可以通过间接路径取消选择。

在全局层面, Wickens(1993)的理论扩展到了 Albin-DeLong 模型, 假设去抑制丘脑皮质环路导致了 Hebb(1949)设想的皮质细胞集合的扩增。Hebb 的假设与此密切相关, 即局部学习规则(Local learning rule; LTP)允许形成细胞集, 作为重复再激活的函数, 就像一个人执行一个动作时可能发生的那样。以细胞集为目标的直接途径通过 BG 的传输将“引发”正反馈, 就像吸引子网络的活跃状态一样, 而通过间接 BG 途径的传输将抑制正反馈并使吸引子美化。大脑皮层的基础神经节控制的复杂平衡需要要求大脑皮层网络在操作和学习时间尺度上稳定运行。Miller 和 Wickens(1991)讨论了这些问题, 并将模型与选择性关注(selective attention)以及可预测和可控事件的表达联系起来。

Houk 和 Wise(1994)提出的运动规划和控制模型除了通过 BG 外, 还包括通过小脑(CB)的回路。该模型的结构建立在图 1 总结的空间分布网络的解剖模块上。循环中的信号处理操作是从行为动物的神经元活动的细胞外记录中推导出来的。该模型的最新版本(Houk 2005)包含了关键的细胞和分子神经科学数据, 这些数据突出了 BG 环路中多棘神经元和小脑皮质环路中浦肯野细胞(Purkinje cell, 其是从小脑皮质发出的唯一能够传出冲动的神经元。浦肯

野细胞是自律性细胞,当有来自窦房结的节律兴奋传过来时,浦肯野细胞就按窦房结的节律兴奋;在没有刺激的情况下,浦肯野细胞就能产生动作电位。)的独特计算特征和特殊学习规则。图 1 所示的每一个 CB 回路都包括一个穿过小脑核的相互兴奋回路,该回路传递正反馈,其行为类似于吸引子网络,类似于 Wickens 的全局模型中的 Hebbian 脑皮质集合。从浦肯野细胞到核细胞的显著抑制性投射被认为是对 CB 吸引子网络(Houk 和 Mugnani 2003)的固定点施加了强有力的控制。该机制可以解决 Wickens 的全局模型所面临的稳定性问题。

Redgrave 等人(1999a)认为,多巴胺能细胞对意外出现的行为显著性刺激的短时间、短潜伏反应,是过程的一个重要组成部分,这些过程负责真实定位注意力和行为资源,使其有利于意外的显著性事件。据推测,这样的爆发产生了一种相对非分化的多巴胺能输入波到纹状体的大面积区域。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”输入/输出响应,即 GPi。作者(Redgrave 等人,1999a)指出,脊椎动物 BG 已经进化成一种中枢选择装置,专门用来解决多个子系统之间的冲突,这些子系统争夺有限的运动或认知资源。Redgrave、Prescott 和 Gurney(1999a)结合了自下而上和自上而下的方法来模拟穿过 BG 的环路。自下而上的模型来源于神经科学数据,类似于过去已经讨论过的某些模型。相反,自上而下的方法从生物体面临的主要行为问题开始。他们运用行为学和控制论的观点来识别和描述一个自主存在所面临的动作选择问题。在这个框架下,选择操作有选择地解除对获胜竞争者“感觉-运动”连接的抑制,同时保持或增加对失败竞争者的抑制控制。多巴胺爆发的另一个隐含功能是,在奖励刺激被定位和消费之前,首先有必要中断正在进行的行为,并切换注意力和行为资源来处理奖励事件。人们发现,意想不到的新奇或强烈的刺激总是会引发短期的多巴胺爆发,从而证实了这种作用。他们的理论是,BG 是脊椎动物解决动作选择问题的方法。这一理论与下面关于强化学习(Reinforcement learning)一节中讨论的 Actor-Critic 模型非常吻合。行为(Actor)的抽象模型在人工智能框架中可以实现动作选择(见 Sutton;Barto(1998)第 2 章)。

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)性能模型。该模型结合了前额皮质(PFC)和 BG 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 BG 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 BG 环。PFC 观察到分类规则的主动维持,而 BG 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。BG 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)。BG 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 BG 的反馈改变其排序规则。请注意,这一假设不同于向 PFC 输入 BG 的假设,因为在前者中,在 WM 中的信息维持不依赖于 BG 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#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，BG间接通路不纳入模型。而且，该模型不能解释对BG的损伤导致WM任务(包括WCST)中出现持续反应的结论。

Gurney, Prescott 和 Redgrave(2001)的基于人口的模型对BG的计算架构给出了不同的解释，特别是图2中的GPe阶段。他们通过BG将动作选择留在直接路径上，通过皮质到STN再到ON的间接路径将去选择(De-selection)留在间接路径上，但他们认为进出GPe的路径是一个控制回路，稳定了先前的选择路径，提高了选择性。在之后的一篇文章研究中，Humphries 和 Gurney(2002)介绍了一个基于人群的丘脑皮质电路模型，并展示了它是如何与BG处理的固有模型兼容，并实际改进了前者的动作选择性能。

Prescott 等人(2002)以动作选择假设为组织原则，提出了一种新的基于选择和控制电路的功能分组方法。该模型的强度已经通过将其嵌入一个移动机器人的控制体系结构中进行了测试，该机器人具有少量的类动物行为(参见Prescott 等人,2002)。这项工作证实了模拟的BG可以在需要适当和及时的行为切换的真实环境中提供有效的动作选择。该机器人模型还提供了一个洞察异常多巴胺调制动作选择的紧急后果。例如，让人想起帕金森病的一些运动症状的是，多巴胺的减少被发现会导致无法选择适当的行为或曾经选择完成的行为。

最近的研究已经为纹状体内的局部抑制提供了证据，这种抑制要么是通过局部间神经元介导的，要么是通过输出细胞之间的相互抑制网络介导的(Oorschot 等人,2002)。Wickens(1997)运用网络模型研究了纹状体神经元这种局部邻域的动态。在拓扑结构和大小不同的假设下，观察到相互抑制通常会导致网络动态竞争，也就是说，最活跃的神经元往往会抑制其较不活跃的邻居的活动。这项研究还解释了模拟多巴胺输入的影响，表明多巴胺低的环境下，网络的动态变化与竞争“同激活”(活动均匀分布于神经元的局部种群中)，这种模式可以为多巴胺缺乏的帕金森患者的肌肉僵硬提供一个模型。

由于许多环路通过BG辅助大脑皮层的认知区域，BG架构显然参与了思考、计划和行动的选择(Prescott、Bryson 和 Seth,2007)。前额叶皮质的PBWM模型及其穿过基底节的环路(Hazy, Frank 和 O'Reilly,2006)解释了在几个认知任务中的决策表现。PBWM是图2中所示的信号处理操作的抽象实现。这些作者并没有在几个潜在的行为之间建立一个选择模型，而是在各种认知任务中模拟“行动”与“不行动”的决定。由于实际上有数千个微观模块组成任何给定的宏观模块，因此大量潜在的行为或想法可能是值得研究的。Frank 已经开始增强具有神经科学特征的PBWM模型(Frank,2005; O'Reilly 和 Frank,2006)。

Rubchinsky、Kopell 和 Sigvardt(2003)以及Humphries、Stewart 和 Gurney(2006)的模型在他们的生物物理学中非常详细。例如，后一论文的作者研究包括了STN和GPe中多巴胺的影响、神经元之间的传输延迟、树突上突触输入的特定分布、实验研究得出的参数。虽然重点是解释了STN和GPe之间的反馈如何导致各种振荡，但该模型还能够以与Gurney 等人的(2001)人口水平模型一致的方式进行动作选择。在正常情况下，STN-GPe反馈回路被持续的多巴胺功能性地解耦，并在多巴胺耗尽后重新耦合，后者模拟了许多体内实验条件和帕金森病的情况。BG细胞或电路开始出现许多类似的详细模型，更多的关注时间细节，即在网络及其模型中存在高频振荡。目前还不清楚高频振荡是简单的表现现象还是大脑实际用于信号处理。

Matell 和 Meck(2000; 2004)的纹状体节律模型(*The striatal beat frequency model, SBF*)依靠低频振荡(~5-15Hz)通过奖励正确的动作来学习秒到分钟的时间间隔。为了总结该模型的操作(参见 Lustig 等人 2005 中的图 2),重置刺激在皮层开始振荡活动,在不同的皮质单位以不同频率反射。多棘神经元学习检测皮质输入向量中的活动模式,大概是通过如图 2 所示的竞争模式分类操作。在不同的时间间隔内会出现不同的空间重合模式,因此多棘神经元在强化学习的基础上学习检测与所需时间间隔相对应的模式。这个模型是通过对比皮质和纹状体神经元同时记录(Matell, Meck 和 Nicolelis,2003)来测试的。数据支持这一电路在间隔时间中的重要作用,但不支持振荡假设。

核内作用选择模型使用解剖学数据显示纹状体中多棘神经元之间的侧氨基丁酸能连接(Wickens,1993,1997;Beiser 和 Houk,1998;Plenz,2003)。在这些模型中,一个“winner-take-all”机制,它构成了许多动作选择方法的基础,是使用神经网络收敛到一个单一的“winner”。假设 MSNs 之间的抑制连接的密集网络抑制相邻神经元,从而维持单个神经元的活动。选择过程发生在纹状体内,然后选择的作用通过不同的途径传递到 BG 的输出层。这些模型也被用作“Actor-Critic”模型(Houk 等人,1995;Baldassarre,2002)中的 Actor 部分,其中纹状体的基质部分使用来自 SNc 的奖励信号来选择正确的行动或更高级别的计划。虽然大多数模型都集中在单个动作的选择上,但是在“软”选择上已经做了一些工作,在“软”选择中可以同时选择多个动作。实施这种模式的模式是“winner-shares-all”模式(Fukai 和 Tanaka,1997)。这个模型是“winner-take-all”机制的一个变种,它依赖于神经元的侧抑制和自我抑制之间的关系,从而使多个神经元变得活跃。这种机制允许从选择模型过渡到更动态的系统,该系统根据依赖于整体输入的变化阈值提供输入。

核间选择模型在其选择机制的运作方式以及不同核与 BG 通路之间的相互作用方面有所不同。其中一个模型描述了 BG 以两种基本方式对“丘脑-皮质”网络的抑制:同时或连续激活相反的途径(Hikosaka 等人,1993)。同时激活时,直接和间接路径共同激活核输出,从而使输出目标的空间聚焦和锐化。在顺序激活(sequential activation)(也称为时间标度(temporal scaling))中,间接途径保持非选择性抑制,随后由直接途径执行选择性“去抑制”。根据“聚焦选择模型(focused selection model)”(Mink,1996),BG 接收来自多运动模式发生器(*multiple motor pattern generators, MPG*)的输入,通过释放(或解除)单个模式,同时抑制其他模式,从而实现正常的运动功能。使用这种“聚焦选择(focusing selection)”,可以防止竞争的 MPG 同时工作并破坏正常的运动功能。与早期的“盒箭”模型不同,该模型并没有将不同的 BG 核描述为统一的实体,而是探讨了同一个核内的表征之间的内在关系。在“winner-lose-all”模型中(Berns 和 Sejnowski,1996),间接途径增强了 GPi 的背景活动,导致抑制了所有的行为,而使用直接途径抑制单个程序则导致了对信号的抑制(因此采用术语 winner-lose-all)。运用不同途径之间的时间差异来阻止竞争行为。该模型假设存在与神经元池相对应的“单元”。这些“单元”是从纹状体经苍白球到丘脑的分离信息流的一部分。与 STN 的扩散激励相比,Gurney(2001)等人提出的模型也是基于纹状体发送的“聚焦信号(focusing signal)”的概念。这就实现了偏离中心(off-center)的环绕模式(on-surround)的激活。然而,在这个模型中,GPe 被添加为一个控制回路,有助于稳定选择路径并提高选择性。这与 GPe 在 Berns 和 Sejnowski 模型中的作用不同,在该模型中,GPe 用于为实际选择过程创建关键的时间延迟。此外,Gurney 的模型支持“硬”和“软”选择(单个操作与多个选定操作)。最后,一些模型试图将 BG 的动作选择网络置于一个更大的框架内。这包括丘脑和丘脑皮质连接作为动作选择网络的一部分的作用(Humphries 和 Gurney,2002)。

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。

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，同一组人模拟了动作选择与  $\gamma$  振荡和  $\beta$  振荡之间的关系(Blenkinsop,Anderson 和 Gurney,2017)。

### 3.3.序列生成(sequence generation)与依序处理

另一种模拟 BG 的方法是将 BG 与选择或生成动作序列而不是单个动作的作用联系起来。BG 的序列生成和检测模型可分为两大类。第一种假设 BG 内部有一种机制，用来维持构成序列的动作或状态的记忆。第二，BG 是序列作用机制的“无记忆”部分。在这类中，状态由皮质维持，并通过皮质-纹状体连接反映到纹状体。依序处理是高阶智能的一个重要特征(Lashley,1951)，是 Beiser 和 Houk(1998)混合模型的研究焦点——该模型主要起到连接作用，但也包括重要的生物物理特征。用 T 型钙离子通道捕捉丘脑神经元对 BG 回路抑制输出的环绕模式(Rebound responses)，运用 SpNs 间的 GABA<sub>A</sub> 脉络抑制进行竞争模式分类。该模型编码感官事件序列的能力来源于三个协同工作的计算特征：(1)通过纹状体内的计算对皮质输入向量中的模式进行分类；(2)皮层丘脑回路模式分类结果的工作记忆；(3)由于环路将先前分类的工作记忆从皮层储存到纹状体的更新输入向量中，所以产生了一种类似递归的操作。更新后的向量不仅表示当前事件，还表示充当时间上下文的先前事件。

BG 的特征解剖特征之一是环路的观察。皮质-BG-丘脑皮质环路提出了一个在串行处理中起作用的概念。早期的循环神经网络模型被用来学习和执行一系列的任务(Jordon,1986)，并且它们已经产生了一些早期的 BG 模型(Brotchie,Iansek 和 Horne,1991)。长期以来，BG 参与运动控制和规划的可能性为预测序列处理能力提供了动力。计划的动作序列不仅需要序列处理能力，还需要序列事件的表示。在电机运行中的作用和 BG 环路的性质结合在一起，为 BG 模型的串行处理提供了强大的动力。

正如我们前面所看到的，序列处理和强化学习在 BG 中的作用是相容的(Suri 和 Schultz,1998;Dominey,Arbib 和 Joseph,1995)。在 Suri 和 Schultz(1998)模型中，环境为序列中的下一个项目提供了线索，而在 Dominey 等人(1995)的模型中，序列的内部表示是通过前额皮质产生的。然后可以使用这个内部表示来重现序列。Beiser 和 Houk(1998)提出了一个直接通路的模型，用于生成这些时间序列事件的表示。他们提出 BG 充当编码器，将连续事件处理成前额叶皮质活动的空间表示。尾状核中的多棘神经元起着背景事件检测器的作用。因为每一个神经元接收多达 10000 个不同的皮质纹状体传入(Kindaid,Zheng 和 Wilson,1998)，这使得这些纹状体神经元处于发挥这种作用的佳位置。模型的前额叶皮质接收(预处理)环境的刺激，从而产生最初的活动模式。尾状核中的中等多棘细胞间的横向抑制只允许有限数量的中等多棘神经元对这种活动模式作出反应。那些确实对抑制丘脑神经元有反应的神经元，在丘脑和皮层之间进行持续的反复活动。丘脑输入，再加上对前额叶皮层的新刺激，改变了它的活动模式。这种新的前额叶活动模式反过来增强尾状核反应。Beiser 和 Houk 演示了这

个持续循环过程如何在前额皮质构建一系列刺激的独特表现。

Berns 和 Sejnowski(1998)的一个模型说明了序列产生和编码过程中短期记忆的皮质位置的一种替代方法。在这个模型中，提出了将下丘脑-苍白球环路作为这种短期记忆的位置。在学习过程中，纹状体反复在 GPe 上施加时间序列模式(通过直接抑制)。由于对 STN 的输入来自 GPe，并且被赋予了显著的连接延迟，因此 STN 上的活动模式至少比施加在 GPe 上的活动模式落后一个时间步长。因此，学习将 STN 中的前一步长按时间顺序与 GPe 中的当前步骤相关联。当 GPe 反馈给 STN 时，这允许学习和生成序列。一旦学习到，STN 的活动(序列中的前一步长)生成当前步骤在 GPe 上的活动模式和苍白球内部段(GPi)的输出单位。然后，GPe 将当前步骤输入到 STN，从而成为上一步，循环继续。学习是通过一个强化算法实现的，它允许以一种简单的方式评估多巴胺的影响。多巴胺的损失是通过学习率的整体降低来模拟的，这会导致序列学习的缺陷。

产生活动模式序列的能力也仅在纹状体的结构中提出。Wickens 和 Arbuthnott(1993)证明纹状体内的横向连通性可用于产生时空模式。在他们的模型中，一个中等多棘神经元对另一个中等多棘神经元的抑制通过轴突络脉发生在彼此足够接近的中等多棘细胞之间(称为“抑制域”)。模型表明，在均匀随机输入的情况下，由于相互抑制，一个区域内通常只有一个中等多棘神经元活跃。此外，由于过度极化后电导缓慢增加，活性神经元最终可能变得不活跃，从而允许域内的另一个神经元开始放电。尽管这些时空模式的产生强烈依赖于外部输入，但该模型证明了侧抑制网络产生局部时空模式的能力。模式的性质取决于后超极化电导的强度。加入抑制性中间神经元可提高时空格局的稳定性。这一观点已被扩展到运用中棘神经元(Wickens 和 Wilson,1998)的奖励特性来控制 BG 皮层序列活动的模型中(Plenz 和 Kitai,1996)。

Connolly 和 Burns(1993a,1993b,1995)描述了一个模型，该模型将纹状体内的基质室表示为一个连接电阻网络。从解剖学上来说，这假设了中轴神经元之间的电耦合。有证据表明，由于染料从细胞内电极扩散到多个神经元，导致弱电子耦合(Onn 和 Grace,1999)。然而，在电子显微镜中，主要观察到 parvalbumin 阳性神经元间的缝隙连接(Kita 和 Kosaka,1990)。在该模型中，皮质纹状体输入在阻力网络上设置目标和边界条件，分别对应于现实世界中的目标位置和障碍位置。目标和边界分别由固定的尖谷电压和峰值电压表示，以建立电压景观。在纹状体肢体位置的直接表示假设下，这会产生一个沿着梯度的活动轨迹，并避免边界(电压峰值)。提出了时间活动序列在边界条件下的最优运动轨迹。这个模型主要依赖于纹状体中肢体位置的直接表示。

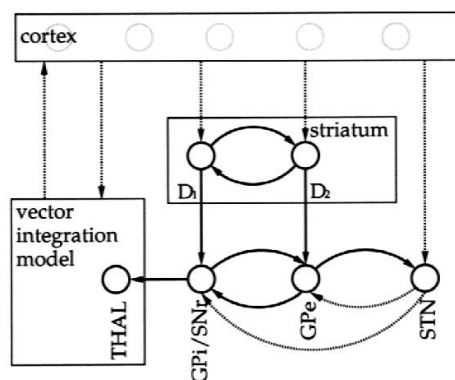


图 14: 丘脑单元在矢量积分模型中起着门的作用，它决定着从肢体当前位置到目标位置的变化速度。实线表示兴奋性连接;虚线表示抑制连接。

Contreras-Vidal 和 Stelmach(1995)提出了一种模型，将 BG 电路直接作为运动的“Go”信号(图 4)。虽然在该模型中，BG 不产生连续活动，但它起到向理论模型提供测量的门控信号



的作用，该理论模型计算目标位置矢量与肢体当前位置矢量之间的差异(Bullock 和 Grossberg,1988)。BG 电路模型的输出控制目标和当前位置矢量之间的整合速率。当系统捕捉到 BG 电路对表示肢体位置变化的向量的影响时，可以直接与实验测量的现象(如反应时间和运动时间)进行比较。该模型包括直接和间接路径，因此，可以测量帕金森病类型条件对这些实验测量参数的影响。根据 Albin 等人(1989)的描述，模拟的帕金森病条件，其中沿直接途径(即与 P 物质相关的途径)的纹状体活性降低，沿间接途径(即与脑啡肽相关的途径)的纹状体活性增强。在这种情况下，模型模拟了某些帕金森症现象。例如，随着多巴胺水平的降低，任务反应时间和运动时间都会增加。对于低水平的多巴胺，该模型不能屏蔽所有的运动和运动类型的行为结果。因此，该模型显示了运动迟缓和运动迟缓的单一机制。

因为这些串联处理的不同模型的基础机制是不同的，所以他们对 BG 疾病如何影响处理的描述也是不同的。例如，Wickens 等人(1991)研究了多巴胺对纹状体抑制区域内活性的影响，如上文所述。Wickens 等人通过模拟纹状体中多棘神经元的特定通道传导，并结合大的多棘胆碱能神经元的影响，证明了在低多巴胺条件下，抑制域的功能发生了戏剧性的变化。如上所述，由于抑制侧方相互作用的竞争，该区域内的单个神经元可能是活跃的。在多巴胺水平降低的条件下，该模型显示了一个区域内神经元的协同激活。有人提出，相邻的神经群可能代表拮抗性肌肉，因此，协同活化可能与僵硬等症状有关。

一个使用内部 BG 状态的模型建立在早期动作选择模型的基础上(Berns 和 Sejnowski,1996)，通过添加序列的选择和编码(Berns 和 Sejnowski,1998)。在该模型中，序列学习和生成是通过 GPe 和 STN 之间的互通性编码的短时记忆来实现的。该循环包含以前活动的痕迹，并有效地作为毫秒级的短期内存工作。长期记忆与输出神经元与输入神经元的连接(数百毫秒的时间尺度)或与前额皮质内的记忆存储(秒的时间尺度)有关。

另一个用于复制和区分序列的模型(Dominey,1995)将排序任务分解为子网络。前额叶皮层编码输入的历史(即动物的内部状态)。“皮质-纹状体”连接作为一种联想记忆成分，参与将历史与期望的输出联系起来。纹状体神经元的放电代表运动输出。这种联系是由从 SNc 接收到的增强信号决定的，它调节“皮质-纹状体”突触。除了纹状体识别序列而不是单个动作之外，该模型与动作选择模型相似。同样，Beiser 和 Houk(1998)提出了一种运用皮质-BG-皮质环编码感觉信息的模型。在这个模型中，纹状体分类了包含动作序列背景事件的皮质输入。

或者，在 Fukai(1999)提出的模型中，实际的动作序列通过振荡存储在额叶皮质中。在这个模型中，BG 的作用是解码序列及其组成部分。纹状体中的一组神经元负责选择序列的初始部分。这是作为一个暂时的赢家采取所有机制实现的横向抑制的一部分。另一个纹状体组保留当前执行的动作。STN 对 GPi 的兴奋信号被认为是 BG 的外部信号，在皮质缓冲区内有计划的信号运动转变。实际动作选择概念化为丘脑中继转接的一部分，可能在纹状体和苍白球中有额外的选择性。

多棘神经元在纹状体(图 2 阴影框)内具有广泛的 GABA-氨基丁酸能抑制性脉络，能够使 SpN 模式分类操作具有竞争力。脉络抑制被认为是在某些模型中调解竞争的有效机制(Beiser 和 Houk,1998; Plenz,2003)，而在其他模型中则被认为是无效机制(Tepper、Koos 和 Wilson,2004)。在 Houk 等人(2007)，模拟了两种脉络抑制，即突触后和突触前。虽然发现突触后抑制是有效的，但突触前抑制更明显。由于纹状体中的突触前抑制似乎是系统学上较新的机制，这可能有助于解释由 Crow(1997)建模的精神分裂症的中心悖论。

在 Botvinick 和 Plaut(2006)联结主义模型(connectionist model)中，序列信息通过简单循

环网络(SRN)体系结构中的持续激活模式进行编码。这个计算简单的模型为即时序列回忆的许多基准特征提供了一个简洁的解释,包括那些被认为排除了在这个领域中循环神经网络应用的数据。与大多数竞争方式不同,该模型自然地处理了有关背景知识在*工作记忆*中顺序回忆作用的发现。

我们的许多连续动作依赖于长期记忆,这些记忆可能被储存在背景或者大脑皮层中。Wickens 和 Arbuthnott(1993)通过 SpN 网络中的交互作用对顺序动作进行建模。相反, Berns 和 Sejnowski(1998)提出了一个基于 GPe 和 STN 之间循环延迟的系统级模型(图 2 中未明确显示)。当猴子进行连续的扫视时,纹状体和额叶皮质之间会产生有趣的协方差,这些协方差可以用来标记所学序列的上落和偏移(Fujii 和 Graybiel,2005)。

大量的证据表明,学习在很大程度上取决于激励因素的不可预测性(RescoRLa 和 Wagner,1972; Dickinson,1980)。只有在某种程度上不可预测地发生的强化才会支持学习,当所有的强化都被完全预测时,学习曲线才会达到渐近线。强化的出现与预测的不一致,称为强化预测中的误差。“误差-驱动”学习被广泛应用于各种神经模型中,并在*时序差分*(TD)算法中得到了特别的阐述,该算法实时连续计算预测误差,并建立了增强预测(Sutton 和 Barto,1990)。

TD 算法可以在一个明确的 Actor-Critic 体系结构中实现(Barto 等人,1983; Sutton 和 Barto,1998)。预测误差由 Critic 计算并作为*教学信号*发出,以适应 Actor 的突触权重,从而指导行为输出(Montague 等人,1993; Friston 等人,1994)。这种结构类似于 BG 的解剖结构: Critic 取代了黑质纹状体多巴胺神经元,以及 Actor 对应于纹状体(Barto,1995; Houk 等人,1995)。Critic 的预测误差信号与中脑多巴胺神经元的活动惊人地相似(Montague 等人,1996; Schultz 等人,1997)。这两个信号都是由不可预测的奖励增加的,不受完全预测奖励的影响,并且由于忽略预测奖励而减少。它们通过经验被转移到最早的奖励预测事件中,因此,在奖励发生之前预测它们,而不是仅在行为之后报告它们。

神经生理学和失活研究表明 BG 在运动序列中的重要参与(Kermadi 和 Joseph,1995; Mushiakhe 和 Strick,1995; Miyachi 等人,1997)。纹状体多巴胺神经传递障碍损害了人类患者的序列运动(Benecke 等人,1987; Phillips 等人,1993)。由于强化学习任务可能构成顺序马尔可夫决策过程(Sutton 和 Barto,1998),Suri 和 Schultz(1998)旨在研究具有类似多巴胺反应特征的 TD 预测误差信号是否可用于学习顺序运动。

### 3.4.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”)，否则按左侧的键(“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 PFCBG 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，BG 的间接通路也不包含在模型中。

基于 *BG* 的 *Go/NoGo* 模型，最近的一个模型模拟了 *BG* 与脊髓相互作用在动作选择中的作用(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)。

### 3.5.单通道模型(Single pathway models)

早期的 *BG* 模型强调了信息的中心途径：从皮层，通过纹状体到 *GPi* 的前馈流。尽管关于信息流动方向和相关核的普遍共识，但关于通过 *BG* 流动的多源信息的相互作用仍存在争议。这两种极端的观点认为，要么是来自多个领域的所有不同信息聚合，要么是通过分离的循环并行地流动信息。

Percheron 等人(1984)运用解剖数据证明了多个纹状体来源在同一个苍白球神经元上的会聚。每一个苍白球神经元都可以通过其树突树接近大量的纹状体轴突。这种解剖结构可以导致感觉运动、联想和边缘信息的融合，因此可以作为整合分散的皮质信息的基础。此外，*GPi* 的神经元(*BG* 的输出部分)的这种收敛性最高。收敛不仅是神经元数量减少的结果(“基本收敛”)，而且主要是释放轴突形状和接受树突树状化(“接收收敛”)的结果(Percheron 和 Fillion,1991)。后来的电生理数据(Yoshida 等人,1993)支持解剖数据，表明尽管输入似乎分成两个平行的环，但在每个环内，从皮层的多个来源到苍白球仍然有很高的收敛性。

相反，在并行分离环模型中，来自不同域的信息被认为是在单独的环中分离的，并且不同的并行路径之间没有相互作用。在这个模型中，不同的环路来自不同的皮质区域，在它们的血流中通过 *BG* 和丘脑保持分离，最后在额叶皮质的不同区域结束。最初的划分分为“马达”回路和“关联”回路(Delong 和 Georgopoulos,1981)。该划分扩展到功能上不同的平行环，这些平行环起源于不同的、功能上相关的皮质区域，从中每个分离环内分别发生一个聚合过程(Alexander 等人,1986)。通过将运动回路细分为三个独立且看似分离的路径，将不同的苍白区域与不同的皮质区域连接起来，进一步细化了划分为五个回路(Hoover 和 Strick,1993)。这种细分使得并行处理的概念扩展到不同循环中的处理，而不仅仅是不同循环之间的关系(Alexander 和 Crutcher,1990)。由于躯体定位结构(Somatotopic organization)和在不同核中的分域投影，提出了在环内划分为子通道的建议。多回路的分离限制了信息的收敛和集成，最多只能在狭窄的子通道中进行。最近的出版物往往描述了大量通过 *BG* 的信息循环，例如已经报道了 10 个循环(Middleton 和 Strick,2000)，具有明显的解剖和生理特征。

这两个极端的观点，一个融合了所有皮质信息的集合漏斗，或者一个平行的路径结构，使信息在整个皮质 *BG* 皮质环中保持分离，被合并到一些后来的模型中，以创建更复杂的功能结构。其中一个这样的体系结构是由开放的互连回路构成的(Joel 和 Weiner,1994)。这种结构依赖于纹状体向 *SNr* 和 *GPi* 的发散，假设尽管单个额叶皮质区的输出被发送到 *BG* 的单个输入区，但它接收至少两个 *BG* 输出核的输入。这导致至少一些穿过 *BG* 的循环返回到作为输入的区域以外的区域(即开环)。这一理论与保持信息隔离的闭合回路的思想不同，而是由于返回信息的多个目标，在不同的皮质区域之间呈现出复杂的互连和信息流动。在上升螺旋模型(Ascending spirals model)中，提出了一种不同的并行回路积分方法(Haber 等人,2000)。在这个模型中，环的积分因子不是额叶皮质而是纹状体-黑体系统。该模型建立在“黑质-纹状体”环的基础上，而不是局部相互作用的闭环。相反，黑质(多巴胺能)投射也延伸到纹状体区域，纹状体区域不受神经支配，从而在不同纹状体区域之间传递信息。因此，循环中的信息流被其他循环的活动所调制。

### 3.6.多路径模型(multiple pathway models)

有关 *BG* 连接复杂性的数据积累将模型的焦点转移到不同核和通路之间的相互作用上。解剖成分及其化学相互作用的图谱构成了 *BG*“盒箭”模型的基础 (Albin 等人,1989;DeLong,1990)(图 1)。在该模型中,不同的核被建模为简化的单位(“框”),根据传入核的速率变化及其连接类型(“箭头”),可以增加或减少它们的总体配置速率。*神经递质*谷氨酸盐的连接是阳性的,*神经递质* GABA 的连接是阴性的。*皮质 BG 皮质环*被认为是一个具有两个竞争要素的反馈环:一个由*直接途径*(*纹状体-GPi*)介导的正反馈机制,而*间接途径*(*纹状体-GPe-STN-GPi*)介导的负反馈机制。因此,*BG* 控制着*丘脑*的整体抑制作用,*丘脑*传递到皮层。这两种途径的相对活性受 SNc 多巴胺能信号的控制,对这两种途径有相反的作用:通过 D1 受体增加*直接途径*的活性,同时通过 D2 受体减少*间接途径*的活性。该模型尤其成功地解释了与 *BG* 相关的低动力和高动力运动障碍(DeLong,1990)。*直接途径*中的活性降低和*间接途径*中的活性增强(由于多巴胺耗尽)导致 *BG* 输出结构中的活性升高。这导致通过*丘脑*中继传递抑制*皮质*,导致运动机能减退(如帕金森氏病)。另一方面,*直接途径*中的活性增强和*间接途径*中的活性降低会导致输出结构的抑制,进而抑制(即激活)皮层,从而导致超动力障碍(如亨廷顿氏病和偏瘫)。电路本身的组织可以包含*平行回路*或*收敛结构*,因此涉及每个回路或整个结构内的*直接/间接*路径平衡。多年来,已经增加了许多新的*解剖学*数据,如*超直接*(*皮质-STN-GPi*)通路和 *GPe* 到 *GPi* 通路(Wichmann 和 DeLong,1996;Chesselet 和 Delfs,1996)。多年来对原始模型所做的各种修正成功地解释了一些新的*解剖学*和*生理学*数据,但也增加了其复杂性。然而,尽管对 *BG* 的缺陷和生存能力存在争议,但“盒箭”模型仍然是理解 *BG* 信息处理的主要模型(Wichmann 和 DeLong,1996;Chesselet 和 Delfs,1996;Parent 和 Cicchetti,1998)。虽然“盒箭”模型是理解 *BG* 相关的低动力和高动力病理学的一个有价值的工具,但它对健康 *BG* 的计算提供了很少的见解。通过对完整核的大体表征和单纯观察放电速率来分析*皮质-BG-皮质环*,势必会忽略复杂的放电模式、每个核内的计算以及同一核内神经元之间的相互作用。

## 四、BG 的计算模型

### 4.1.从机器人到膜

近年来,越来越多的计算模型已经解决了 *BG* 功能的各个方面。构建这类模型的*动机*源于对解释与 *BG* 相关的复杂生物数据的迫切需要,这种需求在不断增加。在过去,微电路的定性信息流(box-and-arrow)模型(Gerfen 和 Wilson,1996)、*BG* 核之间的内部连接(Albin,2001;Mink,1996;Gillies 等人,2002)及其与外部结构(Prescott 等人,1999; Alexander 等人,1986;Alexander 等人,1986)的相互作用对于解释广泛的实验数据是有用的,并指导了许多最近的 *BG* 研究。然而,解剖学、生物化学、生理学、药理学和行为学信息的快速积累暴露了定性模型在解释当前数据和预测未来实验结果方面的不足(图 1)。为了进一步了解 *BG* 内信息处理的功能动力学及其与 *BG* 相互作用,以及它与大脑其他部分的相互作用,需要 *BG* 生物学各个方面的定量模型。随着对计算模型的预期衍化要求,所有这些模型都需要具有不同程度的生物合理性,因此,对于实验工作者和建模者来说,了解模型的不同类型和级别、它们的基本假设和局限性、它们如何相互关联以及如何最好地验证它们都是至关重要的。因此,本文观点的主要目的是提供一个组织框架,在该框架内可建立 *BG* 的广泛应用的计算模型。

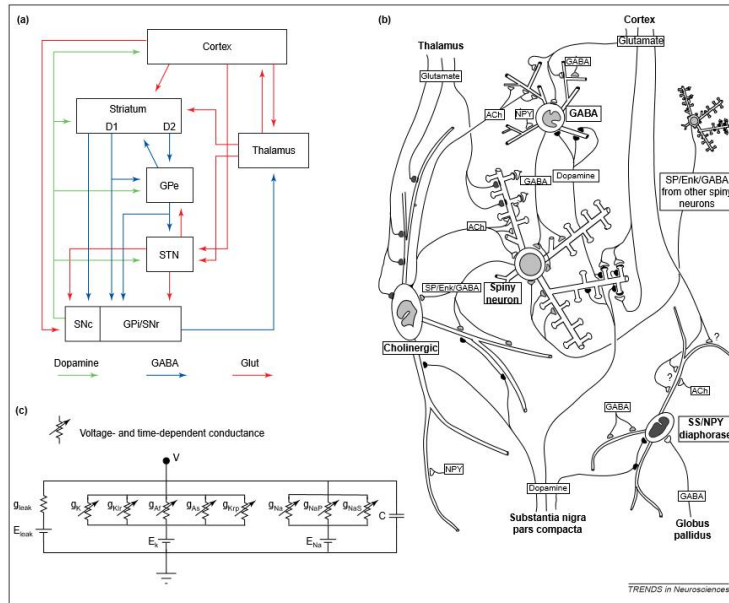


图 15: BG 不同结构水平的生物学复杂性。(a)表示 BG 亚核、大脑皮质和丘脑之间当前已知连接的汇总框和箭头图。缩写: D—多巴胺受体; GLut—谷氨酸连接; GPe—外苍白球; GPi—内苍白球; SNc—黑质致密; SNr—黑质网状; STN—丘脑下核。(b)纹状体中神经元元件之间连接的示意图。经许可, 根据参考文献[59]©(1995)R.G.Landes Co。(c)复制了神经膜的单个等电位贴片或“隔间”的等效电路, 其中表示了多个离子种类的电导[53]。所有面板的复杂性说明, 在每一个层次上都存在这样的细节, 以至于不可能通过神经或电路的动态信息处理“思考”一种方式。

我们提出的框架基于两个基本思想。第一个问题最初由 David Marr(1982)提出, 当他提出大脑功能处理计算问题的解决方案时, 这些问题被分解为三个层次进行分析:

- ①“什么”正在被计算, 为什么要计算——计算任务;
- ②“如何”计算——算法;
- ③“在哪里”计算问题——算法实现。

第二种观点是, 对 Marr 的三层次分析可应用于描述(Churchland 和 Sejnowski,1992)的几个结构层次中的每一个(框注 1)。因此, 计算问题可以在从膜到整个大脑系统的神经组件中解决——因为每个模型都可以提供重要的计算基础, 所以没有首选的建模结构层次。

**框注 1: BG 及其在脑内位置生物学描述的结构层次(Churchland 和 Sejnowski,1992)**

- I 中枢神经系统(全脑)
- II 大脑模块(如基底节、小脑、皮质和海马)
- III 模块内的核(如纹状体、苍白球和黑质)
- IV 小电路和显微解剖(例如相互抑制、收敛和发散)
- V 神经元和信号代码(例如, 中等多棘神经元和中间神经元)
- VI 突触和膜(如脊柱和轴膜、突触前和突触后膜)
- VII 细胞内信号(例如第二信使系统)

该方案的一般适用性将通过讨论处理 BG 不同结构层次计算问题的最新模型的具体实例来证明。我们将从最高级别的系统模型开始, 很明显, 已经制定了两种不同但可能互补的建模策略。然后, 我们将通过实例说明所提出的框架如何帮助评估神经功能的低层微电路和膜模型。

#### 4.1.1 系统层面的模型

构成 BG 的核被认为是在更广泛的大脑结构中形成一个功能性的子系统(图 1a)。试图理

解 BG 的计算单一作用(或多个作用)的模型, 作为一个整体, 通常涉及动作选择(Gurney 等人,2001;Prescott 等人,2002)、运动序列处理( Beiser 和 Houk,1998;Berns 和 Sejnowski,1998;Suri 和 Schultz,1998;Fukai,1999)和/或降维(Dimensionality reduction) (Joel 等人,2002;Bargad 等人,2003)的高层次方面。许多模型的一个共同特点是具有某种工具性学习(条件学习)形式的适应能力, 其中黑质纹状体多巴胺信号起着关键作用。这一发展反映了短潜伏期多巴胺反应的实验研究(Schultz,1998) 和基于奖励预测误差(Schultz 和 Dickinson,2000)的强化学习方法(Schultz,2002)之间的显著融合, 这一研究最初是由机器学习理论家(Sutton 和 Barto,1998)开发的。这项工作产生了一些突出的和完善的系统级模型, 这些模型在其他地方得到了广泛的验证(Joel 等人,2002;Bargad 等人,2003;Schultz,2002;Gillies 和 Arbuthnott,2000;Suri,2002)。这些模型是系统级建模的“自上而下(Top-Down)”策略的主要示例, 我们将与我们和其他人用来研究 BG 潜在功能的主要“自下而上(Bottom-Up)”方法进行对比。为了解释这两种方法之间的差异, 我们将在最近文献中两个对比例子的背景下介绍“机制映射(Mechanism mapping)”和“机制挖掘(Mechanism mining)”的概念——Bar Gad 等人提出的一种强化驱动(Reinforcement-driven)的降维模型(Bargad 等人,2003), 和我们自己实验室的选择模型(Gurney 等人,2001;Gurney 等人,2001)。

#### 4.1.2.1.计算假设(The computational hypothesis)

自上而下和自下而上策略通常都是从某种形式的计算假设开始的, 这种假设可以限制对可能机制的研究。因此, Bar Gad 等人(2003)提出, BG 的主要计算任务是进行强化驱动的降维, 从多个感官、运动、情感和认知来源获取输入, 并将此信息的压缩编码传递给参与执行计划和动作选择的大脑区域(如额叶皮质)。这一过程被视为满足管理数据流入执行系统的需要, 以及以有助于快速有效地状态-动作映射的学习方式对代理-环境状态信息进行编码的要求。相比之下, Gurney 等人(2001)的选择模型基于这样的假设: BG 构成了一种通用的资源选择机制(Redgrave 等人,1999)。在这里, BG 被视为一种生物解决方案, 通过多个功能独立的动作系统来调节对“最终共同运动路径”有限且大部分共享的运动资源的访问。这两个模型在方法上的差异出现在算法描述的层次上。

#### 4.1.2.2.机制映射(Mechanism mapping)

Bar Gad 模型调用高级“透明计算(Procedurally transparent)”算法——主成分分析(PCA)——以执行所需的降维。对于“透明计算程序”这种算法, 可以在一般数学术语中定义, 而无需参考神经架构。在这种情况下, PCA 是一种经典的统计方法, 它以最小的信息损失压缩数据(Oja,1982)。然而, 该模型在透明算法及其(神经)实现之间需要两个映射方面也是具有代表性的。首先, 将高级过程(PCA)映射到一个人工神经网络(ANN), 该网络使用诸如输入加权和以及赫布学习法等机制进行实例化。然后需要第二个映射, 由此在神经网络模型中元素的活动和 BG 生物学的细节之间绘制出平行线。

#### 4.1.2.3.验证自上而下模型(Validating top-down models)

使用透明计算的算法有助于清晰地分析了解由人工神经网络生成的输入-输出的转换。然而, 为实现这一清晰性所付出的代价是, 第二阶段的映射——从人工神经网络到其在生物底物中的假设实现——是个更大的问题(Joel 等人,2002;Pennartz,1996)。神经机制必须通过自上而下的分析来确定确切的规则。因此, 尽管基于程序的模型可以提供对执行高级任务所需操作类型的重要参考, 但其验证的关键方面是成功的“机制映射”, 即从人工神经网络到生物神经电路。这个过程的成功程度揭示了大脑是否可能使用类似的操作。

#### 4.1.2.4.机制挖掘(Mechanism mining)



与受已知数学方法启发的方法不同，Gurney 等人的模型(2001)目前没有程序计算的算法。相反，它包含了一组生物约束的神经元机制，这些机制协同作用以实现计算目标。这种模型的基础方法是基于能够实现总体目标的潜在机制的“挖掘”。就 Gurney 等人的这种情况而言(2001)，模拟和定量分析表明，基于竞争输入通道中的相对显著性，适当的输出目标不受抑制，从而确定 BG 的环通道可以作为一种合理的作用选择基质。

#### 4.1.2.5.子组件功能(Sub-component functions)

除了提供一个测试高层次计算假设的平台外，体系结构受到生物数据严格约束的系统级模型也能够提供对网络子组件功能特性的参考。例如，关于丘脑下核和苍白球之间反馈环路作用的新假设(图 1a)是 Gurney 等人建立的模型(2001)的一个未预料到的结果。在分析和模拟中，该回路确保来自输出核(黑质网状部和脚内核)的信号保持在严格的操作范围内，与活跃竞争通道的数量无关。当无环路时(“损伤”)，来自丘脑下核的兴奋性驱动随着活跃通道的数量增加而增加，最终胜过纹状体施加选择性抑制的能力。这种“容量缩放”的突然出现的特性完全是不可预测的，很难从当前的生物数据中辨别出来。

#### 4.1.2.6.验证自下而上的模型(Validating bottom-up models)

所有模型的输出都应该类似于相应的生物系统。然而，与用于验证已经在这里讨论过的程序启发模型的其他方法不同，需要不同的策略来评估系统级模型，其中模型架构直接从生物数据构建。它们可以列举如下：

①在所有现有模型中，所代表的生物特征都是被高度选择和简化过的，因此，有效的评估策略是添加更多的生物约束细节，并重新测试模型执行指定计算功能的能力。在功能保持完整或实际改进的情况下，原始计算假设将获得强有力的支撑。如果表现不理想，初始的猜想的有效性将受到质疑。控制程序，包括在模型中添加“非生物”特性，预计会损害模型的性能。正是基于这些考虑，Gurney 等人的选择模型已经接受了“添加生物实在论”的进一步测试：BG-丘脑皮质环路的添加 (Humphries 和 Gurney,2001)、从“直接途径”到苍白球的络脉的内含物(Gurney,2001)、苍白球中元素与输出核之间的横向抑制连接的添加 (Humphries 等人,2003) 以及最近的，伴有棘波神经元的漏积分(Leakage integral)元素的置换。在每种情况下，模型的整体选择能力都得到了提高。在我们测试了添加“非生物”版本连接的影响的示例中，模型的整体选择性能受到影响。

②对系统级模型的进一步、更严格的测试是将其嵌入到自主代理的控制体系结构中。在这种情况下，通过机器人与其环境之间的紧密动态耦合，模型将暴露在更真实的感官输入序列中。嵌入到更广泛的系统中还需要考虑模型将如何有助于明显的行为，以及它应该如何与 BG 外部的感觉和运动系统连接。同样，考虑到这些因素，Gurney 等人的生物学约束模型已经成功地展示了如何选择进行模拟觅食任务的移动机器人的动作 (Prescott 等人,2002; Girard 等人,2003)(图 2)。

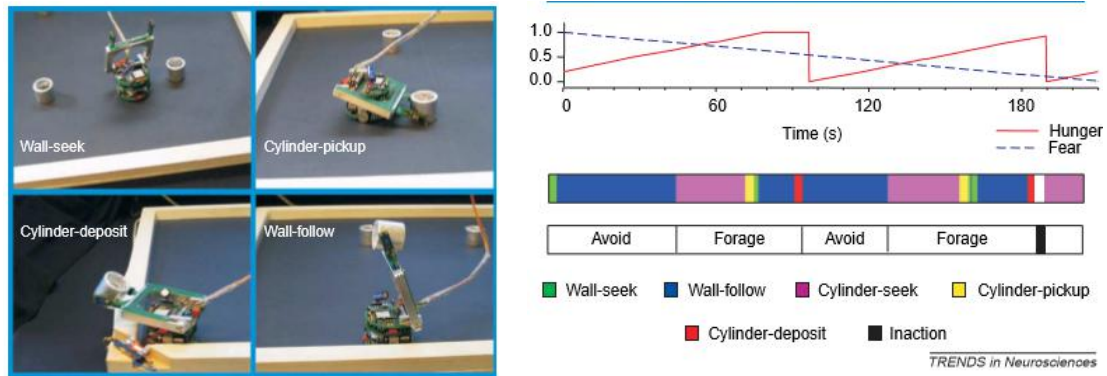


图 16: 由 BG(包括 BG-丘脑皮质环路)的生物约束计算模型控制的机器人的动作选择[24]。行动子系统(照片)之间的选择竞争以 8 个周期  $s^{-1}$  的速度解决。模型的输入根据不同行为的感知和动机提供对其瞬时显著性进行编码。每个循环的结果是从获胜子系统生成的运动命令中去除“黑体抑制”。机器人的行为表现为一系列特定的行为(以颜色编码的时间条显示)和更高阶的综合行为(避免和觅食),中间夹杂着不活动(黑色编码)。行为的时间组织反映了由模拟的“恐惧”和“饥饿”水平编码的机器人的动机优先权,这些水平的相对强度在中心图中随时间以任意值表示。

#### 4.1.2.7.总结

*BG* 系统级建模的两种不同和互补的策略：*自上而下*的策略，其中透明计算算法的性能被“映射”到生物系统，以及更多的*自下而上*的方法，在这些方法中功能能力被“挖掘”到生物受限的架构中。现在我们来了解一下 Marr(1982)提出的分析如何同样适用于结构层次结构 (Churchland 和 Sejnowski,1992)中较低描述级别的计算模型(框注 1)。这些模型通常采用*自下而上*的方法，目的是找出(“挖掘”)潜在机制和计算假设。

#### 4.1.2.微电路模型

目前很少有正式的分析知识可以让我们从神经元的组成及其相互连接的细节中推断出神经电路的运作。即使有电路内部连接的完整示意图，以及对组成神经元生理特性的完整描述，我们通常也无法推断出问题电路的运作特性 (Hopfield 和 Tank,1986)(考虑图 1b)。然而，通过对同一个电路进行建模，就可以检查分布在整个网络上的输入-输出关系，并且可以得出有关正在计算的内容的想法。这种方法是“*机制挖掘*”的另一个实例。反过来，所得到的机制可以用作正在执行的总体计算的线索。这种产生电路级假设的*自下而上*策略可以被看作是一种强有力的补充方法，用于验证来自病变(受损)证据或理论分析的假设。因此，一系列的计算模型被用来研究包括纹状体、苍白球和下丘脑在内的 *BG* 核内和 *BG* 核之间的微电路特性(Gillies 等人,2002; Kotter 和 Wickens,1995; Plenz,2003; Plenz,2003)。为了说明提议框架对这一实质性工作的适用性，我们将考虑纹状体内电路的单一示例。

##### 4.1.3.1.纹状体的模型(Models of the striatum)

纹状体微电路模型最初受到多棘投射神经元及其络脉形态的限制。一般来说，每个多棘神经元轴突在其自身树突状树的区域内形成外延的侧支，与其他多棘神经元形成抑制性突触连接 (Wilson 和 Groves,1980; Somogyi 等人,1981)。最初的模型假设这种解剖结构的排列将产生一种机制，由构成基本计算单元的神经元之间具有强大交互连接的“侧抑制型网络”构成。基于这一假设的模型显示出一种竞争动态，即在高增益条件下，最活跃的细胞抑制其较弱相邻细胞(Wickens 等人,1991)的活性。在完全对称的情况下，这样的网络表现出稳定的高活性且被不活跃的 (Alexander 和 Wickens,1993)谷所包围的峰。这些模型所提出的纹状体电路计算任务与本文所述的选择假设(Redgrave 等人,1999)一致；也就是说，它们从相互竞争、相互排斥的备选方案中选择一个行动。

然而，随后的实验生物学已经表明，棘状投射神经元之间的抑制作用并不像这些早期模

型中假设的那么强。事实上，这些连接在本质上通常是稀疏的——相邻细胞之间的接触概率远远小于 1，而且主要是不对称的 (Plenz,2003;Czubayko 和 Plenz,2002;Tunstall 等人,2002;Taverna 等人,2004;Koos 和 Tepper,1999)。此外，抑制性间神经元虽然在纹状体中占细胞总数小于 1%，但在抑制性突触中所占比例接近 10%，在**前馈抑制**中发挥重要作用(Koos 和 Tepper,1999;Gurney 和 Overton,2004)。基于这些新数据的建模研究表明，对称性的破坏可以产生活动周期和行波(Plenz,2003;Wickens 等人,1996)，这可能在正常和病理条件下的**动作选择**和序列组织中发挥更复杂的作用。例如，其他结果表明，**前馈**神经元可能在初始化网络以执行扩展序列 (Alexander 和 Wickens,1993)中发挥重要作用。

这个例子表明，基于现有的最佳生物数据研究模型的动力学可以产生仅凭生物数据难以直观理解的功能假设。**新纹状体**微电路建模的历史也表明，这些模型的有效性关键取决于所导出的生物观测结果的准确性。因此，详细的定量解剖(Oorschot,1996;Bolam 等人,2000;Ingham,1998;Zheng 和 Wilson,2002) 对于约束可能的连接方案是非常宝贵的。

#### 4.1.3.基于电导的单神经元模型

为了在更低的描述层次上捕捉功能，有必要对代表单个神经元不同部分的隔间中的单个膜电流的动力学进行建模。离子电流动力学通常是用 *Hodgkin-Huxley* 形式来捕捉的，并通过一组常微分方程来表示。在没有模型的情况下，这些电流之间相互作用的复杂性几乎不可能预测结果(图 1c)。因此，提出了 *BG* 中突起细胞类型的一系列基于生物物理电导的模型，包括纹状体中多棘神经元(Kotter 和 Wickens,1995;Wilson,1992;Mahon 等人,2000;Wood 等人,2004)、**苍白球**(Hanson 等人,2004)和**黑质致密部多巴胺**神经元(Canavier,1999;Wilson 和 Callaway,2000)。为了说明 Marr 的方案如何在神经元水平上应用，我们将以纹状体中多棘神经元中观察到的短期促进(STF)效应为例(Mahon 等人,2000)。

##### 4.1.4.1.纹状体神经元的短期促进效应(Short-term facilitation in striatal neurons)

当两个超阈值电流脉冲作用于一个**中等多棘纹状体**细胞时，第二个脉冲与第一个峰值的时间缩短和引出峰值的数量增加相关(图 3)。为了解释这种 STF 效应，Mahon 等人(2000)构建了一个包含多个  $K^+$  和  $Na^+$  电流的中等多棘神经元的单室电导模型。通过采用建模方法，可以使用有意忽略特定电流的模拟来研究不同膜电流对 STF 等现象的影响。使用此策略，Mahon 等人(2000)表明缓慢灭活 A 电流( $I_{As}$ )可能是产生这种促进作用的最重要机制(图 3b)。Mahon 等人(2000)后续建议 STF 在中等多棘神经元中的作用可能是延长检测时间分布输入的窗口，这是一个很好的例子，说明了如何使用**机制挖掘**来生成计算假设。

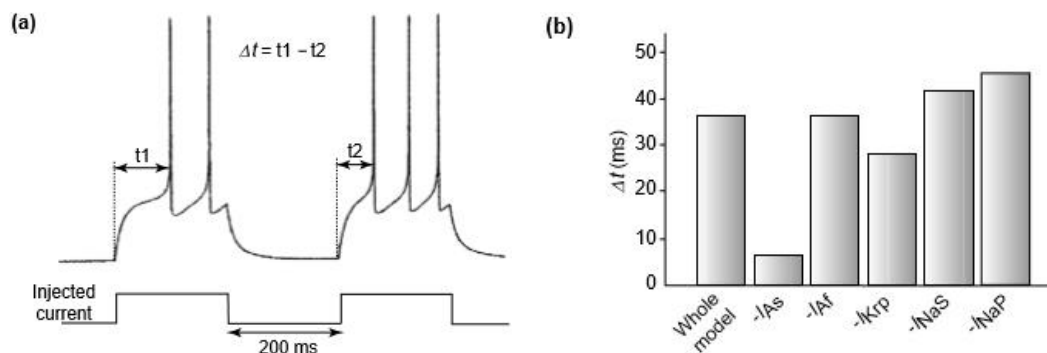


图 17: 一个基于电导模型的模拟结果显示，短期促进(STF)在新纹状体中间多棘投射神经元。经许可，改编自参考文献[53]。(a)由两个相同脉冲组成的电流钳下模型的膜行为。STF 通过时间到第一个尖峰( $\Delta t$ )的变化和每个脉冲引出的尖峰数量来证明。(b)完整模型的  $\Delta t$  和一系列省略单电流的模拟。缓慢灭活的 A 电流( $I_{As}$ )使  $\Delta t$  显著降低，这表明这是最有可能诱导 STF 的机制。所用的  $K^+$  电流包括慢灭活 A 电流( $I_{As}$ )、快速 A 电流( $I_{Af}$ )

#### 4.1.4.2.基于电导的模型的评估(Evaluation of conductance-based models)

生物物理神经元模型的一个特点是，它们很少受到任何初始“自上而下”计算要求的激发；相反，它们试图以受观察生物约束的方式复制膜现象的动态行为。在捕捉这些限制的过程中，一个主要问题是需要确定最能确定生物数据的多个相互作用模型参数的特定配置。Wood 等人(2004)最近解决了这个问题，他们开发了一种确定性参数搜索技术来确定生物物理模型中的最大电导，并将此技术应用于模拟中多棘神经元的膜特性(Wood 等人,2004)。生物数据的精确复制可以看作是一个先验步骤，以便随后对模型中的各个特征进行选择性地调制，以确定对整体现象的相对贡献(参见已经描述的 STF 案例)。因此，这类基于电导的模型尤其有用：

①为已知的膜行为提供可以用实践检验的(机械论)解释；

②通过使用虚拟操作，假设到目前为止，还没有在体内或体外观察到，但可以通过实验测试进一步检验；

③运用①和②中发现的机制，发展有关神经功能的计算假设。

#### 4.1.4.模型间的关系

所有层次的模型应该被评估，主要是通过它们解释 *BG* 所观察到的生理和行为特性的能力。然而，随着领域的发展，模型应该显示出不同层次的一致性。例如，高层次模型的组成部分应该是对低层次模型的合理、现实的抽象。因此，*BG* 神经元的生物物理模型的另一个重要用途将是提供和优化用于更高微电路和系统级模型的“简化”神经元元件的操作特性。虽然这样的分析将导致更高层次的现实主义，但在更高层次的抽象中出现的计算假设将为进一步寻找合理的计算机制提供有用的起点。最后，重要的是要注意到，不同层次可用的信息表示方式可能使某个层次的机制不能有效地转换为另一层次的机制；例如，不能指望使用放电率的模型神经元支持依赖于脉冲编码的机制。基本上，这意味着最高层次的精确模型可能需要显式地包含相对较低层次元素。

#### 4.1.5.分析

在本文中，我们试图展示 Marr(Marr,1982;Marr 和 Poggio,1977)所引入的差异，如何为理解计算模型在 *BG* 所有结构层次上测试的目标、限制和策略提供了有用的基础(框注 1)。此外，在系统层面上，基于从透明程序到神经基底的两阶段映射的模型与基于更直接解剖和生理数据挖掘算法开发的模型之间有着重要的区别。将算法映射到生物基底上的过程，可以从数学科学的角度，用强大的、经过反复试验的视角，阐明我们对 *BG* 的理解。与此同时，这种方法可能导致对这些程序的重新思考或改进，从而为工程学和神经科学带来好处。然而，*机制挖掘*方法似乎更容易适用于许多结构层次，并且通过生成与神经生物学约束密切相关的模型，将有助于解开神经组织中包含的功能机制。这种方法还提供了一种有趣的可能性，即发现迄今未知的程序，最终可能导致所有智能系统科学的进步。最后，我们希望，对各种建模方法的更好理解，将鼓励经常持怀疑态度的实验神经科学家与他们的计算同行合作，更好地解释描述 *BG* 生物学特征的日益复杂的数据集。

#### 4.2.BG 中的动作选择的计算模型：I.一个新的功能解剖

脊椎动物神经网络的一个重要任务是解决大脑内物理上分隔但在行为表达上存在竞争的功能单元之间的冲突。非正式地说，我们如何解决“下一步该做什么”问题。如果有几个这样的单元在争夺同一种资源，例如，与涉及进食和饮水的神经网络竞争，这两个系统都需要

使用口腔肌肉，系统竞争对这种情况就尤为敏感。在其他情况下，可以同时进行多个动作，例如，当动物同时行走和咀嚼时。竞争在行为表达上也可能出现*间接*模式，例如在试图控制*认知*结构资源的系统之间。因此，一般来说，我们假设大脑正在处理大量的感觉和*认知*流或通道，每一个都可能要求采取一些行动。动物需要的任务是抑制大多数这些请求，同时在某些情况下只允许有限数量的表达。这种*动作选择*问题显然对我们理解人类行为至关重要(Duncan,1995)，但这也出现在对人类行为的行为学分析中。动物行为(McFarland,1989)，(通常被称为“*行为转换(Behavioural switching)*”或“*决策(Decision-making)*”问题)*动作选择*是基于行为的机器人控制的核心问题(Arkin,1995)。我们最近认为，在对*动作选择*问题的分析中，在不考虑道德约束的背景下(Prescott 等人,1999)这些问题有许多共同的特征。然而，从生物学的角度，我们可以提出以下问题：①进行*动作选择*的神经基质是什么？②这种*基质*功能性操作是如何进行的？

为了识别用于*动作选择*的神经基质，了解并认识到空间分离的功能系统所要求的动作如何在原则上被选择的一般计算问题是很有帮助的。从结构的角度来看，有两种主要的可能性(Snaith 和 Holland,1990)。首先，竞争模块之间可以*直接*联系，通过模块间处理解决任何冲突。或者，竞争者可以将请求传达给一个中央仲裁机制，该机制随后从中选择行为表达的子集。最近我们回顾了这些可能性(Redgrave 等人,1999)，并认为，就连接性和代谢效率而言，有效的*动作选择*有利于集中的开关单元。此外，我们还提出，在脊椎动物大脑中，BG(一组功能相关的中枢脑结构)非常适合发挥这种作用。BG参与了广泛的过程，包括感知和*认知*(Brown 等人,1997)、*工作记忆*(Levy 等人,1997)以及运动功能的许多方面(Marsden 和 Obeso,1994;Graybiel,1995)。然而，文献中反复出现的一个主题(例如 Cools,1980；Mink,1996;Kropotov 和 Etlinger,1999)是，BG与*动作选择*问题有关。我们最近的工作(Prescott 等人,1999;Redgrave 等人,1999)提出了选择的概念，将其作为BG功能的一个主要统一假设，展示了它与已知解剖和生理学的关系，以及它如何满足几个高级计算要求。

在本文和 Gurney 等人(2001)写的论文中，从此被称为 GPR2，我们解决了上面提出的第二个问题：用于选择作用的神经生物基质是如何运作的？我们的研究采用了BG固有加工的生物学模型。在开发模型时，我们有四个主要目标：首先，定量地阐明“选择”的概念，并说明它如何应用于BG模型。第二，根据一套专门用于选择的神经机制重新解释BG解剖结构。第三，定量解释不同BG选择的功能贡献。第四，考虑到多巴胺在调节BG功能中的重要性(见 Mink,1996)，对多巴胺在选择环境中的作用进行分析。

本文主要讨论这些目标中的前两个；后两个在 GPR2 中被考虑。定量分析允许与实验数据相关联，并用于支持我们对BG新功能结构的主张。然而，导致这一功能方案的推理思路并不依赖于详细的机械特性细节，因此它是基于自身的优点。

完成BG解剖生理和功能的回顾后，第2节表明了BG固有的处理方式可以根据信号选择的更容易处理的问题来解决，并继续将这一过程正式化。接下来，我们用以下观察为基础的方法来探索BG的功能结构：鉴于BG的选择性功能在神经回路中被实例化，它们使用的机制必须属于这些系统可用的全部功能。相反，对这组机制的检查将有助于指导对BG解剖结构可能的功能配置的解释。因此，第3节介绍了一般的神经网络*选择机制*，并且在第4节显示了这些*选择机制*是如何映射到BG解剖结构中的。其结果是将BG的功能解剖结构重新解释为两种新的途径：一种名义上专门用于选择过程本身，另一种则用于调节或调节初级选择途径的操作。

#### 4.2.1. BG 解剖结构和生理机理

构成我们模型基础的解剖学和生理学数据是已知的,并在最近的几篇评论中进行了描述(Gerfen 和 Wilson,1996;Mink,1996;Wickens,1997;Redgrave 等人,1999)。灵长类 BG 亚核及其内在联系如图 1a 所示。灵长类 BG 的主要成分是前脑的纹状体、苍白球(GP)和下丘脑核(STN),和中脑的黑质(SN)。苍白球包括两个亚区,分别是内、外节段(GPi 和 GPe),而黑质包含不同的区域,分别称为致密部(SNc)和网状部(SNr)。类似的结构(尽管通常有不同的名称)在其他脊椎动物类的神经系统被发现(Medina 和 Reiner,1995)。

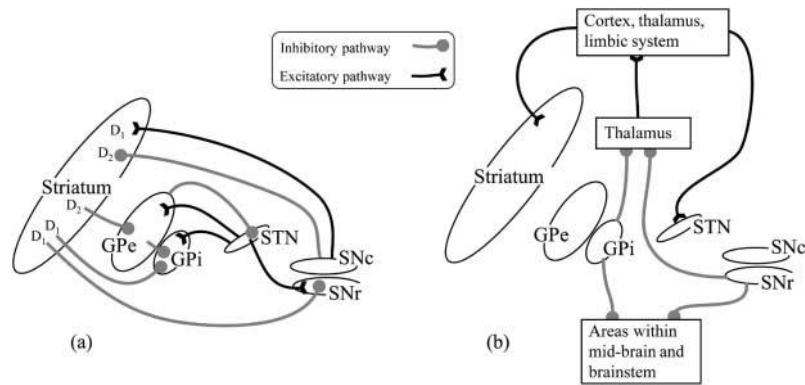


图 18: 基底节解剖结构。兴奋性和抑制性通路分别用实线和灰色线表示: 内部通路, 包括黑质致密部纹状体的多巴胺能神经支配(SNc)。其他核的关键: GPe 和 GPi-苍白球的外部 and 内部片段; STN-下丘脑核; SNr-黑质网状部。D1 和 D2 指多巴胺受体类型; b 显示主要输入和输出以及丘脑中心作用的外部通路。

BG 内的主要激发源是 STN, 而与大多数其他核之间的连接是抑制性的。SNc 向纹状体提供多巴胺能输入, 根据突触后细胞的受体类型判断, 可能会产生抑制或兴奋作用。因此, 当 D1 型受体(Akkal 等人,1996)介导时, 多巴胺能传递主要是兴奋性的。但当激活 D2 受体(Gerfen 等人,1990;Harsing 和 Zigmond,1997)时, 它可以发挥抑制作用。纹状体中绝大多数(90%)的神经元是氨基丁酸能的中等多棘细胞。这些细胞从各种各样的大脑结构接收主要的兴奋性输入到 BG, 也从 SNc 接收多巴胺能输入。它们为 GPi、GPe 和 SNr 提供相位抑制输出。SNc 神经支配的差异效应通过纹状体的传出通过 BG 进行传播。因此, GPi/SNr 仅从与 D1 型受体相关的中等多棘细胞接收投射, 而 GPe 主要从具有 D2 型受体的细胞(Gerfen 等人,1990)接收投射。对 GPi/SNr 的投射也与 GPe 的侧支相关, 但是, 这些细胞往往比那些发源于具有 D2 受体的细胞的分支要轻很多, 而且体积也更小(Kawaguchi 等人,1990)。我们以此作为证据, 具有 D1 型受体的纹状体细胞对 GPe 的投射比其 D2 型受体弱得多。

在默认状态下, 中等多棘的神经元细胞基本上是静默的, 对低水平的输入没有反应。然而, 在接收大量的协调兴奋性输入时, 这些细胞产生一个显著的输出, 其量值随后可能会受到低水平输入的影响, 而低水平输入单独出现时无效。对于这两种操作模式, 分别使用术语“下降状态(Downstate)”和“上升状态(Upstate)”来描述两种操作模式(Wilson,1995)。

BG 的解剖结构背景如图 1b 所示。其主要输入核(纹状体和 STN)几乎从整个大脑皮层、丘脑的中线和丘脑板内核(Intralaminar nuclei of the thalamus)以及边缘系统(主要是杏仁核和海马)接收传入。因此, BG 能够处理各种各样的信息, 这突出了其作为中央选择机制的可能意义。此外, 这种 BG 是通过一系列传入的平行处理流或通道产生输入(Hoover 和 Strick,1993;Alexander 等人 1986)并且, 在涉及运动领域的地方, 它展示了一个躯体组织行为(Delong 等人,1983;Flaherty 和 Graybiel,1993)。主要的 BG 输出核是 GPi 和 SNr, 它们为腹侧丘脑(反过来投射回大脑皮层)、上丘和其他脑干区域的功能相关区域提供广泛分支的氨基丁酸能的效应。

#### 4.2.2. 基底节功能机制

虽然人们对 *BG* 的原子和生理学了解很多,但对功能结构及其内在计算原理的了解却相对较少。这些知识对于理解 *BG* 如何介导选择性功能至关重要。总的来说, *BG* 通过释放抑制而起作用(Chevalier 等人,1985;Deniau 和 Chevalier,1985)。 *BG* 输出核的默认功能是对靶结构施加广泛的强直抑制控制。然后 *BG* 能够通过解除其相关目标结构的抑制来促进作用,同时保持对其他神经节的抑制控制。

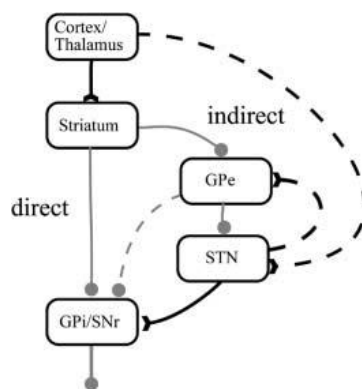


图 19: 基底节的主要功能解释。兴奋性和抑制性通路分别用实线和灰色线表示。此外,在该模型中,受到较少重视的路径显示为虚线。

*BG* 固有的功能结构的主要模型最初由 Albin(1989)等人提出。如图 2 所示。输出核的纹状体调制有两种途径:(1)一种“直接途径”是纹状体的氨基丁酸能的输出能够直接抑制 GPi 和 SNr 中的元素;(2)一种“间接途径”,包括从 GPe 到 STN 再到 GPi/SNr 的双突触途径,它能够对输出核产生整体兴奋性影响。在过去的十年中,这种模式有助于指导研究,但也出现了一些缺点。首先,以上提出的想法仍需作为一个完整的计算模型进行开发,使其能够运行的几个方面仍然不清楚,例如,包括直接和间接路径之间的精确交互方式(Alexander 和 Crutcher,1990)。第二,目前有几种成熟的途径,例如 STN 的外部输入、GPi/SNr 的 GPe 神经支配和 GPe 的 STN 激发(见图 1),尽管 Smith 等人(1998)最近承认这些途径的重要性,但这些途径尚未完全整合到模型中。第三,模型显然无法容纳某些功能数据;例如, GPe 在帕金森病动物中的作用(Chesselet 和 Delfs,1996;Parent 和 Cicchetti,1998)。模型的作者们对这些缺点进行了批判性的评估,他们期待着“找出模型的致命不足之处及其找出完善这些模型的方法”(Albin 等人,1995)。

Mink 和 Hach(1993)提出了 *BG* 固有电路的一个子集的替代解释,该子集原则上可以调节选择功能。他们的组织结构强调 STN 是一个主要的输入核,并依赖于 STN 对输出核 GPi/SNr 的扩散性质(Parent 和 Hazrati,1993,1995)。目前的研究将此方案的修改作为其核心组成部分之一。然而,我们继续定量地探讨其功能,提出 GPe 的新作用,并在 *BG* 功能选择假设的背景下,论证纹状体的多巴胺能调节是如何运作的。

#### 4.2.3. 选择过程

相互竞争的候选行为被认为是分布在整个中枢神经系统的网络中的代表。如 Koehlin 和 Burnod(1996)提出的,给定动作的神经表示的整体活动水平可决定其显著性或执行倾向。我们不直接处理每个动作的神经编码,而是提出 *BG* 与它们的(标量值)显著性有关。因此,根据显著性的“共同流通”在 *BG* 的输入阶段表示动作(机械细节在 4.2 节中进一步讨论)。在 *BG* 输出端,动作是通过释放抑制来调节的,发生的程度定义了另一个标量值信号。因此,在只处理 *BG* 固有的模型中,我们能够将“动作选择”的过程重新定义为“信号选择”的过程,



其中纹状体大信号输入和 STN 选择为 GPI/SNr 低信号输出。就我们的目的而言，一个行为被定义为由 BG 内的特定细胞群决定的行为表达。这些群体就是我们所说的“通道”，可能与任何事物有关，从基本的运动行为到广泛的行为策略(Redgrave 等人,1999)。确定 BG 介导的动作序列，以及低水平动作的选择如何与高水平动作的选择相关，仍是进一步工作的主题。

#### 4.2.3.1.定义信号选择

虽然减少信号选择是一个重要的概念简化，但必须明确这意味着什么。信号选择可以看作是一种输入、输出转换，它将一组输入  $x$  分别划分为“选定”和“未选定”信号的输出集  $S$  和  $\bar{S}$ 。 $S$  中元素选择的标准可能是转换信号大于某个阈值  $\theta$ ，并且所有小于该阈值的转换信号都在  $\bar{S}$  中。但是，每类信号的值可能任意接近  $\theta$ ，因此选择可能容易受到噪声干扰(Horowitz 和 Hill,1989)。

因此，我们改进了定义，并将转换后的信号间隔划分为三个间隔，由两个选择阈值  $\theta_1$  和  $\theta_2$  决定，允许转换后的信号处于  $S$ 、 $\bar{S}$  或不确定集  $Y_0$  中，其中  $y \in Y_0 \Rightarrow \theta_1 < y < \theta_2$ 。此外，我们还需要考虑所选信号被认为是上间隔或下间隔信号的可能性，因此，如果  $y \in S \Rightarrow y \geq \theta_2$  我们归类于大信号编码，如果  $y \in S \Rightarrow y \leq \theta_1$ ，则归类于小信号编码。为了进一步进行，可以方便地给输入和输出信号标记，因此，选择转换由一组映射  $x_i \rightarrow y_i$  定义，其中  $i=1,2,\dots,n$ ， $n=|X|$ ，且  $x_i \in X$ ， $y_i \in S \cup Y_0 \cup \bar{S}$ 。选择还意味着保留了排序关系。因此，如果所有的输入和输出都包含在某个区间  $I$  中，那么一个选择机制(selection mechanism)现在被定义为映射  $G: I^n \rightarrow I^n$ ，以及两个阈值  $\theta_1$  和  $\theta_2$ ，其中  $G(x_1, x_2, \dots, x_n) = (y_1, y_2, \dots, y_n)$ ，并且它遵循以下的顺序关系

$$\begin{aligned} x_i \leq y_i \Rightarrow y_i \leq y_j : \text{large-signal selection} \\ x_i \leq y_i \Rightarrow y_i \geq y_j : \text{small-signal selection} \end{aligned} \quad (1)$$

在这种情况下，我们专门研究了映射到  $S$  中的较大输入。

显然，任何保序函数  $G$  都是选择机制的正式候选者，它仍然需要建立标准，使其在选择背景条件中真正有用。应用于  $X$  的机制的决定性  $D(x)$  由

$$D(x) = 1 - \frac{|Y_0|}{|X|} \quad (2)$$

给出，其中  $0 \leq D(x) \leq 1$ 。 $D(x)=1$  表示对  $X$  的一个极其确定的或“干净利落”的选择，其中所有信号都转换为  $S$  或  $\bar{S}$  中的一个。 $D(x)=0$  表示没有选择，所有信号都在不确定集  $Y_0$  中。接下来，我们指定选择信号的  $\phi(x) = |S|/|X|$  部分，并将其称为选择混乱(selection promiscuity)；或者，我们可以讨论选择性(selectivity)  $1-\phi(x)$ 。一个混杂的机制是对于所有的  $x$ ， $\phi(x)$  都是大的。如果对所有  $\phi(x)$  只选择一个输入( $\phi(x)=1/n$ )，我们就说这个机制实例化了硬切换(hard switching)。然而，如果对于某些  $x$ ， $\phi(x) > 1/n$ ，则我们说这个机制允许软切换(soft switching)。

现在考虑的是小信号编码机制。如果静态状态( $x_i = 0, \forall i$ )满足  $y_i \in \bar{S}$ ， $\forall i$ ，那么，为了选择被输入，则对于一些  $x$  我们必须有  $dy_i/dx_i < 0$ 。另外，如果  $dy_i/dx_i > 0, i \neq j$ ，那么我们就说  $i$  和  $j$  之间存在竞争(competition)。很明显，竞争会提高选择性，因为以前选择的输入可能会以选择新输入为代价而被取消选择。

#### 4.2.4.神经网络选择

在一个神经网络中的信号选择可以通过将每个节点与由两个具有相反影响极性的同心区组成的突触接触的空间模式相关联来完成。该方案可以使用层内链路、横向循环链路或前馈网络进行配置。如果中枢突触区是兴奋性的，那么我们就得到一个中心外的环绕网络。事实证明，这些网络在视觉皮层的自组织特征检测器建模中很受欢迎，最显著的是使用它们的反复实例化(Von der Malsburg,1973)，但也使用它们的前馈形式(Grossberg,1976)。在生物学

的背景下，递归网络必然局限于一个特殊的亚核。虽然这些可能在  $BG$  计算中很重要，但我们主要对解释核间通路感兴趣，因此在这一阶段，我们将重点放在前馈网络上。

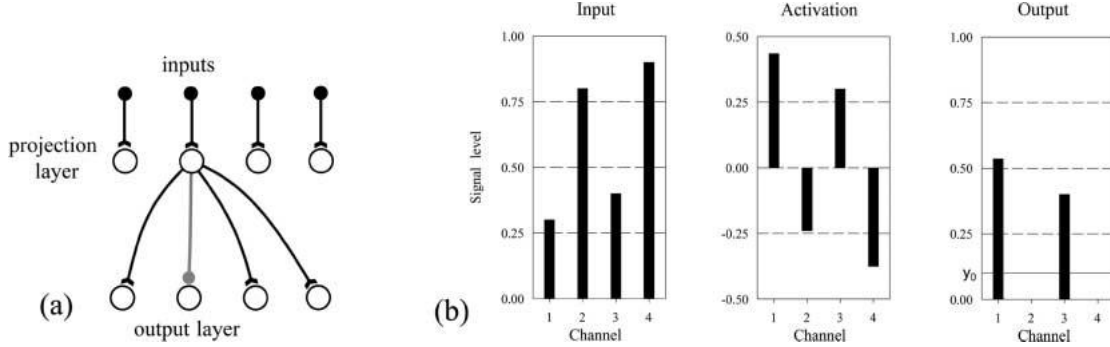


图 20: 非中心环绕神经网络: a 图具有四个通道的网络结构。兴奋链接和抑制链接分别用实线和灰色线表示; b 图对(a)网络模拟, 显示了具有最大输入的通道 2 和 4 的输出的静态。兴奋和抑制量分别为 0.45 和 -1.35。

为了支持中心外环绕结构, 需要在  $BG$  中有一个扩散抑制源, 对此没有证据能证实。另一方面,  $STN$  提供的扩散激发可能有助于非中心环绕方案。图 3a 显示了一个前馈网络的示例, 该网络具有偏心环绕连接。假设以标准化速率编码解释输入  $x_i$ , 其中  $0 \leq x_i \leq 1, \forall i$ 。每个输入节点将其值  $x_i$  发送给所有输出神经元, 但仅将抑制发送给一个分配了相同指标  $i$  的输出神经元。尽管目前还没有对  $BG$  电路的解释, 我们将这个输入和输出节点组合称为一个通道。输出层中的节点是简单的半线性单元, 其激活  $a_i$  是单元输入的加权和。因此, 如果  $w^+$ ,  $w^-$  分别是兴奋性和抑制性权重的大小, 那么有

$$a_i = -w^- x_i + w^+ \sum_{j \neq i} x_j \quad (3)$$

对于两个任意的通道  $p$  和  $k$ , 满足  $a_p - a_k = (w^+ + w^-)(x_k - x_p)$ 。也就是说, 如果是  $x_k \geq x_p$ , 则是  $a_k \leq a_p$ 。给定输出  $y_i$  和激活  $a_i$  之间的单调递增关系, 因此输入和输出之间有一个保序映射, 网络需要实例化一个小信号编码选择机制。然而, 为了使网络发挥一种有用的功能(具有非零的决断性和混乱性), 抑制和兴奋之间的平衡必须谨慎选择。为了研究这一点, 令  $w^+ = \delta w^-$ , 设  $\langle x \rangle_{/i}$  是输入的平均值, 不包括  $i$ 。然后可以将激活(3)的表达式重写为

$$a_i = w^- [\delta(n-1)\langle x \rangle_{/i} - x_i] \quad (4)$$

其中  $n$  是通道数。现在假设  $\delta$  是固定的, 那么  $a_i$  是不受  $n$  的约束的, 除非  $\langle x \rangle_{/i}$  是  $n$  和  $\langle x \rangle_{/i} = O(1/n)$  的函数, 其中,  $f(n) = O(h(n))$  表示当  $n \rightarrow \infty$  时, 以  $f(n)/h(n)$  为界。因此, 在输出  $y$  和激活之间给定一个单调压缩关系  $y = y(a)$ , 除非以这种方式调节输入, 否则所有输出最终都将达到(或接近)其最大值。如果输入条件可以放宽, 替代的是  $\delta = O(1/n)$ 。我们将激励和抑制之间适当比率的输入独立设置称为容量缩放。  $\delta = O(1/n)$  为“硬连接”网络提供一个突触机制。

选择机制描述是通过提供两个阈值  $\theta_1$  和  $\theta_2$  和一个规范函数  $y(a)$  完成的。考虑分段线性压缩映射

$$y_i \begin{cases} 0 & : a < \varepsilon \\ m(a - \varepsilon) & : \varepsilon \leq a \leq 1/m + \varepsilon \\ 1 & : a > 1/m + \varepsilon \end{cases} \quad (5)$$

由于  $y_i$  标准化,  $0 \leq \theta_1 < \theta_2 \leq 1$ 。在静态条件下,  $a_i = 0, \forall i$ , 如果为  $\varepsilon \geq 0$ , 则  $y_i = 0, \forall i$ 。因为选择标准是  $y_i \leq \theta_1$ , 这意味着当没有网络输入时, 所有通道都被选择。这显然是不可接受的, 因此我们将其设为  $\varepsilon < 0$ , 这给出了静默或“强直”输出值  $y_0 > 0$ , 并确保  $0 \leq \theta_1 < \theta_2 \leq y_0$ 。最后注意到, 由于

$da_i/dx_j = w^+ > 0, i \neq j$ ，存在通道间竞争的潜力。所以，如果  $\varepsilon \leq a \leq 1/m + \varepsilon$  则  $dy_i/dx_j > 0$ 。

图 3b 显示了上述类型的 4 通道网络，其中  $w^- = 1.35$ ， $w^+ = 0.35$ ， $\varepsilon = -0.1$  和  $m = 1$ 。输入较小的通道(通道 1 和 3)的输出大于强直的，而具有最大输入(通道 2 和 4)的通道已将其输出抑制到零。这导致选择通道 2 和 4，而其他两个通道则被果断地取消选择。可能选择多个通道意味着前馈网络可以实现软交换。总之，只要具有容量缩放和非零强直输出，就可以在前馈、非中心环绕网络中实现小信号编码选择机制。

#### 4.2.5.模型建立

##### 4.2.5.1.基本假设

在没有明确确定的 BG 功能原则的情况下，首先采用系统级方法是较为合适的 (Churchland 和 Sejnowski,1994)。在这个“粗纹理”或“低倍率”的观点中，我们对细胞的整体行为以及每个核内的整个亚群相互作用的方式产生研究兴趣。此外，我们认为，在研究涉及更广泛的解剖学背景模型(包括传入和传出结构，如皮质、丘脑和上丘)之前，了解基底节内的工作机制是有帮助的。

我们的一般功能假设是，一个活跃的竞争通道将刺激纹状体内的细胞群，从而抑制输出核(GPi/SNr)中的相应细胞群。这抑制了这些细胞对其靶结构施加的强直抑制控制，使其变得活跃并表达与之相关的作用(Chevalier 等人,1985;Deniau 和 Chevalier,1985)。

##### 4.2.5.2.功能解剖学的重新解释

再次考虑图 3 的网络。如果我们将输入与一组离散动作处理通道上的显著值相关联，则可以将其解释为动作选择机制。通道在物理上是不同的，这一假设与 BG 可能划分为不相交的并行处理流(Alexander 等人,1986;Hoover 和 strack,1993)和适当情况下的躯体组织结构相一致(Delong 等人,1983;Flaherty 和 Graybiel,1993)。显然，这种假设在某些情况下可能会不成立，但我们认为，对于许多选择的通道集来说，这是一个很好的近似假设。每个与单个通道相关的 BG 亚核中的神经元群被视为是相同的，因此我们可以对每种情况只研究其中的一个神经元群代表。这是一种在系统层面进行研究的标准方法，并已被采用，例如，在模拟皮层的研究中就采用了类似的方法(Douglas 等人,1989)。

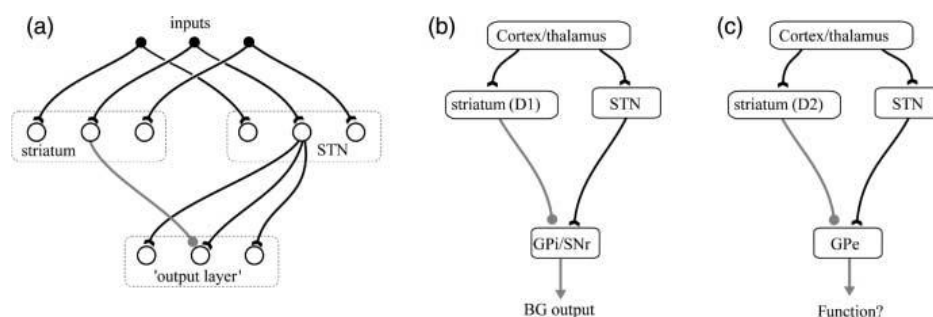


图 21: 新功能体系结构的组成部分。将图 3 中的网络映射到 BG 解剖结构上: (a) 分离输入核进行激发和抑制，但迄今尚未确定输出核; (b) 在(a)中系统的一个特定实例，“选择路径”包括 BG 输出核; (c) 在(a)中的另一个实例构成“控制路径”。以及它的“输出”，GPe，它向其他 BG 核提供控制信号。

虽然图 3 中的网络在原则上具有执行动作选择的能力，但不可能用 BG 电路的任何部分对应地定义网络，因为它在生物学上是解释不通的。因此，投影层同时受到抑制和激发，而扩散兴奋区具有高度局部不连续性。图 4a 显示了克服这些缺陷的更现实的架构实现。该实现具有单独的兴奋和抑制输入结构和均匀分布的兴奋传出。使用适当的权重，可以使该网络的行为与图 3 中的简单对应网络相同。此外，输入结构已被定义为 STN 和纹状体。

现在转到*显著性*输入的运用,我们提出中等多棘神经元的目的是提取其相关通道上的*显著性*动作请求。这可能发生在这些细胞的广泛树突状接触中,这与广泛分布的输入活动的处理模式有关,并且在运动处理涉及的皮质-纹状体连接的收敛模式中很明显(Flaherty 和 Graybiel,1991;Graybiel 等人,1994)。这种机制的细节存在于亚细胞层面,也将涉及对 BG 外的控制结构中的动作的神经编码的理解。考虑到我们在系统层面上工作并处理 BG 内的处理,我们假设在树突棘上提取了*显著性*,并且模型纹状体神经元的总输入仅仅是其当前处理的动作的*显著性*。

STN 可能缺乏成熟的中等多棘细胞提取*显著性*的装置,但这对 STN 在模型中的总体假定作用并不重要。我们假设 STN 将扩散神经支配发送到其目标核(Parent 和 Hazrati 1993、1995),使这些核中的每个种群获得 STN 总输出。因此,单个通道的贡献不必维持在 STN 靶点上,并且 STN 提供与传入 BG 的总活动成比例的兴奋是足够的。尽管广泛的轴突侧支到靶点是刺激再分配的一种可能机制(也是我们在这里采用的机制),它也可以用 STN 中广泛的连通性来实例化(Kita 等人,1983;Afsharpour,1985),这样一个集中的刺激可以激发多个传出。这与观察到的结果一致,在刺激感觉运动皮层(Fujimoto 和 Kita,1992)的单个部位后,STN 产生广泛的兴奋性影响,这一过程由 Gillies 和 Willshaw(1998)模拟。此外,我们提出的体系结构并不依赖于有一个单一的相当于整个 GPi/SNr 的 STN 投影场(PF)。输出核可以由多个 STN 的 PFs 服务,这些 PFs 实例化体系结构的多个副本;主要前提是 STN 通道的 PFs 比纹状体中的对应物更分散。

最后,我们的模型假设纹状体和 STN 都是由一个共同输入的分支侧支所支配。虽然有证据表明这确实发生了(Feger 等人,1991),但它可能并不是对所有情况都是成立的。一个较弱但功能等效的假设是纹状体和 STN 接收到高度相关的输入或相同信号的“拷贝”。

虽然图 4 中的总体方案现在看来是合理的,我们重新解释功能解剖学的关键是认识到这种结构在 BG 内有两个实例。其中一个类似于 Mink 和 Thach (1993)提出的建议,是将 GPi/SNr 用作其“输出层”。然而,还有另一种方法运用 GPe 作为输出。这种区别在根据中棘细胞的多巴胺能调节模式分化中棘细胞的投射靶点方面具有生理基础(图 1a)。因此,GPi/SNr 主要由与 D1 型受体相关的纹状体细胞支配,而 GPe 主要从具有 D2 型受体的细胞接收投射。图 4a 和 c 分别显示了由此假设分别产生的两个“路径”。由于 GPi/SNr 提供 BG 输出,因此可以合理地假设图 4b 所示的系统本身就是*选择机制*,因此我们将其指定为*选择路径(selection pathway)*。

对于另一个途径(图 4c),目前尚不清楚 GPe 在什么意义上是一个“输出”核,因为其输出仅限于其他 BG 核。假设该子系统通过 GPe 提供的信号形成一个控制通路,其功能是调节主*选择机制*的特性,从而解决这一问题。Parent 和 Hazrati(1995)曾提到 GPe 在 BG 内发挥控制作用的一般观点。

为了帮助发现控制路径的精确作用,我们可以进行以下定性观察。首先,GPe 对 STN 的抑制应导致选择途径中 GPi/SNr 的激发降低。在 GPR2 中,我们证明这种抑制性控制与第 3 节中描述的容量缩放要求完全匹配。第二,两种途径中的多巴胺能调节协同作用。根据第 1.1 节的讨论,我们假设 D1 和 D2 受体分别与兴奋性和抑制性多巴胺能调节相关。提高多巴胺水平将增加选择通路中纹状体对 GPi/SNr 的抑制强度,降低控制通路中纹状体对 GPe 的抑制。这反过来又导致 GPi/SNr 输出降低,既*直接*通过 GPe→GPi/SNr 途径,也*间接*通过减少 STN 激发。净效应是两种途径都能抑制基底节输出核。

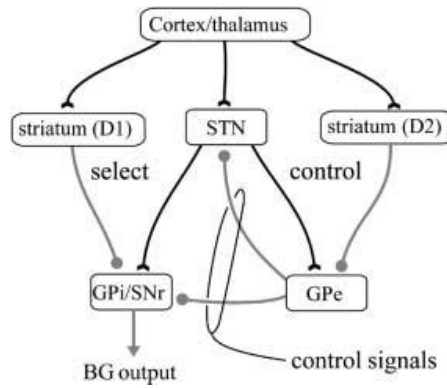


图 22: 新功能架构的最终形式。显示选择和控制路径组合的完整模型。

结合这两个子系统，并结合提供控制信号的 GPe 输出，形成了构成新模型基础的功能架构(图 5)。新的架构应该与在图 2 中所示的“直接/间接”路径解释进行比较(Albin 等人,1989)。这两种方案在结构上完全不同。旧的“直接”通路由纹状体和 GPi/SNr 组成。与此相反，新的选择途径将 STN 作为输入核，其纹状体的作用包括使用 D1 多巴胺受体的中等多棘细胞群。旧的“间接”通路由纹状体、GPe-STN 和 GPi/SNr 组成。新的控制通路不包含 GPi/SNr，其纹状体的作用包括使用 D2 多巴胺受体的中等多棘细胞群。此外，两种新途径(选择和控制的)的功能作用与旧模型的功能作用有很大不同。

#### 4.2.5.3.局部纹状体网络

从功能角度来看，纹状体被证明支持通过抑制介导的局部处理(Brown 和 Sharp,1995)，尽管其解剖学基础尚不清楚(Jaeger 等人,1994)，但纹状体功能的几种模型(Wickens 等人,1991;Alexander 和 Wickens,1993;Houk 等人,1995)假设纹状体支持这种处理方式。新功能体系结构的运行并不依赖于纹状体内抑制的存在，但是，只要满足以下两个条件，模型就能够适应这种情况。首先，通过每个局部网络中的最大反应神经元来提取显著性的梯度测量。在 GPR2 中，我们表明这适用于仅运用抑制作用的侧部递归网络。对于工程神经网络中流行的所谓“赢者通吃”(winner-take-all, WTA)网络来说，情况并非如此，因为这些网络会导致所有局部最大显著性强制最大输出。纹状体中 WAT 网似乎不太可能存在，因为它们需要一个兴奋的邻域，该邻域与中等多棘神经元表达的 GABA 抑制不一致。第二个条件是两个纹状体子系统(由其多巴胺受体类型定义)包含高度显著输入类似表示。虽然每个子系统内局部递归网络的假设与同一受体类型相关的纹状体细胞相互支配的证据一致，但 D1 和 D2 子系统也可能相互作用(Smith 等人,1998)。然而，我们假设这种耦合并不禁止在两个系统中同样忠实地表示高度显著的输入。这可能是由于耦合很弱，或者是因为它被配置为协作处理，从而促进了每个群体中显著性的等效表示。我们不认为局部递归神经网络的最大显著输入与随机从任一人群中选择的单一多巴胺受体类型的细胞有关。

#### 4.2.5.4.硬切换和软切换相结合

计算上，局部纹状体处理可能起到至关重要的作用。在纹状体的体细胞组织下，每个递归网络可能处理与同一体细胞相关的动作请求。然后，鉴于循环网络仅能选择一个信道(因此执行硬交换；见第 2.1 节)，每个网络的理想配置是解决寻求控制单个电机资源的电机程序之间的资源冲突。然而，由于局部经常性竞争而选择的通道可能需要进一步选择。如果没有发生这种情况，可能会出现不良和不适当的动作组合，例如：注意缺陷多动障碍或图雷特综合征，两者都与 BG 功能紊乱有关(Brito,1997;Swanson 等人,1998)。另一方面，选择多个通道的可能性是观察的一个要求，即有可能参与多种明显可分离的行为(例如步行和咀嚼)。

软切换的后续阶段可以满足这些需求,我们建议在新功能架构的*选择路径*中找到软切换的实现。

#### 4.2.6.讨论

我们最初的计算前提是 *BG* 的主要作用之一,即介导*动作选择*。然后,我们认为,考虑到*显著性*的共同输入表征,*BG* 固有的选择功能可被视为信号选择之一。这个过程被量化为一个保序映射,它倾向于将信号分离成两个由间隔@分隔的集合。考虑到 *BG* 由执行信号选择的神经电路组成,我们随后探讨了可能执行这种计算的神经架构。这导致了*对 BG 功能解剖的两种途径的自然解释*。首先,*选择路径*,其中包含最小的神经基础设施小信号编码选择:一个*偏离中心*的环绕*前馈网络*。第二,在*选择路径*内调节处理的控制路径。定性的争论导致了控制通路可能的特定功能:限制 STN 刺激以促进容量缩放,增强多巴胺能调节选择。新的功能体系结构主要关注核间通路及其相互合作的选择过程。图 3b 中的模拟表明,这是软切换的形式。然而,这可以补充由局部纹状体电路执行的硬切换,该电路可以适应模型。我们建议 *BG* 运用两种*选择机制*:硬切换来调解争夺相同运动资源的通道之间的冲突,软切换来确定适当的运动活动组合。

##### 4.2.6.1.模型局限

我们的模型无论是在原子水平(连接方面)还是在生理水平(在每个途径的兴奋性和抑制性方面)都受到生物信息的启发。虽然我们已努力包括已知的主要途径,但有证据表明从 STN 到纹状体的微弱兴奋路径(Kita 和 Kitai,1987)和从 GPe 到纹状体的抑制路径(Staines 等人,1981)。虽然这些“小”通路都不包括在其他强调*直接和间接*通路的模型中,但是 GPe 的纹状体神经支配的存在支持了我们对 GPe 的主张,即 GPe 在 *BG* 内具有主要的控制功能。从 D1 型中棘细胞到 GPe 的相对较弱的投影(Kawaguchi 等人,1990)也被省略。其存在并不影响选择/控制通路的区分,但可能有助于精细化控制通路的主要功能能力。最后,虽然这与我们的系统级方法相一致,但还没有努力将已知存在于纹状体内的每个核内的神经元间的作用结合起来(Kawaguchi,1997)。

##### 4.2.6.2.与其他模型比较

与其他一些脑区(如视皮层)相比,*BG* 在模拟研究学术界的关注度相当低(最近模型的评论见 Beiser 等人,1997)。关注度低的原因也许是 *BG* 模型缺乏对 *BG* 执行的计算功能的普遍一致性。在这方面,我们最近对 *BG* 在*动作选择*中的作用的分析(Prescott 等人,1999;Redgrave 等人,1999)对于确定当前建模工作的方向至关重要。最新的 *BG* 计算模型(Berns 和 Sejnowski,1996)确实解决了*动作选择*的问题,也运用了广泛的 STN 激励。然而, Berns 和 Sejnowski 模型不同于目前的模型,因为 GPe 不是一个控制核,而是通过在“*间接路径*”中提供一个时间延迟在选择过程中起着中心作用。我们的模型也与以前的研究相对比,这些研究主要集中在基底节特定亚核内的内在加工。特别是纹状体在这方面引起了广泛关注(Wickens 等人,1991;Wilson,1995),而且还对 STN 进行了研究(Gillies,1996;Gillies 和 Willshaw,1998)。显然,这些模型将提供有用的信息,因为我们试图进一步的细节层次纳入我们自己的模型中。

##### 4.2.6.3.定量模型的需求

新结构的灵感来自第 3 节中研究的*前馈网络*。所提出的*选择机制*的核心,即*选择路径*,就是这样一个网络,但是仍然需要证明的是,整个模型体系结构以适当的选择性满足*选择机制*的要求。这些要求包括一个保序映射的实例化和容量缩放(GPi/SNr 通常满足对强制输出水平的需求)。目前尚不清楚的是,该架构在控制路径的影响下实现了一个保序映射(基于神经元输

出单调性的一般论据可能被调用,但这些论据将始终保持不确定)。因此, GPR2 进行了定量动力学分析,表明模型确实实例化了一个保序图,而控制通路的功能之一不是阻碍选择,而是提供不依赖于调节抑制/兴奋性突触重量比的自动能力缩放。定量模型也证实了多巴胺能对选择的协同调节,并且允许模型行为与生理和动物行为数据之间的比较。

#### 4.3.BG 动作选择的计算模型。II.行为分析与模拟

在 Gurney 等人(2001)的文章中,下文中把该文章记为 GPR1,我们提出了一种新的 BG 的功能结构,该结构是基于它们在行为控制方面执行选择功能的计算前提。结果表明,在这种功能解释下, BG 固有的处理可能被视为信号选择之一。这是正式定义,并基底节满足执行有用的信号选择的标准。本文定量地阐述了我们的模型,并通过分析论证了该模型的选择机制。进一步,我们在模拟中展示了该模型与某些神经生物学数据的一致性,并且该模型中的多巴胺能调节赋予其与多巴胺耗尽或兴奋时行为变化一致的选择特性。

本文的内容组织结构如下。第 2 节详细介绍了基于 GPR1 中描述的新的基底节功能结构的计算模型。然后,我们对平衡状态下单个核活动的模型进行分析求解(第 3.1 节),并继续使用这些解获得关于模型选择行为的一般结果(第 3.2 节)。分析完成后,检查多巴胺能调节的效果(第 3.3 节)。然后,主要的分析结果将在模拟(第 4 节)中得到证明,当出现大的、短时间的输入刺激时,该模拟也用于显示模型的动力学行为与相应的生理数据之间的相似性。

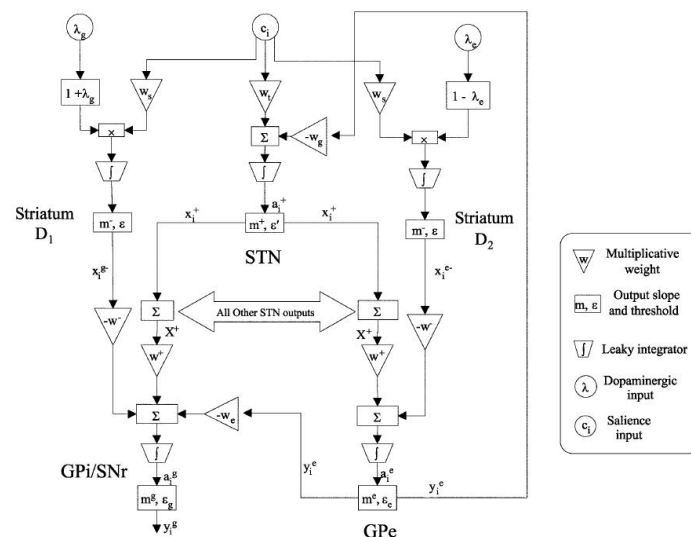


图 23: 模型详细信息和参数(仅一个通道)。

#### 4.3.1.量化新的功能架构

对于本节中的概要参考,图 1 显示了用每个模型参数注释的单个通道的结构。我们首先描述了一般的人工神经模型,然后依次对每个核进行描述。

##### 4.3.1.1.一般神经框架

该模型使用漏积分型(Arbib,1995)的特殊神经元。这是一个最简单的神经模型,它融合了动态膜电位的概念,同时避免了建立大量离子通道模型的需要(Yamada 等人,1989)。因此,它是神经回路生物学上可信模型的一个常见选择。漏积分器是由激活  $a$  的变化率定义的,这可以解释为轴突附近的膜电位。设  $u$  为传入输入所产生的总突触后电位,  $k$  为取决于细胞膜电容和电阻的速率常数,  $\tilde{a}$  是平衡活化,则有

$$\begin{aligned} \dot{a} &= -k(a - u) \\ \tilde{a} &= u \end{aligned} \quad (1)$$



式中  $\dot{a} \equiv da/dt$ 。神经元的输出  $y$  (假定与平均放电率相对应) 是  $a$  的单调递增函数。它将以 0 为下界且以一些最大值  $y_{\max}$  为上界, 这些最大值  $y_{\max}$  可以归一化为 1。我们采用分段线性输出函数的形式

$$y = \begin{cases} 0 & : a < \varepsilon \\ m(a - \varepsilon) & : a \leq \varepsilon \leq 1/m + \varepsilon \\ 1 & : a > 1/m + \varepsilon \end{cases} \quad (2)$$

$y$  的这种形式的选择(比 logistic 函数——激活函数更常见)是基于模型的平衡行为是可分析地可处理的这样一个事实驱动激活的; 模型的激活空间被划分为一组独立的区域, 这些区域的个体行为是线性的, 并且可以精确地确定。

在某些情况下, 可以确保  $y$  永远不会达到 1。然后将输出关系写为

$$y = m(a - \varepsilon)H(a - \varepsilon) \quad (3)$$

其中  $H(\cdot)$  是单位阶跃函数。假设  $y < 1$ , 用于模型的解析解, 并在仿真中得到证明。

#### 4.3.1.2. 纹状体

正如 GPR1 所讨论的, 纹状体模型允许局部复发性抑制的可能性, 但并不取决于其运作。允许这种情况发生的循环网络模型由(1)和(3)描述的一组漏积分器组成。在每个循环网络中, 每个节点通过一个权重为  $w^-$  的抑制链接彼此连接。如果输出关系中的非零斜率为  $m^-$ , 第  $i$  个节点  $x_i^-$  的平衡输出和输出阈值  $\varepsilon$ , 则网络平衡状态由以下一组耦合方程定义

$$\begin{aligned} \tilde{a}_i &= J_i - w^- \sum_{j \neq i} x_j^- \\ x_i^- &= m^- (\tilde{a}_i - \varepsilon) H(\tilde{a}_i - \varepsilon) \end{aligned} \quad (4)$$

现在令  $J_k = \max_i (J_i)$ 。如果  $w^- m^- \geq 1$ , 那么(4)的一个解是

$$x_i^- = m^- (J_i - \varepsilon) H(J_i - \varepsilon) \delta_k \quad (5)$$

其中  $\delta_k$  是克罗内克函数(又称克罗内克  $\delta$  函数, 是一个二元函数)。这个解可以很容易地通过直接替换来验证。在平衡状态下, 如果  $J_k > \varepsilon$  有一个单节点  $k$ , 它有非零输出  $x_k^- = m^- (J_k - \varepsilon)$ 。如果没有输入超过  $\varepsilon$ , 那么所有节点都有零输出。根据 GPR1 中拓展研究的选择理论, 如果  $x_k^-$  大于选择阈值  $\theta_2$ , 那么网络是完全决定性的 ( $D(x) = 1$ ), 与  $\phi(x) = 1/n$  个通道最小混杂, 并实现硬切换。

为了与通道显著性  $c_i$  作为输入的想法联系起来, 我们把  $J_i = w_s c_i$  输入, 其中  $w_s$  是对中等多棘神经元整合其输入时整体突触效率的测量。纹状体应该由(4)所定义类型的许多复发性网组成, 每个网处理多个通道。然而, (5)的解意味着, 只有每个网络中最大的显著性才是进一步处理的竞争者。假设有  $N$  个纹状体网络, 设  $c_{ri}$  为网络  $r$  的第  $i$  个通道的显著性, 令  $c_{rk(k)} = \max_i (c_{ri})$  且  $P = \{c_{rk(r)} : r = 1, \dots, N\}$ , 是一组潜在的活跃通道。现在用网络索引重新标记  $p$  的每个成员, 使每个局部循环网络  $r$  在平衡时服从(5)中形式与其最大突出通道(6)的关系。最大突出通道(6)如下:

$$x_r^- = m^- (w_s c_r - \varepsilon) H(c_r - \varepsilon/w_s) \quad r = 1, \dots, N \quad (6)$$

这里, 阶跃函数的编写方式强调了对显著性的依赖性。然后, 通过给  $\varepsilon$  指定一个正值, 非常简单地模拟中等多棘细胞的上/下状态行为, 从而使从“向下”状态到“向上”状态的转换符合最小正显著性的要求。在选择方面,  $\varepsilon > 0$  确保了一个主过滤机制, 强制所有  $x_i^- < \varepsilon$  都不被选择。

现在观察(6)中的关系也描述了单个神经元在孤立状态下的平衡输出。因此, 在局部复

发网松弛至平衡后获得的情况正式与仅存在  $N$  条纹状体通道从头获得的情况相同。

#### 4.3.1.3.多巴胺能调制

我们现在介绍中棘细胞活性的多巴胺调节。原则上，我们希望建立由 SNc 纹状体神经支配引起的多巴胺水平的动力学模型，尤其是与生物显著刺激开始相关的短潜伏期多巴胺信号(Schultz 和 Romo,1990)。然而，我们将自己局限于处理强直或“背景”多巴胺水平的简单案例，这些多巴胺水平可能会因 SNc 的病理学、动物实验的实验室操作或人类的主观臆断而改变。此外，我们假设这种对纹状体的多巴胺能输入具有足够的扩散性，以至于它在所有通道上基本上都是恒定的。

多巴胺能突触出现在中等多棘细胞的脊髓轴上(Bouyer 等人,1984)。这暗示着一个乘法(而不是加法)过程。然而，不管突触机制的细节如何，总是有可能通过在突触强度  $w_s$  中引入一个乘法因子，以现象学的方式对其进行建模。根据在 GPR1 中讨论的多巴胺能调节的观察到的特性，我们假设选择途径中的兴奋效应(通过 D1 受体)和控制途径中的抑制效应(通过 D2 受体)。因此，对于控制通路，传入突触强度  $w_s$  被修改为  $w_s(1-\lambda_c)$ ，其中  $\lambda_c$  参数化该通路的强直多巴胺调节程度且  $0 \leq \lambda_c \leq 1$ 。(6)中的阶跃函数现在变为  $H[c_i - \varepsilon/(w_s(1-\lambda_c))]$ 。为了简化符号并强调阶跃函数以多巴胺能依赖的方式决定向上状态的转变，我们将其写为  $H_i^\uparrow(\lambda_c)$ 。控制通路第  $i$  通道中纹状体中棘细胞的平衡输出  $x_i^{e-}$  现在为

$$x_i^{e-} = m^- [w_s(1-\lambda_c)c_i - \varepsilon] H_i^\uparrow(\lambda_c) \quad (7)$$

同样，对于选择路径， $w_s$  变为  $w_s(1+\lambda_g)$ ， $0 \leq \lambda_g \leq 1$  因此如果  $x_i^{e-}$  是该路径

$$x_i^{e-} = m^- [w_s(1+\lambda_g)c_i - \varepsilon] H_i^\uparrow(-\lambda_g) \quad (8)$$

中第  $i$  个通道上的平衡输出。其中  $H_i^\uparrow(-\lambda_g) = H[c_i - \varepsilon/w_s(1+\lambda_g)]$ 。

设  $P_c$ ， $P_g \subset P$  为在控制和选择路径中分别具有活动(非零)纹状体输出的一组通道。现在， $H_i^\uparrow(-\lambda_g) \geq H_i^\uparrow(\lambda_c)$ ，这样  $P_c \subseteq P_g$  与平等的关系只有在  $\lambda_c = \lambda_g = 0$  的情况下才能保持。因此，我们将参考一组活跃通道，这是一组与两个通道中非零纹状体输出相关的通道  $P^* = P_g$ 。

#### 4.3.1.4.丘脑底核(STN)

首先考虑 STN 可能包含离散信道群的情况。这与 STN(Delong 等人,1985)中的体细胞学证据一致，并将 STN 作为模型中的其他 BG 核置于同一基础上。对 STN 第  $i$  通道的平衡活化  $\tilde{a}_i$  有三个贡献。第一个贡献是第  $i$  通道 BG 的传入输入，缺乏成熟的中棘细胞模式匹配装置的 STN 可能无法真实地提取显著性  $c_i$ 。然而，任何信道上计算出的传入信号都可以用一个与信道相关的附加噪声项建模为显著值。然后，假设 STN 预测值由靶核、苍白球内段(GPi)和黑质网状部(SNr)以及苍白球外段(GPe)相加，并且假设噪声的预期值为零，那么对于实际大量的信道，噪声的总体贡献变得微不足道。因此，我们从后面的表达式中删除噪声项，以简化工作。我们现在分别考虑潜在的活动通道( $P$  中的通道)和非  $P$  中的通道。对于前者，我们将外部输入到基底节对 STN 活性的贡献称为  $w_i c_i$ ，其中  $w_i$  控制着对 STN 的传入突触强度，非  $P$  中的传入都是那些在局部纹状体竞争中未被选择的基底节传入。我们假设这些对 STN 活动的贡献可以被建模为一个平均值，加上加性噪声。再一次，假设噪声的期望值为零，且此非活跃输入的总和约为常量，我们可以用一个正的常数贡献  $\varepsilon'$  来表示这个。这种对 STN 皮质输入的低水平“背景”将有助于维持观察到的该核的强直输出(Delong 等人,1985;Wichmann 等人,1994)。当然，由于 STN 中细胞的固有膜特性，强直性水平也可能包含一个贡献。然而，这对整个核在功能架构层面上没有任何影响，因此我们将这些贡献归

因到  $\varepsilon$  中。最后，从 GPe 的第  $i$  个通道有一个分量  $w_g y_i^e$ ，其中  $y_i^e$  是其输出， $w_g$  是 GPe 和 STN 之间的突触权重。把这三个部分结合起来，我们就有了  $\tilde{a}_i^+ = w_i c_i + \varepsilon' - w_g y_i^e$ 。如果 STN 中的第  $i$  个通道的输出是  $x_i^+$ ， $m^+$  是输出函数严格递增部分的梯度，则  $x_i^+ = m^+ \tilde{a}_i^+ H(\tilde{a}_i^+)$ 。然后，STN 靶核受到总激发(9)的支配

$$X^+ = m^+ \sum_{i=1}^n (w_i c_i + \varepsilon' - w_g y_i^e) H(\tilde{a}_i^+) \quad (9)$$

考虑到相对较小的 STN 尺寸和大量核内激发的可能性(Kita 等人,1983)，STN 可能不支持离散通道。相反，考虑一个模型，其中所有通道都由一个输出为  $x^+$  的单细胞群“服务”，其中

$$x^+ = m^+ \sum_{i=1}^n (w_i c_i + \varepsilon' - w_g y_i^e) \quad (10)$$

对(9)和(10)的比较表明，尽管它们的结构不同，但对于 STN 目标而言，离散和均匀的 STN 备选方案都具有相似的功能，即以弥散激发的形式整合基底节输入及其再分布。两个模型都进行了分析和模拟，但在它们的一般行为模式上没有显示出什么不同。因此，该模型在 STN 输入集成机制的变化下具有鲁棒性，因此我们只报告离散信道情况下的结果。

#### 4.3.1.5. GPe

设纹状体和 STN 的突触强度分别为  $w^-$  和  $w^+$ ，并令  $w^+ = \delta w^-$ 。假设  $\tilde{a}_i^e$  是 GPe 中通道  $i$  的平衡激活，和  $\varepsilon_e$  通道独立阈值项，则  $\tilde{a}_i^e = w^- (\delta X^+ - x_i^{e-}) + \varepsilon_e$ 。使用(7)中给出的表达式  $x_i^{e-}$ ，则有

$$\tilde{a}_i^e = w^- \left\{ \delta X^+ - m^- [(1 - \lambda_e) w_i c_i - \varepsilon] H_i^+ (\lambda_e) \right\} + \varepsilon_e \quad (11)$$

且输出关系为

$$y_i^e = m^e \tilde{a}_i^e H(a_i^e) \quad (12)$$

#### 4.3.1.6. GPi/SNr

假定 GPi/SNr 的 STN 输入与 GPe 相同。这与这些区域的分支 STN 侧支的证据一致 (Deniau 等人, 1978; Kita 等人, 1983)。此外，我们假设来自 STN 的传入神经具有与 GPe 相同的突触强度。如果  $w_e$  是 GPe 传入的突触强度且  $\varepsilon_g$  为阈值项，那么 GPi/SNr 平衡激活为

$$\tilde{a}_i^g = w^- (\delta X^+ - x_i^{g-}) - w_e y_i^e + \varepsilon_g \quad (13)$$

从(8)中代换  $x_i^{g-}$  得到

$$\tilde{a}_i^g = w^- (\delta X^+ - m^- [w_s (1 + \lambda_g) c_i - \varepsilon] H_i^+ (-\lambda_g)) - w_e y_i^e + \varepsilon_g \quad (14)$$

如果  $y_i^g$  是平衡时的 GPi/SNr 输出，那么  $y_i^g = m^g \tilde{a}_i^g H(a_i^g)$ 。

### 4.3.2. 分析结果

本节总结了最重要的分析结果。在解的研究工作中，将所有输出函数梯度  $m$  归一化是很方便的；这仅意味着可能会重新调整权重和阈值偏移。我们首先求解每个 BG 核内的活性，然后使用这些表达式来探索整个模型的选择特性及其在多巴胺能调节下的行为。所有结果的详细证明可在 Gurney 等人(1998)中找到。

#### 4.3.2.1. 核活性

**GPe:** 设  $\aleph_0$  为一组非 GPe 活动通道(即  $\aleph_0 = P - P_e^*$ )。现在通过  $\aleph = \{i \in P_e^* : y_i^e > 0\}$ ， $\aleph^* = \{i \in P_e^* : y_i^e = 0\}$  定义  $P_e^*$  的两个子集。则有

$$\begin{aligned} i \in \aleph_0 : y_i^e &= w^- \delta X^+ + \varepsilon_e \\ i \in \aleph : y_i^e &= w^- \left\{ \delta X^+ - [(1 - \lambda_e) w_i c_i - \varepsilon] \right\} + \varepsilon_e \\ i \in \aleph^* : y_i^e &= 0 \end{aligned} \quad (15)$$

STN: 从(9)式开始, 可以通过检查静态条件下的信号值来确定参数约束, 即  $c_i = 0, \forall i$ 。如果  $\hat{x}$  是这些情况下的强直 STN 水平, 那么, 由于  $\hat{x} > 0, H(\tilde{a}_i) = 1, \forall i$ , 并且(在梯度归一化条件下)(9)给出  $\hat{x} = n(\varepsilon' - w_g y_i^e)$ 。所有通道都是  $\mathfrak{S}_0$  的元素。因此, 从(15)替换并求解  $\hat{x}$  得出  $\hat{x} = n(\varepsilon' - w_g \varepsilon_e) / (1 + m w_g \delta w)$ 。要使这一点成为正直, 需要要求

$$\varepsilon' > w_g \varepsilon_e \quad (16)$$

对于一般的输入条件, 如果  $S^*$  是  $H(\tilde{a}_i) = 1$  的一组通道, 则有

$$X^+ = \sum_{i \in S^*} (w_i c_i + \varepsilon' - w_g y_i^e) \quad (17)$$

我们现在通过反证法来证明  $S^*$  永远不可能是空的。假设  $S^* = \emptyset$ , 那么  $X^+ = 0$ 。则由(15),  $y_i^e \leq \varepsilon_e, \forall i$ , 从而  $\varepsilon' - w_g y_i^e \geq \varepsilon' - w_g \varepsilon_e > 0$ , 最后一个不等式由(16)式确定, 表明  $\tilde{a}_i > 0, \forall i$ , 因此  $X^+ > 0, S^* \neq \emptyset$ 。所以, 最初的前提是矛盾的, 即前提是错误的。

我们现在求解  $X^+$ 。将(17)式中的求和项除以由 GPe 输出类确定的子集得出

$$X^+ = \sum_{i \in S^*} (w_i c_i + \varepsilon') - w_g \left( \sum_{i \in S^* \cap \mathfrak{N}_0} y_i^e + \sum_{i \in S^* \cap \mathfrak{N}} y_i^e \right) \quad (18)$$

现在定义  $n^s = |S^*|$ ,  $\phi_s^g = |S^* \cap \mathfrak{N}| / n^s$ 。设  $\mathfrak{R} = \mathfrak{N} \cap \mathfrak{N}_0$  且  $\phi_s^m = |S^* \cap \mathfrak{R}| / n^s$ 。定义平均显著值

$$\langle c \rangle_s = \frac{1}{n^s} \sum_{i \in S^*} c_i, \quad \langle c \rangle_{q,s} = \frac{1}{n^s \phi_s^q} \sum_{i \in S^* \cap \mathfrak{N}} c_i \quad (19)$$

则用(15)中的(18)替换并求出  $X^+$

$$X^+ = \frac{n^s}{1 + \delta w_g w n^s \phi_s^m} \left\{ w_i \langle c \rangle_s + w_g w^- \times \phi_s^g \left[ (1 - \lambda_\varepsilon) w_s \langle c \rangle_{q,s} - \varepsilon \right] + \varepsilon' - w_g \phi_s^m \varepsilon_e \right\} \quad (20)$$

GPI/SNr: GPI/SNr 激活的一般表达式由(14)给出。当  $i \notin P^*$  和  $i \in \mathfrak{N}_0$  激活时, 会出现强直状态。

则有

$$\hat{a}^g = w^- \delta X^+ \gamma^e - w^e \varepsilon^e + \varepsilon_g \quad (21)$$

其中  $\gamma^e = (1 - w^e)$ 。(21)式的右侧必须大于零, 才能达到正强直水平。因此, 假设  $\varepsilon_g$  和  $\varepsilon_e$  的大小相似, 则  $\gamma^e > 0$ , 这意味着

$$w^e < 1 \quad (22)$$

#### 4.3.2.2.BG 作为选择机制

本节展示了有关模型基本选择特性的主要分析结果。我们首先证明模型满足实现保序映射的最低要求; 这是 GPR1 中给出的*选择机制*的定义。接下来, 我们展示了该模型的特点, 以便于潜在的有用选择; 自动能力扩展的竞争力, 以及选择性的提升。

##### 4.3.2.2.1.保序映射

如果模型实现了保序映射, 并且赋予选择阈值  $\theta_1$  和  $\theta_2$ , 分别定义了选定和未选定的输出信号集  $S$  和  $\bar{S}$  (如果  $y_i^g < \theta_1$  则  $y_i^g \in S$ ; 如果  $y_i^g > \theta_2$  则  $y_i^g \in \bar{S}$ ), 那么该模型就是在 GPR1 中定义的*选择机制*。此外, 考虑到强直输出  $\hat{y}^g$  应该足够确保无选择, 我们假设  $0 \leq \theta_1 < \theta_2 \leq \hat{y}^g$ 。

如果  $\Delta_{ij} y^g = y_i^g - y_j^g$ , 且  $c_j > c_i \Rightarrow \Delta_{ij} y^g \geq 0$  则信号*排序保留*。为了证明这一点, 最初使用激活更容易。如果  $\Delta_{ij} \tilde{a}^g = \tilde{a}_i^g - \tilde{a}_j^g$ , 则从(13)式出发有

$$\Delta_{ij} \tilde{a}^g = w^- \left\{ [(1 + \lambda_g) w_s c_j - \varepsilon] H_j^\uparrow(-\lambda_g) - [(1 + \lambda_g) w_s c_i - \varepsilon] H_i^\uparrow(-\lambda_g) \right\} - w_e \Delta_{ij} y^e \quad (23)$$

这里， $\Delta_{ij}y_i^s = y_i^s - y_j^s$  且是运用式(15)来定义的，并通过将  $i, j$  作为  $\mathbb{S}_0$ ， $\mathbb{S}$  和  $\mathbb{S}^*$  的元素进行适当的赋值。类似于阶跃函数  $H_j^+(\lambda_g)$ ， $H_i^+(\lambda_g)$  也是根据是否满足  $i, j \in P^*$  来定义的。集合成员的这些组合在所有情况下产生十个类别，并且每个案例的测试给出以下结果：如果  $c_j > c_i$ ，则  $\Delta_{ij}\tilde{a}_i^s \geq 0$ ，当且仅当  $i, j \in P^*$  时，等号成立。使用输出关系(3)，除非  $y_i^s = y_j^s = 0$ ，否则有  $\Delta_{ij}\tilde{a}_i^s > 0 \Rightarrow y_i^s > y_j^s$ ，所以我们有

结果 1:  $\Delta_{ij}y_i^s > 0$ ，当且仅当  $y_i^s = y_j^s = 0$  或  $i, j \notin P^*$  成立时，等号成立。

#### 4.3.2.2.2. 尺度缩放

这在 GPR1 中讨论过，并提到了必须在激励和抑制之间保持适当平衡，以使所门控道集不为空的要求。为了研究这一点，我们需要确定通道输出活动  $\tilde{a}_i^s$  相对于强直水平  $\hat{a}^s$  的一般行为。如果  $\Delta^i = \tilde{a}_i^s - \hat{a}^s$ ，则使用(14)和(21)，可以为  $\Delta^i$  编写一个通用形式，该形式适用于所有情况

$$\Delta^i = \delta \Delta X^+ - L(\lambda_g, \lambda_e)c_i + k_3 \quad (24)$$

其中  $\Delta X^+ = k_1 X^+ - k_2 \hat{X}^+$ ，且  $k_1 \geq k_2$  时取正。此外， $k_1$  和  $L$  不依赖于通道的数量；唯一的显著性依赖项是  $X^+$  和  $c_i$ ，多巴胺依赖性包含在  $L$  中。

现在考虑一下，如果 GPe 对 STN 没有抑制作用，会发生什么，即  $w_g = 0$ 。则  $n^s = n$ ，并且从(20)式出发，有  $X^+ = n(w_g \langle c \rangle_s + \varepsilon)$  和  $\hat{X}^+ = n\varepsilon$ 。因此， $\Delta X^+$  与  $n$  成正比，兴奋性项  $\delta \Delta X^+$  在(24)中占主导地位，容量缩放只能通过确保  $\delta = O(1/n)$  来实现。

现在考虑一下当 GPe 抑制被实例化时的情况。我们首先证明  $X^+$  在上面有界。关于(20)，右手边的花括号中的因子包含的项的大小受权重比例的限制，因此以某个值  $c_0$  和(25)式为界。

$$X^+ \leq n^s c_0 / (1 + \delta w_g w^- n^s \phi_s^m) \quad (25)$$

这不一定是有限的，因为  $n^s$  可能无限增长，而  $n^s \phi_s^m$  保持有界。为了证明这是不可能发生的，我们首先观察到——对于相当大的  $X^+$ ——所有  $y_i^s > 0$  (见(11))。换句话说， $\exists X^+$  使得下式成立

$$X^+ > X_i^+ \Rightarrow y_i^s > 0 \forall i \Rightarrow \mathbb{S}^* = 0 \Rightarrow \phi_s^m = 1 \quad (26)$$

我们现在用反证法继续证明。因此，假设  $X^+$  不在上面有界。然后给定任何一个  $N$ ， $\exists n^s > N$  使得下式成立

$$X^+ > \max(X_i^+, c_0 / \delta w_g w^-) \quad (27)$$

为了(26)成立，且  $\phi_s^m = 1$ 。则从(25)出发， $X^+ \leq c_0 / (1/n^s + \delta w_g w^-) < c_0 / \delta w_g w^-$  与(27)相矛盾，因此  $X^+$  是在有上界的。紧接着， $\delta \Delta X^+$  也是如此。因此，兴奋项  $\delta$  是有界的，没有使  $\delta$  无限小并精确匹配到通道容量。此外， $X^+$  和(25)有界性的一个推论是， $\phi_s^m$  有下界。总之，我们得到了以下结果

结果 2:

- ①如果  $w_g = 0$ ，则只能通过确保  $\delta = O(1/n)$  实现容量缩放。
- ②如果  $w_g \neq 0$ ， $X^+$  有上界，且  $\exists \delta_0 > 0 > 0$ ，如果  $\delta > \delta_0$ ，当  $n \rightarrow \infty$  时，容量缩放将被保持。
- ③  $\phi_s^m$  有下界。

当然，限制  $X^+$  的原因是，STN 的 GPe 抑制。这体现在  $X^+$  表达式(20)中分母项  $\delta w_g w^- n^s \phi_s^m$  中，这对证明  $X^+$  的有界性至关重要。因此，模型中的容量缩放是自动的，因为它是由 GPe  $\rightarrow$  STN 抑制反馈的动态相互作用得到的，并且在结构上不是通过显式设置权重值来确定的。

#### 4.3.2.2.3.竞争力

该模型保留了 GPR1 中讨论的简单前馈机制所显示的竞争力,通过评估  $d\tilde{a}_i^s/dc_j$  可以证明这一点。关于  $P^*$ ,  $\aleph_0$ ,  $\aleph$  和  $\aleph^*$ , 有一组基于  $i$  的集合元素的案例,给出了以下结果

结果 3:

- ①如果  $i \neq j$ , 则如果  $i \in S^*$  有  $d\tilde{a}_i^s/dc_j > 0$ ; 如果  $i \notin S^*$  有  $d\tilde{a}_i^s/dc_j = 0$ 。
- ②如果  $i \notin S^*$ , 除非  $i \in P^*$ , 否则  $d\tilde{a}_i^s/dc_i < 0$ , 在这种情况下  $d\tilde{a}_i^s/dc_i = 0$ 。
- ③如果  $i \in S^*$ , 那么如果  $\delta < 1$  且  $w_i = w_s$ , 除非  $i \in P^*$ , 否则  $d\tilde{a}_i^s/dc_i < 0$ , 在这种情况下  $d\tilde{a}_i^s/dc_i > 0$ 。

一个推论是, 只要  $y_i^s \neq 0$ , 一组相似的结果就可以保持输出  $y_i^s$ 。结果 3 的第 1 条表明, 在外部 BG 刺激下, GPI/SNr 刺激可能增加。这与 GPI/SNr(Mink,1996)的已知特性一致, 并且在通道  $i$  变得越来越突出时, 在通道  $j \neq i$  上的模型中出现。它的起源是通过 STN 的激励, 当  $c_i$  增加时, 通道不是  $i$  接收的。如果  $c_i$  足够大, 通道  $i$  可能会被选中, 并且之前选定的通道  $k$  上的输出可能会被提升到它失去选定状态的点; 在这种情况下, 我们说通道  $i$  中断了通道  $k$ 。结果 3 的第 2 条和第 3 条说明, 通常, 通道上的输出活动倾向于随着该通道的显著性增加而减少。当显著性足以使通道纹状体活跃时, 就会发生异常。然后, STN 刺激在没有相应纹状体抑制的情况下增长, 从而使输出活性增加。

#### 4.3.2.2.4.选择性

再次考虑(24)式给出的 Gpi/SNr 激活变化。假设  $k$  通道上的显著性是固定的, 而其他通道上的显著性是允许增加的。然后根据结果(3),  $X^+$ ——因此  $\Delta X^+$ ——会增加。因此, 其他通道的活动使得通道  $i$  上输出  $\Delta^k < 0$  的任何减少都难以实现。当然, 这仅仅是通道间竞争的结果, 它会限制模型的混乱性(等同地增加选择性)。

选择性增强的程度将取决于显著性增加对(24)中总兴奋项  $\delta \Delta X^+$  的影响率。现在,  $d(\delta \Delta X^+)/dc_i$  与  $\delta \Delta X^+/dc_i$  成比例。为了确定后者, 设  $\chi_s(i)$  和  $\chi_q(i)$  分别作为元素  $S^*$  和  $\aleph$  的特征函数 ( $\chi_s(i) = 1 \Leftrightarrow i \in S^*$ ,  $\chi_q(i) = 1 \Leftrightarrow i \in \aleph$ )。从(19)可得

$$\frac{d\langle c \rangle_s}{dc_i} = \frac{\chi_s(i)}{n^s} \quad \frac{d\langle c \rangle_{q,s}}{dc_i} = \frac{\chi_s(i)\chi_q(i)}{n^s \phi_s^q} \quad (28)$$

并运用(20)得

$$\frac{dX^+}{dc_i} = \frac{\chi^s(i)[w_i + w_g w^- (1 - \lambda_c) w_s \chi_q(i)]}{1 + \delta w_g w^- n^s \phi_s^m} \quad (29)$$

现在假设  $\chi_s(i) = 1$ 。此外, 如果  $w_i = w_s$  (如结果 3 第 3 条中的假设), 则(29)表明

$$\frac{dX^+}{dc_i} = \frac{w_i [1 + w_g w^- (1 - \lambda_c) \chi_q(i)]}{1 + \delta w_g w^- n^s \phi_s^m} \quad (30)$$

对于大的  $n_s$ , 结果 2, 第 3 条表明  $n^s \phi_s^m$  远大于 1, 因此

$$\frac{dX^+}{dc_i} \approx \frac{w_i [1 + w_g w^- (1 - \lambda_c) \chi_q(i)]}{1 + \delta w_g w^- n^s \phi_s^m} \geq \frac{1}{w_g w^- n} \quad (31)$$

现在, 在梯度归一化下, 如果神经元输出函数的线性部分在大多数输入组合下被执行, 最大激活不应该显著大于 1。这同样适用于对每个节点的兴奋性和抑制性贡献, 因为每个节点只有一个抑制性输入, 这对  $w^-$  和  $w_g$  施加了类似的约束。符合此标准的权重称为单位标度。特别是, 如果  $w_g w^- = 1$ , 由(31)得出

$$\delta \frac{dX^+}{dc_i} \geq \frac{w_i}{n} \quad (32)$$

现在假设没有 GPe→STN 路径，那么  $w_g=0$ 。则  $n^s = n$ ，并且如果  $\delta = 1/n$  (获得容量缩减)，则由(30)得

$$\delta \frac{dX^+}{dc_i} = \frac{w_i}{n} \quad (33)$$

与(32)相比，当 GPe→STN 路径被实例化时，激发的增加更快。反过来，这意味着更显著的竞争力和更强的选择性。

#### 4.3.2.3.多巴胺能调制

在本节中，我们分别研究了用  $\lambda_g$  和  $\lambda_e$  参数化的选择和控制路径中，GPi/SNr 输出随多巴胺水平变化的方式。选择路径的结果立即从(14)开始。因此，如果  $i \in P^*$ ，则有

$$\frac{d\tilde{a}_i^s}{d\lambda_g} = -w^- w_s c_i \quad (34)$$

要找到  $d\tilde{a}_i^s/d\lambda_e$ ，需要对  $dX^+/d\lambda_e$  进行评估。根据(20)得

$$\frac{dX^+}{d\lambda_e} = \frac{-n^s w_g w^- w_s \phi_s^s \langle c \rangle_{q,s}}{1 + \delta w_g w^- n^s \phi_s^m} \quad (35)$$

一个完整的分析需要对几个案例进行测试，但要注意  $\tilde{a}_i^s$  依赖  $\lambda_e$ ， $\langle c \rangle_{q,s}$  必须是非零的。结合(35)和(34)得出的结果得出

结果 4:

$$\frac{d\tilde{a}_i^s}{d\lambda_g} \begin{cases} < 0: i \in P^* \\ = 0: i \notin P^* \end{cases} \quad (36)$$

$$\frac{d\tilde{a}_i^s}{d\lambda_e} \begin{cases} < 0: \langle c \rangle_{q,s} \neq 0 \\ = 0: \langle c \rangle_{q,s} = 0 \end{cases} \quad (37)$$

由于  $y_i^s$  是  $\tilde{a}_i^s$  的单调函数，因此对于 GPi/SNr 输出，该结果有一个推论，其中非零导数的条件被扩展到包含  $y_i^s \neq 0$ 。因此，增加任何一种途径中多巴胺水平的一般效应是降低 GPi/SNr 输出水平，从而使选择更加混乱。

控制通路中的多巴胺调节是由 GPe 对 STN 和 GPi/SNr 的影响介导的。然而，GPe→GPi/SNr 通路对信号间间隔  $\Delta_y \tilde{a}_i^s$  的控制至关重要，因为只有当  $w_e \neq 0$  时才可能显示  $d\Delta_y \tilde{a}_i^s/d\lambda_e \neq 0$ 。而且，在大多数情况下，随着多巴胺水平的增加，间隔  $\Delta_y \tilde{a}_i^s$  增加。

#### 4.3.3.模拟结果

##### 4.3.3.1.参数详情

模拟使用了 6 通道模型。所有坡度参数( $m^-$ ,  $m^+$ ,  $m^e$ ,  $m^s$ )均设为 1，这是分析中的梯度归一化假设。应用于  $w_s$ ,  $w_i$ ,  $w^-$  和  $w_g$  的权重的单位缩放，这些权重均设置为 1。结果 3 还要求  $\delta < 1$ ；我们使用了  $\delta = 0.9$  以便  $w^+ = \delta w^- = 0.9$ 。为了满足(22)中的条件， $w_e$  被设置为 0.3。纹状体偏移  $\varepsilon$  被设置为 0.2 以强制“下降状态”，除非激活为阳性。其他的偏移量  $\varepsilon^e \varepsilon_e$  和  $\varepsilon_g$  设置为负值 -0.25, -0.2 和 -0.2。特别是，GPi/SNr,  $\hat{y}^s$  的强直输出约为 0.15。在灵长类动物中， $0 < y^s < 1$  和 GPi/SNr 的强直输出约为 60-80Hz(Georgopoulos 等人,1983;Mink 和 Thach,1991)这意味着生理上合理的最大 GPi/SNr 率为 400-500Hz。最后，设定模拟时间尺度的(1)式中的速率常数  $k$  为 25。

##### 4.3.3.2.基本选择特性



第 3.2 节中描述的一些特性如图 2 所示的模拟结果所示。在图 2a 中，通道 1 和 2 在模拟过程中被刺激，而其余通道的显著性被设置为零。因此，信道 3 代表通道 3-6。多巴胺水平  $\lambda_e$  和  $\lambda_s$  均设为 0.2。最初，在所有六个通道上都有零显著性(注意，输出信号中的初始瞬变是模拟器的假象，因为它建立了一组一致的活动级别)。当  $t=1$  时，通道 1 上的显著性增加到 0.4。这导致该通道上的输出  $y_1^s$  减小，其他四个通道上的输出增加。如果  $y_1^s < \theta_1$ ，则通道 1 已被选中。当  $t=2$  时，通道 2 的显著性变为 0.6。这足以强制  $y_2^s = 0$ ，从而确保选择通道 2，同时将  $y_1^s$  推到强度值  $\hat{y}^s$  之上。通道 3(以及所有其他非活动通道)上的输出进一步增加。因此，如果在  $t=2$  之前选择通道 1，则选择通道 2 会导致其选择中断。接下来的两个事件是由通道 1 上从  $t=3$  到  $t=4$  的显著性瞬间增加引起的。在此期间，通道 1 和通道 2 都具有相同的、相对较高的显著性，但通道间竞争导致的共同输出水平高于单独通道上相同的显著性。请注意，信号的顺序在整个过程中都被保存：如果  $c_i > c_j$ ，则  $y_i^s \leq y_j^s$ 。

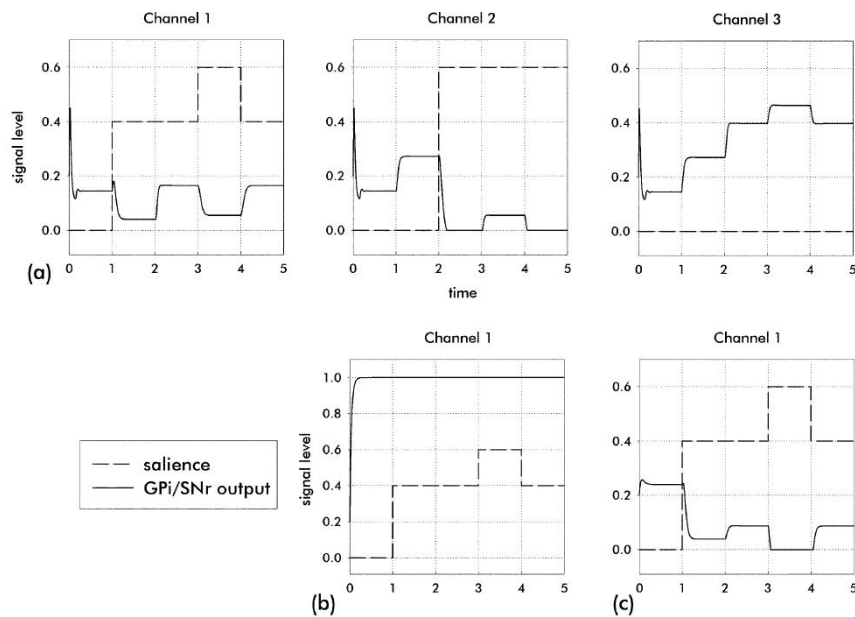


图 24: 六通道模型中三个通道的仿真结果。

只有信道 1 和信道 2 具有非零显著性；信道 3 显示为其他非活动信道的代表。对于每个通道，实线表示 GPI/SNr 输出，虚线表示输入凸度。时间以任意单位测量。(a) 一个完整模型的结果是，STN 输出的突触权重是纹状体的 0.9 倍。(b) “降低” GPe→STN 通路模型的结果，同时将 STN 的突触权重保持在原来的值。(c) 与(b)中相同的损伤，但现在 STN 的权重是以前值的 1/6。

图 2b 显示了容量缩放的现象。该结果是通过一个模型得到的，在这个模型中 GPe→STN 通路被削弱，而与 STN 传入相关的权重保持不变。显著性事件与用于获得图 2a 的事件相同，但仅显示通道 1。在整个刺激期内，输出饱和于 1，无法选择。选择如图 2c 所示，可以通过设置  $\delta = 1/n = 1/6$  来恢复。但是，为了在  $n$  增加时保持这种行为， $\delta$  必须按比例减少。图 2a 和图 c 中通道 1 的图比较也表明没有 GPe→STN 路径的选择性降低。因此，在后一种情况下，两个通道都能被选择为  $3 < t < 4$ ，因为它们的选择在 STN 的 GPe 抑制下是不确定的。

#### 4.3.3.3.多巴胺能调制

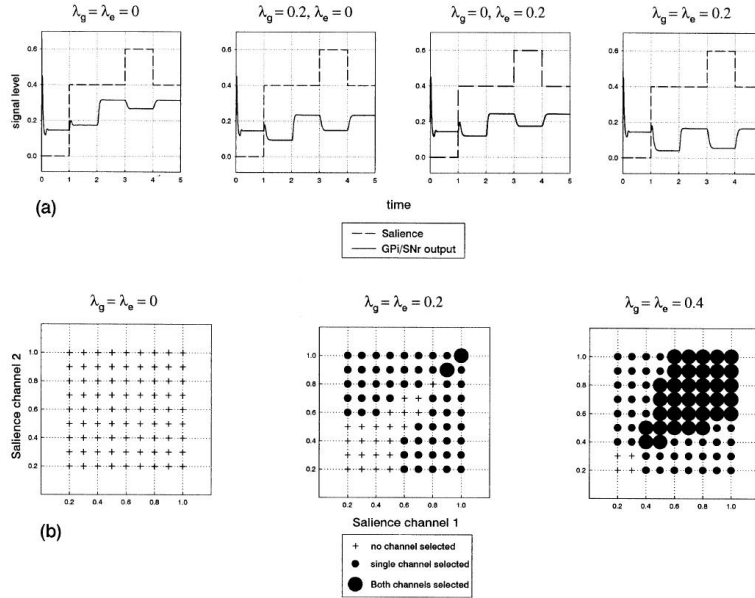


图 25: 多巴胺对模型输出行为的影响。(a)类似于构成图 3 结果基础的模拟(此处仅显示通道 1)。(b)多巴胺对两个竞争通道选择的影响。每个子图显示显著组合排列的通道选择,显著值在 0.2 到 1 之间,步幅为 0.1。在每种情况下,如果一个信道的输出为零,则该信道被视为被选中。两种途径的多巴胺水平相同。

图 3a 显示了使用图 2 结果中使用的突出事件序列对通道 1 进行的模拟。如果没有多巴胺( $\lambda_g = \lambda_e = 0$ ),则不会发生选择。在选择途径中有中等水平的多巴胺,但在控制途径( $\lambda_g = 0.2, \lambda_e = 0$ )中没有。GPi/SNr 输出信号有所减少。这同样适用于多巴胺仅在控制途径( $\lambda_g = 0, \lambda_e = 0.2$ ),但效果不那么明显。在两种途径多巴胺水平相同的情况下,各途径对 GPi/SNr 的影响是协同作用的。与多巴胺相关的通道输出减少的一个后果是,模型中的选择往往更加混乱。图 3b 说明了这一点,其中每个子图显示了在一系列显著性组合中两个竞争信道的选择,以 0.1 为步进,取 0.2 到 1 之间的值。在每种情况下,如果一个通道的输出为零,并且控制和选择路径中的多巴胺水平相同,则该通道被视为“已选择”。在没有多巴胺的情况下,无论有多大的显著值,都不会选择任何一个通道。现在考虑一个中间水平的多巴胺( $\lambda_g = \lambda_e = 0.2$ )的情况。在这里,低于 0.6 的显著性无法启动选择,但对于大于或等于此值的值,通道选择是可能的。近匹配显著性值小于 0.9 时,通道间竞争同样是近匹配的,两个通道均未被选择。但是,如果两个通道都相当突出,则可能强制同时选择。

在高多巴胺( $\lambda_g = \lambda_e = 0.4$ )的情况下,出现了类似的一般结果模式,但选择的显著阈值较小(0.4),同时选择通道的机会明显更多。

#### 4.3.3.4.与神经生理学数据的比较

模型的一个强大的测试是比较外部刺激下的特征信号分布与生物系统中在类似条件下获得的特征信号分布。信号图轮廓可以被认为是与它们的产生相关的结构特征的标志,因此它们在体内和模型中的识别是一个共同体系结构的证据。这种立场类似于系统(控制)理论中所采用的立场,通过检查系统对立体输入的响应,可以对系统进行结构和/或参数识别。特别是,一个系统完全以其“脉冲响应”为特征:它的输出由脉冲输入驱动,脉冲输入的持续时间非常短(理论上为零),振幅非常大(理论上为无穷大)。微电极对神经组织的人工电刺激通常会产生这种输入。Ryan 和 Clark(1991)使用这种技术,在大振幅电流(0.3, 0.7mA)的短脉冲(0.2 ms)下刺激大鼠皮层,然后从 GPe 中记录。他们获得的一些数据如图 4a, c 所示。

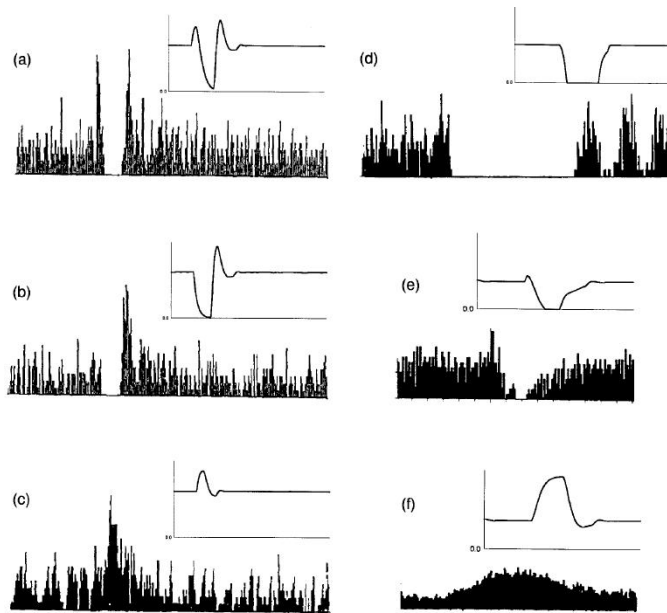


图 26: 单个单位记录(连续刺激直方图)和相应的模型模拟结果(线性图)。(a)-(c)对 BG 皮质输入的单脉冲电刺激后, 苍白细胞单位的反应(Ryan 和 Clark,1991;数据经 Springer-Ver-Lag 许可使用)。模拟均使用短时间(相对于模型的固有时间尺度)和大振幅(大于 1)的矩形凸脉冲。(d)STN 兴奋性输入损伤后的数据和模拟。(e)猴子的 SNr 细胞在一项行为任务中显示出抑制(Schultz, 1986;数据经美国生理学会许可使用)和单个显著通道的 GPi/SNr 输出模型。(f)来自同一个实验的信噪比神经元数据, 该实验将增加其输出以及来自非刺激通道的相应模拟结果。

这些记录显示了与短时间的兴奋和抑制相称的各种组合中活动增加和减少的瞬变过程。每个子图中的插图显示了使用短时间(相对于模型的固有时间尺度)和大振幅(大于或等于 1)的显著脉冲从 GPe 获得的模拟输出。通过将模型输出解释为 GPe 的平均触发率, 可以与电生理数据进行比较; 在每种情况下, 兴奋和抑制瞬变的模式都是可行的。图 4d 显示了与图 4a 所示相同的信号注入位置的数据, 但在禁用 STN 的兴奋性输入之后; 插图显示了相应的模型模拟。注意, 在数据和模拟中, 早期和再束缚激发的减少和抑制阶段的持续时间的延长。除了检查系统对人工脉冲的响应之外, 还需要确认模型能够在其正常工作状态下显示信号行为, 这与行为动物的数据相当。图 4e 显示了猴子在一项行为任务中, 显示活动减少的 SNr 细胞的数据(Schultz, 1986)。本子图中的插图是使用小振幅矩形脉冲输入的单个有源通道的 GPi/SNr 输出模型。数据说明的主要点是 SNr 中去抑制的基本过程, 但要记住, 生物显著性可能没有简单的矩形形式, 在模拟中使用, 观察数据和模拟之间的其他接触点很有趣。因此, SNr 信号的上升时间明显慢于其初始下降时间, 并且解除抑制开始之前有一个短暂的兴奋性瞬变。图 4f 显示了同一个实验中的一个增加输出的信噪比神经元, 以及一个非刺激通道的类似模拟结果。因此, 我们将从我们的模型中预测, 图 4e 和 f 中的数据分别与所采用的行为任务的活动和非活动通道相关。

#### 4.3.4. 讨论

##### 4.3.4.1. BG 结构的功能机制

结果表明, 基于 BG 解剖基本特征的计算模型能够传递与动作选择功能一致的 GPi/SNr 输出信号。信号的选择分为两个阶段。首先, 纹状体内的一组局部解剖地局部循环电路执行硬切换, 以使至少一个通道可用于每个此类电路内的进一步处理。接下来, 一个基于非中心环绕结构的全局前馈网络从当地电路竞争的获胜者中选出。选择的第二阶段满足输入(通道显著性)和输出(GPi/SNr 抑制)之间的保序映射的形式要求。此外, 在允许多通道选择(软切换)的同时, 它还具有通道间竞争, 可能限制选择的混乱性。

全局选择路径的特性受来自 GPe 的控制信号的调制, 假定的控制路径的作用似乎是三

倍。首先, 确保在激发和抑制之间有一个适当的平衡, 使有用的选择发生。这个容量缩放过程是使用从 GPe 到 STN 的负反馈自动执行的, 与活动的 STN 通道  $n^s$  数无关。运用负反馈来稳定系统或自适应地控制系统的工作点是控制工程中的一个常见概念, 这种机制自然地出现在生活系统中, 特别是在 BG 等控制结构中, 这并不奇怪。我们认为, 假设与 STN(由  $\delta$  控制)的突触接触是根据全局系统参数(如通道数)进行调整是不现实的。相反, 我们建议 BG 运用 GPe 到 STN 的负反馈提供的自适应机制来自动缩放 STN 的激励以适应任意数量的通道。控制通路的第二个作用是增强通道间竞争下的选择性。这也可以被看作是模型中激励的动态分配的另一个特征, 该模型也是容量缩放的基础。第三, 这种途径有助于多巴胺能控制选择的表达, 一般来说, 选择和控制途径中的多巴胺协同工作, 促进更混乱的选择, 并增加通道间输出激活间隔。

#### 4.3.4.2. 临床、行为和生理学研究的的关系

多巴胺在选择和控制途径中的协同效应与 Murer 等人(1997)的研究一致, 这表明两种多巴胺受体亚型的同时激活对 SNr 激活率有相互促进的作用。由于多巴胺水平较低, 在广泛的显著性下, 选择是被禁止的。该模型在这些情况下的行为与观察到的人类多巴胺消耗与帕金森病(PD)几种症状之间的关系一致。这些包括运动迟缓(运动执行缓慢)和运动机能减退(自发运动缺乏, 例如, 相对缺乏面部运动、眨眼和行走时摆动手臂(Korczyn, 1995))。在帕金森病中, 这两种症状通常被归入运动缺乏症的范畴, 但无论用什么术语来说, 它们都可以被视为运动动作选择和执行能力的缺陷。在我们的模型中, 帕金森病患者无法在选定的运动通道上适当释放 GPi/SNr 的强直抑制输出; 因此, 运动机能减退会与当前模型选择通道的失败相一致。然而, 为了能够解决运动迟缓(选定运动的减速)的问题, 需要在抑制程度和运动速度之间建立联系。这种机制已成功地应用于一种受生物启发的运动控制模型(Bullock 等人, 1998), 并且是我们动作选择的具体模型(Gonzalez, 2000)中电机门控的基础。我们的模型对动物多巴胺消耗损伤后出现的致盲现象提供了类似的解释(如 Jenner 等人, 1984)。

在我们的模型中, 我们发现, 在多巴胺水平中等的情况下, 模型中的选择限制可以迫使两个高度突出的通道相互对抗, 从而导致两个通道的选择都比只有一个具有相同突出性的通道更弱(抑制释放更少)。这对应于同时执行两个操作的情况。一般来说, 两个任务的性能都比单独执行时差。对这一现象的解释是以有限注意的形式出现的, 但术语“注意”有许多标准(Styles, 1997), 其中一个标准可能与 BG 内的竞争所施加的限制有关。这一可能性得到了 Benecke 等人(1986)的研究证据的支持。这表明 PD 患者(相对于对照组)在双运动任务上的表现比他们在单独任务上的表现更差。这将导致 BG 缺损, 而不是皮质注意力“瓶颈”或缺乏“执行”控制。

随着多巴胺水平超过正常水平, 我们的模型预测同时释放抑制(从而选择通道)的机会也会增加。然而, 当前模型嵌入更广泛的解剖背景(包括皮层、丘脑和运动资源)时, 小显著性通道的同时减少输出可能不稳定。在这样一个扩展的模型中, 我们预计两个相关行为之间会有快速的转换, 而不是真正的同时动作选择。这可能与快速的行为和认知转换以及忽视干扰因素的能力相似, 如在与 BG 紊乱相关的几种临床条件中所见, BG 的高水平多巴胺神经传递受到牵连: 这些包括注意力缺陷多动障碍(Swanson 等人, 1998), 图雷特综合症(Tourette's syndrome)(Brito, 1997), 精神分裂症(Schizophrenia)(calabresi 等人, 1997)和兴奋剂中毒(Stimulant intoxication)(Genova, 1997)。在我们自己的实验室中对丘脑-皮质-纹状体环路和棘突神经元群进行的初步模拟表明, 在低显著性条件下, 快速转换确实是一种可能。

亨廷顿氏病(Huntington's disease)是另一种与 BG 病理学相关的疾病, 但它不是由于多巴胺神经支配丧失引起的, 而是由中等程度的棘细胞退化引起的。Smith 等人(2000)已经证明,

亨廷顿氏病患者无法在运动任务中对手臂轨迹进行必要的修正，尽管最初的运动是准确的。Smith 等人推测 BG 可能通过调节丘脑参与纠错。然而，正如 Lawrence(2000)所指出的，观察到的运动误差可能是这里提出的 BG 主要功能选择的副作用。

从生理学数据来看，模型模拟数据与某些基底节神经信号“信号”之间的对应关系允许用经典线性系统理论解释其中一些信号(假设子核在其输出区域的近似线性部分上运行)。其中一个例子是 Ryan 和 Clark(1991)的数据中观察到的 GPe 的多相(激发、抑制和反弹)响应，这是阻尼反馈系统脉冲响应的典型特征，可能是由包含 STN 和 GPe 的循环电路造成的。此外，低阶系统的一个特点是脉冲响应早期阶段的持续时间较短(与较长的衰减分量相比)，这在图 4 的模拟和数据中都很明显。

然而，仔细观察图 4d 表明，模拟中没有复制生物数据的某些特征。因此，当 GPe 静默时，经过一段时间的抑制后，GPe 在尖峰群中恢复具有迸发模式的活动。突发性放电可能与神经元的固有膜特性有关，在基底节中，一个突出的例子是由 STN(Beurrier 等人,1999)中钙和钾电流的相互作用提供的。显然，这些现象并不适用于我们简单的神经模型。然而，我们最近已经证明，从现象学上模拟这些电流(Humphries 和 Gurney,2000)是可能的，并且可以重现 Plenz 和 Kital(1999)研究的 STN 和 GPe 体外培养中观察到的大量信号行为。我们预计这一(中间)层次的神经模型将被证明是非常有用的补充从这里描述的系统层次模型收集的见解。

我们的结果(见图 4f)也与 Mink(1996)的分析一致，Mink(1996)的分析引起了人们对很少被强调但被观察到的 10 个神经元中，GPi/SNr 中一些输出神经元的放电率增加，这伴随着更常见的报道的放电减少。后者形成了由 Chevalier 和 Deniau(1990)所确定的去抑制的基本过程的基础。我们假设这些观察反映了在我们的模型中选择的过程，其中选择的通道的输出活动被抑制而竞争的被抑制，但是非选择的信道被维持或增加。

#### 4.3.4.3.分析

本文提出的模型和分析首次表明了 BG 的选择性功能(Mink 和 Thach,1993、Hikosaka,1994、Redgrave 等人,1999a)被确认是生物结构计算模型的特征。其次，模型的一些性质，如容量标度和选择性增强，都不是先验的，只有在定量分析了控制路径的影响后才会出现。这些特征提供了一个完全不同于仅由生物研究产生的假设。最后，选择过程被视为使用强化学习将一系列“成功”选择组合成协调行为模式的过程的先决条件。未来将特别有趣的是，如何将当前模型与最近关于加固功能的建议(Schultz,1997;Redgrave 等人,1999b)和序列学习和自动化过程(Hikosaka,1994;Jog,1999)结合起来。

#### 4.4.BG 的机器人模型：动作选择与内部处理

BG 是一组高度互联的中枢大脑结构，对运动和认知有重要影响。对这些结构的兴趣部分源于它们对包括帕金森氏病、亨廷顿氏病、图雷特综合征、精神分裂症和注意力缺陷多动障碍在内的一系列大脑疾病的重要影响，而且对这些结构的研究也推动了一个多世纪的神经生物学研究。这种广泛的研究工作已经产生了大量的相关数据信息，因此迫切需要更好地理解这些结构的功能。因此，BG 是脑功能计算建模最令人期待的前景之一，也是当下用于广泛建模研究的重点(Houk、Davis 和 Beiser,1995；Wickens,1997；Gillies 和 Arbuthnott,2000；Prescott、Gurney 和 Redgrave,2002；Gurney、Prescott、Wickens 和 Redgrave,2004)。

BG 相关文献中反复出现的一个主题是，这些结构的作用是解除所需动作的抑制，同时保持或增加对不希望动作的抑制(Denny Brown 和 Yanagisawa,1976；Cools,1980；Robbins 和

Brown,1990; Hikosaka,1994; Mink,1996; Wickens,1997)。在我们自己的理论工作(Prescott, Redgrave 和 Gurney,1999; Redgrave, Prescott 和 Gurney,1999a)中,我们提出了这样一种观点: *BG* 作为一种*动作选择机制*,解决了大脑内物理分离但在行为表达上存在竞争的功能单元之间的冲突。我们已经展示了这个观点如何与已知的*解剖学*和*生理学*相关理论建立联系,并且满足了有效*动作选择*装置的几个高级计算要求。根据这一假设,我们还设计了一个程序,在几个抽象层次上模拟 *BG* 和相关结构的电路。重点是研究受已知功能*解剖学*约束的 *BG* 系统级模型,在功能*解剖学*中神经群由简单的*漏积分*子单元表示(Gurney、Prescott 和 Redgrave,2001a、2001b; Humphries 和 Gurney,2002; Gurney、Humphries、Wood、Prescott 和 Redgrave,2004)。在较低水平的神经生物学细节中,我们研究了动作电位序列(峰值)编码的信号模式(Humphries 和 Gurney,2001; Humphries,2002),并研究了*纹状体*神经元膜动力学的生物物理模型(Wood、Gurney 和 Wilson,2004)。所有这些层次的研究都产生了互补的结果,证实了 *BG* 的生物学结构可以作为一种有效的*选择机制*发挥作用。然而,在我们看来,这种计算神经科学的方法,即在更广泛的背景下对特定的大脑系统进行建模,仍然留下许多重要的问题没有解答。首先,我们想知道如何最好地解释模型的输入和输出,我们可以选择将输入视为,例如,感觉信号,或输出视为运动信号,但这种分配基本上是毫无根据的。第二,如果不在更广泛的背景中建立模型,我们就无法判断它是否能够在更为全面指定的控制架构中实现其假设的功能角色。第三,如果 *BG* 与感觉和运动系统之间没有任何联系,我们可能会质疑一个模型是否能够处理噪音或模糊的感觉数据,或者作为一个系统的一部分,对协调实际效应器系统的运动提出质疑。最后,如果没有多个需求的背景,例如需要保持物理完整性、避免威胁、*探索*和*运用*资源等,我们将无法确定模型是否能够满足适应性行为的一些基本要求。因此,在本文中,我们描述了一种将基础神经节和相关的*丘脑皮质*连接的系统级模型嵌入到一个小型*移动机器人*的控制结构中,该小型*移动机器人*执行模拟觅食任务,要求机器人在不断变化的感觉和*动机*条件下选择适当的动作,从而生成集成行为序列。我们把现在正在应用的描述这一套方法,作为具体的计算神经科学(embodied computational neuroscience)。机器人模型的初步结果已在(Montes Gonzalez, Prescott, Gurney, Humphries 和 Redgrave,2000)中进行了描述,并且该模型的一个版本已显示出比标准“*赢者通吃*(winner-takes-all)”在机器人生存任务中的*选择机制*具有更好的选择属性(Girard, Cuzin, Gullot, Gurney 和 Prescott,2003)。然而,本文首先全面介绍了机器人嵌入 *BG* 模型,并与相关的神经行为研究进行了比较,对模型行为进行了广泛的评估。我们还展示了机器人模型在面对多个高*显著性*选择时的行为,并与冲突情况下的动物行为研究进行了比较。

文章的其余部分按如下方式展开论述:第2节回顾了基底节的*动作选择*假设和相关的建模工作。机器人基础神经节模型的*动机*、机器人实施的全部细节以及*动作选择*指标的总结,这些都是第3节(以及随附的附录)中进行了描述。接着第4节描述了三个实验的结果:

实验一:使用扩展基底节模型的无实体版本系统搜索特征空间(Humphries 和 Gurney(2002)对该模型的早期分析进行了扩展);

实验二:我们主要研究机器人基底节模型的*动作选择*;

实验三:高特征性选择背景下的机器人行为研究。

## 五、强化学习(Reinforcement learning)

许多模型已经讨论了 *BG* 在学习中的作用,基于神秘的多巴胺能细胞相似物激发在*时序差分*(TD, temporal difference)学习中的*误差信号*的要求。这些模型将 TD 学习的 Actor-Critic 实施映射到 *BG*(Barto,1995;Houk,Adams 和 Barto,1995;Montague,Dayan 和 Sejnowski,1996;

Schultz, Dayan 和 Montague, 1997), 其中大致上, Actor 映射到 BG 的选择功能, Critic 映射到图 2 中的 RL 电路。因此, 多巴胺信号被设想为教学信号, 改变行为的反应, 以最大限度地提高未来的回报。

Joel、Niv 和 Ruppin(2002)评估了 Actor-Critic 模型的解剖和计算视角。他们评估了几个模型, 并得出结论, 它们与当前的解剖数据不兼容。尽管这些模型在计算上很优秀, 但 Critic 对其生物学上的不真实性提出了不同的看法, 理由是:

◎多巴胺细胞增加了对厌恶性(*aversive*)、中性(*neutral*)和奖励刺激(*rewarding stimuli*)的激发, 因此不能唯一地发出奖励预测误差的信号。(Pennartz, 1996, 然而, 参考 Ungless 等人, 2004, 声称, 对厌恶性刺激作出反应的神经元实际上不是多巴胺能)

◎多巴胺细胞在刺激完全被感知之前就启动了, 因此当时无法知道其产生的作用 (Redgrave, Gurney 和 Prescott, 1999b)。

◎刺激奖励学习(*stimulus-reward learning*)可以在没有多巴胺的情况下进行, 因此对于学习来多巴胺的作用说是不必要的(Berridge, 2007)。

受动物行为研究的启发, 经典机器学习理论将术语“学习”分为两部分类型。在第一种类型中, 学习元素, 即“智能体(*Agent*)”, 伴随着一个无所不知的要素, 即“教师”, 它告诉它在每种情况下要采取的正确行动。这种学习形式被称为“监督学习(*supervised learning*)”, 通常由一组例子(任务)及其各自的答案(行动)组成。当将其应用于神经网络时, 通常会实现反向传播体系结构。在这里, 给出了一个例子, 网络根据其先前的知识作出响应。然后, 提供正确的答案, 并将其与网络的输出进行比较。这种比较用于学习: 每个单元的输出和所需输出之间的差异表明应以相反的方式逐层更改哪些连接(Lippmann, 1987)。第二种类型被称为“无监督学习(*unsupervised learning*)”, 因为它缺少“教师”要素。在这种情况下, 只使用输入, 并根据网络动态进行分类。这种学习通常应用于旨在发现输入统计规律的任务。这些系统是自我组织的(*self-organizing*)(Linsker, 1988; Kohonen, 1995), 通过调整权重(*synaptic efficacies*), 通常运用 Hebb 的局部学习规则(Hebb, 1949; Churchland 和 Sejnowski, 1992)。

这两种学习模式在脑功能建模中的应用遇到了一些可行性问题。最明显的与监督学习有关, 因为将全知“教师”引入网络在生物学上是不现实的 (Churchland 和 Sejnowski, 1992)。在另一个极端, 经典的无监督学习(例如原始 *hopfield* 网络(hopfield, 1982)或主成分分析(PCA)网络(Oja, 1982))对于一组有限但毫无疑问重要的精心选择的问题很有用。然而, 根据定义, 这种学习模式永远不会发现与输入的统计结构不相符的关键信息 (Churchland 和 Sejnowski, 1992)。

作为强化的奖励学习行为的现象已经被广泛地建模(Barto, 1995; Sutton, Barto, 1998)。几乎所有将学习作用归因于 BG 的模型都是基于某种强化学习形式(一种指定如何通过序列结束时提供的奖励信号来学习一系列动作的方法, )。最近的实验证据表明, BG 可能介导强化学习的各个方面(Hollerman 和 Schultz, 1998)。在一个典型的调节任务中, 经过多次试验, 多巴胺神经元失去了对主要的奖励刺激的反应, 同时获得了对条件性刺激的反应。这个观察结果导致采用了一种工程起源的强化学习算法, 叫做时序差分(*Temporal Difference, TD*)学习算法, 作为对 BG 学习的描述。

Barto(1995)描述了 TD 算法在 Actor-Critic 模型中的应用, 该模型已适用于 BG 解剖、生理学和细胞化学的某些方面(Houk, Adams 和 Barto, 1995)。Actor-Critic 模型分为两部分: Actor 和 Critic(图 2)。Critic 向 Actor 提供了一个“有效强化”的信号, 然后运用这个信号学习对环境



境的行为反应。Actor 的目标是学习生成或选择最大化主要奖励的操作。然而，Actor 从即时反馈中学习。如果 Actor 直接从环境中获得初级强化(primary Reinforcement)，并且强化的到来在导致强化的行为之后延迟(通常是这种情况)，Actor 将无法将潜在的行为与奖励联系起来。这个问题叫做时序差分问题。Critic 的作用是学会预测主要的奖励并产生一个信号(称为“有效的强化”)，该信号可被 Actor 用作强化，以便 Actor 获得有效的奖励，与刚开始的行动相关联。因此，这里有两个学习系统：一个是学习执行最大化奖励的行为的 Actor，另一个是学习根据正在执行的行为和当前环境状态预测奖励的 Critic。

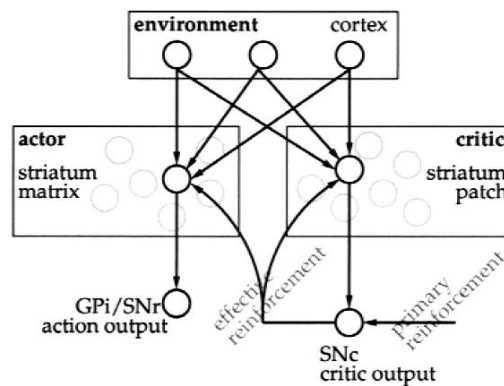


图 40: Actor-Critic 模型与可能的神经生物学实施相关显示。皮质纹状体连接可能具有兴奋性或抑制性影响。Actor 通过 GPI/SNr 的输出单元生成操作。Critic 预测这些行动应该产生的主要强化。预测的强化被用作强化或削弱皮质纹状体连接的奖励信号。

有人提出，Actor 的神经实现与纹状体基质室有关，并且 Critic 的神经实现与纹状斑块室和黑质致密部多巴胺神经元(SNc)有关。来自 SNc 的多巴胺信号起到有效或预测强化的作用(Houk, Adams 和 Barto,1995)。在多巴胺神经元中已经确定了这种作用的特性(Hollerman 和 Schultz,1998)。在上面提到的经典条件反射任务中，多巴胺反应趋向于最早的“奖励预测因子(predictor of reward)”(Mireniewicz 和 Schultz,1994)(例如条件刺激)，它提供了对初级奖赏的类似于 Critic 的预测。尽管图 2 所示的机制通过直接途径起作用，但涉及苍白球外段(GPe)和底丘脑核(STN)的间接途径被提议介导多巴胺反应从初级强化到预测因子的转移(Houk, Adams 和 Barto,1995)。

这种强化学习模式已经在 BG 结构的背景下以各种方式进行了探索。许多类似的模型研究了 SNc 多巴胺神经元反应的类型，如果它们确实产生了有效的强化信号，就会产生预期的反应(Montague,Dayan 和 Sejnowski,1996;Schultz,Dayan 和 Montague,1997)。多巴胺神经元的预测特性包括当预期的奖励未能在预期的时间发生时，就会出现负反应(Montague,Dayan 和 Sejnowski,1996)。类似地，一旦多巴胺反应移到了许多感官线索中最早的一个，每个感官线索都会预测奖赏，移除其中一个中间线索将导致在线索的预期时间内多巴胺反应的抑制(Schultz,Dayan 和 Montague,1997)。已经观察到这两个预测的实验结果(Schultz,1998)。这些模型也做了一些假设。大多数强化学习模型假设感官线索的时序表示。在线索的皮质表示中，有一个时序成分允许从线索开始到奖励的时间的显式表示。这种对纹状体的提示输入表示决不是确定的。此外，模型还需要一个主要的奖励信号输入到黑体多巴胺神经元。这种信号的来源尚不清楚。

Suri 和 Schultz(1998)展示了 Actor-Critic 模型学习模式关联序列的能力。例如，刺激 A 需要动作 Q，之后刺激 B 需要动作 R，依此类推。只有在执行序列中的最终所需操作并且序列中所有之前的操作关联都正确时，才会收到奖励。该模型说明了时序差分算法如何产生长关联模式的学习。他们进一步证明，用初级奖励本身取代有效的奖励信号，即初级奖励的预测，会导致学习长序列关联的失败。这说明了在长序列学习中预测奖励信号的价值。

黑质多巴胺神经元不仅对奖赏有反应，而且对新奇或意外的刺激也有反应 (Schultz,1998)。因此，在这些模型的背景下，新奇的刺激被认为是有益的事件。意外的视觉刺激会导致早期多巴胺反应和眼跳 (Ljungberg, Apicella 和 Schultz,1992)。Redgrave 等人 (1999b)认为，由于多巴胺反应发生在刺激视觉位置的眼跳完成之前，因此在视觉识别之前，此类事件不应视为奖励。然而，由于多巴胺神经元反应不区分不同的奖励方式，Schultz(1998)建议多巴胺反应可能代表奖励警报信号，而不是奖励本身。看来，“Critic”可能需要以这种方式产生偏见，以促进新刺激的关联和对奖励的预测。

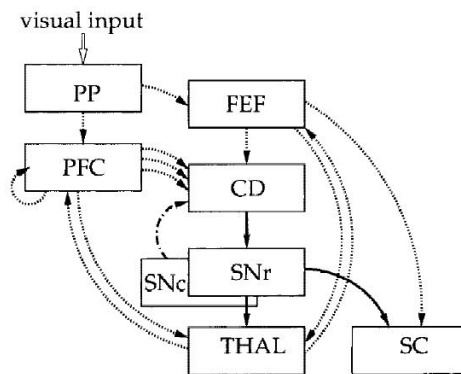


图 41: Dominey 等人序列学习模型中连接的示意图。14 条实线表示抑制连接;虚线表示兴奋性连接。点划线表示多巴胺能投射。PP, 后顶叶皮层;FEF, 前额视野;PFC,前额叶皮层;SC,上丘。

有少量的 *BG* 的强化学习模型不是基于时序差分算法的 (Arbib 和 Dominey,1995; Dominey,1995;Dominey 和 Arbib,1995)。这些模型集中在 *BG* 电路上，使用一种更简单的强化学习形式。Dominey 等人(1995)在皮质、*BG* 和丘脑相互作用的模型中证明了这一点。该模型考虑了视觉空间任务需要的连续动作。图 3 说明了该模型中模拟的电路。该系统由当前视野的各种皮质表示组成。在这个模型中，眼动的控制可以通过一个从额叶视野(FEF)到上丘(SC)的直接投射来实现。*BG* 对所学的皮质背景事件活动进行调节。例如，在一个条件视觉辨别任务中，模型必须学会生成两个提示中的一个(只有一个得到奖励)，*BG* 通过强化学习来解除对上丘中正确反应的抑制(Dominey 和 Arbib,1995)。在学习感觉运动序列时，设计了前额皮质结构模型，使序列中的每个点都具有独特的活动模式。这为纹状体神经元把序列中的一项与紧接着的下一项区分开来提供了一个新的背景。

强化学习已经成为许多当代 *BG* 模型的主要组成部分。然而，这些相同的模型突出了许多有待解决的问题。例如，对于间接途径的作用还没有明确的共识。强化学习范式的简单性和优雅性更容易与作为实现站点的直接路径联系在一起。间接通路可以更好地描述为一个存在复杂动态变化可能性的间接网络(Parent 和 Cicchetti,1998;Parent 和 Cicchetti,1998b)，其在强化学习模型中的作用尚不清楚。

乍一看，强调学习在多巴胺在 *BG* 功能中的作用的模型与临床数据相矛盾。学习与帕金森病明显的神经系统问题有什么关系？当然，没有人认为不能从扶手椅上站起来或者在路中间冻僵是一个学习的问题。解决这一难题的一个办法是在纹状体中赋予多巴胺两个角色。其中一个作用是阶段性的，需要黑质细胞的突然活跃，也许 D1 受体的活动参与了学习。另一个作用是强化剂，低水平的多巴胺是多巴胺细胞静息放电水平的结果。左旋多巴可能对 D2 受体作用更大，复制了这种作用。缺乏这种强直性活动可能是神经病理学的原因。多巴胺的这种双重功能作用在 Dominey 等人的模型中得到了明确的实现。

一个更优雅的方案是假设低多巴胺水平的存在是一个有效的增强信号。例如，在没

有奖励呈现的情况下，但是为期望奖励(即，预测奖励)时，会抑制多巴胺的活动。根据 TD 解释，这是一个误差信号，表明出现了比预期更少的回报。因此，导致这一结果的行动在未来将不太容易被选择。在帕金森病的情况下，这样的解释将把每一个皮质纹状体的选择都标记为产生比预期更少的回报。所有的动作都不太可能被选择，最终正确的反应是没有动作。这种错误比简单的学习损失更为严重，并且与 TD 模型一致。

### 5.1. 强化学习的一般原则(General principles of Reinforcement learning)

这两种学习方法的结合，通常放在监督学习的大标题下，是强化学习。强化学习领域通常涉及具有明确目标的“智能体”对环境的作用。“智能体”的行为改变了环境的状态，环境反过来提供了对其行为的反馈(奖励或惩罚)。在该方案中，外部奖励作为评价信号，反映了网络性能的适宜性程度。因此，一方面，强化学习是监督学习的一种形式，因为网络接收和使用来自环境的反馈信息。然而，该信息是一个标量值，因此是可评估的，而不是有指导意义的(Hertz 等人,1994)。评估信号要么作为一个单一的信息位给出：对或错，或者在连续的情况下，作为一个描述正确程度的值。正确的答案本身对“Actor”来说仍然是未知的(不像监督学习)。因此，强化学习有时被称为“弱监督学习”(Pennartz 等人,2000)。

简言之，强化学习系统除了包括“智能体”和环境之外还包括许多概念(Sutton 和 Barto,1998)：“策略(policy)”描述了“智能体”在各种环境状态下的行为，即一组刺激反应规则或关联。奖励函数将每个状态(或状态动作对)映射到一个数字上，即奖励，指示该状态的即时可取性。值函数具体说明了状态的长期可取性，并考虑到可能遵循的状态及其可获得的奖励。最后，几乎所有的强化学习方法都是围绕着估计值函数来构建的;换句话说，它们试图估计在给定状态下，或者在给定状态下，对某个策略执行给定操作有多好。最后，一些强化学习系统涉及环境的内部模型(Suri,2002)。

各种强化学习方法背后的基本思想源于这样一个概念：根据一条被称为“桑代克效应定律(Thorndike law of effect)”(Thorndike,1911)的规则，如果一个行为改善了行为人的条件，那么产生这种行为的倾向就会加强。根据此规则学习是基于“生成和测试(Generate and test)”方案(Barto,1995)。在该方案中，生成、测试备选方案，并最终将行为导向更好的备选方案(或备选行动的顺序)。强化学习系统的目标是最大化预期收益，这是后续奖励序列的一个函数。如果时间被分为离散的事件，每一个事件都由若干个状态组成，那么它们的总和就是有效的。在有限的情况下，折扣系列的奖励应该最大化。由于强化信号没有提示正确答案应该是什么，但只给出最后一个答案(或一组答案)的接近程度，因此强化学习网络必须包含随机性源，以便能够探索可能的输出集，直到找到最佳输出。这一概念体现了强化学习策略中的一个关键要素：探索与运用：“智能体”人必须运用其已经知道的东西来获得奖励，但也必须探索，以便在未来获得更好的绩效(Sutton 和 Barto,1998)。

在动物和人类的学习以及机器学习中，奖励通常是相对于带来奖励的行为而延迟的。因此，Actor 在收到关于响应正确性的信息之前，必须执行一系列操作(或经历一系列状态)。因此，加强学习的一个常见的困难是暂时的学分分配。“Actor”必须能够将功劳和责任分别分配给序列中的每个动作。因此，时间学分分配对于生物强化学习算法至关重要。

### 5.2. BG 功能的非 RL 观点

尽管图 4 和图 5 所示的 BG 的功能结构与大脑其他关键区域的关系本身就很有趣，也很有说服力，但必须指出的是，当代文献对 BG 功能的观点存在着根本的不同。随着新的实验数据的获得，有人试图添加新的连接，甚至模块到现有的 BG 架构。这样的添加挑战了 BG 功能的早期概念。他们似乎要求彻底修改这个问题，如果可能的话，将新数据纳入现有的框

架。

将 *BG* 模型分为严格 *RL* 模型和广义 *RL* 模型两大类是有帮助的。Schultz 等人(1997)的模型属于严格的 *RL*，其中多巴胺信号被识别为 TD 误差。在基于更普遍、更广泛的 *RL* 概念的模型中，多巴胺只是表示某种形式的奖励、显著性或错误信息。甚至 Prescott、Redgrave、Gurney 等人的模型(Redgrave 等人,1999a, b; Prescott 等人,2002)，其中多巴胺信号表示动作的显著性，并用于动作选择，可以解释为属于这一广泛的 *RL* 范畴。事实上，从目前的情况来看，这种对 *BG* 功能的更广泛看法似乎更接近事实真相。现在假设严格的基于 *RL* 的方法可以解决所有问题还为时过早，因为这种解释并非没有缺点。

虽然 Schultz(1998)最初的电生理实验已经被重复了几次，但它们只在一个实验室中进行过。很明显，其他实验室需要重复这些实验。甚至在建模领域，也有几次尝试扩展多巴胺信号与 TD 误差相等的主题。最近的一个模型提出了强直性 DA 活动和阶段性 DA 活动的不同作用，前者与长期奖励预测有关，后者与试验特异性预测有关(Daw 和 Touretzky,2000)。另一种观点认为，相位 DA 激发与预测误差有关，紧张性 DA 活性成分与预测的不确定性有关(Fiorillo 等人,2003)。

因此，可以对 *BG* 功能给出一个更具包容性的概念总结如下：

*BG* 将多巴胺信号所携带的奖励或显著性或误差相关信息与感觉运动皮层状态相结合，为运动皮层提供适当的纠正。

还有其他一些 *BG* 模型甚至不属于这个更广泛的 *RL* 的范畴。例如，Leblois 等人(2006)将 *BG* 及其相关结构组织成两个回路：第一个回路由“纹状体-皮质-GPi-丘脑-皮质”组成，他们将其称为传统的直接通路；第二个回路由“皮质-STN-GPi-丘脑-皮质”组成，他们将其称为超直接通路。作者认为这两个环路之间的竞争负责动作选择。它们还显示了多巴胺的适度减少是如何导致单配对动作选择的，而更大的多巴胺缺乏会产生振荡。在另一项激进的提议中，Obeso 等人(2006)将 GPe 置于中心位置，皮质-STR-GPi 和皮质-STN-GPi 为两条环绕 GPe 的平行通路(图 6)。作者主要通过实验研究来支持他们的架构，没有提到强化学习。多巴胺信号在 *BG* 功能中的作用也没有明确的描述。看看这样的架构是否可以解释 *BG* 的 *RL* 功能将是一个不小的挑战。

在对 *BG* 功能进行了如此多的重新思考之后，即使经过近 30 年对建模的深入研究，这一课题似乎比以往任何时候都更具挑战性。然而，在这些相互矛盾的建议中，一个基本的观点是 *BG* 是一个皮层下结构的网络，它通过处理中脑多巴胺能信号中与奖赏相关的信息，对皮层控制结构中的决策产生强烈影响。所有激烈的争论都是关于在特定的 *BG* 核中发生的详细信息处理的性质。但是，俗话说，细节决定成败。

### 5.3. 评估值函数和策略(estimating the value function and policy)

系统通过评估每个备选方案的价值来选择策略。一个状态，或者更一般地说，一个状态的价值被定义为从该状态开始并遵循未来所有状态的固定策略时所获得的奖励之和。给定一个精确的值函数，提取最优策略就变成了一个简单的任务。针对这一点的一般解决方案是重复迭代，根据给定的策略改进状态值的估计，然后根据值估计的变化改进策略(Sutton 和 Barto,1998)。因此，强化学习算法的第一个目标是为每个状态确定最佳值函数的估计，从而使该估计的误差为零。

如果遵循强化学习的三个基本理念，价值近似的目标是可以实现的(Barto,1994)。首先，在整个学习过程中，“智能体”的策略必须保持固定。第二，为了使近似成为可能，必须假设

环境状况趋向于重现，因此近似是基于对相同(或类似)情况的回忆。最后，状态/结果组合的可变性可以用概率论来克服。因此，一个状态的价值实际上可以被未来回报的预期总和所取代，即所有可能的未来情景的(加权)平均值。

### 5.3.1. 动态规划(dynamic programming)

假设学习系统对环境动态有充分的了解，可以应用称为动态规划的算法。在这些算法中，*值函数*的逼近自然可以用值迭代算法来实现。这种方法假定情况会再次发生，并且其结果是恒定的。通过对状态空间进行扫描，根据当前的近似误差更新每个状态值的近似值，可以确定最优*值函数*。这些迭代将继续进行，直到不需要更改(即近似误差为零)。接下来，执行策略改进步骤，在该步骤中，如果发现某个操作将改进状态值，则会更改当前策略。重复这一过程，直到没有实质性的改善(Sutton 和 Barto,1998)。

本节讨论了确定性过程的简单化场景。显然，并不是所有的问题都可以归为这样。当给定状态下的给定动作产生一组潜在的后继状态(具有已知的概率分布函数)时，该过程称为非确定性。在这种情况下，值迭代将寻求最大预期值。这个期望值是在给定操作之后，在所有可能的后续状态的值之间计算出来的。动态规划只有在充分了解系统动力学的情况下才可能实现，因此不适合对生物系统进行建模。

### 5.3.2. 蒙特卡罗方法(Monte Carlo methods, MC)

动态规划方法有两个明显的缺点：首先，它们需要一个完整的环境模型，而这并不总是可用的。第二，即使存在这样的模型，计算预期收益仍然是非常复杂的，特别是在大的状态空间中。蒙特卡罗方法克服了这两个问题。在 MC 方法中，价值估计和策略改进的过程是基于样本经验，而不是基于(未知的)全概率分布。基本上，使用 MC 方法的算法平均在访问每个状态后观察到的完整*奖励*。如果执行了许多迭代(即在任意步骤中遇到状态的许多不同事件)，该平均值将收敛到预期的累积未来*奖励*，这正是状态值(Sutton 和 Barto,1998)。

蒙特卡罗方法的学习并不是真正的在线学习：学习系统必须等待事件的完成，而不是一步一步地更新。因此，它只是部分递增的。变化发生在事件之间(或整个任务序列的迭代)，而不是一步一步地(Sutton 和 Barto,1998)。

### 5.3.3. 时序差分法(Temporal difference methods, TD)

*时序差分*(TD)模型是一类处理预先定义的状态序列过程的未来收益预测任务的模型。这对应于一个经典的条件作用范式，在这个范式中，“*智能体*”必须预测与前面事件相关的*奖励*。TD 方法是当今最广泛使用的*强化学习方法*，部分原因是它们将蒙特卡罗方法的一些优点与动态规划方法相结合。它们和蒙特卡罗方法一样使用无模型方法，即*直接从经验中学习*。但是，与蒙特卡罗方法不同，它们更新的估计部分基于以前学习的估计(如动态规划)。在每个步骤中，状态值与随后的(即时)*奖励*加上下一个状态的估计值进行比较。此比较的结果称为 TD 误差，用于更新状态值。因此，与蒙特卡罗不同，TD 自然是在线实现的，因为它是真正的增量：而蒙特卡罗方法必须等到事件的最终状态更新状态值，而在 TD 方法中，只有一个步骤必须通过(Sutton 和 Barto,1998)。尽管 TD 模型计算预测信号而不是选择最佳行动(Suri,2002)，但在下面内容(第 4.3.1 节)中讨论了决策过程的扩展。

### 5.3.4. TD( $\lambda$ )-桥接蒙特卡罗方法和时序差分法

TD 和 MC 方法实际上是一个连续体上的两个极端。它们的区别在于，在一段情节中，为了*奖励*而检查的步骤的数量，与那些使用未来状态估计来估计的步骤的数量不同。如前所

述, TD 学习使用子序列状态的估计值, 等待一个单独的步骤。MC 学习不使用任何其他状态估计。事实上, 可以使用实际奖励状态估算的任何组合(Sutton 和 Barto,1998)。TD( $\lambda$ )是指这样一种组合: 根据衰减因子( $0 < \lambda < 1$ ), 对每个未来状态的奖励和估计进行平均。或者, 可以用相反的方式来看待它: 每个过去的状态都有一个与之相关的衰退的内存跟踪。全局 TD 误差信号触发对所有最近访问状态的比例更新, 根据 $\lambda$ : TD(0)是经典的时序差分法, TD(1)对应蒙特卡罗。

### 5.3.5.探索 and 运用平衡(Balancing exploration and exploitation)

动物必须在不确定、未知的环境中觅食。有两种神经系统被认为可以调节探索性行为, 一种基于海马, 另一种基于 BG (Packard 和 Knowlton,2002)。海马体在大鼠空间导航能力中的作用已被广泛研究。几项研究报告了大鼠海马中存在的位置细胞, 当动物位于熟悉的空间环境中的某个位置时, 这些细胞会被激活(Packard 等人,1989)。在一项实验中, 老鼠被训练从一个固定的起始位置走到 T 形迷宫中有诱饵的端口处(例如指向西方的端口)(Packard 和 McGaugh,1996)。经过几天的训练, 迷宫旋转 180°。目的是观察动物是否表现出位置学习的迹象, 并朝着相同的绝对方向(西方), 或表现出刺激反应(S-R)学习, 使相同的身体转向并朝着错误的绝对方向(东方)。换句话说, 在位置学习中, 动物在视觉空间坐标系中表示目标; 在 S-R 学习中的编码是在就身体位置而言的运动坐标。经过 8 天的训练, 大多数动物表现出位置学习, 但经过 16 天的训练, 大多数动物表现出 S-R 形式的学习。第 16 天背外侧纹状体功能受损, 使动物表现出与 S-R 学习相背的位置学习。因此, 即使保留了位置学习, 动物也会在长时间的训练中从位置学习策略切换到 S-R 策略。这些和其他实验表明, BG 参与了 S-R 形式的导航。

研究了多巴胺能活动对蜜蜂觅食行为的影响(Scheiner 等人,2002)。Montague 等人(1995)提出了一种基于预测的 Hebb 学习和多巴胺神经元的蜜蜂觅食模型。在这个模型中, 模拟的蜜蜂遵循一个简单的觅食策略: 如果多巴胺信号足够高(高于基线), 有机体就会保持航向, 否则(多巴胺细胞活动低于基线), 有机体就会随机摇摆, 这是一种被称为“变向无定位运动”的运动形式。然而, 是什么神经机制驱动有机体的随机转向? 这个问题没有答案(Montague 等人,1995)。虽然在 RL 框架中, “运用”和“探索”同样重要, 但是关于 BG 在 RL 中的作用的文献, 似乎主要集中在奖励信号——它的化学信使, 它的解剖位置, 它在学习中的结果等等, 而只是对探索进行了总结。人们常说, SNc 和/或 VTA 的多巴胺能细胞的活动表明奖赏(Montague 等人,1995)。但是 BG 的哪一部分产生探索所需的随机信号呢?

Daw 等人(2006)也注意到, 虽然对“运用”的解剖位置有相当清楚的认识, 但“探索”的神经底物相对不清楚。Daw 等人(2006)在一项基于成像的研究中, 研究对象玩四臂强盗游戏, 这是一种“探索”与“运用”很好地融合在一起的 RL 游戏, 他们观察到, 当受试者进行探索活动时, 额皮质和顶叶沟优先活跃。刚才提到的两个皮层区域可能是支持探索性决策的皮层区域。但是, 由于其他与 BG 功能相关的 RL 成分已经在 SNc 的 BG 电路——TD 误差、纹状体的 Critic 等中被识别出来, 因此没有相应的 BG 核(或细胞核)被提及作为探索行为的皮质下基础, 这是令人惊讶的。

我们之前曾提出, BG 电路中的 STN-GPe 环路被正确地放置在支持探索行为的位置(Sridharan 等人,2006)。在 RL 模型中, “探索”总是由随机分量建模。从生物学的角度来说, 只有能够产生复杂的尖峰活动的神经系统才有资格成为“Explorer”。STN-GPe 层形成了兴奋-抑制神经元层对, 以其在完整 BG 中的复杂的尖峰活动而闻名, 在多巴胺缺乏或帕金森条件下观察到这种复杂活动的丧失(Nini 等人,1995; Bergman 等人,1994; Brown 等人,2001)。

蒙特卡罗方法和时间差分方法都提出了“充分探索”的问题。这些方法只有在遇到所有状态时才能工作，至少在大量迭代之后。此外，特别是在无模型方法中，估计动作值而不是状态值会很有用。Q-学习算法(Watkins 和 Dayan,1992;Jaakkola 等人,1994)为这两个问题提供了解决方案：它将值(称为 Q 值)分配给状态/动作对。Q 值是在行动和未来给定策略之后收到的(折扣系数 $\gamma$ )强化的总和。因此，一个最优 Q 值是在采取行动和在未来使用最优策略之后得到的强化总数。在实践中，期望的 Q 值是由一个后续状态的单个随机样本估计的。根据所使用的精确算法，Q 学习可被视为 TD 学习或 MC 学习的非策略方法(在一种情况下，当行为在一种策略下，另一种不同的政策被估计(Sutton 和 Barto,1998))。

#### 5.4.强化学习架构(Reinforcement learning architectures)

虽然强化学习的概念不受计算机或生理结构的限制，但人工智能和控制理论专家对动物经典和操作反射作用理论的应用产生了许多具有强大计算能力的学习结构。

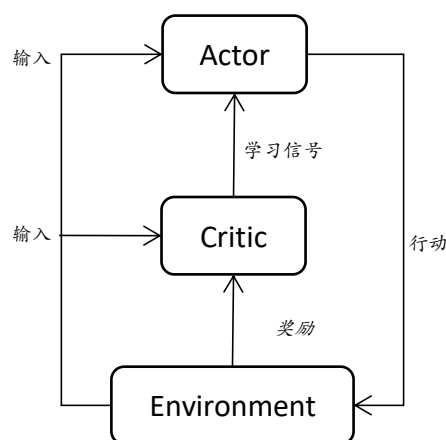
##### 5.4.1.“Actor-Critic”体系结构

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，例如操作条件反射(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 42: 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)。



“Actor-Critic”体系结构为强化学习提供了一个解决时间信用分配问题(the temporal credit assignment problem)的方案(Tesaroo,1994;Barto,1995)。在这个体系结构中，一个独立的单元(Critic)从“智能体”(Actor)接收的环境中的所有信息，即输入和强化，以及与 Actor 输出



相关的信息。Critic 的目的是预测环境产生的强化作用。一旦这个模型单元被训练，它就可以用来计算每个状态的值：因为 Actor 应该尝试最大化累积强化，而不是立即奖励，所以可以估计未来强化的加权平均值，并逐步反馈给 Actor。用于预测强化并将其预测和真实强化之间的差异反馈给 Actor 的网络(或算法)称为自适应 Critic(*adaptive Critic*)(Sutton,1988)，整个系统称为“Actor-Critic”。解决了时间信用分配问题，基于模型对未来强化的预测，为 Actor 提供了即时的评价反馈。因此，一个以最大化即时有效强化为目标的 Actor 实际上是按照最大化未来奖励的某些功能的战略目标行事(Barto,1995)。这也使得“Actor-Critic”架构特别适合学习动作序列(Suri,2002)。“适应性”一词强调这样一个事实，即 Critic(根据其预期)将即将到来的奖励调整为即时的评价反馈。显然，Critic 需要通过经验来改进(Sutton,1988;Barto 等人,1989)。自适应 Critic 使用 TD 误差(见上文)更新其权重：如果 Critic 在相邻时间步骤中的预测已更改，则自适应 Critic 应更改。

事实上，Critic 提供给 Actor 的输出(有效强化)通常也与 TD 误差相同(Barto,1995)。如果考虑到 Actor 的学习目标，这是合理的：已经反复证明，动物和人类的奖励依赖性学习取决于奖励的不可预测程度(Rescorla 和 Wagner,1972)。如果对感官刺激(动作)的反应具有预期的结果(TD 误差=0)，则该反应趋势应保持不变。但是，如果结果比预期更好(TD 误差为正)，或更糟(TD 误差为负)，则应相应地加强或减弱响应。

“Actor-Critic”架构被认为是 BG 功能的基本机制(Houk 等人,1995)。根据该模型，增强信号由多巴胺能神经元产生，这是来自主要增强源(如下丘脑外侧)的输入和来自纹体内 MSNs 的输入的结果。纹状体 MSNs 和多巴胺能神经元之间的相互循环被称为“纹状小体模块”，实际上是作为一个自适应的 Critic 来计算 Actor 的误差信号。另一方面，基质中的 MSNs 通过苍白球和丘脑将信息传递到额叶皮质，从而形成“基质模块”，作为“Actor-Critic”架构的 Actor 部分。

#### 5.4.2.BG 强化学习的 Actor-Critic 模型

##### 5.4.2.1.Houk,Adams 和 Barto(1995)

Houk 及其同事(Houk,1995a;Houk 等人,1995)提出了第一个模型，表明 BG 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。该模型表明，纹状小体(Striosomal)模块充分发挥了适应性 Critic 的主要功能，而周围基质(*matrix*)模块则发挥了 Actor 的作用。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。

纹状小体模块由纹状体的纹状小体、下丘脑核和在黑质致密部(SNc)中的多巴胺能神经元组成。根据该模型，三个输入源相互作用，产生 DA 神经元的激发模式。其中两个输入来自纹状体中的纹状小体，并提供有关预测强化的刺激发生的信息。一个是对 SNc 的直接输入，它会提供长期的抑制，另一个是间接输入，通过提供相位刺激的下丘脑核传导到 DA 神经元。对 DA 神经元的第三个输入，假设是由下丘脑外侧产生的，也是兴奋性的，并提供有关初级奖励发生的信息。在获取过程中，纹状体的纹状小体神经元通过 DA-依赖的皮质纹状体突触增强，学会在刺激预测未来初级强化发生时突然迸发。学习后，奖励预测刺激的呈现将导致由纹状小体间接激发产生的 DA 突然迸发。预期初级奖赏的到来不会导致 DA 反应，因为由纹状小体引起的长期直接抑制会抵消由下丘脑外侧引起的兴奋。根据预测误差的 TD

方程，在 TD 方程中的 *初级强化* 等于对 DA 神经元的 *初级强化*，未来强化的预测  $P(t)$  等于对 DA 神经元的 *间接兴奋输入*，*直接抑制输入* 等于在较早期时间步长的预测  $P(t-1)$ 。

Houk 等人的 Critic 模型不包括精确的时间机制，而是对 DA 神经元的缓慢而持久的抑制。因此，当预期奖励被忽略时，它不能解释 DA 活动的定时抑制。这个问题已经在以后的模型中通过使用网络输入的不同表示来解决。“完整的连续复合刺激”(Montague、Dayan 和 Sejnowski,1996)是刺激的一种表现形式，在刺激呈现期间和呈现之后的一段时间内，刺激的每一个时间点都具有不同的激活成分。一般来说，假设一个刺激的呈现引发了大量的时间表征，学习规则可以选择合适的，也就是说，对应于 *刺激-奖励间隔* 的那些。后面描述的模型使用这个计算原理，但是描述了这个一般解决方案的不同神经实现。

与 Critic 的详细讨论相反，Houk 等人仅提供在 BG 电路中实施该 Actor 的总体方案。根据他们的模型，由 *纹状体基质*、*视丘下核*、*苍白球*、*丘脑* 和 *额叶皮质* 组成的 *基质模块*，产生指挥各种动作的信号，或代表组织其他系统产生实际指挥信号的计划。然而，他们注意到，从感官的角度来看，*基质模块* 产生的信号可能是显著情境发生的信号(另见第 5 节)。

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 *BG 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 BG 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)。

#### 5.4.2.2. Berns 和 Sejnowski(1996)模型

Berns 和 Sejnowski(1996)提出了一个 Actor-Critic 模型，该模型假定 BG 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 *动作选择*。然而，在这个模型中，*动作选择* 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 BG 和 PFC 之间的相互作用，采用时间差分学习和监督学习(Leabra)算法相结合的方法对模型进行训练。该模型假设 BG 辅助动作选择。这个模型没有模拟信息进入 WM 的闸门，尽管同一作者早期的工作假设 BG 辅助信息进入 WM。在这项工作之后，其中一种模型将 BG 在运动节和信息闸门两方面的作用纳入 WM(Moustafa 和 Maida,2007)。

最近的一些模型采用了强化学习方法来模拟 BG 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底体的功能。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 BG 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 *动作选择机制* 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)。

#### 5.4.2.3.Suri 和 Schultz(1998、1999)

Suri 和 Schultz 扩展了 Barto(1995)提出的基础 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，而基底体有利于行为学习。每个隔室使用一个单层网络进行模拟。与基底体和纹状小体的皮层连接是完全连接的。该模型假设在基底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，大概模拟基底体单位之间的横向抑制，选择活性最高的单位。每个基底体单位对应一个行动。

Suri 和 Schultz(1998)模型既提供了 Actor 的神经模型，又修改了刺激表示的 TD 算法，以便在遗漏奖励时重现 DA 活动的定时抑制。时间机制是通过使用一组神经元来表示每个刺激来实现的，每个神经元都被激活了不同的持续时间(而不是在 Barto 的模型中单一的延长的抑制)。Critic 学习规则被修改，以确保只有包含实际刺激-奖励间隔的刺激表示组件的权重被调整，而其他神经元的权重保持不变。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。这些模型允许模型再现 DA 神经元的固定模式，以奖励-预测刺激、预测奖励和忽略奖励(Suri 和 Schultz,1998)。在基础模型的改进中(Suri 和 Schultz,1999)，教学信号进一步丰富，以更好地确定 DA 神经元对新刺激反应的相关生物学数据。

这些模型中的 Actor 由一层神经元组成，每个神经元代表一种特定的 Actor。根据 Critic

提供的预测误差信号,学习“刺激-动作”对。赢者通吃(winner-take-all)规则,可以通过横向抑制神经元之间,确保只有一个行动是选定在一个给定的时间。Suri 和 Schultz(1998、1999)运用这一 Critic 的改进和扩展模型证明,即使是一个简单的 Actor 网络也能有效地解决相对复杂的行为任务。然而,尽管这些作者承认 Actor-Critic 结构和基底节结构之间的一般相似性,并认为时间刺激表征的成分可能与纹状体和皮质神经元的持续活动相对应,但没有试图在已知的 BG 的结构中实现 Critic。此外,TD 算法的扩展包括新颖性响应、泛化响应和奖励预测中的一些时间方面,是通过任意指定模型的特定参数值(例如,用特定值初始化特定突触权重,对不同的突触使用不同的学习率)来实现的。而不是通过在与 BG 解剖和生理学相关的神经网络中更具生物合理性的实现。Contreras-Vidal 和 Schultz(1999)也曾尝试过这种方法。

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 PFCBG 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 BG 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。运用具有噪声阈值的 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。

#### 5.4.2.4. Contreras Vidal 和 Schultz(1999)

Contreras Vidal 和 Schultz(1999)提供了一种与 BG 解剖相关的神经网络结构, 该结构可通过合并一个最初由 Carpenter 和 Grossberg(1987)开发的附加自适应共振神经网络, 解释 DA 对食欲刺激和厌恶刺激的新颖性、归纳和辨别的反应。他们进一步指出, 奖励预测误差有两种类型: 一种表示奖励预测时间误差的信号, 可能与 TD 模型有关; 另一种表示奖励预测类型和数量误差的信号编码, 可能与自适应共振网络有关。虽然对该网络的描述超出了本文的范围, 但我们将简要讨论它们在遗漏奖励时对 DA 活动抑制负责的定时机制的实现。与 Suri 和 Schultz(1998、1999)相似, Contreras Vidal 和 Schultz 假设, 纹状小体神经元产生一系列时间信号, 以响应感官输入(刺激的“完整序列复合”表示)。然而, 在他们的模型中, 与 Suri 和 Schultz 假定的不同持续时间的持续活动相比, 在刺激开始后和有限的一段时间内, 纹状小体神经元被连续激活。在 Suri 和 Schultz 的模型中, 学习规则确保了在初级奖赏传递(即结合 DA 活动)时, 纹状小体神经元的突触活动得到加强, 但在 Contreras Vidal 和 Schultz 的模型中, 纹状小体是纹状体黑质(striatonigral), 而不是皮质纹状体的突触, 这些突触被认为是由学习改变的。(值得注意的是, 尽管有足够的证据表明皮质纹状体突触具有长期可塑性, 但纹状体黑质突触却没有这种证据。)学习后, 由于纹状小体产生的时间性抑制, 通过预测的初级奖励来消除 DA 神经元的兴奋。重要的是, 与 Barto(1995)提出的基于 Critic 总体方案的模型相比, 在该模型中, DA 神经元的激发源被假定为不同于抑制源。因此, 对奖赏预测刺激的阶段性 DA 反应归因于来自前额叶皮质(PFC)的激发, 并通过纹状体基质和黑质网状部分(SNr)传导到 DA 神经元。

#### 5.4.2.5. Brown、Bullock 和 Grossberg(1999)

Brown 等人(1999)提供了另一种尝试来回答什么生物学机制计算 DA 对奖励和奖励预测刺激的反应。与 Contreras Vidal 和 Schultz(1999)相似, 这些作者认为条件刺激的快速兴奋反应和对有回报的无条件刺激反应的延迟、自适应定时抑制是由不同的解剖路径决定的。当预测奖励被忽略时, DA 对预测奖励的抑制和 DA 活性的降低依赖于从背侧和腹侧纹状体的纹状小体到 SNc 的适应性定时抑制投射。然而, 与 Contreras Vidal 和 Schultz(1999)相比, 刺激开始后, 纹状小体神经元的连续迸发取决于细胞内钙依赖性的时间机制。在早期的模型中, 同时发生的纹状小体神经元的尖峰和 DA 突然激发(对主要奖赏的反应)导致活跃的纹状小体神经元上的皮质纹状体突触增强。因此, 选择了一个在预期的奖励发放时间内有效的纹状小体群体, 从而防止 DA 对预期奖励的反应。DA 神经元对奖赏和奖赏预测刺激的激活归因于从脑桥被盖网状核(PPN)到 SNc 的兴奋性投射。DA 激活的阶段性特征被认为是由于投射到 SNc 上的 PPN 神经元的习惯化或调节。

#### 5.4.2.6. Suri、Bargas 和 Arbib(2001)

在最近的一篇文章中, Suri 等人(2001)使用扩展的 TD 模型扩展了 Suri 和 Schultz(1998、1999)使用的 Actor-Critic 模型, Actor 是基于对 BG-丘脑皮质电路的解剖以及 Critic 和 Actor 之间的复杂交互作用。与 Suri 和 Schultz(1998、1999)中的 Actor 类似, 纹状体层中的每个模型神经元被认为对应于能够引发动作的少量纹状体基质神经元。然而, 确保在给定时间内只选择一个动作的机制取决于连接纹状体和 BG 输出核的直接和间接途径之间的相互作用, 以及皮质层的赢者通吃规则。在这个模型中, DA 通过三种依赖于纹状体神经元的膜电位影响 Actor 的作用: 皮质激素传递的长期适应, 以及对纹状体神经元兴奋率和奖励持续时间的短暂影响。Critic 就像在早期的模型中一样接收感官和奖励信息, 此外, 还从 Actor 的丘脑和皮质水平接收有关预期和实际动作的信息。因此, Critic 可以学习刺激——奖励和行动刺激的关联。

Suri 等人表明, 这种扩展的 Actor-Critic 模型能够进行感觉运动学习, Suri 和 Schultz(1998,1999)使用的原始 Actor-Critic 模型也是如此。此外, 该模型还具有规划能力, 即能够形成新的关联链, 并根据这些关联链预测的结果选择其作用。该模型中的规划关键取决于这样一个事实: 对扩展 Critic 的输入包括对未来刺激的预测和有关预期行动的信息(由丘脑提供), 这些信息可用于估计未来预测信号, 并且 Critic 在每个行动步骤中运行两次迭代。总之, 这些特征使我们能够根据在一个 Actor、该 Actor 的感官结果和奖励之间形成的新的关联链来评估预期的 Actor。

Suri 等人还模拟了 DA 神经元的新反应, 即在遇到新刺激时纹状体 DA 的短暂增加。这种新颖的反应增加了纹状体神经元在向上状态下发生反应的可能性, 从而增加了动作的可能性, 从而产生了探索 Actor。DA 神经元的新反应是通过初始权重选择来实现的, 有效地等同于为新的位置/刺激分配较为乐观的初始值。探索 Actor 也源于模型中纹状体神经元奖励的随机转换。下面我们描述另一种控制勘探和开发之间权衡的机制, 这是武装强盗情况(armed bandit situations)的特征。

### 5.5. 强化学习的演变——对 Critic 建模的另一种方法

我们(Niv,Joel,Meilgson 和 Ruppin,2002)采用了一种替代的方法来模拟 RLCritic。运用进化计算技术, 对一个简单的蜜蜂采蜜决策神经网络模型的神经元学习规则进行了进化。为此, 我们形成了一个非常通用的学习规则进化框架, 其中包括所有异突触的赫布型学习规则(Hebbian learning rule), 也允许突触可塑性的神经调节。运用遗传算法(Genetic algorithm), 蜜蜂在不断变化的环境中, 根据其交配能力进行进化。由于环境的不确定性, 有效的觅食只能由有效的 RL 产生, 因此形成了有效的 RL 机制。

为了避免可能混淆的术语, 我们在异突触可塑性的概念(Dittman,Reehr,1997;Skaar,Wu,Sun,1997;Vogt,Nicoll,1999)和神经调节可塑性(Bailey,Giustetto,Huang,Hawkins 和 Kandel,2000;Fely 和 Lister,1998)之间做一个区分。与传统的单突触赫布型学习不同, 异突触赫布型学习允许对突触进行独立于活动的调节, 这样, 当只有突触前或突触后成分被激活时, 突触也可以被更新, 更普遍地说, 即使两者都没有被激活。我们称之为“异突触”模式, 因为它允许一个神经元的激活影响其所有突触, 而不管与之相连的其他神经元的活动如何。突触可塑性的神经调节通过在学习过程中允许一个三因素的相互作用进一步增强了学习规则: 通过神经调节一个神经元的活动可以影响另两个神经元之间突触的可塑性。在神经组织中已经证明了突触可塑性的异突触可塑性和神经调节门控(Bailey 等人,2000;Dittman 和 Regehr,1997;Fellous 和 Linster,1998;Schacher 等人,1997;Vogt 和 Nicoll,1999), 并且已经被认为增加了突触学习的计算复杂性(Bailey 等人,2000;Fellous 和 linster,1998;Wickens 和 Kotter,1995)。通过考虑到异突触学习和可塑性的神经调节, 我们建立了一个很大的搜索空间, 在这个空间中, 进化算法可以搜索最优的突触学习规则。在我们的模型框架内, 我们表明只有一个网络结构能够产生有效的 RL 和以上的随机觅食 Actor。进化网络类似于 Montague,Dayan,Person 和 Sejnowski(1995)早期提出的一种结构, 它由一个感官输入模块、一个奖励输入模块和一个输出单元 P 组成, 该模块编码了感官输入中随时间变化的信息。进化的学习规则确实是异突触的。结合突触可塑性的神经调节(详细描述见 Niv 等人(2002)出版; [http://www.cns.tau.ac.il/t\\_yaeln/adaptivebehavior2002.htm](http://www.cns.tau.ac.il/t_yaeln/adaptivebehavior2002.htm))。

学习机制的演变与适应 Critic 密切相关, 与输出单元的活动和突触可塑性的神经调节有关。类似于 Montague 等人(1995), 模型单位 P 的输出非常准确地捕捉了灵长类和啮齿动物中脑多巴胺能神经元的活动模式的本质(Montague 等人,1996;Schultz、Dayan 和 Montague,1997), 以及蜜蜂中相应的章鱼胺能神经元(Hammer,1997;Menzel 和 Muller,1996)。

由于在进化的网络中，突触权重代表预期回报，输入代表感官输入随时间的变化，网络的输出代表后续时间步骤中预期回报之间的持续比较。在 Critic 模型中，这种比较提供了误差度量，通过它网络更新其权重并学习更好地预测未来的回报。

关于神经调节，这项研究表明，有效的 RL 在很大程度上取决于突触可塑性的神经调节的演变，也就是说，通过第三个神经元的活动来控制两个神经元之间的突触可塑性(一个‘三因素’的赫布型学习规则)。这与皮质纹状体突触中描述的 DA 依赖性可塑性相似(Calabresi 等人,2000;Wickens 等人,1996)。这一学习规则对 RL 的计算最佳性的证明有助于计算模型的尝试，以在 BG-丘脑皮质系统的复杂解剖和生理学之间架起桥梁，以及损伤和影像学研究的结果，这些研究涉及到该系统在程序或刺激反应学习中的作用。

与 Actor-Critic 模型通常采用的单突触学习规则不同，我们衍化出的异突触学习规则能够在突触前或突触后成分(或两者)未被激活的情况下调节突触。这使得不同刺激预测的奖励之间可以进行非平凡的互动。例如，一个刺激预测的奖励量可以根据面对不同刺激时遇到的失望或惊讶进行调整，即使执行了另一个响应，执行某个响应的趋势也会发生变化。在该模型中，这些微观水平的异突触可塑性动态直接导致了觅食 Actor 的勘探与开发特征之间的宏观权衡。小脑(Dittman 和 Regehr,1997)和海马(Vogt 和 Nicoll,1999)突触的证据表明，大脑中确实存在异突触可塑性，但纹状体中尚未证实这种现象。除了 Suri 等人(2001)提出的机制外，这种机制还可以提供另一种控制探索的纹状体内机制。

我们的模型主要反映了 Actor-Critic 框架中的 Critic 模块，并且只包含一个极其简单的 Actor。为了提高模型在 BG 学习中的相关性，并允许更详细地描述该计算模型如何在基底节电路中实现，需要进一步的工作，重点阐述模型的 Actor 部分。

## 5.6.BG 中的 Critic 网络——一个讨论

从模型的早期描述可以明显看出，一个类似 Critic 的功能被纹状小体与 DA 系统的连接所替代。然而，只有三项研究(Brown 等人,1999;Contreras Vidal 和 Schultz,1999;Houk 等人,1995)试图基于这些连接的已知解剖和生理学提供 Critic 的神经网络模型。Brown 等人(1999)和 Contreras-Vidal,Schultz(1999)对这些模型进行了总体比较，尤其是与定时机制的实施有关。在这里，我们要集中讨论两个问题：(1)是否有解剖学基础来支持一致的观点，即纹状小体在 Critic 中起着关键作用？(2)当遇到一个奖励预测刺激时，做 DA 神经元的激发，当一个预测奖励被忽略时，这些神经元的抑制来自一个来源(如 Houk 等人(1995)提出的建议，并在 Suri 和同事的不同模型中有所暗示)。或者他们是来自两个具有不同特征的不同来源(如 Brown 等人,1999;Contreras-Vidal 和 Schultz,1999 所建议的)？

## 5.7.DA 信号在 Actor-Critic 中的作用

### 5.7.1.DA 神经元的兴奋和抑制源

除了 Contreras-Vidal 和 Schultz(1999)提出的模型外，我们所回顾的所有模型都是基于 Barto(1995)的 Critic 架构。在这种结构中，预测误差的计算取决于一个神经元或一组神经元通过奖励预测刺激的激活。这导致 DA 神经元的快速兴奋和延迟抑制(分别对应于 Barto 模型中的  $P(t)$ 和 $-P(t-1)$ )。由于这些模型中的大多数假设激发和抑制源位于纹状小体，因此假设存在从纹状小体到携带这些信号的 DA 系统的解剖路径，如 Houk 等人的模型(见上文)所述。我们已经讨论了假设纹状小体直接抑制整个 DA 系统的问题。然而，Houk 等人的模型在假设从纹状小体经下丘脑核到 DA 系统存在间接途径方面遇到了另一个困难，因为当前的解剖学数据表明，到下丘脑核(通过苍白球)的纹状体投射来自基质神经元而不是来自纹状小体神



神经元(回顾见 Gerfen,1992)。因此,纹状小体不太可能为 DA 神经元提供快速的刺激。

纹状体(不一定是纹状小体)神经元可能是 DA 神经元早期兴奋性和晚期抑制性输入的来源吗?电生理数据(审查见 Bunney、Chiodo 和 Grace,1991;Kalivas,1993;Pucak 和 Grace,1994)和解剖数据(审查见 Haber 等人,2000;Joel 和 Weiner,2000)确实表明,背侧和腹侧纹状体的神经元活动既可以直接抑制 DA 细胞活动,也可以直接地促进 DA 细胞分裂。然而,直接抑制效应可能先于间接兴奋效应,间接兴奋效应由至少两个抑制性突触介导(例如,腹侧纹状体投射到 VP 的 GABA 氨基丁酸能神经元,后者投射到大多数 DA 系统)。这意味着 DA 系统接收到的信号是  $P(t-1)-P(t)$  而不是  $P(t)-P(t-1)$ 。当然,这预测了 DA 神经元与观察到的相反的活动模式。例如,当遇到奖励预测刺激时,它会抑制而不是激发 DA 活性。除时间问题外,纹状体背侧神经元的不同亚群也可能产生抑制和促进作用。关于腹侧纹状体,是否腹侧纹状体神经元投射到 VP 与直接投射到 DA 细胞的神经元不同仍然是一个悬而未决的问题(见 Joel 和 Weiner,2000)。综上所述,单个纹状体神经元群不太可能是大多数 Critic 模型所要求的对 DA 神经元的间接快速刺激和直接延迟抑制的来源。

这种双重输入到 DA 系统的另一个来源是边缘 PFC。Schultz(1998)认为,来自这个皮质区域的输入可能负责 DA 神经元对奖励和奖励刺激的兴奋反应。边缘 PFC 中的神经元对初级奖励和奖励预测刺激作出反应,并在期望奖励期间表现出持续的活动(回顾见 Schultz、Tremblay 和 Hollerman,1998;Zald 和 Kim,2001),大鼠的数据表明边缘 PFC 直接向 DA 神经元投射(回顾见 Overton 和 Clark,1997)。除了边缘(腹侧)纹状体外,边缘 PFC 项目(Berendse、Galís de Graaf 和 Groenewegen,1992;Groenewegen、Berendse、Wolters 和 Lohman,1990;Parent,1990;Uylings 和 van Eden,1990;Yeterian 和 Pandya,1991)。通过后一种途径,边缘 PFC 可以对 DA 神经元产生延迟抑制作用。这与电生理学证据一致,即边缘纹状体中的神经元显示奖赏相关活动,包括期望奖赏和奖赏预测刺激期间的持续活动(Rolls 和 Johnstone,1992;Schultz,Apicella,Scarnati 和 Ljungberg,1992)。在边缘 PFC 和边缘纹状体中持续活动的神经元的确定符合 Suri 和 Schultz(1998,1999)的 Critic 模型中实施的时间机制。如前所述,在他们的模型中,涵盖实际“刺激-奖励”间隔的刺激表示组件的持续活动负责对奖励预测刺激的阶段性 DA 响应,“刺激-奖励”间隔期间 DA 响应不足,以及预期奖励被忽略时 DA 活动的减缓。假设边缘 PFC 的神经元提供了时间持续的活动,它们的直接投射可以提供未来增强  $P(t)$  的时间  $t$ ; 它们的间接投射,通过边缘纹状体,可以提供延迟的预测。然而,我们要注意的,尽管上述建议遵从已知的解剖结构,但它不包含对 DA 细胞的其他重要投射,这些投射可能在 DA 信号的产生中起作用,尤其是来自苍白边缘的投射。

Brown 等人(1999)的模型中也发现了边缘 PFC 是 DA 神经元早期兴奋和晚期抑制的来源的假设。然而,在他们的模型中,皮质神经元的持续活动转化为 DA 神经元的阶段性反应(即,对奖赏预测刺激的反应增加,对预测奖赏的遗漏的反应减少),是通过携带刺激和抑制信号的路径的特定特性来实现的。因此,PPN 的习惯化(这是为其模型中的 DA 神经元提供兴奋性输入的通路中的最后一个站点)确保 DA 神经元在预测奖赏刺激后只接收到短暂的刺激,并且纹状小体中的细胞内自适应计时机制将持续的皮层活动转化为在预期回报时短暂和定时抑制 DA 细胞。

我们想通过结论来结束这一节,尽管 BG 与 DA 系统的连接被认为具有“Critic 式”功能,但基底节连接中基本 Critic 模型的当前实现基于与已知的这些核解剖不一致的假设。希望将来在已知的神经回路中实现这类模型的尝试能够揭示 BG 的功能,并为理论模型提供额外的约束。

### 5.7.2.DA 信号在强化学习和 Actor 转换中的双重作用

在本文中,我们将 DA 对奖励和奖励预测刺激的反应作为增强信号。这一假设是 DA 在学习起核心作用的一个证明(Le Moal 和 Simon,1991;Robbins 和 Everitt,1996; White,1997)。DA 系统的另一个核心功能是在不同 Actor 之间切换(Le Moal 和 Simon,1991;Lyons 和 Robbins,1975;Odes,1985;Robbins 和 Everitt,1982;van den Bos 和 Cools,1989; Weiner,1990)。最近,Redgrave 等人(1999)指出奖励刺激不仅有助于强化他们之前的 Actor,而且还可以打断他们的 Actor,并引发不同的 Actor(例如,在奖励发放之后从杠杆转向接近食品杂志)。基于这一观察结果,作者认为,短潜伏期 DA 对奖励和奖励预测刺激的反应是观察转换而不是学习。

相反,基于条件刺激在增强和转换中的双重作用,Weiner 和 Joel(2002)认为 DA 神经元的相位反应参与了学习和转换。他们进一步指出,这两种功能分别因纹状体 DA 的阶段性增加对皮质纹状体突触传递的长期和短暂影响而失效。因此,RL 在 DA 增加之前被激活的纹状体神经元的皮质纹状体突触的 DA 依赖性增强所替代,而 Actor 转换则被 DA 介导的皮质纹状体传递的促进和减弱所替代,这促进了一组神经元纹状体活动的改变。积极参与不同的一组(参见 Weiner 和 Joel,2002,了解可能构成这些影响的细胞机制)。尽管 Suri 等人(2001)的模拟结果与 Actor 切换问题没有直接联系,但对这一假设的一些支持可以在其模拟结果中找到。Suri 等人(2001)的模型将 DA 对纹状体神经元的长期和短暂影响结合起来。正如所料,这些作者发现前者对于 RL 是必要的。Suri 等人他们还发现,在他们的模型中,DA 的阶段性增加导致了 Actor 输出的增加,并且这种影响是由 DA 对纹状体纤维化的短暂影响所介导的。

在奖励事件的双重作用,即指导学习和促进 Actor 转换的背景下,我们想指出,在学习过程中,条件刺激失去了前者的作用,但不是后者。因此,随着学习的进展,每一个条件刺激都会被先前的刺激和 Actor 所预测,因此丧失了诱导阶段性 DA 反应的能力,从而丧失了支持学习的能力。然而,在学习过程中,由于强化驱动的刺激反应学习,每一个条件刺激都成为目标导向 Actor 下一步行动的诱导者。因此,在所学动作序列的执行过程中,每一个动作都会导致条件刺激的发生,而条件刺激又会在该序列中引发以下动作。

由此可知,条件刺激可能通过至少两种不同的机制引发转换。一种机制依赖于纹状体 DA 的阶段性增加,并且是新情况和早期学习阶段的特征。这种机制要么增加了一般情况下转换的可能性,要么倾向于转换到具有新情况特征(例如定向)的一类 Actor(主要是先天的)。另一种机制依赖于皮质纹状体突触的加强,并且具有学习良好的 Actor 特征。该机制负责终止当前 Actor,并启动后续 Actor,这是具体的和可被学习的(Weiner 和 Joel,2002)。尽管后一种类型的转换发生在纹状体 DA 没有阶段性增加的情况下,但基线 DA 水平被认为在运动开始中起着重要的许可作用(Le Moal 和 Simon,1991;Robbins 和 Everitt,1996; Salamone,1994)。

我们(Joel、Avisar 和 Doljansky,2001)最近在大鼠身上获得了证据,表明 DA 也能调节条件刺激终止先前 Actor 的能力。

上面回顾的模型都没有模拟两种类型的切换。然而,在 Suri 和 Schultz(1998)的模拟中,可以发现逐步获取和失去获取 DA 信号的能力,同时也获得了获取‘相位 DA-独立’切换的能力。在他们对 Actor-Critic 模型获取连续动作的模拟中,奖励发生在正确执行的刺激动作对序列的末尾。在任务获取过程中,每个不同的刺激逐渐获得了诱发 DA 信号和触发正确动作的能力。随着训练的进行,刺激被早期的刺激所预测,结果停止诱发 DA 信号。然而,由于 Actor 的学习,每个刺激继续触发正确的行动。因此,在学习之后,刺激的呈现导致正确的动作的激发,而不增加 DA。

### 5.7.3.纹状小体与适应的 Critic

纹状体的纹状小体隔间内与 DA 系统之间的联系是由 Charles R. Gerfen 的工作引起的，他发现在大鼠中，纹状体背面的纹状小体与一个相对较小的 DA 神经元群之间存在相互联系，这些神经元群位于 SNc 的腹侧部分和 SNr 中(Gerfen,1984,1985;Gerfen、Herkenham 和 Thibault,1987)。目前灵长类动物的数据表明，一组 DA 神经元可能与纹状体背侧的神经元相互连接。然而，没有证据表明这些纹状体神经元的区域起源(见 Joel 和 Weiner,2000)。因此，在灵长类动物中，解剖证据不支持在纹状小体神经元与 DA 系统的连接中实施 Critic。即使只考虑大鼠的解剖证据，这种实现也只能解释一个相对较小的 DA 神经元群的活动。

在不同的模型中，是否还有另一组纹状体神经元可以取代“纹状小体”呢？或者说，是否有一组纹状体神经元，它们与整个 DA 系统有相互联系？最近对灵长类纹状体和 DA 系统之间的联系的两项解剖学数据(Haber、Fudge 和 McFarland,2000;Joel 和 Weiner,2000)和大鼠(Joel 和 Weiner,2000)的荟萃分析得出结论，纹状体和 DA 系统之间的联系的一个重要特征是不对称而不是相互作用。也就是说，边缘(腹侧)纹状体投射到大多数 DA 系统，但由一个相对较小的 DA 神经元亚群支配，而运动纹状体(主要是壳核)则相反，后者由一个比其投射到的 DA 系统更大的区域支配。这种组织的结果是，边缘纹状体使 DA 输入相互作用，使投射到结合核(主要是尾状核)和运动纹状体的 DA 神经元受到神经支配；结合纹状体使 DA 输入的一部分相互作用，使投射到运动纹状体的 DA 神经元受到神经支配，运动纹状体使其 DA 输入(Haber 等人,2000p;Joel 和 Weiner,2000)的一部分相互作用。基于这一组织，这两篇论文的作者提出，除了这些连接在电路内处理中的作用外，纹状体-DA-纹状体的连接可能在 BG-丘脑皮质电路之间的信息传递中起重要作用。

我们的结论是，一个建立在 DA 神经元和另一组神经元之间相互连接的 Critic，不能在 DA 系统和纹状体之间的连接中实现。然而，由于腹侧纹状体(和腹侧苍白质(VP)，见下文)对 DA 系统提供了一个主要的抑制投射，许多腹侧纹状体神经元的活动与奖励和奖励预测刺激有关，这一结构可能是导致 DA 神经元活动模式的机制的一部分。未来的研究有望揭示纹状体和 DA 系统之间连接的地形组织在 BG 计算中的作用。

### 5.7.4.结论

我们对 BG 的 Actor-Critic 模型的选择性回顾提出了几个问题，我们相信未来的模型将不得不处理这些问题。Critic 的模型建立在 DA 神经元活性与 Critic 的 TD 预测误差信号强相似的基础上。从计算的角度来看，这些模型面临着两个相关的挑战：①如何重现奖励的特定时间动态、奖励预测刺激和新颖性。②将 DA 对新颖性、归纳和识别的响应合并到 TD RL 算法中会产生什么计算结果？

从解剖学和生理学的角度来看，很明显，一个建立在 DA 神经元和另一组神经元之间相互连接的 Critic 模型不能在 DA 系统和纹状体之间的连接中实现，因为这些连接的特点是不对称而不是互惠。同样，基于 Barto(1995)架构的 Critic 也不能在这些连接中实现，因为正如 Critic 的模型所要求的那样，不太可能单个纹状体神经元群是间接快速兴奋和直接延迟抑制 DA 神经元的来源。解决这些问题的一个潜在的有效方法是运用进化计算技术的力量，找到在各种解剖和功能约束下最大限度地提高 Critic 功能的候选解决方案架构，然后通过实验检验这些预测。Niv 等人(2002 年出版)的工作，是朝这个方向迈出的第一步。Critic 的未来模型将不得不处理这些问题，此外，还应该涉及到一个问题，即对 DA 系统的单一投射(例如来自 BG)是否对 DA 神经元对奖励和新刺激的反应负责，或者这些反应是否被不同的投射所替代(如由 Contreras Vidal 和 Schultz 于 1999 所建议)。

Actor 的模型建立在纹状体中依赖 DA 的长期突触可塑性与在 Actor 中预测误差信号引导的学习之间的强相似性上。然而，目前的 Actor 模型非常简单，通常在纹状体层次上建模，细节很少。未来研究的目的是以更详细和可靠的方式模拟已知的基底节解剖和生理，并解决 BG-丘脑皮质连接的计算作用问题。目前，这些连接的几种不同神经网络模型为这些问题提供了不同的答案(Berns 和 Sejnowski,1998;Gurney 等人,2001)。我们已经描述了 BG-丘脑皮质连接的模型，这表明 BG 对皮质输入进行了增强偏倚的降维(Bar Gad 等人,2000)。该 RDDR 框架可作为未来 BGActor 模型的基础。

总之，BG 的 Actor-Critic 模型通过将 BG 处理的一些核心方面(DA 信号、纹状体中的 DA 依赖性学习)与学习理论结合起来，有助于我们对 BG 功能的思考。然而，关于这些核的功能以及 RL 的理论方面的许多问题仍未得到解答。我们希望未来的模型，包括 Actor 和 Critic 的组成部分，更受已知的解剖和生理 BG 将回答这些问题。

### 5.8.内部模型(Internal models)

Actor 对环境的完整描述对于规划非常有用(Sutton 和 Barto,1998;Suri,2002)。在需要提前数步计算最优性能和评估复杂的假定环境未来状态的值的条件下，作为预测器的计算单元可以提供很大的帮助。该模型不仅可以预测未来的强化，还可以预测整个状态空间。内部模型方法模拟未来的走法，并使用假设的未来结果来选择最佳走法。该方法已被用于对纹状体和皮层的多巴胺神经元活动和预期神经活动进行建模，以便在需要计划的情况下做出正确的动作选择(Suri 等人,2001)。

## 六、降维(Dimensionality reduction)

降维描述了将输入从高维数据空间投影到低维空间的过程(Haykin,1999)。神经系统中的维数降低可以描述为将大量神经元群编码的信息压缩为数量明显较少的神经元。当原始空间中包含的全部或大部分信息被保留时，就可以实现有效的减少。例如，群体编码经常与神经编码相关(Georgopoulos 等人,1986)。这种编码包括结合神经元群的活动来编码一个特定的变量(如肢体的运动角度)。提取基础变量可能会取代保持或处理整个群体完整和详细活动的需要。因此，我们可以将数据从一个大的数量等于神经元编码数量(数量规模是数据的表面维度)到一个神经元(占数据的内在维度)编码相关的变量(角)的放电率。另一个例子可以从感官世界中得出(Barlow,1989,1992b)。在视觉系统中，由大量视网膜细胞活动组成的输入信号被转换成一个简单细胞的活动信号，这个简单细胞代表皮层 V1 区的一条线。将大量的相关信息转化为最少数量的活动细胞(Barlow,1992a)的想法被认为是大脑信息编码的主要机制。

### 6.1.动机(Motivation)

大脑面临的主要任务之一是使机体能够在连续状态之间进行转换。这种转变可以简化为生物体当前状态和下一步行动之间的转变。对即将到来的动作的计算需要来自不同形式的信息，这些信息存储在大脑的许多不同区域，处理感觉、运动、边缘和认知信息。然而，在将所有这些信息传递到涉及规划的区域(如额叶皮质)的过程中，会出现多个问题。这些问题的根源在于需要传递大量的信息，它们与在完全不同的领域中遇到的问题惊人地相似，即在试图连接地理上分散的信息源时遇到的广泛传播问题。将  $n$  个源与  $m$  个目的端连接起来可能会成为阻碍信息流动的一个巨大问题。这个问题可以分为四个主要的子问题：

- 1、通信线路数：源端和目的端之间必须有  $n \cdot m$  个通信线路。

- 2、通信端点数：每个目的端必须创建  $n$  个连接点，共有  $n \cdot m$  个连接。
- 3、通信灵活性：源点或目的端的任何变化，如新端点、端点的移动或连接的变化都要求通信线路和端点的物理变化。
- 4、通信调制：对信息流的任何调制，如过滤、增强等，必须对每个连接分别进行。

同样的问题在中枢神经系统领域更为重要，中枢神经系统的物理尺寸和可用于创建和维护中枢神经系统的资源自然受到限制。例如，在人类皮质中，109-1010 个神经元 (Shepherd,1998) 的数量级编码有关当前状态的信息必须与额叶皮质中大约相同数量级的神经元相连 (Abeles,1991;Braitenberg 和 Schuz,1991)。通过 109 个传入轴突连接额叶皮质，在每个神经元上建立 109 个突触似乎是不可行的。一个典型的大脑皮层神经元只有 103-104 个突触 (Abeles,1991)，这些突触构成了它接收到的所有信息。此外，在这样一个分散的系统中不可能进行中央调制，使得信息流的任何中央控制机制都不起作用。我们建议中枢神经系统使用一种类似于通信系统的解决方案，即中央开关从大量源元素接收所有信息。交换机通过对数据进行优先级排序和压缩，并通过有限的通道将其传输到目的端，对信息进行编码。

除了连通性问题，大脑的分布式编码不可避免地会导致一个主要的计算问题。“维度灾难(*curse of dimensionality*)”(Belman,1961)，是一个术语，用于描述由高维数据引起的一组计算问题，描述某一空间所需的样本数呈指数增长。神经网络需要与输入空间的超容量成比例的资源，而输入空间的超容量随其维数呈指数增长。在生物学世界中，这意味着处理行动计划和选择的神经元需要大量不切实际的例子来学习从当前状态到特定行动的正确映射。通过将输入映射到包含输入关键特征的低维空间，可以减少从状态到动作的有用映射所需输入的数量。

## 6.2.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  $17 \times 10^6$  个皮质神经元投射到  $17 \times 10^6$ ，会使聚率达到 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。前馈权重采用赫布 (Hebb) 学习法更新，层间横向连接采用反赫布 (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)。

运用这种网络结构，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。输出层活动的非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。

### 6.3.不同降维方法

存在大量的 *降维* 方法, 但 *降维* 的原因各不相同。此外, 根据学习和处理系统设置的约束, 为同一目的执行的 *降维* 可能使用不同的方法。

#### 6.3.1.数据和降维(Data and dimensionality reduction)

*降维* 可以通过两种基本方法实现: 减少数据的维数, 或者减少数据本身。在 *降维* 过程中, 原始数据被嵌入到低维空间中。然而, 在数据简化过程中, 数据与表示它们的有限数量的模式相关联。尽管诸如 PCA 之类的 *降维* 方法通常试图维护与数据相关的特定信息标准, 但诸如矢量量化(VQ)和聚类之类的数据 *降维* 方法则试图消除输入的可变性, 并将其转换为给定的一组可能的输出之一。本文的重点是使信息丢失最小化的 *降维* 方法, 而不是在丢失大部分信息的同时将数据编码成特定模式的方法。

#### 6.3.2.基于监督(Supervised)、无监督(unsupervised)和强化的降维

*降维* 的类型是从学习“正确/对的”约简的方法中推导出来的。在有监督的学习中, “教师”在学习阶段提供最佳的 *降维*。因此, 有监督的算法学习模仿这样一个“老师”所取得的结果。由于缺乏 *教学* 信号, 这些方法在生物学上通常不可信。另一个极端是, 无监督的学习算法不使用任何“教师”。*无监督* 学习算法依靠数据集本身的性质来学习最优约简。重要的是要注意, 有监督和无监督的 *降维* 方法可以达到相同的结果。然而, 得出这一结果的方法将有很大的不同; 例如, PCA 算法可通过监督神经网络(BouRLard 和 Kamp,1988)或非监督神经网络(Foldiak,1990;Kung 和 Diamantaras,1990)实现。

#### 6.3.3.局部和全局降维(Local and global dimensionality reduction)

*降维* 方法之间的一个主要区别是它们感知输入之间相互作用的方式。全局 *降维* 方法(如经典 PCA)假设所有输入共享一个全局相关方案。因此, 所有的输入都可以简化为一个更小的子空间。另一方面, 局部 *降维* 假设数据子集之间存在局部交互, 从而导致每个子集的最佳 *降维* 不同(Chakrabarti 和 Mehrotra, 2000)。当原始输入空间包含几个非常不同的子结构时, 这种方法可以有效地进行约简。局部结构以单独的方式减少, 这是局部最优的, 不考虑整个输入空间的全局条件。

#### 6.3.4.线性和非线性降维(Linear and non-linear dimensionality reduction)

*降维* 方法传统上分为线性方法和非线性方法。线性方法试图找到一个子空间来表示输入的线性组合。更一般的非线性方法试图找到表示输入非线性组合的子空间(Oja 等人,1995;Malthouse 等人,1995)。线性方法通常具有简单和易于实现的优点。然而, 它们的成功完全取决于它们的基本假设, 即存在一个很好的输入线性组合表示。当情况不是这样时,

例如来自球体的输入，则缩减将完全失败。另一方面，非线性方法往往更复杂，但能够更紧凑地表示更复杂的输入分布。

#### 6.4.主成分分析(Principal component analysis, PCA)

主成分分析是一种典型的无监督线性全局降维方法。但是，有一些变体使用监督(BouRLard 和 Kamp,1988)、非线性(Oja 等人,1995;Malthouse 等人,1995)或局部(Chakrabarti 和 Mehrotra,2000)降维方法。虽然主成分分析并非是最有力或最通用的降维方法，但其余的研究将集中在这一方法上。选择 PCA 的原因是它的简单性，事实上它已经被广泛研究，并且它有一个生物学上可信的神经网络实现。

PCA 是一个数学过程，也被称为协方差矩阵的离散 Karhunen-Loève 变换(Karhunen-Loève transform, KLT)和奇异值分解(SVD)。PCA 将  $n$  维输入投影到  $m(m < n)$  正交轴上，其中包含原始输入的最大方差。由于 PCA 具有信息保存特性，因此它主要是有用的。对于给定的  $m$  维约简，PCA 是导致输入输出重构误差最小的线性变换(Gerbrands,1981)，也是保存“香农信息(Shannon information)”最大量的变换(Linsker,1988)。

在过去的 20 年里，用于执行 PCA 的神经网络模型有了很大的发展。由 Oja(1982)出版的最早的著作是基于线性单一单元网络的学习规则。该网络仅提取输入数据的第一个主要成分。神经元的学习规则是基于一个修改过的 Hebb 规则，因此被称为“规范化 Hebb 学习规则”。在标准的 Hebb 规则中(Hebb,1949)，如果输入和输出神经元具有正相关的放电率，则它们的连接强度(权重)会增加;或者，如果它们倾向于以负相关的方式放电，则它们的连接强度会降低。这一标准的 Hebb 规则往往会导致神经元之间连接强度的无限增长。基于这个原因，引入了一个衰减因子，该衰减因子从连接的值减少到一个标准化的 Hebb 学习规则。这种标准化的 Hebb 规则的生物学合理性已经得到证明(Friston 等人,1993)。这种学习过程导致权重收敛到第一个主分量，并通过使用完全局部算法保持输入空间的最大方差。

第一个用于提取多个单一成分的神经网络模型是广义 Hebb 算法(GHA)(Sanger,1989)和随机梯度算法(SGA)(Oja 和 Karhunen,1985)。这些模型将 Oja 规则推广到多单元输出阶段。在这些模型中，有  $n$  个输入神经元， $m$  个输出神经元( $m < n$ )，学习规则导致提取数据的第一个主要成分。提取只使用输入和输出之间的前馈权值，通过引入一个对单个神经元不可用的全局量来执行。这一全局数量显著降低了生物系统神经网络模型的合理性。

Foldiak 扩展了 Oja 的模型，通过抑制连接对称连接输出神经元(Foldiak,1989,1990)。神经元的输出是通过减去由前馈网络产生的活动的横向权重传递的活动来获得的。前馈权值采用与 Oja 学习规则相似的 Hebb 规则训练，抑制性侧权值采用反 Hebb 学习规则训练。反 Hebb 学习规则导致输入和输出神经元结合在一起时连接减少，反之则连接增强。因此，当两个神经元的放电相关时，抑制性侧连接变得更强(更负)。与 Sanger 的模型一样，该模型提取输入数据的  $m$  个主要成分。该模型中的横向权值代替了全局量，并作为输出层神经元之间信息传递的局部机制，使它们对输入的不同分量进行编码。需要注意的是，在该模型中，网络提取包含相同信息的主子空间，但是前馈权重本身不包含主子空间。Kung 和 Diamantaras(以及早期的模型(Rubner 和 Tavan,1989;Rubner 和 Schulten,1990))设计的模型使用输出层神经元之间的分层横向连接(Kung 和 Diamantaras,1990;Diamantaras 和 Kung,1996)，其中神经元  $i$  与神经元  $i+1$  相连。在这个模型中，前馈权值的训练采用和 Oja 相似的 Hebb 规则，抑制侧权值的训练采用反 Hebb 学习规则。该网络收敛到一个解决方案，其中横向权重为零，前馈权重编码第一个  $m$  主分量。第  $i$  个输出神经元的权重等于输入的第  $i$  个特征向量，第  $i$  个神经元的活动等于输入对第  $i$  个主分量的投影。最重要的是，Foldiak 提出的模型和 Kung 和



Dimantaras 提出的模型没有使用任何全局信息，因此保持了它们的生物合理性。

## 七、BG 基础强化驱动降维模型

基础强化驱动降维模型(Bar-Gad 等人,2000)是对基于神经网络的主成分分析模型的扩展。两种不同的无监督神经网络是模型的基础。第一个网络基于 Foldiak(1989)使用对称神经网络的工作，即同一层中的每对神经元通过相互连接而连接。第二个网络使用 Kung 和 Diamantaras(1990)运用的 APEX 网络，该网络使用不对称神经网络，即对于同一层中的每对神经元，只有一个神经元影响另一个。这种类型的连接在某些实现中也称为分层连接。这两种网络都使用输入层和输出层之间的全(all-to-all)连接，并运用相同的基本机制来实现神经元的活动和学习。两个网络中的神经元活动函数都是线性的，即输出是前馈输入的加权和减去横向输入的加权和。学习规则是规范化的 Hebb 规则(Oja,1982)对前向权值的扩展，以及对横向网络的反 Hebb 规则的扩展。

RDDR 模型为两个网络的无监督学习添加了一个强化因子。该增强因子与神经元的输入和输出相乘，形成一种“多 Hebb 学习”算法。在“多 Hebb 学习”规则中，增强信号调节给定输入/输出对的突触连接权重的变化量(Reynolds 和 Wickens,2002)。这一规律受到复杂的皮质-纹状体-多巴胺能突触三联体的启发(另见第 2.5.1 节)。大部分的皮质输入是在位于中层多棘神经元(MSN)的脊髓(Freund 等人,1984;Gerfen,1988)上的突触中接收的，中层多棘神经元是纹状体的投射神经元。多巴胺能(强化)信号从 SNc 传递到同一个神经元，并在脊柱颈部形成突触(Freund 等人,1984;Bouyer 等人,1984)。生理学研究表明，多巴胺能系统可以控制皮质纹状体突触的学习过程(Calabresi 等人,1997;Kerr 和 Wickens,2001)。

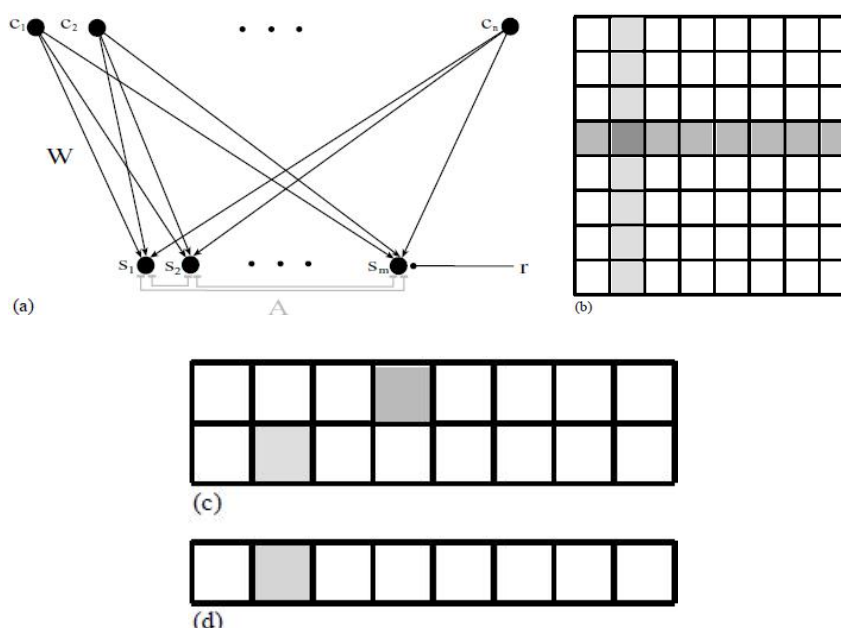


图 43: 基本的 RDDR 模型。(a)基本 RDDR 网络结构, C:输入(皮质)神经元;S:输出(纹状体)神经元;W:输入/输出(皮质纹状体)前馈权重;A:输出层(纹状体)侧权重, r:强化(SNc)信号。(b)从表面尺寸为 64 的输入空间向系统输入的例子。(c)将该实例简化为 16 个固有维数,其中输入的每一行(水平或垂直)由输出的一个单一元素的活动表示。(d)在只对竖线进行加重的情况下,可以把例子缩小到 8 个输出元件的较小尺寸。

### 7.1 模型

基本的 RDDR 网络由两层(输入和输出)构成，其中一个调制标量值代表增强(预测误差)信号。RDDR 模型的输入层中的神经元数量大于输出层中的神经元数量;换句话说，网络的

输入空间比其输出空间具有更高的超表面尺寸(图 2a)。然而,输入实际上可能位于构成输入数据内在维度的输入空间的显著较低子空间内。例如,典型的输入是矩阵中的行(Rumelhart 和 Zipser,1985;Foldiak,1990)。输入空间由大小为  $n \cdot n$ (构成  $n^2$  的超平面尺寸)的矩阵组成。但是,输入仅包含此矩阵中垂直线和水平线的组合。输入矩阵的每个元素都是由  $2n$  维源向量的线性变换(即线求和)生成的,其分量是独立的。输入矩阵的这种定义导致  $2n$  的固有尺寸(远小于  $n^2$ , 对于大的  $n$  值)(图 2b)。在最优编码网络中,假设输出由  $m$  个神经元编码,当神经元数目大于数据的内在维数(即  $m=2n$ )时,网络可以对所有信息进行编码,否则网络将对  $m$  维子空间进行编码,其中包含输入信息的最大的部分(图 2c)。为了研究 RDDR 模型中的选择性增强,针对不同的模式给出了一个差分奖励信号(图 2d)。

输出神经元(s)的活动是其前馈输入( $w \cdot c$ )和横向输入( $a \cdot s$ )的线性加权和:

$$s_i = \sum_{j=1}^{n^2} w_{ij} c_j + \sum_{j=1}^m a_{ij} s_j \quad (1)$$

前馈权值( $w$ )的学习规则是一个归一化的“多 Hebb 学习”规则(Kung 和 Diamantaras,1990),将前馈和增强( $r$ )信号与学习速率常数( $\eta$ )结合在一起

$$\Delta a_{ij} = \eta r [s_i c_j - s_i^2 w_{ij}] \quad (2)$$

横向权值( $a$ )的学习规则是一个归一化的反 Hebb 规则(Kung 和 Diamantaras,1990),除了设置自连接为 0。

$$\Delta a_{ij} = -\eta [s_i s_j + s_i^2 a_{ij}], \quad a_{ii} = 0 \quad (3)$$

当使用非对称组织时,当  $i < j$  时,所有  $a_{ij} = 0$ 。为简单起见,前馈和横向权重均使用相同的学习速率常数( $\eta$ )。为了测量由于 RDDR 过程导致的网络信息丢失,输出层中输入模式的  $m$  维表示被扩展回一个  $n^2$  维空间,以创建重建的、解压缩的模式。重建误差是原始元素和重建元素在所有输入模式上的均方差。

## 7.2.结果

网络显示了一种动态的活动模式,给定来自某个分布的输入,该模式将经历一个学习过程,从而实现稳定。几个关键参数显示学习过程中的重要变化。

### 7.2.1.关联性(Correlation)

网络的输入(Foldiak,1990)包含显著的关联,因为矩阵中同一行和同一列的所有神经元倾向于一起改变它们的速率,模拟来自皮层相邻区域的相关输入(Eggermont,1990)。聚合到网络中的相关输入预期会在输出层内产生相关性。然而,侧抑制连接的动态导致相关性发生更复杂的变化(图 3a)。最初,输出神经元显示相关活动,因为多个输出神经元编码输入的共同方面。然而,随着时间的推移,横向抑制网络导致输出神经元的活动正交化,导致不相关的放电模式(Foldiak,1989)。不相关状态由前馈权值的模式保持。这种不相关的放电模式与动作选择模型的预测非常不同。在这类模型中,任何两个神经元在输出阶段之间的相关性被预测为高度负相关(即,一个细胞的激活率增加导致其所有相邻细胞的激活率降低),并且在网络的稳定状态下保持负相关。关于纹状体(Jaeger 等人,1994;Stern 等人,1998)和苍白球(Nini 等人,1995;Raz 等人,2000;Heimer 等人,2002;Stanford,2003;Bargad 等人,2003)尖峰活动的实验数据表明相关性较弱或不存在,因此类似于 RDDR 模型的预测。

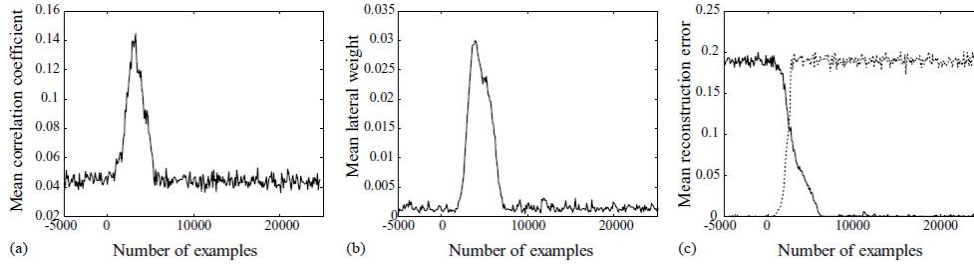


图 44: RDDR 网络属性的动态变化。在得到奖励的输入集在 0 时发生变化后, 网络中发生的变化。(a)输出层神经元间的平均相关;(b)输出层神经元间横向权值的平均值;(c)新奖励模式(实线)和以前奖励模式(虚线)的平均重建误差。模拟 64 个输入神经元、8 个输出神经元、单正前向权值和单负侧方权值。

### 7.2.2. 横向连通性(Lateral connectivity)

网络的横向连通性本质上是抑制性的, 具有反 *Hebb* 学习规律。在这样一个学习规则中, 如果两个神经元结合在一起, 它们的连接效率就会降低(即变得更负或更抑制), 而如果它们不结合在一起, 效率就会增加。因此, 模拟 *BG* 典型的氨基丁酸能侧支的侧权重是一种解除不同神经元输出相关的方法, 使它们编码输入的不同性质。网络学习过程的开始导致输出层神经元的相关活动。这种相关的活动会导致侧边权重的负值。这一过程导致输出的去相关, 最终导致横向权重值降低到非常低的值(图 3b)。在稳态的这个阶段, 主要成分的编码由前馈权重执行(图 4)。因此, 横向权重在设定前馈权重中起着至关重要的作用, 但不在持续输入的稳态编码中起作用。这与使用“winner-take(或 lose)-all”机制的 *动作选择* 模型中横向权重的作用非常不同(Wickens,1993、1997;Berns 和 Sejnowski,1996)。在这种模型中, 在稳态输入的持续处理中, 横向权重是显著的。生理学证据表明, 纹状体(Jaeger 等人,1994;Tunstall 等人,2002;Czubayko 和 Plenz,2002)和苍白球(Stanford,2003)的横向权重对邻近神经元的抑制弱而不对称, 与 RDDR 模型的预期非常相似。

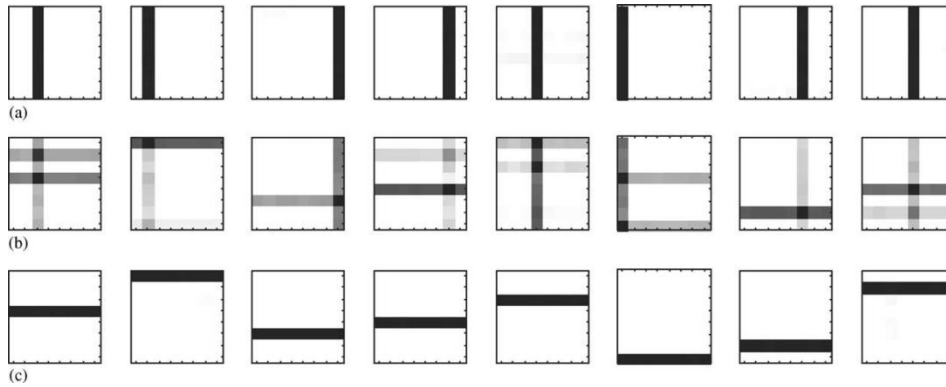


图 45: 前馈权重编码的变化。该网络有水平和垂直两种线。直到时间 0, 竖线与增强信号相关联。在时间 0 之后, 水平线与增强信号相关联。8 个前馈权值的阴影编码在  $8 \times 8$  矩阵中, 次数为:(a) -5000 个样本;(b) +2000 个样本;(c) +10000 个样本。模拟 64 个输入神经元、8 个输出神经元、单正前向权值和单负侧方权值。

### 7.2.3. 信息编码(Information encoding)

网络的信息编码是衡量其保持输入中编码的信息的能力的一个关键指标, 因为这些信息在输出阶段被传递给数量少得多的神经元。在初始状态下, 信息的编码是随机的, 导致无法重构输入。因此, 大部分信息丢失了, 无法从输出神经元的活动中恢复学习过程通过减少重建误差显著改善了信息编码(图 3c)。如果输入空间的内在维度低于输出阶段的神经元数量, 则信息完全编码在输出神经元的活动中, 不会丢失任何信息。因此, 可以从网络输出的活动

中进行完全重建。当输入空间的内在维数大于输出阶段的神经元数时，输出神经元对包含最大信息的输入维数进行编码。因此，可以实现最佳(尽管不是完全)重建。这种编码与*动作选择模型*所建议的数据约简方案有很大的不同。在这种选择模型中，只有一个选择的操作被编码，而所有其他方面都将丢失。这导致网络的信息容量非常低，这会丢失输入的大部分方面。因此，为了简单起见，假设  $m$  个二元输出神经元，*动作选择网络*能够编码  $m$  个模式，而 RDDR 网络能够编码  $2^m$  个模式。

#### 7.2.4.强化信号(Reinforcement signal)

增强信号构成一个控制信号，它调节*前馈网络*的 *Hebb 学习规则*，即改变网络所看到的输入空间结构。它增加了与*奖励活动*相关的维度的可变性，同时减少了与*奖励活动*无关的维度的相对可变性。这导致网络不仅编码其输入空间的最大可变性，而且编码奖励扭曲空间的可变性。因此，有关*奖励*相关活动的信息以最佳形式维护，而非*奖励*信息不被编码，因此无法重建(图 3c)。运用增强信号，网络不再仅仅根据其统计信息以无监督的方式对信息进行编码。相反，它忽略了不重要的信息，更好地编码被增强信号认为重要的信息。

#### 7.2.5.病理学(Pathologies)

BG 的疾病与一些最严重的神经系统疾病如帕金森氏病有关。这些疾病的特点是严重的运动功能障碍和额外的*认知*和*边缘问题*(Sethi,2002)。PD 以中脑多巴胺能神经元的死亡为特征(Ehringer 和 Hornykiewicz,1960)。通常情况下，这些细胞维持的*纹状体*中有稳定状态的多巴胺水平，可能由于好于预期的事件而增加，或由于令人失望的事件而减少(Schultz 等人,1997;Schultz,1998;Fiorillo 等人,2003)。在 PD 期间，由于 SNe 中的神经元死亡，多巴胺释放的正常水平降低，导致*纹状体*中的稳态水平降低(Bernheimer 等人,1973;Schultz,1982; Hornykiewicz 和 Kish,1987)。在 RDDR 模型中，这相当于网络所有输入的负增强信号。所有输入的负增强信号会导致系统编码中断。这反过来导致网络的信息容量非常低，并且输出水平的神经元之间的相关性增加。对抗帕金森病的主要防线是多巴胺替代疗法——通过多巴胺前体左旋多巴或突触后多巴胺激动剂。这种多巴胺替代疗法的非特定(时间和空间)效应最好在背景多巴胺活动框架中理解，代表预测和现实之间的匹配，因此保留当前网络。这些治疗的主要副作用是发展的超活跃障碍——左旋多巴诱导异动症(LID)。RDDR 模型为 LID 的发展提供了自然的解释。多巴胺替代治疗导致血浆(可能在*纹状体*)中多巴胺的“脉动”水平(Shoulson 等人,1975;Nutt 等人,2000)。多巴胺水平不再与受试者的实际表现相关，而只与用药时间相关。因此，网络不断地经历着信息编码的变化，这种变化被错误地认为是加强。这种对信息的错误编码可能导致运动过程中“错误”肌肉的激活，从而导致运动障碍。

## 八、BG 强化驱动降维模型的建立

基本的 RDDR 模型主要是以粗线条表示 BG 的概念，该模型的简单性对于探索其在特定生物系统中的主要想法和预测非常重要。基本模型的假设是一个两层网络，其中所有的神经元都连接到所有其他的神经元，无论是在每层内部还是在层之间。基本模型也作出了简单的假设，即神经元是线性的，它们之间的可调权重不受任何值的限制。重新定义基本模型的这些假设可以对 BG 的功能产生额外的见解。扩展的 RDDR 模型能够包含实验数据的其他方面。对模型的这些添加要么在网络结构级别上，要么在单个神经元级别上。在单个神经元层面上，额外的特征包括对神经元之间连接极性和非线性激活功能的限制。在网络结构层面，附加功能包括多层网络、将纯*前馈*结构改为部分闭环、层间和层内稀疏有序连接以及层间多路径。

## 8.1.约束权重(Constrained weights)

基本 RDDR 模型使用非限制权重, 如基于网络的 PCA 模型(Foldiak,1989;Kung 和 Diamantaras,1990)所示。这些用于前馈和横向连接的权重可能达到不受约束的负值和正值, 甚至可以在正值和负值之间切换。一般来说, 这是一个问题, 因为在大多数情况下, 一个单一的传输器要么是兴奋的, 要么是抑制的。皮质、丘脑和 STN 投射神经元运用谷氨酸, 这被认为是兴奋性的, 而纹状体和苍白球的投射神经元运用 GABA, 这被认为是抑制性的。有一些例外, 其中一种神经递质可以同时起到兴奋和抑制作用, 例如 GABA(Wagner 等人,1997;Chavas 和 Marty,2003)和谷氨酸(Katayama 等人,2003)。然而, 这些例外是罕见的, 纹状体(Tunstall 等人,2002;Czubayko 和 Plenz,2002)和我们对苍白球的初步体外研究(Rav Acha、Bergman 和 Yarom, 未发表的结果)未能确定 GABA 在这些结构中的非抑制作用。

将神经元之间的连接权重限制在单个符号上, 相当于非负矩阵因子分解(NMF)中应用的非负约束(Lee 和 Seung,1999)。对单个符号的约束, 无论是正的还是负的, 意味着每个神经元编码的特征之间的关系只能是相加的。加性编码意味着不同神经元的输出之间不可能有减法。这反过来又导致每个输出神经元对输入空间的一部分进行编码。空间的编码部分是本地编码或稀疏编码的一种形式, 它取代了 PCA 执行的全局编码。非负矩阵因式分解的结果对于它们与 RDDR 模型的关系非常有用, 因为它们除了具有明显的权重约束外, 还能够更好地适应不同层的地形分布和编码的稀疏性。(另见第 7.6 节)。其他研究表明, 这种非负分解可以通过横向抑制来实现(Plumbley,2001)。研究表明, 运用非负约束, 网络可以进行独立分量分析(ICA), 将输入信号分离为非相关分量(如 PCA), 并将其分离为独立分量。独立分量与不相关分量的区别在于, 最小化包括高阶统计量, 而不仅仅是二阶统计量。将权重限制在一个符号上也是一种非线性形式, 通过非线性编码实现效益(第 7.2 节)。

然而, 尽管单符号约束具有优势, 但重要的是要记住, 在真实的生物环境中, 这种约束可能不是“硬”约束。有些输入神经元传递的信息可能与其他输入神经元传递的信息相反。例如, 运动皮层中的神经元编码方向, 一个神经元编码特定方向, 而另一个则编码相反方向(Georgopoulos 等人,1986)。当两个输入神经元(传递特征和相反的特征)使用相同的神经递质(因此相同的约束)连接到相同的输出神经元时, 它们提供补充信息。在这种情况下, 两个输入作为信息的两个对称部分发挥作用。这迫使在学习过程中的任何给定时间, 只有一个输入神经元有效地激活输出神经元。总效应是由于反向输入神经元的互补效应而消除了单信号约束。

## 8.2.非线性元件(Non-linear elements)

基本的 RDDR 模型使用线性神经元进行 PCA。具有线性响应函数的神经元通过简单的求和以线性方式将其加权输入的和转换为放电速率。然而, BG 区的神经元, 就像大脑中的所有神经元一样, 在它们的上端和下端的放电率是有限的。例如, 由于神经元不能转换为负的放电速率, 因此放电速率的下限可能为零。除了饱和外, BG 不同核内的神经元具有不同的反应功能, 这些反应功能不一定是线性的, 甚至在饱和极限之间也是如此。一方面, 苍白球的神经元具有线性反应(Nakanishi 等人,1990、1991;Kita 和 Kitai,1991;Kita,1992)和广泛的放电率(Delong,1972;Miller 和 DeLong,1987)。另一方面, 纹状体的 MSNs 是非线性的, 有两个基本的阈下状态。在向下状态下, MSN 神经元的放电率为零或非常低, 而在向上状态下则可以达到非常高的放电率(Nisenbaum 和 Wilson,1995;Wilson,1995;Wilson 和 Kawaguchi,1996;Stern 等人,1997;Bennet 和 Wilson,2000)。这些神经元的中间动态范围很小, 主要导致极端状态之一。

将非线性添加到网络中可以增加网络执行的计算范围。PCA 的线性版本可能只确定输入之间的线性交互，而形成的最佳连接仅基于二阶统计(活动的协方差)。在输入元素之间的非线性相互作用和相互作用的高阶统计的情况下，线性网络将不能有效地降低输入的维数并保持其中的信息。Oja 等人(1991)已经证明，使用 PCA 网络中使用的具有非线性传递函数(通常是 sigmoid)的相同约束 *Hebb 学习规则*，可以学习嵌入在噪声环境中的复杂非线性输入结构。非线性元素的添加(Jutten 和 Herault,1991;Karhunen 和 Joutsensalo,1994)使提取高阶统计量和在输出层形成独立成分成为可能，从而使由输入表示的混合物能够分离成其独立的潜在原因。

极端非线性计算的一个重要特征是泛化。线性编码将不同的输入转换为唯一的输出。另一方面，0/1 转换或 sigmoid 传递函数的极端非线性情况将一组输入转换为单个输出。因此，根据转换输出将输入划分为组(或集群)。Carlson(1990)实施了这样一个非线性网络。在他的网络中，神经元的活动取决于神经元的阈值和过渡宽度。宽度决定了神经元的非线性：当宽度为 0 时，活动是高度非线性的(二进制)，当它很大时，活动是线性的。阈值和宽度都在学习过程中进行调整。输出神经元是按层次组织的(如 APEX 线性网络和其他(Kung 和 Diamantaras,1990;Rubner 和 Schulten,1990))。在非线性情况下，该算法的结果是将输入划分为簇，而在线性情况下，它实现了连续的线性表示。

### 8.3.多层(Multiple layers)

基本的 RDDR 模型是一个两层(输入和输出)网络。然而，事实上，BG 形成一个多层网络：来自皮层的输入在第一层(BG 输入)纹状体和 STN 内接收，然后从那里传递到下一层(BG 输出)GPi 和 SNr。这种多层前馈网络的好处来自两个方面：顺序执行相同处理的能力和每层执行不同类型处理的能力。

由多层(或分层)降维网络执行的顺序处理具有两个主要优点，这两个优点是能够在多个阶段执行相同的计算(Barlow,1992b)：

- ◎由于上述对突触数量的限制，单个神经元无法直接访问大部分输入。多层结构允许在逐步收敛的过程中访问输入中的大部分信息。
- ◎由于输入空间的巨大，其稀疏性阻碍了对其表示形式的有效学习。因此，这个问题被分解为多个层，其中第一个层学习有限数量的输入模式，而后续层学习输入空间的更大部分。

在 BG 中，这两层中的每一层都能接触到多达  $10^3$ - $10^4$  个输入神经元(Zheng 和 Wilson,2002;Yelnik,2002)。理论上，两层的结合使输出层的神经元(苍白球神经元)能够接触到大部分皮质输入(最多  $10^7$  个皮质神经元)。由于皮质 BG 环的部分闭合结构(第 7.4 节)和多重前馈通路(第 7.6 节)，这个数字可能会进一步增加。

多层降维的第二个方面是能够在每一层中执行不同类型的处理。多层网络可以在单个网络中组合线性和非线性层。Oja(1991)的研究表明，网络包含一个非线性(编码)层，然后是一个执行非线性和线性 PCA 的线性(瓶颈)层。该网络能够执行从输入空间到较小的输出层的任何连续映射。这个网络与 BG 主要前馈通路的多层结构惊人地相似(图 5a)。皮质和丘脑输入到达纹状体的 MSNs，这是非线性的，具有与 Oja 使用的 sigmoid 类似的传递功能。信息从纹状体传递给苍白球，苍白球的特征是神经元具有相对线性的传递功能。Oja 的工作还推测了两个附加层在解码由初始编码层处理的信息中的作用。这些解码层很好地位于皮质 BG 皮质环内，相当于丘脑和额叶皮质。进一步研究(Demers 和 Cottrell,1993;Malthouse 等人,1995)对此类多层非线性网络的实际和理论影响，证明了它们在复杂特征提取方面的增强能力，而复杂特征提取不能由线性网络执行。



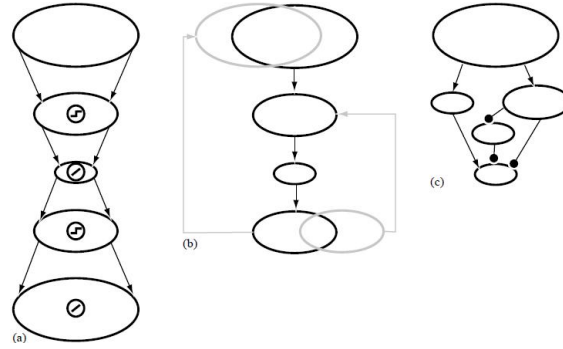


图 45: RDDR 网络的增强。基本网络结构的增强:(a)非线性多层网络, 苍白球和皮层由包含线性神经元的层表示, 纹状体和丘脑由包含非线性神经元的层表示;(b)循环网络, 运用代表皮层和丘脑的各层对纹状体的反馈, 形成部分闭环;(c)具有多个通路的网络, 代表 BG 的直接、间接和超直接通路。

#### 8.4.部分闭环(Partially closed loop)

基础的 RDDR 模型模拟了一个前馈网络, 信息从大部分皮层流向 BG。然而, 这种漏斗结构只是部分闭合回路的一部分(图 5b)。BG 的输出达到丘脑的腹侧、腹前、中嗅和中央核(Haber 和 McFarland,2001)。丘脑的神经元投射回到纹状体(McFarland 和 Haber,2001), 占纹状体输入的 50%(Kemp 和 Powell,1971;Bolam 等人,2000)。此外, 这些丘脑核的输出到达前额、运动前、辅助运动和运动皮质(McFarland 和 Haber,2002;Middleton 和 Strick,2002), 构成纹状体大部分皮质输入(Parent 和 Hazrati,1995)。这种结构形成了一个部分闭合的循环, 其中从 BG 传输的信息作为输入与其他来源的信息一起反馈(Joel 和 Weiner,1994)。

神经回路结构的时间延迟使得不同时间的信息融合成为可能。结果表明, 部分闭环系统是序列学习、序列检测和生成的最佳基础。事实上, 早期的研究已经表明, BG 不同核内的神经元参与动物的顺序活动(Kermadi 和 Joseph,1995;Nakahara 等人,2001)。一些动作选择模型已经扩展到包括顺序行动的选择(Berns 和 Sejnowski,1998;Beiser 和 Houk,1998)(另见第 3.4 节)。

RDDR 模型通过添加一个部分闭合的循环, 为信息处理提供了一个额外的视角。在这种结构中, RDDR 网络不仅接收表示单个时间点的输入, 还接收有关网络先前输入的附加信息。通过创建一个部分依赖于其输出的网络输入, RDDR 网络执行一个简化, 该简化不仅依赖于单个时间事件的属性, 还依赖于先前编码的信息。循环结构可以导致序列的优先表示, 从而导致奖励信号。因此, 网络将在一定的时间内减少信息, 这取决于向网络提供的先前信息。根据先前提供给网络的输入, 给定的输入可能具有非常不同的表示形式。在大多数信息遵循特定顺序(例如正常环境中的自然操作)的情况下, 编码序列而不是单个时间事件是非常有益的。典型的输入不会发生在一个单一的时间点, 而是跟随其他与它们在时间上相关的输入。

部分闭环的第二个方面及其与 RDDR 模型的关系是空间的, 即它能够分布信息并导致其集成。前馈和横向连接的有限空间范围及其固有的连接稀疏性意味着只有一小部分信息到达每个神经元(第 7.5 节)。丘脑和额叶皮质通过环结构提供的反馈信息增加了第一层(纹状体)内每个神经元可获得的信息量, 从而提高了维度的减少。

部分闭环结构参与时空信息整合, 可以补充 BG 多层的整合(第 7.3 节)。多层集成在识别和处理日益复杂的特征方面至关重要(BaLow,1992b)。部分闭合的循环通过在相同的层中再次输入部分处理的信息, 可以减少复杂特征编码所需的层数。这个过程导致在网络中重复使用有限的层, 形成高度压缩的信息。



## 8.5.稀疏连接(Sparse connectivity)

基础的 RDDR 模型假定不同层之间和内部的连接是完全的。完全连接意味着每个神经元与前一层和后一层中的所有神经元以及同一层中的所有神经元之间的解剖连接。然而，BG 的实际解剖和生理学还远没有表现出如此完整的连接模式。特别是 BG 不同核的连接性是不完全的，神经元与其他大多数神经元没有连接。这一方面对于理解神经网络的行为至关重要，通常被称为稀疏连接。

稀疏连接可能出现在两种极端形式之间的任何地方。第一种类型是“有序稀疏性”，在这种稀疏性中，相邻神经元接受类似连接的概率明显高于远端神经元。第二种类型是“无序稀疏性”，所有神经元都有相同的概率接受相似的连接。从解剖学和生理学上检测到的空间组织表明纹状体(Hoover 和 Strick,1993;Kincaid 等人,1998;Zheng 和 Wilson,2002)和苍白球(Hoover 和 Strick,1993)以更接近于“有序稀疏度”的方式接收输入。此外，解剖证据表明，各层内的侧支主要是局部的(Wilson 和 Groves,1980;Parent 等人,2000)，因此，侧支连通性在其空间范围内也主要是有序的和有限的。

横向连接的有序稀疏性导致局部域在全局输入下可能重复出现。同样的约简可以由不同组的非连接(或非常稀疏连接)神经元独立完成，从而导致编码的重复。然而，前馈连接的有序稀疏性可能导致不同的神经元群的形成，这些神经元群高度互联，并对输入的一个子集而不是对核的整体输入进行减少。这反过来会导致形成多个相互连接的小漏斗，而不是一个大漏斗。将其分为多个漏斗可能会降低整体效率。然而，由于输入中的大部分冗余倾向于局部冗余，这可以从邻近皮质神经元相对于远端神经元的相关性增加中看出，因此这种减少可能仍然非常有效。

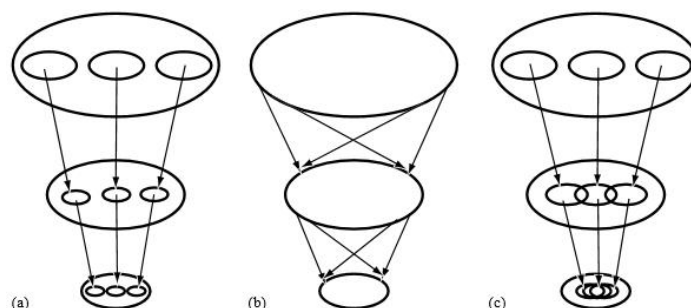


图 46: BG 网络中稀疏连接的模式。前馈网络的组织形式为:(a)反映有序稀疏性的并行隔离路径;(b)反映无序稀疏性的汇聚路径;(c)已知 BG 连通性模式提示的中间结构。

BG 显示的稀疏类型与并行处理和信息收敛问题密切相关，这一问题在过去引起了许多争论(第 3.1 节)。在有序稀疏性中，一组神经元共享相同的输入，这与其他神经元共享的输入不同。因此，它们在隔离的信息通道上执行缩减。多个组与共享信息子集的每个组之间的连接导致了信息的并行处理，而没有交互作用(图 6a)。例如，在这种极端情况下，总共  $10^8$  个皮质神经元中，每  $10^4$  个皮质神经元投射到一组  $10^3$  个纹状体神经元，从而形成  $10^4$  个单独的漏斗，在每个漏斗内进行完全连通的还原。另一个极端是完全无序的稀疏性，即整个输入结构均匀地投射到输出结构中的所有神经元。因此，所有信息在整个层中共享，从而导致信息的聚合(图 6b)。用同样的例子，现在所有  $10^8$  个皮质纹状体神经元都延伸到  $10^7$  个纹状体神经元，从而形成一个非常稀疏(连接概率为  $10^{-4}$ )连接模式的单一漏斗。BG 各层之间的连接模式可能是中介物(图 6c)。这意味着稀疏性在某种程度上是有序的，导致多个(尽管可能不是完全隔离的)具有高连接级别的信息漏斗。

*BG* 形成多层(第 7.3 节)部分闭合环(第 7.4 节)。这种结构可以将多个较小的漏斗集成到较大的漏斗中, 包含更多的总体信息。因此, 尽管这种连接在解剖学上存在局限性, 但可以通过不同级别的局部性和集成来执行降维。此外, 局部连通性结构非常适合通过单个符号权值(第 7.1 节)形成预期的特征提取(或减少输入)。在单符号的情况下, 输入往往是相加的, 因此保持局部方面, 而不是 PCA 建议的全局形式。

#### 8.6. 多种途径(Multiple pathways)

基本的 RDDR 模型只考虑单个路径。这条通路, 也被称为 *BG* 的主轴。(Percheron 等人,1994), 从皮层穿过纹状体, 然后从纹状体直接到达输出核(GPi 和 SNr)。然而, 通过 *BG* 的信息流使用了额外的通路:从纹状体通过 GPe 和 STN 到输出核的间接通路(Albin 等人,1989)和超直接通路(Kita, 1992;Ryan 和 Clark,1992;Nambu 等人,2002b)从皮质直接到 STN, 再从 STN 到输出核(详见 2.2 节和图 1)。将 RDDR 模型扩展为包含多个路径, 可以在基本网络上实现几个主要的改进:

- ◎来自相同来源的正负输入的平衡。
- ◎改进了多输入的整合。
- ◎各种时间延迟线的整合。

纹状体到苍白球的直接通路对 GPi/SNr 有净抑制作用。皮质放电的任何提升都会增加纹状体放电, 从而减少输出核的放电(Tremblay 和 Fillion,1989)。另一方面, 另外两个通路对输出核的皮层产生净的正作用。在超直接通路中, 皮质的增加放电导致 STN 的放电增加, 进而刺激 GPi/SNr。在间接途径中, 皮层增加纹状体的(放电)速率, 从而导致 GPe(放电率)降低, 导致 STN(放电率)增加, 最后导致 GPi/SNr 的激发(Nambu 等人,2000)(图 5c)。运用多个路径, 总的平均速率通常可以保持, 而不会转移到极高或极低的速率(导致输出神经元的非线性激活)。此外, 同样的输入可以通过两个途径以网络兴奋和网络抑制的方式接收, 从而导致对输入的无约束适应。这可能是克服特定神经递质对单信号输入设置的限制的一种方法(见第 8.1 节)。

不同的通路显示了不同程度的皮质信息整合:纹状体直接输出到 GPi/SNr, 比通过 STN 的各种通路的输出呈现出更明显的分离和不同的输入(Mink,1996)。这些不同层次的融合使输出核的神经元能够访问和减少来自不同特定层次的信息组合。信息范围从每种形态的特定信息(低收敛导致的)和全局信息(高收敛导致的)。最后, 不同的路径具有不同的时间延迟:超直接路径最快, 直接路径较慢, 通过间接路径传递的信息最慢(Nambu 等人,2002b)。这使得输出核能够多次接收信息并减少组合信息。这一机制与皮质-*BG* 部分闭环结构(第 7.4 节)相结合, 可以减少时间序列, 而不仅仅是时间上的单一事件。

#### 8.7. 增强驱动的降维——*BG* 奖励偏向表示

与比较先进的 *BG* 处理模型相比, 大多数模型在纹状体层面采用非常简单的信息处理 Actor 系统。例外情况是 Suri 等人(2001)提出的模型中的 Actor, 该 Actor 在 *BG*-丘脑皮质连接中实施。然而, 即使在这个模型中, 纹状体的输出也被假定为以相对直接的方式转化为皮质活动。每一个纹状体神经元都对应一个特定的动作。通过位于苍白球内部节段(GPi)和 SNr 水平的特定神经元, 它抑制一个丘脑神经元, 后者依次投射到一个特定的皮质神经元, 后者的持续激活执行皮质动作。

在这一节中, 我们将提出一个 *BG* 处理的模型, 该模型可能被扩展为 *BG*Actor 模型的基础。根据已知 *BG* 解剖和生理学的几个重要限制条件, 该模型建议 *BG* 对皮质表征(Bar-Gad,

Havazelet-Heimer, Ruppin 和 Bergman, 2000; [www.math.tau.ac.il/~Ruppin](http://www.math.tau.ac.il/~Ruppin)) 进行有效的 *强化驱动的降维*(RDDR)。我们集中在理论部分。有关模型的更详细介绍以及对实验猴进行的电生理实验的描述, 以测试一些模型预测, 请参见 Bar-Gad 等人(2000)的论文。

这一模型是由两个主要的解剖和生理特征的 *BG-丘脑皮质* 电路:

① *BG* 的漏斗状结构。

投射到纹状体的皮质神经元的数量比纹状体神经元的数量大两个数量级(Kincaid, Zheng 和 Wilson, 1998), 纹状体到 GPi 的数量也出现了同样数量的额外减少(Oorschot, 1996; Perceron, Francios, Yelnik, Felon 和 Talbi, 1994)。尽管对苍白质丘脑和丘脑皮质水平的神经元群体的定量研究仍然缺乏, 但大多数解剖学研究表明, 皮质纹-状体-苍白质-丘脑-皮质路在苍白质水平后径逐渐扩大(Arecchi-Bouchioua, Yelnik, Francios, Percheron 和 Tande, 1996; Sidibe, Bevan, Bolam 和 Smith, 1997)。

② 纹状体神经元之间缺乏相互抑制的电生理证据(Jaeger, Kita 和 Wilson, 1994), 尽管存在纹状体广泛横向连接的解剖学证据(Kita, 1996; Yelnik, Francios 和 Tand, 1997)。

解释解剖学和生理学数据与沿皮质基底节-丘脑皮质环的漏斗结构之间明显差异的一个可能的解决方案是假设 *BG* 有效降低皮质活动的维度。术语“*降维*”描述了将输入从高维空间投影到相当小的空间的过程。当原始空间中包含的全部或大部分信息被保留时, 就可以实现有效的减少。

Bar-Gad 等人模型的一个重要假设是, Actor 动物的维数减少不仅应受到输入模式的统计特性的影响, 还应受到其 Actor 特征的影响。投入的相对重要性取决于其新颖性(Redgrave, Prescott 和 Gurney, 1999)、激励显著性(Berridge 和 Robinson, 1998)和预测回报的能力(Robbins 和 Everitt, 1996)。Suri 和 Schultz 在本期的论文回顾了近年来收集的大量证据, 这些证据表明, 这些信号是由 DA 神经元编码的, 并且可以通过其 DA 输入(如 Kotter 等人在这个问题上所述)到达纹状体。

理论研究已经表明, 神经网络可以运用层间连接的竞争性赫布型学习规则(Oja, 1982)和横向抑制层间连接的反赫布型规则(Foldiak, 1989; Kung 和 Diamantars, 1990)进行有效的降维。显然, 这些网络通常具有漏斗状结构。为了检验 RDDR 假设, Bar-Gad 等人研究了一种模拟前馈神经网络, 该网络运用横向抑制提取主成分空间(Foldiak, 1989; Kung 和 Diamantars, 1990)。这个网络由三层组成: 第一层代表皮质输入, 中间层代表纹状体, 输出层代表 GPi。学习前向权值为赫布型, 横向权值为反赫布型。将一个增强信号与中间层的前馈输入相结合, 形成一个三因素的赫布型学习规则, 初步模拟皮质纹状体突触的多巴胺能神经调节。强化信号对奖励相关事件为正, 对非奖励相关事件为零(基线 DA 水平), 与未奖励刺激相比, 为奖励刺激分配更多编码资源。网络权重被限制为正值或负值, 以重新反映已知的神经递质生理学。为了测量 RDDR 过程导致的网络信息丢失, 输出层被扩展回一个输入大小的空间, 重建解压缩后的模式。

这些模拟表明, 将学习过程中大于基线的增强信号归因于“有意义”模式的选定子集确实会导致识别性信息提取, 为选定的、奖励增强的输入提供比基线刺激集更好的重构。Bar Gad 等人证明了增强信号值相对于基线水平的两倍增加导致压缩重建误差几乎减少了五倍。

以新的输入模式呈现网络会导致输出神经元的相关活动。这种相关性导致抑制性侧突触效率的短暂增加和前馈连接效率的短暂变化。这些变化反过来导致输出层内神经元活动的去相关和信息压缩的改善。这些突触变化的短暂性一方面解释了为什么层内突触对编码过程很

重要,另一方面,在编码过程结束时,它们可能获得几乎消失的值。因此,这些网络的学习动态为似乎与纹状体侧抑制连接相关的解剖和生理数据之间的差异提供了可能的解释。这些结果表明,纹状体内部连接性的弱功能性以及纹状体和苍白神经元的低相关性可以通过注意到获得这些结果的大多数实验都是在没有积极参与学习新 Actor 任务的动物身上进行的来解释。

为了实验验证这一预测,Bar-Gad 等人训练一只猴子完成一项按键任务,并记录其在任务执行期间的苍白球的活动,计算 151 对苍白球的神经元的相关系数(Bar-Gad 等人,2000)。相关系数在已知任务执行期间较低,导致预期回报和休息期间。在意外奖励之后、在先前奖励 Actor 停止奖励之后以及在未经培训的奖励 Actor 执行之后,观察到绝对相关值显著增加。增强相关性的时间延长并持续数十秒。在学习过程中发现的高相关性,排除了纹状体和苍白球缺乏相关活动的可能性,这仅仅是皮质-纹状体连接稀疏的结果。这些降低的相关性更像是一个活跃的去相关(decorrelating)过程。

RDDR 模型表明 BG 在整个皮质信息的提取和预处理中起着作用。为什么它在计算上有用?首先,它允许在有限数量的轴突内传输大量信息。Bar-Gad 等人假设 BG 表现出代表动物当前状态的广泛皮质神经活动的维度降低。减少的信息被投射到额叶皮质,额叶皮质运用它来计划未来的行动。因此,RDDR 网络运用每个额叶神经元能接收的有限数量的突触,使额叶皮质执行区的神经元能够接触到最大程度的传入皮质信息。第二,RDDR 网络提供了一种载体,通过该载体,RL 可以在大脑的中央、节俭的位置上进行,通过允许刺激的食欲值来指导它们的存储和表示。这种选择性的 RDDR 存储倾向于将整个网络的响应偏向于有回报的输入刺激。正如 Houk 等人(1995)已经指出的,在制定和实施计划和行动时,对复杂情况的这种有偏见的信号可能是有用的。此外,BG 的部分皮质输入来自额叶皮质,可能代表计划和行动。因此,BG 的输出可能使额叶皮质水平的表现偏向于选择奖励计划和行动。因此,我们建议 RDDR 框架可以作为 BG Actor 模型的基础。

## 九、在 BG 的同步与探索——尖峰网络模型

Chakravarthy,Joseph 和 Bapi(2010)认为,IP 中耦合的兴奋抑制网络 STN-GPe 环可能是探索的底物(Chakravarthy 等人,2010)。众所周知,耦合的兴奋-抑制神经元联合可以表现出丰富的动力学行为,如振荡和混沌(Borisyuk,Borisyuk,Khibnik 和 Roose,1995;Sinha,1999)。这一假设启发了模拟各种 BG 功能的模型,在正常和帕金森病条件下,范围覆盖了连续空间中的动作选择(Krishnan,Ratnadurai,Subramanian,Chakravarthy 和 Rengaswamy,2011)、伸展运动(Magdome 等人,2011)、空间导航(Sukumar,Rengaswamy 和 Chakravarthy,2012)、精确握持(Gupta,Balashubramani 和 Chakravarthy,2013)和步态(Muralidharan,Balashubramani,Chakravarthy、Lewis 和 Moustafa,2013)。运用“速率-编码”神经网络, Kalva,Rengaswamy,Chakravarthy 和 Gupte(2012)表明,探索是从 STN-GPe 系统的混沌动力学中出现的(Kalva 等人,2012)。根据设计,大多数“速率-编码”模型无法捕获动力学现象,如在更现实的脉冲神经元模型中发现的同步现象(Bevan,Magill,Terman、Bolam 和 Wilson,2002; Park,Worth 和 Rubchinsky,2010; Park,Worth 和 Rubchinsky,2011)。自从发现 STN、GPe 和 GPi 神经元在帕金森病状态下表现出高水平的同步性以来,BG 核内的同步性受到关注(Bergman,Wichmann,Karmon 和 DeLong,1994;Bevan 等人,2002;Hammond,Bergman 和 Brown,2007;Tachibana,Iwamuro,Kita, Takada 和 Nambu,2011;Weinberger 和 Dostrovsky,2011)。发现这种振荡活动存在于两个频段,一个在震颤频率[2-4 Hz]附近,另一个在[10-30 Hz]频率附近(Weinberger 和 Dostrovsky,2011)。Park 等人(2011)报告在接受深部脑刺激(DBS)手术(Park 等人,2011)的 PD 患者中,STN 神经

元与其局部场电位(LFP)之间存在间歇性同步性，这在健康对照组中是不存在的。

当前研究的一个关键目标是使用一个二维脉冲神经元模型来理解和关联 STN-GPe 的同步水平与探索。作为第二个目标，我们将上述模型应用于 DAW, O'Doherty, Dayan, Seymour 和 Dolan(2006)以及 Bourdaud, Chavarriaga, Galan 和 del R Millan(2008)(Bourdaud 等人, 2008; DAW 等人, 2006)的 N-臂强盗问题，具体目的是研究 STN-GPe 的动力对探索的贡献。该模型与传统的基于 RL 的 BG 建模方法具有相同的特点。例如，将多巴胺信号与奖励预测误差进行比较(Schultz, 1998)。此外，DA 可以控制皮质纹状体可塑性(Reynolds 和 Wickens 2002)，调节纹状体神经元的增益(Hadipour-Niktarash, Rommelfanger, Masilamoni, Smith 和 Wichmann, 2012; Kliem, Maidment, Ackerson, Chen, Smith 和 Wichmann, 2007)，并通过调制连接影响 STN-GPe 的动态(Fan, Baufreton, Surmeier, Chan 和 Bevan, 2012; Kreiss, Mastropietro, Rawji 和 Walters, 1997)。

### 9.1. BG 脉冲神经元模型

之前描述的 BG 网络模型(Mandali, Rengaswamy, Chakravarthy 和 Moustafa, 2015)用于模拟二元动作选择和 N-臂强盗任务。有关模型及其相关方程式的详细信息，请参阅前面的章节。具体任务及相关措施如下。

### 9.2. 二元动作选择任务

我们模拟的第一个任务是简单的二元动作选择，类似于 Humphries, Stewart 和 Gurney(2006)所描述的那样，其中两个相互竞争的刺激被呈现给模型(Humphries 等人, 2006)。输入频率被认为代表“显著性”，较高的频率代表更高的显著性(Humphries 等人, 2006)。纹状体输出对皮质输入的响应在几十赫兹的范围内(Sharott, Doig, Mallet 和 Magill, 2012)。因此，假设代表 2 个动作的频率约为 4Hz(1 号刺激)和 8Hz(2 号刺激)。纹状体神经元(无输入)的自发输出频率假设为 1 Hz 左右(Plenz 和 Kitai, 1998; Sharott 等人, 2012)。在现有选择中选择更突出的刺激可以被视为“运用”，而选择不突出的刺激可以被视为“探索”(Sutton 和 Barto, 1998)。因此，如果选择 2 号刺激(更突出)，则所选操作定义为“Go-执行”；如果选择 1 号刺激(不突出)，则定义为“探索”；如果未选择 1 号刺激(不突出)，则定义为“NoGo-拒绝”。

在空间上给予输入，使晶格上半部分的神经元接受 1 号刺激和下半部分的另一个(图 6.1)。纹状体的 D1 和 D2 神经元的纹状体输出分别作为 GPi 和 GPe 模块的输入，投射模式如图 6.1 所示。与 1 号刺激相对应的“泊松峰”序列作为神经元(1-1250)的输入呈现，并且它们之间完全相关。同样，与 2 号刺激相对应的“泊松峰”序列呈现为神经元的输入(1251-2500)，它们之间完全相关。刺激 1 和 2 在 100 到 200 ms 之间间隔 100 ms；在其他时间，所有纹状体神经元均呈现 1 Hz 的不相关尖峰序列。

### 9.3. N-臂强盗任务

我们现在描述了用于研究探索性和运用性行为的四臂强盗任务(Bourdaud 等人, 2008; Daw 等人, 2006)。在这项实验任务中，受试者有四个手臂，其中一个手臂将在每个试验中被选中，总共 300 个试验。每个时段的奖励/回报均服从高斯分布，其平均值从实验值变为奖励范围为 0 至 100 的实验值。在第  $k$  次试验中，与第  $i$  台机器相关的奖励  $r_{i,k}$  是从平均值  $\mu_{i,k}$  和标准差  $\sigma_0$  的高斯分布中得出的。回报四舍五入到最接近的整数，范围为[0, 100]。在每次试验中，平均值根据衰减高斯随机游动进行扩散。该试验被定义为“运用性”试验，前提是选择了最高奖励的手臂，而其他试验则被定义为“探索性”试验。

自动控制机械产生的回报计算如下：

$$\mu_{i,k+1} = \lambda_m \mu_{i,k} + (1 - \lambda_m) \theta_m + e \quad (6.1)$$

$$r'_{i,k} \approx N(\mu_{i,k}, \sigma_0^2) \quad (6.2)$$

$$r_{i,k} = \text{round}(r'_{i,k}) \quad (6.3)$$

式中， $\mu_{i,k}$  是第  $i$  台机器在第  $k$  次试验期间具有标准偏差 ( $\sigma_0$ ) 的高斯分布的平均值。 $\lambda_m$  和  $\theta_m$  控制平均值的随机游动 ( $\mu_{i,k}$ )，并且  $e \sim N(0, \sigma_d^2)$  由均值 0 和标准差  $\sigma_d$  的高斯分布得到。 $r_{i,k}$  和  $r'_{i,k}$  分别是四舍五入到最近整数之前和之后的收益。平均收益的初始值， $\mu_{i,0}$ ，被设置为 50。参数  $\lambda_m, \theta_m, \sigma_d, \sigma_0$  的所有值均改编自 (Bourdaud 等人, 2008)。

为了做出一个最佳的决定，受试者需要跟踪与四个手臂相关的奖励。主体对探索或运用的决定取决于这一内部表示，它将与正在获得的实际回报非常相似。仅仅通过观察脑电图和所选的位置数据就很难确定受试者是做出了“探索性”的决定还是做出了“运用性”的决定。需要一个特定的主题模型来对他们的决策进行分类并确定战略 (Bourdaud 等人, 2008; Daw 等人, 2006)。记住这一点，Bourdaud 等人 (2008) 使用了一个“行为模型”，该模型使用了 RL 的归一化指数原理 (softmax principle) 来确定人类受试者的选择模式。调整行为模型的参数“ $\beta$ ”，使最终选择模式与实验中单个受试者的选择模式相匹配 (如下所示)。在行为模型中控制探索水平的参数“ $\beta$ ”进行了调整，以匹配八个受试者中每一个受试者获得的探索百分比 (一个受试者的数据被认为丢弃)；八个受试者中有两个具有相似的探索水平。因此，共考虑了六个受试者的数据，以检查所提出的脉冲 BG 模型的性能。

### 9.3.1 行为模型【改编自 Bourdaud 等人 (2008)】

行为模型将每个试验标记为与“探索性”或“运用性”决策相对应。该模型假设用户使用贝叶斯线性高斯规则 (即卡尔曼滤波器) 估计每台机器的平均收益。使用这些估计，他/她根据归一化指数原理选择一台机器。假设所有受试者共享相同的模型来跟踪回报方式，因此，参数是使用整个可用数据计算的。模型的参数 (对于平均跟踪和机器选择) 是通过最大化模型的可能性来估计的。

在任何给定的试验中，行为模型为所有考虑到先前观察的机器提供平均回报 (即，在先前试验中获得的回报)。将模型对所有机器的估计收益进行比较，将该试验标记为“探索”或“运用”。用户选择具有最高估计平均值的机器的那些试验被标记为与“运用性”决策相对应。

跟踪每台机器回报的目标策略由一个卡尔曼滤波器 (Kalman filter) 建模，该滤波器的参数假定在试验期间保持不变。一旦选择了第  $j$  个机器，在第  $k$  个试验中，估计的回报分配将从其预选值 ( $\hat{\mu}_{j,k}^{pre}, (\hat{\sigma}_{j,k}^{pre})^2$ ) 更新为其后选择值 ( $\hat{\mu}_{j,k}^{post}, (\hat{\sigma}_{j,k}^{post})^2$ )，如下

$$\hat{\mu}_{j,k}^{post} = \hat{\mu}_{j,k}^{pre} + K_k (r_k - \hat{\mu}_{j,k}^{pre}) \quad (6.4)$$

$$(\hat{\sigma}_{j,k}^{post})^2 = (1 - K_k) (\hat{\sigma}_{j,k}^{pre})^2 \quad (6.5)$$

其中

$$K_k = \frac{(\hat{\sigma}_{j,k}^{pre})^2}{(\hat{\sigma}_{j,k}^{pre})^2 + (\hat{\sigma}_0)^2} \quad (6.6)$$

由于用户无法观察到这些机器的回报，因此剩余机器的平均估计值不会因选择而改变。也就是说，

$$\forall i \neq j$$

$$\hat{\mu}_{j,k}^{post} = \hat{\mu}_{j,k}^{pre} \quad (6.7)$$

$$\hat{\sigma}_{j,k}^{post} = \hat{\sigma}_{j,k}^{pre} \quad (6.8)$$

然后，根据扩散规则对估计量进行演化：

$$\hat{\mu}_{j,k+1}^{pre} = \hat{\lambda} \hat{\mu}_{j,k}^{post} + (1 - \hat{\lambda}) \hat{\theta} \quad (6.9)$$

$$(\hat{\mu}_{j,k+1}^{pre})^2 = \hat{\lambda}^2 (\hat{\sigma}_{j,k}^{post})^2 + \sigma_d^2 \quad (6.10)$$

受试者的选择采用归一化指数原理进行建模：即，在每次试验  $k$  中，选择机器的概率为

$$P_{i,k} = \frac{\exp(\beta \hat{\mu}_{i,k}^{pre})}{\sum_j \exp(\beta \hat{\mu}_{j,k}^{pre})} \quad (6.11)$$

其中“ $\beta$ ”是缩放参数。 $\beta$ 值越高，系统就会出现“运用”行为，反之亦然。行为模型  $(\sigma_0, \hat{\theta}, \hat{\lambda}, \hat{\sigma}_d)$  的参数是通过在以下约束条件下最大化日志可能性来估计的。为了加快收敛速度，估计的参数  $(\sigma, \hat{\mu}_{j,0}^{pre} \& \hat{\sigma}_{j,0}^{pre})$  分别初始化为原始模型  $(\sigma_0, \mu_{j,0} \& \sigma_{j,0})$  的参数。修正最后两个参数不会显著影响其他参数的估计，因为它们的影响在几个试验中很快消失。表 6.1 显示了模型的估计值，与机器的实际值一致。

表 3 来自(Bourdaud et al.,2008)的行为模型参数估计

	$\lambda$	$\theta$	$\sigma_d$	$\sigma_0$
真实值	0.9836	50	2.8	4
估计值	0.92	51.37	8.12	N/A
$\beta$ 项	1 2	3 4	5 6	7 8
	0.37	0.19	0.19	0.29
	0.28	0.21	0.29	0.23

### 9.3.2. 自动控制机械选择策略

为了模拟实验，我们运用  $RL$  的概念，结合  $BG$  模型的动态，在每次实验中选择最佳的奖励时段。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 等人,2010; Niv,2009)。由于多巴胺引起的皮质纹状体塑性变化(Reynolds 和 Wickens,2002)通过允许 DA 信号调节皮质纹状体突触的类似 Hebb 塑性而纳入模型(Surmeier、Ding、Day、Wang 和 Shen,2007)。

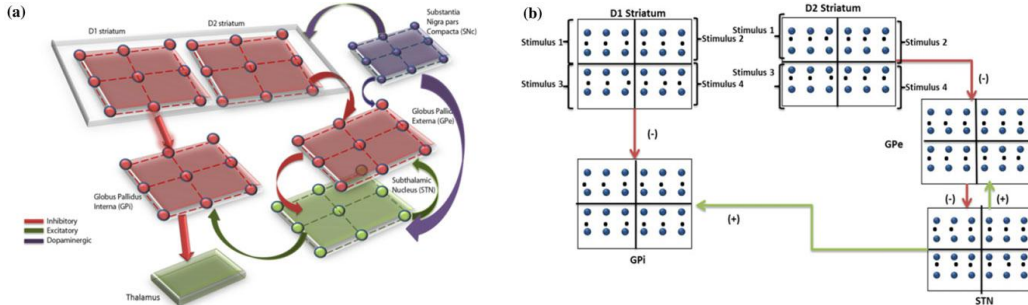


图 47 以纹状体(D1、D2)、STN、GPe、GPi、丘脑等关键核为核心的计算性尖峰  $BG$  模型。兴奋/抑制/调节，谷氨酸/氨基丁酸/多巴胺能投射分别用绿色/红色/紫色箭头表示。b 给出了四层平台对应的  $BG$  模型和每个核内的区域。

拟议网络模型的架构如图 6.1 所示。纹状体的输出(包括 D1 和 D2 部分)被均分为四个象限，接受相应刺激的输入。刺激与 2 个权重  $(w_{i,0}^{D1}, w_{i,0}^{D2})$  相关，初始化值等于 50，代表纹状体中 D1 和 D2 MSNs 的皮质纹状体权重。每一个皮质纹状体权重代表了相应臂的显著性(纹状体峰值率)。这些输出峰值分别由 D1 和 D2 纹状体投射到 GPi 和 GPe。最后一个手臂的选择如第 6.2.4 节所示。所选位置收到的奖励  $r_{i,k}$  从高斯分布中取平均值  $\mu_{i,k}$  和  $SD(\sigma_0)$ (等式 6.3)。

运用输入“ $i$ ”和试验“ $k$ ”获得的奖励，输入到 D1 和 D2 纹状体的位置的预期值，使用以下方程式更新

$$\Delta w_{i,k+1}^{D1} = \eta \delta_k x_{i,k}^{D1} \quad (6.12)$$

$$\Delta w_{i,k+1}^{D2} = -\eta \delta_k x_{i,k}^{D2} \quad (6.13)$$

第  $k$  次试验的预期值( $V_k$ )计算为



$$V_k = \sum_{i=1}^4 w_{i,k}^{D1} * x_{i,k}^{inp} \quad (6.14)$$

第  $k$  次试验收到的回报( $Re_k$ )计算为

$$Re_k = \sum_{i=1}^4 r_{i,k} * x_{i,k}^{inp} \quad (6.15)$$

第  $k$  次试验的误差( $\delta$ )定义为

$$\delta_k = Re_k - V_k \quad (6.16)$$

其中  $w_{i,k}^{D1}$  是在第  $k$  次试验中第  $i$  台机器的 D1 纹状体的皮质-纹状体权重,  $w_{i,k}^{D2}$  是第  $k$  次中第  $i$  台机器的 D2 纹状体的皮质-纹状体权重,  $r_{i,k}$  为第  $k$  次试验中第  $i$  台机器的奖励,  $x_{i,k}^{inp}$  是代表四个自动控制机械的二进制输入向量, 例如, 如果选择第一个自动控制机械, 则  $x_{i,k}^{inp} = [1 \ 0 \ 0 \ 0]$ ,  $\eta (=0.3)$  是 D1 和 D2 纹状体 MSNs 的学习率,  $Re_k$  是所选位置在第  $k$  次试验中收到的回报,  $V_k$  是所选位置在第  $k$  次试验中的预期值。

皮质-纹状体权重更新(公式 6.12 和 6.13)使用误差项“ $\delta$ ”(等式 6.16)。纹状体多巴胺能输入形式的奖励相关信息与误差( $\delta$ )相关(Chakravarthy 等人,2010;Niv,2009)。根据式(6.16)计算的 $\delta$ 既有正值也有负值,没有上下边界,但模型中的起作用的 DA 范围仅限于小正值(0.1-0.9)。因此,从 $\delta$ 到 DA 的映射定义如下:

$$DA = sig(\lambda * \delta_k) \quad (6.17)$$

其中 DA 是范围在 0.1-0.9 内的多巴胺信号,  $\lambda$  是 S 状曲形的斜率( $=0.2$ ),  $\delta_k$  是第  $k$  次试验获得的误差(等式 6.16),  $sig()$  是 S 状曲形的函数。

## 9.4.对策

### 9.4.1.同步

近几十年来,神经同步现象引起了许多计算和实验神经科学家的注意(Hauptmann 和 Tass,2007;Kumar,Cardanobile,Rotter 和 Aertsen,2011;Park 等人,2011;Pinsky 和 Rinzel,1995;Plenz 和 Kital,1999)。据信,部分同步有助于产生各种脑电图节律,如 $\alpha$ 和 $\beta$ (Izhikevich,2007)。由于神经网络在正常功能(四肢协调运动)和病理状态(如癫痫发作期间海马中的 CA3 神经元的同步活动)中的存在,研究神经网络的同步性变得越来越重要(Pinsky 和 Rinzel,1995)。Plenz 和 Kital(1998)提出 STN-GPe 可能作为起搏器(Plenz 和 Kital,1999),在帕金森病等病理条件下产生振荡的来源。Park 等人(2011)报告在接受 DBS 手术的 PD 患者中使用多单元活性电极记录的 STN 神经元与其局部场电位(LFP)之间存在间歇性同步(Park 等人,2011)。他们还通过估计转换率来计算神经元活动中同步和非同步事件的持续时间,转换率是通过使用神经元相位绘制的第一个返回图获得的(Park 等人,2010,2011)。为了观察 STN-GPe 中多巴胺的同步变化,我们计算了在(Pinsky 和 Rinzel,1995)中单个神经元的相位。

第  $j$  个神经元的相位计算如下:

$$\phi_j(t) = 2 * \pi * \frac{(t_{j,k} - t_{j,k})}{(t_{j,k+1} - t_{j,k})} \quad (6.18)$$

$$R^{sync}(t) * e^{i\theta(t)} = \frac{1}{N} \sum_{j=1}^N e^{i\phi_j(t)} \quad (6.19)$$

式中  $t_{j,k}$ 、 $t_{j,k+1}$  为第  $j$  个神经元  $T_{j,k} \in [t_{j,k}, t_{j,k+1}]$  的第  $k$  个和第  $k+1$  个峰值的起始时间,  $\phi_j(t)$  是在时间  $t$  时的第  $j$  个神经元的相位,  $R^{sync}$  为同步测量值  $0 \leq R^{sync} \leq 1$ ,  $\theta$  为神经元的平均相位,  $N$  为网络中的神经元总数。

## 9.5.使用竞赛模型的动作选择

动作选择受 BG 输出核 GPi 的调节，GPi 通过丘脑投射回皮层。我们使用种族模型 (Vickers,1970)进行最终的动作选择,当输出神经元的时间整合神经元活动超过阈值时选择动作(Frank,2006;Frank,Samanta,Moustafa 和 Sherman,2007;Humphries,Khamassi 和 Gurney,2012)。

丘脑神经元的动力学表现为:

$$\frac{dz_k(t)}{dt} = -z_k(t) + f_{GPik}(t) \quad (6.20)$$

$$f_{GPik}^t = \frac{1}{(N * N) / k} \sum_{i=1}^T \sum_{j=1}^{N/k} S_y^{GPik}(t) \quad f_{GPik} = \frac{f_{GPi}^{\max} - f_{GPik}^t}{f_{GPi}^{\max}} \quad (6.21)$$

式中  $z_k(t)$ =第  $k$  次刺激的积分变量,  $f_{GPik}(t)$ =GPi 神经元从纹状体接收第  $k$  次刺激的归一化和逆转平均发放(firing)频率,  $f_{GPi}^{\max}$ =GPi 神经元中的最高发放率,  $S_y^{GPik}$ =接受第  $k$  次刺激的 GPi 神经元的神经元峰值,  $N$ =在 GPi 数组(=50)的单个行/列中的神经元数量,  $T$ =模拟持续时间。

$k$  刺激中第一个超过阈值(=0.15)的神经元( $z_k$ )代表所选的动作。所有代表神经元活动的变量在每次动作选择后立即重置。

## 9.6.结果

我们从神经动力学(STN-GPe)的结果作为 DA 的函数开始,然后给出决策结果。

### 9.6.1.神经动力学

STN 和 GP 的病理振荡与各种 PD 症状相关(Brown,2003; Plenz 和 Kital,1999)。STN 和 GPi 的相关神经功能模式可在多巴胺缺乏的实验条件和帕金森病条件下观察到。在本模型中,我们发现在多巴胺减少的情况下,同步行为增加,类似于帕金森病多巴胺缺乏的情况。通过对  $R_{STN}^{sync}$ 、 $R_{GPe}^{sync}$  和  $R_{STNGPe}^{sync}$  在 DA(0.1~0.9)增加时的值进行估计,研究了 DA 对 STN 和 GPe 神经元同步的影响。

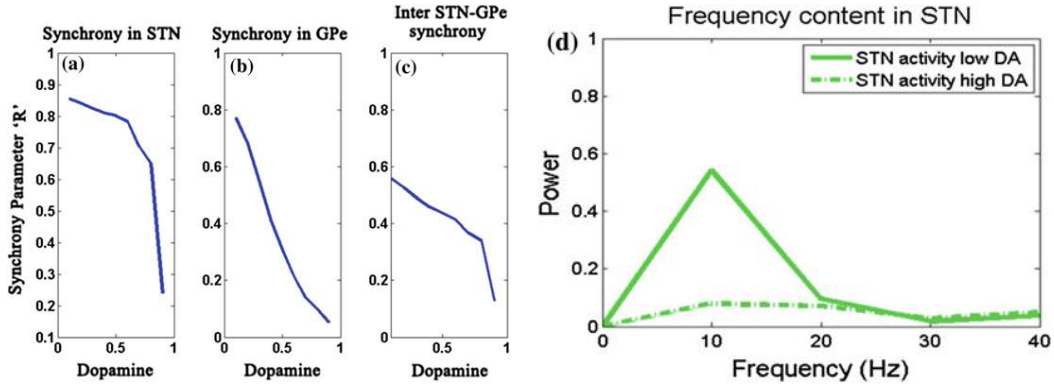


图 48  $R_{STN}^{sync}$  (a)、 $R_{GPe}^{sync}$  (b)和  $R_{STNGPe}^{sync}$  (c)三个同步值的变化,STN 神经元(d)的振荡活动频率随 DA 值(0.1-0.9)的变化。仿真结果表明,随着 DA 的增加,STN 和 GPe 网络之间以及 STN 和 GPe 网络之间的同步降低。

三个“ $R^{sync}$ ”(方程式 6.19)值显示振幅随 DA 水平的增加而减小(图 6.2a-c)。在低 DA 条件下,GPe 活性跟随 STN 活性变化(Plenz 和 Kital,1999),从而形成一种起搏器类型的电路,它可能是 STN-GPe 振荡图 6.2d 的来源。STN 中迸发活性的一个可疑原因可能是在低 DA 水平时,GPe 神经元的抑制减弱(Plenz 和 Kital,1999)。由于较低 DA 水平的 GPe 配置率较小,因此模型捕捉到了这一特征。STN 神经元在低 DA 时的频率约为 10Hz,但在高 DA 水平时则不存在(Kang 和 Lowery,2013)。

### 9.6.2.决策

在神经层次上对模型的性能进行量化后，我们运用两个任务，特别是在“探索性”和“运用性”的动态中，研究了BG在决策中的作用。这项工作延续了我们先前的假设，即探索来自STN-GPe动态(Kalva等人,2012)。第一个任务是一个简单的二元动作选择，类似于Humphries等人(2006)文中描述的那样，其中两个相互竞争的刺激被呈现给模型。输入频率被认为代表“显著性”，较高的频率代表较高的显著性。选择两个可用选择之间具有更高显著性的刺激可以被视为“运用”，而选择不那么显著的一个作为“探索”(Sutton和Barto,1998)。因此，如果选择2号刺激(更突出)，则所选操作定义为“Go-执行”；如果选择1号刺激(不突出)，则定义为“探索”；如果未选择1号刺激(不突出)，则定义为“NoGo-不执行”。对100个试验进行模拟，计算多巴胺水平从低(0.1)到高(0.9)的每种方案(Go-执行、Explore-探索和NoGo-不执行)下选择的行动百分比(图6.3)。我们可以注意到，在没有选择任何作用的情况下，NoGo的概率随着多巴胺的增加而降低；Go的概率随着多巴胺的增加而增加；探索的峰值出现在多巴胺的中间水平(图6.3)。当STN和GPe网络表现出混沌活动时，观测到的DA峰值范围相同。

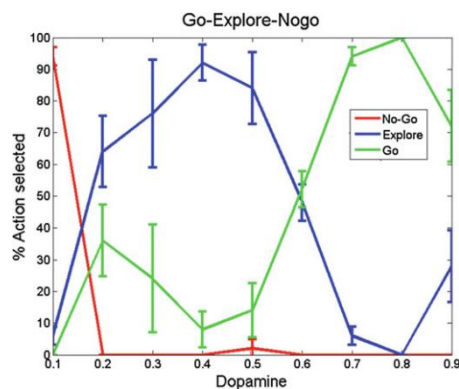


图49 在Go、NoGo中观察到的动作选择的百分比，在 $w_{STN \rightarrow GPe}=1.15$ 且 $w_{STN \rightarrow GPe}=0.8$ 时，平均超过200个试验DP和IP权重值。我们对100个试验进行了模拟，并将其分为4个箱子(每个箱子25个)。然后我们计算了所有DA水平上每种状态的方差。

第二个任务是一个四臂的强盗任务(Bourdaud等人,2008;Daw等人,2006)，类似于现实世界的决策场景。在这项任务中，受试者有四个手臂，其中一个手臂将在每个试验中选择，总共300个试验。每个时段的奖励回报均来自高斯分布，其平均值从实验值变为奖励范围为0至100的实验值。模型的性能(运用率)与行为模型进行了比较，后者代表了N臂强盗任务中的实验数据(图6.4)。控制运用-探索平衡的行为模型的参数“ $\beta$ ”进行了调整，以匹配实验中个体受试者的表现。模型中的探索可以通过增加IP权重(来自STN的影响)或减少DP权重(来自纹状体的影响)来实现。

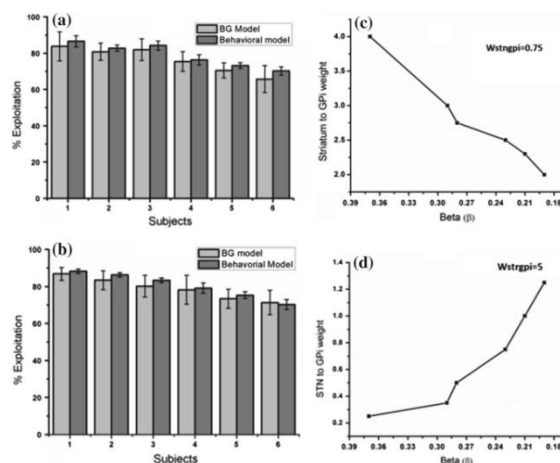


图 50 比较了 BG 模型与行为模型的性能。a. 从 BG 和行为模型中得到的六个被试的运用率。行为模型的  $\beta$  与 DP 权重  $w_{STN \rightarrow GPi}$  的关系, 且  $w_{STN \rightarrow GPi}$  具有恒定的值 (=0.75), 从而得到(a), 如(c)所示。b. 行为模型的  $\beta$  与 IP 权重  $w_{STN \rightarrow GPi}$  的 BG 模型, 且  $w_{STN \rightarrow GPi}$  具有恒定的值 (=5), 从而得到(b), 如(d)所示。Y 轴表示运用百分比, X 轴表示行为模型中特定  $\beta$  值, 和 BG 模型中的 IP 或 DP 权重。

## 9.7. 分析

同步结果与电生理学的一般观察结果一致, 即在中多巴胺水平较高的情况下, STN-GPe 系统显示出不同步的活性, 并且在多巴胺不足的情况下, PD 显示出同步迸发(Bergman 等人,1994; Gillies, Willshaw, Gillies 和 Willshaw, 1998; Park 等人, 2011)。我们观察到 STN 活性呈现振荡活性, 其频率 (=10Hz) 低于实验性 PD 研究中观察到的  $\beta$  频率范围(Weinberger 和 Dostrovsky, 2011)。本研究的目的之一是揭示 STN-GPe 系统的复杂动力学对探索的贡献。为此, 我们首先模拟了二元动作选择任务[类似于 Humphries 等人, (2006)文中描述的那样], 其中显著性被编码为发放率。选择较高的一个被定义为“运用/执行”, 而选择较低的一个被定义为“探索”, 没有选择任何输入作为“不执行”。模型显示, 在低 DA 水平(0.1-0.3)下, “NoGo-不执行”和高 DA 水平(0.7-0.9)下“Go-执行”, 与经典的 BG 函数图一致。除此之外, 还观察到 DA(0.4-0.6)中间水平的“探索”峰值(图 6.3)。为了检查网络中的其他模块是否影响系统的探索, 我们移除了 STN 到 GPi 连接(这有效地消除了 IP)。这种遗漏使得系统只能显示 Go 和 NoGo 机制(不进行探索, 结果不包括在内)。然后我们开始模拟 N 臂强盗任务, 并将模型的性能与实验结果进行了比较。从 BG 模型获得的结果与行为模型(图 6.4)密切匹配, 强化了 STN-GPe 可能是皮质下探索的来源的观点。

## 十、结论(Conclusion)

新的实验数据继续推动 BG 模型的新发展。来自许多领域的的数据, 包括解剖学、生理学、生物化学, 以及临床, 甚至计算证据, 正在为这些核在健康和疾病中功能的其他方面的建模铺平道路。RDDR 模型的建立是为了解释沿纹状体皮质 GPi/SNr 轴神经元数量大幅度减少, 以及与 BG 核内横向连接相关的看似矛盾的解剖学和生理学数据。该模型通过建议 BG 根据强化信号对从皮层接收的信息进行压缩来解决这一矛盾。因此, BG 的漏斗状结构被用来消除皮质神经元活动所跨越的数据空间中的冗余, 并保持最重要的信息(从奖励的角度), 以便有效地规划新的行动。

在本文中, 我们强调了“皮质-STN-苍白球“超直接”通路和基底神经节功能的动态“中心-环绕模型”的功能意义。在下一阶段, 通过观察在执行行为任务时阻断“皮质-STN”神经传递对苍白球活动的影响, 应该验证这种与超直接通路相关的模型是否在生理状态下发挥作用。

净抑制 GPe-STN-GPi 和/或抑制 GPe-GPi 途径对 GPi 活性的贡献也应进行测试。如果这些抑制途径真的起作用的话, GPi 活动应该被 GPe 活动强烈地调节。然而, 迄今为止在皮质刺激和运动任务执行过程中观察到的 GPe 和 GPi 的神经元活动非常相似 (Georgopoulos 等人, 1983; Anderson 和 Horak, 1985; Mitchell 等人, 1987; Hamada 等人, 1990; Nambu 等人, 1990; Mink 和 Thach, 1991; Yoshida 等人, 1993; Turner 和 Anderson, 1997; Nambu 等人, 2000b)。这些神经元的行为就好像它们完全受“皮质-STN-苍白球“超直接”和皮质-纹状体-苍白球“直接”通路的控制。Parent 和 Hazrati (1995) 提出了一个关于“间接”途径存在的问题。为了回答这些问题, 必须更明确地描述“超直接”、“直接”和“间接”通路在行为任务执行期间对苍白球神经元活动的影响。

本文提出的模型可能仍然过于简单化, 忽略了一些神经元连接。丘脑的中心体-丘脑束旁核复合体 (CM-Pf) 接收来自 GPi 和皮质的输入, 大量投射到纹状体, 较少投射到 STN

(Sadikot 等人,1992), 强调纹状体和 STN 作为输入站的另一相似性。虽然这些投射的功能意义没有得到很好的分析, 但它们可以为输入站(即纹状体和 STN) 提供基底神经节的输出, 并且可以作为反馈回路工作。最近, Matsumoto 等人(2001) 已经表明, CM-Pf 中的神经元向纹状体神经元提供有关行为显著的感官事件的信息, 这些信息可以激活纹状体神经元的条件反应。

最后, 我们要指出本文所提出的动态“中心-环绕模型”对于理解基底神经节疾病的病理生理学的有效性。基本运动功能减退的机制或运动过度的障碍目前解释为基底神经节的静态状态的变化即增加或减少意味着 GPi /SNr 神经元的放电频率可能会造成“直接”和“间接”通路之间的不平衡 (DeLong,1990)。然而, 基于动态“中心-环绕模型”(见图 2 b), 帕金森病中观察到的运动缺乏症可以更好地解释为通过时空域的“直接”途径减少了丘脑和皮层的抑制 (Boraud 等人,2000;Nambu 等人,2000a)。基底神经节功能的动态模型可以使我们更好地了解基底神经节疾病的病理生理学, 以及基底神经节的正常功能。

在这一章中, 我们展示了我们使用多个抽象层次上的计算模型对 BG 的“Go-Explore-NoGo”理论的一致性。比较了二元动作选择任务和 N 臂强盗任务两种模型 (Spiking、Hybrid) 的结果。

第一个任务, 二元动作选择任务用于测试模型中多巴胺水平变化时探索状态的存在 (图 9.2)。所有这三个模型都表明, 在较低的 DA 水平上存在“NOGo”机制, 在较高的水平上存在“Go”, 在中等水平上存在“explore”。模型显示, 由于 NoGo 通路的高激活, 低 DA 时选择 NO 输入, 而直接通路的主要影响, 则选择更高的显著刺激。在这两个极端之间, 模型显示了选择较低的显著刺激 (探索), 这被假设是由于 STN-GPe 动力学的原因。为了进一步了解这一点, 我们研究了 STN 和 GPe 神经元的相关性和同步性水平 (ManDALi 和 Chakravarthy, 2015), 并观察到低 DA 水平下的同步性增加。所观察到的 STN-GPe 神经元的非同步动力学通常在低 DA 水平下变化为更同步的 (结果请参考 (ManDALi 等人, 2015))。我们还观察到, STN 络脉的突触强度增加导致 STN-GPe 网络的同步性增加 (图 9.3)。

为了进一步验证 STN-GPe 系统对探索的影响, 我们将 spiking 模型与 RL 方法相结合, 计算平均报酬率, 并对不同突触强度水平下的 n 臂强盗执行集总( $\epsilon$ -greedy)、速率编码和 Izhikevich 模型的结果进行比较(图 9.4)。将不同突触强度的速率编码模型和尖峰模型获得的平均回报与标准 RL“探索-利用”模型 (greedy 和  $\epsilon$ -greedy, 其中  $\epsilon$  控制模型中的探索水平) 进行比较。通过改变突触强度获得的平均回报 (图 9.4a 和 c), 它在内部控制 STN-GPe 动力学, 与不同“ $\epsilon$ ”水平的集总 RL 模型的结果相当。上述结果表明, DA 通过调节 STN-GPe 动力学来控制“探索”状态。此外, STN 络脉的突触强度也控制着探索的百分比。

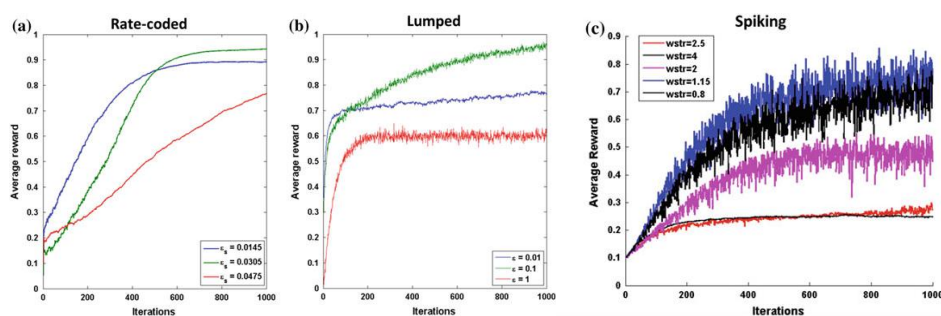


图 51 显示了每个模型的 N 臂强盗任务的平均回报, 当控制探索的横向连接发生变化时, 由速率编码模型得到的平均回报; b: 不同  $\epsilon$  值的集总模型得到的平均回报; 以及 c: 控制探索量的突触权重 ( $w_{str}$ ) 的不同值的 spiking 模型获得的平均奖励

在未来，我们希望将混合模型的 N 臂强盗结果包括在内，由于计算上的限制，我们现在不能这样做。我们计划研究 STN、GPe 神经元单个通道对 N 臂强盗任务中探索的影响。

我们描述了一个基底神经节和相关核的高级模型的机器人嵌入，前提是这些神经回路在动作选择中起着关键作用。该模型面临的任务是在不同的动机和感官输入的情况下，在五个可选的行为子系统之间进行选择，并需要生成机器人行为的一致序列。结果表明，模型基底神经节根据其相对显著性的动态变化，在竞争的子系统之间进行有效的切换。因此，该结构提供了一个存在性证明，证明基底神经节在嵌入物理设备时能够作为有效的动作选择机制发挥作用。此外，通过生成一个行为可以直接观察到的模型（而不仅仅是可以解释为无实体模型），我们能够与动物行为实验的结果进行一些有趣的比较，尤其是关于：

- ①基底神经节在行为排序中的作用；
- ②持续行为过程中基底神经节输入（纹状体和 STN）和输出（SNr）核神经元的活动；
- ③动物在行为冲突情况下的行为。

文献中经常提到的，如本工作所确认的，*Actor-Critic* 体系结构在连续任务中的应用比在离散任务中的应用更困难。关于这个问题，还做了其他几项工作（Doya, 2000）。然而，这些体系结构仍然需要改进，以减少它们的学习时间。

尤其是，我们的“动作规划者”的学习性能似乎仍然远没有真正的小鼠在同一任务中能够达到的学习速度（Albertin 等人，2000），即使我们在模型中使用的高时间常数还不允许进行严格的比较（见附录中的参数表）。这至少可以部分解释为，我们只实施 S-R 学习（或习惯学习），而最近人们知道，大鼠有两种不同的学习系统，与不同的大脑皮层-基底神经节-丘脑环路相关：一种习惯学习系统和目标导向学习系统（Ikemoto 和 Panksepp, 1999; Cardinal、Parkinson、Hall 和 Everitt, 2002）。后者将是快速的，用于学习的早期阶段，并暗示了奖励目标的明确表示或行动结果或有事项的内部表示。前者将非常缓慢，并利用后者，当“动作规划者”达到良好的表现，并能够解决与反应策略（S-R）的任务（Killcross 和 Coutureau, 2003; Yin, Knowlton 和 Balleine, 2004）。

一些理论工作已经开始将 *Actor-Critic* 模型扩展到这一功能差异（Dayan, 2001）。在我们的人工大鼠的实际案例中，这两种系统可以以两种不同的方式使用。

首先，它可能有助于升级探索功能。这个函数可以明确表示环境的不同位置，特别是奖励站点。然后，当动物第一次得到奖励时，探索功能会引导它，尝试让它到达明确记忆的奖励位置的 *Actor*。该函数还可以记住哪些 *Actor* 在不同的领域已经尝试过，但没有成功，因此在探索的情况下，选择未尝试的 *Actor* 而不是随机的 *Actor*。这将加强探索过程，并有望提高动物的学习速度。

目标导向 *Actor* 组件的第二个可能的用途是表示动物正在工作的奖励类型。当一个动物必须处理不同的奖励（食物、饮料）以满足不同的动机（饥饿、口渴）时，这是有用的。在这种情况下，明确选择“动作规划者”作为目标的当前奖励的组件可以选择 *Actor* 的子模块，这些子模块专门用于导致考虑的奖励的行为序列。当人工大鼠 *Psikharpax* 必须在更自然的环境中生存，满足同时发生的动机时，这种改进将成为更现实的验证。

本文综述了 RDDR 模型发展的背景，包括在不同研究领域(第 2 节)和早期皮质 BG 环模型(第 3 节)中获得的实验知识。这一背景，加上强化学习的理论基础(第 4 节)和降维(第 5 节)算法表明，RDDR 模型可以作为我们对健康和疾病中基本神经节功能假设的自然延伸。这一实验和理论背景也是将基本 RDDR 模型(第 6 节)扩展到更全面和更详细模型(第 7 节)的驱动力。基本的 RDDR 模型是一个简单的神经网络，它代表了 BG 作为一个由增强信号调制的



中心降维系统的一般概念。这种简化的观点仍然能够提供对 *BG* 解剖和生理学令人费解的方面的见解，例如其巨大的横向网络，导致神经元之间没有相关性和似乎很少的相互作用。RDDR 模型的不同提升增加了它的应用范围，并被设计成将模型从概念上的模型转移到与 *BG* 生物学紧密相连的模型。这些不同进展的各种计算方面为这些核不同性质的可能作用提供了新的线索。

RDDR 为几个验证实验奠定了基础。第一个实验涉及网络的假设动力学。在学习阶段，神经元之间的相关性(Arkadir 等人,2002)和横向权重的强度预计会显著增加。因此，在这样一个学习期内，这两个参数应该改变，这可能是由于测试一只幼小的动物(Tepper 和 Trent,1993)或一只暴露在全新环境中的动物引起的。不同的(或新颖的)运动和感觉映射预计将完全改变动物(因此 *BG*)体验的输入空间结构。第二个实验是关于学习规则的。Hebb 和反 Hebb 的学习规则还没有被记录在这个大脑区域，因此他们的调制增强信号。

本综述的关键信息是需要使用最先进的分析和计算技术探索新的实验数据。对 *BG* 的新探索以及它们与大脑其他部分的相互作用，对于现有概念的逐步进化是至关重要的。像 RDDR 模型这样的新模型可以满足两个关键需求：作为健康和疾病中 *BG* 行为的一般概念，以及作为生成特定实验和测试预期(或意外)结果的工具。这些新的实验将反过来促进改进模型的形成，这将代表 RDDR 模型的进化，或导致对 *BG* 的新的、不同的和更好的理解的革命。