基底节的 Actor-Critic 模型：新的解剖和计算观点

Actor–critic models of the basal ganglia: new anatomical and computational perspectives

Daphna Joel[a],*, Yael Niv[a], Eytan Ruppin[b]

[a]Department of Psychology, Tel-Aviv University, Ramat-Aviv, Tel Aviv 69978, Israel

[b]Schools of Medicine and Mathematical Sciences, Tel-Aviv University, Tel Aviv 69978, Israel

（Song Jian，translate）

摘要:近年来发展了大量的基底神经节信息处理的计算模型。其中最突出的是基底神经节功能的 Actor-Critic 模型，该模型建立在多巴胺神经元活动与 Critic 中的时间差预测误差信号，以及纹状体中多巴胺依赖的长期突触可塑性和由 Actor 中的预测误差信号引导的学习之间的强相似性上。我们有选择地回顾了几种基底神经节的 Actor-Critic 模型，重点关注两个重要方面：Critic 模型复制多巴胺释放的时间动态的方式，以及 Actor 模型考虑已知基底神经节解剖和生理学的程度。为了补充将基底神经节机制与强化学习（RL）联系起来的已有成果，我们引入了一种替代方法来建模 Critic 网络，该方法使用进化计算（进化计算）（Evolutionary Computation，EC）技术“进化”出一个最优的 RL 机制，并将进化机制与 Critic 家的基本模型联系起来。我们对 Critic 模型的讨论是通过对 Critic 在基底神经节电路中实施的解剖学合理性的批判性讨论来结束的，并且得出这样的实施建立在与基底神经节已知解剖学不一致的假设之上。我们回到 Actor-Critic 模型的 Actor 组件，通常在纹状体级别建模，但细节很少。我们描述了基底神经节的一个替代模型，该模型考虑了基底神经节-丘脑皮质连接的一些重要的和以前被忽视的解剖学和生理学特征，并建议基底神经节执行皮质输入的强化-偏重的降维。我们进一步认为，由于这种选择性编码可能使额叶皮质层的表现偏向于奖励计划和行动的选择，增强-驱动的降维框架可能作为基底神经节因子模型的基础。最后，我们简要讨论了多巴胺信号在 RL 和 Actor 转换中的双重作用。©2002 爱思唯尔科技有限公司版权所有。

关键词：*基底神经节；多巴胺；强化学习；Actor-Critic；降维；进化计算；Actor 切换；纹状小体/斑块*

一、引言

近年来发展了大量的基底神经节信息处理计算模型（Houk、Adams 和 Barto,1995；基底神经节连接的普遍方案见图 1）。最近的一篇评论将这些模型分为三大类（不是相互排斥的）：串行处理模型、动作选择模型和强化学习模型（RL）（Gillies 和 Arbuthnott,2000）。第一类包括在产生活动模式序列中为基底神经节环结构分派为中心作用任务的模型（Berns 和 Sejnowski,1998）。第二类重点关注主要基底神经节输出核对其靶点施加的强直抑制活性，并假设其通过集中的去抑制提供了动作选择（Gurney、Prescott 和 Redgrave，2001 年）。本文着重研究了第三类模型，它们在 RL 中对基底神经节分派了重要作用任务。

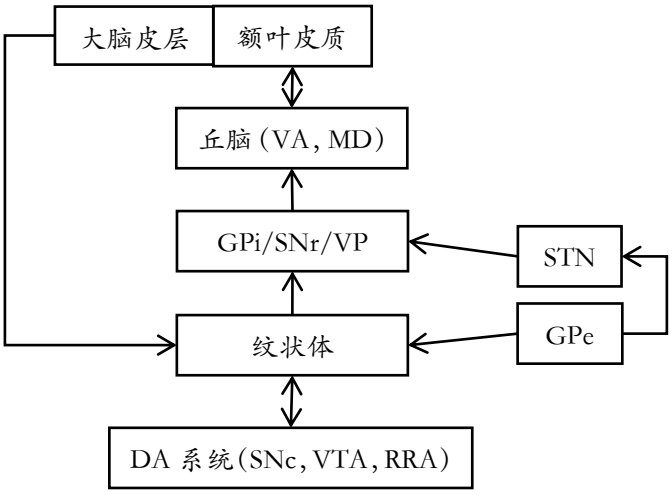

图 1：基底神经节-丘脑皮质连接的一般方案。纹状体是基底神经节的主要输入结构。分为背纹状体（主要由尾状和壳核构成）和腹纹状体（伏隔核和尾状和壳核的腹内侧部分）。纹状体由整个大脑皮层支配，投射到基底神经节、苍白球（GPi）、黑质网状部分（SNr）和腹侧苍白球（VP）的输出核。这些核依次向腹前核（VA）和中嗅丘脑核（MD）突出，它们与额叶皮质相互连接。来自纹状体的信息也可以通过“间接途径”到达输出核，即通过纹状体投射到苍白球的外段（GPe）、GPe 投射到下丘脑核（STN）以及后者投射到 GPi/SNr/VP。纹状体还投射在黑质致密部（SNc）、红质后区（retrorubral area,RRA）和腹侧被盖区（VTA）的多巴胺能神经元。请注意，本计划不涉及两个重要的原则组织的描绘的预测。一种是纹状体背侧的区域组织，分为条纹体（大鼠的斑块）和基质。另一种是将不同层次的投影组织成几个“流”，形成几个神经节-丘脑皮质回路。（有关基底神经节-丘脑皮质连接组织的广泛研究，见 Alexander 和 Crutcher,1990;Gerfen,1992;Joel 和 Weiner,1994,1997,2000;Parent,1990）。

Wolfram Schultz 的开创性研究引起了人们对基底神经节 RL 模型的兴趣，该研究提供了实验证据，表明 RL 在基底神经节处理过程中起着重要作用（Schultz 和 Dickinson,2000;Schultz、Tremblay 和 Hollerman,2000）。Schultz 和他的同事记录了猴子在 Actor 任务获得和执行过程中多巴胺能（DA）神经元的活动，他们发现 DA 神经元对初级奖励有阶段性的反应，随着实验的进展，这些神经元的反应逐渐从初级奖励变回奖励性刺激。DA 神经元的激发模式也被发现可以反映延迟奖励时间的信息（相对于奖励预测刺激），这可以从忽略预期奖励时的准确时间减缓中看到。这种活动模式与 RL 计算算法生成的模式非常相似，特别是时间差（TD）模型（Sutton,1988），如本期另一篇论文（Suri,2002）中所详细描述的。

在基底神经节建模的背景下，TD 学习主要用于 Actor-Critic 模型的框架（Barto,1995;Houk 等人,1995）。在这种模型中，Actor 子网络学习执行操作，以最大化未来奖励的加权和，由 Critic 子网络在每个时间步长计算（Barto,1995）。Critic 是适应性的，因为它学会了根据当前的感官输入和动作的策略预测未来奖励的加权和，通过一个迭代过程，在这个过程中，他将自己的预测与行动代理获得的实际奖励进行比较。自适应 Critic 使用的学习规则是 TD 学习规则（Sutton,1988），其中两个相邻预测之间的误差（TD 误差）用于更新 Critic 家的权重。许多研究表明，使用这种错误信号来训练动 Actor 产生非常有效的 RL（Kaelbling、Littman 和 Moore,1996;Tesarou,1995;Zhang 和 Dieterich,1996）。

基底神经节和 Actor-Critic 模型之间的类比建立在 DA 神经元活性和 TD 预测误差信号，以及纹状体中依赖于 DA 的长期突触可塑性（Calabresi 等人,2000;Wickens,Begg 和 Arbuthnott,1996）之间的强相似性，以及由 Actor 的预测错误信号引导的学习。近年来，基底神经节功能的 Actor-Critic 模型得到了广泛的应用，并提出了几种模型。这些模型之间的比较表明，它们主要在两个重要方面存在差异。Critic 的模型在重现 DA 激发的时间动态的方式上是不同的，也就是说，在负责产生 DA 神经元对不可预测的奖励和奖励预测刺激的短期反应的网络架构中，以及奖励遗漏引起的减缓。Actor 的模型在考虑已知基底神经节解剖和生理的程度上有所不同。

在第 2 节中，我们简要回顾了基底神经节的几个 Actor-Critic 模型，重点介绍了负责再现 DA 激发的时间动态的机制和 Actor 的结构。第 3 节介绍了一种对 Critic 网络建模的替代方法，该方法使用进化计算技术来进化出一个最优 RL 机制。这一机制与第 2 节中提出的更经典的 Critic 模型有关。第 4 节对基底神经节电路中实施适应性 Critic 的解剖学合理性进行了关键性讨论。在第 5 节中，我们回到 Actor-Critic 模型的 Actor 组成部分，并描述了基底神经节的替代模型，该模型考虑了基底神经节-丘脑皮质连接的一些重要的和以前被忽视的解剖和生理特征。该模型将基底神经节的主要计算作用视为一个降维编码-解码皮质-纹状体-丘脑皮质环的关键点。最后，我们简要讨论了 DA 信号在 RL 和 Actor 切换中的双重作用。

二、基底神经节强化学习的 Actor-Critic 模型

2.1 Houk，Adams 和 Barto（1995）

Houk 等人（1995）提出了基底神经节的第一个 Actor-Critic 模型。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。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 等人仅提供在基底神经节电路中实施该 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）。

2.2 Suri 和 Schultz（1998、1999）

Suri 和 Schultz 扩展了 Barto（1995）提出的基础 Actor-Critic 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。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）利用这一 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基底神经节的结构中实现 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。而不是通过在与基底神经节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。

2.3 Contreras Vidal 和 Schultz（1999）

Contreras Vidal 和 Schultz（1999）提供了一种与基底神经节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。

2.4 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 对预测奖励的反应的抑制和 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。

2.5 Suri、Bargas 和 Arbib（2001）

在最近的一篇论文中，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 是基于对基底神经节-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基底神经节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上下状态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上下状态的随机转换。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。

三、强化学习的演变——对 Critic 建模的另一种方法

我们（Niv、Joel、Meiligson 和 Ruppin,2002）采用了一种替代的方法来模拟 RLCritic。利用遗进化计算技术，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。利用遗传算法（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 Listor,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）。

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，以在基底神经节-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。为了提高模型在基底神经节学习中的相关性，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。

四、基底神经节中的 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）？

4.1 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，换句话说，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基底神经节-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基底神经节计算中的作用。

4.2 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、Galis 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,Apiccela，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，我们要注意的是，尽管上述建议尊从已知的解剖结构，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，尽管基底神经节与 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基底神经节的功能，并为理论模型提供额外的约束。

五、增强驱动的降维——基底神经节奖励偏向表示

与比较先进的基底神经节处理模型相比，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 在基底神经节-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，我们将提出一个基底神经节处理的模型，该模型可能被扩展为基底神经节 Actor 模型的基础。根据已知基底神经节解剖和生理学的几个重要限制条件，该模型建议基底神经节对皮质表征（Bar-Gad、Havazelet Heimer、Ruppin 和 Bergman,2000；

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基底神经节-丘脑皮质电路：

①基底神经节的漏斗状结构。

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、Fenelon 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-状状体-苍白质-丘脑-皮质路在苍白质水平后径逐渐扩大（Arecchi Bouchhioua、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基底神经节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，神经网络可以利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，该网络利用横向抑制提取主成分子空间（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 模型表明基底神经节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 等人假设基底神经节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，额叶皮质利用它来计划未来的行动。因此，RDDR 网络利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，基底神经节的部分皮质输入来自额叶皮质，可能代表计划和行动。因此，基底神经节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 框架可以作为基底神经节 Actor 模型的基础。

六、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响应作为增强信号。这一假设是 DA 在学习中起核心作用的观点的一个证明（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，在奖励发放之后从按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。

七、结论

我们对基底神经节的 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利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

（例如来自基底神经节）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。未来研究的目标是以更详细和可靠的方式模拟已知的基底节解剖和生理，并解决基底神经节-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）。我们已经描述了基底神经节-丘脑皮质连接的模型，这表明基底神经节对皮质输入进行了增强偏倚的降维（Bar Gad 等人，2000）。该 RDDR 框架可作为未来基底神经节 Actor 模型的基础。

总之，基底神经节的 Actor-Critic 模型通过将基底神经节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，有助于我们对基底神经节功能的思考。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基底神经节将回答这些问题。

参考文献

Alexander, G. E., & Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: Neural substrates of parallel processing. Trends in Neurosciences, 13, 266–271.

Arecchi-Bouchhioua, P., Yelnik, J., Francois, C., Percheron, G., & Tande, D. (1996). 3-D tracing of biocytin-labelled pallido-thalamic axons in the monkey. Neuroreport, 7, 981–984.

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

Bailey, C. H., Giustetto, M., Huang, Y., Hawkins, R. D., & Kandel, E. R. (2000). Is heterosynaptic modulation essential for stabilizing hebbian plasticity and memory? Nature Reviews Neuroscience, 1, 11–20.

Barto, A. G. (1995). Adaptive critic and the basal ganglia. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), Models of information processing in the basal ganglia (pp. 215–232). Cambridge: MIT Press.

Berendse, H. W., Galis-de Graaf, Y., & Groenewegen, H. J. (1992). Topographical organization and relationship with ventral striatal compartments of prefrontal corticostriatal projections in the rat. Journal of Comparative Neurology, 316, 314–347.

Berns, G. S., & Sejnowski, T. J. (1998). A computational model of how the basal ganglia produce sequences. Journal of Cognitive Neuroscience, 10, 108–121.

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? Brain Research Review, 28, 309–369.

Brown, J., Bullock, D., & Grossberg, S. (1999). How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. Journal of Neuroscience, 19, 10502–10511.

Bunney, B. S., Chiodo, L. A., & Grace, A. A. (1991). Midbrain dopamine system electrophysiological functioning: A review and new hypothesis. Synapse, 9, 79–94.

Calabresi, P., Gubellini, P., Centonze, D., Picconi, B., Bernardi, G., Chergui, K., Svenningsson, P., Fienberg, A. A., & Greengard, P. (2000). Dopamine and cAMP-regulated phosphoprotein 32 kDa controls both striatal long-term depression and long-term potentiation, opposing forms of synaptic plasticity. Journal of Neuroscience, 20, 8443–8451.

Carpenter, G. A., & Grossberg, S. (1987). Self organization of stable category recognition codes for analog input patterns. Applied Optics, 3, 4919–4930.

Contreras-Vidal, J. L., & Schultz, W. (1999). A predictive reinforcement model of dopamine neurons for learning approach behavior. Journal of Comparative Neuroscience, 6, 191–214.

Dittman, J. S., & Regehr, W. G. (1997). Mechanism and kinetics of heterosynaptic depression at a cerebellar synapse. Journal of Neuroscience, 17, 9048–9059.

Fellous, J.-M., & Linster, C. (1998). Computational models of neuromodulation: A review. Neural Computation, 10, 791–825.

Foldiak, P. (1990). Forming sparse representations by local anti-Hebbian learning. Biological Cybernetics, 64, 165–170. Gerfen, C. R. (1984). The neostriatal mosaic: Compartamentalization of corticostriatal input and striatonigral output systems. Nature, 311, 461–464.

Gerfen, C. R. (1985). The neostriatal mosaic. I. Compartamental organization of projections from the striatum to the substantia nigra in the rat. Journal of Comparative Neurology, 236, 454–476.

Gerfen, C. R. (1992). The neostriatal mosaic: Multiple levels of compartmental organization in the basal ganglia. Annual Review of Neuroscience, 15, 285–320.

Gerfen, C. R., Herkenham, M., & Thibault, J. (1987). The neostriatal mosaic II. Patch- and matrix- directed mesostriatal dopaminergic and non-dopaminergic systems. Journal of Neuroscience, 7, 3915–3934.

Gillies, A., & Arbuthnott, G. (2000). Computational models of the basal ganglia. Movement Disorders, 15, 762–770.

Groenewegen, H. J., Berendse, H. W., Wolters, J. G., & Lohman, A. H. M. (1990). The anatomical relationship of the prefrontal cortex with the striatopallidal system, the thalamus and the amygdala: evidence for a parallel organization. Progress in Brain Research, 85, 95–118.

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

Haber, S. N., Fudge, J. L., & McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. Journal of Neuroscience, 20, 2369–2382. Hammer, M. (1997). The neural basis of associative reward learning in honeybees. Trends in Neuroscience, 20, 245–252.

Houk, J. C., Adams, J. L., & Barto, A. G. (1995). A model of how the basal ganglia generate and use reward signals that predict reinforcement. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), Models of information processing in the basal ganglia (pp. 249–270). Cambridge: MIT Press.

Jaeger, D., Kita, H., & Wilson, C. J. (1994). Journal of Neurophysiology, 72, 2555–2558. Joel, D., & Weiner, I. (1994). The organization of the basal ganglia– thalamocortical circuits: Open interconnected rather than closed segregated. Neuroscience, 63, 363–379.

Joel, D., & Weiner, I. (1997). The connections of the primate subthalamic nucleus: Indirect pathways and the open-interconnected scheme of basal ganglia–thalamocortical circuitry. Brain

Research Review, 23, 62–78.

Joel, D., & Weiner, I. (2000). The connections of the dopaminergic system with the striatum in rats and primates: An analysis with respect to the functional and compartmental organization of the striatum. Neuroscience, 96, 451–474.

Joel, D., Avisar, A., & Doljansky, J. (2001). Enhancement of excessive lever-pressing after post-training signal attenuation in rats by repeated administration of the D1 antagonist SCH 23390 or the D2 agonist quinpirole but not of the D1 agonist SKF 38393 or the D2 antagonist haloperidol. Behavioural Neuroscience, 115, 1291–1300.

Kaelbling, L. P., Littman, M. L., & Moore, A. (1996). Reinforcement learning: A survey. Journal of AI Research, 4, 237–285. Kalivas, P. W. (1993). Neurotransmitter regulation of dopamine neurons in the ventral tegmental area. Brain Research Review, 18, 75–113.

Kincaid, A. E., Zheng, T., & Wilson, C. J. (1998). Connectivity and convergence of single corticostriatal axons. Journal of Neuroscience, 18, 4722–4731.

Kita, H. (1996). In C. Ohye, M. Kimura, & J. S. McKenzie (Eds.), The basal ganglia V (pp. 77–94). New York: Plenum Press. Kung, S. Y., Diamantars, K. I (1990). IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 2, pp. 861–864).

Le Moal, M., & Simon, H. (1991). Mesocorticolimbic dopaminergic network: Functional and regulatory roles. Physiological Review, 71, 155–234.

Lyon, M., & Robbins, T. W. (1975). The action of central nervous system stimulant drugs: A general theory concerning amphetamine effects (Vol. 2) (pp. 80–163).

Current developments in Psychopharmacology, New York: Spectrum. Menzel, R., & Muller, U. (1996). Learning and memory in honeybees: From behavior to neural substrates. Annual Review of Neuroscience, 19, 379–404.

Montague, P. R., Dayan, P., Person, C., & Sejnowski, T. J. (1995). Bee foraging in uncertain environments using predictive Hebbbian learning. Nature, 377, 725–728.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framewok for mesencephalic dopamine systems based on predictive hebbian learning. Journal of Neuroscience, 16, 1936–1947.

Niv, Y., Joel, D., Meilijson, I., & Ruppin, E. (2002). Evolution of reinforcement learning in uncertain environments: a simple explanation for complex foraging behaviors. Adaptive Behavior, in press. Oades, R. D. (1985). The role of noradrenaline in tuning and dopamine in switching between signals in the CNS. Neuroscience Biobehavioural Review, 9, 261–282.

Oja, E. (1982). A simplified neuron model as a principal component analyzer. Journal of Mathematical Biology, 15, 267–273.

Oorschot, D. E. (1996). Total number of neurons in the neostriatal, pallidal, subthalamic, and substantia nigral nuclei of the rat basal ganglia: A stereological study using the cavalieri and optical disector methods. Journal of Comparative Neurology, 366, 580–599.

Overton, P. G., & Clark, D. (1997). Burst firing in midbrain dopaminergic neurons. Brain Research Review, 25, 312–334. Parent, A. (1990). Extrinsic connections of the basal ganglia. Trends in Neuroscience, 13, 254–258.

Percheron, G., Francois, C., Yelnik, J., Fenelon, G., & Talbi, B. (1994). The basal ganglia related systems of primates: definition, description and informational analysis. In G. Percheron, J. S. McKenzie, & J. Feger (Eds.), The basal ganglia IV: New ideas and data on structure and function (pp. 3–20). New York: Plenum Press.

Pucak, M. L., & Grace, A. A. (1994). Regulation of substantia nigra dopamine neurons. Critical Review Neurobiology, 9, 67–89. Redgrave, P., Prescott, T. J., & Gurney, K. (1999). Is the short-latency dopamine response too short to signal reward error? Trends in Neuroscience, 22, 146–151.

Robbins, T. W., & Everitt, B. J. (1982). Functional studies of the central catecholamines. International Review of Neurobiology, 23, 303–365.

Robbins, T. W., & Everitt, B. J. (1996). Neurobehavioural mechanisms of reward and motivation. Current Opinion in Neurobiology, 6, 228–236.

Rolls, E. T., & Johnstone, S. (1992). Neurophysiological analysis of striatal function. In C. Wallesch, & G. Vallar (Eds.), Neuropsychological disorders with subcortical lesions (pp. 61–97). Oxford: University Press.

Salamone, J. D. (1994). The involvement of nucleus accumbens dopamine in appetitive and aversive motivation. Behavioural Brain Research, 61, 117–133.

Schacher, S., Wu, F., & Sun, Z.-Y. (1997). Pathway-specific synaptic plasticity: Activity-dependent enhancement and suppression of longterm heterosynaptic facilitation at converging inputs on a single target. Journal of Neuroscience, 17, 597–606.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. Journal of Neurophysiology, 80, 1–27. Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. Annual Review Neuroscience, 23, 473–500.

Schultz, W., Apiccela, P., Scarnati, E., & Ljungberg, T. (1992). Neuronal activity in monkey ventral striatum related to the expectation of reward. Journal of Neuroscience, 12, 4595–4610.

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

Schultz, W., Tremblay, L., & Hollerman, J. R. (1998). Reward prediction in primate basal ganglia and frontal cortex. Neuropharmacology, 37, 421–429.

Schultz, W., Tremblay, L., & Hollerman, J. R. (2000). Reward processing in primate orbitofrontal cortex and basal ganglia. Cereberal Cortex, 10, 272–283.

Sidibe, M., Bevan, M. D., Bolam, J. P., & Smith, Y. (1997). Efferent connections of the internal globus pallidus in the squirrel monkey. 1. Topography and synaptic organization of the pallidothalamic projection. Journal of Comparative Neurology, 382, 323–347.

Suri, R. E. (2002). TD models of reward predictive responses in dopamine neurons. Neural Networks, 15, PII: S0893-6080(02)00046-1. Suri, R. E., & Schultz, W. (1998). Learning of sequential movements by neural network model with dopamine-like reinforcement signal. Experimental Brain Research, 121, 350–354.

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

Suri, R. E., Bargas, J., & Arbib, M. A. (2001). Modeling functions of striatal dopamine modulation in learning and planning. Neuroscience, 103, 65–85. Sutton, R. (1988). Learning to predict by methods of temporal difference. Machine Learning, 3, 9–44.

Tesauro, G. (1995). Temporal difference learning and TD-Gammon. Communications of the ACM, 38, 58–68. Uylings, H. B. M., & van Eden, C. G. (1990). Qualitative and quantitative comparison of the prefrontal cortex in rat and in primates, including humans. Progress in Brain Research, 85, 31–62.

Van den Bos, R., & Cools, A. R. (1989). The involvement of the nucleus accumbens in the ability of rats to switch to cue-directed behaviors. Life Science, 44, 1697–1704.

Vogt, K. E., & Nicoll, R. E. (1999). Glutamate and gama-aminobutyric acid mediate a heterosynaptic depression at mossy fober synapses in the hippocampus. Proceedings of the National Academic Science, USA, 96, 1118–1122.

Weiner, I. (1990). Neural substrates of latent inhibition: The switching model. Psychological Bulletin, 108, 442–461.

Weiner, I., & Joel, D. (2002). Dopamine in schizophrenia: Dysfunctional information processing in basal ganglia–thalamocortical split circuits. In G. Di Chiara (Ed.), Handbook of experimental pharmacology: Dopamine in the CNS, (pp. 417–472). Berlin: Springer.

White, N. M. (1997). Mnemonic functions of the basal ganglia. Current Opinions in Neurobiology, 7, 164–169.

Wickens, J. R., Begg, A. J., & Arbuthnott, G. W. (1996). Dopamine reverses the depression of rat corticostriatal synapses which normally follows high-frequency stimulation of cortex in vitro. Neuroscience, 70, 1–5.

Wickens, J., & Kotter, R. (1995). Cellular models of reinforcement. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), Models of information processing in the basal ganglia, (pp. 187–214). Cambridge, MA: MIT Press.

Yelnik, J., Francois, C., Tand, D (1997). Proceedings if the Third Congress of European Neuroscience Society, Beaurdeax. Yeterian, E. H., & Pandya, D. N. (1991). Prefrontostriatal connections in relation to cortical architectonic organization in rhesus monkeys. Journal of Comparative Neurology, 312, 43–67.

Zald, D. H., & Kim, S. W. (2001). The orbitofrontal cortex. In S. P. Salloway, P. F. Malloy, & J. D. Duffy (Eds.), The frontal lobes and neuropsychiatric illness (pp. 33–69).

Washington, DC: American Psychiatric Publishing. Zhang, W., & Dietterich, T. G. (1996). High performance job shop scheduling with a time delay TD network. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), (Vol. 8) (pp. 1024–1030). Advances in neural information processing systems, Cambridge: MIT Press.