评估自由选择权的强化学习机制

A Reinforcement Learning Mechanism Responsible for the Valuation of Free Choice

Jeffrey Cockburn,[1] Anne G.E. Collins,[1] and Michael J. Frank[1,*]

[1]*Department of Cognitive, Linguistic and Psychological Sciences; Brown Institute for Brain Science, Brown University, Providence, RI 02912, USA*

Accepted: 2014 by Neuron

(translated by zang jie)

摘要：比起同等价值没有选择的选项，人们更喜欢自由选择的选项。然而，驱动这种偏好的神经机制及其功能意义尚待确定。在这里，我们提出了一个模型，在该模型中，由于与自由选择相关的正向奖励预测误差的放大而产生选择偏好。使用概率学习任务的新变体，我们表明选择偏向对主要与积极结果相关的选择具有选择性。发现 DARPP-32 的一个多态性是一个与多巴胺能纹状体可塑性和强化学习中的个体差异相关的基因，可以预测选择的价值效应。我们认为这些选择偏好是信用分配机制的行为副产品，该信用分配机制负责确保向纹状体发送有效的多巴胺能强化学习信号。

1、引言

有机体的适应能力取决于其在寻求报酬时避免危害的能力（Orr，2009）。鉴于此，选择是一种具有极大优势的能力，因为它提供了一个有机体可以根据其需求操纵环境的条件。但是，选择的优势是有代价的。识别需求，机会，候选行动以及在其中进行选择相关的认知需求表明：与简单的刺激驱动的响应相比，选择控制的行为将具有更高的要求。确实，有证据表明，复杂的选择可能令人反感（Iyengar 和 Lepper，2000 年）。然而，人类和动物都表现出对选择（Bown 等，2003；Leotti 和 Delgado，2011，2014）和自由选择的选项（Egan 等，2007；Lieberman 等，2001；Sharot 等，2009，2010）的偏爱。

人们通过认知失调理论的观点来观察对自由选择的选择的偏好，通过重新评估那些选择以支持已选选项，从而消除了选择同等价值的选择所带来的心理紧张（Festinger，1962）。特维尔斯基（Tversky，1972）提出了类似的重新评估思路，但认为选择过程会改变归因于选项特征的重要性，因此，选择后的评价会有不同情况：特征权重会变得有利于所选选项。最近，研究表明，人类不仅喜欢他们已经选择的选择，也喜欢选择本身（Bown 等，2003）。已发现纹状体血氧水平依赖性（BOLD）信号与选择后选项估值的变化（Sharot 等，2009）和选择偏好（Leotti 和 Delgado，2011，2014）相关。但是，尚不清楚这些偏好出现的神经机制，以及其功能意义。在这里，我们好奇选择偏好是否得出更通用的适应机制。

我们旨在确定基底神经节（BG）中的强化学习（RL）过程的计算机制是否可以解释这些发现。我们假设自由选择偏好是涉及 BG 和中脑多巴胺（DA）系统的反馈回路的行为副产物，该机制通过自由选择后优先放大 DA 细胞编码的正向奖励预测错误（RPEs）（图 2A）。我们认为该反馈回路通过提供一条通道，使多巴胺能学习信号优先以 BG 为目标当 BG 产生积极结果的内源性行动时，从而减轻大脑中的信用分配问题。

我们的假设是基于三个主要发现。首先，外源驱动的行为是皮质控制的，而内源性选择驱动的行为则依赖于额外的 BG 回路（Brown 和 Marsden，1998；Francoois-Brosseau 等，2009）。其次，人纹状体中 BOLD 信号的变化与选择的预期（Leotti 和 Delgado，2011，2014）以及对自由选择的选项偏好（Sharot 等，2009）相关。第三，在啮齿类动物中，纹状体而不是额叶的 DA 随选择的增加而增加（StOnge 等人，2012）。总之，这些发现表明选择参与 BG 并影响纹状体 DA 水平。

解剖工作指向 BG 可以调节多巴胺能信号的机制。黑质网状组织（SNr）中的具有调性活性的细胞向黑质致密性组织（SNc）的 DA 细胞上发送抑制性投射（Joel 和 Weiner，2000 年）。SNr 活性的降低（如通过 BG 控制某个动作时发生的情况）会降低 SNr 对 SNc 的抑制作用，从而促进 DA 释放到纹状体中（Lee 等，2004）。换句话说，SNr 中断了 SNc 活动。当 BG 门控一个动作时，此中断被释放，从而在 DA 细胞被额外传入的 SNc 输入驱动爆发时，增加了 DA 释放到纹状体的上限。

这些结构的生物物理模型表明，与动作选择相关的纹状体活性会抑制 SNr，进而抑制 SNc 细胞，从而增加阶段性 DA 爆发（Lobb 等，2011）。此外，已经证明将

这种机制整合到 BG 的生物学约束模型中可以提高学习信号的保真度并改善复杂环境中的性能（O’Reilly 和 Frank，2006 年）。

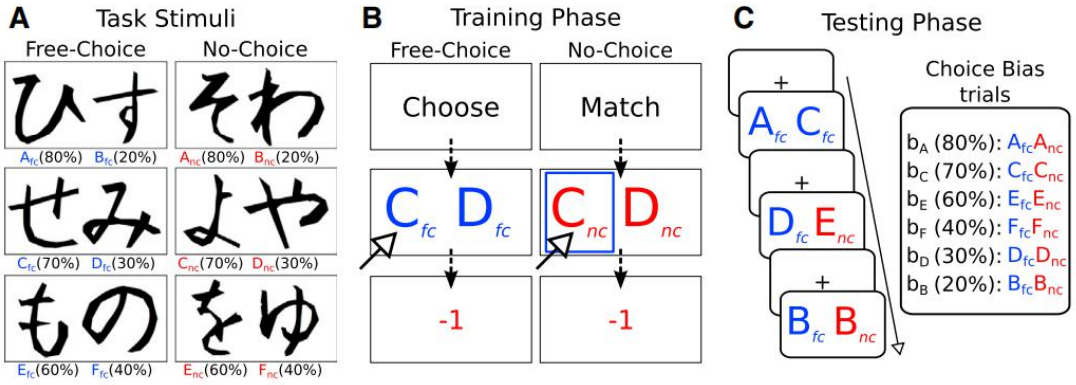

图 1. 实验任务设计

（A）在任务中使用的示例自由选择（fc）和非选择（nc）刺激，并显示了相关的奖励概率。（B）训练阶段：每个试验提供一对刺激。要求参与者选择两个可用选项之一。在进行刺激之前，向参与者发出自由选择（选择）或不选择（匹配）条件的提示。在自由选择试验中，参与者可以自由选择任何一种选择，但在非选择试验中，参与者被迫选择框架刺激。概率反馈跟随选项的选择。（C）测试阶段：反复要求参与者在所有可能的选项配对中选择最佳选项。参与者可以在所有试验中自由选择刺激方案，但没有提供反馈。选择偏倚是根据将相同奖励的自由选择和无选择选项配对的试验的性能进行量化的。

根据这些观察结果，我们假设当与 BG 门控动作相关时，阶段性 DA 发放优先被放大。因此，门控动作应相对于非动作产生虚高的价值，这在行为上会成为对自由选择选项的偏好。这种机制意味着选择偏差的大小应由 RPE 的历史来确定。因此，我们旨在系统地评估一系列选项价值和 RPE 历史上的偏差。如果选择偏好由 BG 中的多巴胺能学习控制，我们还认为多巴能纹状体可塑性和奖励学习的遗传变异应可预测个体选择偏好的差异。具体来说，我们专注于 DARPP-32 基因，该基因与奖励学习和追求（而非避免）选择的个体差异相关（Doll 等，2011；Frank 等，2007，2009）。；Stipanovich 等，2008）。

我们通过施用一种概率学习任务的新颖变体来检验我们的假设，该变体学习任务先前在各种情况下都对纹状体功能敏感（Doll 等人，2011；Frank 等人，2004，2007），并且还允许直接自由选择的选项与非自由选择的选项之间的比较。要求参与者采样并了解六对具有各种期望值的刺激（参见图 1A），并在每次选择后获得概率反馈（获得或失去一个点）（参见图 1B）。在每个培训试验中，参与者随机获得

六个刺激对之一：这些刺激对中的三个允许参与者在两个选项之间自由选择（fc：自由选择），而其他三个刺激对则迫使参与者选择一个预选刺激（nc：无选择）。至关重要的是，无选择试验要与自由选择试验挂钩，以确保相同的抽样并奖励各种情况下的反馈。

在培训阶段之后，测试阶段将探究所学内容。为参加者提供了所有可能的选项配对，并要求他们在每个试验中选择两者中较好的一个（见图 1C）。在这里，参与者可以自由选择所有试验，但不再获得反馈。重要的是，为了在一系列奖励概率中隔离选择的价值，参与者遇到了一些试验，他们必须在具有相同奖励偶然性的自由选择和无选择选项之间进行选择。

我们使用纹状体 RL 的计算模型对假设的行为含义进行形式化。为了更好地表示 BG 的解剖结构，我们扩展了标准的 actor-critic 体系结构，该体系已被建议通过包括有助于产生积极作用的对手角色权重来规范 BG 的某些核心功能（O'Doherty 等人，2004）。每个选项都有'Go'）和否定（“NoGo”）证据。这些不同的动作权重体现了 D1 和 D2 表达的纹状体中棘神经元的功能含义，它们分别参与直接和间接途径（Frank，2005 年）。在该模型中，根据学习率参数 ag 将 RPE 按比例添加到 Go 权重，同时根据学习率参数 an 对 NoGo 权重同时具有相反的减法效果。因此，这个扩展的参与者包括一个对手过程，在此过程中，Go 和 NoGo 权重分别代表正和负结果期望值，并且选择概率是所考虑的每个动作的 Go 和 NoGo 权重之间的相对差的函数。该对手演员模型捕获了与纹状体多巴胺操纵有关学习和激励动机的大量数据，而标准的单个演员模型无法捕获这些数据（Collins 和 Frank，2014）。在这里，我们进一步研究了在此框架中自由选择放大正预测误差的影响（有关模型的详细信息，请参见在线提供的补充信息）。

2、结果

为了研究我们的假设的行为后果，我们增强了核心 BG 模型，使其包含一个参数α_{fc+}，该参数可调节正向自由选择 RPE 对 Go 和 NoGo 权重的影响。然后，我们在系统地更改α_{fc+}的同时将模型暴露于实验任务。图 2B 说明了α_{fc+}对同等价值的非选择选项的偏好对自由选择选项的影响。当 RPE 在选择条件之间保持平衡（α_{fc+}=1）时，自由选择和不选择选项共享相同的 RPE 历史记录，因此，该模型没有任何选择偏差。但是，随着α_{fc+}的增加，它在确定操作权重方面起着更大的作用，尤其是对于奖励与积极 RPE 经常相关的自由选择选项而言，导致奖励自由选择选项的偏好越

来越大。绩效反映了模型的响应模式（图 2C）。与没有选择的参与者相比，参与者对奖励自由选择的选项表现出强烈的偏好（z=6.84，p<0.001），但对于没有奖励的选项则没有这种偏好（z=0.71，p=0.48）。

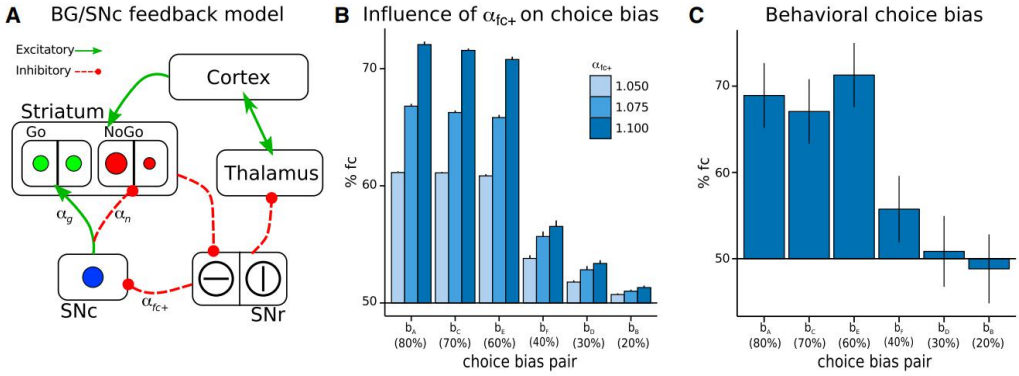

图 2. 积极的 RPE 放大机制和选择偏差模式

（A）BG/SNc 反馈电路的简化图。感觉和运动信息通过皮层皮质投射投射到 BG，通过直接 Go（绿色圆圈）和间接 NoGo（红色圆圈）途径传递给 BG，分别为每个动作提供正面和负面的信息，然后在黑质网状体（SNr）汇聚。此处描述的活动模式说明了两个候选选项的平衡 Go 活动的情况，但 NoGo 活动不同，导致最右边动作的通过。竖线表示对丘脑的门控动作。门控丘脑皮质动作的相同抑制机制也通过 SNr-SNc 投影抑制 SNc 多巴胺能信号，从而使增强信号在 BG 门控动作时被放大。由于这种机制，自由选择的扩增程度被 α_{fc+} 捕获。（B）模型根据奖励偶然性生成了一系列 α_{fc+} 值的选择偏差，计算了选择了自由选择（fc）选项的试验的百分比。（C）选择偏好试验的参与者偏好与奖励意外事件的关系，计算了选择了自由选择（fc）选项的选择偏好试验的百分比。误差线表示 SEM。

在更详细地探讨选择偏好之前，我们首先通过利用行为选择偏好模式来推断关系选项价值结构来建立偏好在各个选项之间是一致的（参见图 3A）。在这里，无选择值取每个选项的真实期望值（例如 $nc_{80\%}=E[A_{nc}]$），而自由选择值则根据每个选项的行为量化选择偏差进行调整（例如 $fc_{80\%}=E[A_{fc}]+b_A$）。然后可以通过比较任何给定选项相对于任何其他选项的偏好来测试图 3A 中所示的结构。

由于自由选择而增加的价值导致了同等报酬的选项之间的差异（例如，$bA=fc_{80\%}-nc_{80\%}$）。这种差异应转化为所有其他选项的一致的自由选择偏好调制（例如，$fc_{80\%}-fc_{30\%}=(nc_{80\%}+b_A)-fc_{30\%}$，$fc_{80\%}-nc_{60\%}=(nc_{80\%}+b_A)-nc_{60\%}$）。我们通过评估涉及最有价值的自由选择和非选择选项，输入根选项（A_{fc}，A_{nc}）和成对选项（C_{fc}，E_{fc}，$.D_{nc}$）

作为逻辑回归的因素的试验的准确性来探索这种预测模式（见图 3B）。该分析表明，所有配对选项的 A_{fc} 总体性能提高都是一致的（根选项的主要影响：$\chi^2(1)=29.23$，$p<0.01$；配对选项的主要影响：$\chi^2(7)=138.02$，$p<0.01$；相互作用：$\chi^2(7)=9.25$，$p>0.2$。通过行为量化的选择偏差来调整 A_{fc} 试验的准确性（图 3B：A_{fc}—b_A）使得性能与 A_{nc} 试验没有区别，这表明 A_{fc} 的性能优势与所有选择对之间的选择偏差均相一致（根的主要影响：$\chi^2(1)=0.15$，$p>0.6$；配对的主要作用：$\chi^2(7)=127.43$，$p<0.01$；相互作用：$\chi^2(7)=9.26$，$p>0.2$）。在涉及最差选择的货币对中也观察到了预期的偏好模式（见图 3C）。

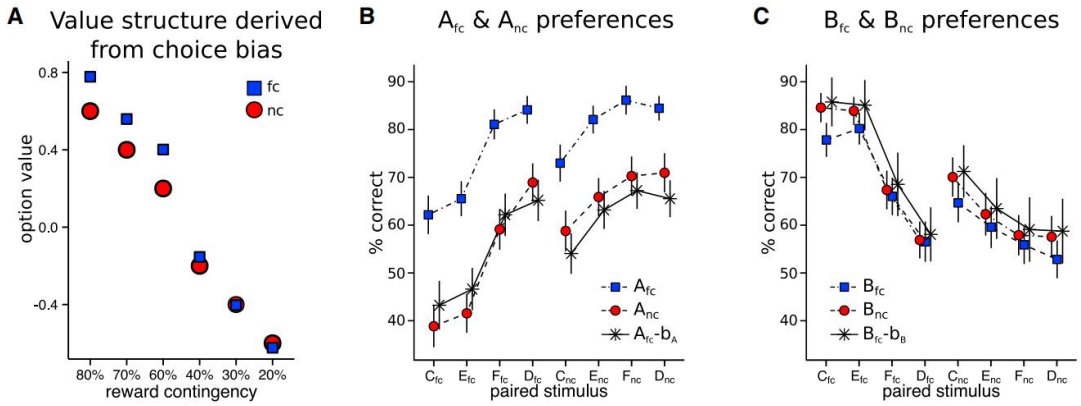

图 3. 派生值结构和隐含偏好模式

（A）从经验上量化的选择偏差得出的选项价值结构。无选择选项（nc）具有真实的期望值。自由选择期权（fc）具有根据每个期权的选择偏差调整的真实期望值。（B）在涉及 A_{fc} 或 A_{nc} 的试验中的正确率（选择更多的奖励选择）。（C）在涉及 B_{fc} 或 B_{nc} 的试验中正确的百分比。所有误差条代表 SEM。

总而言之，参与者行为在一系列独立的选项对中与图 3A 所示的价值结构相一致（更完整的分析请参见图 S2）。这些结果表明，参与者了解了自由选择和不选择选项的相对价值，偏好在激励对之间是内部一致的，并且，正如我们的计算模型所预测的那样，选择偏向效应在奖励选项之间更为明显。

3、奖励概率对选项价值选择放大的影响

效价对选择偏向模式的影响是分类的：正选项的价值被提高，但负选项则没有，但根据奖励概率没有进一步的价值调节。但是，该模型预测，奖励概率会决定动作权重，但对 Go 和 NoGo 权重会产生相反的影响。如图 4A 所示，扩增后的阳性 RPE 对选项权重的影响更大，可获得更多奖励选项（例如 A_{fc}），在这种情况下，更经常

遇到阳性 RPE。这增加了模型选择更多奖励自由选择选项的偏好，这本身就会随着奖励概率的增加（即，$b_A>b_C>b_E$）而带来更大的选择偏好。但是，这会被 NoGo 权重的相反模式所抵消，NoGo 权重的值越大，奖励程度就越适中（例如 E_{fc}）。在这里，扩增的阳性 RPE 会不成比例地减少这些奖励较少的选项的 NoGo 权重。这会降低模型的偏好，避免适度地奖励自由选择权，因为自由选择权本身会导致更大的选择偏差，同时降低奖励的可能性（即 $b_A<b_C<b_E$）。

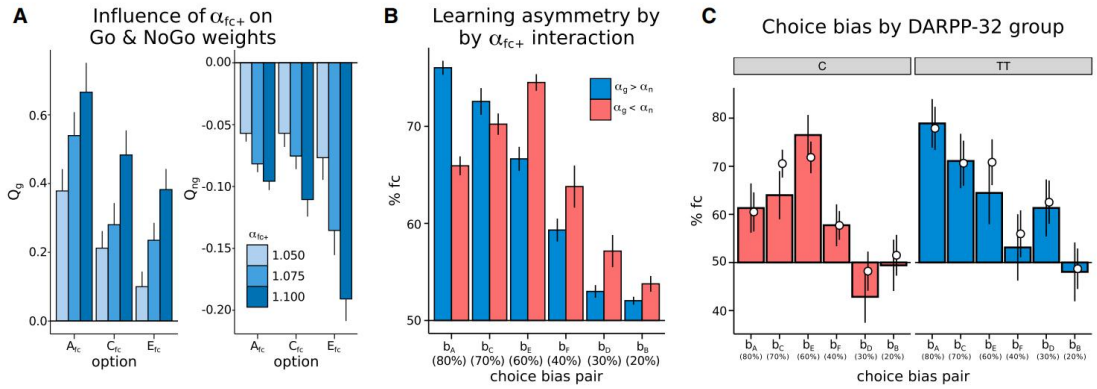

图 4. RPE 正放大对特征权重的影响及其与学习不对称的相互作用

（A）扩增的阳性 RPE 对 Go（Qg）和 NoGo（Qng）权重的影响。优先权最高的选项的权重将被优先放大，从而增加了模型根据放大程度（$A_{fc}>C_{fc}>E_{fc}$）选择这些选项的倾向。收益最小的选项的 NoGo 权重将被优先阻尼，从而根据阻尼程度（$A_{fc}<C_{fc}<E_{fc}$）降低了模型避免使用这些选项的倾向。（B）a_{fc+}与 a_g 不对称性之间的相互作用。（C）根据 DARPP-32 基因组（C 或 TT）的选择偏倚与期望值的关系。条形代表行为数据，点形代表从最佳拟合模型中恢复的期权偏好。误差线表示 SEM。

当 Go/NoGo 学习是对称的时，在 Go/NoGo 权重上产生的相反偏差会在奖励选项之间产生平衡的选择效果（见图 2B）。但是，当 Go/NoGo 学习不对称时，可以通过 a 和学习率参数之间的相对平衡来捕捉每个途径中选择的效果。如图 4B 所示（请参见图 S4），当强调 Go 学习（$a_g>a_n$）时，Go 路径的选择偏好将占主导地位，从而导致对最有意义的选择最强烈的偏好，并根据奖励的概率（$b_A>b_C>b_E$）。当强调 NoGo 学习（$a_g<a_n$）时，就会出现由 NoGo 路径表达的相反选择偏向模式（$b_A<b_C<b_E$）。因此，计算模型预测选择偏好模式应根据学习不对称及其个体差异而变化。

我们试图确定 Go/NoGo 学习不对称的行为后果是否与模型生成的选择偏向模式一致。为此，我们根据 DARPP-32 基因型（与纹状体多巴胺功能相关的基因）分析了行为（Stipanovich 等，2008），以及 Go 与 NoGo 学习中的不对称性（Doll 等，

2011；Frank 等，2007 年，2009 年）。首先，通过将模型参数拟合到逐项试验的行为数据，我们确定 DARPP-32 基因型与可识别的 Go/NoGo 学习不对称性相关。贝叶斯模型选择（Stephanetal。，2009）证明，TT-载体最适合执行 ag>学习速率不对称的模型，而 C 载体最适合执行 ag<学习速率不对称的模型不对称（有关模型的拟合和比较，请参见补充程序和表 S2）。如图 4C 所示，与模型的预测一致，分析显示了通过值交互作用的基因组（$\chi^2(2)=9.88$，p=0.007）。在每个基因组中的孤立分析表明，C 携带者（$a_g<a_n$）表现出 $b_A<b_C<b_E$ 选择偏好模式（z=2.85，p<0.005），而 TT 携带者（$a_g>a_n$）表现出反向 $b_A>b_C>b_E$ 选择偏差模式（z=-1.83，p=0.068）。

4、讨论

与我们的选择会选择性地放大正向 RPE 的假设相一致，在奖励性但非奖励性选择中发现了自由选择偏好。我们还显示出证据表明，扩增的阳性 RPE 具有不同的作用，具体取决于从阳性和阴性结果学习之间的相对平衡。该模型对选择偏向的含义是，RPE 放大会增加奖励选项的 Go 权重并降低 NoGo 权重，同时增加选择收益最高的选项（例如 A_{fc}）的倾向，并降低避免适度奖励选项的倾向（例如 E_{fc}）。从我们的整体样本中可以看出，当学习在 Go/NoGo 途径之间保持平衡时，奖励选择就会出现均衡的选择偏好模式。这支持了将选择偏好与 BG 功能联系起来的先前工作（Leotti 和 Delgado，2011 年，2014 年；Sharot 等人，2009 年），并通过提供由定量行为和建模证据支持的机制解释，扩展了这些发现。这种机制还为观察选择偏向的边界条件提供了自然的解释，从而与更多正预测误差相关的选项表现出更大的自由选择偏向。

我们的结果还表明，在对手途径中学习的相对平衡决定了在 Go 或 NoGo 权重中积累的扩增阳性 RPE 的程度，从而产生不同的选择偏好模式。我们发现，DARPP-32 基因型是与正向和负向 RPE 学习中的结构可塑性和不对称性相关的基因变异（Doll 等，2011；Frank 等，2007，2009；Stipanovich 等）。等（2008 年），预测了个人选择偏向差异。这一结果不仅告知我们其个人权利上的个体差异，而且更广泛地讲，它揭示了根植于 BG 电路中的潜在选择偏差机制，尤其是对我们的模型进行了诊断。重要的是，在 DARPP-32 基因组中观察到的选择偏好模式与选择偏好的注意解释相反，其中出现选择偏好是因为在内源性行动选择过程中参与度更大。确实，有证据表明，在获得奖励时参与度更高，这常常导致奖励与注意力之间的混

淆（Maunsell，2004）。但是，DARPP-32C 载具对更可靠的奖励选择表现出较弱的偏好，这与我们的计算模型是一致的，但与奖励对任务参与的预期影响相反。

据报道，选择偏好的模式相似，相对于被拒绝的选项，选择偏好的偏好更为明显（Sharot 等人，2009），相对于预测损失的偏好，预测收益的选项的偏好更强（Leotti 和 Delgado，2014）。然而，正如 Leotti 和 Delgado（2014）报道的那样，厌恶选择的选择偏好既受情境影响，也受高度可变性影响。的确，我们的样本包括少数参与者（n=80，共 16）显示出对厌恶选择的偏好。但是，这些偏好是没有系统性的，个别参与者对不同的负面选择表现出偏爱和厌恶。此外，我们既不能识别出负面选择偏好的遗传预测因素，也无法识别出其计算或行为的预测因素，这表明，多巴胺能纹状体学习以外的机制在塑造负面选择偏好方面发挥了作用。

我们一直致力于研究选择与学习之间的相互作用。但是，人类通常也会表现出对选择的偏爱（Bown 等，2003；Leotti 和 Delgado，2011，2014），这是我们在此未解决的问题。这种选择偏好可能反映了选择的内在价值，但也可能反映了对于一般选择状态的学习收益。如前所述，自由选择的结果更有可能满足有机体的需求，因此，有机体可以学会偏爱那些可以选择作为奖励的更好预测指标的环境状态。通过时间差异学习，选择也可能会受到青睐，由此通过此处讨论的 BG/SNc 机制放大的增强的选项价值将传播到选项预测状态。尽管这些可能性为将来的研究提供了有趣的途径，但它们似乎都与选择厌恶的报道相矛盾（Iyengar 和 Lepper，2000 年）。我们建议，根据候选选项驱动的选择冲突程度，可以使选择具有竞争性或厌恶性。复杂的选择空间，例如 Iyengar 和 Lepper（2000）所采用的选择空间，可能会潜在地产生足够高的选择冲突，从而可能通过诸如丘脑底核之类的抑制机制来抑制选择（Frank，2006）。

尽管我们的结果表明选择与从积极的 RPE 上获得更好的学习有关，但它提出了一个明显的问题，即为什么应该如此。通常认为 BG 体现了偏向动作选择的门控功能（Ashby 等，2007；Frank，2005；Mink，1996）。这种门控功能体现在背侧纹状体中的多刺神经元的连通性上，它们参与直接 Go 途径或间接 NoGo 途径（Alexander 和 Crutcher，1990）。皮层纹状体投射提出的候选动作的 Go 和 NoGo 活动之间的相对差异决定了哪个动作将进入丘脑，为候选动作提供了选择偏差（Frank，2005）。SNc 的阶段性 DA 信号被认为可通过对 D1 和 D2 受体的下游作用提供发展适当的 Go 和 NoGo 关联所需的学习信号。

但是，动作选择不是由 BG 单独确定的，因此，执行的动作可能与 BG 首选的动作不同。因此，在大脑中均匀地广播 RL 信号提出了一个信用分配问题：所涉及的电路如何确保将补强可靠地传递到对实际执行的动作进行编码的神经系统？解决该问题的方法通常是这样一种观念，即只有最近活跃的神经元才受到 DA 调节的可塑性（Schultz，2002；Wickens 等，1996）。但是，这可以增强决策过程所参与的系统，但是其动作最终没有执行。问题在 BG 本身中进一步复杂化，多巴胺能信号可能会不适当地影响编码被认为但最终未被门控的行为的细胞（见图 2A）。如果 SNc 中的 DA 神经元本身是在 BG 触发某动作时专门门控的，则可以提供 BG 信用分配问题的一种解决方案，该机制可以通过 SNr 的抑制性预测来体现。根据此方案，BG 通过为 SNc 提供动作选通的信息诊断来帮助解决其自身的信用分配问题。该信号引发 SNc 中的 DA 单元，以便每当 BG 参与动作选择过程时，广播到纹状体的阶段性 DA 突发将更加有效。进一步推动这一想法，SNr 可能会向 SNc 提供不仅可以发出动作选通信号的信息，还可以为动作本身提供更丰富的信号诊断信息。然后，SNc 可以整合这些信息，从而以优先针对编码门控作用的纹状体细胞群体的方式构造相位 DA 信号。尽管目前尚不清楚 SNc 的投影体系结构是否能够支持如此丰富的信号，但我们认为这在计算上具有诱人的可能性。

信用分配的问题通常被忽略：以某种方式，大脑的学习信号被传递到回路中的正确地址。最近的工作提出，学习信号在适当的时候会分解成特定于效应子的成分（Gershman 等人，2009），这表明学习信号的确可以构造。我们提出了一种相对简单的机制，通过该机制可以赋予学习信号这种结构，并证明该机制可以解释为什么生物体会偏好其自由选择的选择。简而言之，由于 BG 和 DA 系统之间的反馈回路参与了减轻信用分配的挑战，因此与自由选择的选项相关的学习信号更加有效，这在行为上是一种自由选择的偏好。

4、实验步骤

4.1、参与者

从布朗大学和罗德岛普罗维登斯社区招募了 80 名参与者。六名参与者没有表现出学习任务的能力，因此被从分析中排除（量化为以下涉及 A_{fc} 或 B_{fc} 的试验的机会表现）。但是，当所有参与者都包括在分析中时，此处报告的主要结果仍然有效。布朗大学人类研究委员会批准了所有任务程序。

任务完成后，参与者提供了唾液样本。我们获得了有关 PPP1R1B（DARPP-32）基因的 SNP（rs907094）和 DRD2 基因（rs6277）的 SNP 的基因型数据，这两个基因均与纹状体 DA 功能有关（Hirvonen 等，2009；Stipanovich 等人，2008），以及 COMT 基因的 val158metSNP（rs4680），它与额叶前额叶皮层中的细胞外 DA 水平有关（Huotari 等人，2002；Matsumoto 等人，2003）。DARPP-32 等位基因频率为 7:35:32（C/C：C/T：T/T），DRD2 等位基因频率为 21：45：8（C/C：C/T：T/T）和 COMT 等位基因频率为 13:38:23（MetMet：ValMet：ValVal）。所有 SNP 均处于 Hardy–Weinberg 平衡状态（c2 值<1，p 值>0.4）。通过将最常见的同质等位基因携带者与异质等位基因携带者分组，产生 DARPP-32C：TT（42:32），DRD2CC：T（21:53）和 COMTMet：ValVal 组（51:23）。DARPP-32 与 DRD2T 等位基因频率之间存在正相关（r（72）=0.26，p=0.03），而分类基因组之间存在正相关的趋势（r（72）=0.19，p=0.1）。我们通过将 DRD2 作为协变量纳入所有调查行为遗传预测因子的统计模型中，从而控制了这种相互作用。COMT 与 DARPP-32 或 DRD2 等位基因频率或基因组之间均无相关性（所有 p 值>0.4）。

大多数参与者自认是白种人（49 名参与者）。为了控制人口分层作为潜在的混杂因素，我们在所有统计模型中将种族作为协变量，其中还包括基因组作为因素。但是，当从分析中删除少数群体时，主要论文中报道的结果仍然成立。

4.2、程序

参与者观看了不容易被口头表达的视觉刺激对。在训练阶段，随机出现了六对不同的刺激对，并在选择选项后获得了概率反馈（获得或失去了一个点）。选择选项 Afc 会导致 80％的时间产生正反馈，而选择选项 Bfc 只会导致 20％的时间产生正反馈。$C_{fc}D_{fc}$ 和 $E_{fc}F_{fc}$ 对的可靠性较差（所有奖励意外情况请参见图 1A）。

在自由选择试验中，参与者可以选择提供给他们的任何一种选择。将非选择试验与自由选择试验相结合，以确保条件之间的采样和强化历史相同。记录每个自由选择试验的选择方案和反馈，并将其用于产生无选择的试验。例如，如果在 $C_{fc}D_{fc}$ 试用版中选择了 C_{fc}，并提供了-1 作为反馈，则将生成相应的 $C_{nc}D_{nc}$ 试用版，从而强制选择 C_{nc}（由围绕该选项的蓝色框表示），并提供-1 作为反馈。因此，两种情况下的选项都被采样了相同的次数，并提供了相同的反馈。

参加者完成了至少四个且最多六个训练块。每组包括六个选项对中每一个的 20 个风险承担。在每个模块末尾评估的绩效标准可确保所有参与者在进入测试阶段之

前均达到大致相同的绩效水平（65％的 A_{fc} 选择，60％的 C_{fc} 选择，50％的 E_{fc} 选择）。参加者

在至少完成四个步骤并超过练习标准之后或在六个步骤之后，可以前进到任务的测试阶段。随后以随机顺序对参与者进行了所有可能选项配对（每个选择偏好对的八对配对，以及所有其他配对的四次重复）的完整排列的测试。参与者可以在每个测试试验中自由选择任何一个选项，但不再提供反馈（有关实验设计的详细说明，请参见“补充实验程序”）。

4.3、补充信息

补充信息包括补充实验程序，三个图形和三个表格，可以在本文的 http://dx.doi.org/10.1016/j.neuron.2014.06.035 上找到。

Accepted: June 27, 2014
Published: July 24, 2014

REFERENCES

Alexander, G.E., and Crutcher, M.D. (1990). Functional architecture of basal ganglia circuits: neural substrates of parallel processing. Trends Neurosci. *13*, 266–271.

Ashby, F.G., Ennis, J.M., and Spiering, B.J. (2007). A neurobiological theory of automaticity in perceptual categorization. Psychol. Rev. *114*, 632–656.

Bown, N.J., Read, D., and Summers, B. (2003). The lure of choice. J. Behav. Decis. Making *16*, 297–308.

Brown, P., and Marsden, C.D. (1998). What do the basal ganglia do? Lancet *351*, 1801–1804.

Collins, A.G.E., and Frank, M.J. (2014). Opponent Actor Learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychol. Rev. *121*, 337–366.

Doll, B.B., Hutchison, K.E., and Frank, M.J. (2011). Dopaminergic genes predict individual differences in susceptibility to confirmation bias. J. Neurosci. *31*, 6188–6198.

Egan, L.C., Santos, L.R., and Bloom, P. (2007). The origins of cognitive dissonance: evidence from children and monkeys. Psychol. Sci. *18*, 978–983.

Festinger, L. (1962). A Theory of Cognitive Dissonance. (Stanford: Stanford University Press).

François-Brosseau, F.-E., Martinu, K., Strafella, A.P., Petrides, M., Simard, F., and Monchi, O. (2009). Basal ganglia and frontal involvement in self-generated and externally-triggered finger movements in the dominant and non-dominant hand. Eur. J. Neurosci. *29*, 1277–1286.

Frank, M.J. (2005). Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. J. Cogn. Neurosci. *17*, 51–72.

Frank, M.J. (2006). Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. Neural Netw. *19*, 1120–1136.

Frank, M.J., Seeberger, L.C., and O'reilly, R.C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science *306*, 1940–1943.

Frank, M.J., Moustafa, A.A., Haughey, H.M., Curran, T., and Hutchison, K.E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc. Natl. Acad. Sci. USA *104*, 16311–16316.

Frank, M.J., Doll, B.B., Oas-Terpstra, J., and Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. Nat. Neurosci. *12*, 1062–1068.

Gershman, S.J., Pesaran, B., and Daw, N.D. (2009). Human reinforcement learning subdivides structured action spaces by learning effector-specific values. J. Neurosci. *29*, 13524–13531.

Hirvonen, M.M., Laakso, A., Någren, K., Rinne, J.O., Pohjalainen, T., and Hietala, J. (2009). C957T polymorphism of dopamine D2 receptor gene affects striatal DRD2 in vivo availability by changing the receptor affinity. Synapse *63*, 907–912.

Huotari, M., Gogos, J.A., Karayiorgou, M., Koponen, O., Forsberg, M., Raasmaja, A., Hyttinen, J., and Männistö, P.T. (2002). Brain catecholamine metabolism in catechol-O-methyltransferase (COMT)-deficient mice. Eur. J. Neurosci. *15*, 246–256.

Iyengar, S.S., and Lepper, M.R. (2000). When choice is demotivating: can one desire too much of a good thing? J. Pers. Soc. Psychol. *79*, 995–1006.

Joel, D., and 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.

Lee, C.R., Abercrombie, E.D., and Tepper, J.M. (2004). Pallidal control of substantia nigra dopaminergic neuron firing pattern and its relation to extracellular neostriatal dopamine levels. Neuroscience *129*, 481–489.

Leotti, L.A., and Delgado, M.R. (2011). The inherent reward of choice. Psychol. Sci. *22*, 1310–1318.

Leotti, L.A., and Delgado, M.R. (2014). The value of exercising control over monetary gains and losses. Psychol. Sci. *25*, 596–604.

Lieberman, M.D., Ochsner, K.N., Gilbert, D.T., and Schacter, D.L. (2001). Do amnesics exhibit cognitive dissonance reduction? The role of explicit memory and attention in attitude change. Psychol. Sci. *12*, 135–140.

Lobb, C.J., Troyer, T.W., Wilson, C.J., and Paladini, C.a. (2011). Disinhibition bursting of dopaminergic neurons. Front. Syst. Neurosci. *5*, 1–8.

Matsumoto, M., Weickert, C.S., Akil, M., Lipska, B.K., Hyde, T.M., Herman, M.M., Kleinman, J.E., and Weinberger, D.R. (2003). Catechol O-methyltransferase mRNA expression in human and rat brain: evidence for a role in cortical neuronal function. Neuroscience *116*, 127–137.

Maunsell, J.H.R. (2004). Neuronal representations of cognitive state: reward or attention? Trends Cogn. Sci. *8*, 261–265.

Mink, J.W. (1996). The basal ganglia: focused selection and inhibition of competing motor programs. Prog. Neurobiol. *50*, 381–425.

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

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

Orr, H.A. (2009). Fitness and its role in evolutionary genetics. Nat. Rev. Genet. *10*, 531–539.

Schultz, W. (2002). Getting formal with dopamine and reward. Neuron *36*, 241–263.

Sharot, T., De Martino, B., and Dolan, R.J. (2009). How choice reveals and shapes expected hedonic outcome. J. Neurosci. *29*, 3760–3765.

Sharot, T., Velasquez, C.M., and Dolan, R.J. (2010). Do decisions shape preference? Evidence from blind choice. Psychol. Sci. *21*, 1231–1235.

St Onge, J.R., Ahn, S., Phillips, A.G., and Floresco, S.B. (2012). Dynamic fluctuations in dopamine efflux in the prefrontal cortex and nucleus accumbens during risk-based decision making. J. Neurosci. *32*, 16880–16891.

Stephan, K.E., Penny, W.D., Daunizeau, J., Moran, R.J., and Friston, K.J. (2009). Bayesian model selection for group studies. Neuroimage *46*, 1004–1017.

Stipanovich, A., Valjent, E., Matamales, M., Nishi, A., Ahn, J.-H.H., Maroteaux, M., Bertran-Gonzalez, J., Brami-Cherrier, K., Enslen, H., Corbillé, A.-G.G., et al. (2008). A phosphatase cascade by which rewarding stimuli control nucleosomal response. Nature *453*, 879–884.

Tversky, A. (1972). Elimination by aspects: A theory of choice. Psychol. Rev. *79*, 281.

Wickens, J.R., Begg, A.J., and 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.