《智能简史》6：系统 2 思维

万维钢 • 精英日课 6（年度日更）

不知你是否读过丹尼尔·卡尼曼的名著，《思考，快与慢》？就算没读过，你想必也听说过书中的一个关键思想：我们的思考可以分为两类，一个是「系统 1」，是直觉的、快速的思考；一个是「系统 2」，是理性的、慢速的思考。系统 1 容易让我们犯错误，但系统 2 比较累。卡尼曼书中有个特别有意思的小细节，我看知道的人可能不多，咱们专门说说。

当一个人在进行系统 2 思考的时候，他的瞳孔会放大。

比如你让一个学生做数学题，题目很难。如果他只是凭着本能做，不用力思考，瞳孔并不会扩张。只要他真的努力琢磨这个题，他的瞳孔就会立即扩大大约 50% —— 同时心跳每分钟增加 7 次，但最明显的是瞳孔。如果他思考一段时间感觉这题实在做不出来，决定放弃了，瞳孔又会缩小到原样。

这个现象是如此之灵敏，以至于研究者可以精确判断这个学生在什么时候放弃思考。研究者会问他：你放弃了吗？学生很惊讶，说你怎么知道？研究者说：因为我有一个通往你心灵的窗口。

我第一次听卡尼曼讲这个现象，只是觉得很神奇。卡尼曼没有太多解释为何如此。事实上就连大脑在解剖学上哪些部分是系统 1、哪些是系统 2，卡尼曼也没有细说。这可能是因为卡尼曼是个心理学家而非脑神经科学家，也可能是因为那时候我们还没有很明确的理论。

现在有了。

时隔多年之后，麦克斯·班尼特的《智能简史》这本书，也讲到了瞳孔扩大的现象。

我们前面讲了，新皮层就如同 AI 神经网络，有时候处于接收信息的状态，有时候处于生成 —— 也就是想象 —— 状态，而这两种状态不能同时进行。这里有一个关键特点：当一个人处在想象状态，正在头脑里模拟一个世界的时候，他的瞳孔是扩大的。

因为那时的大脑专注于内部想象，不再处理视觉数据输入，他变成了一个假盲人。

把两本书联系在一起是读书人的一大乐趣。既然都涉及到瞳孔放大，那我们是不是可以说，新皮层的想象状态，就是卡尼曼所说的系统 2 思考呢？

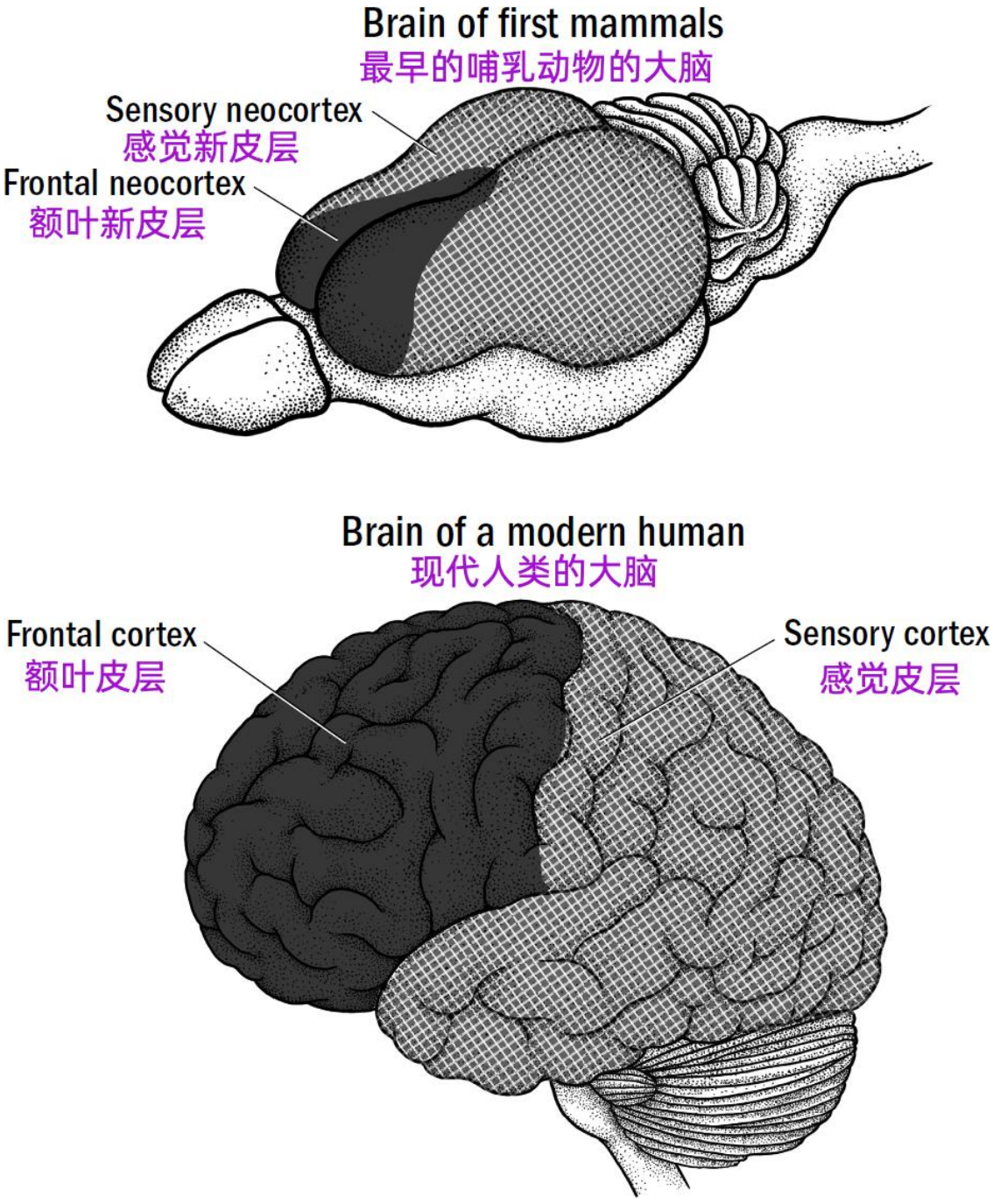
没错，正是如此！这一讲我们会把「哺乳动物的想象力」、「卡尼曼的系统 2 思维」和 AI 的「基于模型的强化学习」这三个东西统一起来。

你了解和思考了多年的几件事儿，原来是一回事儿，这难道不是很神奇吗？

✵

前面我们讲了，哺乳动物的一个新能力是面临两难选择时，能犹豫一下，把不同的局面模拟一番，再做出选择。现在的问题是，大脑是怎么决定要暂停自动化，要犹豫一下的呢？是谁下的命令？

我们先看大脑的解剖图。

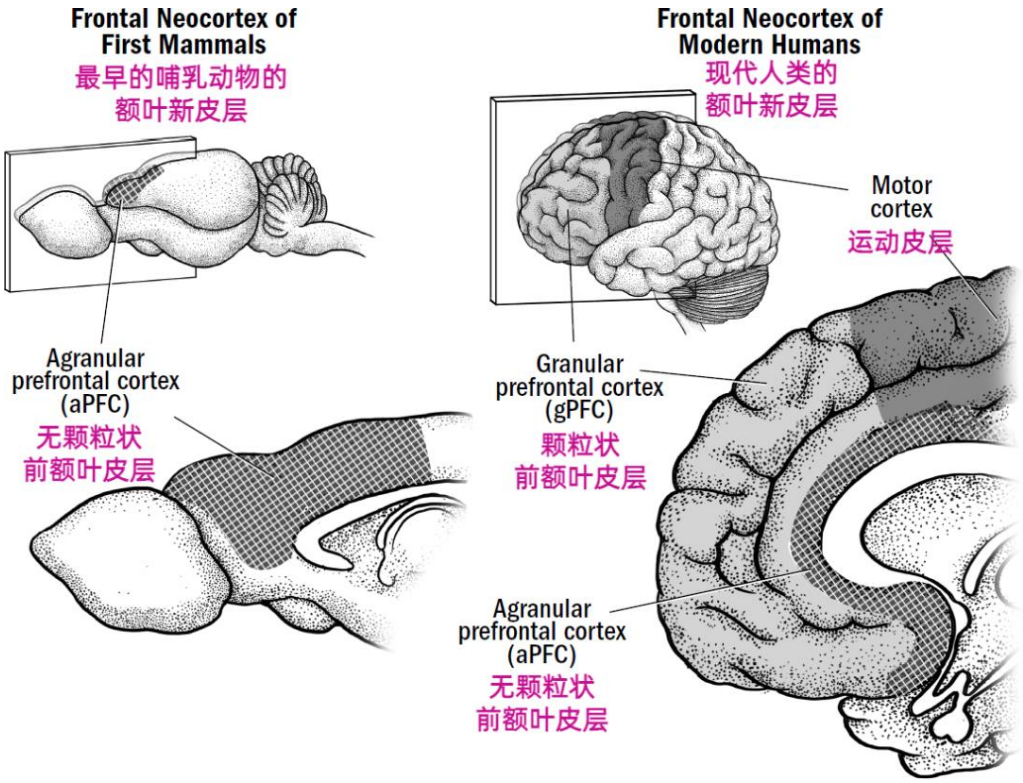


所有哺乳动物的大脑新皮层都可以分成前后两部分。后半部分叫「**感觉新皮层**（sensory neocortex）」，用来对接外部的触觉、视觉、听觉、嗅觉等等，既处理这些信号也模拟这些信号，负责想象。前半部分叫做「**额叶**

新皮层（frontal neocortex）」，或者对人来说可以简单地就叫「**额叶皮层**（frontal cortex）」，就是它，负责决定*要不要*停下来进行想象。

更准确地说，是额叶皮层中的「**无颗粒状前额叶皮层**（aPFC）」这个区域负责决定要不要进行想象。

额叶皮层可以分成三部分：**运动皮层**（motor cortex），**颗粒状前额叶皮层**（gPFC）和**无颗粒状前额叶皮层**（aPFC）。



咱们单说这个无颗粒状前额叶皮层（aPFC），这是我们人脑和最早的哺乳动物共有的区域，我们专栏以前讲的所有的做决定、实施注意力的脑区，都是说的这个区域。我们以下简称它为"前额叶皮层"。

前额叶皮层为啥能决定要不要开启系统 2 思维，也就是暂停直觉行动，做一番模拟计算呢？其实它的工作原理跟感觉皮层是一样的！

正如感觉皮层接收感觉信号，前额叶皮层接收的信号则是来自大脑内部的海马体、下丘脑和杏仁核。特别是，它一直在关注基底神经节。

我们前面讲强化学习的时候说了，强化学习的结果体现在基底神经节上。你可以认为基底神经节负责直觉运算，负责做出近乎本能的快速反应 —— 简单说，基底神经节负责系统 1 思考。

前额叶皮层一直在观察基底神经节，它像视觉皮层对视觉信号建立模型那样，对基底神经节建立了一个模型！然后它要根据这个模型做出预测。

它预测的，是动物自己的行动意图。

比如说，一只老鼠的前额叶皮层看到基底神经节指挥身体前往有水的方向，它就会想，“我之所以往这边走，是为了去喝水。”它会预测下一步的行动是喝水。

这是一种建模。正是因为前额叶皮层的建模，我们才有了「意图」这个东西。换句话说，意图是大脑想象出来的东西。

老鼠本能地前往有水的地方，就如同扫地机器人本能地前往充电插座，这个行为原本只是强化训练的结果，根本谈不上什么意图和目的 —— 是哺乳动物的前额叶皮层没事儿找事儿，非得对这种原始冲动做出建模，提供解释，才发明了意图。

有了意图，才可能有自我意识。这就是为什么现阶段的 AI 没有意识，因为它们只是自动反应，它们还没有前额叶皮层。

✵

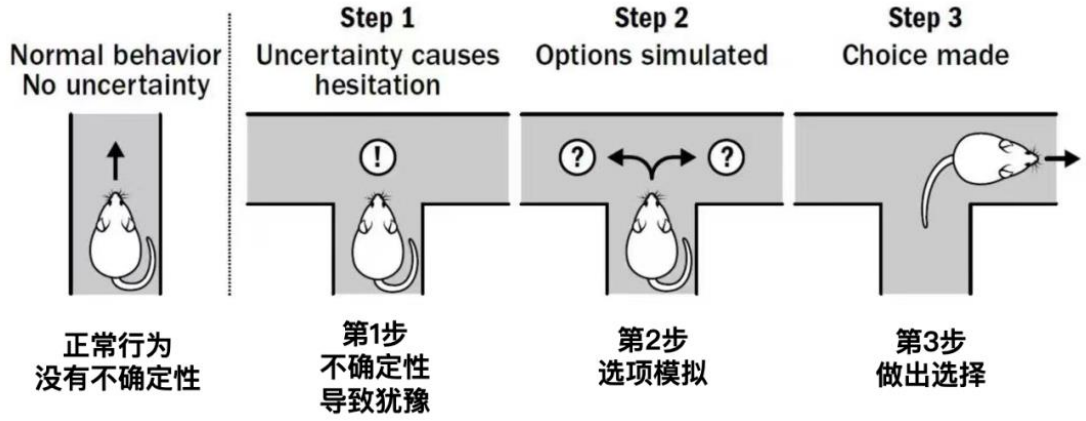
上世纪八十年代，脑神经科学家安东尼奥·达马西奥（Antonio Damasio）接治了一位女中风病人，代号 L。L 中风的脑区正好是前额叶皮层，这使得她完全失去了意图感。她的身体各方面都没问题，能正常运动也能理解别人说的话，但是她懒得跟人说话，什么都不想做，失去了所有的主动性。六个月后，L 在新皮层的其他区域重新映射了一个形成意图的区域，主动性才恢复了。

这个案例生动地证明了前额叶皮层对意图的重要性。这大概也是达马西奥后来形成自己的意识理论的关键启发，我们专栏讲过他的理论 [1]。

前额叶皮层和感觉皮层都属于新皮层，它们的功能没有本质区别，都是模拟感官、建立模型和做出预测。只不过前额叶皮层模拟的是大脑自身的直觉。

我们大多数行为都只是出于本能，害怕时逃跑，渴了想喝水，都是系统 1 思维，基底神经节就能协调得很好，前额叶皮层只是默默旁观而已。如果老鼠的前额叶皮层预测的意图正在被实现，它不会干涉。

只有当前额叶皮层的预测产生矛盾的时候，它才会兴奋起来，出手干预。比如老鼠走在一个岔路口，前额叶皮层中的一部分预测它当前的意图是想喝水，这意味着应该往左转；另一部分却预测它此刻想吃东西，应该往右转，这就是矛盾时刻。



前额叶皮层会以某种机制向基底神经节发出信号，要求暂停行动。它安排感觉皮层对两条路分别模拟一番，看看会发生什么 —— 正如我们上一讲说的 AlphaZero 下围棋时基于模型的强化学习。模拟结果出来之后，前额叶皮层会把结果展示给基底神经节，促使它采纳某一个选项，比如说向右转。

丹尼尔·卡尼曼说的系统 1，也就是快思考，其实就是强化学习带来的本能反应，由基底神经节自动选择；卡尼曼所说的系统 2，慢思考，其实就是前额叶皮层感觉到了冲突，先暂停自动反应，发起模拟再做选择，也就是基于模型的强化学习。

爬行动物全都是系统 1 思维。我们日常大部分时候也都是系统 1 思维。这很好，这使得我们做开车、走路、吃饭喝水这些日常动作都不需要思考，我们很轻松。只在矛盾时刻，我们才需要调用昂贵的新皮层算力去进行模拟。

*

前额叶皮层和基底神经节之间的配合可以解释很多现象。

做陌生的事情，我们总要小心翼翼地想想怎么做，就必须调用系统 2；一旦熟练了，新皮层就可以放手，全交给基底神经节。

在一个实验中，先训练老鼠一摇杠杆就会得到食物。后来实验人员在食物里加入了让老鼠感到恶心的药物。那你说老鼠还会不会去摇那个杠杆呢？

答案是之前训练次数较少，还没有形成习惯的老鼠会减少摇杠杆的次数，因为这个动作对它们已经没意义了；可是那些训练超过 500 次，形成了自动习惯的老鼠，哪怕明知得到的食物自己不喜欢，也仍会去摇杠杆 —— 他们的前额叶皮层没有机会介入，基底神经节完全接管了对杠杆的行动。

人不也是如此吗？现在很多人动不动就把手机拿出来看，哪怕多数情况下看手机并没有什么效用。那不是我们深思熟虑的选择，那只是基底神经节的自动动作。

再者，我们所有的意图、目标、人生的意义，都是前额叶皮层想象出来的。而正是这些想象出来的东西能强硬地指导我们的行动。

如果你没有目标，你不会保持注意力。如果你不是主动记得要做什么事，你不会保持工作记忆。如果你认为人生毫无意义，你不会自我控制。这些都是前额叶皮层对基底神经节不断说服的结果。

有意思的是，并不是说前额叶皮层有比基底神经节更高的命令权 —— 其实它所做的只是把想象出来的各种可能性展示给基底神经节看，让基底神经节相信为什么这个选项是对的。

前额叶皮层只是本分地行使新皮层的职能而已，只不过它负责想象的是大事。

*

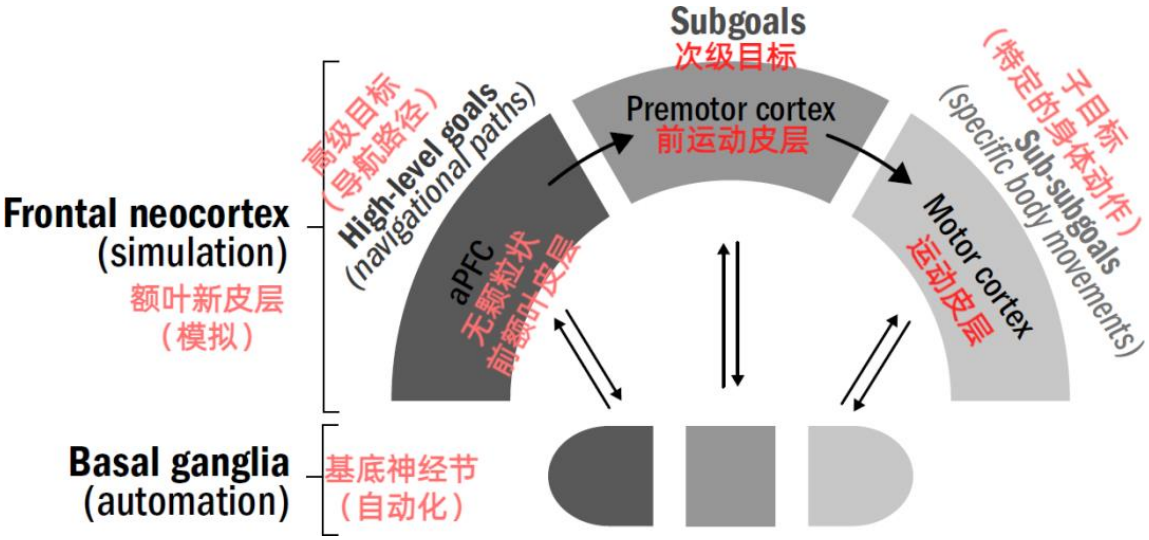
额叶皮层的另一个区域，运动皮层，则专门负责想象小事。特别、特别小的事，比如此时前爪应该放在哪里。

通常的中风都爱发生在运动皮层。或者因为血栓让血流流不进去，或者是因为脑出血导致血流中断，反正中风就是某个脑区缺血，失去功能。一旦运动皮层中风，人的四肢就不受控制，严重情况下可能会瘫痪。所以运动皮层的作用是控制身体的运动，对吗？

原本不是如此。现实是运动皮层中风导致瘫痪这个现象只发生在灵长类动物身上。猫狗这些小动物，即便运动皮层受损也不会瘫痪，它们还能走路、觅食。这是为什么呢？

因为运动皮层的作用其实跟其他新皮层一样，也是模拟和预测：它模拟和预测的是身体的运动。

前额叶皮层规划宏观的路线，运动皮层规划四肢的具体动作，它们都跟基底神经节相连……下面图中，越往左边，越是负责更宏观的目标，越往右越是负责具体的行动。



这张图给机器人研究提供了一个启发。我们需要把要做的事情分解成大目标、小目标和具体的行动，各自有负责想象和规划的模块，有负责执行的模块。对哺乳动物来说，正是新皮层和基底神经节的优雅分工，使得大脑能够在不同层次上完成任务。

✳

OpenAI 迟迟没有发布 GPT-5，但我们听到一些传闻，说 GPT-5 将会拥有「系统 2 思维」。经过这一讲，我们大概能猜测那意味着什么 ——

- 你需要对一个问题建立多个智能体（agents），让每个智能体各自生成答案；

- 你需要一个前额叶皮层之类的机制，对各个答案进行评估，选择最合适的一个，再输出；

- 这两步加起来就是系统 2 思维；

- 为了节省算力，你需要随时判断什么时候一次直觉输出就够了，什么时候需要调用系统 2。

而现今的大语言模型基本上只是系统 1 思维，纯直觉输出。但我们可以想见，跨越到系统 2 在技术上一点都不难，难的只是算力而已 —— 毕竟一切都是新皮层。

我们用了三讲才说清楚新皮层的作用，但是人类的智能比这个还要厉害。相对于老鼠，我们的大脑还有两次突破。

注释

[1] 意识红色胶囊 4：情感神经网络

1.

系统 1、快思考，是强化学习带来的本能反应，由基底神经节自动选择；系统 2、慢思考，是前额叶皮层感觉到了冲突，先暂停自动反应，发起模拟再做选择，也就是基于模型的强化学习。

2 做陌生的事情，我们总要小心翼翼地想想怎么做，就必须调用系统 2；一旦熟练了，新皮层就可以放手，全交给基底神经节。

3.我们所有的意图、目标、人生的意义，都是前额叶皮层想象出来的。而正是这些想象出来的东西能强硬地指导我们的行动。