



## 方程式入侵！数学如何解释大脑如何学习

Kate Nussenbaum 和 Alexandra O. Cohen\*

纽约大学心理学系（美国，纽约）

### 少年审稿人

BROWN  
UNIVERSITY  
PRE-COLLEGE  
STEM I  
PROGRAM  
年龄：12-13



想象一下，每当看见美丽的夕阳，你就会拍下来，然后发在社交媒体上。你的前几张照片有很多人“赞”，但是有一天你发的照片只得到了很少的“赞”，那么你还会继续发日落照片吗？我们会做出选择，然后看看发生了什么，这会影响我们未来的选择。神经科学研究人员已经开发出数学模型，来解释人们如何从他们选择导致的正面和负面结果中学习。这些模型通常被称为强化学习模式。在本文中，我们将解释在现实世界中**使用强化学习的情景**，数学方程式如何帮助我们了解这一过程，以及大脑如何通过从经验中学习做出好的选择。

### 什么是强化学习？

强化学习是一个过程，在这个过程中，我们利用过去的经验帮助我们做出好的选择，以便得到好的结果。让我们详细看看文章摘要中描述的例子。当你第一次发布日落图片时，你会得到大量的“赞”，这让你很开心——接受所有这些“赞”的积极体验让你更有可能发布另一张照片。你发了第二张日落照片并得到了同样多的“赞”；你发布了第三个并且得到了更多“赞”。但是有一天，你发布了你认为很漂亮的图片但几乎没有人“赞”。相反，有人对你的照片写了一个刻薄的评论。突然间，并非所有与发布日落图片相关的都是正面体验了。于是你需要更

新体验，接收刻薄评论的负面体验。下次你发布日落的可能性会更低一些。

我们选择的结果可以被视为强化信号。如果我们做出选择并体验很好——积极的强化信号——我们将来更有可能重复这一选择。但是，如果我们做出决定并体验不好，我们可能会在下次选择不同的选项。这个过程叫做强化学习。

### 为什么使用数学方程来理解学习和决策

一般来说，积极的体验（比如在社交媒体上受到大量的“赞”）会引起我们对增加对奖励的期待，而负面体验（如受到刻薄评论）会导致我们减少对奖励的期待。但是，什么程度的反馈会导致我们期待的变更呢？例如，假设你发布了九张收到许多“赞”的图片和一张收到讨厌评论的图片。一个讨厌的评论会在多大程度上影响你对发布图片的奖励的预期？你将来发布类似图片的可能性有多大？如果没有数学方程来描述学习过程，我们就无法回答这些问题。

此外，研究人员可以编写具有不同组成部分的方程式，每个部分代表思考或决策中涉及的不同过程。然后，通过变更方程式中的每一部分，我们就可以看到其他部分如何变化，来帮助我们理解不同的思考过程如何影响学习。

### RESCORLA - WAGNER 模型

多年来，研究人员提出了不同的数学方程——或“**计算模型**”，来解释人们如何从积极和消极的经历中学习。最先被提出的模型之一，称为瑞思考勒-瓦格纳（Rescorla-Wagner）模型 [1,2]。

研究人员罗伯特·瑞思考勒和艾伦·瓦格纳希望更好地理解另一位研究人员巴甫洛夫所做的著名实验。在这些实验中，巴甫洛夫每次摇铃（它实际上是一种发声装置——一个节拍器——但为了简单起见，我们称之为铃声），就会给狗食物。起初，这些狗没有将食物的奖励与铃声联系起来。但巴甫洛夫发现，经过几次重复这一过程后，即使没有给他们食物，狗也会在听到铃声时开始流口水。这些研究表明，随着时间的推移，狗会学会将铃声与食物联系起来，每当他们听到铃声时，他们都会预期会得到食物（并且流口水！）。

然而，如前所述，学习往往是一个渐进的过程。这意味着每次巴甫洛夫的狗听到铃铛并获得食物时，它们就将铃铛与食物的联系增强一点，并且他们更有可能在钟声响起时流口水。

#### 计算模型

(Computational model)

计算模型用简单数学表示一个更复杂的过程。本文中描述的强化学习模型是一个可以表示某些类型学习中涉及的思维过程的等式。

我们可以预测狗在听到铃声时对食物的期待有多强吗？我们能否知道在新的经历后狗的预期会如何变化？瑞思考勒和瓦格纳想要提出一个数学方程来回答这些问题。但他们不能随意写出等式——他们需要编写一个能准确反映学习过程的方程式。要做到这一点，他们首先需要了解动物如何学会通过经历建立联想。

瑞思考勒和瓦格纳并没有尝试使用狗，铃铛或食物，而是使用声音和电击对老鼠进行实验。在这些实验中，老鼠会听到声音然后受到电击。通常，老鼠在笼子里跑来跑去。但是老鼠不喜欢被电击，所以当他们认为即将发生的电击时，他们经常会僵住。老鼠将声音与电击联系起来的强度，可以通过老鼠在听到声音后僵住的时间长短来测量。例如，当老鼠第一次听到声音时，不会期待电击，所以它会继续正常地移动。但是，如果它听到声音然后受到电击，那么老鼠就会开始意识到声音和电击相关。下一次听到声音，老鼠会移动得更少，僵住时间更长。

瑞思考勒和瓦格纳注意到，在老鼠刚刚开始建立声音与电击之间的联系时，它们僵住的时间和频率变化较大。而在多次电击之后，他们有了更稳定的行为模式——还是会僵住，但僵住的时间在每次电击后仅增加了一点点。

这一实验以及许多其他实验使得瑞思考勒和瓦格纳发现强化学习由“意外”驱动。换句话说，动物在遇到他们没想到的东西时会学到更多。这类学习过程中的意外被称为“预测误差”，因为它们代表了动物期望经历的事物与它实际经历的事物之间的差异。例如，巴甫洛夫的狗第一次听到铃声，他们没有理由期待得到任何食物。当他们收到食物时，他们会出现意外，或者说较大的预测误差，因为实际发生的事件（食物！）与他们预测会发生的事情（没有食物）大不相同。然后它们便意识到在听到铃声时有可能得到食物。当下次听到铃声时确实得到了食物，它们的预测误差较低，因为它们对结果不太惊讶。当他们继续听到铃声并得到食物时，他们对听到铃声后得到食物的期待继续增加，但程度没有前几次高，因为当时他们对得到食物更加惊讶。

我们总是在经历“预测误差”。例如，你可能认为自己不喜欢吃西兰花。但是有一天，你可能会决定试一试，发现它确实味道不错！在这种情况下，你吃美味西兰花的经历会与你预期西兰花难吃不同。你会遇到预测误差，这会让你学习一些关于西兰花的东西，并改变你对它的味道的看法。

如果没有预测误差，我们便不会通过强化来学习任何东西。比如说喜欢吃披萨。有一天，你可能会在从学校回家的路上停下来吃披萨。它很美味！在这种情况下，你吃美味披萨的经历与你的预期没有差别。没有遇到预测误差，因此你不会学到任何东西。你会继续认为披萨味道很棒。

### 预测误差

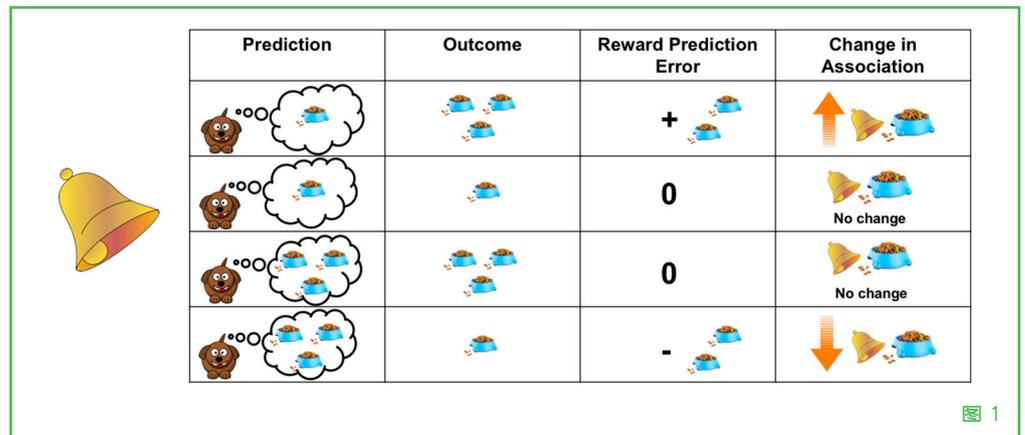
#### (Prediction error)

预测误差代表意外。它们代表了动物期望体验的内容与实际体验的内容之间的差异。

瑞思考勒和瓦格纳写了一个数学方程来描述这个学习过程。他们的方程式表明，动物认为某些事物与奖励之间有关联（比如铃声和食物），这个联系增强多少，是由它对奖励的预期与事实之间的差距决定的。（图 1）

图 1

学习是由预测误差驱动的。该表显示了狗的预测和经验结果如何影响其学习。狗会在多大程度上增强铃声和食物之间关联（显示在第四列中），取决于它听到铃声时实际收到的食物（显示在第二列中）和它预测会收到多少食物（显示在第一列中）之间的差距。这种差异显示在第三列中。



Prediction	Outcome	Reward Prediction Error	Change in Association
		+	
		0	No change
		0	No change
		-	

图 1

这个等式可以告诉我们动物关联两件事情的强烈程度，或某项决定（如在社交媒体上发布图片）可能会带来多少回报。

方程中除了预测误差之外，还有一个称为“学习率”的术语。学习率能告诉我们动物在每次经历后会在多大程度上更新预期，并与预测误差相乘。我们可以将学习率视为每只动物学习的速度。如果动物具有较高的学习率，那么当它遇到预测误差时会更新它的预期。但是，如果动物的学习率较低，那么它可能会更大程度地依赖于过去的所有经历，并且每次经历预测错误时只会稍微改变预期。

### 学习率

(Learning rate)

动物的学习率表示根据新信息更新其认知的速度。

### 多巴胺

(Dopamine)

多巴胺是大脑中一种特殊类型的化学物质，称为神经递质，通常参与信号奖励。

## 大脑如何从强化中学习？

强化学习模型有助于我们了解大脑的学习方式。大脑由大约 1000 亿个称为神经元的脑细胞组成。神经元释放出称为神经递质的化学物质，它可以帮助神经元相互发送信息。多巴胺是大脑中一种重要的神经递质。多巴胺神经元会对我们受到奖励的体验产生反应。

通过使用前面描述的实验，科学家们已经证明，多巴胺神经元的活动在显示大脑预测误差方面起着关键作用。研究发现，多巴胺神经元会对预示着奖励的事物（比如响铃）作出反应，活性会增加，而这是在真正收到奖励之前。如果动物认为会得到奖励，但却没有，多巴胺神经元活性则会降低。在强化学习模型方面，我们可以将多巴胺视为多巴胺神经元的预测误差信号——多巴胺神经元的活动能够表明预期和真实之间的差距 [3]。这有助于我们从强化中学习，并最终帮助我们利用利用经历做出能够获得奖励的选择（图 2）。

图 2

该图显示了在学习之前和之后狗的大脑中发生了什么。多巴胺神经元会对奖励和对奖励的预期做出反应。图中的点表示多巴胺神经元伴随时间的活动。对于一个特定时间点，点上方的线的高度表示其正下方的点的数量。多巴胺神经元会对奖励响应，如食物（左上）。在狗学会将铃声与食物联系起来之后，多巴胺神经元可以对预示奖励的东西响应，如铃声（左下）。请注意，在这种情况下，多巴胺神经元不会对食物本身产生反应，因为它不再令人惊讶。然而，如果预期的奖励没有发生，多巴胺神经元就变得不那么活跃（右下）。源自 Schultz 等人 [3]。

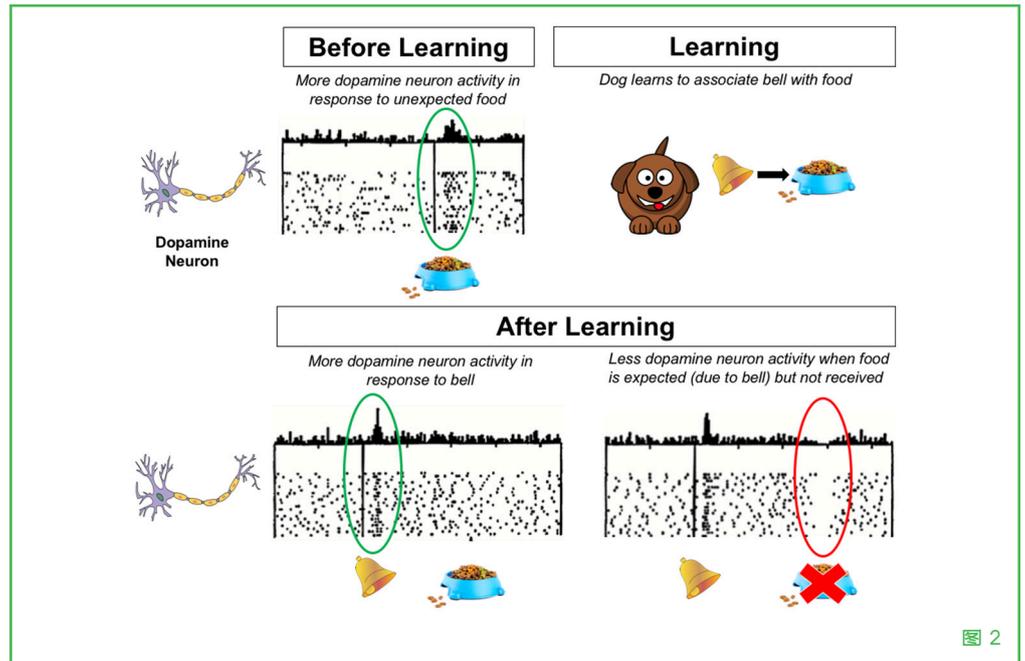


图 2

大脑的许多不同部分显示出与多巴胺神经元预测误差信号相似的活动模式。其中一个位于大脑深处的一组区域，称为基底神经节。基底神经节不仅能帮助我们学习，而且对控制我们的动作和习惯很重要。基底神经节中最大部分是纹状体。纹状体是多巴胺释放的主要部位，是控制我们对奖励的反应的大脑系统的核心部分。

许多关于动物和人类研究表明，纹状体活动与预测误差有关，在强化学习中起着重要作用 [4]。与预测误差相关的大脑活动模式也可以在大脑的额叶皮层中看到，这是一个参与决策的区域。纹状体和额叶皮层之间有许多联系，这些联系对帮助我们快速计算强化学习方程至关重要，也可以解释我们如何快速从经验中学习，并利用这些知识帮助我们做出未来的决策（图 3）。

图 3

强化学习中涉及的大脑区域。位于大脑前部（前额后方）的额叶皮层（左侧）在决策中起重要作用，纹状体（右侧）显示与预测误差相关的活动。版权所有神经科学学会（2017 年）。要使用 3D 地图进一步探索大脑，请访问：<http://www.brainfacts.org/3d-brain>

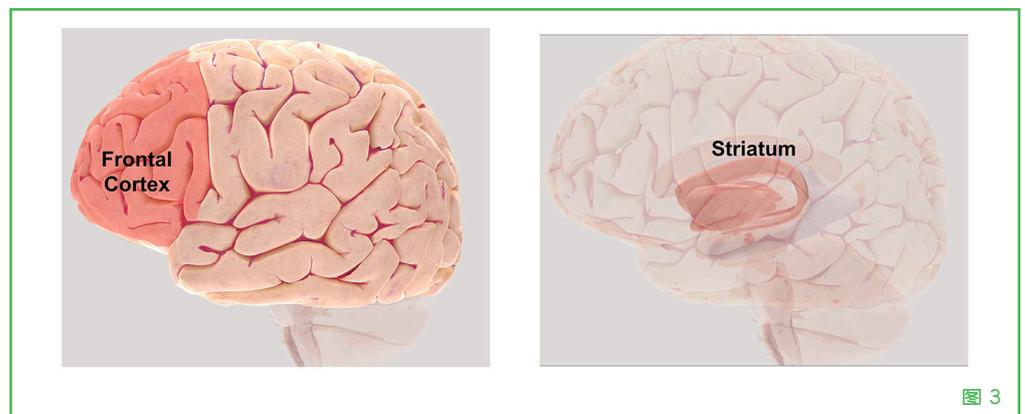


图 3

尽管如此，在纹状体和额叶皮层中显示的活动变化只是这个难题的一小部分！重要的是要记住，大脑由许多不同的部分组成，这些部分协

同工作，来帮助我们思考和做复杂的事情，比如学习。科学家们还在继续研究纹状体，额叶皮层和其他大脑区域如何协同工作，以帮助我们强化中学习，并最终利用我们学到的信息做出最佳决策。虽然我们本文中介绍了经典学习模型——瑞思考勒-瓦格纳（Rescorla-Wagner）模型，但还有更多的数学模型，帮助我们了解学习以及学习时大脑中发生了什么。

所以，下次当你考虑是否将图片发布到社交媒体时，请记住你的大脑正在快速解决数学问题——而你甚至没有意识到！

### 参考文献

1. Rescorla, R. 1998. Pavlovian conditioning. *Am Psychol.* 43:151–60.
2. Rescorla, R., and Wagner, A. R. 1972. “A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement,” in *Classical Conditioning II: Current Research and Theory*, eds A. H. Black and W. F. Prokasy (New York, NY: Appleton-Century-Crofts). p. 64–99.
3. Schultz, W., Dayan, P., and Montague, P. R. 1997. A neural substrate of prediction and reward. *Science* 275:1593–9.
4. Dayan, P., and Niv, Y. 2008. Reinforcement learning: the good, the bad and the ugly. *Curr Opin Neurobiol.* 18:185–96. doi: 10.1016/j.conb.2008.08.003

线上发布: 2023 年 4 月 27 日

编辑: David L. Sheinberg

科学导师: Alexander More 和 Ayumi Tsuda

引用: Nussenbaum K 和 Cohen AO (2023) 方程式入侵！数学如何解释大脑如何学习. *Front. Young Minds.* doi: 10.3389/frym.2018.00065-zh

英文原文: Nussenbaum K and Cohen AO (2018) Equation Invasion! How Math Can Explain How the Brain Learns. *Front. Young Minds* 6:65. doi: 10.3389/frym.2018.00065

**利益冲突声明:** 作者声明, 该研究是在没有任何可能被解释为潜在利益冲突的商业或财务关系的情况下进行的。

**版权:** © 2018 © 2023 Nussenbaum 和 Cohen. 这是一篇依据 [Creative Commons Attribution License \(CC BY\)](#) 条款发布的开放获取文章。根据公认的学术惯例, 在注明原作者和版权所有者, 及在标明本刊为原始出处的前提下, 允许使用、传播、复制至其他平台。如违反以上条款, 则不得使用、传播或复制文章内容。

## 少年审稿人



BROWN UNIVERSITY PRE - COLLEGE STEM I PROGRAM. 年龄: 12 - 13

这些中学生审稿人在布朗大学参加了 STEM I Pre-College 的课程。更准确地说, 这些学生在两周的课程中专注于神经科学和动物行为生物课程。为了进一步丰富他们的学习, 他们热情地加入了“少年审稿人”团队。他们在论文中提供了很好的反馈, 非常投入并对审稿过程感兴趣。审稿于布朗大学校园 2018 年 7 月。

## 作者



KATE NUSSENBAUM

我是纽约大学的研究生, 对人类如何学习和记住来自世界各地的信息, 以便在将来做出正确的决定感兴趣。我特别喜欢思考学习过程如何随着我们度过童年和青春期而变化。



ALEXANDRA O. COHEN

我是纽约大学的神经科学家, 对大脑如何随着人们成长而变化很感兴趣。特别是积极和消极的经历如何影响我们在童年到成年过程中的学习和记忆方式。当我不在实验室时, 我通常会教学, 或给孩子, 青少年或任何有兴趣的人讲解科学!  
[\\*ali.cohen@nyu.edu](mailto:ali.cohen@nyu.edu)