《冲浪不确定性：预测、行动与具身心智》

SURFING UNCERTAINTY

驾驭不确定性

预测、行动与具身心智

安迪 · 克拉克

image

image

牛津大学出版社是牛津大学的一个部门。它通过在全球范围内出版来推进大学在研究、学术和教育方面的卓越目标。

牛津 纽约

奥克兰 开普敦 达累斯萨拉姆 香港 卡拉奇

吉隆坡 马德里 墨尔本 墨西哥城 内罗毕

新德里 上海 台北 多伦多

在以下地区设有办事处

阿根廷 奥地利 巴西 智利 捷克共和国 法国 希腊

危地马拉 匈牙利 意大利 日本 波兰 葡萄牙 新加坡

韩国 瑞士 泰国 土耳其 乌克兰 越南

Oxford是牛津大学出版社在英国和某些其他国家的注册商标。

在美国由以下出版社出版

牛津大学出版社

纽约州纽约市麦迪逊大道198号 10016

美国国会图书馆出版物编目数据

克拉克，安迪，1957—

驾驭不确定性：预测、行动与具身心智 / 安迪 · 克拉克。

页 厘米

包括参考书目和索引。

ISBN 978－0－19－021701－3（布面： alk. paper) eISBN 978－0－19－021703－7 1. 预测（逻辑） 2. 元认知(Metacognition)。 3. 不确定性。I. 标题。 BC181.C534 2016 128’.2—dc23

致克里斯汀 · 克拉克和阿莱克萨 · 莫尔科姆编码器、解码器以及介于两者之间的一切

目录

参考文献

索引

序言：会预测的肉体

‘它们是肉做的。’

‘肉？’

‘肉。它们是肉做的。’

‘肉？’

“毫无疑问。我们从地球的不同地方选了几个样本，将它们带到我们的侦察飞船上，进行了全面的探测。它们完全是肉做的。”

这是那些困惑不解的非碳基外星人的开场白，他们的对话记录在科幻作家特里·比森(Terry Bissom)的精彩短篇小说《异族》(*Omni*, 1991)中。当得知这些肉质的陌生生物甚至不是由非肉类智能体建造，也没有在肉质外表内部隐藏任何简单的非碳基中央处理器时，外星人的困惑进一步加深。相反，它是*完全的肉质结构*。正如其中一个外星人惊呼的，甚至大脑也是肉做的。结论令人震惊：

“是的，会思考的肉！有意识的肉！会爱的肉。会做梦的肉。肉就是全部！你明白了吗？”

由于无法克服最初的惊讶和厌恶，外星人很快决定继续他们的星际旅程，用不可避免的俏皮话”谁想要见肉呢？“将我们这些短命的肉质大脑抛在一边。

撇开这种恐肉症不谈，外星人的困惑确实有道理。会思考的肉，会做梦的肉，有意识的肉，能够理解的肉。至少可以说，这似乎不太可能。当然，如果我们由硅或其他任何物质构成，那也同样令人惊讶。谜团一直存在：纯粹的物质如何设法产生思维、想象、梦境，以及整个心理、情感和智能行动的盛宴。会思考的物质，会做梦的物质，有意识的物质：这就是难以理解的事物——无论它由什么构成。但有一个新兴的线索。这是众多线索中的一个，即使它是一个好线索，也不会解决所有问题和困惑。尽管如此，它是一个真正的线索，也为我们考虑（在某些情况下重新发现）许多先前的线索提供了一个便利的保护伞。

这个线索可以用一个词来概括：预测(prediction)。为了快速流畅地处理一个不确定和嘈杂的世界，像我们这样的大脑已经成为预测的大师——通过实际上试图保持领先于嘈杂和模糊的感官刺激波浪来冲浪。熟练的冲浪者保持”在口袋里”：靠近但略微领先于波浪破裂的地方。这提供了力量，当波浪破裂时，不会被卷入其中。大脑的任务与此类似。通过不断尝试预测传入的感官信号，我们能够——以我们很快将详细探讨的方式——了解周围的世界并在思维和行动中与这个世界互动。成功的、与世界互动的预测并不容易。它关键地依赖于同时估计世界的状态和我们自己的感官不确定性。但如果做对了，主动的行为者既能了解又能行为上地与他们的世界互动，安全地驾驭一波又一波的感官刺激。

当物质被组织得不得不尝试（并一次次地尝试）成功预测冲刷其能量敏感表面的复杂能量变化时，它具有许多有趣的特性。我们将看到，如此组织的物质被理想地定位于感知、理解、梦想、想象，以及（最重要的）行动。感知、想象、理解和行动现在被捆绑在一起，作为同一个潜在的预测驱动的、不确定性敏感的机制的不同方面和表现而出现。

然而，为了这些特性的完全显现，需要满足更多条件。需要预测其时变（和行动相关）扰动的能量敏感表面需要是众多且多样化的。在我们人类身上，它们包括眼睛、耳朵、舌头、鼻子，以及整个有些被忽视的感觉器官——皮肤。它们还包括一系列更”内向”的感觉通道，包括本体感觉(proprioception)（对身体部位相对位置和所部署力量的感觉）和内感受(interoception)（对身体生理状况的感觉，如疼痛、饥饿和其他内脏状态）。对这些更内向通道的预测将在行动的核心解释以及解释感觉和意识体验方面证明至关重要。

也许最重要的是，预测机制本身需要在一个独特复杂的、多层次的、多样化的内部环境中运作。在这个复杂的（且可重复重新配置的）神经经济体中，交易的是概率预测，在每个层次上都被我们对自身不确定性的变化估计所影响。在这里，不同的（但密切相关的）神经元群体学会预测在许多空间和时间尺度上获得的各种对生物体显著的规律性。这样做时，它们锁定指定从线条和边缘、到斑马条纹、到电影、意义、爆米花、停车场，以及你最喜爱的足球队攻防特征模式等一切的模式。由此揭示的世界是一个为人类需求、任务和行动量身定制的世界。这是一个由可供性(affordances)——行动和干预机会——构建的世界。这是一个被反复利用的世界，通过巧妙的行动例程来减少神经处理的复杂性，这些例程改变了具身的、预测性的大脑的问题空间。

但是，你可能会问，所有这些对我们自身感官不确定性的预测和估计从何而来？即使基于预测的与传感器上能量变化的相遇确实能够——正如我将论证的那样——揭示出一个适合参与和行动的复杂结构化世界，这些预测所反映的知识仍然需要得到解释。在一个特别令人满意的转折中，事实证明，不断尝试（使用多级内在组织）预测（部分自发的）感官数据变化的肉体，恰好处于学习这些规律本身的有利位置。因此，学习和在线处理使用相同的基本资源得到支持。这是因为，如果这个故事是正确的，感知我们的身体和世界涉及学习预测我们自己不断演化的感官状态——这些状态既对行动中的身体作出反应，也对世界作出反应。预测这些不断变化的感官状态的好方法是了解导致变化的世界（包括我们自己的身体和行动）。因此，预测感官刺激变化的尝试本身就可以逐渐安装使预测成功的模型。正如我们将看到的，预测任务因此是一种”引导天堂(bootstrap heaven)“。

像这样的肉体也是想象和做梦的肉体。这样的肉体能够使用使其能够将传入的感官数据与结构化预测相匹配的知识和连接，“自上而下”地驱动自己的内在状态。而做梦和想象的肉体（至少在潜力上）是能够利用其想象进行推理的肉体——思考它可能或可能不会执行的行动。结果是一个令人信服的”认知套餐协议”，其中感知、想象、理解、推理和行动从预测性、不确定性估计的大脑的轰鸣和研磨中共同涌现。基于这种预测的潜在流动进行感知和行动的生物是与结构化和有意义的世界保持丰富认知接触的主动、知识渊博、富有想象力的存在。那个世界是由期望模式构成的世界：一个意外缺失与任何具体事件一样在感知上突出的世界，一个我们所有心理状态都被对自身不确定性的微妙估计所着色的世界。

然而，为了完成这幅图景，我们必须将内在预测引擎定位在其适当的家园中。那个家园——正如冲浪形象也有力地暗示的那样——是位于物质和社会结构的多重赋权网络中的移动化身代理。为了与像我们这样的代理的思考和推理建立完整而令人满意的联系，我们必须考虑到我们学习、行动和推理所处的复杂社会和物理”设计师环境”的无数影响。没有这种环境，我们这种对世界的选择性反应永远不可能出现或得以维持。正是在丰富的身体、社会和技术背景下运作的预测大脑，将像我们这样的心智引入物质领域。在这里尤其是，对预测的关注带来了丰厚的回报，为思考神经、身体和环境资源在每时每刻协调成有效的瞬时问题解决联盟提供了新的有力工具。在我们故事的结尾，预测大脑将被揭示出来，不是作为一个孤立的内在”推理引擎”，而是一个面向行动的参与机器——连接大脑、身体和世界的密集相互交换模式中的一个使能节点（尽管，碰巧是肉质的）。

A.C.

爱丁堡，2015年

致谢

本书受益于众多人士的帮助和建议。特别感谢Karl Friston、Jakob Hohwy、Bill Phillips和Anil Seth。你们的耐心和鼓励使这整个项目成为可能。同时也感谢Lars Muckli、Peggy Series、Andreas Roepstorff、Chris Thornton、Chris Williams、Liz Irvine、Matteo Colombo以及预测编码研讨会的所有参与者（爱丁堡大学信息学院，2010年1月）；感谢Phil Gerrans、Nick Shea、Mark Sprevak、Aaron Sloman以及在牛津大学哲学系举办的英国心理网络首次会议的参与者（2010年3月）；感谢Markus Werning、Albert Newen以及2010年欧洲哲学与心理学学会会议的组织者和参与者（德国波鸿鲁尔大学，2010年8月）；感谢Nihat Ay、Ray Guillery、Bruno Olshausen、Murray Sherman、Fritz Sommer以及知觉与行动研讨会的参与者（新墨西哥州圣菲研究所，2010年9月）；感谢Daniel Dennett、Rosa Cao、Justin Junge和Amber Ross（被飓风艾琳阻挡的2011年认知巡航的船长和船员）；感谢Miguel Eckstein、Mike Gazzaniga、Michael Rescorla以及加州大学圣巴巴拉分校心理研究圣哲中心的教职员工和学生，作为2011年9月的访问学者，我有幸在那里试讲了大部分材料；感谢我2013年《行为与脑科学》论文的所有评论者，特别提及Takashi Ikegami、Mike Anderson、Tom Froese、Tony Chemero、Ned Block、Susanna Siegel、Don Ross、Peter König、Aaron Sloman、Mike Spratling、Mike Anderson、Howard Bowman、Tobias Egner、Chris Eliasmith、Dan Rasmussen、Paco Calvo、Michael Madary、Will Newsome、Giovanni Pezzulo和Erik Rietveld；感谢Johan Kwisthout、Iris van Rooij、Andre Bastos、Harriet Feldman以及2014年5月在荷兰莱顿举办的洛伦兹中心研讨会”人类概率推理的视角”的所有参与者；感谢Daniel Dennett、Susan Dennett、Patricia和Paul Churchland、Dave Chalmers、Nick Humphrey、Keith Frankish、Jesse Prinz、Derk Pereboom、Dmitry Volkov以及莫斯科国立大学的学生们，他们在2014年6月在格陵兰冰山之间的难忘船旅中讨论了这些（以及许多其他）主题。同时感谢Rob Goldstone、Julian Kiverstein、Gary Lupyan、Jon Bird、Lee de-Wit、Chris Frith、Richard Hensen、Paul Fletcher、Robert Clowes、Robert Rupert、Zoe Drayson、Jan Lauwereyns、Karin Kukkonen和Martin Pickering对部分材料的信息丰富且富有启发性的讨论。感谢我的牛津大学出版社编辑Peter Ohlin的持续关注、帮助和支持，感谢Emily Sacharin对图表的耐心工作，感谢Molly Morrison在最后阶段的编辑支持，以及Lynn Childress出色的文稿编辑。衷心感谢我的美妙伴侣Alexa Morcom、我的了不起的母亲Christine Clark、整个Clark和Morcom家族、Borat和Bruno（两只猫），以及我们在爱丁堡内外的所有朋友和同事。最后，这本书也纪念我美好的兄弟James（“Jimmy”）Clark。

本书大部分内容是新写的，但有些章节重现或借鉴了以下已发表文章的材料：

接下来是什么？预测大脑、情境代理和认知科学的未来。《行为与脑科学》，36(3)，2013，181-204。

精确性的多面性。《心理学前沿》，4(270)，2013。doi:10.3389/fpsyg.2013.00270。

作为预测的知觉。见D. Stokes、M. Mohan和S. Biggs（编辑），《知觉及其模式》。纽约：牛津大学出版社，2014。

期待世界：知觉、预测和人类知识的起源。《哲学杂志》，110(9)，2013，469-496。

具身预测。对T. Metzinger和J. M. Windt（编辑）《开放心智项目》的贡献。法兰克福：心智集团开放获取出版。2015，在线地址：http://open-mind.net

感谢编辑和出版商允许在此使用这些材料。图表来源在图例中注明。

冲浪不确定性

引言

[猜谜游戏]

这是一本关于像我们这样的生物如何认识世界并在其中行动的书。在这种认知性参与的核心处（如果这些故事正确的话）存在着一个简单但极其强大的技巧或策略。这个技巧就是试图利用你对世界的了解来预测即将到来的感官刺激。失败的猜测会产生”预测误差”，然后被用来招募新的更好的猜测，或者为更缓慢的学习和可塑性过程提供信息。植根于自组织动力学的这些”预测处理”(predictive processing, PP)模型为感知、行动和想象性模拟提供了令人信服的解释。它们为人类体验的本质和结构提供了新的解释。它们将一个自我驱动的循环因果交互过程置于中心舞台，在这个过程中，行动持续地选择新的感官刺激，沿途整合环境结构和机会。因此，PP为近期关于具身心智的工作提供了完美的神经计算伙伴——如果我的论证正确的话——这项工作强调世界通过感知-运动活动的循环而持续参与。如果这是正确的，预测性大脑不是一个孤立的推理引擎，而是一个面向行动的参与机器。此外，它是一个完美定位的参与机器，能够选择节俭的、基于行动的例程，减少对神经处理的需求，并提供快速、流畅的适应性成功形式。

当然，预测是一个难以捉摸的东西。即使在这些页面中，它也以许多微妙（和不那么微妙）的不同形式出现。预测在其最熟悉的表现形式中，是一个人所从事的活动，目的是预期未来事件的形态。这样的预测是知情的、有意识的猜测，通常提前很久就做出，由前瞻性代理者为了其计划和项目而产生。但这种预测，这种有意识的猜测，并不是我将要呈现的故事核心所在的那种。那个故事核心处的是一种不同的（虽然最终并非无关的）预测，一种不同的”猜测”。它是一种自动部署的、深度概率性的、无意识的猜测，作为支撑和统一感知与行动的复杂神经处理例程的一部分而发生。在后一种意义上，预测是大脑为使具身的、环境情境化的代理者能够执行各种任务而做的事情。

这种对预测的强调在心智科学中有着悠久的历史。但只是在过去十年左右，关键要素才汇聚到一起，提供了（至少是潜在的）第一个真正统一的感知、认知和行动解释。这些要素包括预测驱动学习的力量和可行性的实际计算演示、补充计算框架的新神经科学框架的出现，以及大量实验结果，表明在内在经济中，预测、预测误差信号和我们自身感官不确定性的估计发挥着重要且以前被低估的作用。这样的工作跨越了强调内在模型建构活动重要性的解释与认识大脑、身体和世界之间精细劳动分工的解释之间曾经坚固的分界线。

如我将要描述的，PP最好被视为Spratling (2013)所称的”中间层模型”。这样的模型对神经实现的许多重要细节不做具体规定，而是旨在”识别在神经系统不同结构中运作的共同计算原理，并提供对经验数据的功能性解释，这些解释可以说是与神经科学最相关的”。因此，它提供了一套独特的工具和概念，以及一种中层组织性草图，作为三角定位感知、认知、情感和行动的手段。PP框架特别有吸引力，因为它深刻阐明了神经经济在更大的具身的、涉及世界的行动关系网中的嵌套。应用于各种正常和病理情况及现象，PP为理解人类体验的形式和结构提供了新的方式，并开启了与自组织、动力学和具身认知工作的有趣对话。

这幅图景表明，像我们这样的大脑是预测引擎，不断试图猜测传入感官阵列的结构和形态。这样的大脑是持续主动的，不懈地寻求使用传入信号为自己生成感官数据（在对许多传统智慧的令人惊讶的颠倒中），主要作为检查和纠正其最佳自上而下猜测的手段。然而，至关重要的是，所有内在猜测的形态和流动都被对传入信号不同方面的相对不确定性（因此是我们对其的信心）的变化估计所灵活调节。结果是一个动态的、自组织的系统，其中信息的内在（和外在）流动根据任务的需求以及内在（内感知的）和外在环境的变化细节而不断重新配置。

这些描述与人类体验本身的形式和结构形成了引人注目的联系。这种联系很明显，例如，这类模型能够轻松适应意外感觉的感知奇异性（比如当我们强烈期待咖啡时却喝到了茶），或者遗漏现象的显著突出性（比如当一个在预期音乐序列中突然缺失的音符，在被强烈的特定缺失感替代之前，似乎几乎存在于体验中）。PP模型还阐明了各种病理学和障碍，从精神分裂症和自闭症到”功能性运动综合征”（其中期望和改变的置信度分配（精确度(precision)）导致疾病或损伤的虚假感觉”证据”）。

更一般地说，PP框架为熟悉的人类体验提供了一个引人注目且统一的描述，如产生心理意象的能力、对可能的未来选择和行动进行”离线”推理的能力，以及理解其他主体意图和目标的能力。我们将看到，所有这些能力都自然地从使用自上而下的”生成模型”(generative model)（稍后将详细介绍）中产生，作为智能猜测（预测）跨多个时空尺度的感觉数据演变的手段。同样的装置为理解意义本身的性质和可能性提供了坚实而直观的把握。因为能够在多个时空尺度上预测感觉数据的演变，或者如我将论证的，正是将世界作为意义的场所来遇见。这就是在感知、行动和想象中，遇见一个结构化的世界，这个世界由对有机体具有突出意义的远端原因所填充，并倾向于以某些方式演化。如果PP是正确的，感知、理解、行动和想象会通过我们持续尝试猜测感觉信号而不断地共同构建。

这种猜测策略具有深远的重要性。它提供了将感知、行动、情感和对环境结构的利用结合成一个功能整体的共同货币。在当代认知科学术语中，这种策略依赖于”多层概率生成模型”的获取和部署。

这个短语初次遇到时有点令人望而生畏。但基本思想并不复杂。可以立即用一个例子来说明，以我的一位真正的哲学和科学英雄丹尼尔·丹尼特(Daniel Dennett)告诉我的一个故事为出发点，那是2011年夏末，我们因飓风艾琳而相当精彩地被困在他位于缅因州的农舍里。早在1980年代中期，丹尼特遇到了一位同事，一位著名的古生物学家，他担心学生们通过简单地复制（有时甚至描摹）他真正想让他们理解的地层学图纸来在作业中作弊。地层学图纸——字面意思是层次的绘制——是那些显示岩层和分层的地质横截面之一，其作用是揭示复杂结构如何随时间积累。然而，成功描摹这样的图纸很难算是你地质学掌握程度的良好指标！

为了解决这个问题，丹尼特设想了一个后来被原型化并称为SLICE的设备。SLICE由软件工程师史蒂夫·巴尼(Steve Barney)命名和构建，在原始的IBM PC上运行，本质上是一个绘图程序，其操作不像我们许多人在童年时玩过的Etch-a-Sketch设备。不同的是，这个设备以更复杂和有趣的方式控制绘图。SLICE配备了许多”虚拟”旋钮，每个旋钮控制一个基本地质原因或过程的展开，例如，一个旋钮会沉积沉积物层，另一个会侵蚀，另一个会侵入熔岩，另一个会控制断裂，另一个会褶皱，等等。

作业的基本形式如下：学生被*给予*一幅地层学图纸，*必须重新创建*这幅图片，不是通过描摹或简单复制，而是通过以正确的顺序转动正确的旋钮。实际上，学生在这里没有选择，因为这个设备（不像Etch-a-Sketch或当代绘图应用程序）不支持像素级或逐行控制。让地质描绘出现在屏幕上的*唯一方法*是找到正确的”地质原因”旋钮（例如，沉积沉积物，然后侵入熔岩）并以正确的强度部署它们。这意味着以正确的顺序和正确的强度（“体积”）转动正确的旋钮，以重新创建原始图纸。丹尼特的想法是，如果一个学生能够做到这一点，那么她确实对隐藏的地质原因（如沉积、侵蚀、熔岩流和断裂）如何合作产生不同地层图所捕获的物理结果有相当多的理解。在我将在本书其余部分使用的术语中，成功的学生必须掌握一个”生成模型”，使她能够基于对可能起作用的原因以及它们如何相互作用以产生目标图纸的理解，为自己构建各种地质结果。因此，目标图纸扮演了学生需要使用她对地质领域的最佳模型重新构建的感觉证据的角色。

我们可以更进一步，要求学生指挥一个*概率性*生成模型。对于单张展示的图片，通常会有多种不同的方式来组合各种旋钮调整来重现它。但其中一些组合可能代表了比其他组合更可能的序列和事件。因此，要获得满分，学生应该部署与最*可能*带来观察结果的事件集（“隐藏地质成因”集）相对应的调整集。更高级的测试可能会在显示图片的同时明确排除最常见的成因集，从而迫使学生找到另一种方式来实现该状态（迫使她找到*次最可能*的成因集，以此类推）。

SLICE允许用户运用她对地质成因（沉积、侵蚀等）及其相互作用的了解来自我生成地层图像：一个与作业中设定图像相匹配的图像。这阻止了作弊。通过调整旋钮来匹配给定图片（毕竟只是一组像素），这些旋钮通过侵蚀、沉积和断裂等隐藏成因的精确控制混合来创建图片，这*就是*对地质学和地质成因的深入理解。

这是一个很好的——尽管有限的——基本技巧的例证，如果我将要考虑的模型是正确的，大脑使用这个技巧来理解从世界接收到的感官信号的持续变化（实际上只是撞击的能量）。这表明，我们通过识别最有可能产生当前撞击我们众多（外感受性、本体感受性和内感受性）感官受体的能量模式的相互作用的世界成因集来感知世界。通过这种方式，我们通过（如果你愿意的话）猜测世界来看世界，使用感官信号在进行过程中完善和细化猜测。

注意，现实世界的感知匹配任务针对的不是单一的静态结果（如SLICE中），而是一个不断演化的现实世界场景。因此，在我们将要考虑的情况中，匹配传入信号需要了解场景元素如何在多个空间和时间尺度上演化和相互作用。这可以通过预测导向神经组织的多层次特性来实现——我们将在后续章节中对此类多层次架构进行更多讨论。

为了完成这个例证，我们需要从等式中移除学生和尽可能多的先验知识！产生的设备是SLICE*：SLICE的自给自足版本，它获得了自己关于地质隐藏成因的知识。至少在微观层面，使用分层（深层、多层）架构中的预测驱动学习，这是可以做到的。关键思想——这个思想似乎在当代认知科学中以各种形式出现——是我们也通过尝试使用高级生物大脑特有的大量递归连接，从上而下为自己生成传入的感官数据来学习*世界。这之所以有效，是因为好的模型能做出更好的预测，我们可以通过缓慢修正模型（使用已充分理解的学习程序）来改进模型，从而逐步提高它们对感官流的预测能力。

对于被动感知的简单但不现实的情况（见下文），核心思想现在可以总结如下。感知世界就是用恰当的多层次预测流来迎接感官信号。这些预测旨在使用关于相互作用远端成因的存储知识来”从上而下”构建传入的感官信号。以这种方式容纳传入的感官信号本身就已经对世界有了相当多的理解。部署这种策略的生物学会成为自身感官刺激的知识渊博的消费者。它们了解自己的世界，以及居住在其中的实体和事件类型。部署这种策略的生物，当它们看到草以某种特定方式颤动时，*已经在期待*看到美味的猎物出现，*已经在期待*感受到自己肌肉紧绷准备扑击的感觉。一个对其世界有这种把握力的动物或机器，已经深入到理解该世界的事业中。这整个关于感知和学习的基础故事在本书的第一部分中呈现。

但在这个被动感知的简洁图景中，缺少了一个关键要素。缺少的是行动，而行动改变了一切。我们大量的递归神经元集合不仅仅是在不断地嗡嗡作响试图预测感觉流。它们通过引起身体运动来不断地*产生*感觉流，这些运动选择性地收获新的感觉刺激。感知和行动因此锁定在一种无尽的循环拥抱中。这意味着我们需要做出进一步的——也是认知上至关重要的——修正。我们的新玩具系统是一个机器人（称之为Robo-SLICE），它必须以响应其接收到的感觉刺激的方式行动。也就是说，它必须以适合身体和环境原因组合的方式行动，这些原因（它估计）使当前感觉数据最有可能。世界参与行动现在成为解释的核心，使Robo-SLICE能够主动寻求和选择自己的感觉刺激，将其感受器暴露于对其生存以及它所调谐的目标和目的而言重要的能量输入类型。此外，Robo-SLICE能够利用对世界的行动来降低自身内部处理的复杂性，选择节俭、高效的例程，以运动和环境结构换取昂贵的计算。

想象Robo-SLICE是一个艰难的任务，我们这个小思想实验的局限性很快就显现出来了。因为我们没有为SLICE指定任何生活方式、生态位或基本关注点，所以我们不知道什么可能构成对感觉输入的*恰当行动*。我们也还没有展示持续的感觉信号预测尝试如何*导致*这样的代理恰当地行动，以旨在使感觉信号逐渐与其自身感觉预测的某个特殊子集一致的方式采样其世界。这个巧妙的技巧将我们的一些感觉预测转化为自我实现的预言，这是本书第二部分的主题。

我们还没有到达那里。为了完成这个图景，我们需要赋予Robo-SLICE改变其自身社会和物质环境长期结构的能力，以便栖居在一个”重要的能量输入”在需要时更可靠地提供的世界中。这种世界结构化，一代又一代地重复，也使像我们这样的存在能够构建越来越好的思考世界，允许冲击的能量引导越来越复杂的行为形式，并使思想和理性能

够渗透到以前”禁区”的领域。这就是情境化Robo-SLICE——一个自主的、主动的、学习系统，能够以改善其思维并服务（和改变）其需求的方式改变其世界。这是本书第三部分的主题。

我想通过提及一些关键特征和吸引力来结束这个简短的引言，这些特征和吸引力从上面的草图中刚好可见（我希望如此）。

一个特征是认知共现(co-emergence)。多级感觉预测策略同时支持丰富的、揭示世界的感知形式，对学习友好，并且看起来很有希望将想象（以及我们稍后将看到的更有方向性的心理模拟形式）引入生物学舞台。如果我们通过使用关于世界的储存知识”自上而下”地生成传入感觉数据来感知世界，以重新创造那些感觉模式的显著方面，那么感知本身涉及一种理解形式：它涉及知道事物是什么样的，以及它们倾向于如何随时间演化。想象也在那里，因为自我生成（至少是）感觉信号近似的能力意味着能够以这种方式感知世界的系统也可以离线为自己生成类似感知的状态。这种自我生成只是使它们能够用恰当的感觉预测迎接传入感觉刺激的相同生成模型式知识的另一种用途。

这样的解释与最近支持所谓”贝叶斯大脑假设”的实验结果爆炸式增长形成了深刻而有启发性的联系：大脑以某种方式实现处理，这种处理近似于权衡新证据与先验知识的理想方式的假设。找到使当前感觉数据最有可能的隐藏原因集合对应于贝叶斯推理(Bayesian inference)。

当然，这样的大脑不会在所有时候都把一切都弄对！我最近被英国陆军北极探险队队长亨利·沃斯利中校的以下描述所震撼：

白化天气很棘手。那是当云层变得如此之低以至于遮蔽地平线的时候。阿蒙森称之为”白色黑暗”。你对距离或高度没有概念。有一个关于他的故事，他以为看到地平线上有个人。当他开始走路时，他意识到那只是在他前面三英尺处的一块狗粪。

那个”人”的感知很可能在全局上（即，总体上，在我们栖居的世界类型中，根据我们的信息状态）是”贝叶斯”最优的，因为阿蒙森相信他正在看向远方地平线。也就是说，沃斯利上校的大脑可能一直在以最好的方式将先验知识和当前证据结合在一起。尽管如此，地平线上那个人的感知实际上只是在跟踪一块粪便。每当我在本书中使用令人担忧的”最优”一词时，我的意思只是指这种”狗粪最优性”。

另一个大规模特征是整合。要探索的视角将使我们能够以统一的方式看待许多核心认知现象（感知、行动、推理、注意力、情感、体验和学习），并将为”具身和情境”认知科学的许多主张提供定性甚至定量理解的方法。这些后续整合之所以成为可能，是因为有一种认知共同基础，即”使感官数据流变得越来越可预测的方式”。我们可以调节生成模型的参数来匹配感官数据。但我们也可以改变数据，通过调节参数使其更容易被捕获——这些改变可能通过我们的即时行动，以及更长期的环境重构来实现。我相信，这种将概率神经处理工作与具身和行动作用研究统一起来的潜力，是新兴框架最吸引人的特征之一。

这些相同的解释为思考人类体验的形态和本质开辟了新途径。通过突出预测和（重要的是）神经系统对这些预测可靠性的估计，它们为包括精神分裂症、自闭症以及功能性运动和感官症状在内的各种病理学和障碍投下了新的光芒。它们还帮助我们欣赏神经典型人类体验的广阔而复杂的空间，并可能提供暗示（特别是当我们纳入关于我们自己不断演变的内脏状态的内感受预测时）关于意识感受和体验的机制起源。

尽管如此，也许值得强调的是，预测并不是大脑认知工具包中的唯一工具。因为预测，至少在即将探讨的相当特定的意义上，涉及在相当短的时间尺度内，使用在线处理期间计算的预测误差信号对传入感官信号进行自上而下的近似。这是一种强大的策略，很可能是各种认知和行为效应的基础。但它肯定不是大脑可用的唯一策略，更不用说主动代理了。适应性反应是一个多层面的事物，多种策略必须结合起来，使主动代理与它们复杂的、部分自我构建的世界保持联系。

但即使在这个更广阔的竞技场上，预测也可能发挥关键作用，有助于我们许多不同内外资源的时时刻刻的*协调*，以及构建与世界的核心智能接触形式。出现的图景是，基于预测的处理（正如我们将看到的，通过可变的”精度权重”）选择瞬态神经元集合。这些瞬态集合招募并不断被身体行动招募，这些行动可能利用各种环境机会和结构。通过这种方式，预测大脑的愿景与具身的、环境情境化的心智愿景建立了完整而富有成效的联系。

最后，重要的是要注意，在目前的工作中确实有两个故事。一个是大脑作为多层次概率预测引擎的极其广泛的愿景。另一个是更具体的提案（“分层预测编码”或预测处理(PP)），关于如何讲述这样的故事。完全可能的是，即使更具体提案(PP)的许多细节被证明是错误的或不完整的，广泛的故事仍将被证明是正确的。追求更具体提案的附加价值是双重的。首先，该提案代表了目前可用的更一般故事的最彻底阐述版本。其次，它是一个已经应用于大量——且不断增长的——现象的提案。因此，它作为某种故事处理广泛议题潜力的强有力说明，阐明了感知、行动、推理、情感、体验、理解其他代理，以及各种病理学和失调的性质和起源。

这些都是令人兴奋的发展。我认为，它们的结果不是又一个”心智的新科学”，而是潜在的更好的东西。因为出现的实际上只是许多以前方法的最佳部分的汇合点，结合了连接主义和人工神经网络工作的元素、当代认知和计算神经科学、处理证据和不确定性的贝叶斯方法、机器人技术、自组织，以及对具身的、环境情境化心智的研究。通过将大脑视为不安分的、主动的器官，不断受驱动去预测并帮助带来感官刺激的作用，我们可能正在瞥见一些核心功能，这些功能使得大约三磅重的移动的、基于身体的脑肉，沉浸在人类社会和环境漩涡中，能够了解和参与其世界。

第一部分

预测的力量

1

预测机器

1.1 感知咖啡的两种方式

当我与同事短暂交谈后重新进入办公室，视觉感知到我留在桌上等待的那杯热气腾腾的咖啡时，会发生什么？一种可能是，我的大脑接收到大量视觉信号（为了简化，想象成一个激活像素阵列），这些信号迅速指定许多基本特征，如线条、边缘和色块。然后这些基本特征被前向传递、逐步累积，并在适当时绑定在一起，产生越来越高层次的信息类型，最终形成形状和关系的编码。在某个时刻，这些复杂的形状和关系激活存储的知识体系，将感觉的前向流动转化为揭示世界的感知：看到装在时髦复古绿色杯子中的热气腾腾的美味咖啡。这样的模型虽然在此处表达得过于简单，但相当准确地对应传统认知科学方法，该方法将感知描述为”自下而上”特征检测的累积过程。[1]

以下是另一种情况。当我重新进入办公室时，我的大脑已经控制着一套涉及咖啡和办公室的复杂期望。瞥向我的桌子，一些快速处理的视觉线索引发一连串视觉处理，其中传入的感觉信号（被称为”驱动”或”自下而上信号”）与向下（和横向[2]）预测流相遇，这些预测涉及这个小世界最可能的状态。这些预测反映了我们持续神经元处理中大部分嗡嗡作响、积极主动的性质。那股向下流动的预测洪流致力于预先指定沿着适当视觉（和其他）路径的各种神经元群的可能状态。向下（和横向）流动的预测涉及展开遭遇的所有方面，不仅限于形状和颜色等简单视觉特征。它可能包括丰富的多模态关联和（正如我们将在后续章节中看到的）运动和情感预测的复杂组合。随后发生快速交换（多个自上而下和自下而上信号之间的活跃舞蹈），其中不正确的向下流动”猜测”产生错误信号，这些信号横向和向上传播，并被用来促成更好更好的猜测。当预测流充分解释传入信号时，视觉场景被感知。随着这个过程的展开，系统试图（在多个空间和时间尺度上）为自己生成传入的感觉信号。当这成功时，并且建立了匹配，我们体验到结构化的视觉场景。

我认为，这就是我实际看到咖啡的方式。这个基本提案需要细致入微、精炼，并在我们故事展开时反复增强，它让人想起那个朗朗上口（但如我们将在第6章中看到的，可能有点扭曲）的格言：*感知是受控的幻觉*。[3] 我们的大脑试图猜测外面有什么，在那种猜测适应感觉冲击的程度上，我们感知世界。

[1.2 采纳动物的视角]

所有这些知识——驱动构成感知基础的预测的知识，以及（正如我们稍后将看到的）行动的知识——最初是如何产生的？我们当然必须在获得做出关于世界的预测所需的知识*之前*，先感知体验世界？在这种情况下，感知体验毕竟

不可能需要基于预测的处理。

要解决这个担忧，我们需要坚定地区分可能被视为通过感官对能量模式的单纯转换，与那些丰富的、揭示世界的感知之间的区别，后者产生于（如果这个故事是正确的）当且仅当那种转换能够与恰当的自上而下期望相遇时。问题然后变成：基于单纯的能量转换，恰当的期望如何能够形成并发挥作用？这个故事的一个吸引人的特征是，*完全相同的过程*（试图预测当前感觉输入）可能最终成为学习和在线反应的基础。

一个很好的起点（遵循Rieke等人, 1997和Eliasmith, 2005）是对比某个系统的外部观察者视角与动物或系统本身的视角。外部观察者可能能够看到，例如，青蛙大脑中的某些神经元只有在存在特定的视网膜刺激模式时才会放电，这种模式最常出现在某种多汁的猎物（比如苍蝇）在舌头可及范围内的时候。这种神经元活动模式可能被称为”表征”猎物的存在。但这样的描述，虽然有时有用，可能会让我们忽视一个更紧迫的问题。这就是青蛙或任何其他我们感兴趣的系统如何能够对世界有所把握的问题。为了更好地看清这个问题，我们需要采用（在稍后将解释的无害意义上）的不是外部观察者的视角，而是青蛙本身的视角。做到这一点的方法是*只考虑青蛙可获得的证据*。事实上，即使这样也可能产生误导，因为它似乎邀请我们从某种青蛙视角来想象世界。相反，以所讨论意义上的动物视角，就是将自己限制在只能从撞击青蛙感觉器官的能量刺激流中了解到的东西。这些能量刺激确实可能由我们作为外部观察者所认识的苍蝇引起。但对青蛙大脑来说，唯一可获得的东西就是由从世界流经其众多受体的能量对其感觉系统造成的扰动。这意味着，正如Eliasmith (2005, p. 102)指出的，“可能刺激的集合是未知的，动物必须根据各种感觉线索推断所呈现的内容”。我要补充的是（为了预期我们后续的一些讨论），“推断所呈现的内容”与选择恰当的行动密切相关。在这个意义上，动物的视角由通过感觉受体状态的变化向动物大脑提供的信息所决定。但”处理”这些信息的全部意义在于帮助选择合适的行动，考虑到动物的当前状态（例如，它有多饥饿）和由那些撞击能量所索引的世界状态。

同样值得强调的是，这里使用的”信息”一词仅仅是对”能量传递”的描述（见Eliasmith, 2005；Fair, 1979）。也就是说，信息的谈论最终必须简单地以撞击感觉受体的能量来兑现。如果我们要避免再次将观察者的视角非法地引入我们对有信息的观察者如何自然成为可能的解释中，这是至关重要的。这样使用的信息谈论不对信息是关于什么的做任何假设。这是必要的，因为弄清楚这一点正是动物大脑需要做的事情，如果它要作为控制环境适应性反应的赋权资源。因此，（具身的、情境化的）大脑的任务是将这些能量刺激转化为行动指导信息。

Eliasmith指出，以这种方式”采取动物视角”的早期例子可以在Fitzhugh (1958)的工作中找到，他探索了尝试仅从动物神经纤维的反应来推断环境原因性质的方法，故意忽略他作为外部观察者对这些纤维可能响应的刺激类型的所有了解。通过这种方式：

正如大脑（或其部分）从感觉信号推断世界状态一样，Fitzhugh试图确定世界中有什么，一旦他知道了神经纤维对未知刺激的反应。他故意将使用的信息限制为动物可获得的信息。通过观察者视角获得的”额外”信息只在事后用于”检查他的答案”；它不用于确定动物正在表征什么。(Eliasmith, 2005, p. 100)

Fitzhugh面临着一项艰巨的任务。然而，这本质上就是生物大脑的任务。大脑必须在没有任何形式的直接接触信号来源的情况下，发现关于撞击信号可能原因的信息。它在任何直接意义上”了解”的全部，就是它自己的状态（例如，脉冲序列）如何流动和改变。这样的状态也在具身有机体中产生影响，其中一些（外部观察者可能注意到）是对感觉换能器本身运动的影响。通过这种方式，主动代理能够构建它们自己的感觉流，影响它们自己能量刺激的潮起潮落。我们稍后会看到，这是信息的一个重要额外来源。但它并不改变基本情况。说系统直接接触的全部就是它自己的感觉状态（感觉受体上的刺激模式）仍然是正确的。

仅仅基于感觉受体上的刺激模式，具身的、情境化的大脑如何能够改变和适应，以作为有用的节点（承担相当大代谢费用的节点）来产生适应性反应？注意这种概念与将问题设定为建立环境状态和内部状态之间映射关系的概念有多么不同。任务不是找到这样的映射，而是仅从变化的输入信号本身推断信号源（世界）的性质。

[1.3 自举天堂中的学习]

预测驱动学习提供了一种非常强大的方式，在这样最初看似不乐观的条件下取得进展。理解这一点的最佳方式是首先回顾通过为系统提供适当”教师”所能实现的学习类型。当然，这里的教师通常不是人类代理，而是某种自动化信号，它告诉学习者在给定当前输入的情况下应该做什么或得出什么结论。依赖于这种信号的系统被称为”监督”学习者。最著名的这类系统是依赖所谓”误差反向传播”的系统（例如，Rumelhart, Hinton, & Williams, 1986a,b；以及Clark, 1989, 1993的讨论）。在这些类型的”连接主义”系统中，当前输出（通常是对输入的某种分类）与正确输出（在某些标记或预分类训练数据中设置）进行比较，编码系统专业知识的连接权重被缓慢调整，使未来的响应越来越符合要求。这种缓慢自动调整的过程（称为梯度下降学习）能够使从随机连接权重分配开始的系统逐渐达到预期水平——如果一切顺利的话[4]。

连接主义系统的发展和完善是通向预测处理(PP)模型的漫长谱系中的关键步骤，我们很快就会考虑这些模型。事实上，预测处理（以及更广义的分层贝叶斯）模型可能最好被视为同一广泛谱系内的发展（讨论见McClelland, 2013和Zorzi et al., 2013）。在此类工作之前，人们很容易[5]简单地否认有效和基础的学习是可能的，因为感官证据的获得显然微乎其微。相反，人类知识的大部分可能简单地是先天的，在数千年来逐渐安装在我们神经回路的形状和功能中。

连接主义学习模型对此类论证提出了重要质疑，表明从我们实际遇到的统计上丰富的感官数据中学到很多东西实际上是可能的（综述见Clark, 1993）。但标准（反向传播训练）连接主义方法在两个方面受到阻碍。第一个是需要提供足够数量的预分类训练数据来驱动监督学习。第二个是在多层形式中训练此类网络的困难[6]，因为这需要以难以确定的方式在所有层中分布对错误信号的响应。预测驱动学习，应用于多层设置中，解决了这两个问题。

让我们先讨论训练信号。思考预测驱动学习的一种方式是将其视为提供一种无害（即生态可行）的监督学习形式。更准确地说，它提供了一种自监督学习形式，其中”正确”响应以一种持续滚动的方式由环境本身反复提供。因此，想象你是那个大脑/网络，忙于转换来自世界的信号，只能检测你自己感官寄存器中的持续变化。在这样帮助自己获得”动物视角”的同时，你能做的一件事就是忙于尝试预测这些寄存器的下一个状态。

然而，这里的时间故事实际上比看起来要复杂得多。以离散时间步长的预测来思考这个过程是最容易的。但是，实际上，我们将要探索的故事描绘了大脑参与连续的感官预测过程，其中目标是一种滚动的现在。一旦我们看到感知本身是一种预测驱动的构造，它总是植根于过去（系统知识）并在多个时间和空间尺度上预期未来，“预测现在”和”预测非常接近的未来”之间的界限就消失了[7]。

预测任务的好消息是世界本身现在将提供你需要的训练信号。因为你的感官寄存器的状态*将*改变，以系统性地由传入信号驱动的方式，随着你周围的世界发生变化。通过这种方式，你自己感官受体的演变状态提供了一个训练信号，允许你的大脑”自我监督”自己的学习。因此：

> 预测形式的学习特别引人注目，因为它们提供了无处不在的学习信号来源：如果你尝试预测接下来发生的一切，那么每一个时刻都是学习机会。这种普遍的学习例如可以解释婴儿如何似乎神奇地获得对世界如此复杂的理解，尽管他们看似惰性的外显行为（Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996）——他们正在越来越熟练地预测接下来会看到什么，因此，正在发展越来越复杂的世界内部模型。(O’Reilly et al. (submitted) p. 3)

因此，如此构想的预测任务是一种自举天堂(bootstrap heaven)。例如，要预测句子中的下一个词，了解语法（以及更多内容）会很有帮助。但学习关于语法（以及更多内容）的惊人数量知识的一种方法是寻找预测句子中下一个词的最佳方式。这正是世界可以自然提供的那种训练，因为你的预测尝试很快就会得到对应于（你猜对了）句子中下一个词的语音形式的跟进。因此，你可以使用预测任务来自举你通往语法的道路，然后在未来的预测任务中使用这种

语法。如果处理得当，这种自举（它实现了”经验贝叶斯”的一个版本；参见Robbins, 1956）确实被证明提供了一个非常有效的训练机制。

因此，预测驱动学习利用了一个丰富、免费、持续可用、自举友好的教学信号，即不断变化的感官信号本身。无论任务是生态基础性的（例如，预测不断演化的视觉场景以发现捕食者和猎物）还是更高级的生态任务（例如，检测咖啡杯或预测句子中的下一个词），世界都可以被依赖来提供训练信号，使我们能够将当前预测与实际感知到的能量输入模式进行比较。这允许成熟的学习算法挖掘关于实际构造传入信号的相互作用的外部原因（“潜在变量”）的丰富信息。但在实践中，这需要一个额外且至关重要的成分。那个成分就是多层次学习的使用。

[1.4 多层次学习]

在分层（多层）设置中运行的预测驱动学习很可能是学习我们这种世界的关键：一个高度结构化的世界，在许多空间和时间尺度上显示规律性和模式，并由各种相互作用且复杂嵌套的远端原因填充。正是在那里，感官预测和分层学习相结合，我们定位了一个相对于先前工作的重要计算进步。这一进步源于亥姆霍兹(Helmholtz)（1860）对感知作为概率性、知识驱动推理过程的描述。从亥姆霍兹那里来了一个关键思想：感官系统从事着从其身体（感官）效应推断世界原因的棘手业务。因此，这是一种对外界存在什么的押注，通过询问世界必须如何存在才能使感官器官以其当前方式受到刺激来构建。使这变得棘手的部分原因是，单一的这种感官刺激模式将与许多不同的世界原因集合一致，仅通过它们相对的（且依赖于上下文的）发生概率来区分。

亥姆霍兹的洞察启发了MacKay (1956)、Neisser (1967)和Gregory (1980)的有影响力的工作，作为被称为”分析-综合”(analysis-by-synthesis)的认知心理学传统的一部分（回顾参见Yuille & Kersten, 2006）。[8] 在机器学习中，这些洞察帮助启发了从恰当命名的”亥姆霍兹机器”(Helmholz Machine)工作开始的一系列关键创新（Dayan et al., 1995; Dayan and Hinton, 1996; 另见Hinton and Zemel, 1994）。亥姆霍兹机器是多层架构的早期例子，可以训练而不依赖于实验者预分类的例子。相反，系统通过尝试使用自己的向下（和侧向）连接为自己生成训练数据来”自组织”。也就是说，系统不是从分类（或”学习识别模型”）数据的任务开始，而是首先必须学习如何使用多层系统为自己生成传入数据。

这似乎是一个不可能的任务，因为生成数据需要系统希望获得的那种知识。例如，要生成某种公共语言特有的语音结构，你需要已经了解很多关于各种语音和它们如何被发音和组合的知识。[9] 同样，如果系统已经掌握了该语言中语音结构化语音的生成模型，它就可以学习执行分类任务（以声音流为输入并提供语音解析作为输出）。但是，在两者都缺乏的情况下，你从哪里开始？答案似乎是”逐渐地，同时在两个地方”。这个僵局在原则上至少通过开发新的学习程序得到解决，这些程序对”自举天堂”进行迭代访问。

使这一切成为可能的关键发展是发现了诸如”清醒-睡眠算法(wake-sleep algorithm)“([Hinton et al., 1995])这样的算法，该算法利用每个任务（识别和生成）逐步引导另一个任务。这个算法允许系统通过以交替方式训练两组权重，在”迭代估计”过程中同时学习识别和生成模型。清醒-睡眠算法使用自身的自上而下连接为隐藏单元提供期望的（目标）状态，从而（实际上）使用一个试图为自己创造感官模式的生成模型（在”幻想”中，如有时所说）来自监督其感知”识别模型”的发展。重要的是，这种过程即使从整个网络的小随机权重分配开始也能成功（有用的综述见[Hinton, 2007a]）。

在这种非常具体的意义上，生成模型旨在通过推断能够产生该结构的因果矩阵来捕获某些观察输入集合的统计结构。在引言中，我们遇到了SLICE*，其获得的生成模型结合了隐藏的地质成因，如断裂和熔岩侵入，以最佳地解释（通过自上而下生成）目标地质（地层）图像中的像素模式。一个良好的视觉概率生成模型同样会寻求捕获较低层次视觉模式（最终是视网膜刺激）如何由推断的远端成因相互作用网络生成的方式。在给定上下文中遇到的某种视网膜刺激模式，因此可能最好使用一个生成模型来解释，该模型（作为一个诚然简化的示例）将相互作用的主体、

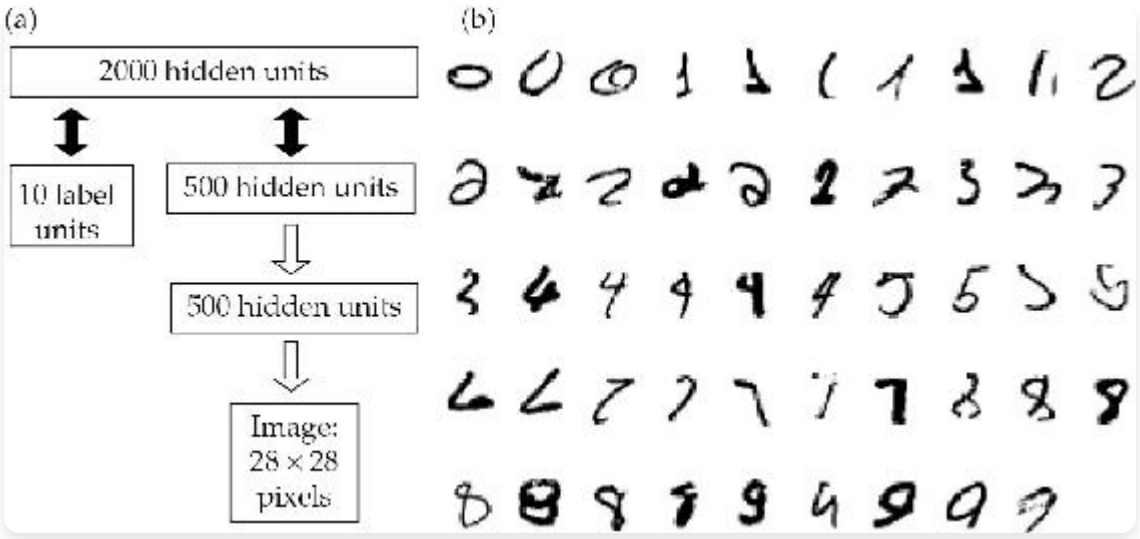
对象、动机和运动的顶层表征与多个中间层结合，这些中间层捕获颜色、形状、纹理和边缘如何结合并在时间上演化。当这些隐藏成因（跨越多个空间和时间尺度）的组合稳定为一个连贯的整体时，系统就使用存储的知识自生成了感官数据，并感知到一个有意义的结构化场景。

值得再次强调的是，对结构化远端场景的这种把握必须仅使用从动物视角可获得的信息来生成。也就是说，它必须是一种完全根植于任何可能由于动物进化历史而存在的预结构化（大脑和身体的）与感觉受体记录的能量刺激模式相结合的把握。实现这种把握的系统方法是通过持续尝试使用多层次架构自生成感官信号来提供。在实践中，这意味着多层次系统内的自上而下和横向连接开始编码在多个空间和时间尺度上运作的相互作用成因的概率模型。如果这些方法是正确的，我们通过找到最可能的相互作用因子（远端成因）集合来识别对象、状态和事务，这些因子的组合将生成（因此预测并最好地解释）传入的感官数据（例如，见[Dayan, 1997]；[Dayan et al., 1995]；[Hinton et al., 1995]；[Hinton & Ghahramani, 1997]；[Hinton & Zemel, 1994]；[Kawato et al., 1993]；[Mumford, 1994]；[Olshausen & Field, 1996]）。

1.5 解码数字

考虑一个我们许多人每天都要解决的实际问题，通常无需太多有意识的努力：识别手写数字的问题。诚然，这样的情况越来越少了。但当有人确实在浴室镜子上留下那张匆忙潦草的便条时，区分这些数字可能是至关重要的（或者至少是有约会或没约会的问题）。我们是如何做到的呢？

机器学习理论家Geoffrey Hinton描述了一个能够进行手写数字识别的基准机器学习系统（见[Hinton, 2007a, b]；[Hinton & Nair, 2006]；另见[Hinton & Salakhutdinov, 2006]）。该系统的任务就是简单地对手写数字图像进行分类（手写1、2、3等的图像）。也就是说，系统旨在将高度可变的手写数字图像作为输入，并输出正确的分类（将数字识别为1、2、3…等的实例）。设置（见[图1.1]）涉及三层特征检测器，在未标记的手写数字图像语料库上进行训练。但系统并不是试图直接训练多层神经网络来分类图像，而是学习并部署了上述类型的概率生成模型。它学习了一个多层生成模型，能够使用其自上而下的连接（followed by some additional fine-tuning）为自己产生此类图像。学习的目标因此是逐步”调整自上而下连接的权重，以最大化网络生成训练数据的概率”([Hinton, 2007a], p. 428)。成功感知的路径（在这种情况下是手写数字识别）因此经由一个实际上更接近于数字主动生成（例如，在计算机图形学中）的学习策略。



image

图1.1 学习识别手写数字

a. 用于学习数字图像和数字标签联合分布的生成模型。(b) 一些测试图像，网络能正确分类这些图像，尽管它以前从未见过这些图像。

来源: Hinton, 2007a。

结果令人印象深刻。训练后的网络能正确识别图1.1中显示的所有（通常书写很糟糕的）示例，尽管这些示例都不在训练数据中。该网络使用包含60,000个训练图像和10,000个测试图像的基准数据库进行了广泛测试。它的表现超过了所有更标准的（反向传播训练的）人工神经网络，除了那些专门为此任务”手工制作”的网络。它的表现也几乎与更计算密集的”支持向量机”方法相当。最重要的是，它使用了一种学习程序来实现这一点，如果我们将要考虑的理论是正确的，这种学习程序反映了大脑功能组织的一个关键方面：使用自上而下的连接来生成系统试图响应的数据的版本。

这种方法可以应用于任何结构化领域。Hinton自己的变体（我必须强调，这与我们即将关注的”预测处理”模型在一些非常重要的方面有所不同）已成功应用于文档检索、预测句子中的下一个词以及预测人们会喜欢什么电影等各种任务（见Hinton & Salakhutdinov, 2006; Mnih & Hinton, 2007; Salakhutdinov et al., 2007）。为了开始理解这种方法的潜在力量，有助于注意到整个数字识别网络，Hinton指出，只有”大约0.002立方毫米小鼠皮层的参数数量”，“数百个这种复杂度的网络可以容纳在高分辨率fMRI扫描的单个体素内”（Hinton, 2005, p. 10）。Hinton谦虚地使用这个比较，作为戏剧化机器学习仍需走多远路程的手段。但从另一个角度来看，它邀请我们欣赏像我们这样复杂的大脑，采用某种版本的强力学习策略，可能对围绕我们的世界获得多么深刻的掌握。

1.6 处理结构

预测驱动的多层学习还解决了早期（基于误差反向传播的）连接主义(connectionist)方法的另一个关键缺陷——它们缺乏处理结构的原则性方法。这是表示和处理”复杂的、铰接结构”（Hinton, 1990, p. 47）的需要，如部分-整体层次结构：元素形成整体的结构，这些整体本身可以是一个或多个更大整体的元素。“经典人工智能”的工作为这个问题提供了一个相当（过于）直接的解决方案。传统的符号方法使用”指针”系统，其中一个本质上任意的数字对象可以用来访问另一个对象，后者本身可能用来访问另一个对象，等等。在这样的系统内，符号可以被视为”对象的一个小的（通常是任意的）表示，它提供到同一对象的更完整表示的’远程访问’路径”。通过这种方式，“许多（小）符号可以组合在一起，创建某些更大结构的’完全铰接’表示”（两个引用均来自Hinton, 1990, p. 47）。这样的系统确实可以以允许元素轻松共享和重新组合的方式表示结构化（嵌套的、通常是层次的）关系。但它们在其他方面证明是脆弱和不灵活的，无法显示流畅的上下文敏感响应，并且在需要在时间压力下的现实世界环境中指导行为时会陷入困境。

需要以原则性方式处理结构化领域推动了早期对连接主义替代使用经典的、句子样内部表示形式的怀疑（例如，Fodor & Pylyshyn, 1988）。但跳到2007年，我们发现Geoffrey Hinton，一位不喜欢夸张的机器学习理论家，写道”反向传播学习的局限性现在可以通过使用包含自上而下连接的多层神经网络来克服，并训练它们生成感觉数据而不是分类数据”（Hinton, 2007a, p. 428）。对结构的担忧得到直接解决，因为（正如我们将在文本中经常看到的）预测驱动的学习，当它在这些多层设置中展开时，倾向于分离出在空间和时间的不同尺度上运作的相互作用的远端（或身体）原因。

这非常重要，因为结构化域在自然界和人类构建的世界中无处不在。语言表现出密集嵌套的组合结构，其中词汇形成子句，子句形成完整的句子，而这些句子本身通过在更大的语言（和非语言）环境中定位来理解。每个视觉场景，如城市街道、工厂车间或宁静的湖泊，都嵌入着多个嵌套结构（例如，商店、店门、门口的购物者；树木、树枝、树枝上的鸟、叶子、叶子上的图案）。音乐作品表现出这样的结构：总体序列由重复和重组的子序列构建，每个子序列都有自己的结构。我们可以合理地认为，世界被我们人类（无疑大多数其他动物也是如此）认识为一个有意义的舞台，由清晰分明和嵌套的元素结构所填充。这种结构化的认知形式是通过预测驱动学习成为可能的（我们即将探讨其方式），在这种学习中，自上而下的连接试图使用关于在多个空间和时间尺度上运作的世界原因的知识来构建感官场景。

1.7 预测处理

正是这种转变——使用自上而下连接的策略，试图通过深层多级联的方式，利用世界知识生成感官数据的虚拟版本——构成了”分层预测编码”感知方法的核心（Friston, 2005；Lee & Mumford, 2003；Rao & Ballard, 1999）。分层预测编码（或”预测处理”（Clark (2013)））结合了自上而下概率生成模型的使用和对这种影响如何以及何时运作的特定愿景。借鉴商业”线性预测编码”的工作，该愿景描述了神经信号的自上而下和横向流动持续（不仅在学习期间）旨在预测当前的感官轰炸，只留下任何未预测的元素（以残余”预测误差”的形式）在系统内向前传播信息（参见Brown et al., 2011；Friston, 2005, 2010；Hohwy, 2013；Huang & Rao, 2011；Jehee & Ballard, 2009；Lee & Mumford, 2003；Rao & Ballard, 1999）。

转置到神经域（以我们即将探讨的方式），这使得预测误差成为任何尚未解释的感官信息的一种代理（Feldman & Friston, 2010）。这里的预测误差报告了由遇到的感官信号与预测信号之间的不匹配所引起的”惊讶”。更正式地——为了将其与正常的、经验性的惊讶感区分开来——这被称为惊讶度(surprisal)（Tribus, 1961）。如前所述，我将描述这样的系统从事”预测处理”。在如此使用”预测处理”而不是止步于更常见的用法”预测编码”时，我意在强调区分这些方法的不仅仅是使用被称为预测编码的数据压缩策略（稍后会详细介绍）。相反，它是在分层（即多级）系统部署概率生成模型的非常特殊的背景下使用该策略。这样的系统展现出强大的学习形式——正如我们稍后将看到的——提供丰富的上下文敏感处理形式，并能够在多层级联中灵活地结合自上而下和自下而上的信息流。

预测编码最初是作为信号处理中的数据压缩策略开发的（历史见Shi & Sun, 1999）。因此考虑一个基本任务，如图像传输。在大多数图像中，一个像素的值通常预测其最近邻居的值，差异标记重要特征，如对象之间的边界。这意味着丰富图像的代码可以（对于适当知情的接收者）通过仅编码”意外”变化来压缩：实际值偏离预测值的情况。最简单的预测是相邻像素都共享相同的值（例如，相同的灰度值），但也可能有更复杂的预测。只要有可检测的规律性，预测（因此这种特定形式的数据压缩）就是可能的。偏离预测的情况随后携带”新闻”，量化为实际当前信号与预测信号之间的差异（“预测误差”）。这在带宽上提供了重大节省，这种经济性是1950年代贝尔实验室的詹姆斯·弗拉纳根(James Flanagan)等人开发这些技术的驱动力（评论见Musmann, 1979）。

通过有信息预测的数据压缩允许相当简单的编码被重构成原始视觉和声音的丰富精美呈现。这种技术在例如视频的运动压缩编码中占据重要地位。这是一个特别有效的应用，因为重构视频序列当前帧中图像所需的大部分信息已经存在于之前处理过的帧中。以移动物体对比稳定背景的情况为例。在这种情况下，当前帧的大部分背景信息可以假定与前一帧相同，预测误差信号表示被遮挡内容的变化或摄像机平移。这种技术也不仅限于如此简单的情况。移动物体的可预测变换本身也可以被纳入考虑（只要速度，甚至加速度保持不变），使用所谓的运动补偿预测误差(motion-compensated prediction error)。因此，构建一个非常简单的移动图像第2帧所需的所有信息可能已经存在于第1帧中，并应用运动补偿。要接收第二帧，你只需要传输一个简单的消息（例如，非正式地说”与之前相同，只是将所有内容向右移动两个像素”）。原则上，每一个系统性和规律性的变化都可以被预测，只留下真正意外的偏差（例如，出现意外的、之前被遮挡的物体）作为残余误差的来源。

因此，诀窍是用智能和知识来换取当日编码和传输的成本。请注意，这里没有任何东西要求接收者参与*有意识*的预测或期待过程。重要的只是接收系统能够以充分利用已检测到的或被证明有用假设的任何规律性的方式重构传入信号。通过这种方式，像我们这样的动物可能通过使用我们已经知道的信息来预测尽可能多的当前感官数据，从而节省宝贵的神经带宽。当窗帘以恰当的方式开始移动时，你似乎几乎看到了你心爱的猫或狗（即使在这种场合下，只是风在起作用），你可能一直在使用训练有素的预测机制来开始完成感知序列，节省带宽并（通常）因此更好地了解你的世界。

因此，预测处理(predictive processing)在多层次双向级联中结合了”自上而下”的概率生成模型和高效编码传输的核心预测编码策略。如果预测处理的故事是正确的，那么感知确实是一个过程，在这个过程中我们（或者更确切地说，我们大脑的各个部分）试图猜测外面有什么，使用传入信号更多地作为调整和细化猜测的手段，而不是作为世界状态的丰富（且带宽成本高昂）编码。当然，这并不意味着感知体验只有在所有前向流动的错误被消除后才发生。这里丰富完整的感知只有在向下预测在许多层次上匹配传入的感官信号时才形成。但这种匹配（正如我们稍后将看到的）本身是一个零碎的过程，在这个过程中，对场景的一般性质或”要点”的快速感知可以通过训练有素的前馈扫描来完成，这种扫描对简单的（例如，低空间频率）线索敏感。然后，更丰富的细节与相对于随后的自上而下预测波计算的残余误差信号的逐步减少同时出现。如果这些模型是正确的，正在进行的感知过程是大脑使用存储的知识以逐步更精细的方式预测当前感官刺激引发的多层神经反应模式的问题。这反过来强调了我们期望的结构（有意识和无意识的）可能决定我们看到、听到和感觉到的大部分内容的惊人程度。

在本书的其余部分，我们将因此探索两个不同但重叠的故事。第一个是关于大脑（特别是新皮层）基本上作为概率预测内在引擎的一般性且日益得到支持的愿景（参见，例如，Bubic等人，2010；Downing，2007；Engel，Fries等人，2001；Kvergaga等人，2007）。另一个是一个具体提议（分层预测编码(hierarchical predictive coding)，或”预测处理”），描述了多层概率预测核心过程的可能形状和性质。这个提议在概念上优雅，在计算上有充分根据，似乎有合理的前景在神经上实现。因此，它正在被广泛应用，新现象以惊人的（有时甚至是令人担忧的）速度被纳入其范围。它提供了一个非常全面的愿景。然而，我们不应该忘记，在这个一般领域有许多可能的模型。

1.8 信号新闻

为了在这些理论基础上添加一些基于实例的内容，首先考虑一个关于视网膜基本预测编码策略的演示（Hosoya et al., 2005）。该论述的起点是一个已确立的观点，即视网膜神经节细胞参与某种形式的预测编码，因为它们的感受野显示出中心-周围空间拮抗作用，以及一种时间拮抗作用。这意味着神经回路基于局部图像特征预测空间和时间中附近点的可能图像特征（基本上假设附近点会显示相似的图像强度），并从实际值中减去这个预测值。因此，编码的不是原始值，而是原始值与预测值之间的差异。通过这种方式，“神经节细胞发出的信号不是原始视觉图像，而是在空间和时间均匀性假设下偏离可预测结构的部分”（Hosoya et al., 2005, p. 71）。这节省了带宽，同时标记了输入信号中最”有新闻价值”的内容（使用Hosoya等人的表述）。

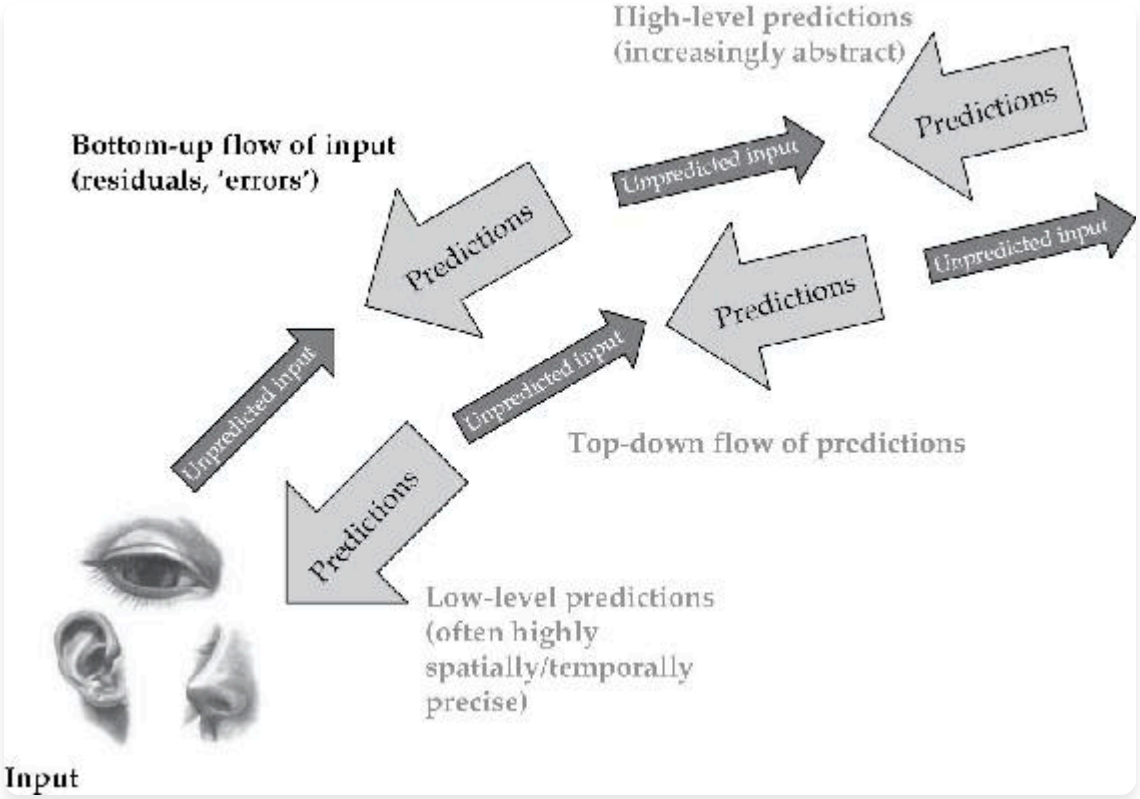
这些预测显著性的计算可能仅基于平均图像统计数据。然而，这种方法在许多生态现实情况下会导致问题。考虑”墨西哥步行鱼”面临的问题，这是一种经常在水环境和陆地之间移动的蝾螈。空间和时间中附近点在图像强度上通常相似的空间尺度在这种情况下变化很大，因为不同类型场景的统计特性不同。在不太戏剧性的情况下也是如此，例如当我们从建筑物内部移动到花园或湖泊时。因此，Hosoya等人预测，为了有效且适应性强的编码，视网膜神经节细胞的行为（具体来说，它们的感受野特性）应该因适应当前场景或环境而变化，表现出他们所称的”动态预测编码”。

将蝾螈和兔子置于不同环境中，并记录它们的视网膜神经节细胞活动，Hosoya等人证实了他们的假设：在几秒钟内，大约50%的神经节细胞改变了它们的行为，以跟上不同环境变化的图像统计特性。然后提出并使用执行反赫布学习(anti-Hebbian learning)形式的简单前馈神经网络测试了一种机制。反赫布前馈学习中，单元间的相关活动导致抑制

而非激活（例如，参见Kohonen, 1989），能够创建所谓的”新颖性过滤器”，学习对输入中最高度相关（因此最”熟悉”）的特征不敏感。当然，这正是学习忽略输入信号中最具统计可预测性元素所需要的，正如动态预测编码所建议的那样。更好的是，有神经上合理的方式来实现这样的机制，使用无长突细胞突触来介导可塑抑制连接，进而改变视网膜神经节细胞的感受野（详情见Hosoya et al., 2005, p. 74），以抑制刺激中最相关的成分。总之，视网膜神经节细胞似乎正在进行一个计算上和神经生物学上可解释的原始图像输入动态预测重编码过程，其效果是”从视觉流中剥离可预测的、因此较少有新闻价值的信号”（Hosoya et al., 2005, p. 76）。

[1.9 预测自然场景]

预测处理将这种对新闻价值的生物学强调推进了几个步骤，为皮层组织本身提供了新的视角。在预测处理方案中，传入的感觉信号被使用多层向下和横向影响构建的”猜测”流所迎接，残余不匹配以错误信号的形式向前（和横向）传递。这些提议的核心是前向和后向路径之间的深层功能不对称性——功能上讲，“在寻求解释的原始数据（自下而上）和寻求确认的假设（自上而下）之间”（Shipp, 2005, p. 805）。在这样的多层级层次系统中，每一层都将下层的活动视为感觉输入，并尝试用适当的自上而下预测流来迎接它（关于这个基本图式，见图1.2）。文献中有几个这方面的具体例子（见Huang & Rao, 2011的综述）。



image

图1.2 **基本预测处理图式**

大脑中信息传递的预测处理观点的高度图式化视图。自下而上的输入在来自层次结构中更高层级的先验（信念/假设）背景下被处理。输入中未被预测的部分（错误）沿层次结构向上传递，导致后续预测的调整，循环继续。

来源：改编自Lupyan & Clark, In Press

Rao和Ballard (1999)提供了开创性的概念验证。在这项工作中，基于预测的学习针对从自然场景中提取的图像片段，使用多层人工神经网络。该网络除了使用向下和横向连接来匹配输入样本与成功预测这一任务外，没有预设任务，它发展出了具有简单细胞样感受野的单元嵌套结构，并捕获了各种重要的、经验观察到的效应。在最低层，存在某种能量刺激模式，由感觉受体从当前视觉场景产生的环境光模式中转导（我们假设）。然后这些信号通过多级级联进行处理，其中每一级都试图通过反向连接预测其下一级的活动。反向连接允许处理某一阶段的活动作为另一个输入返回到前一阶段。只要这能成功预测较低级别的活动，一切都很好，不需要采取进一步行动。但当出现不匹配时，就会发生”预测误差”，随之产生的（指示错误的）活动会横向传播到更高层级。这会自动在更高层级招募新的概率表征(probabilistic representations)，使得自上而下的预测能更好地抵消较低层级的预测误差（产生快速感知推理）。同时，预测误差被用来调整模型的长期结构，以减少下次的任何差异（产生较慢时间尺度的感知学习）。因此，层级间的前向连接仅传递”残余误差”(Rao & Ballard, 1999, p. 79)，即分离预测与实际较低层级活动的差异，而反向和横向连接（传达生成模型）则传递预测本身。改变预测对应于改变或调整你对较低层级活动隐藏原因的假设。在一个有具身的活跃动物的背景下，这意味着它对应于改变或调整你对如何应对世界的把握，考虑到当前的感官冲击。在皮层区域双向层级中并发运行这种预测误差计算，使得关于不同空间和时间尺度规律性的信息能够融入一个相互一致的整体中，其中每个这样的”假设”都被用来帮助调整其余部分。正如作者所说，“预测和纠错周期在整个层级中同时发生，因此自上而下的信息影响较低层级的估计，而自下而上的信息影响较高层级对输入信号的估计”(Rao & Ballard, 1999, p. 80)。在视觉皮层中，这样的方案表明从V2到V1的反向连接将携带对V1中预期活动的预测，而从V1到V2的前向连接将向前传递指示残余（未预测）活动的错误信号。前向和反向连接在作用上的这种功能不对称是预测处理(PP)愿景的核心。

为了测试这些想法，Rao和Ballard实现了一个这样的”预测估计器”的简单双向层级网络，并在从五个自然场景派生的图像片段上训练它（见图1.3）。使用逐步减少链接级联中预测误差的学习算法，并在暴露于数千个图像片段后，系统学会了使用第一级网络中的响应来提取诸如定向边缘和条纹等特征，而第二级网络则捕获这些特征的组合，对应于涉及更大空间配置的模式——例如，斑马的交替条纹。通过这种方式，层级预测编码架构仅使用从自然图像派生的信号的统计特性，就能够诱导出输入数据结构的简单生成模型。它学习了线条、边缘和条纹等特征的存在和重要性，以及这些特征的组合（如条纹），以能够更好地预测在空间或时间中接下来期望什么。用贝叶斯术语来说（见[附录1]），网络最大化了生成观察到的状态（感官输入）的后验概率，这样做时，诱导出了信号源中结构的一种内部模型。

image

图1.3 用于训练文中描述的三级层级网络的五个自然图像

来源: Rao & Ballard, 1999.

Rao和Ballard模型还展现了许多有趣的”非经典感受野”效应，如端点抑制。端点抑制（参见Rao & Sejnowski, 2002）发生在神经元对落在其经典感受野内的短线条产生强烈反应，但（令人惊讶地）随着线条变长，反应却逐渐减弱。这些效应（以及我们稍后将看到的一整套”情境效应”）自然地从分层预测机制的使用中出现。当线条变长时反应逐渐减弱，因为较长的线条和边缘是网络在训练中接触的自然场景中的统计常态。训练后，较长的线条因此是第二层网络首先预测的（并作为假设反馈）。因此，第一层”边缘细胞”在被较短线条驱动时的强烈放电并不反映这些细胞成功的特征检测，而是反映了一个更早阶段的错误或不匹配，因为短线段最初没有被更高层网络预测。

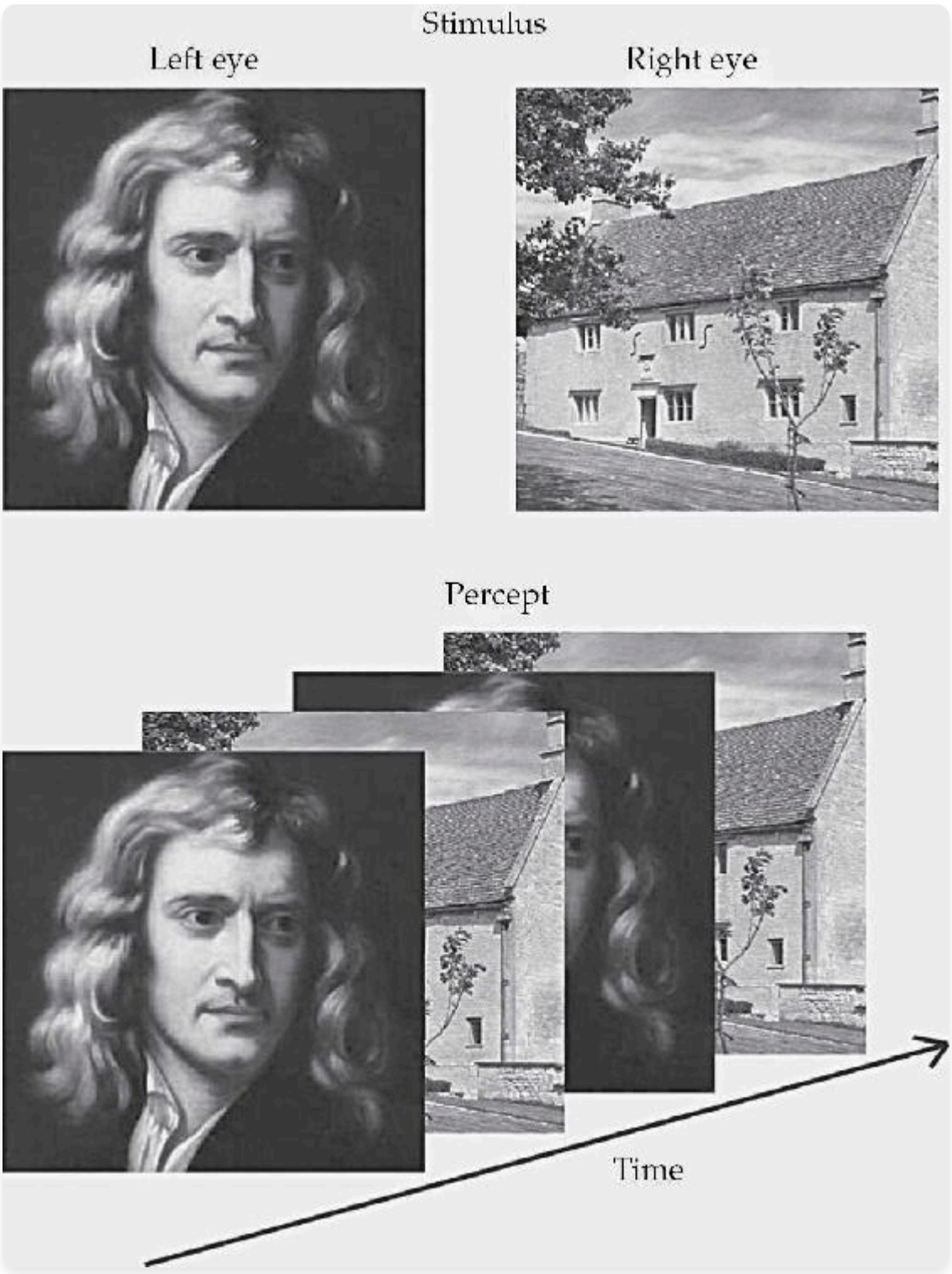
这个例子很好地说明了以简单累积特征检测流程思考的危险性，以及将处理流程重新思考为自上而下期望和自下而上错误校正混合的优势。它还突出了这些学习程序如何抓住由训练数据指定的世界结构。端点抑制细胞只是对训练中使用的自然场景统计的一种反应，反映了这些场景中线条和边缘的典型长度。在一个截然不同的世界中（如某些海洋生物的水下世界），这些细胞会学会非常不同的反应。

这些方法假设环境通过嵌套相互作用的原因产生感觉信号，感知系统的任务是通过学习和应用分层生成模型来反转这种结构，从而预测展开的感觉流。这种广泛类型的学习程序已在许多领域成功应用，包括语音感知、阅读以及识别自己和其他智能体的动作（参见Friston, Mattout, & Kilner, 2011；Poeppel & Monahan, 2011；Price & Devlin, 2011）。这并不奇怪，因为底层的理论基础是相当通用的。如果你想要预测某些感觉信号集如何随时间变化和演化，一个好的做法是学习这些感觉信号如何由相互作用的外部原因决定。而学习这些相互作用原因的好方法是尝试预测感觉信号如何随时间变化和演化。

[1.10 双眼竞争]

到目前为止，我们关于预测处理的例子都局限于一些相对低级的现象。然而，作为最后一个开场例证，这个例子很好地整合了到目前为止介绍的许多关键要素，让我们考虑Hohwy等人（2008）的双眼竞争分层预测编码模型。

双眼竞争[18]（参见图1.4）是一种引人注目的视觉体验形式，当使用特殊的实验装置时，每只眼睛（同时）接收不同的视觉刺激时就会发生。这可以通过使用红色和青色图形渲染的两个重叠图像来实现，使用一个红色镜片和一个青色镜片的特殊眼镜观看（这种装置被称为anaglyph 3D，曾经用于观看3D漫画或电影）。通过这些眼睛特定的滤镜，右眼可能接收到房子的图像，而左眼接收到面孔的图像。在这些（极其重要的是，人工的）条件下，主观体验以一种令人惊讶的”双稳态”方式展开。受试者报告的不是看到（视觉体验到）房子和面孔信息的持续融合，而是一种在看到房子和看到面孔之间的感知交替。过渡本身并不总是突然的，受试者经常报告另一个图像的元素在占主导地位之前逐渐突破（参见，例如，Lee等，2005）先前的图像，之后循环重复。



image

图1.4 双眼竞争图解

不同的图像被呈现给左眼和右眼（“刺激”）。受试者体验从一个图像（面孔）的感知到另一个图像（房子）的转换。注意”混合知觉”（由两个图像的部分组成）也会暂时体验到（“零碎竞争”）。

来源：Schwartz等，2012。经皇家学会许可。

正如Hohwy等人提醒我们的，双眼竞争已被证明是研究意识视觉体验的神经相关性的有力工具，因为传入信号保持恒定，而知觉在两者之间来回切换（Frith等，1999）。然而，尽管受到如此关注，这里起作用的精确机制并不为人所熟知。Hohwy等人的策略是退后一步，尝试从第一原理解释这种现象，以一种对许多看似不相关的发现都有意义的方式。特别是，他们追求他们称为”认识论的”方法：其目标是揭示双眼竞争作为对生态上不寻常刺激条件的合理（知识导向的）反应。

他们故事的起点再次回到大脑作为使用分层生成模型进行预测的器官这一新兴统一愿景。回想一下，在这些模型中，感知大脑的任务是通过匹配的自上而下预测来解释（适应或”解释掉”）传入或”驱动”的感觉信号。匹配越好，传播到层次结构上层的预测误差就越少。因此，更高层次的猜测充当了较低层次处理的先验，采用了所谓”经验贝叶斯”(empirical Bayes)的方式。

在这样的多层级设置中，视觉感知是由跨越（双向）处理层次结构多个层次的预测过程确定的，每个层次都关注不同类型和尺度的感知细节。所有通信区域都锁定在相互一致的预测编码体制中，它们的交互平衡最终选择关于视觉呈现世界状态的最佳整体（多尺度）假设。这是”做出最佳预测并考虑先验因此被分配最高后验概率”的假设（Hohwy et al., 2008，第690页）。在那一刻，其他整体假设被简单地排挤出去：它们被有效抑制，在最好解释驱动信号的竞争中败北。

然而，请注意这在预测处理级联背景下意味着什么。自上而下的信号将只解释（通过预测）驱动信号中符合（因此被预测）当前获胜假设的那些元素。然而，在双目竞争案例中（见图1.4），驱动（自下而上）信号包含的信息暗示视觉呈现世界的两种不同且不兼容的状态，例如，时间t位置x处的面孔和时间t位置x处的房屋。当其中一个被选为最佳整体假设时，它将解释假设预测的驱动输入的所有且仅有的那些元素。结果，该假设的预测误差减少。但与驱动信号中暗示替代假设的元素相关的预测误差并未因此被抑制，所以现在传播到层次结构上层。为了抑制*那些*预测误差，系统需要找到另一个假设。但在这样做之后（因此，将主导假设翻转到另一种解释），将再次出现大的预测误差信号，这次来自那些未被翻转解释解释的驱动信号元素。用贝叶斯术语来说（见[附录1]），这是一个没有独特稳定假设结合高先验和高似然的场景。没有单一假设能解释所有数据，所以系统在两个半稳定状态之间交替。它表现为双稳定系统，在Hohwy等人描述为包含双井的能量景观中最小化预测误差。

使这个解释不同于其竞争对手（如Lee et al., 2005）的是，后者假设输入之间存在一种直接的、注意力介导但本质上前馈的竞争，而预测处理解释假设连接假设集合之间的”自上而下”竞争。这种竞争的效果是选择性地抑制与当前获胜假设（“面孔”）适应的驱动（感觉）信号元素相关的预测误差。但这种自上而下的抑制保持与驱动信号剩余（房屋信号）元素相关的预测误差不受影响。这些误差然后传播到系统上层。为了解释它们，整体解释必须切换。这种模式重复，产生了在不一致刺激的双眼视觉期间经历的独特交替。

但为什么在这种情况下，我们不是简单地体验到组合或交织的图像：例如，一种房屋/面孔混合？虽然这种部分组合的感知确实会发生，并可能在短时间内持续，但它们从不完整（每个刺激的部分缺失）或稳定。这种混合不构成一个可行的假设，考虑到我们对视觉世界的更一般知识。因为那种一般知识的一部分是，例如，房屋和面孔不会在同一时间、同一尺度占据同一地方。这种一般知识本身可能被视为系统先验，尽管是在相对高的抽象程度上设定的（这种先验有时被称为”超先验”(hyperpriors)，我们将在后续章节中对它们有更多说明）。在手头的案例中，因此捕获的事实是”房屋和面孔在时间和空间上共定位的先验概率极小”（Hohwy et al., 2008，第691页）。事实上，这可能是某些更高层次假设之间竞争存在的深层解释——这些假设必须竞争，因为系统已经学会”只有一个物体可以同时存在于同一地方”（Hohwy et al., 2008，第691页）。

尽管具有这些吸引力，这里呈现的双眼竞争情景仍然是不完整的。特别是，显然这里存在强烈的注意力成分，其处理需要额外的资源（将在第2章中介绍）。此外，对所呈现场景的主动视觉探索——特别是我们只能以适合一种”解读”的方式进行视觉探索这一事实——可能在塑造我们的体验中起着重要作用。[20] 因此，这种增强将需要我们稍后（第二部分）称为”行动导向的预测处理”的更大装置。

[1.11 抑制和选择性增强]

如果这些模型是正确的，在知觉中成功表征世界关键依赖于消除感觉预测误差。因此，知觉涉及通过匹配一系列在各种空间和时间尺度上投射的预测来适应驱动（传入）感觉信号。在这种匹配成功的程度上，驱动感觉信号中被很好预测的方面被抑制或减弱——正如有时所说的，信号的这些方面被”解释掉了”。[21]

这种”解释掉”是重要和核心的，但需要非常谨慎地处理。它之所以重要，是因为它反映了预测处理模型的一个特征属性。这个特征是这些模型表现出的编码效率的根源，因为随后需要通过系统向前传递的只是残余误差信号（表示尚未解释的感觉信息），这是预测和驱动信号匹配后剩下的部分。[22] 但随后发生的系统展开不仅仅是抑制和减弱。因为除了抑制之外，PP还提供了锐化和选择性增强。

从根本上说，这是因为PP假设了一种双重架构：在每个层次上结合输入表征与误差估计和（见第2章）感觉不确定性。根据这个提议，真正被抑制、“解释掉”或抵消的是误差信号，在这些模型中，误差信号被描述为由专门的”误差单元”计算。这些单元与编码感觉输入原因的所谓表征单元相关但又不同。通过抵消一些误差单元的活动，一些横向相互作用的”表征”单元（横向和向下传递预测）的活动实际上可能最终被选择和锐化。

这样，预测处理理论避免了与假设感觉信号选定方面的自上而下*增强*的理论（如Desimone & Duncan, 1995的偏向竞争模型）的任何直接冲突。它避免了这种冲突，因为：

高层预测解释掉预测误差并告诉误差单元”闭嘴”[同时]编码感觉输入原因的单元通过与误差单元的横向相互作用被选择，这些相互作用介导经验先验。这种选择…在横向竞争表征中锐化响应。（Friston, 2005, p. 829）

这些效应通过我们将在第2章遇到的注意力（“精度加权”）机制进一步促进。目前要注意的是，PP理论与早期皮质响应（不同方面）的抑制*和*选择性增强都是一致的。[23]

预测处理提议最独特的地方（也是与传统真正分离的地方）是它将信息的前向流描述为仅传达误差，将后向流描述为仅传达预测。因此，PP架构在熟悉和新颖之间实现了相当微妙的平衡。仍然存在特征检测级联，具有选择性增强的潜力，并且越来越复杂的特征由距离感觉外围更远的神经群体处理。但感觉信息的前向流现在被预测误差的前向流所取代。这表示尚未解释的感觉信息。用稍后（第二和第三部分）将占据我们的更加行动导向的术语来说，这是尚未被利用来指导与世界恰当接触的感觉信息。

这种在抑制和选择性增强之间的平衡行为在架构上可能相当苛刻。在标准实现中，它需要假设存在”两个功能不同的子群体，分别编码感知原因的条件期望[表征、预测]和预测误差”（Friston, 2005, p. 829）。功能区别当然不需要意味着完全的物理分离。但这个文献中的一个常见推测描述浅层锥体细胞（前向神经解剖连接的主要来源）扮演误差单元的角色，横向和向前传递预测误差，而深层锥体细胞扮演表征单元的角色，横向和向下传递预测（基于复杂的生成模型）（见，例如，Friston, 2005, 2009; Mumford, 1992）。

重要的是要记住，“错误神经元”(error neurons)尽管有这样的标签，但同样可以被理解为一种表征神经元——只不过它们的功能作用是编码尚未解释的（或者更广泛地说，尚未适应的）感觉信息。因此，它们编码的内容只是相对于预测而言的。例如：

在早期视觉皮层中，预测神经元编码关于视野中某一点的预测方向和对比度的信息，错误神经元则发出观察到的方向和对比度与预测的方向和对比度之间不匹配的信号。在IT[下颞叶]皮层中，预测神经元编码关于物体类别的信息；错误神经元发出预测和观察到的物体类别不匹配的信号（den Ouden et al., 2012; Peelen and Kastner, 2011）。（Koster-Hale & Saxe, 2013, p. 838）

无论它如何实现（或可能不会实现）——关于一些关键可能性的有用概述，请参见Koster-Hale and Saxe (2013)——预测性处理需要预测编码和预测错误编码之间的*某种*形式的功能分离。[24]这种分离构成了架构的一个核心特征，使其能够将预测编码产生的抑制元素与多条自上而下信号增强路径相结合。

[1.12 编码、推理和贝叶斯大脑]

如果分层预测性处理理论被证明是正确的，神经表征编码的是以概率生成模型形式存在的”概率密度函数”(probability density functions)，推理流程遵循贝叶斯原理(Bayesian principles)（简要概述请参见[附录1]），该原理在先验期望与新的感觉证据之间保持平衡。这（Eliasmith, 2007）是对传统内部表征理解的一种背离，其全部含义尚未得到充分理解。这意味着神经系统从根本上适应处理不确定性、噪声和歧义，并且需要某些（也许是多种）具体的内部表征不确定性的方法。这里的非排他性选择包括使用不同的神经元群体、各种”概率群体编码”(probabilistic population codes)（Pouget et al., 2003）和相对时间效应（Deneve, 2008）（非常有用的综述请参见Vilares & Körding, 2011）。

因此，预测性处理理论共享了Knill and Pouget (2004, p. 713)所描述的”皮层处理的贝叶斯理论成功或失败的基本前提”，即”大脑以概率方式表征信息，通过编码和计算概率密度函数或概率密度函数的近似值”（p. 713）。这种表征模式意味着，当我们表征世界的状态或特征（如可见物体的深度）时，我们不是使用单一的计算值，而是使用条件概率密度函数，该函数编码”在给定可用感觉信息的情况下，物体位于不同深度Z的相对概率”（p. 712）。

这样的系统在什么意义上是真正贝叶斯的？根据Knill和Pouget的观点，“对贝叶斯编码假说的真正检验在于，导致感知判断或运动行为的神经计算是否考虑了处理每个阶段可用的不确定性”（2004, p. 713）。也就是说，合理的测试将涉及系统如何很好地处理其实际成功编码和处理的信息所特有的不确定性，以及（我想补充的）它用来做到这一点的策略的一般形态。

有越来越多的（尽管主要是间接的——见下文）证据表明，生物系统在多个领域中近似于这样理解的贝叶斯特征。仅举一个例子，Weiss et al. (2002)——在一篇题目颇具启发性的论文”运动错觉作为最优感知”中——使用最优贝叶斯估计器（“贝叶斯理想观察者”）表明，包括许多运动”错觉”在内的各种心理物理学结果（见6.9后续部分）自然地源于人类运动感知实现了这样一个估计器机制的假设。

例子可以不断增加（平衡的综述请参见Knill & Pouget, 2004）。至少在低级别、基本的和适应性关键的计算领域，生物处理可能非常接近贝叶斯最优性。但研究人员通常发现的不是我们人类在某种绝对意义上——相当令人惊讶地——是”贝叶斯最优”的（即相对于刺激中的绝对不确定性做出正确反应），而是我们在考虑我们实际掌握的信息所特有的不确定性方面通常是最优的或接近最优的：即我们实际部署的感知和处理形式所提供的信息（见Knill & Pouget, 2004, p. 713）。这意味着要考虑我们自己感觉和运动信号中的不确定性，并根据（通常非常微妙的）上下文线索调整不同线索的相对权重。最近的研究证实并扩展了这一评估，表明人类在感知和行动中，在各种领域内都表现为理性的贝叶斯估计器（Berniker & Körding, 2008; Körding et al., 2007; Yu, 2007）。

当然，系统的响应模式呈现这种形状的事实，并不能明确证明该系统正在实施某种形式的贝叶斯推理。在有限的领域内，即使是简单地将线索与响应关联的查找表也能（Maloney & Mamassian, 2009）产生与”贝叶斯最优”系统相同的行为表现。尽管如此，如果预测处理的故事是正确的，它将相当直接地支持神经系统近似于真正版本的贝叶斯推

理这一主张。[25] 一些最近的电生理学研究为这种广泛的可能性提供了强有力的支持，揭示了贝叶斯更新和预测性惊讶的独特皮层响应特征，并进一步表明大脑编码和计算加权概率。总结这些研究，作者们得出结论：

我们的电生理学发现表明，大脑充当贝叶斯观察者，即它可能会调整概率性内部状态，这些状态包含关于环境中隐藏状态的信念，在感觉数据的概率生成模型中。（Kolossa, Kopp, and Fingscheidt, 2015，第233页）。

[1.13 获取要义]

概率贝叶斯大脑不会简单地表示”猫在垫子上”，而是会编码一个条件概率密度函数，反映给定可用信息下这种情况（以及任何在某种程度上得到支持的替代方案）的相对概率。这一估计既反映了来自多个感官通道的自下而上的影响，也反映了各种类型的先验信息。值得停下来检视这种精妙的自上而下/自下而上舞蹈可能展开的诸多方式。

在处理的早期阶段，PP系统会避免对任何单一解释做出承诺，因此通常会有初始的错误信号爆发。这些信号合理地解释了早期诱发响应的主要组成部分（通过使用头皮电极的脑电图记录测量），因为竞争的”信念”在系统中上下传播。这通常紧接着是对主导主题的快速收敛（如”自然场景中的动物”），随后协商进一步的细节（“几只老虎在大树的阴影下安静地坐着”）。这种设置因此有利于一种递归协商的”一览要义”模型，我们首先识别总体场景，然后是细节。这提供了一种”先见森林，后见树木”的方法（Friston, 2005；Hochstein & Ahissar, 2002）。正如Bar, Kassam, et al. (2006)所建议的，这些早期出现的要义元素可能基于快速处理的（低空间频率）线索被识别。这些粗糙的线索可能表明我们是否面对（例如）城市景观、自然场景或水下场景，它们也可能伴随着早期出现的情感要义——我们是否喜欢所看到的？参见Barrett & Bar, 2009以及后面5.10的讨论。

因此，想象一下你被绑架、蒙上眼睛，并被带到某个未知地点。当眼罩被摘除时，你大脑对场景的第一次预测尝试肯定会失败。但快速处理的低空间频率线索很快让预测大脑进入正确的总体范围。在这些早期出现的要义元素的框架下（在训练有素的系统中，这些元素甚至可能通过超快速的纯前馈扫描被识别，参见Potter等人，2014[26]），后续处理可以由与早期填充场景细节尝试的特定不匹配来指导。这些允许系统逐步调整其自上而下的预测，直到它确定一个连贯的总体解释，在时间和空间的多个尺度上确定细节。

然而，这并不意味着语境效应总是需要时间来出现并向下传播。因为在许多（实际上是大多数）现实生活情况下，当新的感官信息到达时，大量的语境信息已经到位。因此，一套恰当的先验通常已经处于活跃状态，准备立即影响处理过程而无需进一步延迟。

这很重要。在生态学正常情况下，大脑不是突然”开启”然后交付一些随机或意外的输入进行处理！所以通常在刺激呈现之前就已经有大量的自上而下影响（主动预测）。[27] 然而，无论在什么时间尺度上，终点（假设我们形成丰富的视觉感知）都是相同的。系统将稳定到一组状态，对场景的许多方面做出相互交织的预测，从总体主题一直到关于部分、颜色、纹理和方向的更加时空精确的信息。

[1.14 大脑中的预测处理]

第1.9节已经通过计算模拟的形式展示了大脑预测性处理的一些间接证据，这些模拟重现并解释了观察到的”非经典感受野效应”，如端点抑制(end-stopping)。另一个这样的效应（见Rao & Sejnowski, 2002）发生在定向刺激产生皮层细胞强烈反应时，但当周围区域被相同方向的刺激填充时，这种反应会被抑制，而当中央刺激的方向与周围区域的方向正交时，反应则会增强。Rao和Sejnowski（2002）提出，对这一结果的有力解释再次表明，观察到的神经反应在这里信号传递的是*错误*而不是良好猜测的内容。因此，当中央刺激从周围刺激高度可预测时，反应最小，而当它被周围环境积极反预测时，反应最大。同样，Jehee和Ballard（2009）提供了”双相反应动力学”的预测性处理解释，在这种情况下，驱动神经元的最佳刺激（如外侧膝状体LGN中的某些神经元）可以在短时间（20毫秒）内反转（例如，

从偏好明亮转为偏好黑暗）。这种转换再次被很好地解释为单元功能角色的反映，即作为错误或差异检测器而不是经典特征检测器。在这种情况下，预测编码策略得到了充分体现，因为：

低级视觉输入被输入与来自高级结构预测之间的差异所替代…高级感受野…代表对视觉世界的预测，而低级区域…信号传递预测与实际视觉输入之间的错误。（Jehee & Ballard, 2009，第1页）

更一般地，考虑”重复抑制”的情况。多项研究（最近的综述见Grill-Spector等，2006）表明，刺激诱发的神经活动会因刺激重复而减少。Summerfield等（2008）操纵了刺激重复的局部可能性，显示当重复不太可能/意外时，重复抑制效应本身会减少。受青睐的解释是（再次）重复通常减少反应，因为它增加了可预测性（第二个实例因第一个而变得更可能），从而减少了预测错误。因此，重复抑制也成为大脑预测性处理的直接效应，因此其严重程度可能会根据我们的局部感知期望而变化（正如Summerfield等发现的那样）。总的来说，预测编码故事为各种低级上下文效应提供了非常简洁统一的解释。

有一个新兴的支持性fMRI和EEG研究体系，可以追溯到Murray等（2002）的开创性fMRI研究，该研究也揭示了预测性处理故事所假设的那种关系。在这里，当高级区域确定了对视觉形状的解释时，V1中的活动被抑制，这与成功的高级预测被用来解释（消除）感觉数据是一致的。最近的研究证实了这种总体模式。Alink等（2010）使用表观运动错觉的变体发现对可预测刺激的反应减少，而den Ouden等（2010）使用在实验过程中快速操纵的任意偶然性报告了类似结果。为这些发现添砖加瓦的是，Kok, Brouwer等（2013）实验性地操纵了受试者对简单视觉刺激可能运动方向的期望。这些研究使用与运动点存在预测关系的听觉线索，显示受试者的隐式期望（通过听觉线索操纵）在感觉处理的最早阶段就影响了神经元活动。此外，这些效应超越了简单的加速或锐化反应，改变了实际主观感知的内容。作者们得出结论，完全符合（如他们所指出的）预测性处理，“我们的结果支持将感知作为概率推理过程的解释…其中自上而下和自下而上信息的整合发生在皮层层次结构的每一层”（Kok, Brouwer等，2013，第16283页）。

接下来，考虑P300，这是一种与意外刺激的发生相关的电生理反应。在最近的一项详细模型比较研究中，P300振幅的变化最好的解释(Kolossa et al., 2013)是自上而下期望与传入感觉证据之间残余误差的表达。相关地，预测处理为”失匹配负性”(MMN)提供了令人信服的解释——这是一种特征性的电生理大脑反应，也由意外(“异常”)刺激的出现，或在学习序列中某些预期刺激的完全缺失所引发。因此(引用Hughes et al., 2001; Joutsiniemi & Hari, 1989; Raij et al., 1997; Todorovic et al., 2011; Wacongne et al., 2011; 和Yabe et al., 1997)，最近有评论指出”听觉系统最显著的特性之一是它能够对缺失但预期的刺激产生诱发反应”(Wacongne et al., 2012, p. 3671)。基于缺失的反应(以及更普遍的异常反应)因此为预测处理风格的模式提供了进一步的证据，在这种模式中”听觉系统[获得]听觉输入规律性的内部模型，包括抽象的规律性，用于生成关于传入刺激的加权预测”(Wacongne et al., 2012, p. 3671)。这样的反应(绝不仅限于听觉模态)一旦我们将它们视为预测误差信号的瞬时爆发的指标，就会恰当地落入到位——这种信号发生作为传入信号被识别的正常过程的一部分(见Friston, 2005; Wacongne et al., 2012)。这里的PP解释与正常人类体验的显著特征直接接触。意外缺失的体验影响(比如熟悉序列中遗漏了一个音符)在感知上可能和包含意外音符一样引人注目和突出。这是一个否则令人困惑的效应，通过假设感知体验的构建涉及基于我们对可能发生事情的最佳模型的期望，得到了巧妙的解释。我们在下面的3.5节中回到这个话题。

在更加架构的层面上，基于生成模型的预测的核心作用既解释了反向神经连接的普遍性，也解释了前向和反向连接之间明显的功能差异——预测处理表明，这些差异反映了预测误差信号和概率预测的不同功能角色(关于这些功能不对称性的一些详细讨论，见Friston, 2002, 2003; 关于这个话题的一些最近实验工作，见Chen et al., 2009)。

1.15 沉默是金?

这个广泛领域的早期工作(比如上面描述的Rao & Ballard的开创性工作)遇到了一些困惑。这也许不足为奇，因为基本故事与从简单到复杂特征检测的前馈(即使是注意调节的)级联的更标准图景根本不同。这种困惑在Christoph Koch和

Tomaso Poggio的一篇评论中得到了著名的概括，副标题是”沉默是金”。这段话如此完美地表达了一些非常常见的第一印象，我希望读者能原谅这个长篇摘录：

在预测编码中，感觉神经元作为检测某些”触发”或”偏好”特征的常识观点被颠倒了，转而支持通过缺失放电活动来表征对象。这似乎与[表明神经元的数据]不符，这些神经元从V1延伸到下颞叶皮层，对越来越复杂的对象做出强烈的活动反应，包括以正确方式扭曲并从特定角度看到的个体面孔或回形针。

此外，所有来自人类的功能成像数据显示特定皮层区域对特定图像类别(如面孔或三维空间布局)做出反应，这又如何解释呢？这种活动可能主要由…细胞的放电主导，这些细胞积极表达错误信号，即该脑区期望的输入与实际图像之间的差异吗？(两个引用均来自Koch & Poggio, 1999, p. 10)

这里表达了两个主要担忧：首先，担心这些解释抛弃了表征而支持沉默，因为信号中被很好预测的元素被压制或”解释掉”了；其次，担心这些解释似乎与高级区域中活动标记的越来越复杂表征的强有力证据存在张力。

这两个担忧最终都不成立。要了解为什么不成立，回忆刚才概述的架构故事。我们看到，现在每一层都必须支持两种功能上不同的处理。为了简单起见，让我们跟随Friston (2005)，将此想象为每一层包含两种功能上不同的细胞或单元类型[30]：

—“表征单元”，编码该层当前的最佳假设(在其偏好的描述水平上)，并将该假设作为预测向下传递到下面的层。

—“误差单元”，当局部层内活动未被来自上层的传入自上而下预测充分解释时，向前传递激活。

这意味着随着沿层级向上移动，确实形成并在处理中使用了越来越复杂的表征。只是表征信息的*流动*（预测），至少在最纯粹的版本中，都是向下（和横向）的。然而，预测误差的向上流动本身就是一个敏感的工具，承载着关于非常具体的匹配失败的细粒度信息。这就是为什么它能够在更高区域诱导复杂假设（一致的表征集合），然后可以针对较低级别状态进行测试。因此，Koch和Poggio提出的两个早期担忧都站不住脚。存在”一路向上”的表征群体，更高级别的细胞仍然可以响应越来越复杂的对象和属性。但它们的活动是由错误信号的前向（和横向）流动以及它们选择的状态决定的。

然而，Koch和Poggio可能还在暗示另一种不同的担忧。这种担忧是，大脑作为”旨在实现沉默”（通过完美预测感觉输入）的基础”预测编码”形象，可能与动物生活本身的基本特征不协调！因为这种特征，肯定是移动和探索，永远寻求需要新一轮神经活动的新输入。简单地说，担忧是预测编码策略可能看起来像是寻找一个黑暗角落并待在那里的方法，正确预测不动性和黑暗，直到所有身体功能停止。

幸运的是（正如我们将在第8章和第9章中详细看到的），这里的威胁完全是表面的。因为在我们将要探索的解释中，感知的作用仅仅是驱动适应性有价值的行动。因此，我们许多时刻的预测实际上是对不安的感觉运动轨迹的预测（更多内容见第6章），它们的工作是让我们在世界中移动，以保持我们得到食物和温暖，并服务于我们的需求和项目。因此，对于像我们这样的生物来说，最能诱发预测误差的状态是所有活动停止、饥饿和口渴开始占主导地位的状态。在当前论述结束时，我们将看到预测误差最小化的基础策略，当它在活跃的、进化的、渴求信息的适应性智能体中展开时，如何本身强化所有我们认识和喜爱的不安、好玩、搜索和探索形式的行为。

1.16 期待面孔

然而，现在让我们回到Koch和Poggio提出的第二个（更具体的）担忧——担忧神经活动，随着处理的进行，看起来并不受表达错误的细胞放电所主导。再次考虑感知的标准模型，即通过越来越复杂的特征检测流处理的产物，使得较高级别的响应反映复杂、不变项目（如面孔、房屋等）的存在。我们现在可以澄清，预测处理故事建议的不是我

们放弃该模型，而是通过在每一层内添加专门用于编码和传输预测误差的细胞来丰富它。因此，每个级别的一些细胞响应身体和世界的状态，而其他细胞则记录相对于这些状态预测的错误：从上一级横向和向下流动的预测。这是正确的吗?

这里的证据才刚刚出现，但似乎符合”预测处理”的特征。因此，考虑已经确立的发现（Kanwisher等，1997），当显示面孔而不是（比如）房屋时，梭状面孔区域(fusiform face area, FFA)的活动增加。批评者可能会说，这最好通过简单假设FFA中的神经元已经学会成为面孔的活跃复杂特征检测器来解释？然而，考虑到PP故事允许FFA确实可能包含专门表征面孔的单元，以及专门检测错误（到达FFA的自上而下预测与自下而上信号之间的不匹配）的单元，这立即变得不再简单。因此，区别在于，如果预测编码故事是正确的，FFA*还应该*包含错误单元，这些单元基于横向和自上而下的预测编码与预期（面孔）活动的不匹配。FFA中表征单元和错误单元的预测存在为一些有说服力的实证测试提供了很好的机会。

Egner等人（2010年）比较了简单特征检测模型（有注意力和无注意力）和预测处理模型在FFA记录反应中的表现。简单特征检测模型预测，正如Koch和Poggio所建议的，FFA反应应该简单地与呈现图像中面孔的存在成比例。然而，预测处理模型预测的是相当复杂的事情。它预测FFA反应应该”反映与预测（‘面孔期望’）和预测错误（‘面孔惊讶’）相关的活动总和”（Egner等人2010，第1601页）。也就是说，它预测从FFA记录的（低时间分辨率）fMRI信号应该反映两种假定类型细胞的活动：那些专门负责预测的（‘面孔期望’）和那些专门负责检测预测错误的（‘面孔惊讶’）。然后通过从FFA区域收集fMRI数据来测试这一点，同时独立地改变呈现的特征（面孔vs房屋）并操纵受试者无意识的面孔期望程度（低、中、高），从而改变他们适当的”面孔惊讶”程度。为了做到这一点，实验者概率性地将面孔/房屋的呈现与250毫秒的先行彩色框架提示配对，该提示给出25%（低）、50%（中）或75%（高）的概率表示下一张图像是面孔。

结果很明确。FFA活动显示了刺激和面孔期望之间的强相互作用。FFA反应仅在低面孔期望条件下表现出最大分化。确实，令人惊讶的是，在高面孔期望条件下，给定*任一*刺激（面孔或房屋）的FFA活动*都无法区分*。因此，在很真实的意义上，FFA可能（如果首先使用预测处理范式进行研究）会被称为”面孔期望区域”。

作者得出结论，与任何简单的特征检测模型相反，“[FFA]反应似乎是由特征期望和惊讶决定的，而不是由刺激特征本身决定的”（Egner等人，2010，第16601页）。作者还通过进一步的模型比较控制了注意力效应的可能作用。但无论如何，这些都不可能做出太大贡献，因为是面孔惊讶，而不是面孔期望，占了BOLD（fMRI）[31]信号的更大部分。事实上，最佳拟合的预测处理模型使用了一个权重，其中面孔惊讶（错误）单元对BOLD信号的贡献大约是面孔期望（表示）单元的两倍[32]，这表明通常使用fMRI记录的大部分活动可能是在发出预测错误信号，而不是检测到的特征。

这是一个重要的结果。用作者自己的话说：

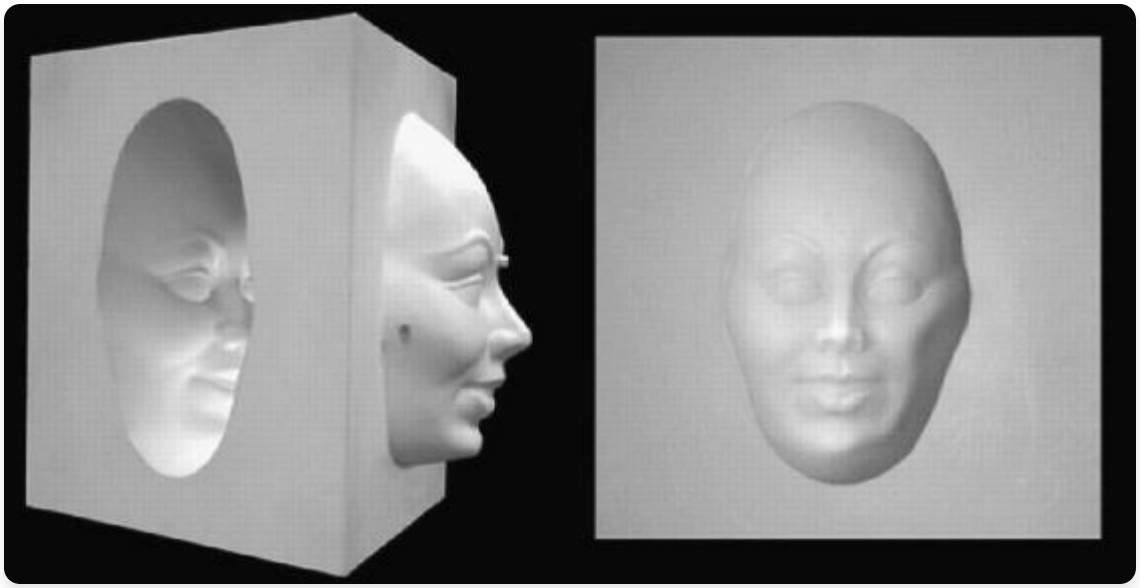
据我们所知，目前的研究是第一项正式和明确证明视觉皮层中的群体反应实际上更好地被描述为特征期望和惊讶反应的总和，而不是自下而上的特征检测（有或没有注意力）的研究。（Egner等人，2010，第16607页）

[1.17 当预测误导时]

当然，所有这些高效的基于预测的反应都有缺点，这在熟悉的视觉错觉中得到了很好的说明，比如空心面孔错觉。在这里，3D面具的凹陷内表面在某些条件下看起来像正常的面孔：凸出的，鼻子向外伸展。为了更好地了解这看起来如何，可以尝试在http://www.michaelbach.de/ot/fcs_hollow-face/的简短评论中嵌入的视频片段。

更好的是，使用真正的三维面具亲自体验这种错觉，就像你在万圣节使用的那种。拿起面具并将其反转，这样你就在看空心的内侧而不是凸出的（面孔形状的）一侧。如果观看距离正确（不要太近：至少需要大约3英尺远）并且面

具从后面轻柔地照明，面具看起来就不是空心的。你会”看到”鼻子向外伸出，而实际上，你在看的是面部印象的凹陷反面。图1.5显示了在这种条件下旋转面具的外观。



图像

图1.5 **空心面具错觉**

最左边和最右边的图像显示了在支架上旋转的面具的空心凹陷侧。当从几英尺外观看并从后面照明时，凹陷侧看起来是凸出的。这证明了自上而下预测（我们”期望”面孔是凸出的）对感知体验的影响力。

来源：Gregory (2001)，经英国皇家学会许可。

在神经典型受试者中，空心面具错觉是强大而持久的。然而，在精神分裂症受试者中，这种效应被显著减少——预测处理装置的特定干扰也可能（见第7章）有助于解释这种效应。空心面具错觉首先被神经科学家理查德·格雷戈里（见，例如，Gregory，1980）用来说明”自上而下”、知识驱动的对感知影响的力量。这种效应直接来自前面章节中讨论的基于预测的学习和处理原则的操作。我们在日常生活中与无数凸面面孔的统计上显著的经验安装了对凸性的深层神经”期望”：这种期望在这里胜过了许多其他视觉线索，这些线索本应告诉我们我们看到的是一个凹陷面具。

你可能合理地怀疑，空心面具错觉虽然令人印象深刻，但实际上只是某种心理学奇异现象。诚然，我们关于人脸可能是凸起的神经预测似乎特别强烈和有力。但如果预测处理方法是正确的，这种一般策略实际上遍及人类感知。像我们这样的大脑不断尝试使用它们已经知道的东西来预测当前的感官信号，使用传入信号来选择和约束这些预测，有时使用先验知识来”胜过”传入感官信号本身的某些方面。这种胜过是有良好适应意义的，因为使用你所知道的来超越传入信号似乎在说的某些内容的能力，在感官数据嘈杂、模糊或不完整时可能极其有益——这些情况实际上在日常生活中几乎是常态。

这的一个有趣结果是，正如1.12中提到的，许多视觉错觉可能最好被理解为”最优感知”。换句话说，考虑到我们居住的世界的结构和统计特性，对世界状态的最优估计（代表对传入信号的最佳可能理解的估计，基于系统已经知道的内容）将是在某些场合会出错的估计。因此，一些局部失败只是我们在一个被模糊性和噪声笼罩的世界中，大部分时间能够获得正确结果所付出的代价。

1.18 颠倒的心智

预测处理将传统的感知图景颠倒过来。根据那个曾经标准的图景（Marr, 1982），感知处理由通过各种感官受体从世界转导的信息的前向流动主导。传统的感知神经科学也随之而来，视觉皮层（最被研究的例子）被视为由底向上驱动的神经特征检测器的层次结构。这是将感知大脑视为被动的、刺激驱动的观点，从感官接收能量输入，并通过某种逐步构建的方式将它们转化为连贯的感知，沿途以某种乐高积木的方式累积结构和复杂性。这种观点可能与过去几十年神经科学和计算研究中追求的日益”主动”的观点形成对比，包括最近关于内在神经活动的工作爆发——即使在没有正在进行的任务特定刺激的情况下也发生的自发的、相关的神经元激活的不断嗡嗡声。所有这些都表明，大脑活动的大部分既是持续的，也是内源性产生的。

预测处理很可能代表了从被动的、输入主导的神经处理流观点中撤退的最后一步。根据这类新兴模型，自然智能系统不会被动地等待感官刺激。相反，它们不断活跃，试图在感官刺激流到达之前预测（并主动引出，见第二部分）它们。在”输入”出现在场景中之前，这些主动的认知系统已经忙于预测其最可能的形状和含义。这样的系统已经（几乎不断地）准备好行动，它们需要处理的只是感知到的与预测状态的偏差。正是这些与预测状态的计算偏差（“预测误差”）承担了大部分信息处理负担，告诉我们在密集的感官轰炸中什么是显著的和新闻价值的。

正如我们将在第二部分中看到的，行动本身需要重新构想。行动不再是对输入的”反应”，而是选择下一个输入的巧妙而有效的方式，驱动滚动循环。这些超活跃系统不断预测它们自己即将到来的状态，并主动移动以实现其中一些状态。因此，我们行动是为了产生不断演化的感官信息流，这些信息流使我们保持可行性并服务于我们日益深奥的目标。在纳入行动后，预测处理实现了传统（自下而上、前向流动）模式的全面逆转。对持续神经响应的最大贡献是向下流动的神经预测的不断预期嗡嗡声，它在循环因果流中驱动感知和行动。传入的感官信息只是扰动那些不安的主动海洋的另一个因素。

作为永远活跃的预测引擎，这类大脑根本不是在”处理输入”的业务中。相反，它们在预测其输入的业务中。这种主动神经策略使我们保持行动准备状态，并且（正如我们稍后将看到的）允许移动的、具身的智能体干预世界，带来使它们保持可行性和满足的感官流类型。

如果这些故事是正确的，那么被动前向流动模型的几乎每个方面都是错误的。我们不是认知上的沙发土豆，无所事事地等待下一个”输入”，而是主动的预测食者——自然界自己的猜测机器，通过冲浪传入的感官刺激波浪，永远试图保持领先一步。

2

调节音量（噪声、信号、注意力）

2.1 信号识别

如果我们去寻找，大多数人都能在云朵中找到隐藏的变化面孔。我们可以在图案壁纸中看到昆虫的形状，或者在地毯的彩色漩涡中看到蛇的身影。这种效应不需要摄入改变心智的物质。像我们这样的心智本身就是自我改变的专家。当我们在杂乱的桌子上寻找车钥匙时，我们会以某种方式改变感知处理过程，帮助从其他物品中分离出目标物品。实际上，发现（真实的）车钥匙和”发现”（不存在的）面孔、蛇和昆虫可能没有太大区别，至少在底层处理形式方面是这样。这种发现反映了我们不仅能够改变行动例程（例如，我们的视觉扫描路径），还能够修改自己感知处理的细节，以便更好地从噪音中提取信号。这种修改在调整支撑我们与世界接触的内置概率预测机器（包括长期和短期调整）中起着真正重要的作用。本章探讨了这种在线修改的空间和性质，讨论了它们与注意力和期望等熟悉概念的关系，并展示了一种可能的机制（预测误差的”精确度加权”），这种机制可能涉及广泛的信号增强效应。

听到Bing

Tom Stafford和Matt Webb那本极其引人入胜的书《Mind Hacks》中的第48个技巧叫做”在确定性边缘检测声音”。基于Merckelbach和van de Ven（2001）之前的实验工作，这个技巧邀请读者首先听一个30秒的音频文件。读者被告知音频文件包含Bing Crosby的”White Christmas”的隐藏片段，但这个片段非常微弱，可能在音频文件的第一、第二或第三个十秒段中开始。勇敢的读者可能想在继续之前尝试一下，点击：http://mindhacks.com/book/links/ 的第48个技巧。

Merckelbach和van de Ven（2001）用本科生进行了这个实验，发现几乎三分之一的学生报告检测到了歌曲的开始。实际上，正如你现在可能已经猜到的，噪音中任何地方都没有隐藏White Christmas。一些人”检测”熟悉歌曲的能力只是一般感知搜索和感知意识中核心能力的表达（在这种情况下，是一种过度延伸）：忽略信号某些方面并将其视为”噪音”，同时强调其他方面（从而将其视为”信号”）的能力。这种能力在强烈期望微弱”难以检测”的熟悉歌曲片段的影响下部署，使许多完全正常的受试者享受到实际上是听觉幻觉的体验。事实证明，这种效应甚至可以通过压力和咖啡因的组合得到放大（Crowe et al., 2011）。

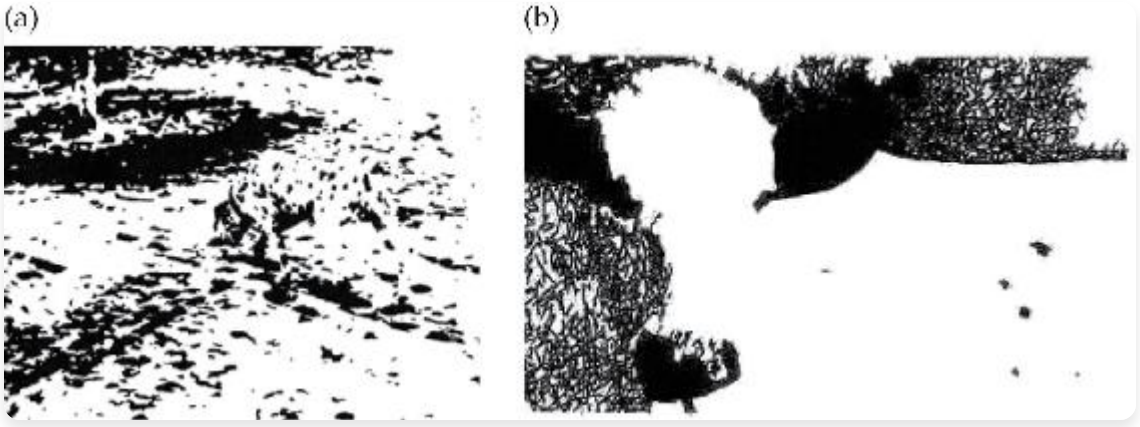
现在考虑第二种情况：正弦波语音。正弦波语音（Remez et al., 1981; Remez & Rubin, 1984）是录制语音的降级复制品，剥离了大部分正常的语音属性和声学特征。正弦波复制品只保留了一种骨架轮廓，其中语音信号动态变化的核心（相当粗糙的）模式被编码为一组纯音调哨声。你可以通过点击：http://www.mrc-cbu.cam.ac.uk/people/matt.davis/sine-wave-speech/ 的第一个扬声器图标听到一个例子。

你很可能听不懂你所听到的：对我来说，它听起来像BBC电台音响工作室在1960年代早期开创的那种科幻哔哔声。其他人听到的像是英国儿童节目The Clangers中月亮鼠角色那种难以理解的抑扬顿挫的哨声。但现在点击下一个扬声器，听原始句子，然后重新听正弦波复制品。这一次，你的体验世界发生了改变。它变得（更多内容在后续章节

中）有意义：一个清晰可懂的语音世界。要获得这样的精彩演示选择，请尝试：http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/SWS/。

记住要先点击SWS（正弦波语音）版本。一旦你知道句子是什么，就几乎不可能以原来的方式”重新听到”它。一个恰当的比较是听你理解的语言和你不理解的语言的语音。在前一种情况下，几乎不可能仅仅将语音声音听作声音。接触原始（非正弦波）口语句子以类似的方式帮助你做准备。随着时间的推移，你甚至可能在正弦波语音感知方面变得足够专业，无需事先接触特定的声学正常句子就能成功。在那时，你已经成为具有更一般技能的专家（正弦波语音的”本地听者”）。

Davis and Johnsrude (2007)将正弦波语音的感知描述为*自上而下影响*对感觉处理这一更普遍现象的一个实例。如果第1章中描述的解释是正确的，这种影响根植于概率生成模型的创建和部署，这些模型忙于预测感觉输入的流动。我们在所有感觉模式内部和跨感觉模式都能看到这种影响。图2.1(a)中展示了一个经典的例子。乍一看，大多数人只能看到光影图案。但一旦你发现了那只有斑点、带阴影的达尔马提亚犬，这种知识就会改变你余生对这幅图片的看法。再看一个不太熟悉的例子，请看图2.1(b)。在这些情况下，[1]我们对世界的知识（我们的”先验信念”，由大脑控制的生成模型实现）在感知的构建中起着重要作用。（值得重申的是，“信念”一词在这个文献中被广泛使用，涵盖指导感知和行动的生成模型的任何内容。这些信念不需要被反思性主体有意识地获取。实际上，在大多数情况下，它们将包含各种亚个人状态，其最佳表达是概率性的而非命题性的[2]）。



image

图2.1 隐藏图形

a. 隐藏在黑白噪声中的是一幅图像（一旦你发现就很清楚），是一只达尔马提亚犬。提示：头部在图像中心附近，正在检查地面。

b. 同一现象的一个不太知名的例子。这次是一头牛。提示：牛有一个大头；它面向你，鼻子在图片底部，两只黑耳朵在左上半部分。

来源：约翰·麦克罗恩的隐藏牛，CC-BY-SA-3.0。http://creativecommons.org/licenses/by-sa/3.0，通过维基媒体共享资源。

文献中已经积累了其他不太明显的跨模式影响例子。一个特别引人注目的发现是，葡萄酒的感知颜色会对人们（包括葡萄酒专家）如何描述该酒的味道产生很大影响（见Morrot et al., 2001；Parr et al., 2003；以及Shankar et al., 2010——后者有着相当精彩的标题”葡萄期望”）。在这些实验中，被人工着色为红色的白葡萄酒，即使是专家也会用红酒描述词来形容，如梅子、巧克力和烟草。先验期望的影响并不止于此。牡蛎似乎在伴随着海洋声音（即使在内陆餐厅深处）的情况下食用时味道更好（Spence & Shankar, 2010）。

预测性处理(predictive processing)提供了一个强大的框架，用于处理和理解基于知识和情境效应对感知推理的整个体系，因为它使我们对世界的了解（有意识的和更常见的无意识的）成为感知体验构建本身的主要参与者。我们将在第7章中回到有关意识体验构建的问题。目前，我想关注一些更抽象但对所提供的解释相当基础的东西。这个东西对感知成功（如发现达尔马提亚犬或听到正弦波语音）、感知游戏（如在云中找到面孔形状）和一些感知失败（如幻听”白色圣诞节”的声音）都至关重要。它是通过形成和部署我们自己感知不确定性的聚焦和细粒度估计来灵活地从噪声中提取信号的能力。[3]这种能力（本章其余部分的重点）是预测性处理(PP)注意力处理的核心，在解释与世界的正常和异常接触形式方面起着重要作用。

自上而下和自下而上之间的微妙舞蹈

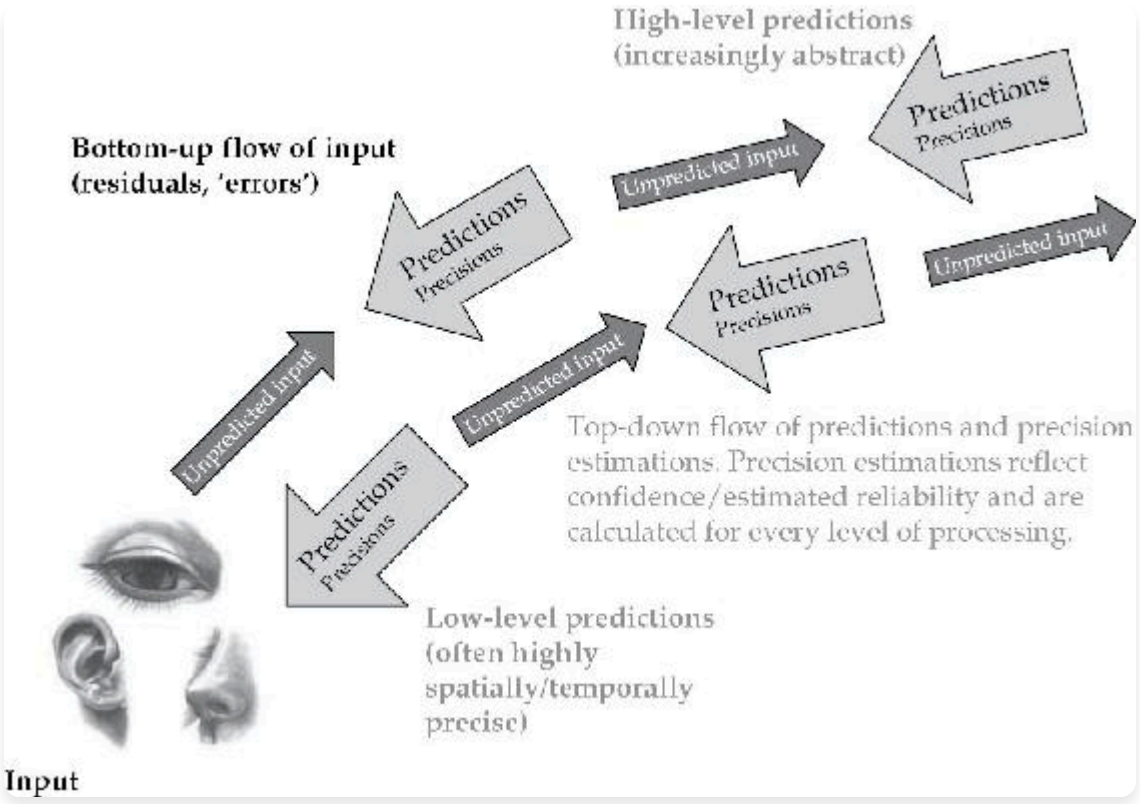
我们在日常生活中面临的感知问题在对我们的要求上差异很大。对于许多任务，最好部署大量先验知识，使用这些知识驱动复杂的主动注视固定模式，而对于其他任务，最好退后让世界尽可能多地起主导作用。哪种策略（更多输入驱动或更多期望驱动）最好也受到众多情境效应的影响。在浓雾中沿着非常熟悉的道路行驶时，有时让详细的自上而下知识发挥重要作用是明智的。在不熟悉的蜿蜒山路上快速行驶时，我们需要让感觉输入起主导作用。概率预测机器如何应对?

PP 建议，大脑通过持续估计和重新估计自身的感知不确定性来应对这种挑战。在 PP 框架内，这些感知不确定性估计会调节感知预测误差的影响。这本质上就是预测处理的注意力模型。因此，注意力被理解为通过考虑其所谓的”精确度”来可变地平衡自上而下和自下而上影响之间强力相互作用的手段，其中精确度是对其估计确定性或可靠性的度量（对统计学专家来说，是方差的倒数）。这是通过相应地改变误差单元的权重（增益或”音量”，使用一个常见的类比）来实现的。其结果是”控制不同层次先验期望的相对影响”（Friston, 2009, p. 299）。更高的精确度意味着更少的不确定性，并反映在相关误差单元的更高增益上（见 Friston, 2005, 2010; Friston et al., 2009）。如果这是正确的，注意力仅仅是某些误差单元响应被赋予增加权重的手段，从而更容易驱动反应、学习和（我们稍后将看到的）行动。更一般地说，这意味着自上而下和自下而上影响的精确组合不是静态或固定的。相反，给予感知预测误差的权重是根据信号被认为的可靠程度（噪声程度、确定性或不确定性程度）而变化的。

我们可以用之前的例子来说明这一点。在雾中，视觉输入被估计为对远程领域状态提供嘈杂且不可靠的指导。在其他条件相等的情况下，视觉输入在晴朗的日子里应该提供更好的信号，这样任何残余误差都应该被认真对待。但这种策略显然需要比建议的更精细的调整。因此，假设雾（如经常发生的那样）从视觉场景的一个小片区暂时散去。那么我们应该被驱动优先从那个较小的区域采样，因为那现在是高精度预测误差的来源。这是一个复杂的问题，因为该小区域存在的证据（就在那里！）仅来自（最初低权重的）感知输入本身。这里没有致命的问题，但这个案例值得仔细描述。首先，现在出现了一些低权重的惊讶，相对于我对视觉情况的当前最佳理解（类似于”在均匀的大雾中”）。输入的各个方面（在清晰区域中）没有按照那种理解（那个模型）预测的方式展开。然而，我的雾模型包括关于偶尔出现清晰片区的一般期望。在这种条件下，我可以通过切换到”雾加清晰片区”模型来进一步减少整体预测误差。这个模型融入了一套新的精确度预测，允许我信任为清晰区域（仅此）计算的细粒度预测误差。那个小区域现在是视觉系统可以信任以招募清晰可靠知觉的高精度预测误差的估计来源。来自清晰区域的高精度预测误差可能随后迅速保证招募一个能够描述当地环境某些显著方面的新模型（小心那台拖拉机！）。

这就是 PP 分配给感知注意力的角色的缩影：“注意力可以被视为对相对于模型预测具有高精度（信噪比）的感知数据的选择性采样”（Feldman & Friston, 2010, p. 17）。这意味着我们不断参与预测精确度的尝试，即预测我们自己感知预测误差的上下文变化可靠性，并且我们相应地探测世界。这种”基于预测精确度”的探测和采样也支撑着（我们将在第二部分中看到）PP 对总体运动活动的解释。目前要注意的是，在这个嘈杂和模糊的世界中，我们需要知道何时何地认真对待感知预测误差，以及（更一般地）如何最好地平衡自上而下的期望和自下而上的感知输入。这意味着要知道何时、何地以及在多大程度上信任特定的预测误差信号来选择和细化指导我们行为的模型。

一个重要的结果是，使人类知觉成为可能的知识不仅涉及（行动相关的——稍后详述）远程世界的分层因果结构，还涉及我们自己与那个世界的感知接触的性质和上下文变化的可靠性。这样的知识必须构成整体生成模型的组成部分。因为该模型必须能够预测冲击感知信号的形状和多尺度动态，以及信号本身的上下文可变可靠性（见图2.2）。现在，“注意力”这一熟悉概念自然而然地被定位为命名精确度预测调整和影响感知采样的各种方式，允许我们（当事情按应有的方式运作时）被信号驱动而忽略噪声。通过主动采样我们期望（相对于某些任务）最佳信噪比的地方，我们确保我们感知和行动所依据的信息适合其目的。



image

图2.2 基本预测处理模式，这次加上了精确度权重

这是在第1章中显示的 PP 架构的相同高度简化视图，但加入了精确度权重。现在，选定预测误差信号的影响通过对其当前可靠性和显著性的变化估计进行调节

来源: 改编自 Lupyan & Clark, 2014。

2.4 注意力、有偏竞争和信号增强

如果这些理论是正确的，注意力指的是有机体增加那些被估计为相对于某些当前任务、威胁或机会提供最可靠感觉信息的预测误差单元的增益（权重，因此是向前流动的影响）的手段或过程。更正式地说，建议是”注意力是在层次推理过程中优化突触增益以表示感觉信息（预测误差）精度的过程”（Feldman & Friston, 2010，第2页）。因此，总体思路是神经元激活模式（在所谓的”表示单元”中）编码关于世界任务相关状态的系统性猜测，而相关误差单元[4]上的增益变化（即权重或”音量”的变化）反映了大脑对自上而下”猜测”和自下而上感觉信息相对精度的最佳估计。因此，精度加权提供了系统对感觉信息本身可信度的最佳估计。这意味着”自上而下的预测不仅涉及较低层次表示的内容，还涉及我们[大脑]对这些表示的信心”（Friston, 2012，第238页）。据认为，这些自上而下的精度估计改变了预测误差单元上的突触后增益（通常被认定为浅层锥体细胞；例如，参见Mumford, 1992；Friston, 2008）。因此我们读到：

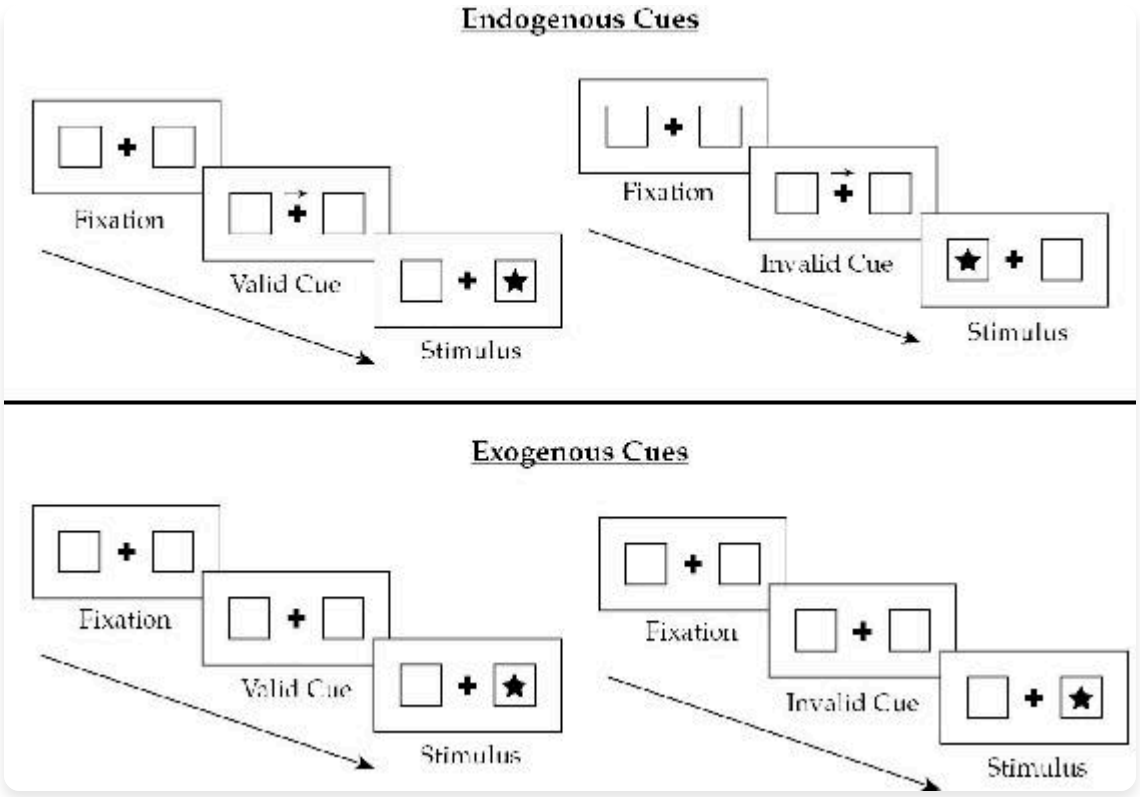
从生理学角度来说，精度对应于报告预测误差的细胞的突触后增益或敏感性（目前认为是发送前向型外在传出的大主细胞，如皮层中的浅层锥体细胞）。（Friston, Bastos, et al., 2015，第1页）

总之，这些增益变化[5]追踪所选预测误差的估计可靠性（统计学上，即逆方差）。这些误差编码所有仍需解释的感觉信息（或尚未被用于控制行动的信息）。因此，精度估计传递消息信号的可靠性，重复第1章中使用的便利隐喻。

估计精度并相应地改变预测误差上的增益带来了即时且极其重要的好处。它允许PP方法流畅地结合表面上矛盾的效应（见1.11）：信号抑制和信号增强。信号抑制当然是预测编码数据压缩方法的熟悉效应。被获胜的高层模型很好预测的低层活动被抑制或”解释掉”（因为那里没有新闻），因此没有信号通过系统向前传播。表面上矛盾的效应是基于显著性的信号增强。这里观察到的效应包括促进（诱发反应的加速；参见Henson, 2003）和锐化（其中一些细胞停止活跃，允许其他细胞主导反应；参见Desimone, 1996）。精度加权允许PP以非常灵活的方式结合这些效应，因为突触后增益的增加实现了促进效应，然后”提升预测误差，这些误差为关于感觉输入原因的最佳假设提供信息（Gregory, 1980），同时抑制替代假设；即*它锐化神经元表示*”（Friston, 2012a，第238页，原文斜体）。Kok, Jehee, and de Lange (2012)发现了这样的锐化效应，揭示了早期感觉反应某些方面基于期望的增强，同时（正如PP模型所建议的）伴随着神经活动的整体减少。这种期望诱导的锐化被证明在行为上是有效的，在涉及检测刺激方向细微差异的简单任务中产生了更好的表现。

这种锐化是”有偏竞争”模型的熟悉支柱（Desimone & Duncan, 1995）。这些模型假设——正如名称所暗示的——存在一种争夺上游神经元表示的竞争，其中只有”获胜的”下层细胞（具有小感受野）被允许驱动上层细胞（具有大感受野）。有偏竞争模型建议，注意力应该被认定为这种竞争过程：一种其结果由任务性质和竞争表示资源的刺激特性共同决定的竞争。许多观察到的效应（例如，Reynolds et al., 1999；Beck & Kastner, 2005，2008）明确符合有偏竞争模型。例如，一些电生理学（ERP）成分（参见Bowman et al., 2013）在目标反复出现在同一位置时会增加。此外（再次参见Bowman et al., 2013），在视觉搜索实验中，干扰项尽管罕见，但产生的诱发反应很少，而预先描述的、频繁出现的目标则产生很大的诱发反应。这些效应能否直接通过注意力调节的残余误差精度加权来解释?

Kok et al. (2012)的一项fMRI研究为此类效应的预测处理模型提供了优雅的支持，它表明这些正是精度加权预测误差模型所建议的预测与注意力之间的相互作用类型。特别是，Kok等人表明，未被注意到且与任务无关的预测刺激会导致早期视觉皮层活动的减少（对预测的”静默”，正如简单预测编码所要求的），但”当刺激被注意到且与任务相关时，这种模式会发生逆转”（Kok et al., 2012，第2页）。该研究通过使用独立的预测和空间线索来操控空间注意力和预测（更多细节请参见Kok等人的原始论文），发现注意力逆转了预测对感觉信号的静默效应，正如精度加权账户所指定的那样。因此，当注意力和预测一致时（当独立的注意力线索选择了预测刺激实际出现的空间半场时），注意力增强了V1、V2和V3中对预测刺激相对于未预测刺激的神经反应。当它们不一致时（即预测刺激没有出现在被注意的空间位置），没有发生增强，V1中对预测刺激的反应减少了。此外，无论未预测刺激出现在被注意的一侧还是未被注意的一侧，对它们的反应都是相同的。最后，对于在被注意半场中意外缺失的刺激，V1、V2和V3中出现了大的反应。作者认为，整个模式最好用注意力调节的预测误差精度加权来解释，其中注意力增加了选定预测误差单元的下游影响。因此，注意力和期望看起来像是在PP所建议的推理级联中作为不同元素运作的。注意力增强（增加增益）与选定预测误差相关的神经反应，而期望则抑制那些符合期望的神经反应。


Endogenous Cues
Fixation
Valid Cue
Stimulus
Fixation
Invalid Cue
Stimulus
Exogenous Cues
Fixation
Valid Cue
Stimulus
Fixation
Invalid Cue
Stimulus

image

图2.3 Posner范式

来源：根据知识共享署名3.0许可证授权 en.wikipedia.org/wiki/File:Posner_Paradigm_Figure.png。

PP账户包含各种形式注意力增强的能力也通过Posner范式的计算机模拟得到了证明（Posner, 1980）。在Posner范式中（见图2.3），受试者注视一个中心点（因此实验探测所谓的”隐蔽注意力”），并呈现一个视觉线索，该线索通常（但不总是）指示即将出现的目标刺激的位置。例如，该线索在80%的试验中可能是有效的。具有有效线索的试验被称为”一致试验”，而不是这种情况的试验（线索无效，因此不能正确预测刺激）被称为”不一致试验”。因此，该范式操控我们的上下文期望，因为线索创造了一个上下文，在其中刺激在线索位置出现变得更加可能。主要发现，不出所料，是促进作用：有效线索加速了对目标刺激的检测，而在不一致试验中呈现的目标被感知得更慢，且信心较低。Feldman and Friston (2010)提出了一个详细的、基于模拟的模型，其中精度调节的预测误差被用来优化感知推理，以重现在人类受试者中发现的ERP和心理物理反应。有效线索通过增加与被提示空间位置相关的预测误差单元的增益来建立有时被称为”注意力定势”的状态。这然后构成了对来自该空间区域信息的良好信噪比的系统性”期望”，从而加速了一旦目标出现后招募正确假设（大致是”目标在那里”）的过程，因此重现了促进效应。无效提示的目标产生低权重的早期预测误差，因此需要显著更长的时间来招募正确的假设（“目标在那边”），并以较低的信心被感知。

这种关于注意力的一般观点在现象学上是令人信服的。试着长时间专注于这一页上的一个单词。对我来说，这种体验最初似乎是局部清晰度的增加，紧接着是在保持警觉的同时清晰度衰减的状态。在那时，有一种驱动行动的倾向，也许使用隐蔽注意力的转移或微眼跳来进一步探索被注视的单词。所有这些持续得越久*而不*出现任何新的、不同的或更清晰的信息，维持注意过程就变得越困难。[6]因此，注意力在体验上表现为与对新的和更好信息的期望和搜索密切相关。

2.5 感觉整合和耦合

预测误差的精度加权被证明是一个非常多用途的工具，并且在我们的故事展开过程中将发挥多种作用。目前我仅仅提到两个这样的额外作用，不作太多阐述。第一个涉及感觉整合(sensory integration)。通常，在面对世界时，大脑会从多种来源接收感觉信号。例如，我们可能看到并听到一辆驶近的汽车。在这种情况下，两个感觉输入源需要在确定我们对显著（实际上，往往与生存相关！）环境状态（如汽车的位置和接近速度）的感知体验中发挥微妙平衡的作用。对两个感觉信号相对精度的自动估计使大脑能够整合两个信息源，在给定的更大背景下以最佳方式使用每个信息源。这种整合既依赖于关于典型汽车视觉和听觉的特定（亚个人层面的）期望，也依赖于更一般的期望，比如这样的期望（一个系统性的”超先验(hyperprior)“）：每当对信号源空间位置的听觉和视觉估计相当接近时，最佳的总体假设是存在一个单一来源——在这种情况下，是一辆快速移动的汽车。这种超先验也可能产生误导，如腹语师木偶的声音投射所证明的。但在生态学上的核心情况下，它们使多种感觉数据源的最优组合成为可能。因此，在假设选择过程和各种感觉输入源的精度加权之间存在有力的相互作用。

估计精度还有助于确定神经区域之间信息的瞬时流动（从而有助于确定”有效连接性(effective connectivity)“的变化模式；见Friston, 1995, 2011c）。当我们后来考虑对可变神经（实际上还有额外神经，见第三部分）资源混合的上下文敏感和任务特定的招募时，第二个作用将被证明是重要的。因此，举一个非常简单的例子，有时最好允许视觉信息在选择行为反应时占主导地位（例如，在已知听觉干扰存在的情况下），这可以通过给听觉预测误差分配低精度而给视觉预测误差分配高精度来实现。[7] 例如，den Ouden et al. (2010) 提供了对皮层区域间耦合强度（即影响）变化的解释，描述了可变精度加权的预测误差作为根据（情境化的）任务需求”即时”控制这种耦合的关键工具。同样

的广泛机制也可能在能够确定在线反应的多个系统（如前额皮层和背外侧纹状体系统）之间进行仲裁。正是在这种意义上，Daw, Niv, and Dayan (2005, p. 1704) 将这些系统描述为受到”根据不确定性的贝叶斯仲裁原则”的约束，使得目前估计能提供最准确预测的子系统得以驱动行为和选择。当我们考虑（在第三部分）预测处理框架与我们称之为”心智”的整个具身的、文化化的、环境嵌入的认知架构的形状和性质之间的可能关系时，这些原则将变得重要。

[2.6 行动的初探]

然而，精度（和精度期望）在预测处理故事中的完整作用如果不至少预览对行动的处理就无法被充分理解。这并不奇怪，因为（正如我们将看到的）PP对感知和行动之间复杂循环相互作用的认知中心性提出了强有力的建议。事实上，这种相互作用如此复杂、核心和循环，以至于感知将会显现（第二部分）为与行动不可分离，感觉和运动处理之间的理论划分本身也将受到质疑。

所有这些都在我们面前。就目前而言，介绍那个更丰富、更面向行动的故事的一个核心要素就足够了。这个要素（在上面关于选择行动的评论中已经暗示过）涉及行动作为*基于精度期望的感觉采样*工具的作用。这很难解析，有点拗口，但这个想法既简单又惊人地强大。考虑这样一种情况：有两个模型在激烈竞争来解释感觉信号。一个模型比另一个更多地减少了预测误差，但它减少的预测误差被估计为不可靠。另一个模型，虽然它减少的绝对误差较少，但减少的误差被估计为高度可靠。在这种情况下，“最佳选择”通常（尽管见Hohwy, 2012的一些重要警告）是支持减少更可靠误差信号的模型。如果两个竞争模型简单地说是”门廊里的猫”和”门廊里的窃贼”，[8]这可能具有一些实际重要性。

但是我们如何确定信号的可靠性呢？正是在这里，行动（一种特殊的行动）发挥了关键的认知作用。我粗略地描述为”门廊里的窃贼”的生成模型的一部分，包括了关于*采样*环境的最佳方式的期望，以便产生关于这种可能性的可靠信息。例如，它包括关于扫描场景的最佳方式的期望，先注视一个位置，然后注视另一个位置，以便*减少关于该假设的不确定性*。假设这个假设是正确的（那里确实有一个窃贼），这个过程将产生一系列精确的预测误差，既完善又确认我的可怕怀疑，可能会显露出金属的闪光（手电筒？枪？）和深色高领毛衣的轮廓。这个过程会迭代，因为现在需要评估”手电筒”和”枪”的假设。在那里，我的生成模型也包括关于参与（采样）感官场景以减少不确定性的最佳方式的期望。这些期望以一种与PP感知论述完全连续的方式（正如我们在第二部分中将看到的）调动行动。感知和行动在这里形成了一个良性的、自我推进的循环，其中行动提供可靠的信号，这些信号招募感知，这些感知既决定又在行动中得到确认（或否认）。

注视分配：做自然而然的事

这里还有一个更大的故事，涉及在执行自然任务(natural tasks)期间注意力的分配方式。正如我使用这个术语，自然任务几乎是我们在日常活动过程中可能执行的任何熟练任务。因此，自然任务包括烧水、遛狗、购物、跑步和吃午餐。这些任务的重要之处（以及使它们区别于许多基于实验室的实验范式的地方）在于，它们提供了我们在那些特定情况下通过学习期待的完整、丰富的感官线索集合。这很重要，因为它允许任务特定知识在驱动（例如）主动视觉扫视(saccades)方面发挥更重要的作用，在这种扫视中，我们的眼睛预期性地移动到下一步将找到相关信息的地方，并为许多其他形式的主动干预开辟了道路，这些干预的共同目的是及时产生更好的信息来指导相关行动。现在看来很清楚，人类在自然任务上的表现无法用简单的自下而上模型来解释，在这些模型中，凝视固定（注意力顺序配置的自然相关物）由低级视觉显著性决定。这与早期的建议相反，早期建议认为简单的刺激特征在预注意提取后，可能会驱动我们的凝视/注意力在场景中移动。当然，这样的特征（绿点海洋中的红点、突然的闪光或水平线海洋中的垂直线）会捕获注意力。但是试图用这种本质上自下而上的术语（例如，Koch & Ullman, 1985）来定义所谓的”显著性地图(salience maps)“，在解释正常日常任务执行期间凝视和注意力的配置方面几乎没有提供什么帮助。

使用现实世界行走和虚拟（但相当真实）环境中行走的混合，Jovancevic et al. (2006)和Jovancevic-Misic and Hayhoe (2009)表明，简单的基于特征的显著性地图无法预测凝视何时何地会在场景中转移。Rothkopf, Ballard, and Hayhoe (2007)也获得了类似的结果，他们表明简单的显著性地图做出了错误的预测，并且在几乎所有情况下都无法解释观察到的固定模式。实际上：

人类主要注视物体，只有15%的固定指向背景。相比之下，显著性模型预测超过70%的固定应该指向背景。（Tatler et al., 2011，第4页）

Tatler等人进一步指出：

在球类运动中，基于特征的方案的缺陷变得更加明显。扫视会启动到球在不久的将来会到达的区域（Ballard & Hayhoe, 2009; Land & McLeod, 2000）。关键是，在固定目标位置的时候，没有任何东西在视觉上将该位置与场景的周围背景区分开来。即使没有定量评估，也很清楚任何基于图像的模型都无法预测这种行为。（Tatler et al., 2011，第4页）

向前看，看向当前空的（没有相关刺激存在）位置是自然任务执行期间凝视分配的普遍特征，并且已经在包括泡茶（Land et al., 1999）和制作三明治（Hayhoe et al., 2003）在内的任务中得到实验确认。在三明治案例中（下次你切三明治时检查一下！）被试看刀与面包第一次接触的地方，然后随着刀向前移动，继续看刀刃前方的位置。

面对这种在自然任务中无法解释日常表现形态的普遍失败，Tatler等人指出，一种回应是保留低级显著性图谱，但添加某种自上而下的调节机制。Navalpakkam and Itti (2005) 和 Torralba et al. (2006) 提出了这种混合方法。其他研究试图用其他结构替代低级显著性图谱，例如所谓的”优先级图谱”(Fecteau & Munoz, 2006)，它以任务特定的方式（即依赖先验知识的方式）流畅地整合低级和高级线索。然而，最有前景的（我认为）是那些从根本上重新定向讨论的方法，将感知和行动紧密耦合(Fernandes et al., 2014)，并将不确定性降低作为注视分配和注意力转移的驱动力。主要例子包括Sprague et al. (2007)、Ballard and Hayhoe (2009)和Tatler et al. (2011)，以及本章回顾的关于注意力和精度加权的不断增长的研究成果。[10] 所有这些方法的核心都是一个简单而深刻的洞察：

观察者已经学会了世界动态属性的模型，这些模型可以用来定位眼球注视以预期预测事件[并且]行动控制必须基于预测而非感知进行。(Tatler et al., 2011, p. 15)

这些模型随着经验而发展。Tatler等人指出，学习驾驶的新手在转弯时将注视分配到汽车前方不远处，而经验丰富的驾驶员则看得更远，注视路面位置比他们的行驶速度提前3秒(Land & Tatler, 2009)。板球运动员同样预测球的弹跳(Land & McLeod, 2000)。所有这些”主动扫视”案例（提前落在正确位置的扫视）都依赖于代理掌握和部署任务特定知识。这些知识体系（PP以概率生成模型的形式表示）首先反映动态环境本身的属性。它们也反映个体代理的行动能力（包括反应速度等）。环境属性发挥主要作用，这通过执行相同熟练任务的不同个体扫描模式之间的大量重叠得到证明(Land et al., 1999)。此外，Hayhoe et al. (2003) 表明信息通常在行动的及时时刻被获取，以将信息留在环境中直到恰当时机的方式（另见Clark, 2008第1章和后续第8章的讨论）。

精度加权PP账户理想地将所有这些元素整合到一个统一的故事中：一个将神经预测和不确定性降低置于中心舞台的故事。这是因为PP将行动、感知和注意力视为（实际上）形成单一机制，用于上下文和任务依赖的自下而上感觉线索与自上而下期望的结合。关键的是，这些自上而下的期望现在包括精度期望，这驱动行动系统以在重要的地方和时间降低不确定性的方式对场景进行采样。因此，注视分配由学习的生成模型驱动，这些模型将对展开事件的期望与关于如何最好地采样场景以在任务关键时刻降低不确定性的行动引导期望相结合。

PP账户还统一了外源性和内源性注意力的处理，揭示了低级”突出”效应与高级内部基于模型的效应在概念上的连续性。在前一种情况下，注意力被强烈、异常（绿色海洋中的红点）、明亮、突然等刺激捕获。这些都是进化系统

应该”期待”良好信噪比的情况。学习的效果在概念上是相似的。学习提供了如何以任务特定方式采样环境以产生高质量感觉信息的掌握。这降低了不确定性并简化了任务表现。正是这种后一种知识在内源性注意力中发挥作用，也许（见Feldman & Friston, 2010, pp. 17 – 18）通过增加选定神经元群体的基线放电率。

在继续之前，我应该对”自然任务”提出一个重要的警告。为了简单起见，我在这里专注于一些熟练学习的（可能是过度学习的）任务，如驾驶和制作三明治。但PP账户在由已知元素和结构构建的新情况中也能提供流畅快速的学习。这意味着我们可以快速成为（适度）全新场景的”专家观察者”。例如，在观看戏剧表演时，我们迅速掌握舞台上人物和物体的新颖安排，学习什么是情节相关的，因此在哪里（以及何时）我们最需要降低不确定性，相应地主动分配注视和注意力。

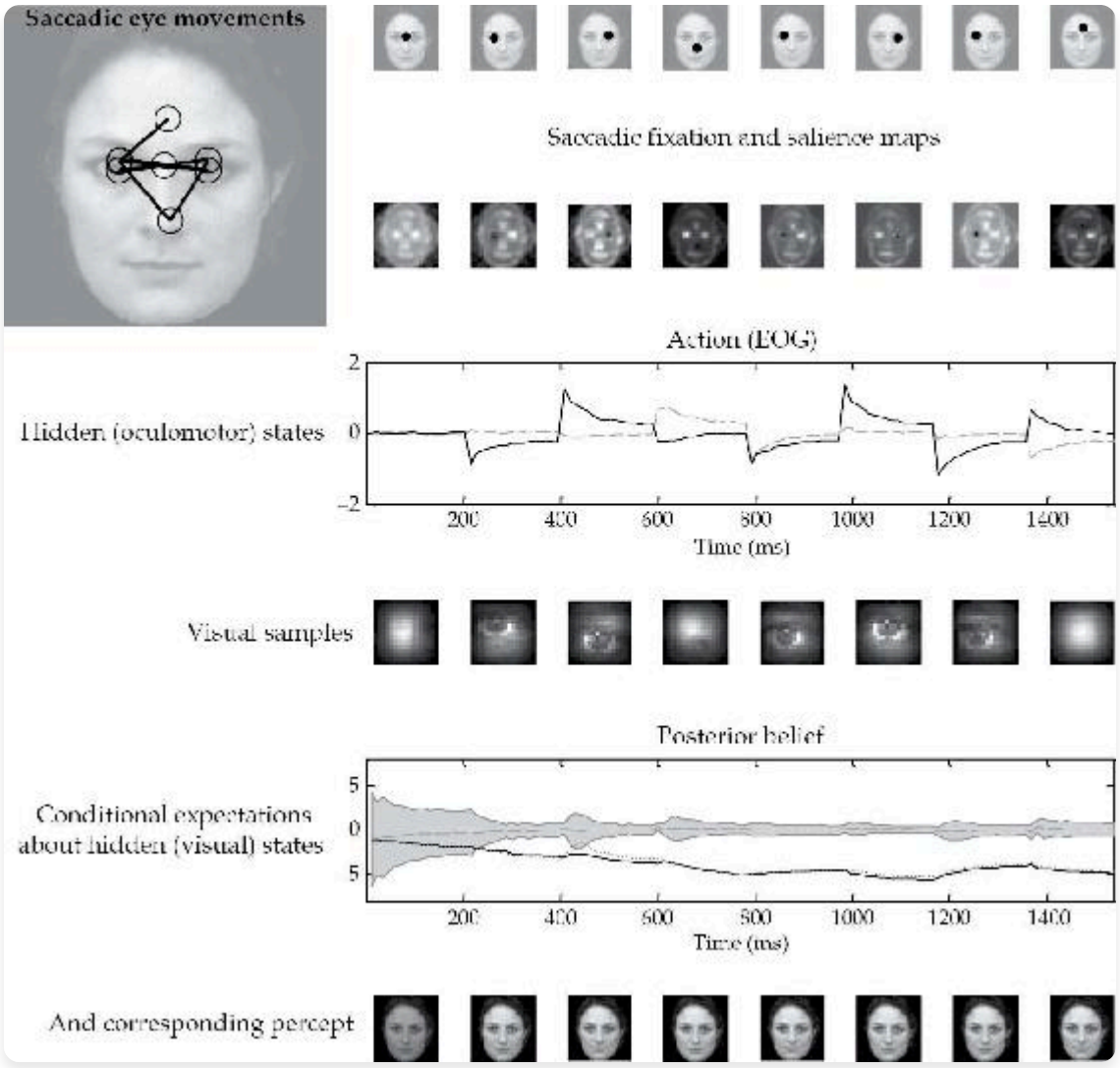
[2.8 感知-注意力-行动循环中的循环因果关系]

所有这些的一个重要结果是，感知背后的生成模型包含了关于*前瞻性确认*的关键行动驱动期望。也就是说，它包含了(次个人的)期望，这些期望涉及假设某个当前感知假设(决定我们正在进行的感知意识的那个)是正确的，事物应该如何展开。这些期望既关注如果我们按照假设对世界进行采样会发生什么(即，会产生什么感知输入)，也关注会产生什么信噪比。在后一种情况下，大脑押注于可以称为*前瞻性精度*的东西，即通过移动我们的眼睛、其他感觉器官，甚至我们的整个身体来采样场景后预期的信噪比。因此，一系列向预期能提供某个特定感知假设所预测的(而不被邻近竞争对手预测的)高精度信息的位置进行的扫视，为该假设是正确的并值得保持活跃和”处于主导地位”提供了极好的证据。但如果事情未能如期发生(如果感知”实验”的结果似乎证伪了假设)，这些错误信号可以用来招募不同的假设，如前面描述的方式。

这有一个直接而有趣的后果，随着故事的展开将继续占据我们的注意力。这意味着感知、注意力和具身行动协同工作，驱动主体进入自我推动的主动感知循环，在这个循环中，我们根据关于我们自己的行动即将揭示的东西的系统性”信念”来探测世界。这导致了Friston, Adams等人(2012)所描述的”感知背后的循环因果关系”，即：

唯一能在连续扫视中持续的假设是正确预测被采样的显著特征的那个。…这意味着假设规定了自己的验证，只有当它是世界的正确表征时才能存活。如果没有发现其显著特征，它将被丢弃以支持更好的假设。(Friston, Adams等人，2012，第16页)

“显著特征”是当被采样时，最小化关于当前感知假设的不确定性的特征(它们是当事情如预期展开时，最大化我们对假设信心的特征)。因此，主动主体被驱动去采样世界，以便(试图)确认它们自己的感知假设。当前获胜的感知应该能够”通过选择性地采样支持其自身存在[正确性]的证据来维持自己”(第17页)。这种采样确实意味着一种”显著性地图”：但这不是由低级的、吸引注意力的视觉特征决定的地图，而是由关于世界和显著、精确的感觉信息分布的相对高级知识决定的地图。

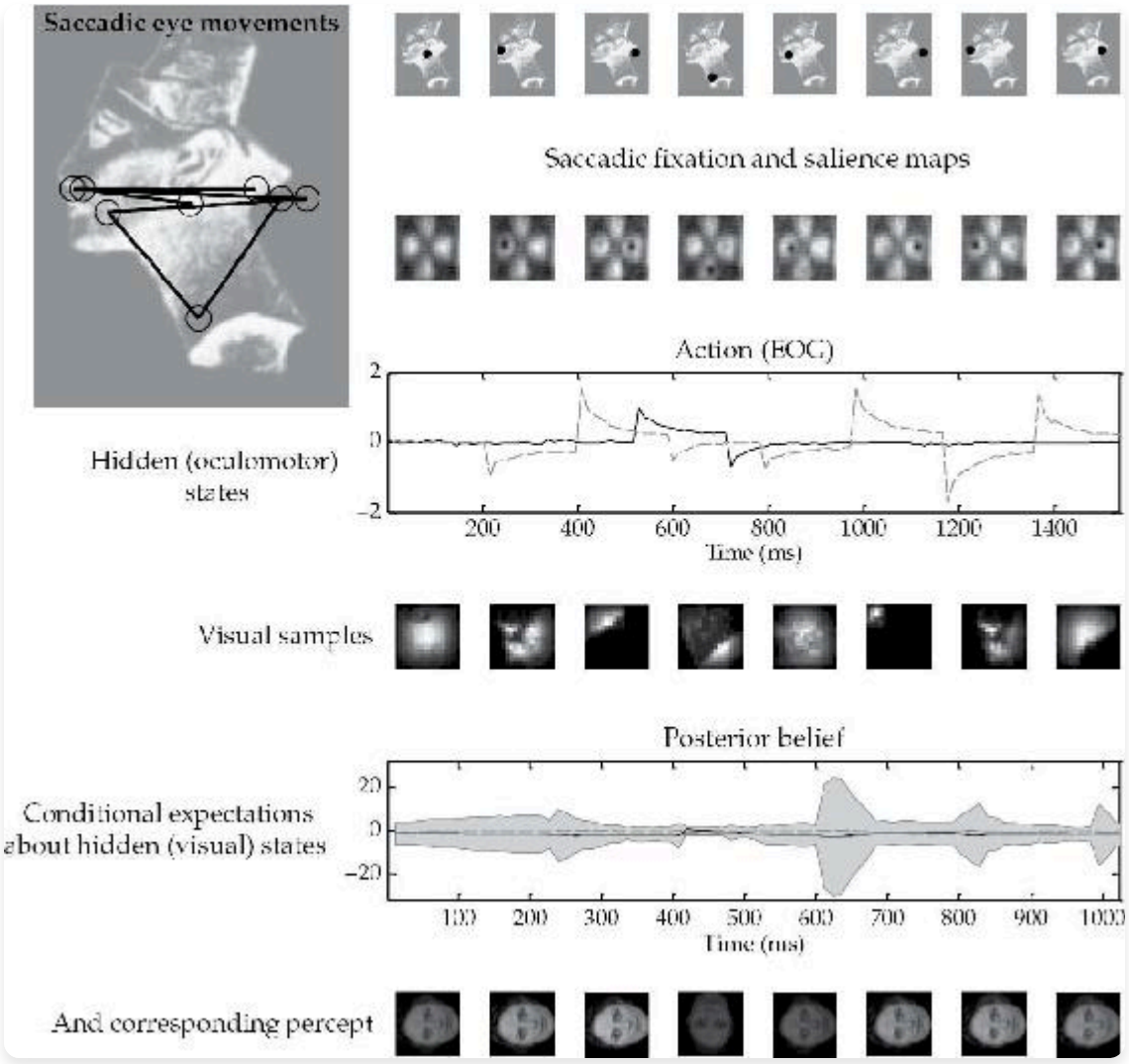


image

图2.4 该图显示了Friston, Adams等人(2012)第一次模拟的结果，其中向一个主体呈现了一张脸，该主体的反应使用文本中描述的PP模式进行模拟。在此模拟中，主体有三个关于它可能采样的刺激的内部图像或假设(一张正立的脸、一张倒立的脸和一张旋转的脸)。向主体呈现了一张正立的脸，并在16个(12毫秒)时间段内评估其条件期望，直到发出下一次扫视。这重复了八次扫视。随之产生的眼动在上排显示为每次扫视结束时位置(外在坐标)的点。相应的眼动序列显示在左上角的插图中，其中圆圈大致对应于采样图像的比例。这些扫视由基于第二排显著性地图的关于凝视方向的先验信念驱动。注意这些地图随着连续扫视而变化，因为关于隐藏状态(包括刺激)的后验信念变得逐渐更加确信。还要注意，在前一次扫视中被注视的位置的显著性被耗尽了。这些后验信念提供视觉和本体感觉预测，分别抑制视觉预测误差和驱动眼动。眼球运动反应在第三排中以对应于垂直和水平位移的两个隐藏眼球运动状态的形式显示。相关的图像采样部分(在每次扫视结束时)显示在第四排。最后两排分别以充分统计量和刺激类别的形式显示后验信念。这里的后验信念以条件期望和关于真实刺激的90%置信区间的形式绘制。这里需要注意的关键是，关于真实刺激的期望优于其竞争期望，结果，关于刺激类别的条件信心增加(置信区间收缩到期望)。这说明了在选择最能解释感觉数据的假设或感知时证据积累的性质。有关实验和结果的完整细节，请参见Friston, Adams等人2012年的原始论文。

来源：经Friston, Adams等人2012年许可转载。

Friston、Adams等人通过一个简单的仿真实验证明了这一核心效应（见图2.4），在该实验中，人工智能体(agent)在各种感知假设驱动下采样视觉场景。在这里，智能体控制三个模型来尝试拟合刺激，最终选定能够正确预测在某种扫视模式下产生的感觉数据的模型。经过几次早期探测后，仿真智能体依次注视那些确认输入源为正立人脸假设的点。图2.5显示了系统在呈现不符合其已知模型的图像时的行为。在这种条件下，没有模型（没有假设）能够规定一种能够自我确认的注视模式，因此感觉不确定性无法消除，也无法选择模型。然后没有知觉能够”通过选择性地采样支持其自身存在[正确性]的证据来维持自身”（Friston, Adams等人，2012年，第17页）。在这种不利条件下，场景以游荡的方式被采样，不会产生清晰稳定的知觉。然而，这种失败（假设大脑相信它正在获得高质量的精确感觉信息）会推动增加的可塑性，允许获得和应用新模型（见第2.12节）。



图像

图2.5 此图使用与前图相同的格式，但显示了呈现未知（无法识别）面孔——古埃及女王奈费尔提提的图像的结果。由于仿真智能体没有内部图像或假设能够对需要凝视的显著位置产生真实预测，它无法解决其感觉输入的原因，无法将视觉信息同化为关于刺激的精确后验信念。扫视运动由显著性地图生成，该地图基于关于刺激的所有内部假设的混合来表示最显著的位置。无论智能体看向何处，它都找不到能够解释感觉输入的后验信念或假设。结果，对世界状态的后验不确定性持续存在，无法自我解决。随之而来的知觉形成不良，随着连续的扫视而零星变化。

来源：来自Friston, Adams等人，2012年，经许可。

总结起来，预测处理(PP)假设了核心的感知-注意-行动循环，其中世界的内部模型及其相关的精度期望发挥关键的行动驱动作用。这些共同决定了一个（经常自我实现的）探索性、认识论要求的感知和行动过程：一个获胜假设（对世界的获胜”理解”）使我们以既反映假设本身又反映我们自身感觉不确定性情境变化状态的方式来采样场景的过程。

2.9 相互确保的误解

然而，所有这些也有可能的阴暗面。当对精度的微妙误导性估计导致我们以阻碍形成事物真实状况良好（真实）图像的方式收集感觉信息时，阴暗面就会出现。Siegel（2012）描述了正是这样一种可能的情景。这是一个”吉尔毫无根据地相信杰克对她生气…当她看到杰克时，她的信念使他在她看来显得生气”的情景。在这种情况下，我们主动的自上而下模型使我们丢弃信号的某些元素（将它们视为仅仅是”噪音”）并放大其他元素。通常——如上述案例所示——这会在嘈杂和模糊的环境中导致更准确的感知。然而，在看起来愤怒的杰克的案例中，我们的信念使我们准备部署一个（部分通过改变我们分配给预测误差信号各个方面的精度）“发现”虚假且无根据的杰克愤怒假设的视觉证据的模型。这就像在噪音中听到”隐藏”的歌曲”白色圣诞节”一样。结果是我们的视觉体验本身（不是某种附加判断）然后将杰克表现为看起来愤怒，为我们早先的怀疑火上浇油。

在这里，行动和感知被锁定在一个*相互误导*的循环中。这是因为被激活的”愤怒的杰克”假设得以控制（以我们将在后面章节中更详细探讨的方式）随后探测世界以寻找确认杰克愤怒证据的*行动*。我们在杰克的脸部周围扫视寻找微妙的证据，我们寻找他肢体动作中的紧张，他用词选择中的异常等等。由于我们提高了携带关于愤怒微妙”迹象”信息的信号的精度，（从而）降低了对正常性真实迹象的精度，我们很可能找到我们正在寻找的”证据”。在现实世界环境中，Teufel、Fletcher和Davis（2010）表明，我们对他人当前心理状态和意图的主动自上而下模型确实影响我们如何在物理上感知他们，影响我们对他们凝视方向、运动开始、运动形式等的基础感知（关于自上而下知识对感知影响的更多例子，见Goldstone，1994；Goldstone & Hendrickson，2010；Lupyan，2012）。

为了巩固这种悲剧，杰克和吉尔都是PP智能体（因此都是感知深度受预测渗透的存在）这一事实可能会迅速让事情变得更糟。因为吉尔的探察和猜疑对杰克本人来说并非无形，而且她的肢体语言有些紧张。杰克（错误地）想：“也许吉尔在生我的气？”。现在这个场景再次重演。对杰克来说，吉尔现在看起来有些生气，听起来也有些生气。然后吉尔在杰克身上察觉到更多紧张的迹象（也许现在这些迹象是真实的），这种相互（最初错位的）预测的循环不断升级。正如我们稍后将看到的，相互预测可以极大地增进人际理解。但当与期望对感知和行为的深刻影响相结合时，它也可能为自我实现的心理社会结(knots)和纠缠提供一个令人担忧的配方。

这种阴暗面也不仅限于多个相互作用的人类智能体的情况。越来越多的最佳工具和技术正在从事预测我们自己的需求、请求和使用模式的业务。谷歌根据过去的模式和当前位置信息预测搜索请求，甚至在我们提出要求之前就提供建议和选项。亚马逊使用强大的协作过滤技术根据过去的购买记录提出建议。这些创新将相互预测的领域扩展到包括人类和机器的网络，每一方现在都忙于预测对方。除非得到检查或控制，这可能会导致，正如[Pariser (2011)]中描述的所谓”过滤泡沫”场景所示，对机会空间的探索日益受限。

2.10 关于精确性的一些担忧

注意力的基本PP理论的简要概述出现在[Clark (2013)]中。这是同行评议期刊《行为与脑科学》的一篇目标文章，因此伴随着该领域领军人物的各种评论。[Bowman et al. (2013)]就是这样一篇评论。除了对有偏竞争的一些担忧（见2.4节）外，Bowman等人还担心基于精确性的解释似乎最适合解释空间注意力，而非基于特征的注意力。Bowman等人指出，基于特征的注意力允许我们增强对给定特征的反应，即使它出现在未预测的位置。因此，借用他们的例子，

寻找粗体字实例的命令可能会导致注意力被附近的空间位置捕获。如果我们然后（正如PP所建议的）增加来自该空间位置的预测误差的精确性权重，这难道不意味着选择性预测误差信号的精确性权重是注意的*结果*而非其*因果机制*吗？

这是一个很好的难题，它揭示了所提供装置的重要特征。因为我认为，解决这个难题在于操纵处理体系不同层级上的精确性权重。基于特征的注意力直观地对应于增加与刺激的身份或配置相关的预测误差单元的增益（例如，增加报告与四叶草独特几何模式相关的预测误差的单元的增益）。提升这种反应（通过给相关类型的感觉预测误差增加权重）应该能增强对该特征线索的检测。一旦线索被暂时检测到，受试者就可以注视正确的空间区域，现在处于”四叶草在那里”的期望条件下。然后，该特征在该位置的残余误差被放大，如果你幸运的话，就可以获得对四叶草存在的高置信度！注意，注意错误的空间区域（例如，由于不一致的空间线索）在这种情况下实际上会适得其反。因此，精确性权重预测误差能够包含纯空间和基于特征的信号增强。

[Block and Siegel (2013)]提出了额外的担忧，他们认为预测加工(predictive processing)无法为一些非常基本的结果（如对感知对比度的注意力增强[Carrasco, Ling, & Read, 2004]）提供任何合理或独特的解释。特别是，Block和Siegel认为PP模型未能捕获由于注意而产生的*先于*误差计算的变化，并且错误地预测了注意后续变化的放大（由于提高某些预测误差的增益而产生）。值得详细看看这个案例。

Carrasco, Ling, and Read (2004) 报告了一项实验，受试者注视中心点，左右两侧显示对比度光栅。光栅具有不同的绝对（实际）对比度。但当受试者被提示要注意（即使是隐蔽地）较低对比度的光栅时，他们对该处对比度的感知会增加，产生（错误的）判断，例如，一个被注意的70%（实际值）对比度光栅与一个未被注意的82%光栅相同。Block和Siegel认为预测处理解释无法解释这里的初始效应（对隐蔽注意的70%对比度光栅的82%对比度的错误感知），因为唯一的错误信号——这正是他们误解故事的地方——是稳定的注意前70%记录与注意后82%记录之间的差异。但这种差异直到注意完成其工作后才可用！更糟糕的是，一旦这种差异可用，它不应该再次被放大吗，因为相关错误单元的增益现在增加了？

这是一个巧妙的挑战，但它基于对精度加权(precision-weighting)提议的一个revealing误解。PP并不假设基于未注意对比度（记录为70%）和随后注意的对比度（现在看起来是82%）之间差异计算的错误信号。相反，注意改变的是对来自被注意空间位置的精确感觉信息的期望。精度是方差的倒数，正是我们的”精度期望”被注意在这里改变了。在手头的情况下，似乎正在发生的是，我们隐蔽地注意左侧（比如说）光栅这一事实增加了我们对精确感觉信号的期望。在这种条件下，对精确信息的期望诱导对感觉错误的夸大加权，结果我们对对比度的主观估计被扭曲了。[14]

重要的一点是，错误不是像Block和Siegel似乎暗示的那样，作为某个先前（在这种情况下是未注意的）感知与某个当前（在这种情况下是注意的）感知之间的差异来计算的。相反，它直接为当前感觉信号本身计算，但根据我们对来自该位置的精确感觉信息的期望进行*加权*。根据PP，精度期望是对比度光栅实验所操纵的，因此PP为效应本身提供了令人满意的（且独特的）解释。这种相同的机制解释了注意对空间敏锐度的一般效应。

Block和Siegel还论证说，至少一旦代理人醒着、警觉并掌握了周围环境，“将错误信号视为感觉输入是没有意义的”。但当然，这个声明并不是说代理人感知错误信号。（同样，没有传统理论家应该说代理人通常感知感觉信息流本身，而不是它所提供的世界。）根据PP，代理人感知周围的东西，但这样做*依靠*错误的前向（和横向）流动以及预测的向下（和横向）流动。

总之，预测处理将注意描述为增加选定预测错误的增益。因此，注意构成了构成正常感知过程的推理级联的一个组成部分。内源性注意(Endogenous attention)在这里对应于影响与某些任务相关特征（例如，四叶草的形状）或某个选定空间位置相关的预测错误增益的意志控制过程。外源性注意(Exogenous attention)对应于更自动的处理，在熟练任务的流畅执行期间，或响应某些生态学显著线索（如闪光、运动瞬变或突然噪音）时，增加选定预测错误的增益。这些生态学显著线索往往产生强烈的感觉信号，这些信号被隐含地”期望”显示高信噪比。因此有一种超先验

(hyperprior)在起作用：对更强信号精度的期望，这合理地要求增加相关预测错误的增益（见Feldman & Friston, 2010, p. 9；另见Hohwy, 2012, p. 6）。最后，精度期望也被看到指导探索性行动，确定（例如）跟踪场景中最可能找到更精确信息的区域的扫视模式。

在单一的自我驱动循环中结合感知和行动，精度估计因此使自下而上的感觉信息（通过预测错误传递）和自上而下的基于生成模型的期望能够灵活地根据任务变化组合。

[2.11 意外的大象]

理解精度和精度期望的作用对于揭示非意识（“亚个人的”）预测与个人层面日常体验的形状和流动之间的复杂联系可能特别重要。例如，神经惊讶（“惊讶值”(surprisal)：给定世界模型，某些感觉状态的不可能性）与代理人惊讶之间似乎存在初始的脱节。这从一个简单事实中显而易见：总体上最能最小化惊讶值（因此最小化预测错误）“对于”大脑的感知很可能对我这个代理人来说是一些高度令人惊讶和意外的事态——例如，想象一下专业魔术师优雅地将一头巨大而忧郁的大象偷偷带到舞台上的突然揭示。然而，这里根本脱节的表象是虚幻的，正如稍微详细的解释所揭示的。

随着魔术师挥去遮布，粗糙快速处理的视觉线索招募出最能最小化感觉预测误差的假设（大象）。感知/行动回路立即参与其中，驱动一系列视觉扫视，以大象特有的方式扫描场景（例如，凝视象鼻应该在的位置）。如果假设正确，这种视觉搜索将产生对该假设的高精度确认。假设扫视满足了所有系统期望。智能体现在掌握了一个可靠的模型，该模型经受住了高精度预测误差的严峻考验。此时大象感知是最符合认知系统对世界的了解和期望，以及它对自身干预世界结果（这里是视觉扫视）的了解和期望的感知。因此，舞台上的大象感知是*给定*当前驱动输入、精度期望和分配精度（如我们所见，反映大脑对感觉信号的置信度）组合下的获胜假设。

给定正确的驱动信号和足够高的精度分配，初始智能体意外类型的顶层理论因此可以获胜，以解释掉高度加权的传入感觉证据潮流。悲伤大象的景象作为给定输入、先验和感觉预测误差估计精度下可用的最佳（最可能、最少”惊讶性”的）感知出现。尽管如此，系统先验并没有使该感知在事前非常可能，因此（也许）智能体实际惊讶感对其是有价值的。也就是说，*惊讶感*可能是保存有用信息的一种方式，否则这些信息将被丢弃——这些信息表明，在当前证据驱动的推理回合之前，感知到的事态被估计为高度不可能。

这（通常）都是好消息，因为这意味着我们不是期望的奴隶。成功的感知需要大脑使用存储的知识和期望（贝叶斯先验）来最小化预测误差。但我们仍然能够在大脑为感觉预测误差分配高可靠性（因此为驱动感觉信号分配高可靠性）的条件下看到非常（智能体）惊讶的事物。重要的是，这需要其他高层理论，虽然是初始智能体意外类型的，获胜以解释掉高度加权的感觉证据。

2.12 精度的一些病理现象

但是，如果这种平衡行为出错会发生什么？如果精度加权机制出现故障，自上而下期望和自下而上感知之间的平衡受到损害会发生什么？在这里，我认为预测处理情境为思考人类心理的广阔且多样的空间提供了有前景的新方式。我们将在后续章节中看到更多这方面的内容。但我们已经可以在一项令人印象深刻的近期工作中瞥见这种潜力，该工作涉及精神分裂症中的妄想和幻觉（Corlett, Frith, et al., 2009; Fletcher & Frith, 2009）。

回想一下前一节中描述的意外大象目击。在这里，系统已经掌握了一个恰当的模型，能够”解释掉”指定悲伤灰色存在的驱动输入、期望和精度（预测误差的权重）的特定组合。但情况并非总是如此。有时，处理持续的、高度加权的感觉预测误差可能需要逐渐形成全新的生成模型（就像在正常学习中一样）。正如Fletcher和Frith（2009）所建议的，这可能是更好理解精神分裂症中幻觉和妄想起源（这两个所谓的”阳性症状”）的关键。这两种症状通常被

认为涉及两种机制，因此有两种破坏，一种在”感知”中（导致幻觉），一种在”信念”中（允许这些异常感知影响顶层信念）。因此Coltheart（2007）正确且重要地指出，仅仅感知异常通常不会导致妄想受试者中发现的奇异复杂信念体系。但我们因此必须将感知和信念形成成分视为严格独立的吗?

如果感知和信念形成，如当前故事所建议的，都涉及尝试将展开的感觉信号与自上而下的预测相匹配，那么可能的联系就会出现。重要的是，这种尝试匹配的影响是精度介导的，因为残余预测误差的系统效应根据大脑对信号的置信度而变化。考虑到这一点，Fletcher和Frith（2009）调查了对分层贝叶斯系统扰动的可能后果，使得预测误差信号被错误生成，更重要的是——高度加权（因此被赋予不当的显著性来驱动学习）。

存在许多潜在机制，它们复杂的相互作用一旦在预测误差最小化的总体框架内得到处理，可能会共同产生这种干扰。主要候选机制包括缓慢神经调节剂的作用，如多巴胺、血清素和乙酰胆碱(Corlett, Frith, et al., 2009; Corlett, Taylor, et al., 2010)。此外，Friston (2010, p. 132)推测，神经区域之间快速、同步的活动也可能在增强同步群体内预测误差的增益方面发挥作用。[16]然而，无论如何实施，关键思想是理解精神分裂症的阳性症状需要理解预测误差产生和（特别是）权重分配中的干扰。建议是该复杂经济体系内的故障（可能根本上源于异常的多巴胺功能）产生一波又一波持续且高权重的”虚假误差”，然后传播到整个层次结构，在严重情况下（通过随之而来的神经可塑性浪潮）强制对我们的世界模型进行极其深度的修正。不可能的事情（心灵感应、阴谋、迫害等）然后变成最不令人惊讶的，而且——因为感知本身受到先验期望自上而下流动的调节——错误信息的级联传播回到下层，使虚假感知和奇异信念固化为一个连贯且相互支持的循环。

这样的过程是自我强化的。随着新的生成模型站稳脚跟，它们的影响向下流动，使传入数据被新的（但现在严重误导的）先验所塑造，以”符合期望”(Fletcher & Frith, 2009, p. 348)。因此，虚假感知和奇异信念形成了一个认识论上隔离的自我确认循环。那么，这就是一种高度有效认知策略的阴暗面。预测处理模型将感知、信念和学习融合在一个单一的总体经济体系中——通常是富有成效的——在这个体系中，多巴胺以及其他机制和神经递质控制着预测误差本身的”精确性”（权重，因此对推理和学习的影响）。但当事情出错时，虚假推理螺旋上升并反馈到自身。妄想和幻觉然后变得根深蒂固，既被共同决定又共同决定。我们在科学(Maher, 1988)和日常生活中到处都能看到这种现象的温和版本。我们倾向于看到我们期望的东西，并用它来确认既在产生我们期望、又在塑造和过滤我们观察以及对其可靠性估计的模型。

同样的广义贝叶斯框架可以用来(Corlett, Frith, et al., 2009)帮助理解不同药物在给予健康志愿者时如何暂时模拟各种形式的精神病。在这里，关键特征也是预测编码框架能够解释学习和体验中的复杂改变，这些改变取决于驱动感觉信号如何通过精确度加权的预测误差与先验期望和（因此）持续预测相结合的（药理学可修改的）方式。例如，氯胺酮(ketamine)的拟精神病效应据说可以用预测误差信号的干扰（可能由AMPA上调引起）和预测流（可能通过NMDA干扰）来解释。这导致持续的预测误差，并且——关键的是——对相关事件重要性或显著性的夸大感知，这反过来又驱动短暂妄想样信念的形成(Corlett, Frith, et al., 2009, pp. 6 – 7; 另见Gerrans, 2007)。作者继续解释其他药物（如LSD和其他血清素致幻剂、大麻和多巴胺激动剂如苯丙胺）的不同拟精神病效应，认为它们反映了分层预测处理框架内其他可能的干扰类型。[17]

在我看来，这种层次的流畅跨越构成了当前框架的关键吸引力之一。我们在这里从对人类体验正常和改变状态的考虑，通过计算模型（突出精确度加权的基于预测误差的处理和生成模型的自上而下部署），到大脑中突触电流、神经同步和化学平衡的实施网络。希望是通过提供推理、期望、学习和体验之间复杂、系统性相互作用的新的多层次解释，这些模型有朝一日可能提供对我们自己代理者级别体验的更好理解，甚至超过”民间心理学”基本框架所提供的理解。这样的结果（另见第7章）将构成对以下主张的证实(P. M. Churchland, 1989, 2012, P. S. Churchland, 2013)：采用”神经计算视角”有朝一日可能引导我们对自己生活体验的更深刻理解。

[2.13 超越聚光灯]

注意力经常被描绘为一种心理聚光灯（参见，例如，Crick, 1984），其部署反映了竞争（由于有限的资源）以获得高质量的神经处理。注意力的预测处理模型与聚光灯模型有一些共同特征，但在其他方面有所不同。它们都将注意力描绘为与寻找精确（低不确定性）感官信息相关联。指向聚光灯创造了（如Feldman & Friston, 2010所指出的）能够从空间位置获得高质量感官信息的确切条件。但PP认为，注意力本身不是一个*机制*，而更多的是一个更根本资源的*维度*。[18]它是我们（我们的大脑）用来预测感官数据流的生成模型的一个普遍维度。但这是一个特殊的维度，因为它不仅涉及传入感官数据的外部原因（信号）的性质，还涉及感官信息本身的精度（统计学上，逆方差）。

生成模型通过包含当前精度的估计以及视觉扫视和其他行为将产生的精度估计，直接带动大量信息收集行为。它不仅对信号应该如何演变（如果世界确实是这样的话）做出预测，还对应该主动征求哪些传入信号并在处理展开时给予最大权重做出预测。正是通过改变这种权重分配，我们能够在多模态处理期间偏向选择感官通道，灵活地改变神经区域之间信息的即时流动，以及（最一般地）改变自下而上的感官信号和自上而下期望之间的权力平衡。这种改变完成了本章早期部分描述的各种”特效”（在云中看到面孔，听到正弦波语音，甚至幻听”白色圣诞节”）。

将我们自己感官不确定性的精度编码估计添加到新兴图景中，还允许我们以流畅灵活的方式结合两个表面上对立世界的精华。一个是信号抑制的世界，这是标准预测编码的核心特征。在这里，预期的信号元素被”解释掉”并被剥夺了向前流动的因果效力。另一个是信号增强和偏向竞争的世界。这是一个”关键任务”信号元素被放大和增强，其向前流动的效应被放大的世界。通过根据预期精度对向前流动的预测误差信号进行加权，PP框架结合了这两个世界的精华，增强一些响应同时抑制其他响应。

注意力、行动和感知现在在相互支持、自我驱动的循环中结合起来。加权预测误差信号驱使我们以既反映又测试生成预测的假设的方式对世界进行采样，这些预测正在驱动着行动。感知、注意力和行动的这种亲密关系形成了本著作的核心主题之一，并为我们提供了迄今为止关于能够阐明大脑、身体和世界深刻认知纠缠的神经处理账户的最佳希望。

[3]

[想象空间][]

[3.1 构造产业]

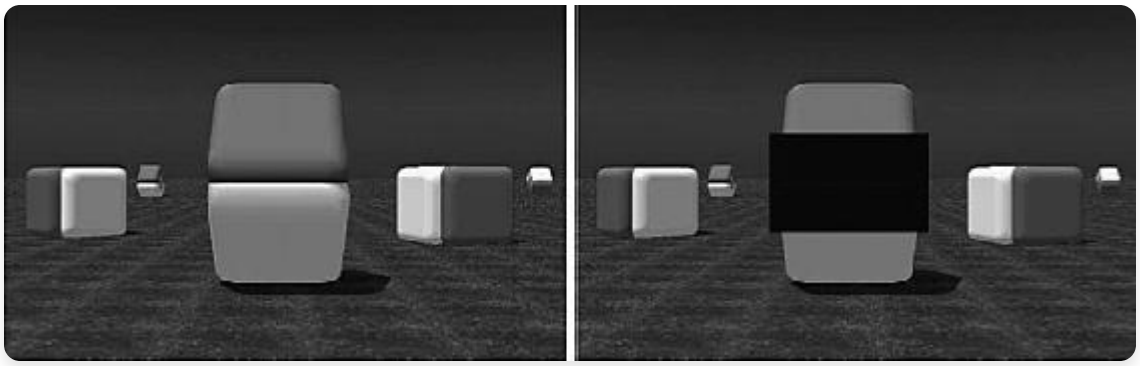
我们的故事表明，感知是一个既具有构造性又深受预测影响的过程。这种类型的感知——那种揭示了相互作用的远端原因的结构化世界的感知——有一个重要且（大多数情况下）促进生活的附带效果。因为这样的感知者因此也是想象者：他们是准备好不仅通过感知和总体物理行动，还通过意象、梦境和（在某些情况下）故意的心理模拟来探索和体验他们世界的生物。

当然，这并不是说每个我们可能直觉上认为与其世界有某种形式感官接触的系统都能够做这些事情。毫无疑问，存在许多简单的系统（如追随光线的机器人或追随化学梯度的细菌）使用感官输入来选择适当的响应，而不部署内部表示的模型来预测传入信号的形状。我将论证，这样的系统不会享受到对丰富结构化外部世界的感知体验，也不会具有梦境或想象等心理状态。但是，如果PP是正确的，像我们这样的感知者可以使用存储的知识来生成一种多层次虚拟类比的驱动感官信号，因为它在多个层次和类型的处理中展开。

想象力和梦境的联系就在眼前，因为这样的系统掌控着一个生成模型(generative model)，能够利用关于世界中相互作用因果关系的知识来重构感觉信号。这种重构过程在感觉信号存在时得到调节和部署，为彻底的构造过程铺平了道路，能够在缺乏常规感觉流的情况下形成和演化。附近还有参与某些理论家所称的”心理时间旅行”的能力：记忆（重构）过去和预测未来的可能形态。这些各种”构造产业”协同工作，让我们能够做出更好的选择和选择更好的行动。因此，从基于生成模型的在线感知的简单种子中，涌现出一种引人注目的（且惊人地熟悉的）认知形式。这是一种形式，其中感知、想象力、理解和记忆作为一种认知套餐交易出现——这种套餐交易将现在定位在其体验归属的地方，即过去影响与知情未来选择之间富有成效的交汇点。

简单视觉

考虑图3.1中的图像。这就是所谓的”康斯威特错觉”。对大多数人来说，中央配对的瓦片看起来是非常不同的灰色阴影——这种外观，正如第二张图片所揭示的，是虚幻的。这种错觉发生是因为（正如我们在第1章和第2章中看到的）我们的视觉体验不仅仅反映当前的输入，而是很大程度上受到关于世界的”先验”（先验信念，通常以无意识预测或期望的形式出现）的影响。在这种情况下，先验是表面倾向于具有相同的反射率，而不是朝着它们自己的边缘逐渐变亮或变暗。因此，大脑的最佳猜测是中央配对涉及两个不同反射率的表面（两种不同的灰色阴影）被不同数量的光照亮。错觉发生是因为图像显示了照明和反射特性的高度非典型组合，大脑使用它从典型照明和反射模式中学到的知识来推断（在这种情况下是错误的）两个瓦片必须是不同的灰色阴影。在我们实际生活的世界中，这些特定的先验信念或神经期望可证明是”贝叶斯最优”的——也就是说，它们代表了从环境感觉证据中推断世界状态的全局最佳方法（Brown & Friston, 2012）。因此，大脑通过结合先验知识（包括，正如我们在第2章中看到的，关于上下文的知识）与传入的感觉证据来生成我们的感知体验。



image

图3.1 **康斯威特错觉设置**

第一张图像（*左*）描绘了一个典型的康斯威特错觉设置。构成中央配对的两个瓦片的中心看起来是不同的灰色阴影。第二张图像（*右*）揭示它们实际上是相同的灰色阴影。

来源：D. Purves, A. Shimpi, & R. B. Lotto (1999)。康斯威特效应的实证解释。*神经科学杂志*，19(19)，8542-8551。

跨模态和多模态效应

这种基本效应解释了各种各样令人惊讶的熟悉感知现象。其中一种现象是跨模态和多模态上下文效应对早期”单模态”感觉处理的广泛存在。这些效应的发现构成了当代感觉神经科学的主要发现之一（见，例如，Hupe et al., 1998; Murray et al., 2002; Smith & Muckli, 2010）。因此，Murray等人（2002）展示了高级形状信息对早期视觉区域V1中细胞反应的影响，而Smith和Muckli（2010）甚至在完全未受刺激（也就是说，未通过驱动感觉信号直接刺激）的视觉区域上显示了类似的效应（使用部分遮挡的自然场景作为输入）。此外，Murray等人（2004）显示V1中的激活受到自上而下尺寸错觉的影响，而Muckli等人（2005）和Muckli（2010）报告了V1中与表观运动错觉相关的活动。即使是看似”单模态”的早期反应也受到（Kriegstein & Giraud, 2006）来自其他模态的信息的影响，因此通常会反映各种多模态关联。引人注目的是，甚至对相关输入将在一种模态（例如，听觉）而不是另一种模态（例如，视觉）中出现的期望也被证明能提高性能，可能通过增强”自下而上输入对给定感觉通道上感知推理的权重”（Langner et al., 2011, p. 10）。

这一系列上下文效应的全貌非常自然地从预测处理(PP)模型中流淌而出。如果所谓的视觉、触觉或听觉感官皮层实际上是使用来自更高层级的反馈级联来主动预测展开的感官信号(这些信号最初通过视觉、声音、触觉等各种专用受体库转导)，那么我们不应该对在”早期”感官反应中发现广泛的多模态和跨模态效应(包括这些”填充”类型)感到丝毫惊讶。产生这种情况的一个原因是，“早期”感官反应的概念在某种意义上现在是误导性的，因为期望诱导的上下文效应将简单地在整个系统中传播，启动、生成和改变”早期”反应，一直向下到V1。任何在”元模态”(或至少是日益信息整合的)区域内记录的统计有效相关性，这些区域位于处理层次的顶部，都能为预测提供信息，然后通过之前被认为是更加单模态的区域，一路级联到接近感官外围的区域。这些效应与将V1视为使用具有固定(上下文不灵活)感受野的细胞进行简单的、刺激驱动的、自下而上特征检测的场所的观念不一致。但它们完全符合(实际上是被要求的)那些将V1活动描绘为基于自上而下预测和驱动感官信号的灵活组合不断协商的模型。反思这种”早期”感官处理的新视角，Lars Muckli写道：

可以想象，V1首先是皮层反馈的目标区域，其次才是将皮层反馈与传入信息进行比较的区域。感官刺激可能是皮层的次要任务，而其主要任务是…尽可能精确地预测即将到来的刺激。(Muckli, 2010, p. 137)

[3.4 元模态效应]

视觉词汇形式区域(VWFA)是腹侧流中的一个区域，它对适当的字母串做出反应：这种字母串在给定语言中可能合理地形成一个词。这个脑区域的反应已知独立于表面细节，如大小写、字体和空间位置。在一项重要的神经成像(fMRI)研究中，Reich等人(2011)发现证据表明VWFA实际上在追踪比视觉词汇形式更加抽象的东西。它似乎在追踪词汇形式，无论转导流的模态如何。因此，同样的区域在先天性盲人被试进行盲文阅读时被激活。这里的早期输入是触觉而非视觉的事实对VWFA的招募没有影响。这支持了(Pascual-Leone & Hamilton, 2001)将这些脑区域视为”元模态操作器”的观念，这些操作器”由给定的计算定义，无论接收到何种感官输入都会应用该计算”。

这很好地契合了，正如Reich等人(2011, p. 365)自己注意到的，PP图像中皮层层次的更高层级学习追踪”隐藏原因”，这些原因解释并因此预测远端事态的感官后果。Reich等人推测VWFA中的大部分活动可能因此反映了关于词汇感官后果的跨模态预测。也就是说，VWFA似乎在使用跨模态的词汇性模型生成自上而下的预测。VWFA的元模态性将”解释其将自上而下预测应用于视觉和触觉刺激的能力”(Reich等人, 2011, p. 365)。

另一个很好的例子，这次来自动作领域，由Wolpert, Miall, 和 Kawato (1998)提供，他们注意到即使使用不同的效应器(如右手或左手，甚至脚趾)，个体手写风格的要素仍然得以保持。[1]抽象的高级运动命令必须以不同的方式展开，因为级联预测越来越接近效应器系统本身。但在更高层级，似乎有大量的运动信息以跨效应器的形式编码。

总之，PP框架为适应各种跨模态、多模态和元模态对感知的影响提供了一种强有力的方式。它将感官描绘为共同工作，为一组连接的预测设备提供反馈，这些设备试图跨多个空间和时间尺度追踪世界的展开状态。这提供了对高效多模态线索整合的非常自然的解释，并允许自上而下的效应渗透到感官处理的最低(最早)要素。(如果这在认识论上让你担心——也许因为你怀疑太多的自上而下影响会让我们看到我们期望看到的，而不是”真正存在的”——不要害怕。实际提供的是一个非常微妙的平衡行为，正如我们将在第6章中看到的。)

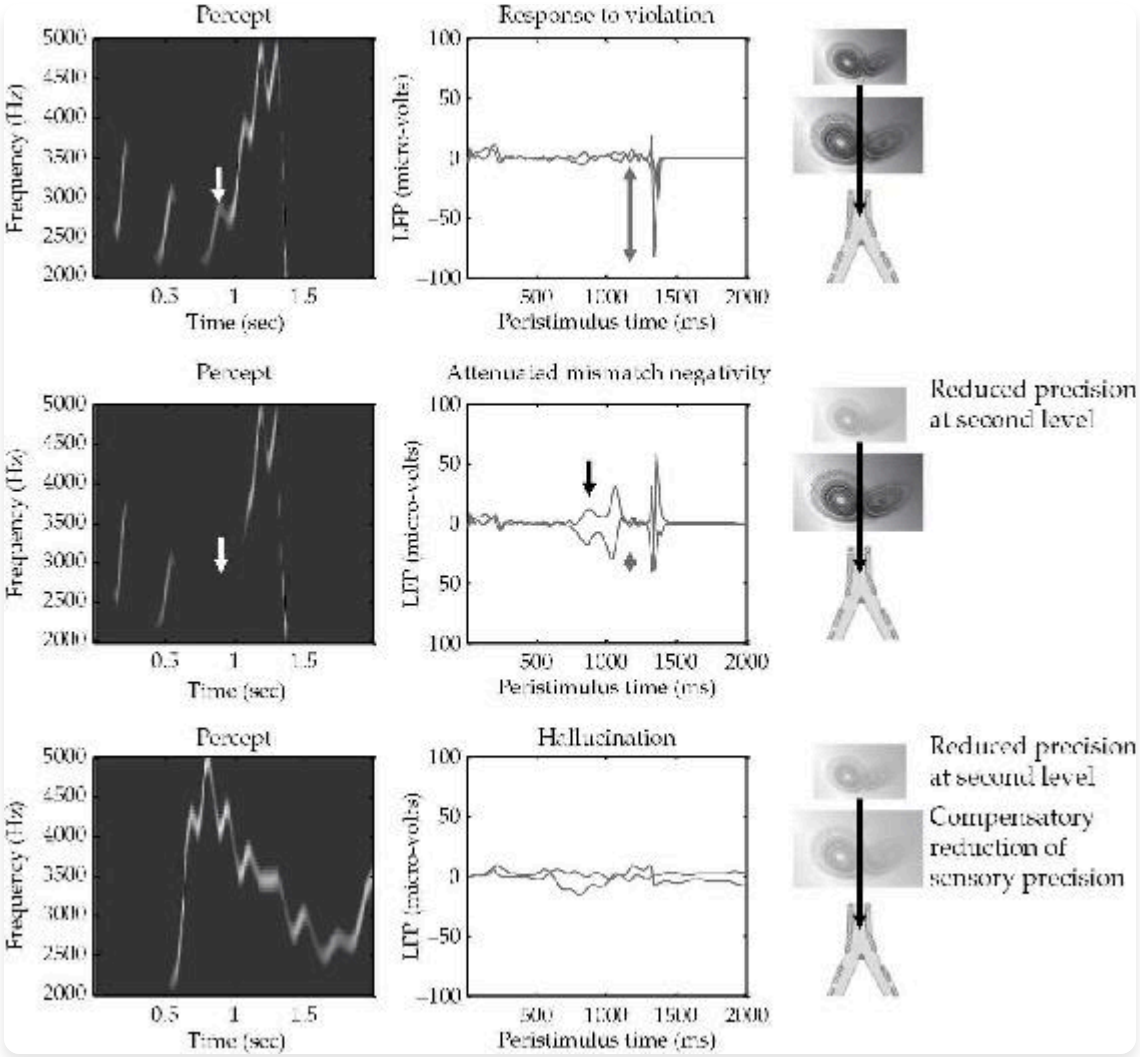
[3.5 感知遗漏]

预测处理理论的另一个优势（如1.14中提到的）是它为”遗漏相关反应”的全谱提供了强有力的解释。这类反应的理论重要性很早就被苏联心理学家尤金·索科洛夫(Eugene Sokolov)在定向反射的开创性研究中注意到——定向反射是环境中意外变化通常引发的立即”注意”反应。索科洛夫注意到重复暴露会导致反应减弱，并将这种效应称为”习惯化”(habituation)。人们可能认为这是由某种形式的低级感觉适应引起的某种粗暴的物理效应。然而，索科洛夫注意到，即使是某个已习惯化刺激强度的*减少*也能引起”去习惯化”(dishabituation)并促发新的反应。索科洛夫得出结论，神经系统必须学习和部署一个”神经元模型”，该模型不断与传入的刺激相匹配，因为现在吸引动物注意的是物理信号本身的减少。

这种情况的极端版本发生在预期信号完全未能实现时。例如，如果我们听到一系列规律的节拍，然后一个节拍被省略，我们在感知上意识到（相当生动地意识到）它的缺失。此外，还有一种熟悉的”几乎体验到”被省略项目开始的感觉——仿佛我们开始听到（或看到，或感受到）那个瞬间后我们生动地注意到并未发生的事物。

那些假设使用生成模型进行”自上而下”处理以通过恰当的期望来迎接传入感觉信号的理论，在解释对遗漏的反应性和遗漏的特殊现象学方面处于理想（也许是独特的）有利地位。Adams等人(2013)使用鸟鸣生成和识别的仿真研究提供了一个令人信服的例子。在这些实验中（见图3.2），一个分层预测处理网络使用前几章描述的多层预测机制，对模拟啁啾声的短序列（显示特征频率和音量的序列）做出反应。然后重复仿真，但省略了原始信号的一部分（最后三个啁啾声）。在第一个缺失的啁啾声处，网络以强烈的预测误差爆发作出反应。作者注意到，这种强烈的误差爆发是在完全没有任何指导性感觉输入的情况下产生的，因为”此时没有感觉输入需要预测，预测误差完全由自上而下的预测产生”(Adams等人，2013，第10页)。此外，对网络反应的更仔细分析显示，在第一个缺失啁啾声本应发生的那一刻，系统产生了一个*瞬时（幻觉性）感知*。这个感知（系统对世界状态的最佳猜测）并不强烈，但时机相

对于缺失的啁啾声是正确的。换句话说，网络首先模糊地”感知”（想象）了缺失的啁啾声，然后在这种信号的实际缺失变得明显时立即以强烈的错误信号作出反应。这些结果很好地模拟了(Adams等人，2013，第10-11页)所谓的”失配负波”(mismatch negativity)——在使用奇异球或省略刺激的EEG研究中发现的P300神经元反应——这一结果在生理学上也是有意义的，因为这类研究对那些最可能涉及报告预测误差的细胞类型（浅层锥体细胞）的反应最为敏感。



image

图3.2 **遗漏相关反应**

左侧面板显示基于后验期望的预测声谱图，而右侧面板显示感觉层面的相关（精度加权）预测误差。顶部面板显示由于精确的自上而下预测在第一个缺失啁啾声未被听到时被违反而产生的正常遗漏相关反应。当第二层的（对数）精度降低到2时，这种反应被减弱（中间行）。这使得自上而下的预测对自下而上的感觉证据更加敏感，感觉预测误差在减少的自上而下约束下得到解决。同时，基于自上而下（经验）先验信念本应被预测的第三个啁啾声被错过，导致感觉预测误差几乎与省略引起的预测误差幅度相匹配。下方一行显示当感觉对数精度从2补偿性地降低到负2时的预测和预测误差。这里，感觉预测误差未能引导高级期望，随后在没有任何刺激的情况下持续存在错误推理。

来源：经Adams, Stephan等人，2013年许可转载。

在进一步的巧妙操作中（再次参见图3.2），Adams等人降低了多层网络上层（第2层）感觉预测误差的精度。正如我们在第2章中看到的，这样做的效果是降低系统对自身自上而下预测的信心。在这些条件下，之前最难检测到的啁啾声（第三个啁啾声）完全被错过，并产生了预测误差。然而，由于系统（第2层精度降低）现在对其预测的信心较低，这个误差不会像正常条件下那样大。作者指出，这可能对应于精神分裂症患者对异常和遗漏刺激的神经元（和行为）反应减弱的现象。对这种反应的解释很有趣，因为它表明

慢性精神分裂症中减弱的不匹配或违反反应可能不是反映未能检测到令人惊讶的事件，而是反映未能检测到不令人惊讶（可预测的）事件。换句话说，它们可能反映了这样一个事实：每个事件都令人惊讶。（Adams et al., 2013，第11页）

系统可能尝试补偿这种泛惊讶性的一种方式是有效地降低其对感觉信号本身的信心。降低感觉信号的估计精度会产生复杂的效果，我们将在后续章节中进一步探讨。在简单的鸟鸣研究中，这种减少导致了完全取消遗漏相关反应，以及从感觉信号正确推断远端环境结构的根本性失败。在这种情况下，听觉遇到的歌声只能被粗略跟踪，结构和频率都出现扭曲。这是不可避免的，因为在这些条件下，“感觉信息没有被赋予约束或引导自上而下预测所需的精度”（Adams et al., 2013，第12页）。这对应于幻觉的产生，这里幻觉作为准知觉状态出现，不足以被自上而下的预测和对我们自身感觉不确定性的恰当估计所控制。

[3.6 期望与意识知觉]

PP模型对研究意识感觉意识的神经基础具有影响（更多内容见第7章）。

我们可以从一些平常的思考开始接近这个话题。直觉上很明显，例如，用劣质收音机播放的熟悉歌曲会比陌生歌曲听起来清晰得多。虽然我们可能在简单的前馈特征检测框架内将此视为某种*记忆*效应，但现在将其视为真正的*知觉*效应似乎同样合理。毕竟，清晰听起来的知觉与模糊听起来的知觉是以同样的方式构建的，只是使用了更好的自上而下预测集合（在故事的贝叶斯翻译中为先验(priors)）。也就是说——我建议——熟悉的歌曲确实听起来更清晰。这不是记忆*稍后*进行某种填补，以回顾的方式影响我们对歌曲听起来如何的判断。相反，自上而下的效应在处理的最早阶段就起作用，让我们几乎没有概念空间（至少在我看来）将这些效应描述为除了增强但真正的知觉之外的任何东西。因此，想象我们发现一种生物，其听觉装置高度调谐于检测某种生物学相关的声音。还想象这种调谐主要由对该声音的强先验组成，使得生物能够在环境信号中有相当大的噪音的情况下检测到它（一种鸡尾酒会效应(cocktail party effect)）。我们肯定会简单地将此描述为敏锐知觉的情况吧？那么，我认为，我们也必须对从低质量收音机听熟悉歌曲的音乐爱好者说同样的话。

当我们逐渐降低驱动信号并提高期望时，我们能避免滑坡效应吗？那个幸运的想象者，其虚构恰好完美地预测了外部世界，根本没有真正*感知*她的世界。她只是一个幸运的猜测者。

两个因素共同拯救我们，使我们不必被迫接受这样的主体进入真正感知者的行列。首先，我们应该考虑反事实(counterfactuals)。如果你只是幸运地认为远端世界目前如预测的那样，那么*如果*世界状态不同，你就无法跟踪它们。这已经区分了幸运预测者与正常预测处理主体。其次，我们必须添加注意力(attention)的可用性。正如我们在上一章中看到的，注意力提高了误差信号各个方面的增益(gain)。这意味着我们确实可以（如果我们决定这样做）专注于劣质收音机声音的模糊性，提高选择性感觉预测误差的增益以揭示声音流的更精细形式。然后PP主体可能会同意收音机已经过时，迫切需要更换。反事实稳健性加上基于注意力的感觉预测误差增益的可用性，从而使我们能够区分”幸运幻觉”与基于真实预测的驱动知觉。

预测在构建有意识感知体验中的作用在Melloni et al. (2011)的研究中得到了很好的展示。Melloni等人表明，形成可报告的有意识感知所需的启动时间会根据我们的期望而变化——换句话说，他们表明期望可以加速有意识觉知。使用脑电图(EEG)特征，计算得出对于预测良好的刺激，有意识感知可以快达100毫秒，因此”可见性的特征并不局限于具有严格潜伏期的过程，而是取决于期望的存在”(Melloni et al., 2011, p. 1395)。Melloni等人建议，这样的结果最好通过分层预测编码框架来解释，在该框架中”有意识感知是假设检验的结果，该检验迭代直到信息在高层和低层区域之间保持一致”(p. 1394)。

[3.7 作为想象者的感知者]

如果预测处理理论是正确的，那么能够形成丰富的、揭示世界的感知的动物，就是理解其世界并准备好想象它们的动物。这个论证很直接。为这类方法提供动力的内部模型的一个重要特征是它们本质上是*生成性的*。也就是说，在上层[4]编码的知识(模型)必须使该层的活动能够预测下层的响应模式。这意味着N+1层的模型在更大系统的背景下运行时，能够为自己生成N层(下层)的感官数据(即，输入在那里被表示的方式)。由于这个过程一直向下应用到试图预测早期处理区域活动的层，这意味着这样的系统完全能够为自己生成感官数据的”虚拟”版本。

从某种意义上说，这并不令人意外。正如Hinton(以及类似观点，见Mumford, 1992)所指出的，“生动的视觉意象、做梦，以及语境对局部图像区域解释的消歧效应…表明视觉系统可以执行自上而下的生成”(Hinton, 2007b, p. 428)。从另一种意义上说，这是相当了不起的。这意味着感知——至少在像我们这样的生物中——与功能上类似于想象的东西共同涌现。我这里所说的”像我们这样的生物”，指的是能够进行丰富的、揭示世界的感知的生物：能够感知由相互作用的隐藏原因构成的复杂远端环境的生物。在我自己的情况下，这样的隐藏原因包括暴雨、樱草花和扑克手牌。在我的两只猫(Bruno和Borat)的情况下，它们似乎包括[5]猫粮、老鼠和飞蛾。我认为，Bruno、Borat和Clark都在部署生成模型来捕获其感官输入在多个空间和时间尺度上的规律性。显然，一个简单的向光源移动的机器人不需要，也可能不应该部署多层生成模型来做到这一点。相反，当系统必须处理以噪声、模糊性和不确定性为特征的领域中复杂的隐藏原因结构时，对生成模型的需求最为明显。

我希望更仔细地为之辩护的主张因此是：能够使用预测驱动学习的特征资源感知复杂外部世界中相互作用原因的动物[6]，将是能够内生生成类感官状态的动物。认为做梦、想象和心理意象因此作为传递我们对结构化(对有机体显著的)外部世界掌握的同一认知包的一部分而变得可用，这似乎并不牵强。这并不意味着每个这样的动物都可以通过某种刻意的意志行为来带来这样的想象。实际上，对于大多数生物来说，刻意想象的行为(我怀疑可能需要通过语言进行自我提示)很可能是根本不可能的。但是这样被赋能感知结构化世界的生物拥有从自上而下生成这些相同感官状态近似值的神经资源。因此，在线感知(由预测处理架构启用)和内生生成准感官状态的能力之间出现了深层的二元性。

[3.8 意象和感知期间的”大脑读取”]

Reddy et al. (2010)的研究中出现了支持这种二元性的强有力的功能性磁共振成像(fMRI)证据。该研究的起点是一组众所周知的结果，表明心理意象和在线视觉感知激活许多相同的早期处理区域(例如，Kosslyn et al., 1995; Ganis et al., 2004)。这样的结果已经被多次复制，并且扩展到包括外侧枕叶皮层(LOC)等区域。这是一个纹外区域，对形状和物体有强烈反应，包括”X”和”O”等字母形式，相对于简单纹理或混乱物体更偏好它们。Stokes et al. (2009)表明，当受试者感知和想象字母”X”和”O”时，LOC都是活跃的。

这些结果为感知和想象之间存在深层计算对偶性的观点提供了直观支持，但它们也与许多较弱的解释相符。它们表明了粗略地理位置的重叠（许多相同区域在在线感知和离线想象及回忆过程中”点亮”），但这尚未确立PP类模型所预测的更深层功能重叠。

Reddy等人的研究直接解决了这个问题，建立在有时被称为”大脑读取”的最新成功基础上。在大脑读取中（例如，Haxby et al., 2001; Kamitani & Tong, 2005; Norman et al., 2006），研究者试图从fMRI数据（追踪血流动力学反应的BOLD信号）中重建刺激的属性，这些数据涉及刺激引发的神经活动。这意味着绘制多体素[7]反应模式并使用它们来推断（解码）产生这些模式的刺激属性。

实验者在这里大致处于生物大脑本身的位置。她的任务——由强大的数学和统计工具使之成为可能——是获取神经激活模式[8]，并仅基于此推断刺激的属性。这些属性范围从识别刺激（通常是图像）所属的类别（例如，它是面孔、水果还是工具？），到选择预定义集合中的哪个特定图像引发了反应，再到（最近且最令人印象深刻的）尽可能实际重建所呈现的图像本身。我们很快就会看到第一种类型的例子。第二种类型（基于fMRI的图像选择）的一个很好例子可以在Kay et al. (2008)中找到，他们能够推断出被试在扫描时感知的是哪个新颖的自然图像（从120个图像集合中）。第三种（主动重建）类型的例子可以在Miyawaki et al. (2008)中找到。

有趣的是，用于执行第三项任务——图像重建任务——的工具和方法越来越像是在重现生物大脑本身使用的策略类型。因此，最有前景的方法使用贝叶斯方法，该方法将测量反应中的信息与关于自然图像结构甚至语义内容的先验信息相结合——例如，见Naselaris et al. (2009)。使用这种先验信息（就像在预测处理中一样）对图像重建的质量产生了巨大而有益的影响。更进一步，van Gerven et al. (2010)使用了在第1章讨论的数字识别例子中使用的架构版本（“深度信念网络”；见Hinton et al., 2006）来从fMRI数据中重建感知的手写灰度数字。作者得出结论（第3139页）：“分层生成模型可用于神经解码，并为了解大脑提供了新的窗口”。

然而，Reddy等人的实验并不涉及图像选择或图像重建。它解决的是更简单的图像分类问题。第一个目标（符合先前的工作）是使用模式分类技术来解码关于*观看*图像的类别信息，确定被试在扫描时是否感知到工具、食物、面孔或建筑物的图像。第二个目标是使用相同的技术来确定被试在扫描时是否在*想象*工具、食物、面孔或建筑物。假设这被证明是可能的，第三个也是最终目标是确定想象物体的体素级”代码”与实际感知到的”相同”物体的代码如何相关。对于解码，实验者使用了一种广为理解的方法（线性支持向量机）来学习体素模式与四个类别（食物、工具、面孔和建筑物）之间的映射。这对感知和想象的物体都进行了，记录既来自早期视觉区域（V1、V2）也来自更高区域（FFA、PPA和一些分布式记录）。

两种形式的解码（解码所看到的和所想象的）都被证明是可能的，尽管——我们很快会回到这一点——从最早的、视网膜拓扑映射区域的解码仅在实际观看期间可能，而在想象期间不可能。相比之下，在腹侧颞皮层，解码在两种条件下（实际观看和想象）都被证明是可能的。然后Reddy等人解决了第三个（对我们来说最有趣的）问题：想象条件下涉及的神经状态与感知条件下涉及的神经状态之间存在什么关系（如果有的话）。这个问题直接关系到我们之前关于感知和想象深层对偶性的推测。

为了解决这个问题，Reddy等人使用了一种巧妙的方法。他们取用了训练好的感知分类器，将其作为表象条件下的解码器，反之亦然（取用训练好的表象分类器来解码在线感知）。令人惊讶的是，每个分类器都适用于另一种条件。换句话说，可以使用”表象解码器”来分类当前观看的项目，也可以使用”感知解码器”来分类仅仅想象的项目。这表明两个任务不只是共享粗略的神经资源，而是共享这些资源的精细使用方式。更具体地说，它显示了在感知场景和仅仅想象场景时，编码这些场景的精细多体素激活模式（在腹侧-颞皮层中）之间存在大量重叠。额外的分析显示，各种体素的作用（它们对给定类别中分类成功的加权贡献）是相似的，两种条件（表象和在线感知）共享关键的”诊断体素”（第6页）。作者得出结论：

模式分类技术的使用…表明实际观看和心理表象在对象响应性腹侧-颞皮层的精细多体素激活模式水平上共享相同的表征[从而证明]感知和自然对象类别表象所涉及的精细表征之间具有高度相似性。（Reddy et al., 2010，第7页）

这些结果为预测处理核心理念提供了强有力的支持，即感知在很大程度上依赖于自上而下的生成能力。

尽管如此，感知和纯粹由自上而下驱动的过程（如心理表象(mental imagery)，也许还有做梦）之间在体验和功能方面显然存在许多差异。Reddy等人研究的另一个方面，前面简要提到过，在这方面很有启发性。尽管在腹侧-颞皮层中证明了感知和表象的重叠编码，但从较早的（V1和V2）视网膜拓扑映射群体进行解码，虽然在感知条件下是可能的，但在表象条件下却不可能。换句话说，只有当受试者实际参与在线观看而不是仅仅想象时，那些早期区域的活动才是fMRI可”读取”的，属于四个图像类别之一。这可能与（正如Reddy等人自己暗示的）心理表象总体上似乎不如在线感知生动和详细（不如真实）这一事实有关。一个可能的解释，与关于V1等区域参与心理表象能力的一系列表面上相当冲突的结果（见，例如，Cui et al., 2007; Wheeler et al., 2000）一致，是可以从自上而下驱动V1，但这只有在任务本身需要精细的想象细节时才会发生。

也许在更典型的情况下，表象（不像丰富的幻觉形式）只涉及生成模型的较高层次？在PP中，以预测误差的精度加权形式，很容易获得调节这种效应的可能机制（见第2章）。为处理的早期（高空间和时间分辨率）阶段计算的预测误差分配低精度，意味着不会花费系统性努力来使这些状态与向下流动的预测保持一致。在这种条件下，系统似乎合理地会生成一个稳定的感知，它简单地忽略较低层次的细节，只有当任务需要时才（通过提高相关的精度权重）引入它们。

在线感知也可能具有特殊特征。可能地，我们可以在非常高的细节水平（粒度）上解决在线感知中的预测误差，比如当我们注意复杂壁纸的精细图案细节或树皮时。这种稳定、丰富的粒度在心理表象的标准情况下可能根本不可用。

然而，其他（“更钝”的）低层次反应可能更容易被引入。Laeng and Sulutvedt (2014)令人惊讶地显示，想象的行为甚至可以影响瞳孔的扩张和收缩。在这项工作中，受试者暴露于不同亮度的三角形图像。在暴露期间，受试者的瞳孔以通常的方式反应，当图像较暗时扩张（变宽），当图像较亮时收缩。当被要求想象同样的三角形时，同样的瞳孔扩张和收缩反应发生了。这个结果是惊人的，因为瞳孔大小是大多数受试者无法施加任何形式的意识控制的东西，导致实验者评论说”观察到的对想象光线的瞳孔调节为心理表象作为基于类似于感知中产生的大脑状态的过程的解释提供了强有力的证据”（第188页）。作者建议，这种反应可能用来为预期的（也许是潜在有害或危险不足的）光线水平准备眼睛。

3.9 梦工厂内部

感知与想象之间如此密切的联系对于梦境也具有启发意义。它们最明显地暗示，梦境状态，就像意象一样，涉及自上而下（基于生成模型）激活许多与普通感知过程中发生的相同状态。然而，这样的观点需要谨慎处理。对于神经系统来说，在缺乏”假设检验行动”的可用性和持续驱动的外部输入的情况下，将无法支持日常感官接触所提供的那种稳定性和丰富的体验细节。

在缺乏驱动性感官信号的情况下，没有关于低级感知细节的稳定持续信息（以可靠的、估计为高精度的预测误差的形式）可用来约束系统，因此没有压力在较低的处理层级创建或维持稳定的假设。相比之下，在清醒状态下，持续存在的外部场景会根据精度期望被反复采样，以提供重要的稳定压力，并帮助创建（正如我们在第2章中看到的）独特的、自我维持的感知。在缺乏可靠感官输入的情况下，对这种低级状态的估计精度将大大降低。由于精度加权涉及提升处理级联的某些方面而不是其他方面，这意味着其他（更高级别）状态的预期精度会增加。因此，整体效果是暂时将展开的内部预测与针对感官状态的现实检验隔离开来。这样，“内部大脑动态就与感觉系统分离开来”（Hobson & Friston, 2012，第87页）。

在睡眠过程中，这一过程伴随着大脑化学状态的一些戏剧性变化。人类大脑的三种主要状态是清醒、REM（快速眼动）睡眠和非REM（NREM）睡眠。每种状态都有明确的生理、药理和体验相关性。在清醒状态下，我们可以处于

许多状态，从闭眼的意象沉思到睁眼警觉地与外部环境接触。在REM睡眠中，我们的梦境（至少根据后续报告证明）是生动的，但其逻辑性很弱。以下是一个典型的报告：

我在一个会议上，试图去吃早餐，但食物和排队的人不断变化。我的腿不能正常工作，我发现举着托盘非常费力。然后我意识到了原因。我的身体正在腐烂，液体从中渗出。我想我可能在一天结束前完全腐烂，但我想如果我还有力气的话，我仍然应该去喝点咖啡。（摘录引自Blackmore, 2004，第340页）

这里是另一个描述，这次来自海伦娜·博纳姆·卡特，当时她正与电影导演蒂姆·伯顿怀孕："我梦见我生了一只冷冻鸡。在我的梦里，我对一只冷冻鸡感到非常满意"（引自Hirschberg, 2003）。在NREM睡眠中，如果我们做梦的话，梦境（同样根据清醒时的报告证明）更像是微弱而平凡的想法或模糊的回忆。所有这些状态（清醒、REM睡眠、NREM睡眠）都与特定的神经化学活动模式相关。显示这种模式的一个有用工具是霍布森的AIM模型（Hobson, 2001）。AIM模型将不同状态描述为三维空间中的点，其轴为：

1. 激活能量

2. 输入源

3. 调节

正常的清醒状态的特征是高激活（例如通过脑电图测量）对应于相当强烈的体验，外部输入源（大脑正在接收和处理来自世界的丰富感官信号流，而不是关闭并主要回收自己的活动），以及独特的模式。这里的调节指的是大脑化学物质之间的平衡，特别是胺类和胆碱类。胺类是神经递质，如去甲肾上腺素和血清素，已知它们的作用对正常的清醒意识至关重要（它们对使我们能够集中注意力、推理和决定行动的过程至关重要）。当这些被关闭，而其他神经递质（胆碱类，如乙酰胆碱）占主导时，我们会体验到妄想和幻觉（如果我们是清醒的）以及生动、不加批判的梦境（如果我们在睡觉）。这样，主要是胺类/胆碱类平衡决定了信号和信息（无论是外部还是内部生成的）将如何被处理和加工。在REM睡眠中，胺能系统被停用，胆碱能系统过度活跃。这是一种高度改变的认知状态。只有极端形式的精神病或严重的医疗或娱乐性药物使用才能在非睡眠的人类中诱发这种状态。

这并不是说（远非如此）人类大脑的最佳状态应该是几乎完全胺能主导的状态。实际上，清醒人类智能的力量、微妙性和美感似乎在很大程度上与两个系统之间不断变化的平衡的精确细节有关。但在正常清醒状态下，模式（定义为两个系统活动之间的比率）倾向于胺能系统。在REM睡眠中，随着乙酰胆碱占主导，体验变得越来越分离，不受感官输入锚定，并超出意志控制。

从预测处理的角度来看，神经调节平衡中这些变化的作用是调节(可能通过精度权重的变化；参见第2章)预测误差的内部流动。这很好地解释了清醒和做梦体验之间截然不同的特征，至少在大体上是如此。因此，

当我们上床闭上眼睛时，感觉预测误差单元的突触后增益下降(通过减少氨能调节)，高级皮质区域误差单元的精度相应增加(由增加的胆碱能神经传递介导)。… 随之而来的睡眠状态是内部预测与感觉约束隔离的状态。(Hobson & Friston, 2012, p. 92)

类似地，Fletcher和Frith建议

也许梦境状态源于层次…处理的中断，使得感觉放电不受自上而下的先验信息约束，而推断由于从低层到高层的预测误差信号的衰减而被毫无质疑地接受。(Fletcher & Frith, 2009, p. 52)

Hobson和Friston (2009, 第4.2.1节)进一步推测，睡眠状态为大脑提供了进行"突触后修剪"的机会——移除冗余或低强度的连接，以降低生成模型本身的复杂性。这里的想法(在第8章和第9章中会有更多论述)是，在清醒和警觉状态下

减少预测误差有时会产生虽然能够捕获感觉模式但过度复杂的模型。这些模型实际上将过多的信号视为数据，而将不够的部分视为噪声。因此它们对特定数据”过度拟合”，从而无法泛化到新情况。

由于刚才描述的平衡改变，睡眠提供了补救这一问题的机会。在睡眠期间，大脑的模型与进一步的感觉测试隔离，但仍可以通过简化和精简得到改善。这是因为大脑实际最小化的量是(正如我们将在第9章中看到的)预测误差加上模型复杂性。在睡眠期间，不会产生精确的预测误差，因此平衡转向减少模型复杂性。因此，睡眠可能允许大脑进行突触修剪，以改善(使更强大和更可泛化)生成模型中体现的知识(参见Tononi & Cirelli, 2006；Gilestro, Tononi, & Cirelli, 2009；Friston & Penny, 2011)。[13] 睡眠与良好认知整理之间的联系是直观的，可能为那些觉得7小时根本不够的人提供特别的安慰！因为如果Hobson和Friston是对的，那么”让大脑离线修剪清醒时建立的过度联系，可能是我们拥有复杂认知系统的必要代价，这种系统能够从感觉样本中提炼出复杂而微妙的联系”(Hobson & Friston, 2012, p. 95)。

[3.10 PIMMS和过去]

到目前为止，我们故事的大部分都集中在使用存储知识来预测可能被认为是一种”滚动现在”的东西。显然，这些预测过程严重依赖于过去的经验。但这种依赖性(目前)并不涉及对过去经验的实际回忆。相反，过去只存在于它结晶成代理不可访问的形式——改变的概率密度分布，用于应对和组织传入的感觉流。然而，像我们这样的生物似乎受益于进一步的技巧。这个技巧是(不时地)能够回忆起可能与手头任务相关的特定具体事件。这里的一个关键接触点是观察到这种”情节回忆”涉及项目与时空情境之间的学习联系。从情境预测项目以及从项目预测情境所涉及的约束和机会，然后提供了工具，当以正确的混合方式部署时，可能实现基于预测的情节记忆重构。

因此考虑Henson和Gagnepain (2010)最近关于多重记忆系统的预测处理解释。Henson和Gagnepain关注的是通常被称为”回忆”和”熟悉感”的对比记忆系统。当受试者面对测试项目时，如果能回忆起他们过去接触该项目的情节情境，就发生了回忆。这样的受试者可能报告原始遭遇的场合和方式或其他周围细节。相比之下，熟悉感出现在受试者无法回忆这些细节，但仍然意识到他们之前遇到过那个特定项目的时候。因此，熟悉感和回忆都不同于(尽管与之交织)仅仅凭借知道对象是什么而存在的语义记忆(例如，“它是一个衣刷”)。回忆和熟悉感似乎(尽管参见Johnson et al., 2009)涉及不同的神经子系统，海马体在前者中起特殊作用，而海马旁回皮质(在内侧颞叶)在后者中起关键作用(参见，例如，Diana et al., 2007)。

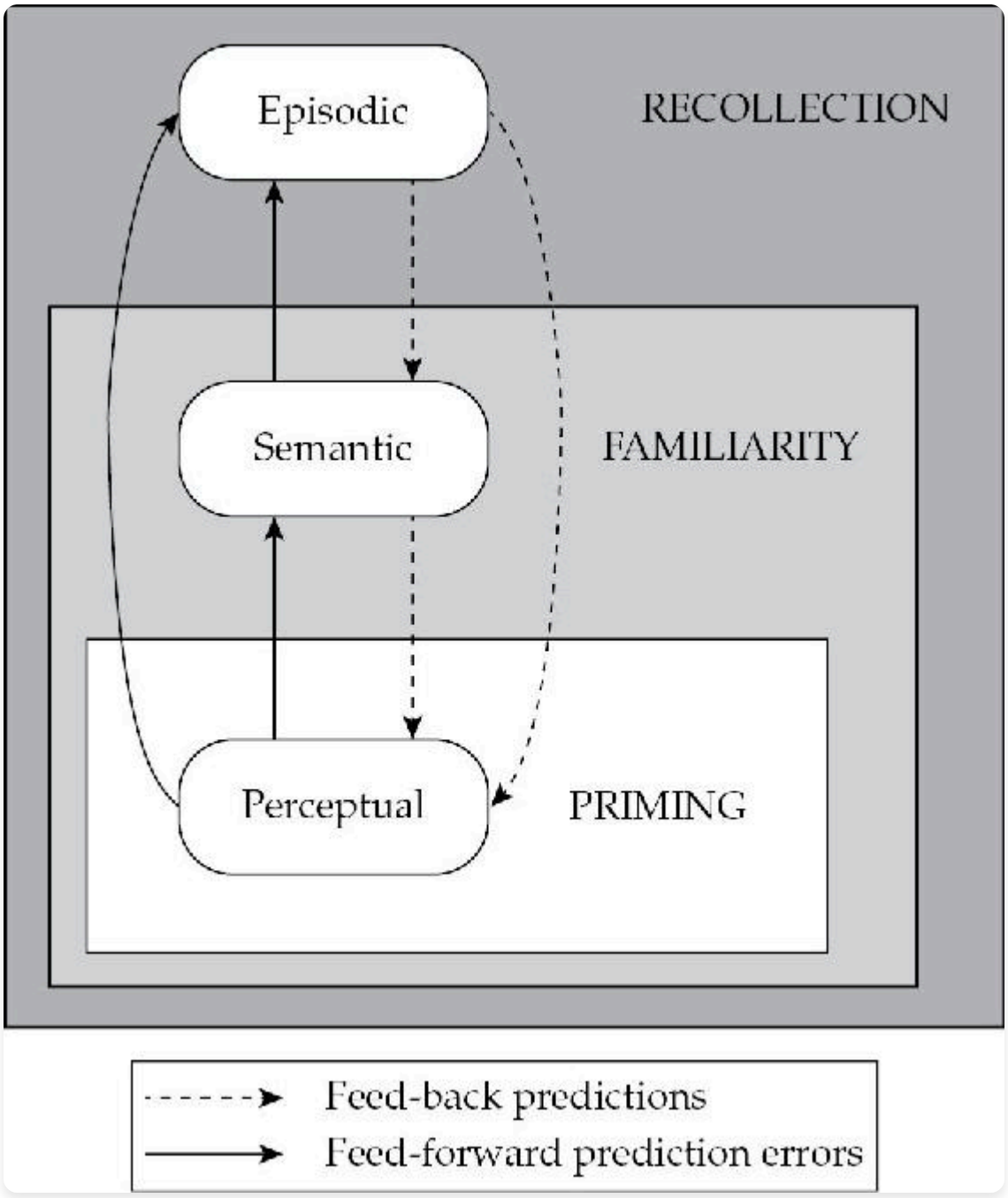
然而，Henson和Gagnepain的核心关注点并非这些不同角色本身，而是区域间相互作用的模式。他们的观点是，不同的相互作用模式（不同的有效连接模式，因此也是不同的功能耦合[14]模式）可以帮助解释与回忆和熟悉性相关的不同行为和神经影像特征。基于此，他们制定并为PIMMS辩护：一个”预测性交互多记忆系统”模型。该模型提出三个”记忆系统”，主要通过它们专门处理的表征内容类型来区分。它们被标记为（遵循Tulving & Gazzaniga, 1995）“情景性”（这里与回忆相关，在生理上与海马体相关）、“语义性”（这里与熟悉性相关，在生理上与嗅周皮质相关）和”知觉性”（与枕颞皮质相关，因此与特定感觉通道如视觉腹侧通路相关）。PIMMS模型的关键创新在于，持续的反馈在编码和检索过程中连接这三个系统，两个阶段的不同递归交互模式解释了行为和生理数据中观察到的差异。

PIMMS将回忆和熟悉性的效应描述为预测处理层次结构内信息流的不同模式所解释的，在这个层次结构中，以现在熟悉的方式”一个系统反馈的作用是预测这个层次结构中’较低’系统的活动”（Henson and Gagnepain, 2010，第1319页）。这个层次结构将海马体置于顶部，嗅周皮质在下方，枕颞皮质在更下方。预测双向层次结构内的不同层级开始专门化（如我们所见）进行不同类型的预测，捕获不同时空尺度的规律性。在这样的架构中，PIMMS模型将海马体描绘为顶层，关注于”优化项目（在嗅周皮质中表征）和情境（presumably在多个区域中表征，取决于情境类型）之间的相互可预测性”（第1321页）。这种优化——这是关键举措——使项目从情境中可预测，情境从项目中可预测。一个在新情境中的熟悉对象因此会引发高预测错误，因为相互可预测性很低。他们认为，海马体预测错误驱

动情景编码，这通过改变海马体与适当的（例如，嗅周）皮质群体之间连接的突触权重来实现。在训练好的层次结构中，反向连接然后允许从情境预测具体内容，而前向流动的错误则驱动编码和检索。

情景和语义记忆系统，如果这是正确的，在相互内部预测的网络中相互连接。在这个网络中，海马体中编码的情境指定信息试图预测嗅周皮质中的基于项目的表征和枕颞皮质中更多的”知觉”表征。预测错误和预测错误解决的不同模式然后实现各种熟悉性和回忆的特征。熟悉性发生在呈现的项目在专门处理项目识别的区域中引发低预测错误（因此高”处理流畅性”，Jacoby and Dallas, 1981）时，但——重要的是，尽管这在PIMMS中没有明确建模——这种流畅性伴随着一种（统计上二阶的）评估，认为这种流畅性是令人惊讶的。[15] 相比之下，回忆发生在项目与特定情境之间存在高相互可预测性时。如果海马体的功能如建议的那样，是优化项目和情境之间的相互可预测性，各种fMRI数据（详见Henson & Gagnepain, 2010，第1320-1322页）也能很好地落入其中。

PIMMS模型既不完整又具有推测性。[16] 我在这里包含它只是作为一些更一般想法和原则的说明。最重要的是，它表明为不同功能（这里是不同种类的记忆）服务的多个、独特神经系统的表面外观可能是微妙误导的。我们可能面对的不是不同系统的简单混合，而是一个统计敏感的相互影响网络，它将情境与内容结合，平衡专门化与整合。在这个网络中，时刻性能依赖于任务特定的有效连接模式的创建和维持（这里连接语义、知觉和情景子系统；见图3.3）。这些模式本身可能源于各种预测错误的估计（任务相关）精确度。通过这种方式，精确度加权预测错误的计算和使用可能构成神经功能的一般原则，不仅服务于驱动和细化知觉识别，还选择和编排整个神经[17]（有时是超神经的；见第三部分）资源集合。



图片

图3.3 **PIMMS记忆模型**

编码、存储和检索是并行和交互的。回忆和熟悉性涉及这些多重记忆系统之间的相互作用。

来源：Henson & Gagnepain, 2010。

3.11 走向心理时间旅行

心理时间旅行（mental time-travel）(Suddendorf & Corballis, 1997, 2007）发生在智能体回忆过去事件或想象未来事件时。Suddendorf和Corballis认为，这种能力可能基于一种更普遍的能力，即使用Hassabis和Maguire (2009, p. 1263)所描述的”大脑的构建系统”来想象体验。这种方法很有吸引力，并且与认知神经科学中的两个融合主题非常吻合。第一个是当代对记忆的观点，即记忆是一个重构过程，当前目标和情境以及先前的回忆事件都极大地影响着被回忆的内容。第二个是大量的成像数据，表明用于回忆过去和想象未来的神经机制之间存在实质性的重叠——尽管并非完全重叠（见Okuda et al., 2003; Szpunar et al., 2007; Szpunar, 2010; Addis et al., 2007）。这种重叠被Ingvar (1985)很好地展现出来，他谈到的”记忆未来”突出了情节记忆中涉及的神经结构在想象可能的未来场景中的作用。正如我们刚才看到的，情节记忆是一种在某种意义上涉及”重新体验”过去经历的记忆（比如当我们记起与邻居的狗的某次特定的、也许痛苦的遭遇时）。它通常与”语义记忆”形成对比（Tulving, 1983），后者涉及概念、特征和属性（狗通常有四条腿，会叫，有各种各样的形状和形式）。语义记忆也根植于我们的过去经验，但它塑造我们对世界的当前把握，而不是在心理上将我们向前或向后传送到时间中。

心理时间旅行到过去和未来共享神经基质的进一步证据来自对记忆障碍的研究。某些形式的失忆症与想象未来的问题相关。Hassabis et al. (2007)报告说，五名海马失忆症患者中有四名在想象新颖事件方面受损——当被要求构建日常场景的新版本时，他们的努力产生的细节较少，而且这些细节缺乏良好组织成连贯的空间结构。Schacter et al. (2007)报告说，回忆中特定的年龄相关恶化模式（情节特定细节的稀疏性；见Addis et al., 2008）与年龄相关的未来思维中的类似模式同步进行。这样的证据使他们为”建构性情节模拟假说”辩护，该假说涉及一个共享的神经系统，支持”将过去事件的细节灵活重组为新颖场景”。他们认为，正是这个面向未来的系统，而不是情节记忆本身，才是适应性价值的真正承载者。他们得出结论，大脑是”一个根本上前瞻性的器官，旨在使用来自过去和现在的信息来生成对未来的预测”（Schacter et al., 2007, p. 660）。这可能是情节记忆易碎、不完整和重构性的深层原因，因为”简单存储死记硬背记录的记忆系统不适合模拟未来事件”（Schacter and Addis, 2007a, p. 27；另见Schacter and Addis, 2007b）。

Schacter和Addis，就像Suddendorf和Corballis一样，对情节记忆与某种形式的”个人的、情节性的”未来思维之间的关系特别感兴趣：在这种思维中，我们通过模拟自己可能的未来经历来在心理上将自己投射到未来。我认为我们现在可以将此标记为从PP视角来看已经显得相当基本的感知、回忆和想象之间对齐的另一个重要而独特的表现。这种对齐直接源于——或者说我一直在论证的——基于预测和生成模型的感知基本视角：这一视角可能因此提供一个更广阔的框架，在其中概念化回忆（各种类型）和想象（各种类型）之间的关系。

更普遍地说，似乎正在出现的是将记忆视为与神经预测和（因此）想象的建构过程密切相关的观点。正如一位记忆理论的领军人物所评论的：

如果记忆是易错的并且容易出现重构错误，那可能是因为它至少与面向未来一样面向过去……类似的神经系统参与自传记忆和未来思维，两者都依赖于一种想象形式。（Fernyhough, 2012, p. 20）

3.12 认知套餐交易

PP提供了一个有吸引力的”认知套餐交易”，其中感知、理解、做梦、记忆和想象都可能作为同一潜在机制策略的变体表达而出现——这种策略是用匹配的自上而下预测来迎接传入的感觉数据。套餐的核心是使用向下连接自我生成类似感知的状态的能力。完全相同的”感知”机制，从自上而下驱动但与驱动感觉信号的牵引隔离，然后解释意象和做梦，并可能为”心理时间旅行”铺平道路，因为我们组合能够重构过去和预构建未来的线索和情境。这也为更深思熟虑的推理形式铺平了道路，正如我们稍后将看到的。

我们核心心理能力之间的密切关系令人瞩目。当过去经验的概率残留遇到传入的感官信号并进行匹配预测时，就会产生感知(丰富的、揭示世界的感知)。这种预测可能是单薄和单一维度的，也可能是丰富结构化的——在许多时间和

空间尺度上捕捉多模态规律性。在其最复杂的表达形式中，它可能涉及丰富时空背景网络的重构(或想象性预构建)。因此，局部的、狭隘的感知逐渐转变为越来越丰富的理解形式，能够支持新的能动性(agency)和选择形式。PP模型并没有在感知和各种认知形式之间做出明确区分，而是假设了自上而下和自下而上影响的不同混合，以及构建预测的内部模型在时间和空间尺度上的差异。[18] 具备这种能力的生物对其世界有一种*结构化把握*：这种把握不在于准语言”概念”的符号编码，而在于用于预测传入感官信号的多尺度概率期望的纠缠质量。

然而，这样的图景在根本上是不完整的。关键任务——我们现在转向的任务——是将预测的神经引擎定位在它们真正所属的地方：嵌套在主动身体的更大组织形式中，并与我们物质、社会和技术世界的变革结构交织在一起。

第二部分

体现预测

4

预测-行动机器

试着感受仿佛你在弯曲小指，同时保持它伸直。一分钟后，它会因为想象中的位置变化而真的感到刺痛；然而它不会明显移动，因为它没有真正移动也是你心中想法的一部分。放下这个想法，纯粹简单地思考运动，释放所有制动，瞧！它毫不费力地发生了。

—威廉·詹姆斯[1]

4.1 保持在突破之前

要冲浪感官刺激的波浪，仅仅预测当下是不够的。相反，我们天生就是要与世界互动的。我们天生就是要以对过去偶然性敏感的方式行动，并积极创造我们需要和渴望的未来。一个猜测引擎(分层预测机器)如何将预测转化为成就？我们将要探索的答案是：通过预测其自身运动轨迹的形状。在解释行动时，我们因此从预测滚动的当下转向预测近未来，以我们自己肢体和身体尚未实际的轨迹形式。预测处理表明，这些轨迹由其独特的感官(特别是本体感受的)后果来指定。正如我们即将探索的，预测这些(非实际的)感官状态实际上有助于实现它们。

这样的预测起到自我实现预言的作用。期望当你移动身体以使冲浪板保持在那个滚动甜蜜点时*将会*产生的感觉流，结果(如果你恰好是专家冲浪者)就是那种流动，将冲浪板定位在你想要的地方。对世界的专家预测(这里是动态不断变化的波浪)与对在该背景下表征所需行动的感官流的专家预测相结合，从而实现该行动。这是一个巧妙的技巧。它与强大而节俭的运动控制计算模型相交，并且具有将在接下来几章中占据我们注意力的扩展和含义。这些扩展和含义一直延伸到对能动性和体验的解释，以及对精神分裂症和自闭症中发现的紊乱或非典型状态的解释。

随着这些关于行动和能动性的解释展开，一件奇怪的事情发生了。曾经看起来像是理解心智和行为竞争模板的方法，现在显现为单一总体认知策略的互补方面。从这个角度重新审视熟悉的主题，我们发现强调具身性(embodiment)、行动和利用身体和环境机会的计算节俭解决方案，从涉及级联推理、内部生成模型和对我们自身不确定性持续估计的预测处理(PP)框架中很自然地浮现出来。这些方法经常被呈现为[2]对人类(和动物)心智深刻对立的愿景。但从所提供的观点来看，它们越来越多地被揭示为单一适应性整体中协调的(和相互协调的)元素。

4.2 痒痒的故事

为什么你不能给自己挠痒？这是Blakemore, Wolpert, and Frith (1998)著名提出的问题。[3] 他们的答案借鉴了大量关于感觉运动学习和控制的先前工作，[4]调用了两个基本元素，每个元素都以不太受限的形式出现在第一部分所追求的感知解释中。

第一个基本要素是（现在已经熟悉的）生成模型概念，这里表现为”运动系统的前向模型”，用于预测自我产生的运动的感觉后果。第二个是”预测编码”提案的一个版本，根据该提案，预测良好的感觉输入的系统性影响被减少或消除。将这两者结合起来处理试图自我挠痒的特殊情况，提出了一个简单但令人信服的模式，其中”自产生触觉刺激的衰减是由于运动系统内部前向模型做出的感觉预测”（Blakemore, Wolpert, and Frith (2000), p. R11）。

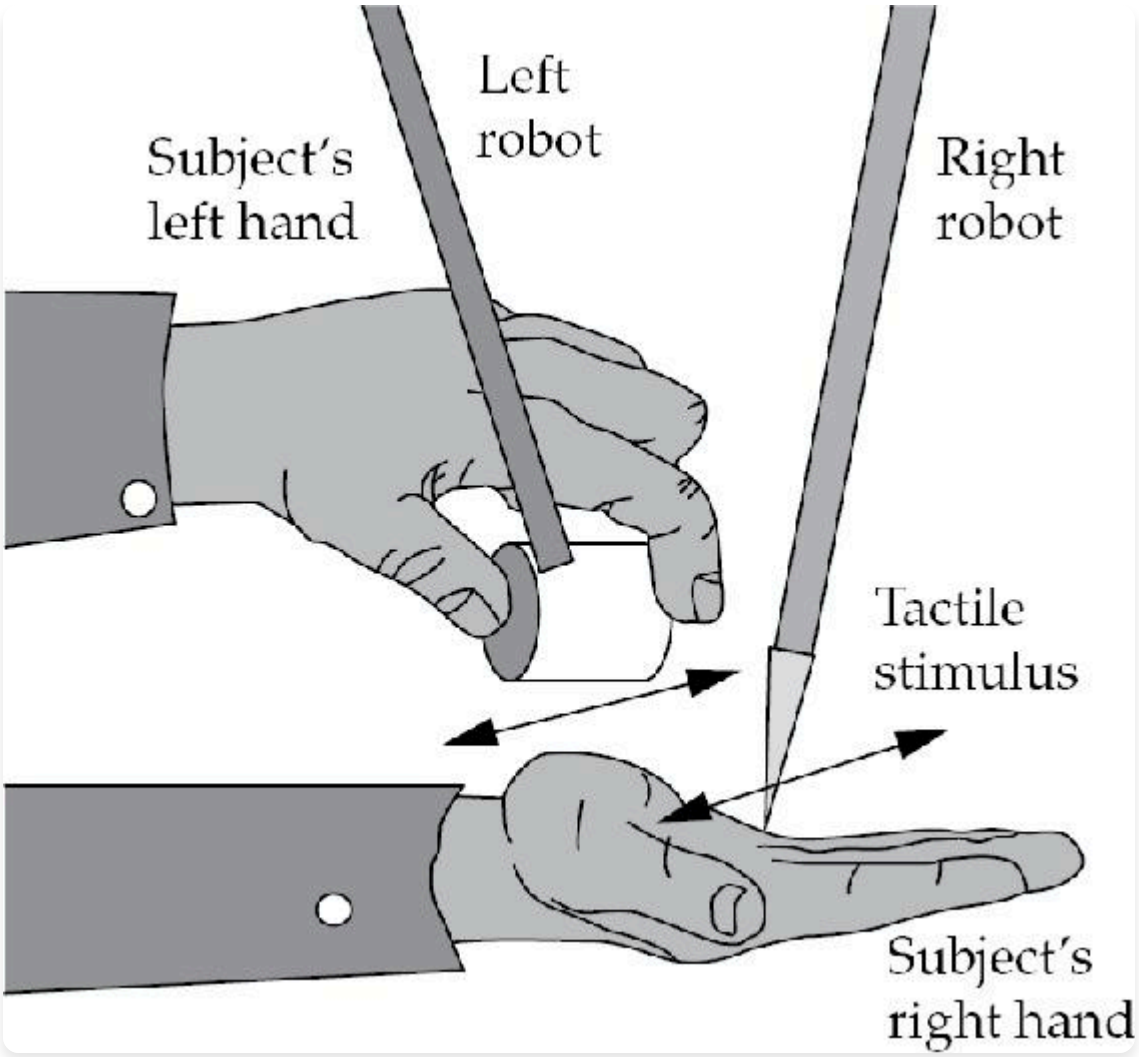
Blakemore等人认为，试图自我挠痒的人掌握着一个”前向模型”，该模型预测她自己运动指令的可能感觉后果。当她开始自我挠痒时，运动指令的副本（称为”传出副本”；Von Holst, 1954）使用前向模型进行处理。该模型捕获（或”模拟”；见Grush, 2004）运动装置的相关生物动力学，能够快速预测来自感觉外围的可能反馈。它通过编码运动指令和预测感觉结果之间的关系来实现这一点。运动指令通过传出副本被捕获，输入前向模型后产生感觉结果的预测（有时称为”随伴放电”）。因此能够比较实际和预测的感觉输入，这些比较为区分自我诱导的运动（其感觉结果将被非常精确地预测）和根植于外部因素和力量作用的感觉效应提供了有用信息的潜在来源。这种比较还能让神经系统抑制甚至消除归因于我们自己自我诱导运动的感觉反馈成分，正如我们感知视觉场景基本稳定时所发生的那样，尽管头部和眼部运动造成了相当大的持续感觉波动（经典讨论见Sperry, 1950）。[5]如果，正如直觉所示，痒感需要一定的惊喜元素（不是关于被挠痒这一事实，而是关于刺激的详细持续形状），我们现在有了解释自我诱导挠痒难以实现的基本框架。

这表明，自我挠痒的障碍类似于给自己讲笑话的障碍：无论多么有趣，笑点永远不会有足够的惊喜。通过部署一个精确的模型来映射我们自己的运动指令到感觉（身体）反馈，我们剥夺了自己以足够不可预测的方式进行自我刺激的能力，并抑制了我们对持续刺激的感觉反应。

这种抑制确实被广泛观察到。例如，一些鱼类产生电场并感知这些电场中指示猎物存在的扰动（Sawtell et al., 2005; Bell et al., 2008）。为了做到这一点，它们需要忽略由自己运动产生的更大扰动。解决方案再次似乎涉及使用预测性前向模型和某种形式的伴随感觉衰减。

同样的一对机制（基于前向模型的预测和对由此产生的预测良好感觉的抑制）被用来解释”力量升级”的令人不安的现象（Shergill, Bays, Frith, & Wolpert, 2003）。在力量升级中，身体交换（游戏场打斗是最常见的例子）通过一种阶梯效应相互升级，其中每个人都相信对方打击得更重。Shergill等人描述的实验表明，在这种情况下，每个人都如实报告了自己的感觉，但这些感觉被自我预测的衰减效应所扭曲。因此，“自产生的力被感知为比相同强度的外部产生的力更弱”（Shergill et al., 2003, p. 187）。这通过实验得到证明，实验中外部设备对受试者的（左食）指尖施加力，然后要求受试者用右食指推压左食指（通过力传感器精确测量施加的力）来匹配刚才受到的力。受试者反复高估了获得匹配所需的力（因此该论文有一个令人难忘的标题”以眼还眼”）。差异是惊人的：“尽管刺激在外围感觉水平上是相同的，但当力是自我产生时，力的感知大约减少了一半”（Shergill et al., 2003, p. 187）。当两个主体进行（他们各自认为的）针锋相对的打击交换，或者其他类型的身体互动时，这种对自我产生力量的感知减弱的滚雪球效应是容易想象的。

在这种情况下，提高准确性的一种方法是要求被试者使用更间接的方法做出反应，从而避开伴随正常身体动作出现的精确前向建模（forward modelling）（和随之而来的感觉抑制）。当要求被试者用手指移动操纵杆来控制力量输出以匹配力量时，被试者能够更好地（Shergill et al., 2003, p. 187）匹配原始力量。类似的操作方法也适用于想要自我挠痒的人。Blakemore, Frith, and Wolpert (1999) 使用了一个机器人接口（如图4.1所示）来插入时间延迟并改变连接被试者自身动作与产生刺激的运动轨迹。随着这些延迟和变化的增加，被试者对产生刺激的”痒痒评级”也随之增加。这些操作试图欺骗精确的前向模型，迫使被试者对不可预测的外部刺激做出反应。



image

图4.1 **实验设置示意图**

一个触觉刺激装置由附着在机器人操纵器末端的泡沫构成，位于被试者右手掌上方。被试者用左手的拇指和食指握住一个圆柱形物体。这个物体直接放在触觉刺激装置上方，并连接到第二个机器人设备。在外部产生触觉刺激的条件下，右侧机器人被编程在被试者的右手上产生正弦波（平滑、重复、振荡）触觉刺激运动。在所有自我产生触觉刺激的条件下，被试者需要正弦波式地移动左手握住的物体，通过两个机器人在右手上方产生触觉刺激的相同运动。可以在左手的运动和右侧机器人的结果运动之间引入延迟和轨迹扰动。

来源：摘自Blakemore, Frith, & Wolpert, 1999。

有趣的是，精神分裂症患者的自我预测感觉的正常抑制被扰乱了。精神分裂症被试者在力量匹配任务上比神经典型者表现更准确（Shergill et al., 2005），也更能够进行”自我挠痒”（Blakemore et al., 2002）。正如我们之前在1.17节中提到的，他们对空心面孔错觉的敏感性也较低。精神分裂症患者感觉衰减的减少可能有助于解释他们出现的各种关于能动性的妄想，比如感觉自己的行为受到另一个代理人控制（Frith, 2005）。由自我产生的运动导致的感觉衰减不足会异常”令人惊讶”，因此可能被错误归因于外部影响。更一般地说，在正常被试者中，感觉衰减的程度与形成妄想信念的倾向之间存在负相关（Teufel et al., 2010）。

4.3 前向模型（巧妙处理时间）

为什么要费心开发前向模型呢？显然不是作为力量升级的进化机制或防御自我挠痒这种可疑做法的手段。相反，使用前向模型对于克服各种信号延迟至关重要，否则这些延迟会阻碍流畅的运动。这是因为：

感觉运动系统的所有阶段都存在延迟，从接收传入感觉信息的延迟，到肌肉对传出运动指令做出反应的延迟。感觉信息的反馈（我们认为包括关于世界状态和我们自身行为后果的信息）受到受体动力学产生的延迟以及神经纤维和突触继电器的传导延迟的影响。

根据Franklin和Wolpert的说法，结果是：

我们实际上生活在过去，控制系统只能获得关于世界和我们自己身体的过时信息，并且不同信息源的延迟各不相同。（两段引用均来自Franklin & Wolpert, 2011, pp. 425 – 426）

前向模型为这些问题提供了强大而优雅的解决方案，使我们能够生活在当下，控制我们的身体（和经过良好练习的工具；见Kluzik et al., 2008）而无需太多持续斗争或努力的感觉。此外，这些模型可以使用开篇章节中回顾的基于预测的学习方案来学习和校准，因为”前向模型可以使用预测误差进行训练和更新，即通过比较运动指令的预测结果和实际结果”（Wolpert & Flanagan, 2001, p. 729）。

最后，为什么使用这些前向模型进行细粒度预测成功的感觉会被抑制，而那些逃脱这种预测的感觉会被增强呢？标准答案（我们之前提到过）是自我预测使我们能够过滤感觉数据的冲击，增强外部产生的数据，在头部和眼部的小幅运动面前提供稳定的知觉，并抑制对更可预测、因此可能生态学紧迫性较低的刺激的反应（见例如Wolpert & Flanagan, 2001）。因此Stafford and Webb (2005)总结了挠痒和力量升级案例揭示的基于模型的抑制效应的进化理由，评论道：

我们的感觉系统不断受到来自环境的感觉刺激轰炸。因此，过滤掉无趣的感觉刺激（比如我们自身动作的结果）是很重要的，这样才能挑选出并关注到具有更多进化重要性的感觉信息，比如有人触碰我们。……预测系统保护着我们，而挠痒可能只是一个意外后果。(Stafford & Webb, 2005, p. 214)

在类似的思路下，Blakemore、Frith和Wolpert总结了基于模型的抑制在自我产生运动中的作用，他们认为：

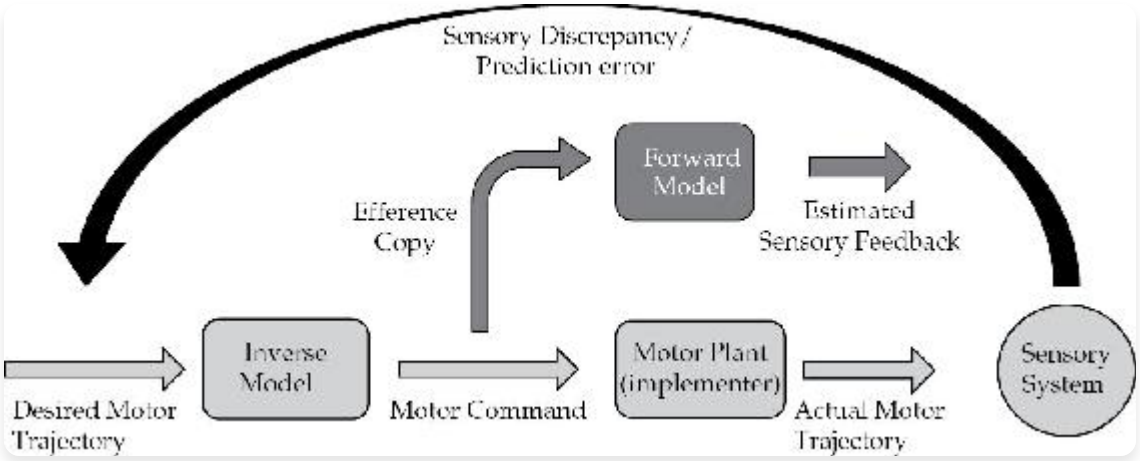
基于预测的调节充当传入感觉信号的过滤器，可以增强传入-再传入比率(akin)（类似于提高信噪比）。这种对传入感觉输入的调节可能具有突出重要特征（例如，由外部事件引起的特征）的效果。(Blakemore, Frith, & Wolpert, 1999, pp. 555-556)

当然，这些都是推动更一般的”预测处理”提案本身的理论依据的版本。该提案以分层生成模型为基础，通过额外的精度加权预测误差策略（见第2章）使其变得灵活，提供了一个更大的框架，能够吸收和重现经典前向模型和运动控制工作中的许多关键见解。但更重要的是，它以一种揭示感知和动作之间更丰富连接网络的方式重现这些见解，并且（正如我们稍后将看到的）修复了我们刚才考虑的解释中的一个显著问题。

这些解释的问题在于，使用来自前向模型的真实预测来减弱预测误差并不能充分解释感觉衰减本身。如果预测误差被来自前向模型的自上而下预测所减弱，那么一旦这些预测到位，感觉刺激仍应在感知上被记录。成功预测我在高度熟悉的场景中扫视时的视觉状态流，并不会以任何方式使我在体验上失明！一个更完整的解决方案（正如我们将在第7章中看到的）不仅仅依赖于前向模型的作用，还依赖于可变精度加权的另一个（较少探索的）效果。然而，目前我们关注的是运动控制本身的一些核心问题。

4.4 最优反馈控制

运动控制，至少在主导的”基于内部模型”的表述中，需要开发和使用的不仅仅是前向模型，还有所谓的*逆模型*(inverse model)(Kawato, 1999)。前向模型将当前运动命令映射到预测的感觉效果，而逆模型（也称为控制器）“执行相反的转换……确定实现某些期望结果所需的运动命令”(Wolpert, Doya, & Kawato, 2003, p. 595)。根据这些”辅助前向模型”(Pickering & Clark, 2014)解释，动作命令向动作的前向模型发送传出副本(efference copy)。在这样的模型中，动作命令作为输入给出，这些命令的预期感觉后果作为输出生成。这个前向模型可能只是涉及一个查找表，但更可能涉及计算（例如，力学定律的近似），这些计算通常在动作执行之前进行。作为一个简单的类比，我把散热器从”关闭”调到一半。在散热器加热之前，我预测（基于对我的中央供暖系统的反复经验）它需要5分钟加热10°C（使用非常简单的方程，例如，每分钟增加2°C，每转动30°）。我可以立即根据预测行动（例如，脱掉外套）或将预测与结果进行比较，并通过我的*逆模型*从任何差异中学习（例如，进一步转动旋钮）。这样的解释（“辅助前向模型”架构，见图4.2）因此假设两个不同的模型：一个逆模型（或最优控制模型），将意图转换为运动命令，以及一个前向模型，将运动命令转换为感觉后果（与实际结果进行比较，用于在线错误纠正和学习）。



image

图4.2 **辅助前向模型(AFM)架构**

在这种架构中，逆模型的输出是运动命令，复制到前向模型，用于估计感觉反馈。

来源：来自Pickering & Clark, 2014。

然而，学习和部署适合某些任务的逆模型通常比学习前向模型更加困难，需要解决复杂的映射问题（将期望的终状态连接到非线性相互作用运动命令的嵌套级联），同时在不同坐标方案之间进行转换（例如，视觉到肌肉或本体感觉，见例如Wolpert, Doya, & Kawato, 2003, pp. 594-596）。

关于”最优反馈控制”的最新研究（综述见Franklin & Wolpert, 2011, pp. 428 – 429）代表了这一框架的精密且成功的发展。它广泛使用所谓的”混合成本函数”(mixed cost-functions)[7]作为从众多（实际上是无限多的）能够达成目标的轨迹或运动中选择一个的手段，并以高效的方式结合前馈和反馈控制策略（一些不错的例子，见Todorov, 2004; Harris & Wolpert, 2006; Kuo, 2005）。特别是，这种策略允许运动的规划和执行同时完成，因为”反馈控制法则用于解决每时每刻的不确定性，允许系统在每个时间点都能最好地响应当前情况”（DeWolf & Eliasmith, 2011, p. 3）。这与更传统的方法不同，传统方法中规划和执行是不同的过程。

反馈控制策略的另一个优势是它识别出一个”冗余子空间”(redundant sub-space)，在其中变异性不会影响任务完成。反馈控制器只会费心纠正那些使系统偏离这个允许变化空间的偏差。这就是Todorov (2009)所谓的”最小干预原则”(minimum intervention principle)。这种系统也能够最大限度地利用它们自身的内在或”被动”动力学(passive dynamics)。我们将在第三部分回到这个话题，但关键点是它们可以将动作的成本计算为系统在有控制信号和没有控制信号情况下会做什么（运动工厂如何表现）之间的差异。完善这一系列优点，该范式的扩展允许组合预学习的控制序列来处理新情况，通过”快速且廉价地从之前学习的最优运动中创建最优控制信号”（DeWolf & Eliasmith, 2011, p. 4）。结果是一种预学习运动命令的组合语法。在分层设置中运行（其中更高层级编码轨迹和可能性的压缩表示），这种系统能够使用高效且可重组的神经资源控制极其复杂的行为。形式上，最优反馈控制理论（特别见Todorov & Jordan, 2002; Todorov, 2008）将运动控制问题显示为在数学上等价于贝叶斯推理(Bayesian inference)（见[附录1]）。非常粗略地说——再次，详细说明见Todorov 2008——你将期望的（目标）状态视为已观察到的，并执行贝叶斯推理来找到能让你到达那里的动作。对我们的目的而言，这里关于贝叶斯推理重要的只是它是一种概率推理形式，考虑了数据的不确定性，将其与关于世界和运动系统的先验信念（由生成模型编码）相结合，以便提供（这里，相对于某个成本函数）最优控制（见，例如，Franklin & Wolpert, 2011, pp. 427 – 429）。

感知和动作之间的这种映射也出现在一些关于规划的最新工作中（例如，Toussaint, 2009）。这个想法与这些简单运动控制方法密切相关，即在规划中我们将未来目标状态想象为实际的，然后使用贝叶斯推理来找到让我们到达那里的中间状态集合（现在它们本身可以是完整的动作）。因此正在出现一套根本统一的计算模型，正如Toussaint (2009, p. 28)评论的，“不区分传感器处理、运动控制或规划的问题”。这种理论表明感知和动作在某种深层意义上是计算上的兄弟姐妹，并且：

通过感知解释传入信息的最佳方式，与通过运动动作控制传出信息的最佳方式在深层上是相同的……所以认为有一些可指定的计算原则支配神经功能的观念似乎是合理的。（Eliasmith, 2007, p. 380）

[4.5 主动推理]

第一部分中介绍的PP模型与这些新兴的动作和运动控制方法非常自然地结合在一起（同时提出了一些具有启发性的转折）。[8]最优反馈控制的研究利用了运动系统（像视觉皮层一样）显示复杂层次结构的事实。这种结构允许复杂的行为在更高层级以紧凑的方式被指定，其含义可以在较低层级逐步展开。然而，直观的区别是，在运动控制的情况下，我们想象信息向下流动，而在视觉皮层的情况下，我们想象信息向上流动。在这种直观图景中，视觉接收复杂的能量刺激并将其映射到越来越紧凑的编码上，而运动控制则接收某种紧凑编码并逐步将其展开为复杂的肌肉命令集。传统图景表明，运动皮层中的下行通路在功能上应该对应于视觉皮层中的上行通路。然而，情况并非如此。在运动皮层内，向下的连接（下行投射）在”解剖学和生理学上更像视觉皮层中的反向连接，而不是相应的前向连接”（Adams等，2012，第1页）。这很有启发性。我们可能想象层次运动系统的功能解剖结构是感知系统的某种镜像，但两者似乎更加紧密对齐。[9]PP表明，解释是向下的连接在两种情况下都承担着本质上相同的任务：预测感觉刺激的任务。

正如我们在第一部分中看到的，PP已经颠覆了关于感知的传统图景。紧凑的高层编码现在是试图预测感觉表面上能量变化的装置的一部分。PP表明，同样的故事也适用于运动情况。区别在于运动控制在某种意义上是*假设性的*。它涉及预测如果我们执行某个期望动作将会产生的非实际本体感觉轨迹(proprioceptive trajectories)。然后减少对这些非实际状态计算的预测误差（以我们即将探索的方式）使它们变为实际。我们预测自己动作的本体感觉后果，这带来了动作的发生。

结果是，在运动和感觉皮层中，向下（和横向）的连接都携带复杂的预测，向上的连接携带预测误差。这解释了运动皮层的功能回路似乎没有相对于感觉皮层倒置这一”矛盾”（Adams, Shipp, & Friston, 2013，第611页）事实。相

反，运动和感觉皮层之间的区别被消解了——两者都在从事自上而下的预测业务，尽管它们预测的事物类型（当然）是不同的。运动皮层在这里最终出现为一个多模态感觉运动区域，在本体感觉和其他模态中发出预测。

核心思想（Friston, Daunizeau等，2010）因此是生物智能体可以通过两种方式减少预测误差。第一种（如第一部分所见）涉及找到最好地适应当前感觉输入的预测。第二种是通过*执行使我们的预测成真的动作*——例如，四处移动和采样世界，以产生或发现我们预测的感知模式。这两个过程可以使用相同的计算资源来构建（我们将看到）。在正常情况下，它们无缝地协同工作，正如第2章（2.6到2.8）中关于注视分配讨论的微观缩影所示。结果是：

感知和运动系统不应被视为独立的，而应被视为试图在所有域中预测其感觉输入的单一主动推理机器：视觉、听觉、体感、内感受，以及在运动系统的情况下，本体感觉。（Adams, Shipp, & Friston, 2013，第614页）

“主动推理”（Active Inference）(Friston, 2009; Friston, Daunizeau, et al., 2010)是感知和运动系统共同减少预测误差的联合机制，它采用两种策略：改变预测以适应世界，以及改变世界以适应预测。这个一般架构也可以——也许更透明地——被标记为”面向行动的预测处理”（action-oriented predictive processing）(Clark, 2013)。在运动行为的情况下，关键的驱动预测现在具有假设性色彩。Friston及其同事建议，它们是对本体感受模式的预测，如果执行该动作，这些模式就会随之而来。“本体感受”（Proprioception）是告知我们身体各部位相对位置以及正在施加的力量和努力的内在感觉。它要与外感受（即标准感知）通道区分开来，如视觉和听觉，也要与告知我们饥饿、干渴和内脏状态的内感受通道区分开来。关于后者的预测在我们后来考虑情感和情绪的构建时将发挥重要作用。然而，目前我们关注的是简单的运动行为。为了实现这种行为，运动植物以取消本体感受预测误差的方式运作(Friston, Daunizeau, et al., 2010)。这之所以有效，是因为本体感受预测误差信号表示身体植物当前配置与执行期望动作时配置之间的差异。因此，本体感受预测误差将持续存在，直到运动植物的实际配置能够产生（逐时刻地）预期的本体感受输入。通过这种方式，对与某个动作执行相关联的展开本体感受模式的预测实际上带来了该动作。这种场景被Hawkins and Blakeslee (2004)恰当地描述，他们写道：

听起来很奇怪，当涉及你自己的行为时，你的预测不仅先于感觉，它们还决定感觉。思考进入序列中的下一个模式会引起对你应该接下来体验什么的级联预测。随着级联预测的展开，它产生了实现预测所必需的运动命令。思考、预测和行动都是沿着皮质层次结构向下移动的相同序列展开的一部分。(Hawkins & Blakeslee, 2004, p. 158)

然而，Friston及其同事更进一步，建议（精确的）本体感受预测直接引发运动行为。这意味着运动命令已被本体感受预测所替代（或者我更愿意说，由本体感受预测实现）。根据主动推理，主体以主动寻求其大脑期望的感觉后果的方式移动身体和传感器。感知、认知和行动——如果这个统一的观点被证明是正确的——共同工作，通过选择性采样和主动塑造（通过运动和干预）刺激阵列来最小化感觉预测误差。

这消除了感知和行动控制之间的任何基本计算界限。当然，在适应方向上仍然存在明显（且重要）的差异。这里的感知将神经假设与感觉输入相匹配，涉及”预测现在”，而行动使展开的本体感受输入与神经预测保持一致。正如Anscombe (1957)著名地评论的那样，[10]这种差异类似于查阅购物清单来选择要购买的物品（从而让清单决定购物篮的内容）与列出一些实际购买的物品（从而让购物篮的内容决定清单）之间的差异。但尽管在适应方向上存在这种差异，潜在神经计算的形式现在被揭示为相同的。实际上，根据这个解释，运动皮质和视觉皮质之间的主要差异更多在于预测什么样的东西（例如，运动轨迹的本体感受后果）而不是如何预测。结果是：

初级运动皮质不比纹状体（视觉）皮质更多或更少地是运动皮质区域。运动皮质和视觉皮质之间的唯一差异是，一个预测视网膜拓扑输入，而另一个预测来自运动植物的本体感受输入。(Friston, Mattout, & Kilner, 2011, p. 138)

这里感知和行动遵循相同的深层逻辑，并使用相同计算策略的版本来实现。在每种情况下，系统性命令保持相同：减少持续的预测误差。在感知中，当自上而下的级联成功匹配传入的感觉数据时，这种情况就会发生。在行动中，当物理运动通过产生产生某些预测本体感受状态序列的轨迹来消除预测误差时，这种情况就会发生。因此，行动显

现为一种自我实现的预言，其中神经回路预测所选动作的感觉后果。然而，这些后果不会立即获得，所以预测误差随之而来：然后通过移动身体以产生预测的感觉序列来消除这种误差。

然而，这些表达方式可能会让人觉得感知和行动是分别展开的，各自忙于追求自己的匹配方向。这是一个错误的理解。相反，PP智能体不断尝试通过招募一个感知和适当世界参与行动的交织网格来适应感觉流。如果这是正确的，我们的感知并不是关于世界的行动中立的”假设”，而是持续尝试以适合参与世界的方式解析世界。可以肯定的是，并非所有预测误差都能通过行动来*解决*——有些必须通过更好地理解事物的本质来解决。但这种练习的目的是让我们以一种能够选择更好行动的方式与世界接触。这意味着即使是感知层面的东西也深深地具有”行动导向性”。这并不令人惊讶，因为感知推理的唯一目的就是规定行动（这会改变感觉样本，从而影响感知）。因此，我们遇到的是一个由行动可供性构成的世界。这将在文本的其余部分更清楚地显现出来。目前要注意的是，即使在所谓的”感知”情况下，预测误差最好也被视为编码了尚未被利用来控制适当世界参与行动的感觉信息。

4.6 简化控制

这些基于预测的行动控制方法与前面描述的关于前向模型和最优反馈控制的重要工作分享许多关键见解。共同点是核心强调基于预测的学习，学习一个能够预期行动感觉后果的前向（生成式）模型。共同点还有（正如我们稍后将更详细地看到的）对代理体验的独特视角：这种视角，正如”挠痒痒故事”已经开始暗示的那样，部分追溯到预测和实际感觉流之间匹配的精妙性。但是由Friston和其他人正在开发的主动推理(active inference)（见，例如，Friston, 2011a; Friston, Samothrakis, & Montague, 2012）在两个关键方面与这些方法不同。

首先，主动推理摒弃了逆模型或控制器，同时也摒弃了对运动命令传出副本的需要。其次，它摒弃了对成本或价值函数作为强制速度、准确性、能量效率等手段的需要。[11] 这一切听起来相当戏剧性，但在实践中主要相当于现有职责的重新分配：一种重新分配，其中成本或价值函数被”折叠”到同时规定识别和行动的上下文敏感生成式模型中。尽管如此，这种重新分配在概念上是有吸引力的。它与来自现实世界机器人技术和情境行动研究的重要见解完美契合，可能有助于我们更好地思考行动选择和运动控制复杂问题的解决方案空间。

行动在这里被重新概念化为关于运动轨迹期望（跨越多个时间和空间尺度）的直接后果。结果是”环境导致关于运动的先验信念……而这些信念导致采样环境”（Friston & Ao, 2012, p. 10）。这种方法突出了一种循环因果关系，它将智能体所知道的（在生成式模型中出现的概率”信念”[12]）与选择输入以确认这些信念的行动联系起来。我们的期望在这里”导致采样环境”，正如Friston和Ao所说，但只是在形而上学无害的意义上，即驱动选择性揭示预测感觉刺激的行动。

正是通过这种方式，智能体通过行动召唤出她所知道的世界。[13] 正如我们将在第三部分中看到的，这使得行动导向的预测处理与自组织动力系统的工作产生密切而富有成效的联系，为所谓”能动论”(enactivist)愿景的核心要素提供了新的视角：一种心智是它们所揭示世界的积极构造者的愿景。在短时间尺度上，这就是前面描述的主动采样过程。我们以反映并寻求确认构成采样的世界把握的方式采样场景。这是一个只有”适合的”假设（假设在适当的行动导向意义上理解）才能生存的过程。在更长的时间尺度上（见第三部分），这是我们构建设计师环境的过程，这些环境安装新的预测，决定我们如何行为（我们如何采样那个环境）。因此，我们构建世界，这些世界构建心智，这些心智期望在那种世界中行动。

然而目前最重要的是要注意到，前向运动模型现在仅仅是一个更大更复杂的生成模型的一部分，该模型将预测与其感觉后果联系起来。这里的运动皮层指定的不是传统意义上理解的运动命令，而是运动的感觉后果。这里特别重要的是关于*本体感觉*感觉后果的预测，这些预测隐含地最小化了各种能量成本。在层次化自上而下处理的完整级联作用下，一个简单的运动命令展开为关于本体感觉效应的复杂预测集合。这些驱动行为，并使我们以当前获胜”假设”所规定的方式对世界进行采样。这样的预测可以在更高层次上，用外在（以世界为中心、以肢体为中心）坐标

指定的期望状态或轨迹来表述。这是可能的，因为所需的向内在（基于肌肉）坐标的转换随后被下放给本质上是经典反射弧的机制，这些反射弧被设置来消除本体感觉预测误差。因此：

如果运动神经元被连接来抑制脊髓背角的本体感觉预测误差，它们实际上实现了一个逆模型，从期望的感觉后果映射到内在（基于肌肉）坐标中的原因。在这种对传统方案的简化中，下行运动命令变成了对由初级和次级感觉传入神经传递的本体感觉的自上而下预测。(Friston, 2011a, p. 491)

对独特逆模型/最优控制计算的需要现在似乎已经消失了。取而代之的是一个更复杂的前向模型，它将关于期望轨迹的先验信念映射到感觉后果，其中一些（“底层”本体感觉）使用经典反射弧自动实现。而且，如前所述，在这些方案中也不需要传出拷贝(efference copy)。这是因为下行信号已经（就像在感知情况下一样）在预测感觉后果的业务中。所谓的”*推论放电*”（编码预测的感觉结果）因此是普遍存在的，并且渗透在向下的级联中，因为”大脑中的每个向后连接（传递自上而下预测）都可以被视为推论放电，报告某种感觉运动构造的预测”(Friston, 2011a, p. 492)。

4.7 超越传出拷贝

这个提议在初次遇到时，可能会让读者觉得相当激进。对传出拷贝功能意义的理解不是当代认知和计算神经科学的主要成功故事之一吗？事实上，通常被认为支持该解释的大部分（也许是全部）证据，在更仔细的检查下，仅仅是前向模型和推论放电的普遍和关键作用的证据——也就是说，这是对传统故事中那些被保留（实际上被预测处理(PP)使得更加核心）部分的证据。

例如，Sommer和Wurtz具有影响力的(2008)综述论文，其焦点是允许我们区分自己运动的感觉效应和环境变化导致的效应的机制，很少提及传出拷贝本身。相反，它广泛使用了推论放电这一更一般的概念——尽管正如这些作者也注意到的，这两个术语在文献中经常互换使用。一篇更近期的论文，Wurtz et al. (2011)，只提到传出拷贝一次，然后也只是将其与推论放电的讨论合并（推论放电在文本中出现了114次）。同样，有充分的理由相信（正如标准故事所表明的）小脑在这里起着特殊作用，而且这个作用涉及制作或优化关于即将到来的感觉事件的感知预测(Bastian, 2006; Roth et al., 2013; Herzfeld & Shadmehr, 2014)。但这样的作用当然完全符合预测处理的图景。我认为，道德是前向模型和推论放电的一般概念，而不是我们之前定义的传出拷贝这一更具体的概念，目前享有来自实验和认知神经科学最清晰的支持。

当然，传出拷贝在一套特定的计算提议中占据突出地位。这些提议涉及（本质上）前向模型和推论放电在一个假定的更大认知架构中的定位，该架构涉及多个配对的前向和逆模型。在这些”配对前向-逆模型”架构中（见例如Wolpert & Kawato, 1998; Haruno, Wolpert, & Kawato, 2003），运动命令被复制到一个独立前向模型堆栈中，用于预测动作的感觉后果。但正如即使是其最强支持者也承认的那样，获取和部署这样的架构提出了各种极其困难的计算挑战（见Franklin & Wolpert, 2011）。预测处理的替代方案巧妙地避开了许多这些成本。

PP提议认为，预测感官后果的一个子集（预测的本体感觉轨迹）已经作为运动命令发挥作用。因此，不存在独立的运动命令需要复制，也不存在传出拷贝(efference copies)本身。但我认为，我们同样可以将那些基于前向模型的本体感觉轨迹预测描述为”隐式运动命令”：这些运动命令（本质上——下文将详述）通过指定结果而非指定精细的肢体和关节控制来运作。这些隐式运动命令（本体感觉预测）也会影响对即将到来的动作的外感官感觉后果的更广泛预测。

通常归因于传出拷贝作用的大部分功能因此得以保留，包括基于前向模型对核心现象的解释，如时间延迟的精细调节(Bastian, 2006)和眼球运动时视觉世界的稳定性(Sommer & Wurtz, 2006, 2008)。不同之处在于，通常由传出拷贝、逆模型和最优控制器完成的繁重工作现在转移到预测（生成）模型的获取和使用上（即，获得正确的先验概率”信

念”集合）。如果（但仅当）我们能够合理地假设这些信念”在分层感知推理过程中作为自上而下或经验先验自然涌现”(Friston, 2011a, p. 492)，这种转变就具有潜在优势。因此，计算负担转移到获取正确的先验集合（这里指轨迹和状态转换的先验），也就是说，负担转移到获取和调节生成模型本身。

4.8 摒弃成本函数

第二个重要差异（与”最优反馈控制”模式相比）是主动推理避开了将成本或价值函数作为选择和塑造运动反应手段的需要。它再次通过本质上将这些折叠到生成模型中来实现这一点(Friston, 2011a; Friston, Samothrakis, & Montague, 2012)，该生成模型的概率预测与感觉输入结合产生行为。

可能应用于塑造和选择运动行为的成本或价值函数的简单例子包括最小化”急动度”（某行为期间肢体加速度的变化率）和最小化扭矩变化率（关于这些例子，分别见Flash & Hogan, 1985和Uno et al., 1989）。如前所述，最优反馈控制的最新研究最小化更复杂的”混合成本函数”，这些函数不仅涉及身体动力学，还涉及系统噪声和所需结果的准确性（见Todorov, 2004; Todorov & Jordan, 2002）。

如Friston (2011a, p. 496)所观察到的，这些成本函数有助于解决困扰经典运动控制方法的多对一映射问题。使用身体实现某个目标有很多方式，但动作系统必须从中选择一种方式。然而，在所提供的框架内不需要这样的设备，因为”在主动推理中，这些问题通过关于轨迹的先验信念（可能包括最小急动度）得到解决，这些信念唯一地确定了（外在）运动的（内在）后果”(Friston, 2011a, p. 496)。因此，简单的成本函数被折叠到决定运动轨迹的期望中。

但故事并未就此结束。因为这里使用的相同策略适用于各个层面的期望后果和奖励概念。因此我们读到”关键的是，主动推理不调用任何’期望后果’。它仅依赖于经验依赖的学习和推理：经验诱发先验期望，这些期望指导感知推理和行动”(Friston, Mattout, & Kilner, 2011, p. 157)。除了以向下流动的预测形式出现的某种推论放电的繁盛之外，我们在这里似乎面对着某种荒漠景观：一个价值函数、成本、奖励信号，甚至可能是欲望都被复杂的相互作用的期望所取代的世界，这些期望指导感知并引导行动。但我们也可以说（我认为这是表达这一点的更好方式）奖励和成本函数的功能现在只是被吸收到更复杂的生成模型中。它们隐含在我们的感官（特别是本体感觉）期望中，通过规定其独特的感官含义来约束行为。

内在奖赏的”趋向性”刺激(以最明显的例子)因此不应该从我们的本体论中被消除——相反，它们被简单地重新概念化为一旦被识别，就”引发强制性意志和自主反应”的刺激(Friston, Shiner, et al., 2012, p. 17)。从概念上讲，这里重要的是行为被描述为我们的信念(亚个人的概率期望网络)与环境相互作用的结果。奖赏和愉悦然后是这些相互作用中某些的结果，但它们不是(如果故事的这个——诚然相当具有挑战性的——部分是正确的)这些相互作用的原因。相反，是复杂的期望驱动行为，使我们以可能经常带来奖赏或愉悦的方式探索和采样世界。通过这种方式”奖赏是行为的感知(享乐)结果，而不是原因”(Friston, Shiner, et al., 2012, p. 17)。

请注意，这种职责重新分配不会获得整体计算优势。事实上，Friston本人明确表示：

用先验信念替换成本函数没有免费午餐[因为]众所周知[Littman et al., 2001]当将问题表述为推理问题时，问题的计算复杂性并不会降低。(Friston, 2011a, p. 492)

尽管如此，这种重新分配(其中成本函数被视为先验)很可能在概念上和策略上具有重要后果。例如，使用关于路径和轨迹的先验信念来指定整个路径或轨迹是很容易的！相比之下，标量奖赏函数指定点或峰值。结果是成本函数可以指定的一切都可以通过轨迹先验来指定，但反之则不然，并且(更一般地说)成本函数可以有用地被视为结果而不是原因。

相关的关切已经导致许多从事机器人学工作的人认为明确的基于成本函数的解决方案是不灵活的和生物学上不现实的，应该被那些以隐式利用具身智能体(embodied agents)复杂吸引子动力学的方式吸引行动的方法所取代(参见，例如，Thelen & Smith, 1994; Mohan & Morasso, 2011; Feldman, 2009)。大致想象这类广泛解决方案的一种方式(更长的讨论

见Clark, 2008, 第1章)是思考你如何通过简单地移动连接到特定身体部位的线来控制木制牵线木偶。在这种情况下，“关节间运动的分布是…施加到末端执行器的力和不同关节’顺应性’的’被动’结果”(Mohan & Morasso, 2011, p. 5)。这种解决方案旨在(与PP一致)“规避运动学逆转和成本函数计算的需要”(Mohan, Morasso, et al., 2013, p. 14)。作为原理验证，Mohan, Morasso等人在使用类人iCub机器人的一系列机器人仿真中实施和测试了他们的想法，注意到在这些实验中行动本身是由一种内部前向模型驱动的仿真推动的。所有这些都表明PP方法与追求计算节俭的运动控制手段之间存在诱人的汇合。我们将在第8章中对这种故事有更多要说的。

最大限度利用学习或内置”协同效应”(synergies)和身体装置复杂生物力学的解决方案可以使用主动推理和(基于吸引子的)生成模型的资源非常流畅地实施(见Friston, 2011a; Yamashita & Tani, 2008)。例如，Namikawa et al. (2011)展示了具有多时间尺度动力学的生成模型如何实现流畅和可分解的(另见Namikawa & Tani, 2010)运动行为集合。在这些仿真中：

行动本身是符合…关节角度本体感受预测的运动的结果[并且]…感知和行动都试图在整个层次结构中最小化预测误差，其中运动最小化了本体感受感觉水平的预测误差。(Namikawa et al., 2011, p. 4)

另一个例子(我们之前简要遇到过的)是使用向下流动的预测来避免将期望的运动轨迹从外在(以任务为中心)坐标转换为内在(例如，以肌肉为中心)坐标的需要：据说这是一个复杂且不适定的”逆问题”(Feldman, 2009; Adams, Shipp, & Friston, 2013, p. 8)。在主动推理中，指导运动行为的先验信念已经将在外在参考框架中(在高层次)表述的预测映射到在肌肉和效应器上定义的本体感受效应，这仅仅是普通在线控制的一部分。通过这种方式：

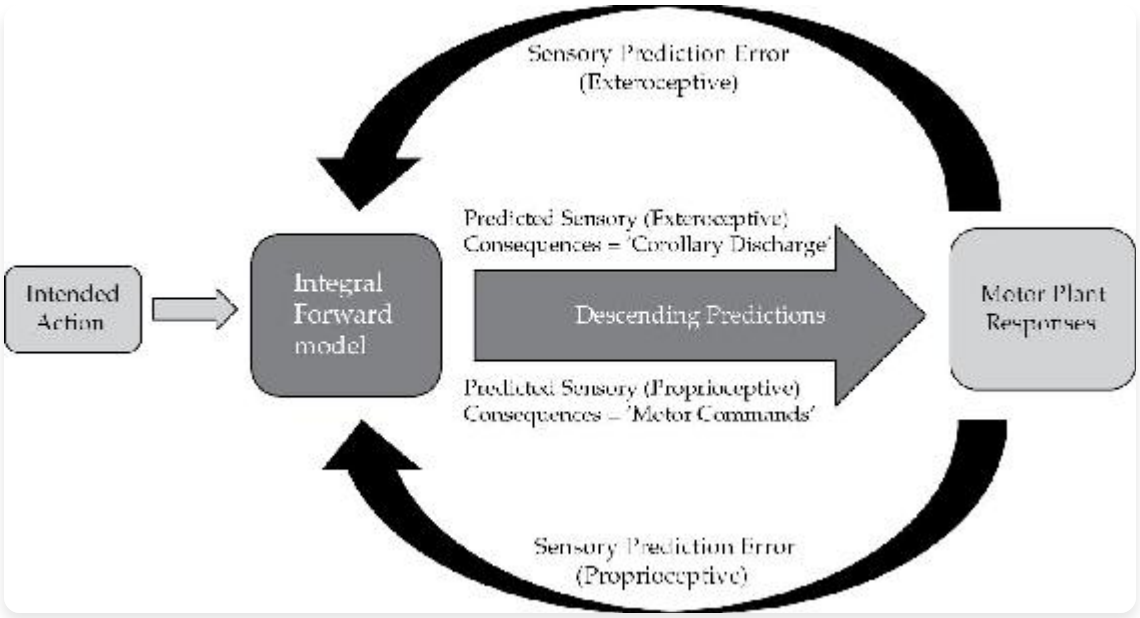
主动推理通过注意到分层生成模型可以将外在坐标中的预测映射到内在(本体感受)参考框架来避免这个困难的[逆向]问题。这意味着逆向问题变得几乎是微不足道的——要激发特定伸展感受器的放电，只需收缩相应的肌肉纤维即可。简而言之，逆向问题可以降级到脊髓水平，使来自M1[初级运动皮层]的下行传入作为预测而非命令——并使M1成为分层生成模型的一部分，而非逆向模型的一部分。(Adams, Shipp, & Friston, 2013, p. 26)

因此，运动命令被替换(见图4.2)为下行本体感受预测，其起源可能位于最高(多模态或元模态)水平，但其渐进的(上下文敏感的)展开一直进行到脊髓，在那里最终通过经典反射弧兑现(见Shipp et al., 2013; Friston, Daunizeau, et al., 2010)。

通过将成本函数重新构想为运动轨迹期望体系中的隐含因素，这种解决方案避免了在在线处理过程中解决困难(通常难以处理)的最优性方程的需要。[16] 此外，借助更复杂的生成模型，这些解决方案能够流畅地适应信号延迟、感觉噪声以及目标和运动程序之间的多对一映射。可以说，那么，涉及明确计算成本和价值的更传统方法对在线处理提出了不现实的要求，未能利用物理设备的有用(例如被动动态)特性，并且缺乏生物学上合理的实现。

然而，这些各种优势是以熟悉的代价换来的。因为在这里，PP故事也将大部分负担转移到了这些先验”信念”的获取上——多层次、多模态的概率期望网络，它们共同驱动感知和行动。实际上，PP的赌注是，这是一个值得的权衡，因为PP描述了一个生物学上合理的架构，最大程度地适合通过与世界的具身交互来安装并随后调整基于生成模型的预测的必要套件。

我们现在可以总结这些运动控制方法之间的主要差异。PP假设一个单一的整合前向模型(见图4.3)驱动行动，而更标准的方法(图4.2)将与行动相关的前向模型描述为一种附加资源。根据更标准的(“辅助前向模型”，见Pickering & Clark, 2014)解释，前向模型与实际驱动在线行动的装置截然不同。它是该装置某些效应的(简化)模型。这种模型可以在形式上与控制智能体真实运动学的任何东西有相当大的差异。此外，前向模型的输出实际上并不引起运动：它们只是用来巧妙处理和预测结果，以及用于学习。然而，根据PP(“整合前向模型”，见Pickering & Clark, 2014)解释，前向模型本身通过一个确定反射设定点的下行预测网络来控制我们的运动行为。


Sensory Prediction Error
(Exteroceptive)
Predicted Sensory (Exteroceptive)
Consequences = 'Corollary Discharge'
Intended
Action
Integral
Forward
model
Descending Predictions
Motor Plant
Responses
Predicted Sensory (Proprioceptive)
Consequences = 'Motor Commands'
Sensory Prediction Error
(Proprioceptive)

image

图4.3 整合前向模型(IFM)架构

在这种架构中，来自前向模型的预测充当行动命令，因此不需要传出副本。

来源：来自Pickering & Clark, 2014。

4.9 面向行动的预测

注意，生成模型中固有的许多概率表征现在将是，用Clark (1997)的术语来说，“面向行动的”。它们将以一种方式表征事物的状态，一旦被预测误差的精度加权适当调节，也规定了(凭借它们预测的感觉流)如何行动和响应。因此，它们是(正如我们将在后续[第8-10章]中更详细地看到的)*可供性(affordances)*的表征——环境中对有机体有意义的行动和干预机会。在这样的图式中，行动提供了一种强大的基于预测的信息流结构化形式(Pfeifer et al., 2007; Clark, 2008)。行动在概念上也是首要的，因为它提供了唯一的方式(一旦良好的世界模型就位并被恰当激活)来实际改变感觉信号以减少预测误差。[17] 智能体可以通过改变她的预测来减少预测误差而无需行动。但只有行动能够通过系统性地改变输入本身来减少误差。这两种机制必须协调工作以确保行为成功。

值得注意的是，这个关于动作的非常宽泛的故事，甚至可以被那些可能希望拒绝本体感觉预测充当运动指令这一特定模型的人所接受——也许是因为他们希望保留更熟悉的传出拷贝、成本函数以及配对的前向和反向模型装置。因为这里关于预测和动作的更广泛观点所断言的全部内容是：(i) 动作和感知都依赖于概率层次生成模型，(ii) 感知和动作在以复杂循环因果流为特征的状态下协同工作，以便最小化感觉预测误差。这种观点表明，动作和感知类似地且持续地围绕预测误差的演化流动而构建。我认为，这是预测性大脑研究关于动作所提出的基本洞察。直接利用本体感觉预测作为运动指令只是提供了这一更广泛模式的一种可能的神经元实现——尽管Friston及其同事认为考虑到运动系统生理学的已知事实，这种实现是高度合理的(Shipp et al., 2013)。

4.10 预测性机器人学

考虑到这一点，值得探索基于预测的处理程序作为移动机器人获得运动和认知技能工具的一些更广泛应用。这项工作的大部分都是在”认知发展机器人学”(CDR)范式内进行的(参见Asada et al., 2001, 2009)。这里的核心思想是，作为代理”大脑”的人工控制结构应该通过与代理环境(包括其他代理)持续具身交互的过程来发展。因此，我们读到：

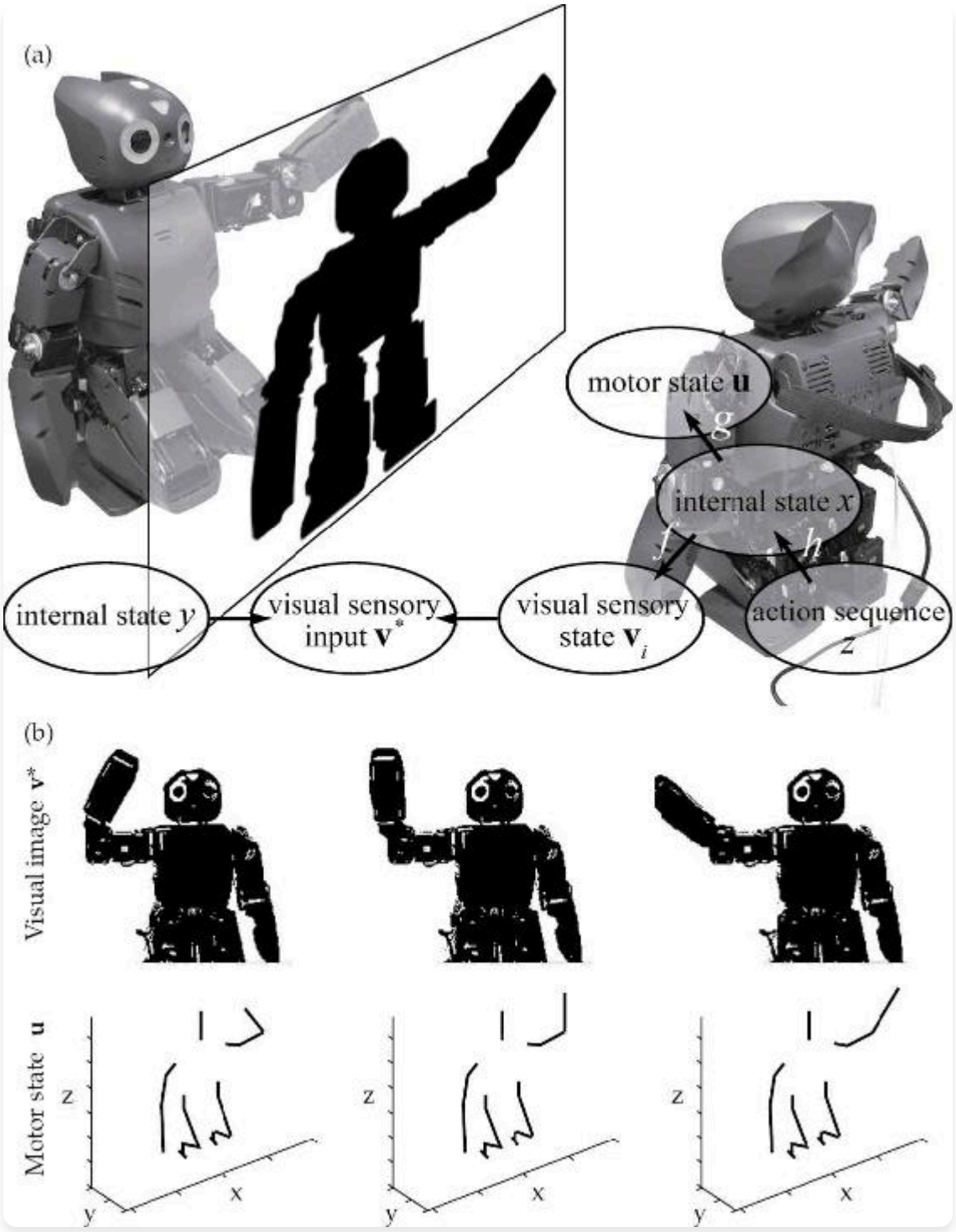
CDR的关键方面是其设计原则。现有方法通常明确地在机器人的”大脑”中实现一个控制结构，该结构来源于设计者对机器人物理学的理解。根据CDR，该结构应该反映机器人通过与环境交互来理解的自身过程。(Asada et al., 2001, p. 185)

让我们首先看看简单的运动学习。Park et al. (2012)描述了使用人形机器人AnNAO的工作。在这项工作中，简单的运动序列通过层次(贝叶斯)系统中的预测误差最小化来学习。机器人开始”体验”随机运动，类似于人类婴儿的所谓”运动咿呀学语”(Meltzoff & Moore, 1997)。在运动咿呀学语中，婴儿通过实际上随机发出运动指令然后感知(看到、感觉到、有时品尝)会发生什么来探索其个人动作空间。正如Caligiore et al. (2008)在该领域的另一项机器人研究中所指出的，这种学习是Piaget (1952)所称的”初级循环反应假说”的一个实例，根据该假说，早期的随机自我实验建立了目标、运动指令和感觉状态之间的关联，使得后来有效目标导向动作的出现成为可能。标准形式的赫布学习(Hebb, 1949)可以调节这种连接的形成，导致获得一个将动作及其预期感觉后果相关联的前向模型。Park et al. (2012)然后在这种早期学习的基础上，训练他们的机器人产生三个目标动作序列。这些是使用一系列期望动作状态定义的运动轨

迹，绘制为内部状态转换序列。机器人能够学习和重现目标动作序列，并且尽管在训练序列内的子轨迹之间存在潜在混淆的重叠，仍然做到了这一点。

在这个实验的第二阶段，机器人使用了一个层次(多层)系统，其中更高层最终学习更长的序列，运动由自上而下的预测和自下而上的感知的结合产生。因此，给定层级上最可能的转换受到关于运动所属的更长序列的自上而下信息的影响。使用这种分层方法，机器人能够学习”物体恒存性”的简化版本，预测视觉呈现物体(移动点)的位置，即使当它被另一个物体暂时遮挡时也是如此。

第三阶段将故事扩展到包含（一个非常简单版本的）通过运动模仿进行学习。这是机器人学和认知计算神经科学中一个极其活跃的领域（有关很好的介绍，请参见Rao, Schon, & Meltzoff, 2007; Demiris & Meltzoff, 2008）。Park等人使用两个相同的类人机器人（这次使用DARwin-OP[22]机器人平台构建），它们被放置在各自的视觉系统能够捕捉到对方动作的位置。一个机器人充当教师，移动手臂以产生预编程的运动例程。另一个机器人（“婴儿”）必须仅通过观察来学习这个例程。这被证明是可能的，因为婴儿机器人被训练来发展一个”自我形象”，将其自身总体运动的视觉图像与内部动作命令序列联系起来。一旦这个（薄弱的）自我形象建立起来，从教师机器人观察到的运动就可以与记忆的自我形象匹配，从而与内部动作命令联系起来（图4.4）。然而，这个例程只有在目标系统（教师）与学习者（“婴儿”）充分”相似”的情况下才能工作。这种”像我一样”的假设（Meltzoff, 2007a, b）可能会在智能体的生成模型在复杂性和内容上增加时得到放松，但它可能是启动模仿学习过程所必需的（有关一些优秀的讨论和各种进一步的机器人研究，请参见Kaipa, Bongard, & Meltzoff, 2010）。



image

图4.4 **模仿学习的结构**

a. 目标系统（左侧机器人）从内部状态序列(*y*)产生图像序列(*v*)。智能体系统（右侧机器人）通过将图像序列(*v*)映射到已知内部动作状态*x*的记忆自我形象(*vi*)来跟随。目标的视觉序列在智能体中产生一系列内部动作状态。智能

体训练这个序列来构建动作序列(z)并将动作重现为真实的运动状态u。(b) 智能体看到目标机器人的视觉图像（上），智能体的运动状态从图像中推导出来（下）。

来源：摘自Park et al., 2012。

因此，基于预测的学习为在简单感觉运动技能和更高认知成就（如规划、模仿和行为的离线模拟）之间架起桥梁提供了特别有力的资源（参见第5章）。

4.11 感知-认知-行动引擎

熟悉科幻小说（或电影）《安德的游戏》的读者会记得，起初看似仅仅是模拟的东西实际上被秘密地用来在真实战斗情况下驾驶物理星舰。在我看来，这是思考关于行动的核心PP提案的一种方式。因为如果故事的这个方面是正确的，行动是作为基于前向模型的模拟的直接结果产生的。

基本预测处理故事的面向行动的扩展，正如威廉·詹姆斯的开篇引言所暗示的，因此与有时被称为”观念运动”(ideomotor)的行动账户有很多共同点。根据这些账户（Lotze, 1852; James, 1890），运动的想法本身，当不受其他因素阻碍时，就是带来运动的原因。换句话说：

在观念运动观点中，在某种意义上，存在于现实世界中的因果关系在内在世界中是颠倒的。对行动预期效果的心理表征是行动的原因：这里不是行动产生效果，而是（效果的内部表征）产生行动。（Pezzulo et al., 2007, p. 75）

在Friston及其同事所支持的方法中，这表现为我们学会将自己的运动与其独特的本体感受后果联系起来的想法。因此，行动是通过本体感受预测来控制和实现的，通过移动身体以适应预测来消除本体感受预测误差。

这些方法广泛使用了经典运动控制工作中的前向模型构造，但现在将其重新定义为更全面的生成模型的一部分。这复制了使用前向模型的许多好处，同时使用与（在第一部分中）用来解释感知、理解和想象的完全相同的装置来处理运动控制。在第3章结尾宣布的”认知一揽子交易”因此得到了丰富，运动控制从基于生成模型的感觉预测的相同核心架构中流出。这暗示了用于动作生成和推理可能动作（无论是我们自己的还是其他智能体的）的共享计算机制的可能性——这将是下一章的重要主题。

这是一个有吸引力的包装，但它也带来了成本。现在是我们获得的期望（生成模型所隐含的次个人预测的复合体）承担了大部分解释性和计算负担。然而，这可能会被证明是一个赋权而非限制的因素。因为PP描述了一种生物学上合理的架构，几乎最大程度地适合通过与我们遇到、扰动以及在几个较慢时间尺度上主动构建的训练环境进行具身交互来安装必要的预测套件。这是一个特别有效的配方，因为大部分”高级认知”（我将在第三部分论证）只有通过我们与由我们自己文化创造的”设计师”环境所产生的越来越奇异的感官流的遭遇历史才成为可能。

新兴的图景是感知、认知和行动是单一适应性机制的表现，该机制旨在减少对有机体重要的预测误差。感觉和运动处理之间曾经清晰的界限现在消解了：行动源于预测感官信号的知觉，其中一些信号引导行动，而行动又招募新的知觉。当我们用感官接触世界时，知觉和行动配方现在共同出现，将运动处方与持续的认知和理解努力结合起来。因此，行动、认知和感知是持续共同构建的，同时根植于构成、测试和维持我们对世界掌握的级联预测中。

5

精确工程：塑造流动

5.1 双重代理

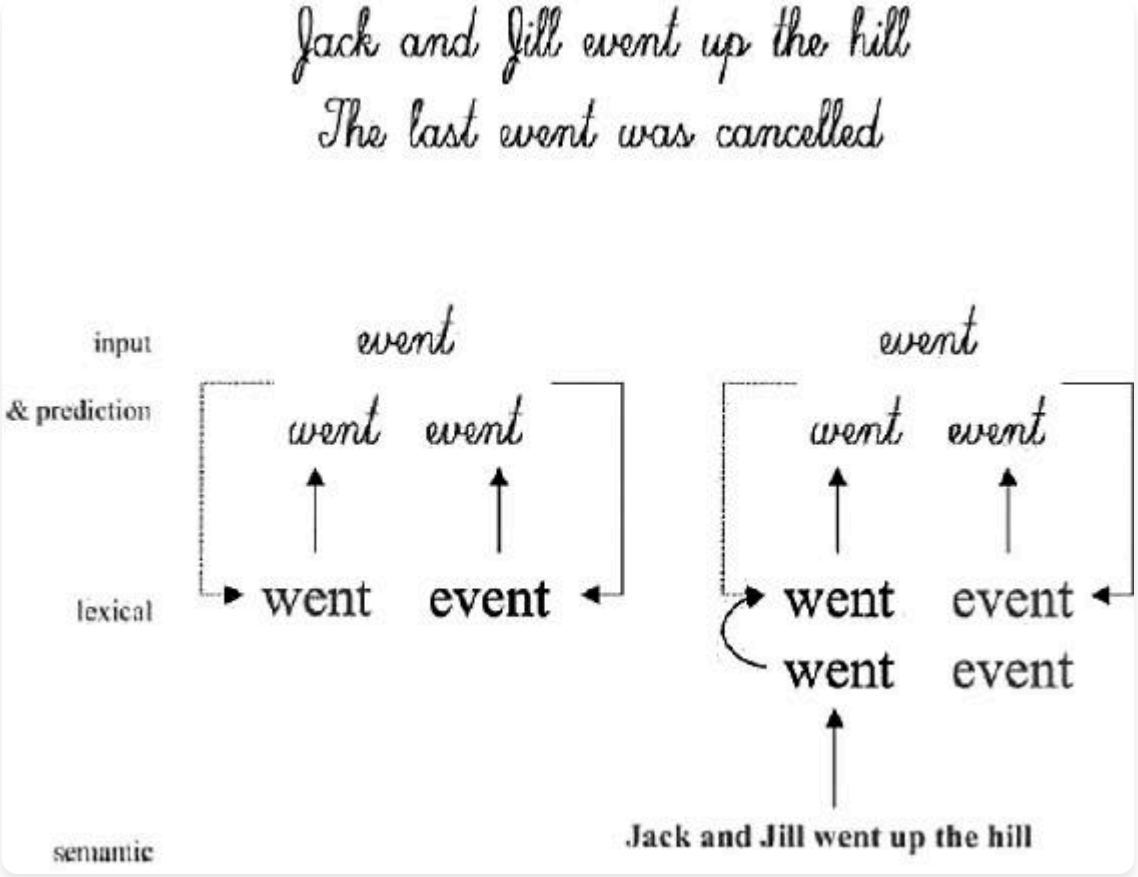
将大脑视为概率预测机器的形象将情境和行动置于中心舞台。它要求我们放弃”输入-输出”模型的最后残余，根据该模型，环境刺激反复冲击一个组织丰富但本质上被动的系统。取而代之的是一个持续活跃的系统，不断地从一个期望状态转移到另一个，将感官状态与收获新感官状态的预测在滚动循环中匹配。

在这个复杂、变化的网络中，行动过着某种双重生活。行动（像世界中的任何其他规律性一样）需要被理解。但如果第4章所述的故事是正确的，我们自己的行动也是我们部署的生成模型中编码的感官期望的后果。这产生了一个机会。也许我们对其他代理的预测可以通过构建我们自己行动和反应模式的同一个生成模型的信息？这表明，我们有时可能通过部署适当转换的、支撑我们自己行为的多层期望集合来掌握其他代理的意图。因此，其他代理被视为我们自己的情境细化版本。这为”镜像神经元”和（更一般地）“镜像系统”的发展和部署提供了洞察：既参与行动执行又参与观察他人执行”相同”行动的神经资源。

本章探讨了这一策略，将其用作更一般和强大事物的核心例证——在PP内使用精确度的改变分配来重新配置大脑内有效连接的模式。

5.2 朝向最大情境敏感性

我们可以从考虑情境敏感反应的熟悉案例开始，如图5.1所示。



image

图5.1 **说明先验在偏向输入的一种表征或另一种表征中的作用的示意图**

（左）单词”event”被选为视觉输入最可能的原因。（右）单词”went”被选为最可能的单词，它（1）是感官输入的合理解释，并且（2）符合基于语义情境的先验期望。

来源：Friston, 2002。

在图5.1中，尽管单词”went”渲染得很糟糕，我们还是毫不费力地将顶部句子读作”Jack and Jill went up the hill”。实际上，它在结构上与第二句中渲染得更好的单词形式”event”相同。然而，这一事实只有在相当仔细的检查下才明显。这是由于强烈的自上而下先验的强大影响，它们帮助确定最佳整体”拟合”，调和感官证据与我们的概率期望。在忽略（由于来自上层的强预测）顶部句子中”went”的结构缺陷时，“我们容忍较低层的小误差以最小化总体预测误差”（Friston, 2002，第237页）。关于这种自上而下对知觉外观影响的另一个例子，见图5.2。

image

图5.2 **局部情境线索建立先验期望的另一个例子**

在阅读A的情境中，B假设使原始视觉数据最可能。在阅读12的情境中，13假设使完全相同的原始视觉数据最可能。关于此例的进一步讨论，见Lupyan & Clark，出版中。

这种效果相当常见。它们是早期人工神经网络所展现的上下文敏感性的例子，这些网络采用了连接主义的”交互激活”范式。[2]在预测处理范式中，这种上下文敏感性变得（在下面要探讨的意义上）普遍存在且”最大化”。[3]这是由于分层形式与灵活的”精度加权”相结合的结果，正如在第2章中介绍的那样。这种结合使得上下文敏感性变得根本性、系统性和普遍性，这对神经科学以及我们对智能反应的本质和起源的理解具有重大意义（另见Phillips & Singer (1997)，Phillips, Clark, & Silverstein (2015)）。为了逐步理解这一点，请思考（在上面给出的简单例子中）：“如果我们在对’事件’的自上而下期望下记录’went’单元，我们可能会得出结论，它现在对’事件’具有选择性”（Friston, 2002，第240页）。因此，在预测处理设置中，向下流动的影响对较低层级反应的选择性产生重大影响（更多例子见Friston & Price, 2001）。结果，“任何神经元、神经元群体或皮层区域的表征能力和固有功能都是动态的和

上下文敏感的，并且在任何给定皮层区域中，神经元反应可以在不同时间代表不同的事物”（Friston & Price, 2001，第275页）。

预测处理架构在这里以一种普遍而流畅的方式结合了神经组织的两个最显著特征。这两个特征是功能分化(functional differentiation)（有时被误导地称为”专门化”）和整合(integration)。功能分化意味着局部神经集合将呈现不同的”反应模式”，这些模式反映了”内在局部皮层偏向和外在因素（包括经验以及与大脑其他区域功能交互的影响）的结合”（Anderson, 2014，第52页）。这些反应模式将有助于确定该集合可能被招募执行的任务类型。但正如Anderson正确强调的那样，这不一定意味着在更标准意义上的专门化（例如，存在专门执行面部识别或心理阅读等固定任务的区域）。整合（神经经济体展现的那种相当深刻的整合）意味着那些功能分化的区域以动态方式相互作用，使得短暂的任务特定处理机制（涉及神经资源的短暂联盟）能够出现，因为上下文效应不断重新配置信息和影响的流动。

这种效应作为基本预测处理模型所隐含的分层组织的直接结果而变得普遍和系统化。在该模型中，多个功能分化的子群体交换信号以找到最佳整体假设，来自每个更高群体的信号基于其自身的概率先验（“期望”）为直接下方的层级提供丰富的上下文化信息。正如我们刚才看到的，这种向下（和横向）的预测流对接收它的单元的瞬时反应性产生了巨大差异。此外，正如我们在第2章中看到的，特定自上而下或自下而上影响的效力*本身*可以通过系统性精度估计来修改，这些估计提高或降低特定预测误差信号的增益。这意味着向下流动影响的模式（以我们即将探索的方式）本身根据任务和上下文动态可重新配置。结果是预测和预测误差信号的流动实现了一个灵活的、动态可重新配置的级联，其中来自每个更高层级的上下文信息可以在塑造选择性和反应方面发挥作用，“一直向下”。

[5.3 层次结构的重新考虑]

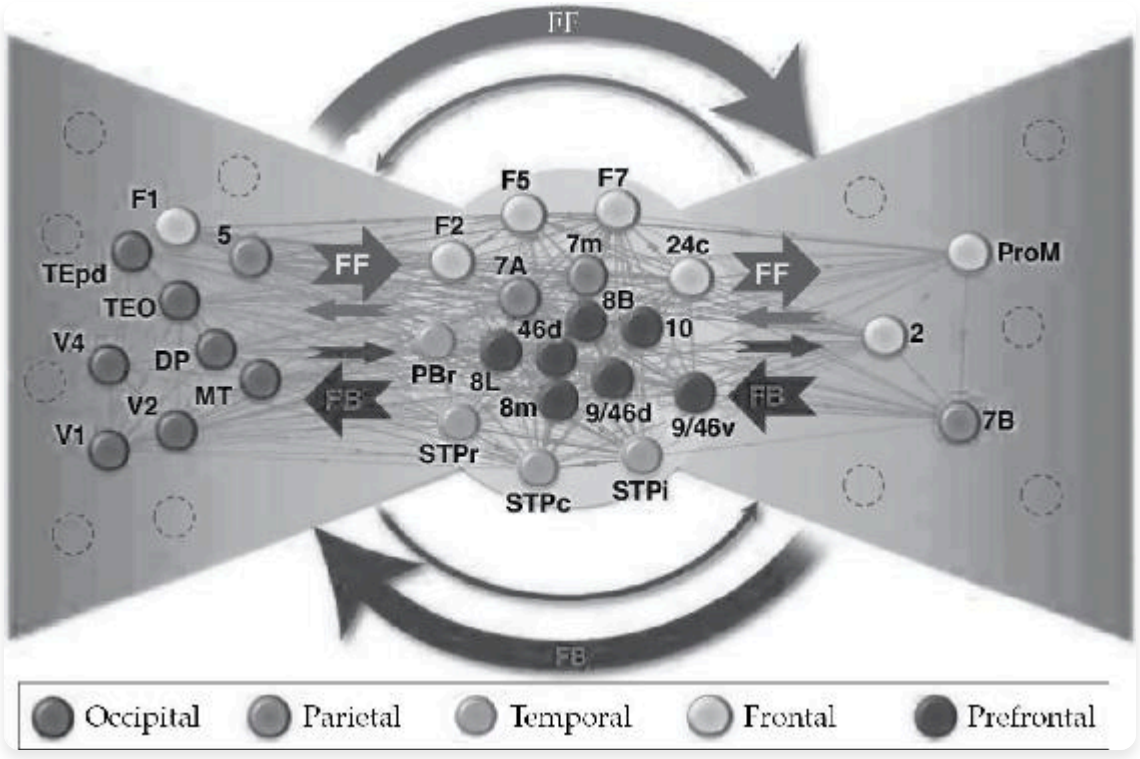
回忆一下，在预测处理的标准实现中[4]，更高层级的”表征单元”横向（在层级内）和向下（到下一个层级）发送预测信号，从而为下级层级的活动提供先验。通过这种方式，反向（自上而下）和横向连接结合起来”对皮层变换的较低或等效阶段施加调节影响，并定义皮层区域的层次结构”（Friston & Price, 2001，第279页）。这种皮层层次结构支持（正如我们在第1章中看到的）引导经验先验的自举式学习。[5] 这样的层次结构简单地*由*这些交互模式定义。核心要求只是存在一个具有不对称功能角色的反馈和前馈连接的相互连接结构。更详细地说，所需要的是神经元群体使用不同的前馈、反馈和横向连接交换信号，并且在这些影响网络中，功能不对称的资源处理预测和预测误差信号。

在一项开创性研究中，Felleman and Van Essen (1991) 描述了一个解剖层级结构（针对猕猴视觉皮层），其结构和特征很好地映射到这个广泛模式所需要的特征（讨论见 Bastos et al., 2012, 2015）。这样的层级结构为额外的——以及功能上显著的——复杂性留下了充足的空间。例如，相邻层级级别之间局部循环信号交换的概念与多个平行流的存在是一致的，这些平行流传递 Felleman and Van Essen (1991) 称之为”分布式层级处理”的内容。在这样的方案中，多个区域可能共存于层级的单一”级别”上，也可能存在完全跳过某些中间级别的长程连接。

然而，Felleman 和 Van Essen 的开创性研究受到了缺乏层级距离测量的限制。这给从研究中出现的区域排序引入了实质性的不确定性（见 Hilgetag et al., 1996, 2000），这一不足后来通过使用连接性新数据得到了弥补（Barone et al., 2000）。最近，Markov et al. (2013, 2014) 使用神经追踪数据和网络模型来探索猕猴视觉皮层中区域间以及前馈/反馈连接的网络。他们的研究表明，反馈和前馈连接在解剖学和功能上确实都是不同的，前馈和反馈通路”遵循明确定义的距离规则”（Markov et al., 2014, p. 38），从而确认了预测处理(PP)对层级结构和反馈/前馈功能不对称性的基本要求。

尽管如此，这些研究也为固定皮层层级中反馈和前馈连接的任何简单形象引入了实质性复杂性，揭示了显示”蝴蝶结”结构的连接网络，该结构将高密度局部通信（在一种”核心”中——图5.3所示蝴蝶结中心的结）与到皮层其余

部分的更稀疏长程连接相结合。这些长程连接允许密集连接的局部处理包（“模块”）进入临时的任务和上下文变化联盟（见 Park & Friston, 2013; Sporns, 2010, Anderson, 2014）。这样的组织形式也与更高级别的”富人俱乐部”组织一致（Van den Heuvel & Sporns, 2011），其中某些连接良好的局部”枢纽”本身彼此高度相互连接（就像精英和实权人物的专属乡村俱乐部）。出现的是多尺度动力学复杂性的令人生畏的图景。

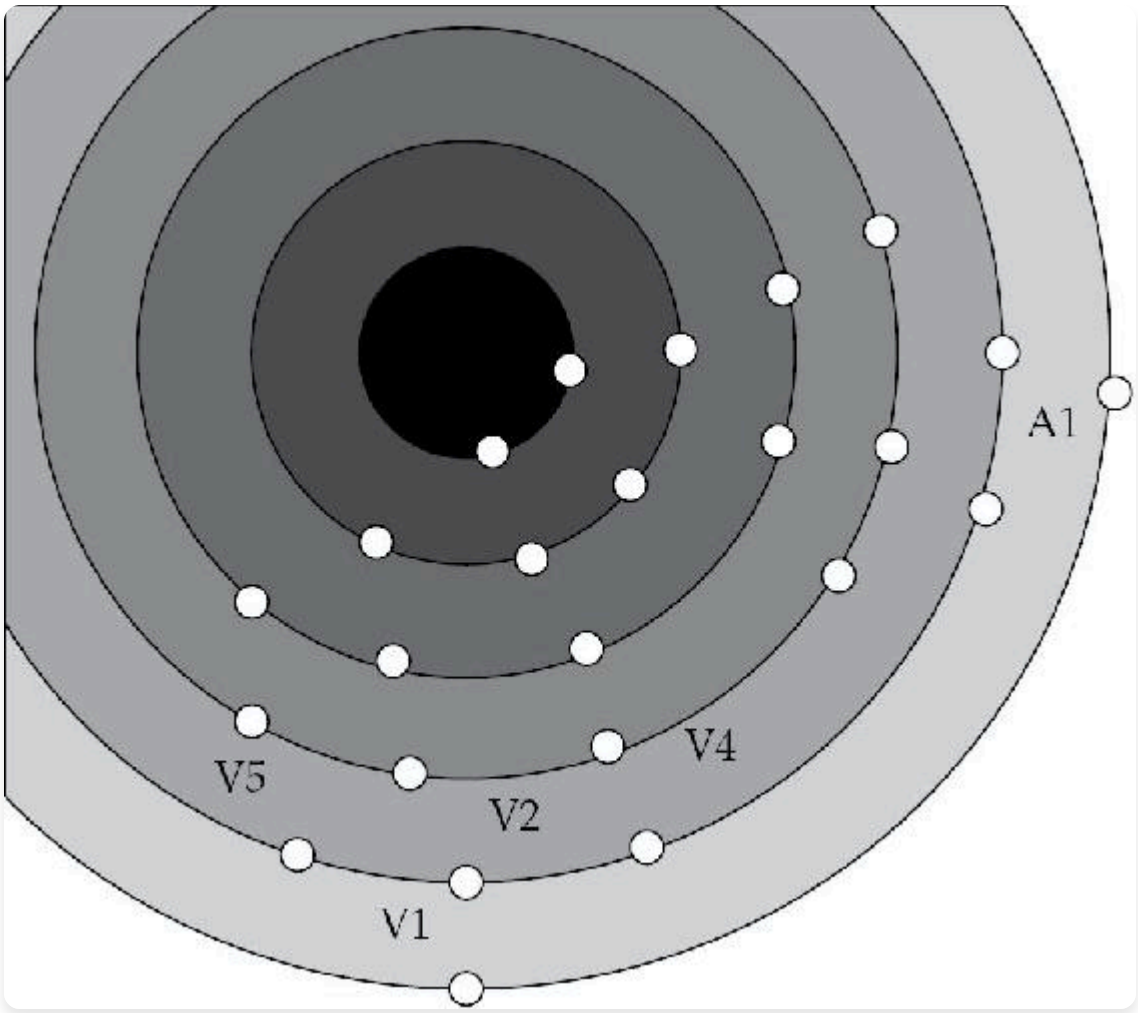


image

图5.3 高密度皮层矩阵的蝴蝶结表示

来源: Markov et al., 2013，经许可使用。

这很重要。皮层层级内处理的简单形象可能看起来暗示着一种刚性、固定、串行的信息流——一种神经阶梯，具有不可避免的有问题的”顶层”端点。值得强调的是，预测处理(PP)架构具有非常不同的含义。与传统的前馈模型（被 Churchland et al., 1994 正确批评的那种）不同，预测处理(PP)架构支持持续的、并发的双向信息流。这意味着任何给定较高级别的处理不会”等待”下面级别的处理完成后才开始发挥其影响。此外，感知处理层级最好被想象（Mesulam, 1998; Penny, 2012）为一种球体（见图5.4）而不是阶梯。感官刺激在外围扰动球体，并遇到一系列预测，其募集是上下文和任务依赖的。在球体内，有结构和结构的结构。但信息和影响的演化流动不是固定的。相反，预测处理(PP)建议（如我们即将看到的）多种有效机制，用于根据任务和上下文重新配置时刻”有效连接性”模式。因此，预测处理(PP)对层级的使用与大脑作为一个复杂的永远活跃的动力学系统的图景高度一致：一个其时刻信号传递结构是上下文敏感的、流动的、多重可重新配置的，并且在许多相互作用的结构和时间尺度上不断变化的系统（Singer, 2013; Bastos et al., 2012, 2015）。灵活的精度加权提供了，正如我们即将更详细地看到的，关键的系统工具，允许将基础双向层级模型与上下文敏感的”假设”生成，以及上下文可变的级别间（和区域间）影响模式相结合。



image

图5.4 **描绘中心多模态区域和外围单模态感觉处理区域的皮层架构**

视觉区域显示在底部，听觉区域在右侧。

来源: Penny, 2012，基于 Mesulam, 1998。

在这样一个持续活跃的系统中，高层级经常被描述为（参见 [Sherman & Guillery, 1998]）对低层级的活动进行”调节”。然而，越来越不清楚的是，“调节”是否真的是描述持续概率预测对处理流程和响应产生诸多戏剧性影响的最佳方式。最近的证据表明，当两个区域在层级上接近时（如V1/V2），“反馈连接可以像前馈连接一样强烈地驱动其目标”（[Bastos et al., 2012]，第698页）。在预测处理框架内，这意味着自上而下的预测在适当的情况下，能够以激进地修正甚至撤销”驱动性”前馈影响的方式强制下游响应。同时，“调节”这一概念在某些方面仍然适用，因为向下（和横向）流动预测的功能作用是提供必要的情境化信息。

5.4 塑造有效连接

在预测处理框架内，基础情境效应不可避免地源于使用高层级概率期望来指导和细化低层级响应。这种指导涉及对下一层级最可能的活动展开模式的期望。但正如我们在[第2章]中看到的，这些期望与基于情境的感觉信息本身不同方面的可靠性和显著性评估交织在一起。这些可靠性和显著性的评估决定了在不同处理层级对预测误差信号不同方面给予的*权重*（精度）。这提供了一种强大的手段来塑造”有效连接”的更大模式，根据任务和情境修改内部影响和信息流。

“有效连接”[6]指的是”一个神经系统对另一个神经系统施加的影响”（[Friston, 1995]，第57页）。它与结构连接和功能连接都有所区别。“结构连接”指的是物理连接的总体模式（纤维和突触网络），这些连接可能与更分散的”体积信号”机制（Philipides et al., 2000, 2005）协同工作，允许神经元在空间和时间上相互作用。“功能连接”描述神经事件之间观察到的时间相关模式。密切相关的”有效连接”概念旨在反映神经事件之间短期的因果影响模式，从而超越对无向且有时无信息价值的相关性的简单观察。思考功能连接和有效连接之间关系的一个有用方式是将：

[电生理学]有效连接概念……视为依赖于实验和时间的、能够复制所记录神经元之间观察到的时间关系的最简单可能电路图。（[Aertsen & Preissl, 1991]，引用自[Friston, 1995]，第58页）

当我们执行认知任务时，功能连接和有效连接模式会迅速改变。相比之下，结构变化是一个较慢[7]的过程，因为它实际上是在重新配置可重新配置的网络本身（通过改变支持其他更快速、瞬时重新配置形式的底层通信骨架）。

近年来，神经影像技术与新分析技术的结合使得有效连接模式的研究变得越来越可行。这些技术包括结构方程建模、“格兰杰因果关系”的应用，以及动态因果建模（DCM）[8]。在一个相当令人满意的转折中，DCM（[Friston et al., 2003]；[Kiebel et al., 2009]）采用了大脑用来建模世界的相同核心策略（如果预测处理是正确的话），并将其应用于神经影像数据本身的分析。DCM依赖于生成模型来估计（推断）给定某些影像数据的神经源，并使用贝叶斯估计来揭示有效连接的变化模式。通过这种方式，DCM首先估计源之间的内在连接，然后估计由于某种形式的外部（通常是实验性）扰动导致的连接变化。

DCM的非线性扩展（[Stephan et al., 2008]）不仅允许估计有效连接如何随实验操作（任务和情境的改变）而变化，还允许估计这些新的有效连接模式是如何产生的，即”两个神经单元之间的连接如何被其他单元的活动启用或门控”（[Stephan et al., 2008]，第649页）。这种门控涉及[Clark (1997]，第136页）所称的”神经控制结构”，这些可以定义为任何其作用是控制内部经济形状而不是（直接）跟踪外部事态或控制身体行动的神经回路、结构或过程。正是在这种意义上，[Van Essen et al. (1994)]提出了与现代工厂过程分工的类比，在工厂中，必须花费大量努力和能量来确保材料的适当内部流通，然后才能构建任何实际产品。

神经门控假说有多种形式，包括假设存在特殊的信息路由”控制神经元”群体(Van Essen et al., 1994)，巧妙利用再入式处理(Edelman, 1987; Edelman & Mountcastle, 1978)，以及”汇聚区”的发展(Damasio & Damasio, 1994)。后者本质上是许多反馈和前馈循环汇聚的枢纽，因此能够”指导解剖学上分离的区域同时激活”(第65页)。这些推测与一个新兴的研究领域自然契合，该领域证明了”大规模脑系统能够以情境依赖的方式重新配置其功能架构”的惊人程度(Cocchi et al., 2013, 第493页; 另见 Cole et al., 2011; Fornito et al., 2012)。

在预测处理(PP)框架内，门控主要通过操控分配给特定预测误差的精度权重来实现。这样做的主要效果(如我们在第2章中所见)是通过增加选定误差单元的增益(“音量”)来系统性地改变自上而下与自下而上信息的相对影响。这提供了一种实现丰富注意机制集合的方法，其作用是偏置处理过程，以反映对感觉信号和生成模型本身(不同方面)的可靠性和显著性的估计。[9] 但这些相同的机制为在皮层群体间实现流畅灵活的大规模门控形式提供了有前景的手段。要理解这一点，我们只需注意到精度极低的预测误差对持续处理几乎没有影响，并且无法招募或细化更高层次的表征。因此，改变精度权重的分布实际上等同于改变当前处理的”最简单电路图”(Aertsen & Preissl, 1991)。注意的神经机制在这里与改变有效连接模式的神经机制是相同的。

这是一个直觉性的结果(另见Van Essen et al., 1994)，特别是如果我们考虑到实现这种改变的具体手段很多，而且它们在大脑不同部位的详细功能意义可能有所不同。预测误差精度权重(在PP中等同于控制突触后增益)的可能实现机制包括各种”调节性神经递质”的作用，如多巴胺、血清素、乙酰胆碱和去甲肾上腺素(Friston, 2009)。振荡频率也必须发挥重要作用(见Engel et al., 2001; Hipp et al., 2011)。例如，同步的突触前输入似乎会导致突触后增益增加。特别是，有人提出”伽马振荡可以通过让同步神经元放电对下游神经元的放电率产生更大影响来控制增益”(Feldman & Friston, 2010, 第2页)。这些机制也相互作用，因为(仅举一例)伽马振荡对乙酰胆碱有反应。总的来说，似乎可能自下而上的信号传递(在预测处理中编码预测误差，假设源于浅层锥体细胞)可能使用伽马范围频率进行传播，而自上而下的影响可能通过贝塔频率传达(见Bastos et al., 2012, 2015; Buffalo et al., 2011)。因此，虽然通过”精度权重预测误差”塑造有效连接模式的概念足够简单，但实现这些效果的机制可能是多重和复杂的，它们可能以重要但尚未充分认识的方式相互作用。

最近一项使用非线性DCM分析的fMRI研究为这一总体思想（即基于精确度的大规模有效连接模式重新配置的思想）提供了进一步支持。在这项研究中（den Ouden et al., 2010），一个（纹状体）神经区域的特定预测误差信号修改了其他（视觉和运动）区域之间的耦合。在这个实验中，听觉线索（高或低”哔哔声”）以随时间变化的方式差异化地预测视觉目标。受试者的任务是快速辨别（通过运动反应）由听觉线索预测的视觉刺激（预测方式随时间变化）。随着可预测性的增加，速度和准确性提高（正如人们所期望的）。使用DCM（并假设一个贝叶斯学习模型，该模型在考虑模型复杂性的情况下为数据提供了最佳拟合；见den Ouden et al., 2010，第3212页），实验者发现预测失败（由变化的偶然性引起）系统性地改变了视觉运动耦合的强度，这种改变”由壳核编码的预测误差程度控制”，并且”壳核的预测误差反应[调节]从视觉到运动区域的信息传递……与……纹状体的门控作用一致”（两个引用，第3217页）。因此，纹状体计算的预测误差的数量和精确度精细地控制着视觉运动连接的强度（效力），协调着视觉和运动区域之间时刻变化的相互作用。这是一个重要的结果，证明了”特定区域中逐试次的预测误差反应调节其他区域之间的耦合”（den Ouden et al., 2010，第3217页）。

因此，预测层次结构内持续活动的最重要影响是，它支持将大脑视为不安静的：几乎持续处于某种（变化的）主动期望状态，我们才刚刚开始理解这种状态对感觉输入流动和处理的影响。

[5.5 瞬态组合体]

我们已经看到，PP架构将功能分化与多种（普遍且灵活的）信息整合形式相结合。这为备受争议的认知”模块性”概念提供了新的视角（见，例如，Fodor, 1983，讨论见Barrett & Kurzban, 2006；Colombo, 2013；Park & Friston, 2013；Sporns, 2010，Anderson, 2014）。神经群体之间（以及更大尺度区域之间）的变化影响模式在这里由精确度加权的预测误差信号决定，因此由显著性估计和相对不确定性估计决定——对于给定时间的给定任务——与不同神经区域和不同神经元群体中的活动相关。这样的系统显示出极大的上下文敏感性，同时受益于一种涌现的”软模块性”。独特的、客观可识别的[10]局部处理组织现在在一个更大的、更整合的框架内涌现和运作，在这个框架中，功能分化的群体和亚群体以不同的方式参与和细化，以服务于不同的任务（关于这种一般的多用途图景的更多信息，见Anderson, 2010，2014）。

因此，PP实现了一种理想适合支持Anderson (2014)巧妙地称为TALoNS——瞬态组装的局部神经子系统——的形成和解散的架构。TALoNS在某些方面像模块或组件一样起作用。但它们是”即时”形成和重新形成的，它们的功能贡献根据它们在更大处理网络中的位置而变化。PP实现了这样一种完全灵活的认知架构，并提供了一幅高度敏感的神经动力学图景，在多个时间尺度上，既对变化的任务需求敏感，也对特定自上而下期望和自下而上感觉输入的估计可靠性（或其他方面）敏感。[11]

这里的神经表征”成为远端皮质区域输入的函数，并依赖于这些输入”（Friston and Price, 2001，第280页）。这是灵活性的有力来源，因为来自这些区域的输入流本身*也*受到大脑其他地方预测误差信号的快速重构影响。当这些特征

结合时，结果是一种架构，其中有不同的、功能分化的组件和回路，但其不断变化的动力学是（借用Spivey, 2007的一个短语）“交互主导的”。这样构建的高度可协商的影响流本身是动作响应的（强制执行连接知觉和动作的各种形式的”循环因果关系”），动力学可能性的空间进一步通过各种身体和世界技巧来丰富（正如我们将在第三部分中看到的），这些技巧用于构造我们自己的输入和重构问题空间。因此支持的表征经济牢固地建立在感觉运动经验基础上，但受益于（正如我们即将看到的）分层学习导致的各种抽象形式。结果是一个令人生畏的复杂系统：一个将深度上下文灵活的处理制度与丰富的脑-身体-世界回路网络相结合的系统，产生可协商的、极其（几乎无法想象地）可塑的信息和影响流。

5.6 理解行动

在我看来，这种日益增长的复杂性在我们理解自己和他人行动的能力上表现得最为明显。人类婴儿在4岁左右，不仅具有将自己视为具有特定需求、欲望和信念的个体代理人的意识，还具有将他人视为拥有*自己的*需求、欲望和信念的独特代理人的意识。这是如何实现的呢？“镜像神经元”的发现似乎为许多人提供了答案的重要部分。然而，镜像神经元的存在可能更多的是一种症状而非解释，灵活的、情境敏感的预测性处理提供了一种更基本的机制。如果这是正确的，理解他人的行动只是灵活、情境敏感反应这一更一般能力的一种表现。

镜像神经元最初在猕猴的F5区域（前运动区）被发现（Di Pellegrino等，1992；Gallese等，1996；Rizzolatti等，1988，1996）。当猴子执行某个动作时（例如从盒子里拿苹果，或用精确的抓握动作拾起葡萄干），这些神经元会产生强烈反应。然而，实验者惊讶地发现，当猴子仅仅*观察*另一个代理人执行同样动作时，这些相同的神经元也会产生强烈反应。具有这种双重特征的神经元也在猴子的顶叶皮层中被发现（Fogassi等，1998，2005）。此外，“口部镜像神经元”（Ferrari等，2003）在猴子从分配器（注射器）中吸取果汁时以及当猴子看到人类执行相同动作时都会做出反应。在更大的尺度上，使用fMRI等神经影像技术在人类大脑中发现了”镜像系统”（用于产生、观察和模仿动作的重叠资源）（Fadiga等，2002；Gazzola & Keysers，2009；Iacoboni，2009；Iacoboni等，1999；Iacoboni等，2005）。

镜像神经元（及其参与的更大的”镜像系统”）吸引了认知科学家的想象力，因为它们提出了一种使用我们自己行动”意义”的知识作为理解他人行动的杠杆的方法。因此，假设我们承认，当我用精确抓握伸手去拿葡萄干时，我知道（以某种简单的、一阶的方式）我的行动完全是为了获得并吃掉这个有吸引力的小食物。那么，如果当我看到*你*用这样的精确抓握伸手去拿葡萄干时，完全相同的镜像神经元亚群被激活，也许我因此获得了关于*你的*目标和意图的信息——坦率地说，就是你对那颗葡萄干的渴望。这样一个窥视其他代理人内心的窗口将非常有用，使我能够更好地预测你的下一步行动，甚至可能阻挠它们，通过某种快速干预为自己获得葡萄干。以某种这样的方式，镜像神经元被认为为解释Gallese、Keysers和Rizzolatti（2004）所称的我们对”他人行动的体验性理解”提供了”基本机制”，特别是（见Rizzolatti & Sinigaglia，2007）他们的目标和意图。

这里的”体验性理解”指的是某种深层的、原始的或”具身的”理解，它使我们能够通过涉及我自己对相同行动的运动表征的”直接匹配”（Rizzolatti & Sinigaglia，2007）或”共振”（Rizzolatti等，2001，第661页）过程来理解观察到的行动的意义。如果这是正确的，观察你的行动会让我模拟或部分激活（在我身上）会导致观察到的活动的目标/运动行为例程。以这种方式，据声称，“我们通过自己的’运动知识’理解他人的行动[并且]这种知识使我们能够立即为他人的动作赋予意向性意义”（Rizzolatti & Sinigaglia，2007，第205页）。

然而，这一切都不能像听起来那么简单。这是因为执行这项任务涉及解决一个特别复杂的所谓”逆问题”版本。这个问题（我们在4.4节已经遇到过一个简单版本）是要获取一个结果指定输入（这里是观察到另一个主体的一系列运动动作），并用它来找到产生这些结果的指令（这里是指定各种高级目标和意图的神经状态）。通常，问题在于观察到的运动与产生它们的高级状态（编码目标和意图）之间可能存在多种映射关系。因此，“如果你看到街上有人举手，他们可能是在叫出租车或者拍黄蜂”（Press, Heyes, & Kilner, 2011）。或者，重复Jacob和Jeannerod（2003）的生动例子，那个穿白大褂拿着刀对着人类胸部的男人是想进行残忍的谋杀还是挽救生命的手术——他是杰基尔医生

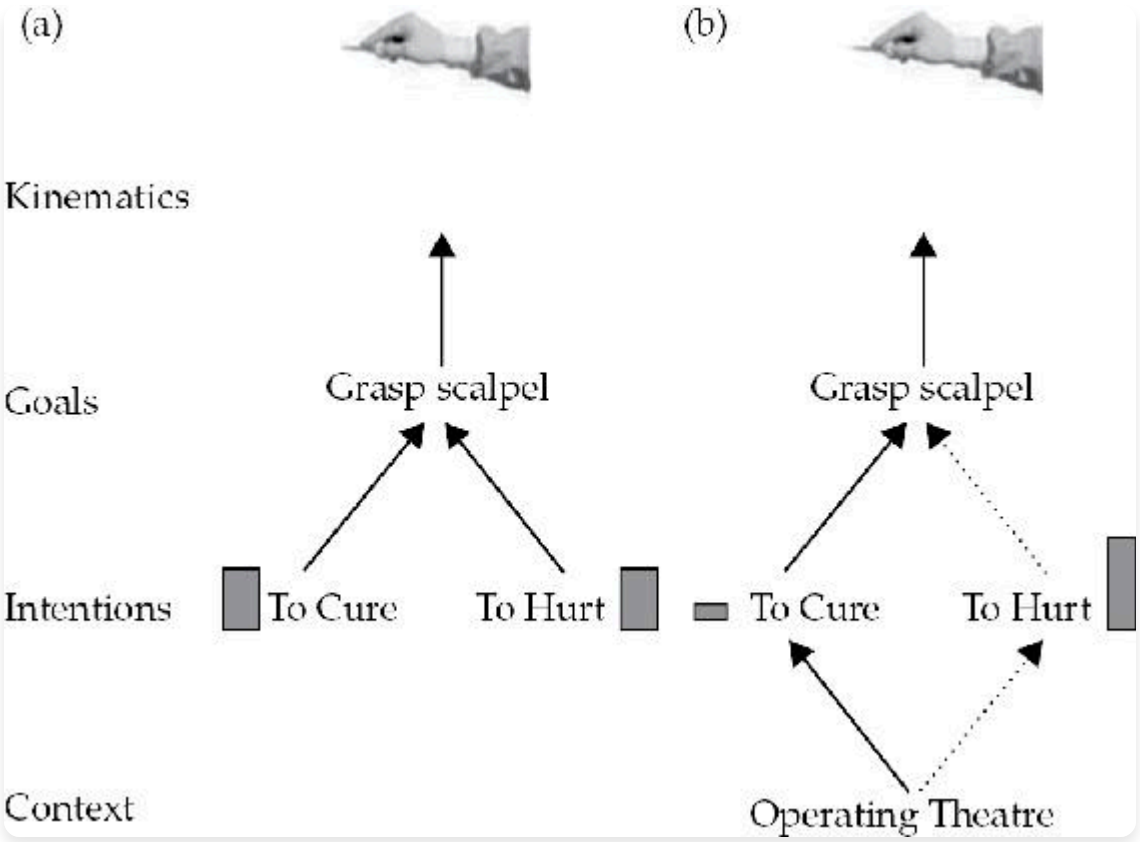
还是海德先生？这样的意图在仅仅的运动序列中并不透明，因为这些序列与其背后的意图之间没有唯一的映射关系。Jacob和Jeannerod（另见Jeannerod, 2006, p. 149）因此担心简单的基于运动的匹配机制必然无法把握他们所说的”先验目标和意图”。

这表明，无论什么机制可能支撑假设的”直接匹配”或”共振”过程，它都不能是仅仅依赖感觉信息前馈（“自下而上”）流动的机制。相反，从基本观察到的运动学到对主体意图的理解的路径必须由系统的先验状态非常灵活地调节。实现这一点的一种方法是用反映观察者已经了解的关于其他主体运动产生的更大背景的向下流动活动来迎接传入的感觉信息流。现在回想一下自我产生行动的图景（第4章）。当我们行动时，如果这个图景是正确的，我们预测将产生的感觉数据流。这种预测涉及一个多层次”调节”过程，其中许多神经区域交换信号，直到达到一种总体协调（最小化所有层次的误差）。这种协调无疑是不完美和暂时的，因为误差永远不为零，大脑也在不断变化中。但是当它（或多或少）达到时，编码基本运动指令信息（低级运动学）、由此产生的多模态感觉输入以及我们自己正在进行的目标和目的的区域之间就有了协调。这些目标和目的同样被编码为跨越许多处理层次的分布式模式，必须包含”局部”目标（如转动开关）和更远程的目标（如照亮房间），甚至更远程的目标（如点亮房间）。正是这整个相互支撑的结构网络，分布在许多神经区域中，其可能的配置由学习到的生成模型指定，使我们能够预测自己行动的感觉后果。

PP对镜像系统活动的观点现在应该变得更加清晰了。因为假设我们部署同样的生成模型（但请参见5.8节的一些调整和注意事项）来迎接指定另一个主体活动的感觉信息流？那么在这里，大脑也将被迫找到一组相互一致的活动，跨越许多神经区域，同时适应先验期望和感觉证据。将这个图景应用于杰基尔和海德的谜题案例，Kilner等人（2007）指出：

在这个方案中，从观察行动推断出的意图现在取决于从背景层次接收到的先验信息。换句话说，如果在手术室观察到这个行动，那么”伤害”意图的预测误差会很大，而”治疗”意图的预测误差会较小。在层次结构的所有其他层次上，两种意图的预测误差是相同的。通过最小化总体预测误差，MNS（镜像神经元系统）将推断观察到的运动的意图是治疗。因此，即使两种意图导致相同的运动，MNS也能够推断出唯一的意图。（Kilner等人, 2007, p. 164）

在缺乏所有背景指定信息的情况下，没有机制（当然）能够区分治疗和伤害的意图。然而，PP提供了一个合理的机制（如图5.5所示），允许我们已经知道的东西（体现在用于预测我们自己行动感觉后果的生成模型中）对其他（相似）主体行动背后的意图进行反映背景的推断。


(a)
(b)
Kinematics
Goals
Grasp scalpel
Grasp scalpel
Intentions
To Cure
To Hurt
To Cure
To Hurt
Context
Operating Theatre

图像

图5.5 镜像神经元系统(MNS)预测编码解释的示例

在这里，我们考虑在MNS示例层次结构中的四个归因(attribution)层次：运动学、目标、意图和情境。在(a)中，在没有情境的情况下考虑动作观察，在(b)中，观察相同的动作但现在是在手术室的情境中。条形图描述了预测误差的程度。在(a)中，两种意图预测相同的目标和运动学，因此两种方案中的预测误差相同。在这种情况下，模型无法区分导致该动作的意图。在(b)中，情境对”伤害”目标造成大的预测误差，对”治愈”目标造成小的预测误差。在这种情况下，模型可以区分两种意图。

来源：[Kilner et al., 2007]，第164页。

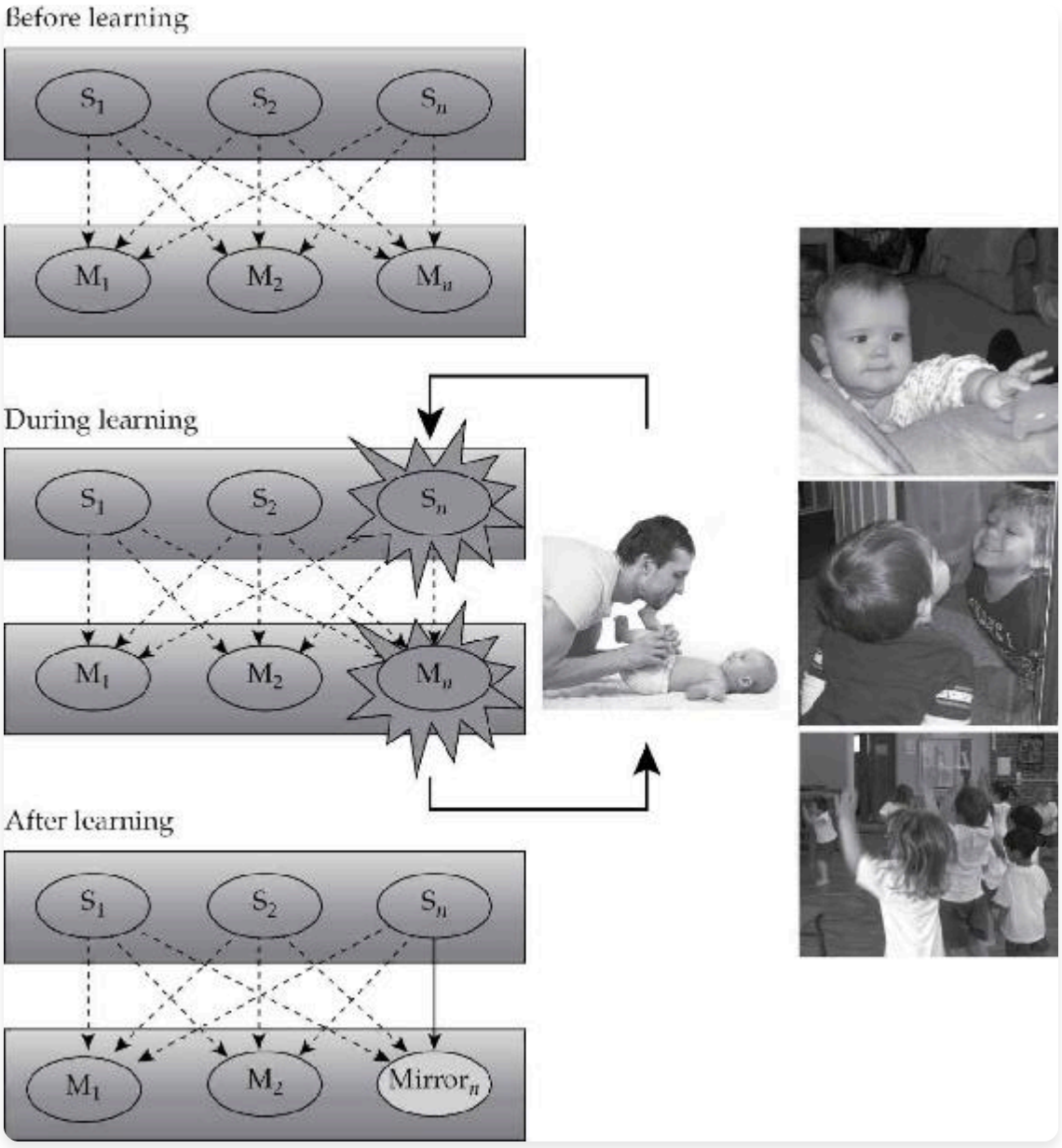
制造镜像

所有这些都表明对镜像神经元本身的某种紧缩观点。紧缩观点([Heyes, 2001]，[2005]，[2010])将个体神经元的”镜像特性”描述为本质上是联想学习过程的直接结果。根据这一观点，“每个镜像神经元都是通过感觉运动经验锻造的——观察和执行相同动作的相关经验”([Heyes, 2010]，第576页)。

这种经验很丰富，因为我们经常观察我们自己正在执行的动作。因此：

每当猴子在视觉引导下执行抓握动作时，运动神经元(参与抓握的执行)和视觉神经元(参与抓握的视觉引导)的激活是相关的。通过联想学习，这种相关激活赋予抓握运动神经元额外的匹配特性；它们成为镜像神经元，不仅在执行抓握时放电，在观察抓握时也放电。([Heyes, 2010]，第577页)

同样，我们可能伸手去拿杯子并观察由此产生的手形，或者吹喇叭并听到发出的声音。在这种条件下(见图5.6)，运动神经元和感觉神经元的相关活动使一些神经元变得多重调谐，既对执行又对被动观察产生反应。这种联想(associations)为我们用来产生和理解自己动作的生成模型提供信息，然后当我们观察其他(足够相似的)主体的动作时可以使用该模型。



image

图5.6 **联想序列学习**

学习前，对观察动作的不同高级视觉特性有反应的感觉神经元(S1、S2和Sn)与一些在动作执行期间放电的运动神经元(M1、M2和Mn)之间存在微弱且无系统的连接(虚线箭头)。产生镜像神经元的学习类型发生在对相似动作各自有反应的感觉神经元和运动神经元存在相关(即连续且偶然)激活时。

来源：[Press, Heyes, & Kilner, 2011]；父亲和婴儿照片 ©Photobac/Shutterstock.com。

这表明(见[Press, Heyes, & Kilner, 2011])镜像神经元和镜像系统如何可能有助于灵活理解其他主体的动作。它们不是通过直接(但神秘地)指定他人的目标或意图来做到这一点，而是通过参与使我们能够在许多空间和时间尺度上预测演化

感觉信号的相同双向多级级联。当所有层次的误差都最小化时，系统已经确定了一个复杂的分布式编码，它包含低级感觉数据、中间目标和高级目标：Jekyll被看作是以某种特定方式挥舞手术刀，意图切割，并希望治愈。

谁干的？

然而，有一个复杂情况(但其中隐藏着一个机会)。当我伸手去拿咖啡杯时，我的级联神经预测包括作为主要组成部分的该伸手动作特有的多种本体感觉张力和延伸感觉。正如我们在第4章中看到的，这些预测是(PP声称)带来伸手动作的原因。但这些预测不应该随意地延伸到我观察另一个主体动作的情况。

这类问题有两种广义的解决方案。第一种涉及创建一个全新的模型（生成模型片段），专门用于预测目标事件。这是一个昂贵的解决方案，尽管我们有时可能会被迫采用这种方案（例如，如果我正在观察某个极其异类的生物或细菌的行为）。然而，最大限度地利用构建我自己行动的生成模型与理解他人行动所需模型之间的任何重叠，会更加高效。这在概念上至少不像听起来那么困难，因为我们已经获得了所需的主要工具。这个工具再次是预测误差信号各个方面的精度加权(precision-weighting)。我们已经看到，精度加权实现了一系列自动和有意注意机制，以及在”自上而下”期望和”自下而上”感官证据之间改变平衡的机制。但正如我们在5.4节中看到的，它还提供了一种通用且灵活的上下文门控手段，允许不同的神经元群体形成软装配联盟（有效连接模式），以响应当前的需求、目标和环境。有了这个工具，适用于自我生成行动情况的预测与适用于观察和理解其他行为主体的预测之间的差异，可以通过改变我们的精度期望来系统性地处理，从而将各种自我/他人区别视为上下文的进一步层次。

这种改变的主要目标是我们自己的本体感受预测。当一些线索告诉我正在观察另一个行为主体时，相对于观察场景的那些方面，本体感受预测误差的精度加权（增益）应该设置得很低。因为根据PP理论，正是本体感受预测误差的最小化直接驱动我们自己的行动，因为这些高精度预测通过运动系统得到满足。当本体感受预测误差的增益设置得很低时，我们就可以自由地部署面向产生我们自己行动的生成模型，作为预测他人行动的视觉后果和理解他们意图的手段。在这种条件下，生成模型其他方面之间的复杂相互依赖性（那些将高层次目标和意图与近端目标以及展开动作的形状联系起来的方面）保持活跃，允许皮质层次结构中的预测误差最小化确定关于观察行为”背后”意图的最佳整体猜测。

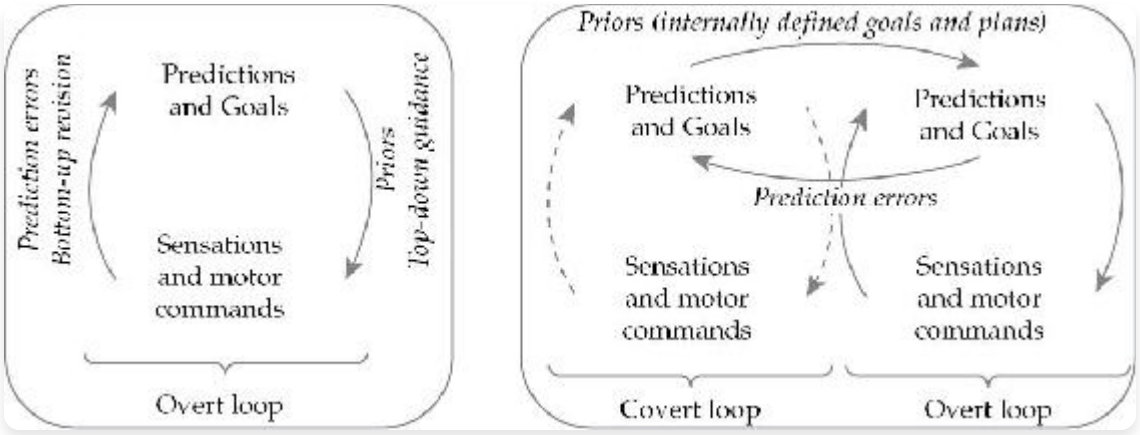
结果是”我们可以在行动或观察时使用相同的生成模型，通过选择性地关注视觉或本体感受信息（取决于视觉运动是由我们自己还是他人引起的）“（Friston, Mattout, & Kilner, 2011，第156页）。相比之下，当进行自我生成的行动时，相关本体感受误差的精度加权必须设置得很高。当本体感受预测误差被高度加权，但被一堆自上而下的预测（其中一些反映我们的目标和意图）适当地解决时，我们感觉自己是自己行动的行为主体。备受讨论的”行动感”的核心方面（在具有正常功能本体感受系统的正常受试者中）依赖于此，预测误差生成和向此类误差分配精度加权的错误，被越来越多地认为是许多行动和控制错觉的根源，我们将在第7章中详细看到。

更一般地说，这意味着支撑我们自己有意运动行为的神经表征与我们建模其他行为主体运动行为时活跃的表征在很大程度上是相同的，并且”完全相同的神经表征可以作为自我生成行动的处方，同时在另一种上下文中，它编码另一个人意图的感知表征”（Friston, Mattout, & Kilner, 2011，第150页）。功能上的明显差异在这里不是归因于核心表征，而是归因于影响其效果的精度估计，反映了发挥作用的不同上下文。因此，镜像特性可能是分层预测处理机制运作的结果，该机制假设感知和行动的共享表征，在其中”大脑不会分别表征预期的运动行为或这些行为的感知后果；大脑中表征的构造既是意图性的又是感知性的，具有感官和运动相关性”（第156页）。这些表征本质上是将目标和意图与感官后果联系起来的跨模态高层次联想复合体。这些状态根据它们发生的上下文具有不同的模态特定含义星座（一些是本体感受的，一些是视觉的，等等）：这些含义通过改变预测误差信号不同方面的精度加权来实现。

5.9 机器人未来

这个相同的广泛技巧可以让我们——如在第3章中介绍的心理时间旅行案例——想象我们自己未来的行动过程，以服务于规划和推理。因为在这里也出现了类似的问题。假设有一个动物掌握了丰富而强大的生成模型，使其能够在多个时间和空间尺度上预测感官信号。这样的动物似乎已经处于有利位置，可以”离线”使用该模型（见Grush, 2004），从而进行心理时间旅行（见[第3章]）想象可能的未来展开并相应地选择行动。但是感知和行动之间的深度紧密关系在这里产生了一个显著的问题。因为根据前面概述的过程模型，预测某个手臂运动轨迹的本体感受后果（举一个非常简单的例子）正是我们实现该轨迹的方式。

解决方案可能再次在于精明的（习得的）精确度加权部署。假设我们再次降低对本体感受错误信号（选择方面）的权重，同时进入一个高级神经状态，其粗略的民间心理学解释可能是”杯子被抓住了”。运动行为被高精度的本体感受期望牵引，在这里不能随之产生。但在这里，生成模型中所有其他相互交织的元素仍然准备以通常的方式发挥作用。结果应该是对伸手动作的”心理模拟”和对其最可能后果的理解。这种心理模拟提供了一种吸引人的方式，从基本形式的具身反应到规划、考虑和”离线反思”能力之间架起桥梁。[18]这种模拟构成了Pezzulo (2012)所描述的”隐蔽循环”。隐蔽循环（见图5.7）：



image

图5.7 隐蔽循环

隐蔽循环允许运行想象中的行动，产生一系列”虚构行动”，从而产生相对于未来（而非现在）事态的预测。

来源：Pezzulo, 2012。

通过抑制显性感官和运动过程离线工作（在主动推理框架中，这需要抑制本体感受）。这允许运行想象中的行动，产生一系列虚构行动和相对于未来（而非现在）事态的预测。虚构行动和预测可以通过自由能[预测错误]最小化进行优化，但无需显性执行：它们不仅仅是”思维漫游”，而是真正受控地朝向在更高层级水平指定的目标。因此，前瞻和规划是支持产生远距离和抽象目标（以及相关计划）的优化过程，超越了当前的可供性(affordance)。（Pezzulo, 2012，第1页）

在这里，核心思想仍然独立于运动命令通过本体感受预测实现的完整过程模型。比这更根本的是这样一个概念：行动产生、行动理解和模拟可能行动的能力可能都得到基于智能体自身感觉运动技能库的单一生成模型的上下文细致调整的支持。

这个想法已经使用各种模拟在微观层面得到研究。因此，Weber等人(2006)描述了一个运动皮层的混合生成/预测模型，提供了这种多重功能。在这项工作中，使机器人能够执行行动的生成模型同时充当模拟器，使其能够预测可能的感知和行动链。然后使用这种模拟能力来实现一个简单但具有挑战性的行为，其中机器人必须以能够抓取视觉检测到的物体的方式靠近桌子。由于大多数靠近桌子的模式都不适合该任务，这为基于机器人自己的”心理模拟”结果的靠近行为提供了一个很好的机会（见图5.8）。

image

图5.8 来自Weber等人(2006)的机器人执行”靠近”动作

注意如果它以一定角度接近桌子，由于其短抓手和侧柱，它无法够到橙色水果。这可能对应于手指、手臂或手处于不适合抓取位置的情况。

来源：Weber等人，2006。

相关想法被Tani等人(2004)和Tani (2007)所追求。Tani及其同事描述了一系列使用”带参数偏置的循环神经网络”(RNNPBs)的机器人实验：一类实现基于预测的分层学习的网络。指导思想是基于预测的分层学习在这里解决了一个关键问题。它允许系统将处理其世界的真实感觉运动掌握与更高层次抽象的出现相结合，这些抽象（关键地）与该掌握协同发展。这是因为这里的学习在更高处理层次产生表征形式，允许系统预测支配在较低层次存在的神经模式（它们自己响应感觉外围的能量刺激）的规律性。

这些都是”基础抽象”(grounded abstractions)的例子（关于这个一般概念，参见Barsalou, 2003；Pezzulo, Barsalou, et al., 2013），它们为更具组合性和策略性的操作打开了大门，如解决新颖的运动问题、模仿其他智能体的观察行为、参与目标导向的规划，以及在离线想象中预测试行为。这种基础抽象并不脱离其在具身行动中的根源。相反，它们构成了可以被视为那些交互的一种”动态编程语言”：在这种语言中，例如，“连续的感觉-运动序列被自动分割成一组可重复使用的行为原语”(Tani, 2007, p. 2)。Tani et al. (2004)表明，配备了这样的原语（作为学习驱动的自组织结果）的机器人能够部署它们来模仿另一个智能体的观察行为。在另一个实验中，他们表明这样的原语也促进了行为到简单语言形式的映射，使机器人能够学会遵循命令，例如，以针对指定对象或空间位置的方式进行指向（使用其身体）、推动（使用其手臂）或击打（用其手臂）。

这组研究在Ogata et al. (2009)中得到进一步扩展，他们使用RNNPB模拟解决了视角转换的重要问题，其中一个机器人观察然后模仿另一个智能体的物体操作行为，将一组学习到的变换应用于自己的自我模型。在这个”认知发展机器人学”实验中，“另一个个体被视为一个动态对象，可以通过投射/转换自我模型来预测”(Ogata et al., 2009, p. 4148)。

这样的演示虽然范围有限，但很有启发性。“可重复使用的行为原语”的出现表明，组合性、可重复使用性和可重新组合性等特征（曾经与经典人工智能的脆弱、笨重的符号结构相关联）可以作为分层设置中概率预测驱动学习的结果而自然产生。但由此产生的抽象现在丰富地基于智能体的过去经验和感觉运动动力学。

[5.10 不安的、快速响应的大脑]

看起来，语境就是一切。不仅仅是”杰基尔对海德”，甚至”自我对他者”在这里都显现为一种根据我们所处语境来重塑和细化我们自己处理程序的能力的表现。但是，语境本身当然必须被识别，而它通常是在语境中被识别的！

这里没有不幸的无限回归。在典型情况下，我们被提供了许多可靠的线索（外部和内部的），这些线索通过残余预测误差信号招募正确的神经资源子集进行微调。清晰的外部线索（在杰基尔对海德的情况下的手术室线索）是明显的例子。但我自己正在进行的神经状态（编码关于我的目标和意图的信息）提供另一种类型的线索，已经建立了各种各样的语境化影响，我自己身体运动对感觉信息作用的许多细粒度效应也是如此（参见O’Regan & Noe, 2001）。两个进一步的成分完成了这幅图画。一个是超快速形式的”要点处理”的可用性，能够在遇到新场景的几百毫秒内提供语境化线索。另一个，真的不能被过分强调，就是不安的、永远期待的大脑的持续活动。

我们在第1章（第13节）中早就看到，PP模式如何支持一个递归协商的”一瞥要点”模型，我们首先识别总体场景，然后是细节。我们也强调，语境的指导作用（在生态正常情况下）很少从我们的心态中缺失。我们几乎总是处于某种或多或少合适的感觉期待状态。如此理解的大脑是一个不安的、主动的(Bar, 2007)器官，不断使用其自己最近和更遥远的过去历史来组织和重组其内部环境，以便为对传入扰动的响应设置场景。

此外，即使在罕见的情况下，我们被迫（也许由于某种巧妙的实验设计）处理一系列不相关的感觉输入，也有精明的技巧和策略支持场景的广泛意义或”要点”的超快速提取。这样的超快速要点提取甚至可以在其他难以捉摸的情况下提供语境，相对于这个语境可以计算合适的精度期望：这样的语境因此锻造有效连接性网络，能够聚集新的、软装配的神经资源联盟，既选择又细化用于满足感觉信息前向流的模型。

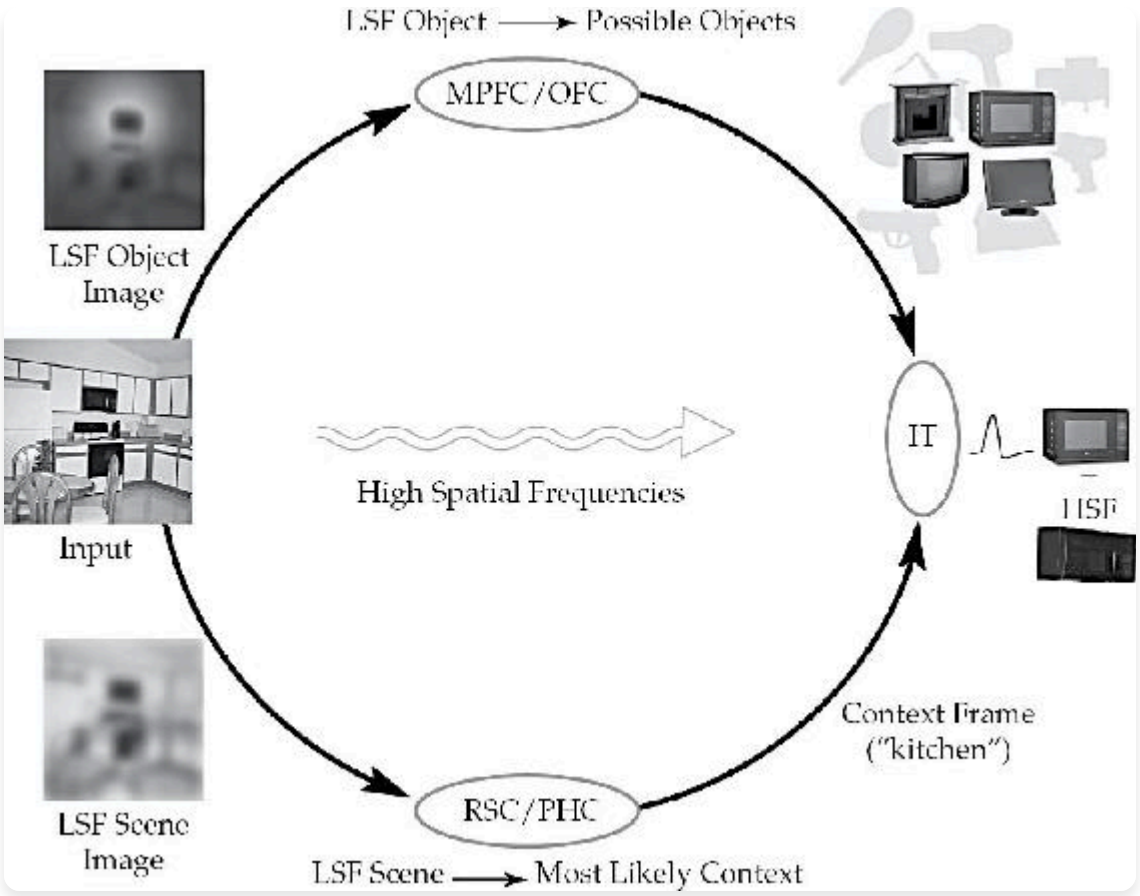
超快速要点提取

超快速要点提取(Ultra-rapid gist extraction)绝非视觉模式的专利。但哺乳动物的视觉系统特别容易理解，这里受益于两个相当不同的处理流的综合资源：快速的巨细胞通路(magnocellular pathway)，其从V1的投射构成所谓的”背侧视觉流”，以及较慢的小细胞通路(parvocellular pathway)，其从V1的投射创建所谓的”腹侧视觉流”。这些是Milner和Goodale在他们的”双视觉系统”理论中使著名的视觉流（参见，例如，Milner & Goodale, 2006）。在预测驱动的神经经济背景下，人们认为这些视觉流分别提供：

一个快速、粗糙的系统，基于部分处理的视觉输入启动自上而下的预测，以及一个较慢、更精细的子系统，由快速激活的预测引导并根据较慢到达的详细信息加以完善。(Kveraga et al., 2007, p. 146)

还有第三个小细胞流(konicellular stream)，尽管在撰写本文时其作用仍不明确（参见Kaplan, 2004）。巨细胞流和小细胞流都显示了前面描述的那种层次组织。但它们也密集且反复地相互连接，创建了一个令人眼花缭乱的前向、后向和侧向影响网络(DeYoe & van Essen, 1988)，其综合效果是（Kveraga等人建议）允许背侧流快速处理的低空间频率信息提供上下文暗示信息来指导物体和场景识别。

这些快速”猜测”的早期阶段产生粗糙且现成的”类比”（用Bar, 2009的词汇）来处理当前输入。作者的意思只是，快速处理的线索支持基于过去经验检索一种高级骨架：这种骨架可以（在大多数情况下）建议关于场景可能形式和内容的*足够*信息，以允许流畅地使用残余误差快速揭示任务要求的任何附加细节（参见图5.9）。这种骨架内容不限于关于场景性质的简单事实（城市场景、办公室场景、运动中的动物等），而且将包括（Barrett & Bar, 2009）基于我们先前情感反应快速检索的场景或事件的”情感要点”元素。通过这种方式：



图片

图5.9 使用低空间频率信息进行快速猜测

与图像细节沿视觉通路的自下而上系统进展并行，存在低空间频率(LSF)信息的快速投射，可能通过巨细胞通路。这种粗糙但快速的信息足以生成关于上下文和其中物体的”初始猜测”。这些基于上下文的预测通过较高空间频率(HSFs)的逐渐到达得到验证和完善（Bar, 2004）。MPFC，内侧前额叶皮层；OFC，眶额皮层；RSC，楔前叶复合体；PHC，海马旁回皮层；IT，下颞叶皮层。图中的箭头是单向的，以强调在所提议分析期间的流向，但所有这些连接在性质上都是双向的。

来源：来自Bar, 2009，经许可使用。

当大脑检测到当前时刻来自眼睛的视觉感觉，并试图通过生成关于这些视觉感觉在世界中指代或代表什么的预测来解释它们时，它不仅使用先前遇到的声音、触觉、嗅觉和味觉模式，以及语义知识。它还使用情感表征——这些外部感觉如何影响来自身体的内部感觉的先前经验。(Barrett & Bar, 2009, p. 1325)

随着处理的进行，情感和内容在这里被共同计算：在解决场景的连贯、暂时稳定的解释过程中交织在一起。这表明，体验世界不仅仅是在许多空间和时间尺度上达成连贯的理解。它是达成一种连贯的、多层次的、情感丰富的理解。这种理解直接准备好引发适当的行动和反应。当我们从一个上下文转到另一个上下文时，我们的大脑不断活跃地试图准备我们处理和回应每种情况。这意味着激活一系列”心理状态”(Bar, 2009, p. 1238)，使用更粗糙的线索招募更详细的猜测，并启动关于可能提供的行动和干预机会的性质和形状的丰富存储知识体。

Bar (2009)指出，这里基于预测的解释与近期认知科学中的其他两个主要研究传统相联系。第一个涉及通过简单线索（如单词或面部表情）自动激活影响行为的刻板印象（参见，例如，Bargh et al., 1996，以及丰富的回顾，Bargh, 2006）。在这里，联系是直接的：广泛预测集合的快速和自动激活提供了一种能够涵盖这些效应以及（正如我们所见）许多其他效应的机制。

第二个问题，稍微复杂一些，涉及所谓的”默认网络”(default network)。这是一组神经区域，据说在我们没有从事任何特定任务时（当我们让思维自由漫游时）会保持强烈活跃，而当注意力集中于外部环境的特定元素时，其活动就会受到抑制（参见Raichle & Snyder, 2007；Raichle et al., 2001）。对这种”静息状态活动”特征的一种可能解释是，它反映了构建和维持一种背景性、滚动式”心态”的持续过程，为我们未来的行动和选择做准备。这种持续活动将反映我们的整体世界模型，包括我们作为主体特有的”需求、目标、欲望、情境敏感的惯例和态度”集合（Bar, 2009，第1239页）。这将提供基线期望集合，当我们处理来自外部（或内部）环境的最粗糙、最粗略的感官线索时，这些期望已经处于活跃状态。这种持续的内源性活动在功能上是强有力的，已被用来（举一个不同的例子）解释为什么受试者对完全相同的刺激[19]会产生不同反应，这些反应方式与自发的刺激前神经元活动系统性相关（Hesselmann et al., 2008）。综合所有这些，我们得出这样一幅图景：大脑从不被动，即使在驱动超快速要点识别的粗糙线索到达之前也是如此。如此理解的”静息状态”绝非安静。相反，它也反映了神经机制的不懈活动，其强迫性预测活动使我们保持在恒定但不断变化的期望状态中。[20]

[5.11 赞美瞬变性]

预测处理(PP)描绘了一个复杂但可快速重新配置的认知架构，其中内部（和外部，见第三部分）资源的瞬时联盟在多种神经增益控制机制的影响下形成和消散。这些机制实现”神经门控”体制，其中脑区之间的影响流动是动态可变的，其中自上而下和自下而上信息的相对影响可以根据我们自身感觉不确定性的估计而不断变化。[21]结果是一个能够将功能分化的回路与高度情境敏感（和”交互主导”）的处理和反应模式相结合的架构。

支撑这种强大组合的是复杂的、习得的”精度期望”(precision expectations)体系，其作用是改变各种系统元素之间获得的影响模式。在观察和理解其他主体的情况下，这些精度期望的最重要作用是降低本体感觉预测误差的权重，允许多层次预测展开而不直接驱动行动。这种降权重也可能提供”虚拟探索”的手段，允许我们想象非现实场景作为推理和选择的指导。

最重要的是（也是最普遍的），可变精度加权雕塑和转移了信息和影响的大规模流动。因此，它们提供了重复重新配置有效连通性模式的手段，使得支撑神经反应的”最简单回路图”本身成为一个移动目标，不断响应快速处理的线索、自生成的行动、变化的任务需求以及我们自身身体（如内感受）状态的改变而变化。如此理解的大脑是一个变形的、嗡嗡作响的动力系统，永远重新配置自身以更好地应对传入的感觉冲击。

[6]

[超越幻想][]

6.1 期待世界

如果大脑是概率预测机器，这对心智-世界关系有什么启示？这是否意味着，正如一些人所建议的，我们体验的只是一种”虚拟现实”或”受控幻觉”？或者，由于所有那些预测重型机械的帮助，我们是否更直接地接触到可能（模糊而有问题地）被称为”世界本身”的东西？此外，我们*似乎*感知到的与概率内部经济之间的关系是什么？我们似乎感知到一个由确定对象和事件组成的世界，一个由狗和对话、桌子和探戈填充的世界。然而，在这些感知背后（如果我们的故事是正确的）是复杂相互交织的概率分布编码，包括对我们自身感觉不确定性的估计。

在不充分考虑行动重要性和我们自身行动技能库的情况下接近这些问题，会让我们严重误入歧途。因为指导世界参与行动，而不是产生”准确”的内部表征，才是预测误差最小化例程本身的真正目的。这改变了我们应该如何思考心智-世界关系以及概率内部经济的形状和范围。认识我们的世界现在必须作为一个更大系统矩阵的一部分落实到位，该矩阵的轴心和核心是具身行动以及适应性成功所需的快速、流畅反应类型。在这个更大的背景下探索预测处理是本书其余部分的任务。

6.2 受控幻觉和虚拟现实

克里斯·弗里斯(Chris Frith)在他关于预测性贝叶斯大脑的精彩著作（2007）中写道：

我们的大脑构建世界模型，并根据到达感官的信号不断修改这些模型。因此，我们实际感知到的是大脑对世界的模型。它们不是世界本身，但对我们而言，它们已经足够好了。你可以说我们的感知是与现实相符的幻想。(Frith, 2007, p. 135)

这让人想起我们在第1章遇到的口号：“感知是受控的幻觉”。[1] 这是受控的幻觉，这种观点认为，因为它涉及使用存储的知识来生成”最佳多层次自上而下的猜测”。这是在多个空间和时间尺度上定义的猜测，最好地解释了传入的感官信号。正是在这种情况下，Jakob Hohwy写道：

这个理论告诉我们关于心智的一个重要且可能不受欢迎的事实是，感知是间接的……我们感知的是大脑的最佳假设，体现在关于外部世界原因的高级生成模型中。(Hohwy, 2007a, p. 322)

在后来的著作中，Hohwy使用”虚拟现实”的概念来描述这种关系。Hohwy认为，意识体验：

产生于大脑对当前感官输入进行最佳理解的渴望，即使这意味着高度重视先验信念。这符合意识体验就像是为了抵御感官输入而构建的幻想或虚拟现实的想法。它不同于真正的幻想或虚拟现实的意识体验，我们在心理意象或梦境中享受的体验，因为这些体验并不是为了抵御感官输入。但它仍然与它所代表的真实世界保持一定距离。(Hohwy, 2013, pp. 137 – 138)

所有这些都有其正确之处，也有一些（我将论证）根本错误的地方。正确的是，这些理论将感知描述为*在某种意义上*是一个推理过程（最初由Helmholtz, 1860提出；另见Rock, 1997）：这个过程不能不在原因（如感官刺激或远端对象）和结果（知觉、体验）之间插入*某些东西*（推理）。这些过程可能出错，由此产生的幻想、妄想和错误状态往往被视为”间接”观点（见，例如，Jackson, 1977）的有力证据，即我们与世界的感知接触是间接的。

此外，如果预测机器模型是正确的，我们正常、成功的日常感知接触世界的大部分——在很大程度上取决于我们对感知场景的期望，就像取决于驱动信号本身一样。更令人惊讶的是，这里感官信息的前向流[2]仅包括错误信号的传播，而内容丰富的预测向下和横向流动，通过相互连接的网络以复杂的非线性方式相互作用。正如在第1章中提到的，这种影响模式的一个关键结果是神经编码使用效率大大提高，因为：“预期事件不需要明确表示或传达给高级皮层区域，这些区域在事件发生之前已经处理了其所有相关特征”(Bubic et al., 2010, p. 10)。在生态正常的情况下，与世界的瞬时感知接触的作用因此”仅仅”是检查并在必要时纠正大脑对外界存在什么的最佳猜测。这是一个具有挑战性的愿景。它将我们的（主要是非意识的）期望描绘为我们感知内容的主要来源[3]：然而，这些内容不断被对不断演变的感官输入敏感的预测误差信号检查、调整和选择。

尽管如此，我认为我们应该抵制这样的说法：我们感知的东西最好理解为一种假设、模型、幻想或虚拟现实。在我看来，这样想的诱惑建立在两个错误之上。第一个错误是将基于推理的适应性反应路径设想为在主体和世界之间引入一种表征面纱。相反，只有从预测驱动学习中提炼出的结构化概率诀窍才能使我们*看穿表面统计的面纱*，看到远端相互作用原因的世界本身。[4] 第二个错误是未能充分考虑行动的作用，以及生物体特定的行动库在选择和不断测试持续的预测流本身中的作用。预测驱动学习并不旨在揭示客观领域的某种行动中性图像，而是提供对*可供性(affordances)*的把握：环境为给定主体提供的行动和干预可能性。[5] 综合考虑这些要点，表明大脑中的概率推理引擎并不构成主体与世界之间的*障碍*。相反，它提供了一个独特的工具，用于遭遇一个充满意义的世界，一个充满人类可供性的世界。

6.3 结构化概率学习的惊人范围

基于预测的学习的一个自然结果是，学习者能够揭示（当一切正常运作时）相互作用的远端原因的网络，这些原因——基于她的行为库存和兴趣——描述了学习发生的可交互环境的特征。通过这种方式，基于预测的学习带来了对结构化外部世界的认识，这个世界由持续存在的（尽管经常在时间上演变的）物体、属性和复杂嵌套因果关系构成。因此，“识别系统’继承’了环境的动力学，并能够准确预测其感官产品”（Kiebel, Daunizeau, & Friston, 2009，第7页）。

在主动智能体中，层次化预测驱动学习的全部力量和范围仍有待确定。它受到时间、数据的限制，以及——也许最重要的——所涉及的神经近似的性质的限制。然而，已经很清楚的是，可处理形式的层次化预测驱动推理能够揭示深层结构，甚至是那些曾经似乎需要提供大量先天知识才能理解的抽象、高层次规律性。这里起作用的一般原则现在已经很熟悉了。我们假设环境通过嵌套相互作用的（远端）原因产生感官信号，感知系统的任务是通过学习和应用层次化生成模型来反转这种结构，以便预测展开的感官流。感觉流（正如我们所见，与行为流保持恒定的循环因果交流）的可预测性正好与该流中的空间和时间模式相对应。但这种模式是世界属性和特征以及智能体的需求、形式和活动的函数。因此，从观察到的足球比赛到达眼睛的感官刺激模式是照明条件、结构化场景以及观察者头部和眼部运动的函数。它也是各种更抽象的相互作用特征和力量的函数，包括每支球队特有的攻防模式、当前比赛状态（例如，当一支球队大比分落后时会有策略性改变）等等。前面章节中描述的计算神经科学和机器学习各波浪工作的美妙之处在于，它们开始展示如何在不需要（尽管可能容易利用）广泛先验知识的情况下学习这种复杂的相互作用原因堆栈。这应该从根本上重新配置我们对先天主义与经验主义辩论的思考，以及对”按自然关节切割”的性质和可能性的思考。

因此，考虑Tenenbaum等人（2011）对层次贝叶斯模型(Hierarchical Bayesian Models, HBMs)的描述。[6] HBM是一种多层处理以特别有效的方式相互激活的模型，每一层都试图解释下一层的激活模式（编码某些变量的概率分布）。当然，这正是在所谓的”过程”层面[7]由层次化预测编码描述的那种架构。当这样的系统启动并运行时，所有多个层面的小假设会稳定到最佳解释传入感官信号的相互一致的集合中，考虑到系统已经学到的内容和当前的感官证据，包括我们所见的系统对该证据可靠性的最佳估计。用第1章中介绍的贝叶斯术语来说，每一层都在学习下一层的”先验”。这整个多层过程由传入的感官信号调节，并实现被称为”经验贝叶斯”的策略，允许系统在学习过程中从数据中获取自己的先验。

这种多层学习有一个额外的好处，即它非常自然地适合于将数据驱动的统计学习与早期联结主义和人工神经网络工作的反对者长期坚持的那种系统性生产性知识表示结合起来。[8] HBMs（与那些早期形式的联结主义不同）实现了适合表示复杂、嵌套、结构化关系的处理层次（有关一些很好的讨论，见Friston, Mattout, & Kilner, 2011；Tani, 2007；以及5.9中的讨论）。要在微观层面看到这一点，回想一下引言中描述的理想化地层学程序SLICE。*SLICE*有效地体现了关于地质原因的生产性和系统性知识体系。因为SLICE*可以产生其生成模型中表示的隐藏原因的可能组合和重组所允许的全套地质结果。

通过将多层生成模型的使用与强大的统计学习形式相结合（实际上，使用这种学习来诱导这些模型本身），我们因此获得了早期连接主义（‘联想主义’）和更经典的（‘基于规则’）方法的许多好处。此外，不需要固定在任何单一形式的知识表示上。相反，每一层都可以自由使用最能使其预测和（因此）解释下一层活动的任何形式的表示。在许多情况下，似乎出现的（正如Tenenbaum等人努力强调的）是结构化的、生产性的知识体系，但这些知识体系仍然是基于由原始训练数据中可见的统计规律驱动的多阶段学习而获得的。这里的早期学习诱导出总体期望（例如，关于在给定领域内成功分类最重要的事物类型的非常广泛的期望）。这种广泛的期望然后约束后续学习，减少假设空间并实现对特定案例的有效学习。

使用这些例程，HBM最近被证明能够基于本质上的原始数据学习许多领域的深层组织原则。例如，这些系统已经学会了所谓的’形状偏见’，根据这种偏见，属于同一对象类别的项目（如起重机、球和烤面包机）往往具有相同的形状：这种偏见不适用于物质类别，如金、巧克力或果冻。它们还学会了什么样的语法（上下文无关或正则）最能解释儿童导向语音语料库中的模式，关于将未分段语音流正确解析为单词，以及一般性地了解许多不同领域中因果关系的形状（例如，疾病导致症状，而不是相反）。最近的工作还显示了当同化到现有类别需要过于复杂——因此实际上是’临时性’——映射时，如何产生由新因果模式定义的全新类别。这些方法还被证明能够通过汇集在广泛案例中获得的证据，快速学习高度抽象的领域通用原则，如对因果性的一般理解。综合来看，这项工作证明了使用HBM学习的意外力量。这些方法允许系统通过将适当的多层系统暴露于原始数据来推断特定于领域的高层结构，甚至推断支配多个领域的高层结构。

需要注意的一个重要点是，HBM在这里允许学习者在’填充’关于个别样本的细节之前，获得一个领域特征性的图式关系。例如，通过这种方式：

语法归纳的分层贝叶斯模型可能能够解释儿童如何对语法的某些属性变得自信，即使支持这一结论的大多数个别句子都理解得很差。

类似地，对象的形状偏见可能在学习任何个别对象的名称之前就被学会了。这种偏见作为最佳的高层模式早期出现，一旦建立，它就能够快速学习属于该组的特定样本。这在’儿童能够获得大量…噪声观察[使得]任何个别观察可能难以解释，但综合起来可能为一般结论提供强有力支持’的情况下是可能的。因此，作者继续说，在形成关于球、圆盘、毛绒玩具等个别具体对象的任何想法之前，人们可能有足够的证据表明视觉对象往往是’有凝聚力的、有界的和刚性的’。

当然，这正是容易被误认为是先天世界知识影响证据的早期学习模式类型。这种错误是自然的，因为高层知识是针对领域量身定制的，并允许后续学习比其他情况下预期的更容易和更流畅地进行。但HBM式学习者不是依赖丰富的先天知识体系，而是能够从数据中诱导出这种抽象的结构化知识。正如我们所看到的，核心技巧是以一种多阶段的方式使用数据本身。首先，数据被用来学习编码关于领域大规模形状期望的先验（Tenenbaum等人称之为领域内’结构形式’）。通过这种大规模（相对抽象）期望结构的适当支撑，关于更详细规律的学习变得可能。通过这种方式，HBM积极发掘抽象的结构期望，这些期望使它们能够使用原始数据学习越来越细粒度的模型（支持越来越细粒度的期望集合）。

这些系统——就像更广泛的PP系统一样——也能够从数据中诱导出它们自己所谓的”超先验”(hyperpriors)。超先验(hyperpriors)（在这里与”过度假设”(overhypotheses)可互换使用，见Kemp et al., 2007）本质上是”先验之上的先验”，体现了关于世界非常抽象（有时几乎是”康德式”）特征的系统性期望。例如，一个高度抽象的超先验可能要求每组多模态感官输入都有一个最佳解释。这将强制概率分布具有单一峰值，从而确保我们总是将世界视为处于一种或另一种确定状态，而不是（比如）处于等概率状态的叠加。这样一个极其抽象的超先验可能是先天规范的良好候选。但它同样可以留给早期学习，因为需要使用感官输入来驱动行为，以及同时以两种截然不同的方式行动的物理不可能性，可以设想地推动HBM提取这一点作为支配推理的一般原则。

HBM，以及可能实现它们的各种过程模型（包括PP），使贝叶斯理论家免于在成功学习之前需要预先设置正确先验的明显罪责。相反，以经验贝叶斯的方式，多层系统可以从数据中学习自己的先验。这也提供了最大的灵活性。因为虽然现在很容易将反映知识的抽象领域结构（以各种超先验的形式）构建到系统中，但系统也有可能获得这样的知识，并在它既简化又使之成为可能的更详细学习之前获得这样的知识。这样构想的先天知识在发展上仍然部分”开放”，因为它的某些方面可以通过使用相同多层过程的数据驱动学习来平滑和完善，甚至完全消除（对于一些很好的讨论，见Scholl, 2005）。

当然，正如李尔王著名地评论的那样，“无中不能生有”，如上所述，即使是最精简的学习系统也必须始终从某些偏见集合开始。更重要的是，我们的基本进化结构（大体神经解剖学、身体形态等）本身可以被视为一套特别具体的内置（体现的）偏见，它们构成我们整体世界”模型”的一部分（见Friston, 2011b, 2012c，以及8.10中的讨论）。尽管如此，多层贝叶斯系统已经证明能够获得抽象的、领域特定的原则，而无需构建许多以前被认为对不同领域流畅学习至关重要的相当具体的知识类型（例如，关于形状对学习物质对象重要性的知识）。这样的系统可以从原始感官数据中获得相当抽象的组织原则的知识——这些原则然后允许它们对那些数据做出越来越系统的理解。

6.4 为行动做好准备

然而，6.3中的讨论是根本不完整的。它根本不完整，因为当前大多数关于HBM的工作将认识世界视为被动感知的模型。但是感知和行动，正如使用PP模式构建的，被看作是共同决定和共同确定的（见上面的第2章、第4章和第5章）。在这些更广泛的框架中，我们所做的取决于我们所感知的，而我们所感知的不断受到我们所做的条件限制。这导致了2.6和第4章中描述的相当特定的循环因果关系形式。在这里，高层预测引发的行动既测试又确认预测，并帮助塑造招募新高层预测的感官流（以此类推，在期望、感官刺激和行动的滚动循环中）。

我们应该如何看待这个滚动循环？在我看来，错误的模型是将滚动循环描绘为经典的感知-思考-行动循环的稍微花哨（因为是循环的）版本，在这个版本中，感官刺激必须被充分处理，外部对象的结构化世界必须被揭示，然后才能选择、计划和（最终）执行行动。这样的”感知-思考-行动”愿景已经为认知心理学和认知科学的大量工作提供了信息。它被Cisek (2007)很好地描述（然后被彻底拒绝），他指出：

根据这种观点，感知系统首先收集感官信息来构建外部世界中对象的内部描述性表示(Marr 1982)。接下来，这些信息与当前需求的表示和过去经验的记忆一起被用来做出判断并决定行动方案(Newell & Simon 1972; Johnson-Laird 1988;

Shafir & Tversky 1995)。然后使用所得计划来生成运动的期望轨迹，最后通过肌肉收缩来实现(Miller et al. 1960; Keele 1968)。换句话说，大脑首先使用独立于行动的表示来构建关于世界的知识，然后这种知识被用来做出决策、计算行动计划并最终执行运动。(Cisek, 2007, p. 1585)

对这样的模型保持警惕有很多理由（参见 Clark, 1997; Pfeifer & Bongard, 2006，以及 第三部分 中的进一步讨论）。但最令人信服的理由之一是需要准备好对不断展开的——以及可能快速变化的——情况做出流畅响应。这种准备状态对于那些必须随时准备抓住机会、避开危险的生物来说，在生态学上似乎是必需的，这些生物可能在与他人（有时包括自己的同类）竞争中行动。配备了永远活跃的预测性大脑的生物，当然已经（非常字面意义上地）“领先于游戏”，因为这样的大脑——正如我们所看到的——不断猜测持续的感觉输入流，包括应该由它们自己的下一步行动和世界干预产生的输入。但故事并没有就此结束。

一个强有力的策略，它与永远活跃的预测性大脑的形象很好地结合（我将论证），涉及将经典的感知-思考-行动循环重新思考为一种马赛克：一种马赛克，其中每个碎片都结合了（可能在经典意义上被认为是）感知和思考的元素以及相关的行动处方。这种马赛克愿景的核心（其根源在于主动视觉范式，参见 Ballard, 1991; Churchland et al., 1994）在于同时计算多个概率感染的”可供性(affordances)“：多种对生物体显著的行动和干预可能性。

这一观点的旗舰陈述是”可供性竞争假说”（Cisek, 2007; Cisek & Kalaska, 2010）。这样的观点受到大量异常神经生理学数据的推动（完整回顾请参见 Cisek & Kalaska, 2010），包括：

1. 持续未能找到完整”被动重建”模型所预测的那种世界内在表征，在这种模型中，例如视觉处理的目标是生成一个单一的丰富、统一、行动中性的场景表征，适合随后用于规划和决策。

2. 注意调节的普遍影响，根据任务和情境导致对持续神经活动不同方面的增强和抑制（因此神经响应似乎适应当前行为需求，而不是构建外部世界状态的行动中性编码）。

3. 越来越多的证据表明，参与持续规划和决策的神经群体也参与运动控制，并且（更一般地说）区域皮质响应未能遵守感知、认知（例如推理、规划和决策）和运动控制之间的经典理论划分。

因此，视觉神经科学的大量工作表明，多个不同的信息体持续并行计算，并且只有当（且仅当以及在何种程度上）某些当前行动或响应需要时才部分整合。这里一个熟悉的例子是将视觉信息分离为不同的（尽管重叠的）腹侧和背侧流：这些流，越来越清楚地，通过持续的、任务敏感的信息交换模式联系起来（例如，参见 Milner & Goodale, 1995, 2006; Schenk & McIntosh, 2010; Ungerleider & Mishkin, 1982）。在神经经济中，这种分离和部分性似乎是规则而不是例外，不仅表征每个流内部的处理以及流之间的处理，也表征大脑其他部分的处理（例如，参见 Felleman & Van Essen, 1991; Stein, 1992）。

还有充分的证据表明注意的普遍影响，反映情境和任务，以及内部情境，形式为内感受状态如饥饿和无聊。这些影响已被证明调节皮质层次结构各个层面的神经响应，也包括一些皮质下（例如丘脑）区域（Boynton, 2005; Ito & Gilbert, 1999; O’Connor et al., 2002; O’Craven et al., 1997; Treue, 2001）。

最后，越来越多且极具启发性的证据挑战了这样一种观点：核心认知能力（如规划和决策）在神经生理学上与感觉运动控制回路截然不同。例如，眼球运动决策和眼球运动执行在外侧顶内叶区(LIP)、额叶眼动区(FEF)和上丘中招募了高度重叠的回路——正如Cisek and Kalaska (2011)巧妙地指出的那样，上丘是”一个脑干结构，距离移动眼球的运动神经元只有两个突触的距离”（第274页）（相关研究见Coe et al., 2002；Doris & Glimcher, 2004；以及Thevarajah et al., 2009）。同样地，一项感知决策任务（其中决策通过手臂运动报告）显示了运动前皮层内与决定反应过程相对应的明显反应（Romo et al., 2004）。更一般地说，无论何时决策需要通过（或以其他方式调用）某种运动行为来报告，感知-运动处理和决策制定看起来都是交织在一起的，这使得Cisek和Kalaska提出”决策，至少是那些通过行动报告的

决策，是在负责规划和执行相关行动的相同感觉运动回路内做出的”（Cisek & Kalaska, 2011，第274页）。在后顶叶皮层(PPC)等皮层联合区域中，Cisek和Kalaska进一步论证，活动似乎完全不遵循感知、认知和行动之间的传统划分。相反，我们发现神经元群体从事的是感知、决策和行动的变化性和情境响应性组合，其中甚至单个细胞都可能参与许多这样的功能（Andersen & Buneo, 2003）。

Selen et al. (2012)的研究为这种观点提供了进一步支持。在这项研究中，受试者被展示移动点的显示（“动态随机点显示”），并被要求判断这些点主要向左还是向右移动。已知这类决策非常敏感地依赖于点运动的连贯性和持续时间，实验者改变这些参数，同时通过在刺激开始后的不可预测时间要求决策来探测受试者的决策状态。受试者的任务是在显示停止后立即做出反应，并通过运动反应（将手柄移动到目标）来完成。此时，对受试者的肘部施加小的扰动，引起拉伸反射反应，使用肌电图(EMG)技术测量，这是一种记录与肌肉活动相关的电位的技术。这提供了探测时运动反应效应器（手臂）状态的可量化测量。重要的是，这里的决策任务本身表现得相当良好，因此受试者的选择与不断发展的证据的精细细节（移动点显示的连贯性和持续时间的精确混合）紧密相关。实验者表明，变化的肌肉反射增益和决策变量（代表连贯性和持续时间的综合效应）以一种与经典”顺序流”模型完全不兼容的方式共同演化。在顺序流模型中，报告决策的运动行为被认为独立于决策本身，并且在决策之后计算。相比之下，Selen等人发现每个时刻的反射增益都反映了不断演化的决策状态本身。结果与”可供性竞争(affordance competition)“的概念非常吻合（下文将详述），在这种竞争中，两种可能的运动反应都在同时准备，并且”人脑不会等待决策完成后才招募运动系统，而是传递部分信息，以分级方式为可能的行动结果做准备”（Selen et al., 2012，第2277页）。

也就是说，反射增益”不仅仅反映决策的结果，而是在决策形成过程中参与大脑的考虑过程”（第2284页）。

从决策过程到运动系统的持续流动是有意义的，如果我们假设总体目标是尽可能主动地准备执行不断演化的证据所建议的任何反应。这种主动准备，要真正有用，必须是多重的和分级的。它必须允许许多可能的反应同时部分准备，准备程度取决于当前的证据平衡——包括对我们自己感觉不确定性的估计——随着越来越多信息的获取（关于应用于语音选择的类似情况，见Spivey et al., 2005，另见Spivey et al., 2008）。

在一个具有启发性的结论评论中，Selen等人推测，这可能不仅仅源于这种行动准备的实用优势，还可能源于”大脑评估证据的装置与运动功能控制之间更深层的联系”，并补充说”我们实验中展示的流动可能是决策制定和运动控制基础大脑过程之间更大的双向相互作用的一部分”（第2285页）。同样，Cisek和Kalaska评论道：

感知、认知和运动系统之间的区别可能并不反映感觉引导行为背后神经计算的自然分类[而且]串行信息处理的框架可能不是大脑全局功能架构的最佳蓝图。(Cisek & Kalaska, 2010, p. 275)

作为替代蓝图，Cisek和Kalaska探索了”可供性竞争假说”(Affordance Competition Hypothesis)(由Cisek, 2007引入)，根据该假说：

大脑处理感觉信息以并行地指定当前可用的几种潜在行动。这些潜在行动相互竞争以获得进一步处理，同时收集信息来偏向这种竞争，直到选择出单一反应(Cisek, 2007, p. 1585)

这里的想法是，大脑不断地计算——部分地和并行地——一大套可能的行动，这种部分的、并行的、持续的计算涉及神经编码，这些编码不遵循感知、认知和行动之间熟悉的区别。其原因是，正如Cisek和Kalaska (2011, p. 279)所说，所涉及的神经表征是”实用的”(pragmatic)，因为”它们适应于产生良好控制，而不是产生感觉环境或运动计划的准确描述”。所有这些都具有良好的生态意义，允许时间紧迫的动物部分地”预计算”多种可能的行动，其中任何一种都可以在短时间内被选择和部署，并且只需最少的进一步处理。

大量神经生理学数据为这种观点提供了支持。例如，Hoshi和Tanji (2007)发现猴子前运动皮层的活动与双手协调伸手反应任务中任一手的潜在运动相关，在该任务中猴子必须等待指示使用哪只手的提示(另见Cisek & Kalaska, 2005)。类

似的结果已在视觉扫视的准备中获得(Powell & Goldberg, 2000)，以及使用人类受试者伸手行为的行为和损伤研究(Castiello, 1999; Humphreys & Riddoch, 2000)。此外，正如我们在前一节中看到的，有令人着迷的证据表明”关于行动的决策在定义这些行动物理属性并指导其执行的相同细胞群中出现”(Cisek & Kalaska, 2011, p. 282)。

这里出现的图景是神经编码*根本上*从事行动控制的图景。这种编码以与如何作用于世界的信息在多个层面纠缠的方式表征世界如何存在。可供性竞争假说表明，许多这样的”面向行动的”(action-oriented)(见Clark, 1997)对世界的理解在每时每刻都在被准备，尽管只有少数能超越实际运动反应控制的阈值。[13]

[6.5 实现可供性竞争]

所有这些洞察都被巧妙地容纳了，或者我现在要建议的是，使用神经组织预测处理模型的独特资源。为此，我们利用预测处理框架的三个关键属性。第一个涉及支持感知和行动的表征的概率性质。第二个涉及感知、认知和行动的计算亲密性。第三个涉及由此产生的有机体与环境之间独特的循环因果互动形式。可供性竞争然后作为概率性面向行动预测的自然结果出现。

回想(1.12)概率贝叶斯大脑编码条件概率密度函数，反映了在可用信息给定的情况下某种事态的相对概率。因此，在每个层面上，表征的基本形式仍然是完全概率性的，编码一系列关于”外面是什么”和(我们当前的重点)如何最好地行动的深度纠缠的赌注。因此，多种竞争的行动可能性不断被计算，尽管只有获胜的(高精度)本体感觉预测才能作为运动命令本身发挥作用。

在Friston及其同事建议的行动模型中(见第4章和第5章)，高精度本体感觉预测误差产生运动行动。当本体感觉预测误差被给予高精度时，参与行动产生的完全相同的神经群体可能因此被”离线”部署(再次，见第5章)作为产生适合推理、选择和规划的运动模拟的手段。这为行动控制中涉及的神经群体与推理、规划和想象中涉及的神经群体之间的许多重叠提供了令人信服的解释。然而，在规划中，我们必须减弱或抑制通常会(在主动推理模型上)驱动我们肌肉的下降预测误差。这有效地将我们与世界隔离，允许我们使用分层生成模型在反事实(“如果”)模式下进行预测。感觉减弱的问题也将在自制行为和能动性归因的背景下变得重要(正如我们将在第7章中看到的)。

最后，也许最重要的是，行动（回顾第4章）现在涉及一种强效的循环因果形式，其中被招募来解释当前感官刺激的表征同时决定行动，这些行动导致新的感官刺激模式，招募新的运动反应，如此循环。这意味着我们面临着——正如可供性竞争假说所建议的——一个经济体，其中多个竞争性概率赌注不断进行，本质上是在一个循环因果的感知-行动机器内。

预测处理(PP)因此实现了Cisek和Kalaska所描述的独特循环动力学，他们使用了美国实用主义者约翰·杜威的一句名言。杜威拒绝刺激引发反应的”被动”模型，而支持一个主动和循环的模型，其中”运动反应决定刺激，就像感官刺激决定运动一样真实”（Dewey, 1896，第363页）。这个想法在同一篇文章中的另一段引文中得到了很好的澄清，杜威将看见描述为一个”不间断的行为”：

正如所体验的，它不再仅仅是感觉，也不仅仅是运动（尽管旁观者或心理观察者可以将其解释为感觉和运动），它绝不是刺激伸手的感觉；正如我们已经充分指出的，我们只有*行为*协调中的连续步骤。但是现在考虑一个孩子，在伸手去拿明亮的光（即行使看见-伸手协调）时，有时有愉快的体验，有时找到好吃的东西，有时烫伤了自己。*现在反应不仅是不确定的，刺激也同样不确定；一个不确定只是因为另一个不确定*。真正的问题可以同样恰当地表述为发现正确的刺激、构成刺激，或者发现、构成反应。（Dewey, 1896，第367页，原文强调）

杜威的描述优雅地预示了预测处理所突出的复杂相互作用，即改变我们的预测以适应证据（“感知”）和寻找证据以适应我们的预测（“行动”）之间的相互作用。但它也表明（我认为是正确的），在预测处理框架内，我们真的

不应该将这些视为竞争策略。[14]相反，这两个方面不断协同工作，揭示一个在某种意义上在行动中构成的世界。因为行动现在揭示导致更多行动的证据，我们对世界的体验由这个持续的循环构成。

这些循环因果回路发挥两个经常重叠的作用。一个作用当然是实用的。作为迎面而来的车辆在错误车道上的高级感知状态将招募快速移动方向盘的运动命令，导致新的感知状态，根据其内容必须微调或试图否定所选择的行动方案。另一个作用是认知的。头部和眼部的运动被快速部署来测试和确认假设（迎面而来的车辆在碰撞轨道上）本身。只有能够承受这种自动生成测试的假设才会被维持和加强（见Friston，Adams等人，2012）。通过这种方式，我们采样世界以最小化对我们自己预测的不确定性。

要看到这在实践中的样子，请思考早期、模糊的感官刺激激流将产生预测误差，招募多个竞争的感知假设。然而，这些假设并非行动中性的。相反，每个假设已经涉及刚才描述的两种行动形式。也就是说，每个假设包括关于如何作用于世界以确认或否认假设的信息，以及（由于反映上下文的精确期望主体）关于假设证明正确时如何在世界中行为的信息。受任务和时间各种约束的限制，一个好的策略将是允许最有希望的假设启动一些便宜的认知行动（如快速序列的跳视），能够确认或否认假设。随着这种循环展开，感知将与各种形式的运动规划和行动选择密切结合，导致前面描述的独特神经生理特征。

Spivey（2007）描绘了这种连续循环因果网络的丰富图景，并将其动力学描述为朝着不断变化、永远无法完全达到的稳定终点的持续旅程。作为例证，Spivey要求我们考虑眼部运动（运动产出，尽管在快速、小规模上）与可能被说成——在某些方面误导性地，正如我们即将看到的——指导它们的认知过程之间的相互作用。在现实世界环境中，Spivey指出：

大脑不是实现稳定感知，然后进行眼部运动，然后实现另一个稳定感知，然后进行另一个眼部运动，如此循环。眼睛经常在试图实现稳定感知的过程中移动。这意味着在感知能够完全稳定到稳定状态之前，眼球运动输出通过将新的和不同的信息放置在中央凹上来改变感知输入。（Spivey，2007，第137页）

因此，视觉感知不断受到视觉运动行为的影响，而视觉运动行为也不断受到视觉感知的影响。就成功行为而言，重要的是由此产生的感知运动轨迹(perceptuomotor trajectories)。正是这些轨迹，而不是沿途产生的感知的稳定性甚至真实性，构成了主体性行为(agentive behaviour)，并决定了我们与世界互动尝试的成功或失败。

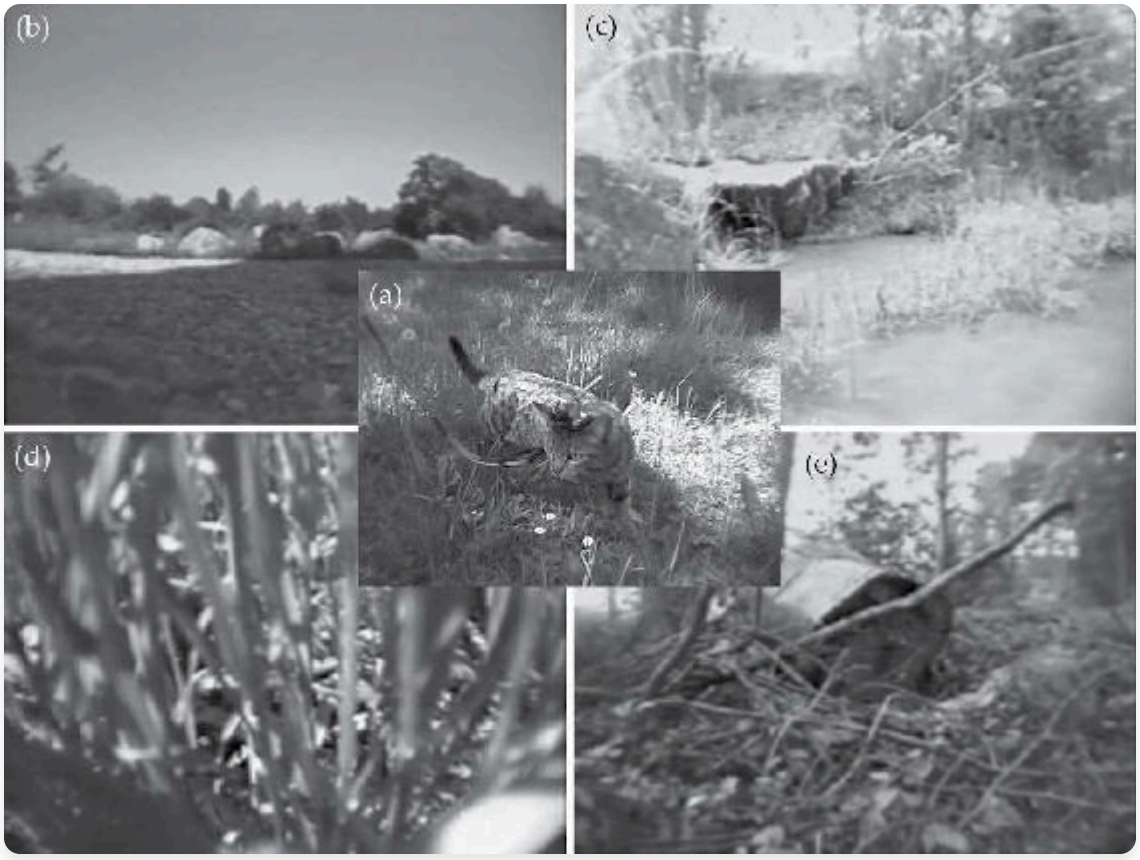
总之，感知运动轨迹在循环因果网络中产生并得以维持。在这些网络中，估计的不确定性和行动需求调节着强烈的可供性竞争(affordance competition)。这是因为精度估计(precision estimations)及其暗示的实用性和认识性行动发挥作用（Cisek & Kalaska, 2011，第282页），“增强环境中最具行为显著性的信息，使感觉运动系统偏向于最具行为相关性的可能行动”。精度估计将假设定位为获得行为控制权，此时它必须变得自我驱动（参与确认性循环因果交换）或消失。精度估计在竞争假设之间的横向（层级内）抑制背景下工作，将假设定位为获得行为控制权，此时它必须变得自我驱动（参与确认性循环因果交换）或消失。我认为，所有这些都提供了一种特别清晰和生动的意义，即许多神经表征必须是”实用性的”，同时建立了一个更大的框架，在这个框架中，可供性竞争作为基于概率的行动导向预测的自然结果而出现。

[6.6 自然界中基于互动的连接点]

所有这些都对关于我们与世界感知接触本质的辩论产生了影响。基于概率预测驱动的学习提供了一种机制，能够（当一切进展顺利时）看穿感觉信号中的表面噪音和模糊性，揭示远端领域本身的形状。在这种（受限的）意义上，它提供了一种”在连接点处切割自然”的强大机制。但现在看来很清楚，许多这样的连接点是*基于互动的*：它们相对于一个具有特定需求和行动与干预可能性的主动有机体而定义。因此，我们对世界的感知”理解”不断受到我们自己的”行动库”(König et al., 2013)与我们的需求、项目和机会相互作用的影响。

这个简单（但深刻）的事实通过帮助在任何给定时刻选择要处理的特征和要尝试预测的事物，大大降低了计算复杂性。从给定冲击性能量流的解析世界的巨大可能方式空间中，我们的大脑试图预测服务于我们需求并适合我们行动库的模式。这很可能导致（正如我们将在第8章中详细看到的）使用简单模型，其力量恰恰在于它们未能编码感觉阵列中存在的每个细节和细微差别。这不是与世界真正接触的障碍——相反，这是其先决条件。因为认识世界，在对进化有机体唯一重要的意义上，意味着能够在那个世界中行动：能够快速有效地响应显著的环境机会。

在像我们这样的大脑需要预测的许多事物中，一个重要的子集涉及我们自己行动的感觉后果。一旦接受了这一点，正如König et al. (2013)所指出的，感觉和运动处理的密切关系就不足为奇了。此外，我们自己行动的感觉后果深受关于我们具身性的基本事实的影响，比如我们的大小、传感器的位置和效应器的触及范围等等。这种影响通过一项分析（Betsch et al., 2004；Einhäuser et al., 2009）得到了戏剧性的展示，该分析研究了从自由移动的猫探索一些户外环境的视角获得的视觉输入的统计结构（使用头戴式摄像头）。当从猫的视角探索时（见图6.1），同样的总体外部领域产生了自然图像序列，其统计结构包括水平轮廓的主导地位、对比度的空间分布改变，以及归因于猫头部运动惊人速度的各种效应。



图像

图6.1 **从猫的视角看世界**

a. 猫在其散步过程中正在探索一个户外公园区域。电缆与牵引绳缠绕在一起，连接到背包中的录像机。(b-e) 显示了从视频中拍摄的四张典型图片。(b) 地平线将图像分为明亮、低对比度的上部图像区域（天空）和较暗、高对比度的下部区域（石头）。(c) 猫对池塘的视角显示了丰富详细的植物结构和低对比度的水域。(d) 通过猫的近距

离观察，草叶均匀分布在整个图像上。(e) 在附近森林中散步时，上半部分由明亮天空前的深色垂直取向树木主导。代表森林地面的图像下半部分由许多以各种可能方向排列的物体（树枝、树叶）组成。

来源：Betsch et al. 2004。

自然场景的统计特性，正如某种动物在行动中遇到的那些场景一样，也被写入到该动物所表现的皮层活动模式中（包括自发性和诱发性活动）。这一点在Berkes等人（2011）关于清醒雪貂V1活动的研究中得到了巧妙的证明。该研究在雪貂发育的各个阶段对这种活动进行了分析，并在三种条件下进行：观看自然场景电影、在黑暗中以及观看非自然场景电影。研究发现，自发性和诱发性反应之间的相似性随着年龄的增长而显著增加，但*仅仅*是在由自然场景诱发的反应方面。作者认为，这种结果模式最好的解释是”内部模型在神经层面上对自然刺激统计特性的渐进适应”（Berkes等人，2011，第83页）。换句话说，雪貂的自发性神经活动模式在发育时间内缓慢适应，以反映”内部模型的先验期望”（第87页）。采用贝叶斯视角，作者提出自发性皮层活动反映了构成内部模型的先验期望的多层结构，而刺激诱发活动则反映了后验概率——某种特定环境原因组合引起感觉输入的概率，因此是动物对当前世界状态的”最佳猜测”。这一诊断通过对暴露于非自然场景电影的成年雪貂的测试得到了强化，在这些测试中记录到了自发性和诱发性活动之间更大的差异。Berkes等人总结道，自发性皮层活动在这里显示了雪貂世界的逐渐适应内部模型的所有特征。

这样的模型可能如何被用来指导行动呢？在预测处理(PP)中，从给定主体的行动库中进行选择是通过反映上下文和任务的精度分配来完成的。这样的分配根据对感觉信号不确定性的估计来控制突触增益。特别是（见Friston, Daunizeau, Kilner, & Kiebel, 2010，以及第4章和第5章中的讨论），本体感觉预测误差的精度加权被认为实现了一种”运动注意”，这种注意必然参与到任何运动行为的准备中。因此，注意力起到”在运动准备期间提升本体感觉通道增益”的作用（Brown, Friston, & Bestmann, 2011，第2页）。同时，正是对感觉预测误差所有方面的精度加权共同决定了哪些感觉运动回路赢得了行为控制的竞争。结果，正如可供性竞争假说(Affordance Competition hypothesis)所描述的那样，是从当前上下文和有机体状态已经建议（因此部分激活）的各种行为反应中选择一种。实现本体感觉预测误差精度加权的机制因此用于选择”具有可供性的显著表征，即预测感知和行为后果的感觉运动表征”（Friston, Shiner, 等人，2012，第2页）。该模型提出，我们所做的事情是由精确的（高度加权的）预测误差决定的，这些误差有助于在竞争的高层假设中进行选择（同时响应这些假设），每个假设都暗示着一整套感觉和运动预测。这样的高层假设本质上是充满可供性的：它们既代表世界是什么样的，也代表我们可能如何在那个世界中行动（因此它们是Millikan（1996）所称的”Pushmi-pullyu表征”的一种：具有描述性和命令性内容的状态）。感知通过招募显著的充满可供性的表征，使我们接触到一个已经为行动和干预而解析的世界。

可以说，正是*因为*我们遇到的世界必须为行动和干预而解析，我们才能在体验中遇到一个相对明确、确定的世界。如果去掉对行动的需要，广义的贝叶斯框架可能看起来与关于意识感知体验的现象学事实相当不符：可以说，我们的世界看起来并不像是被编码为一组相互交织的概率密度分布。相反，它看起来是统一的，在晴朗的日子里是明确的。然而，在一个主动参与世界的系统的背景下，这样的结果具有适应性意义。因为所有这些概率投注的目的是驱动行动和决策，而行动和决策没有能够无限期地保持所有选项开放的奢侈。相反，可供性竞争必须不断得到解决，行动必须被选择。精度加权预测误差提供了一个通过选择最显著的感觉运动表征——那些最适合驱动行为和反应的表征——来偏向处理的工具。

如前所述，生物系统可能受到各种学习或先天的”超先验”(hyperpriors)信息的指导，这些信息涉及世界的一般性质。其中一个超先验可能是世界通常处于一种或另一种确定状态。为了实现这一点，大脑可能使用一种概率表示形式，尽管存在持续的竞争，但每个分布都有一个单一峰值（意味着每个整体感觉状态都有一个最佳解释）。我们的大脑似乎只考虑单峰（单个峰值）后验信念的一个根本原因可能是——归根结底——这些信念的作用是指导行动和行为，而我们一次只能做一件事。使用这种表示形式相当于部署一个隐式的形式化超先验[15]（形式化是因为它涉及概率表示本身的形式），表明我们的不确定性可以用这样的单峰概率分布来描述。考虑到上述关于行动的基本事实

（例如，我们一次只能执行一个行动，选择抓笔写字或扔笔，但不能同时做两件事），这样的先验具有适应性意义。

6.7 证据边界和对推理的模糊诉求

预测误差最小化发生在Hohwy (2013, 2014)所描述的”证据边界”(evidentiary boundary)内。他论证说，我们对世界的主体性访问受到预测误差最小化例程的限制，该例程应用于内感受、外感受和本体感受信号流。这些考虑导致Hohwy将PP描述为对认知心智施加了一个坚固的以神经为中心的边界。因此我们读到：

预测误差最小化应该让我们抵制[心智-世界]关系的概念，在这种概念中，心智在某种基本方式上对世界是多孔的，或者被视为具身的、延展的或能动的。相反，心智似乎与世界隔离，它看起来更像是以神经为中心的颅骨束缚，而不是具身的或延展的，行动本身更像是对感觉输入的推理过程，而不是与身体和环境的能动耦合。(Hohwy 2014, p.1)

Hohwy (2013, pp. 219 – 221, 2014)提供了各种相互关联的考虑，旨在支持这种隔离的、以神经为中心的心智观。所有这些考虑的核心是这样的观察：预测误差最小化例程是在感觉信号上定义的，因此”从颅骨内部，大脑必须推断其感觉输入的隐藏原因”(Hohwy 2013, p.220)。指导主题因此是推理隔离——论证认为，心智是在转换感觉信息面纱后运作的，推断复杂的”隐藏原因”作为对变化的（和部分自我诱导的）感觉刺激模式的最佳解释。相比之下，Hohwy建议：

强调开放性、具身性和主动延展到环境中的心智和认知观点似乎对这种推理概念的心智有偏见。(Hohwy 2014, p.5)

但是具身观点当然不是对这个（肯定无可争辩的）主张有偏见，即当主体以灵活、适应性、智能反应的方式与世界接触时，大脑正在做一些重要的事情。那么所谓的紧张关系可能在哪里呢？它主要在于Hohwy反复强调的观念，即大脑所做的最好被理解为一种推理形式。但在这里我们需要非常小心。因为这里起作用的推理概念实际上远没有最初看起来那么苛求。

为了看到这一点，考虑早期关于视觉和具身心智的辩论中的争议。根据1990年代中期发表的一篇典型综述论文：

关键洞察…是视觉的任务不是构建周围3D现实的丰富内在模型，而是在真实世界、实时行动的服务中有效且经济地使用视觉信息。(Clark 1999, p. 345)

一个替代方案——主要由”生态心理学”(ecological psychology)研究项目追求——是使用感知作为一个通道，允许我们锁定感觉流中的简单不变量[16]。以这种方式使用，感知提供了对世界的基于行动的*握持*，而不是适合分离推理的行动中性重构。这种握持可能内在地涉及有机体行动，例如——以McBeath and Shaffer (1995)的著名例子为例——棒球外野手奔跑以保持球的图像在视网膜上静止。通过以持续抵消任何明显光学加速度的方式行动，外野手确保（见Fink, Foo, and Warren 2009）她将在球下降到球场的地方接到球。在这种情况下，行为成功不是在外部世界的某种内在复制品上定义的推理结果。相反，它是通过保持感觉刺激在某些界限内运作的感知/行动循环的结果。

这些情况将在第8章中进一步讨论。就目前而言，重要的是这些策略类型是*根本性非重构的*。它们不使用感知在每时每刻构建一个内在模型来重现真实世界的结构和丰富性，因此无法代替那个世界来进行规划、推理和行动指导。相反，此时此地的行为是通过以上述特殊方式使用感知来实现的——作为一个通道，使有机体能够将其行为与远端环境的特定方面协调一致。

这种感知的非重构角色经常与推理性的、隔离的视觉形成鲜明对比。因此Anderson (2014)将非重构方法描述为主流（推理性、重构性）方法的替代，在主流方法中，感知被类比为科学推理，其中：

从不完整和碎片化的数据中，人们生成关于世界真实本质的假设（或模型），然后根据进一步传入的感官刺激对这些假设进行测试和修改。(Anderson 2014, p. 164)

Anderson继续说道，这种传统方法将认知描述为”后感知的……表征丰富的，并与环境深度解耦”。

感知作用的非重构解释提出了一种可行的替代方案，Anderson认为这显著改变了我们对自身认识论状况的理解。非重构解决方案展示了如何通过在感知和行动之间维持精妙的舞蹈来实现行为目标，而不是基于在感知的封闭门后构建的丰富内在模型来参与世界。这种基于把握的非重构舞蹈的一个特征是，它暗示了我们对感知和行动关系的普通思维方式的有力逆转。与其将感知视为对行动的控制，不如将行动视为对感知的控制(Powers 1973, Powers et al., 2011)。因此，问题变成了”不是……根据给定刺激选择正确反应，而是……根据给定目标选择正确刺激”(Anderson, 2014, p. 182 – 3)。

但是，正如Hohwy自己正确指出的，PP视觉中绝对没有任何东西与这种以收获感知为作用的行动视觉相冲突，或者（更一般地）与作为促进行为成功的一种手段的非重构策略理念相冲突。事实上，这种策略很自然地被容纳，因为最小化长期预测误差的最佳方式往往既节俭又涉及行动。因此我们读到：

仅仅因为大脑只进行推理，就认为它必须像遵循严肃的物理学教科书那样构建其内部模型，这是错误的。只要平均和长期最小化预测误差，由哪个模型来做这件事并不重要。因此，预测线性光学轨迹的模型是完全可行的，并且很容易优于更繁琐的一系列计算。如果它是一个参数更少的不太复杂的模型，这尤其如此，因为长期预测误差通过最小复杂性得到帮助。(Hohwy, 2014, p. 20)

这很有启发性。Hohwy在这里（以及其他地方[17]）认识到，PP框架往往会与更多”智识主义”的故事相对立，后者将每时每刻的行为成功描述为在丰富内部模型上定义的推理的产物，这些模型的作用是让认知者能够”抛弃世界”。相反，内部模型的作用在许多情况下是识别一些更节俭的、涉及行动的程序将起作用的上下文（关于这种混合图景的更多内容，参见第8章）。这意味着”推理”在PP故事中的功能并不被迫提供承载丰富重构内容的内部状态。它不是为了构建一个能够代表外部世界全部丰富性的内在领域。相反，它可能提供高效、低成本的策略，这些策略的展开和成功精细而持续地依赖于外部领域本身的结构和持续贡献，这些通过各种形式的行动和干预得以利用。

相关地，Hohwy经常谈到PP风格的系统寻找最好解释感官信息的*假设*。从某种角度听，这又是正确的。最小化预测误差系统必须找到最好地*容纳*（我现在这样说）当前感官冲击的多层次神经元状态集合。但我认为，这远比谈论”找到正确假设”要好，因为这种说法再次引入了不必要的和潜在误导性的包袱。容纳当前感官冲击可能采取多种形式，其中一些涉及选择重塑感官信号或维持其在预设界限内的低成本行动方法。因此，容纳传入信号不必暗示任何类似于确定对外部情况的描述，或找到最好描述或预测那些传入信号的命题或命题集合。实际上，在最基本的层面上，PP系统的任务不是检索恰当的描述。基本任务是使用预测误差作为杠杆，找到通过参与世界的行动最成功地容纳当前感官状态的神经活动模式。

6.8 不要害怕恶魔

为什么霍维(Hohwy)尽管经常强调对预测处理(PP)进行”非知识主义”解读的重要性，却坚持认为它促进了一种以神经为中心的、隔离的心智视角？原因似乎在于，他将这种隔离的、推理的视角与实践性具身认知科学讨论中相当不同且（我认为）相当陌生的东西联系起来。他将其与邪恶恶魔式全球怀疑主义的纯粹可能性联系起来——即我们可能被愚弄相信自己是在真实世界中行动的具身主体，而”实际上”我们只是被喂养了维持这种幻觉所需的感官信号序列的大脑。正是这种纯粹的可能性，在霍维的处理中，足以建立一个强健的”转导面纱”，将我们所知的世界置于一个重要的、主体无法穿透的证据边界的远侧。

因此，针对预测处理与上述描述的使用感知的快速而节俭策略的使用相一致（并且确实积极预测）的建议，霍维写道：

传入的视觉信号驱动行动，但…这种驱动实际上确实依赖于转导面纱，即证据边界，在其内部有充分的推理，而在其之外只有推理的原因。(2014, p. 21)

为了证明这一点，霍维援引了笛卡尔式（邪恶恶魔式）怀疑主义的幽灵。但在我看来，这似乎是一个转移注意力的错误论证。怀疑论断言仅仅是这样的断言：如果大脑接收和（显然）收获的感官刺激的作用保持固定，我们对世界的体验也会如此。那么，据我们所知，我们的物理身体可能悬挂在某种类似黑客帝国的能量网中，保持活力并被喂养任何使我们看起来像是在跑去接飞球和争论邪恶恶魔力量所需的感官刺激。但这种纯粹的可能性（即使被接受）决不会对与具身认知科学工作相关的关键断言产生怀疑。考虑跑去接那个飞球。这（在黑客帝国/缸中）将涉及向大脑提供复杂的、行动敏感的展开感官流，这些感官流通常会在具身主体实际跑动以抵消球的光学加速时产生。仅仅这是所需要的事实支持了这里真正重要的东西，即飞球拦截的非重构性说明。

因此，怀疑论挑战所暗示的是与感知和行动辩论中重构性和非重构性方法之间争议问题中的”推理隔离”非常不同的意义。因为那些辩论（关于感知-行动联结形状的辩论）不是关于我们是否可能被某种巧妙的操纵愚弄，误解我们自己的世界情况。相反，它们是关于如何从我们当前的科学视角最好地理解感官流在使适当的世界参与行动成为可能方面的作用。争议的问题是适当的行动是否总是且到处都通过使用感知来获得足够的信息进入系统来计算，以允许它通过探索远端世界的丰富的、内部表征的概述来规划其响应。非重构性方法（更多内容见第8章）证明了替代的、计算上更节俭的、行为交互的解决方案的可行性。它们并不意味着——它们也不寻求意味着——怀疑论假设的虚假性。这是一个正交问题，需要完整的哲学处理本身。

因此，心智作为隔离在推理帷幕后的形象是重要的模糊的。如果它仅仅意味着世界，就我们所知和体验的而言，是通过（部分自我诱导的）感官刺激的持续流动在体验上被指定和积极参与的，那么预测处理确实要求某种隔离。但隔离，在这种相当有限的意义上，并不意味着感知的丰富重构性模型，根据该模型，我们的行动是通过在丰富内部模型内容上定义的推理过程选择的，这些内部模型的作用是用一种内部模拟替代外部世界。

因此，神经处理围绕预测误差最小化例程组织的纯粹事实对位于最近具身心智工作核心的断言没有施加真正的压力。因为那项工作最根本拒绝的是感知的丰富重构性模型。冲突的表象源于推理和隔离概念本身的模糊性。因为这些概念可能似乎意味着远端环境的丰富内部概述的存在，伴随着行动作用的相应降级和在该内部模型上定义的推理作用的升级。然而，预测处理中的任何东西都不要求这一点。相反，预测处理强烈暗示像我们这样的大脑将尽可能利用严重依赖世界参与行动的简单策略，及时提供新的感官刺激以支持行为成功。这种策略是第8章的焦点。

6.9 你好世界

PP模式不仅没有在智能体和世界之间设置任何令人担忧的障碍[20]，它还提供了必要的手段，使结构化的世界首先进入视野。因此，考虑语音处理过程中对句子结构的感知。在这里（参见，例如，Poeppel & Monahan, 2011），我们也可能依赖存储的知识来指导对当前声音流的形状和内容的一系列猜测：这些猜测不断与传入的信号进行比较，允许残余误差在竞争的猜测之间做出决定，并（在必要时）拒绝一组猜测并用另一组替换它。这种对现有知识的广泛使用（驱动猜测）如我们所见，有许多优势。它使我们能够在嘈杂的环境中听到所说的话，在每个与单纯声音流一致的替代可能性之间进行裁决，等等。很可能只有通过部署丰富的概率生成模型，听者才能从撞击的声音流中恢复语义和句法成分。这是否意味着感知的句子结构是”对输入面纱背后内容的推断幻想”？当然不是。在恢复正确的相互作用远端原因集合（主语、宾语、意义、动词从句等）时，我们透过粗暴的声音流看到语言环境本身的多层结构和复杂目的。

不过，我们必须小心行事。当我们（作为母语使用者）遇到这样的声音流时，我们听到一系列单词，用间隙分隔。然而，声音流本身是完全连续的，正如声谱图非常戏剧性地揭示的那样。这些间隙是由听者添加的。我们在感知中遇到的在这个意义上是一个构造。但它是一个跟踪信号源中真实结构的构造（其他智能体产生不同有意义单词的字符串）。这里的预测大脑让我们透过嘈杂的感官信号看到引起感官刺激波的世界中与人类相关的方面。这可能是PP模型上感知通常所做的一个相当好的图景。如果是这样，那么我们在感知中遇到的世界不比我们在用母语说出的句子中听到的单词结构更多（也不更少）是虚拟现实或幻想。

预测处理在这里让我们透过感官信号看到远端世界中与人类相关的方面。从这个角度来看，预测处理故事与所谓的”直接”（例如，Gibson, 1979）感知观点有很多共同之处（至少在我看来如此）。因为它提供了一种真正的形式——也许是自然可能的唯一真正形式——的”对世界的开放性”。然而，必须承认的是，对自上而下级联的广泛依赖有时使真实感知在很大程度上依赖于先验知识。

我不会试图在这里进一步裁决这个微妙的问题（参见Crane, 2005）。但如果需要一个标签，有人建议隐含的形而上学观点最安全地被称为”非间接感知”[21]。这种类型的感知是”非间接的”，因为我们感知的*内容*本身不是假设（或模型、或幻想、或虚拟现实）。相反，我们感知的（当一切进展顺利时）是结构化的外部世界本身。但这不是”如其所是”的世界，这里暗示的是一个有问题的概念（另见9.10），即独立于人类关注和人类行动库表征的世界。相反，它是根据我们特定于有机体的需求和行动库解析的世界。因此揭示的世界可能充满了诸如隐藏但美味的猎物、扑克牌、手写数字和结构化、有意义的句子等项目。

在这里，感知对象也没有任何被当作类似”感觉材料”(Moore, 1913/1922)的意义，这些被构想为介入感知者和世界*之间*的代理。所涉及的内部表征在我们*内部*起作用，而不是*被*我们遇到。它们使我们能够在噪声、不确定性和模糊性的生态常见条件下遇到与有机体相关的世界。所有这些表明，我们在感知中遇到我们的世界，因为像我们这样的大脑是统计引擎，能够锁定非线性相互作用的原因，这些原因的特征有时可能深深埋藏在感官噪声和能量流动中。结果是，远端领域的智能体相关结构反映在神经架构本身的大规模形状和（见第9章）自发活动模式中[22]。然而，因此揭示的不是某种行动中性的远端领域。相反，它是从特定人类（和个体，见Harmelech & Malach, 2013）形式的行动和干预诱发的感官冲击统计中提炼出的世界。

6.10 幻觉作为不受控制的感知

在这些理论描述中，内容固定在本质上是（认知上的）外在主义的。感知状态的功能是估计与生物体相关的远端环境的属性和特征（为此目的，包括我们自己身体的状态和其他主体的心理状态）。但是这些状态是通过参考实际采样的世界来区分的。为了理解这一点，考虑一个案例（Hinton, 2005），一个训练好的神经网络的高级内部状态被”夹住”，即被实验者强制置于某种特定配置。然后活动在生成级联中向下流动，导致一种（如果你愿意这样说的话）实验者诱导的幻觉状态。但是这种状态的内容是什么？Hinton论证说，所表示的内容最好通过询问*世界*必须如何才能使这样的级联构成真实感知来把握。因此，如此描述的感知状态只不过是”一个假设世界的状态，在这个世界中，高级内部表征将构成真实感知”（Hinton, 2005，第1765页）。

这些考虑对感知作为”受控幻觉”的概念提出了一个转折。因为我认为，更好的做法是将幻觉描述为一种”不受控制的（因此是模拟的）感知”。在幻觉中，感知的所有机制都被调动起来，但要么完全没有感觉预测误差的指导，要么（见2.12和第7章）预测误差回路出现故障。在这种情况下，主体确实进入了Smith（2002，第224页）所称的”模拟感觉意识”状态。

最后，请注意，由主动的、可供性敏感的预测所传递的感知内容，现在本质上是有组织的和外向的。我的意思是，它揭示了——并且不能不揭示——一个结构化的（因此在某种弱意义上是”概念化的”[23]）外部世界。这是一个外部舞台，由远端的、因果相互作用的项目和力量填充，它们的联合作用最好地解释了（给定先验知识）当前的感觉刺

激套件。这正是智能主体必须拥有的那种对世界的把握，如果她要恰当地行动的话。当这样的主体看到世界时，他们看到的是远端相互作用原因的确定结构，适合于像他们这样的生物的行动和干预。感知体验的所谓”透明性”[24]——即在正常的日常感知中，我们似乎看到桌子、椅子和香蕉，而不是我们感觉表面的近端兴奋，如视网膜上的光线变化——很自然地从这种模型中产生。我们似乎看到狗、猫、目标、抢断和获胜的扑克牌，因为这些特征位于对人类选择和行动重要的相互作用的、嵌套的远端原因结构中。

[6.11 最优幻觉]

当然，事情有时会（并且确实会）出错。正如Paton等人（2013，第222页）雄辩地论证的那样，人类心智”总是岌岌可危地被摆脱预测误差的冲动所挟持[并且这]当世界不以其通常的、统一的方式合作时，强加给我们非常不可能和奇幻的感知”。例如，令人惊讶的是，很容易诱导（即使在完全清醒的正常成年人中）这样的错觉：放在你面前桌子上的橡胶手是你自己的。这种错觉是通过确保受试者能够看到有人在敲击逼真的橡胶手，同时（在视线之外）他们自己的手被完全同步地敲击而产生的（Botvinick & Cohen, 1998）。Ramachandran和Blakeslee（1998）描述了一个类似的错觉，其中蒙眼受试者的手臂被伸展，他们的手指被让去敲击坐在他们正前方的另一个受试者的鼻子，同时实验者使用间歇节奏完全同步地敲击他们自己的鼻子。在这里，预测性的贝叶斯大脑也可能被愚弄而产生错误感知——在这种情况下，是你有一个两英尺长的鼻子！有许多方式可能导致这种错误，涉及先验期望和驱动感觉信号之间的不同平衡（一些很好的讨论，见Hohwy, 2013，第1章和第7章）。但就目前的目的而言，重要的是时间同步性的事实起到了关键作用（正如实验操作清楚显示的那样）。当感觉信号开始揭示这种意外的（生态上罕见的，因此通常高度信息丰富的）同步性时，为了”解释掉”预测误差，产生了奇怪和不可信的感知。那么，这告诉我们关于我们与世界的普通的、日常的感知接触什么呢?

在某种意义上，这似乎表明（正如Paton等人所论证的），大脑用来跟踪和接触外部世界的常规机制存在某种脆弱性。这些常规机制确实可能被劫持和胁迫，从而产生误导。然而，一个值得思考的问题是：“对于某种特定的感知错误，避免这种错误的代价是什么？”因为在许多情况下，这种代价可能是我们感知生活（或更广泛地说，我们心理生活）中其他方面的大量错误。正如我们在第1章中提到的，Weiss等人（2002）使用最优贝叶斯估计器(optimal Bayesian estimator)证明，多种运动”错觉”直接由人类运动感知实现最优估计器的假设所推导出来。他们得出结论：“许多运动’错觉’并不是视觉系统各个组件草率计算的结果，而是在合理假设下最优的连贯计算策略的结果”（Weiss等人，2002，第603页）。这表明，至少在某些情况下，即使是”错觉性”的感知体验也构成了对最可能的真实世界来源或属性的准确估计，给定噪声感官证据和相关样本内真实世界原因的统计分布。因此，一些局部异常可能是我们为全局优化性能付出的代价（Lupyan，出版中）。

这是一个重要发现，现在已在许多领域得到重复验证，包括声音诱导的闪光错觉（Shams等人，2005）、腹语术效应（Alais & Burr，2004）以及图形-背景凸性线索在深度感知中的影响（Burge等人，2010）。此外，Weiss等人（2002）对一类静态（注视依赖性）运动错觉的贝叶斯最优解释现在已扩展到解释在平滑追踪期间主动眼球运动存在下产生的更广泛的运动错觉集合（见Freeman等人，2010，以及Ernst，2010中的讨论）。因此，即使在这些错觉情况下，感知体验看起来也在真实地跟踪感官数据与其最可能的真实世界来源之间的统计关系。这再次表明，中介机制没有在心智和世界之间引入令人担忧的障碍。偶尔偏离轨道只是我们大多数时候把事情做对所付出的代价。

或者考虑最后一个更具争议性的案例——“大小-重量错觉”。这被引用（Buckingham & Goodale，2013）作为对人类心理物理学中最优线索整合假定普遍性的挑战。在大小-重量错觉中，外观相似的物体似乎经过重量调整，使我们判断较小的物体比较大的物体感觉更重，尽管它们的客观重量相同（一磅铅确实比一磅羽毛感觉更重）。Buckingham和Goodale调查了关于大小-重量错觉的最新研究，注意到虽然贝叶斯处理确实能够掌握举重行为本身，但它们未能解释主观比较效应，有些人将其描述为”反贝叶斯的”，因为先验期望和感官信息在那里似乎是对比的而不是整合的（Brayanov & Smith，2010）。他们认为，这为更加分裂和防火墙化的认知经济提供了证据：一个显示”用于运动控制和感知/认知判断的独立先验集合，它们最终服务于完全不同的功能”的经济（第209页）。

然而，有一个有趣的（尽管仍然高度推测性的）替代方案。Zhu和Bingham（2011）表明，相对重量的感知与最大距离可投掷性的功能可供性(affordance)精细地同步。那么，也许我们简单地标记为”重量”体验的东西，在某种更深层的生态意义上，是最优重量-大小比以提供远距离可投掷性的体验？如果这是真的，那么Buckingham和Goodale描述的体验重新出现为可投掷性的最优感知，尽管我们通常将其误解为简单但错误的相对物体重量感知。从一个角度看起来是分裂的、脆弱的和断开的认知经济，因此在更深入的考察下，可能证明是一个强健的、良好整合的（尽管绝非同质的）机制，适应于*不是*提供简单的行动中性的世界描述，而是让我们接触环境中的行动相关结构。

6.12 更安全的渗透

这些考虑还有助于揭示为什么低级处理过程被高级预测和期望的大量”渗透”并不会对我们的认识论情况构成深刻威胁。这里的担忧（见[Fodor, 1983]、[1988]）是，我们（认为我们）感知到的东西可能——由于所有这些自上而下的影响——太容易被我们期望感知到的东西所感染，这将破坏科学调查本身的基础。我们希望我们的观察能够检验我们的理论和期望，而不是简单地与它们保持一致！福多尔论证说，对我们来说幸运的是，感知并不是这样可渗透的，这一点可以从视觉错觉的持续性得到证明（福多尔声称），即使在我们了解了它们的错觉性质之后也是如此。例如，经典版本的缪勒-莱尔错觉中的相等线条看起来仍然长度不等，即使我们已经亲自测量过它们。福多尔将此作为证据，证明感知总体上是”认知不可渗透的”，也就是说，不会直接受到任何类型高级知识的影响（[Pylyshyn, 1999]）。

正确的诊断，我们现在可以看到，实际上相当不同。我们应该说的是，感知可以被自上而下的影响渗透，当且仅当这种渗透在足够广泛的训练实例范围内赢得了其存在的价值。许多错觉尽管有相反的语言形式知识仍然持续存在的深层原因是，感知系统的任务是最小化Lupyan（出版中）有用地描述为”全局预测误差”的东西。相对于感知系统需要处理的全套情况，线条长度不等的假设是最佳假设（[Howe & Purves, 2005]）。从那个更全局的角度来看，我们对错觉的易感性实际上根本不是认知失败。因为如果系统要推翻传递这个判决的许多精细交织的中级处理层，结果将是在许多其他（更生态正常的）情况下真实感知的失败。

这里对我们的认识论情况没有威胁。总的来说，我们的感知系统作为调解感官刺激和行动之间的设备是校准良好的，它们的输出（虽然可能因广泛的重新训练而改变）不会简单地被我们对诸如”是的，这两条线确实长度相等”之类句子的认可所推翻。认可这样一个句子（见[Hohwy, 2013]）并不能充分解释构成那个特定感知的低级预测和预测误差信号的全谱，所以它无法推翻系统的长期学习。简单地接触句子最可能对感知体验产生影响的情况是感官证据模糊的情况。在这种情况下（见9.8），听到一个句子可能会使系统倾向于对场景的一种解释——这种解释真正影响场景如何呈现给代理（关于这种类型的例子，见[Siegel, 2012]，以及2.9中的讨论）。

总之，各种类型的自上而下影响可能在每个低级水平上影响处理，但只有当这些影响模式在全局上（而不仅仅是局部上）是有效的时候。结论是：

感知系统在这种渗透最小化全局预测误差的程度上是可渗透的。如果允许来自另一个模态、先前经验、期望、知识、信念等的信息降低全局预测误差，那么这些信息将被用来指导低级水平的处理。例如，如果听到声音可以消除原本模糊的视觉输入的歧义…那么我们应该期望声音影响视觉。如果知道特定的线条集合可以被解释为有意义的符号能够改善视觉处理，那么这种知识将被应用于低级视觉处理。（Lupyan，出版中，第8页）

这对科学来说是好消息。它使我们能够对可能否定我们自己理论的感官证据保持开放，同时也使我们能够成为专家感知者，能够在令人生畏的噪音和模糊背景下发现表明希格斯玻色子作用的微弱痕迹。

6.13 谁来评估评估者？

最后，严重形式的精神紊乱，如精神分裂症和各种其他形式精神病特征的妄想和幻觉怎么办？在这些情况下，通常用于平衡感官输入、自上而下期望和神经可塑性的精细平衡机制已经严重失常。如果2.12中探索的假设是正确的，系统性故障（可能根植于异常的多巴胺能信号传导）在这里破坏了预测误差信号本身的产生和权重。这是一种特别具有挑战性的破坏形式，因为（正如我们所看到的）持续的、高权重的预测误差将看起来像是在发出显著外部结构、威胁和机会的信号。如果不解决，它将因此驱动系统改变和适应生成模型，启动一个恶性循环，其中错误感知和错误信念共同出现，相互提供虚假支持。

更糟糕的是，正如Hohwy 2013第47页正确指出的那样，系统本身没有简单的方法来评估其自身精度分配的可靠性。因为对预测误差的精度加权已经反映了系统对每个处理层级信号可靠性的系统性估计。显然，任何系统都无法承担无尽的”计算自我怀疑”螺旋，即试图估计其对自身信心分配的信心、对自身可靠性评估的可靠性，等等。此外，考虑到现在争议的是证据和对该证据信心度量的可靠性，系统需要使用什么类型的证据来计算这些元元度量尚不清楚。[28]

我们可以得出结论，精度问题对任何形式的理性自我纠正都具有异常的抗性。当然，这正是在许多形式的精神病中发现的（否则相当令人困惑的）特征。精度估计的破坏使与世界的概率预测接触既不可靠又极难纠正。这种复杂的干扰是下一章的主题。

6.14 引人入胜的故事

如果概率预测机器视觉是正确的，那么感知是一个涉及对我们自身不断演化的神经状态进行滚动（次个人）预测的主动过程。这种彻底内向的自我预测过程最初可能看起来不太有希望作为感知如何接触世界的模型。然而，如果本章探讨的论点是正确的，那么主动适应我们自身变化感觉状态的压力提供了我们对结构化、有机体相关的外部世界的把握。

在主动动物中，这种把握并不根植于对客观外部现实的某种动作中性图像。相反，最小化预测误差就是最小化识别世界呈现的动作可供性(affordances)的失败。在这里，一个好的策略是在每个时刻提供对多个竞争可供性的部分把握：一种”可供性竞争”，可能只有在动作需要时才合理地得到解决。随着这个过程的展开，决策和动作准备过程不断交织，多个反应以其表达变化概率分级的方式准备。这一切都表明，我们对世界的感知把握从根本上是基于交互的：它是在世界参与动作的存在中锻造的，并致力于为世界参与动作服务的把握。

这种把握并不完美。它让我们容易受到幻觉、错误，甚至是精神分裂症和其他形式精神病特有的全面破坏。这是否意味着即使是正常运作的系统也只能提供与”虚拟现实”的接触？最终，我们可能不应该过分担心我们在这里使用的词汇。但深度和持久断开的含义是误导性的。运作良好的感知系统并不产生某种粗暴地插在心灵和世界之间的虚拟现实，而是驱散表面统计和部分信息的迷雾。所揭示的是一个由人类需求和可能性塑造的显著、有意义模式的世界。

[7]

[期待自我（悄然接近意识）][]

7.1 人类体验的空间

我们已经涵盖了一个广阔而多样的领域。我们的故事从通过试图预测我们自身变化的感觉状态来了解世界这一巧妙技巧开始。我们继续探索在多层级设置中使用该技巧来指导感知、想象、动作以及关于世界和其他主体的基于模拟的推理。我们看到了如何对与不同神经群体活动相关的相对不确定性的持续估计可以进一步转化这种故事的力量和范围，使处理流程动态可重配，并在真正宏大的规模上提供上下文敏感性。我们已经开始了理解主动的、具身的主体如何通过创建和维持反映有机体需求和环境机会的感知-动作循环来利用这些资源的关键和持续任务。通过这样的增强，我相信我们的故事具有阐明人类思想、体验和动作全谱所需的资源。

要兑现这样的声明——或者甚至使这样的声明真正可理解——我们现在需要将这个相当理论化的、大规模的图景与人类体验的形状和性质更紧密地联系起来。如果你愿意的话，我们需要开始在持续的、多层级预测的旋涡中认识*我们自己*。在那时，我们图景中许多更实际的——以及人类意义重大的——方面开始显现，揭示了人类心灵复杂空间的某些东西。在那个空间内，一些关键原则和平衡（涉及预测误差及其在动作展开中的微妙作用）可能决定正常和非典型形式的人类体验的形状和性质。

当然，我无法完全兑现这一点。不幸但不令人意外的是，对人类体验全貌及其机制性（和社会文化）根源的令人信服的解释显著超出了我们的能力范围。但新兴的文献提供了一些有希望的线索、几个简化的模型，以及少量有趣的（但投机性的）提议。那么，预测处理(predictive processing)能够告诉我们关于意识、情感和人类体验多样性的什么呢？

7.2 警告灯

随着本章的展开，焦点落在了人类体验变得结构化、被干扰或以微妙方式受到影响的广泛案例上，这些案例可以通过预测处理(predictive processing)的独特机制来阐明（我将如此建议）。在每个案例中，预测处理机制的一个方面发挥着核心作用。再次强调，这个方面是特定预测误差信号的精度(precision)，因此是对不同证据体可靠性的估计：这些证据包括外感受感觉信号、内感受和本体感觉信号，以及先验信念的整个多层次谱系。正如我们在前面章节中反复看到的，这种可靠性（等价于不确定性）的估计为这些解释提供了关键的附加维度，使得特定预测误差信号的影响能够根据任务、上下文和背景信息而改变。更一般地说，这些精度估计构成了一种根本性的元认知策略。这种估计是元认知的(metacognitive)，因为它们涉及对我们自己心理状态和过程的确定性或可靠性的估计（大多数是非意识和亚个人的）。但这是一种元认知策略，可以说是感知和行动基本机制的基本组成部分，而不是仅在高级”高层次”推理中才出现的东西。

估计我们自己预测误差信号的可靠性（或其他方面）显然是一个微妙而棘手的事情。因为正如我们经常注意到的，预测误差信号会”传达消息”。然而，在这里，大脑的任务是确定预测误差信号本身的正确权重。估计某个（假定的）消息项目的可靠性从来都不容易，任何遇到过不同媒体对同一事件的截然不同报道的人都知道这一点！面对这种熟悉的困难时，一个常见的策略是优先考虑某个特定的新闻来源，比如你最喜欢的报纸、频道或博客。但是假

设，在你不知情的情况下，该来源的所有权在一夜之间易手了。你预先倾向于非常认真对待的信息流现在（让我们想象）严重误导。来自你选择的可靠来源的信息现在严重受污染，而且是以你从未预料到的方式。你可能现在选择探索许多其他不太可能的选项（“火星人真的降落了：白宫新闻办公室这样说”），你本来会忽略或立即拒绝的选项，而不是相信这一点。正如我们即将看到的，这就是一类新兴的非典型心理状态解释的独特形状的大致轮廓。这些解释将关键失败定位为精度估计失败，因此作为其任务是估计我们自己信息来源可靠性的机制的失败。正如我们将看到的，这种失败可能会根据复杂的先验和感觉证据经济的哪些方面受到最大影响而产生非常复杂和不同的效果。

为了了解总体风格，考虑Adams等人（2013）描述的”警告灯”场景。我完整引用这个案例，因为它巧妙地捕捉了几个对我们后续讨论来说重要的因素：

想象一下你汽车的温度警告灯太敏感（精确），报告超过某个温度的最轻微波动（预测误差）。你自然推断你的汽车有问题，并将其送到车库。然而，他们没有发现任何故障——但警告灯继续闪烁。你的第一直觉可能是怀疑车库未能识别故障——甚至开始质疑推荐它的《好车库指南》。从你的角度来看，这些都是符合你可获得证据的合理假设。然而，从从未见过你的警告灯的人的角度来看，你的怀疑会有一种非理性和略带偏执的味道。这个轶事说明了妄想系统如何可能因为赋予感觉证据过多精度而被详细阐述。（Adams等人，2013，第2页）

Adams等人随后补充说：

这里的主要病理学本质上是元认知的：从某种意义上说，它基于一个关于信念的信念（警告灯报告精确信息）（发动机过热）。关键是，在形成预测或预测误差方面没有必要的损伤——问题在于它们被用来告知推理或假设的方式。

在继续讨论一些实际案例之前，对以上所有内容做两个简要评论。首先，在这里层层叠叠地增加复杂性并不会带来明显的好处。假设我们为汽车装配一个进一步的设备：警告灯故障警告灯！我们所做的只是将问题进一步推后。如果两个灯都闪烁，我们现在必须确定哪一个传达的是最可靠的信息。如果只有一个灯闪烁，据我们所知，它传达的信息可能仍然可靠，也可能不可靠。在某个点上（虽然对于所有任务和所有时间来说，未必是同一个点），信任的回归必须停止。而无论在哪里停止，它都可能形成虚假或误导性推理的自我强化螺旋的起点。其次，请注意精确度加权提供的本质上是一种塑造推理和行动模式的手段，因此它在增加（比如说）先验信念的精确度或降低感官证据的精确度这一直观差异上奇怪地保持中性。重要的只是影响的相对平衡，无论这种平衡是如何实现的。因为正是这种相对平衡决定了主体反应。

7.3 推理和经验的螺旋

回忆一下关于精神分裂症中妄想和幻觉（所谓的”阳性症状”）出现的PP解释（Fletcher & Frith, 2009），这在2.12中有所描述。基本思想是这两种症状可能都源于一个单一的潜在原因：错误生成且高度加权（高精确度）的预测误差波。因此，关键的扰动是元认知的扰动——因为分配给这些误差信号的加权（精确度）使它们在功能上如此有效，使它们能够驱动系统进入可塑性和学习状态，形成并招募越来越奇异的假设，以适应持续不断的（表面上）可靠且显著但持续无法解释的信息波。由此产生的高层假设（如心灵感应和外星人控制）在外部观察者看来显得奇异且毫无根据，但从内部来看，现在构成了最好的，因为是唯一可用的解释——很像上面演示的警告灯例子中对《优秀车库指南》的怀疑。一旦这种高层故事站住脚，新的低层感官刺激可能会被错误地解释。当这些新的先验占主导地位时，我们可能因此经历看似证实或巩固它们的幻觉。这在根本上并不比先验期望使空心面具看起来坚实凸起（见1.17）或白噪音听起来像”白色圣诞节”（见2.2）更奇怪。在那时（Fletcher & Frith, 2009，第348页），错误的推理提供了错误的感知，这些感知为产生它们的理论提供了虚假的支持，整个循环变得恶性自我确认。

那么”明显的”高层解释呢，朋友或医生甚至可能向受影响的主体建议，即主体本身在认知上受到了损害？这确实应该构成一个可接受的高层解释，但严重受影响的受试者发现它不令人信服。在这种情况下，值得注意的是预测误差信号不是经验的对象（或经验的实现者）。因此，类比中的”红色警告灯”不是对预测误差信号的体验。PP的建议不是我们像这样体验我们自己的预测误差信号（或它们相关的精确度）。相反，这些信号在我们内部起作用，招募适当的预测流，揭示一个由远端对象和原因组成的世界。然而，持续未解决的预测误差信号可能产生”显著奇异性”的模糊感觉，在这种感觉中，受试者发现自己被（对其他人来说）似乎只是偶然巧合的事情强烈影响，等等。在分层设置中，这相当于（Frith & Friston, 2012）低层的持续扰动，其唯一的解决方案在于奇异的、抵抗反证的顶层理论化。Frith和Friston建议，这与第一人称报告相符，比如Chadwick（1993）的报告，他是一位受过训练的心理学家，患过一次偏执性精神分裂症发作。Chadwick回忆说，他”必须对所有这些不可思议的巧合给出某种意义，任何意义”，他”通过根本性地改变[他的]现实概念来做到这一点”。对此，Frith和Friston评论道：

用我们的术语来说，这些不可思议的巧合是由具有不适当高精确度或显著性的预测误差产生的错误假设。为了解释它们，Chadwick必须得出结论，其他人，包括广播和电视主持人，能够看透他的思想。这就是他必须在现实概念中做出的根本性改变。（Frith & Friston, 2012，第8节）

7.4 精神分裂症和平滑追踪眼动

这样的推测既有趣又合理。但PP解释的一个主要吸引力在于它还为各种其他、不那么戏剧性但同样具有诊断意义的特征提供了令人信服的解释。其中一个特征涉及精神分裂症受试者表现出的”平滑追踪眼动”中的一些异常。这项工作的背景是正常和精神分裂症受试者在平滑追踪暂时视觉遮挡目标期间的一种稳健的差异模式。

平滑追踪眼动[3]可以与扫视眼动形成对比。人眼能够在视觉场景中进行扫视，快速跳跃式地从一个目标移动到另一个目标。但当出现移动物体时，眼睛可以”锁定”该物体，平滑地追踪其在空间中的运动（除非它移动得太快，在这种情况下会启动所谓的”追赶扫视”）。平滑追踪眼动能够追踪缓慢移动的物体，将其图像保持在高分辨率的中央凹上。在平滑追踪中（见Levy et al., 2010），眼睛的移动速度小于每秒100度，并且（在这个范围内）眼睛的速度与目标的速度密切匹配。一个常见的例子，至今仍被用作体检中的神经学指标，就是在医生面前来回移动手指时，你用眼睛跟随医生移动的手指，而不移动头部或身体。（你可以自己进行同样的测试，将手臂伸直，追踪食指尖端，同时左右移动手部。如果在这种情况下你的眼睛出现抖动，你在”现场清醒测试”中会得分很低，可能会被怀疑受到酒精或氯胺酮的影响。）

平滑追踪眼动包括两个阶段：启动阶段和维持阶段（分别由开环和闭环反馈区分）。在维持阶段，称为”追踪增益”（或等效地称为”维持增益”）的量测量眼睛速度与目标速度的比率。这个比值越接近1.0，目标速度与眼睛速度之间的对应关系就越好。在这种条件下，目标的图像在中央凹上保持稳定。当两者出现偏差时，可能会发生追赶（或后退）扫视，使两者重新对齐。

精神分裂症患者明显表现出各种平滑追踪障碍（也称为”眼动追踪功能障碍”），特别是当追踪目标被遮挡或改变方向时。根据Levy et al. (2010)的权威综述，“眼动追踪功能障碍(ETD)是精神分裂症中重复性最高的行为缺陷之一，在精神分裂症患者临床未受影响的一级亲属中过度代表”（Levy et al., 2010，第311页）。

特别地，我们将重点关注（遵循Adams et al., 2012）三个能够可靠区分正常和精神分裂症患者表现的差异。它们是：

1. 视觉遮挡期间的追踪障碍。精神分裂症患者产生较慢的追踪，这在被追踪项目被遮挡（视野中被遮蔽）时尤其明显。因此，虽然神经典型受试者的追踪增益约为85%，但在精神分裂症人群中平均约为75%。更引人注目的是，当移动目标暂时被遮挡时，神经典型受试者能够以60-70%的增益进行追踪，而精神分裂症患者的追踪增益为45-55%（见Hong et al., 2008；Thaker et al., 1999，2003）。

2. 矛盾性改善。当目标意外改变方向时，精神分裂症患者短暂地超越神经典型受试者，在新轨迹的前30毫秒内产生更好的目标/眼睛速度匹配（见Hong et al., 2005）。

3. 重复学习障碍。当目标轨迹重复多次时，神经典型受试者达到最佳表现，而精神分裂症患者则不能（见Avila et al., 2006）。

这整个复杂的令人困惑的效应（矛盾性改善以及双重缺陷）同时出现，是分层预测和精度加权预测误差经济体系单一扰动的结果，我们接下来将会看到。

7.5 模拟平滑追踪

Adams et al. (2012)回顾了大量证据，表明平滑追踪眼动的预测成分对精神分裂症型扰动最敏感，因此比单纯的维持增益测量本身提供更大的洞察力（和诊断潜力）。例如（Nkam et al., 2010），对随机移动刺激的平滑追踪在精神分裂症型和神经典型受试者之间没有区别。一旦运动在某种程度上变得可预测，差异开始出现。但随着预测成分的增加，差异变得越来越明显。当移动物体暂时被遮挡时，预测成分很大，两个人群之间的差异（如我们之前所见）最大。

为了突出预测成分，引入了一个称为”平均预测增益”的测量（Thaker et al., 1999）。平均预测增益是遮挡期间的平均增益。遮挡也导致所有受试者的眼睛先减速，然后增速回到目标速度。为了考虑这一点，“残余预测增益”测量平均预测增益减去一般减速期。在一个大样本中，涵盖严重和不太严重的精神分裂症型受试者亚型，所有人都显示出残余预测增益减少，无症状的精神分裂症型亲属也是如此。相比之下，只有严重受影响的个体在总体上显示出维持增益减少。

这样的证据揭示了精神分裂症型平滑追踪最显著的影响模式与运动刺激可预测性之间的强烈联系。任务的预测要求越高，与神经典型反应和追踪模式的差异就越大。在这方面，关于平滑追踪的证据恰好融入了一个更大的结果和推测拼图中，涉及精神分裂症型对某些错觉的反应、著名的”自我挠痒”研究（第4章）以及控制妄想。我们将在后续章节中回到所有这些主题。

为了探索预测性处理系统的干扰对平滑追踪眼动的可能影响，Adams et al. (2012)部署了一个简化的分层生成模型，涉及感知和运动控制的链接方程。启发式地，[4]该模型”相信”其凝视和目标物体都被吸引到一个在单一（水平）维度上定义在外在坐标中的共同点。因此，“生成模型…基于这样的先验信念：凝视中心和目标被吸引到视觉空间中的一个共同（虚构）吸引子”(Adams et al., 2012, p. 8)。这种简单的启发式规则可以支持令人惊讶的复杂适应性反应形式。在当前情况下，在该启发式”信念”影响下运行的模拟智能体，即使在存在遮挡物体的情况下也显示出平滑追踪。追踪在遮挡物体干预下仍继续进行，因为网络现在表现得好像一个单一的隐藏原因同时吸引着眼睛和目标。然而，重要的是，生成模型还包括足够的分层结构，允许网络表示涉及周期性轨迹的目标运动（即目标周期性运动的频率）。最后（也是至关重要的），处理的每个方面和层级都涉及相关的精确度期望，编码模拟智能体对进化信号该元素的信心：要么是感觉输入本身，要么是关于感觉输入随时间演化的期望——在这种情况下，是关于目标周期性运动的期望。

这个模型虽然简化，但捕获了正常平滑追踪眼动的许多关键方面。在运动目标持续存在的情况下，眼睛在短暂延迟后平滑追踪。当目标第一次被遮挡时，系统在大约100毫秒后失去对隐藏运动的控制，并且当目标出现时必须产生追赶性跳视（这只是近似建模）。但当相同序列重复时，追踪显著改善。此时，网络的第二层可以预期运动的周期性动力学，并能够利用这种知识向下一层提供恰当的上下文固定信息，实际上使持续感觉刺激的缺乏（当目标被遮挡时）变得不那么令人惊讶。这些结果与来自人类受试者的数据提供了良好的定性拟合（例如，Barnes & Bennett, 2003, 2004）。

[7.6 扰乱网络（平滑追踪）]

回想一下精神分裂症受试者平滑追踪眼动的三个显著特征（7.4）。这些是视觉遮挡期间的追踪受损、轨迹意外变化时的悖论性改善，以及重复学习受损。Adams et al. (2012)论证，这些影响中的每一个都可以追溯到基于预测的内在经济中的单一潜在缺陷：实际上，这是同一个缺陷（见4.2），被援引来解释精神分裂症在力匹配任务上的表现，以及备受关注的自我挠痒能力改善（见，例如，Blakemore et al., 1999; Frith, 2005; Shergill et al., 2005；以及第4章中的讨论）。

因此假设精神分裂症在一个长而复杂的因果链的开始附近，实际上涉及先验期望相对于当前感觉证据的影响减弱（见Adams, Stephan, et al., 2013）。这可能让读者感到奇怪。我听到你说，肯定相反的情况必须是真的，因为这些受试者似乎允许奇异的高层信念压倒他们感官的证据！然而，因果性的箭头向另一个方向移动的可能性似乎越来越大。先验期望相对于感觉输入的影响减弱可能导致，正如我们稍后将看到的，异常的感觉体验，其中（例如）自我产生的行为对智能体来说似乎是外部引起的。这反过来可能导致形成越来越奇怪的高层理论和解释（见Adams, Stephan, et al., 2013）。

影响感官证据和高层信念之间关键平衡的一个重要因素是，智能体在自主运动期间减少（衰减）感官证据精确度的能力。这种能力在功能上至关重要（我们很快会探讨其原因）。这种能力的减弱（即感官衰减(sensory attenuation)能力降低）可以解释，正如在第4章中提到的，精神分裂症患者在自我挠痒和准确匹配体验到的力量与自我产生的力量方面表现出超出正常水平的能力。在探讨这些问题时，重要的是要记住，低层和高层状态精确度之间的*平衡*在功能上是重要的，因此增加低层感官预测误差的精确度（从而增加其影响）和降低与高层预测相关的误差精确度（从而减少其影响）——至少就推理过程而言——是等同的。

在平滑追踪眼动的情况下，降低处理层级中高层（具体来说，是Adams等人2012年简单模拟中的第二层）预测误差的精确度会导致前面描述的特定效应组合。为了证明这一点，Adams等人降低了图7.5中勾勒的模拟平滑追踪网络第二层预测误差的精确度。这样做的直接效果是减少了关于目标周期性运动的预测误差的影响。在较低速度下，当运动物体在视线内时，两个网络表现出相同的行为，因为”降低高层精确度”网络（简称RHLP网络）此时依赖感官输入来指导行为。但当物体被遮挡时，这种依赖变得不可能，RHLP网络相对于其”神经典型”同类网络表现受损。随着周期数增加，这种效应变得越来越明显。这是因为第二层精确度受损不仅导致相对于感官输入，关于运动的期望的直接影响减少，还导致学习能力受损——在这种情况下，无法从持续接触中学习周期性运动的频率（见Adams等人，2012年，第12页）。最后，RHLP网络还表现出微妙的”矛盾性改善”模式，当未遮挡的目标意外改变方向时，其表现优于神经典型网络。这整个效应模式很自然地源于高层精确度降低的存在，因为在这种条件下，当构建良好的预测改善表现时（例如，在遮挡物后面和较高速度时），网络表现会更差；当预测产生误导时（例如，当预期轨迹突然改变时），表现会更好；并且在从经验中学习方面会受损。

这里建模的这种干扰在生理学（Seamans & Yang, 2004）和药理学（Corlett等人，2010）上都是合理的。如果精确度确实是通过影响浅层锥体细胞误差报告增益的机制编码的，如果高层（视觉或眼部）误差报告细胞主要（如看起来可能的那样）存在于前额皮质的额叶眼区，那么文献中报告的多巴胺能、NMDA和GABA能受体异常提供了一个清晰的途径，通过这个途径，在前额皮质中作为突触增益实现的高层精确度可能会受损。这种异常会选择性地损害高层期望的获得和使用，减少与使用上下文信息来预测感官输入相关的益处和（在罕见条件下）成本。

7.7 重新审视挠痒

如此巧妙地解释平滑追踪眼动障碍的解释机制也对第4章中勾勒的精神分裂症患者增强自我挠痒能力的标准解释提出了重要修正。正如我们将看到的，这些修正最具启发性的方面是，它们更好地将感官效应与运动障碍和妄想信念的出现联系起来，从而使用单一机制解释了一个复杂的观察效应组合。

回想一下，精神分裂症患者在自我挠痒方面表现出比神经典型对照组更强的能力。也就是说，精神分裂症患者对自我产生的挠痒刺激的真实痒感评级高于神经典型对照组(Blakemore et al., 2000)。这种效应确实发生在感觉层面，而不仅仅是言语报告的某种异常，这一点通过精神分裂症患者在力量匹配任务(见4.2节)中的表现得到证实，在该任务中言语报告被尝试匹配参考力量所取代。如前所述，在这里，神经典型个体”过度匹配”参考力量，施加更大的自我产生的力量，这种方式在多主体场景中导致施加压力的持续升级。这种效应在精神分裂症患者中有所减少，他们能够更准确地执行任务(Shergill et al., 2005)。因此，这里也存在一种”矛盾性改善”，即精神分裂型知觉比神经典型个体的知觉更加准确。神经典型个体对许多形式的自我产生刺激表现出感觉衰减(sensory attenuation)，包括自我产生触觉的愉悦感和强度(评级为比通过其他方式提供的相同刺激更不愉快和强度更低)，甚至对自我产生的视觉和听觉刺激也是如此(Cardoso-Leite et al., 2010; Desantis et al., 2012)。因此，相当普遍地说，自我产生的感觉在神经典型情况下被衰减(减少)，而这种衰减在精神分裂症患者中*本身*也被减少了。

Brown et al. (2013)提供了一个可能的解释，同时解释了关于能动性(agency)特征性妄想信念的出现(另见Adams et al., 2013; Edwards et al., 2012)。更标准的解释(我们在[第4章]中遇到的那个)是，准确的前向模型通常允许我们预期自己施加的力量，结果这些力量似乎更弱(衰减)。如果这样的模型受到损害，我们自己行动的效果将(标准模型提示)显得更加令人惊讶，因此更可能被归因于外部原因，从而导致关于能动性和控制的妄想的出现。Brown等人注意到这个标准解释的三个重要缺陷：

1. 成功预测与知觉*强度*降低(例如，在力量匹配或挠痒任务中)之间的联系并不清楚。正如我们在[第1-3章]中看到的，信号中被很好预测的元素被”解释掉”，因此不会施加压力来选择新的或不同的假设。但这对当前获胜假设所提供的知觉体验的强度或其他方面没有任何说明。

2. 操纵自我产生感觉的可预测性似乎不会影响所体验到的感觉衰减程度(Baess et al., 2008)。换句话说，预测误差的大小看起来与所体验到的感觉衰减程度无关。

3. 最重要的是，即使对于*外部产生*的刺激(例如，由实验者产生)，只要它们被施加到正在进行自我产生运动或个体期望移动的身体部位，感觉衰减仍然会发生(Voss et al., 2008)。Voss等人指出，这种对外部施加刺激的衰减无法通过前向模型和传出拷贝(efference copy)的正常机制来解释。相反，他们提供了”仅基于高级运动准备的预测性感觉衰减的证据，排除了基于运动指令和(再)传入机制的解释”(Voss et al., 2008, p. 4)。

为了适应这些发现，Brown等人首先将我们的注意力引向第4章中描述的PP行动解释的一个有些令人困惑的复杂性。如果那个故事是正确的，当动作的感觉(本体感觉)后果被强烈预测时，运动就会随之而来。由于这些后果(被指定为本体感觉的时间轨迹)尚未实际发生，预测误差就会出现，然后通过动作的展开被压制。这是Friston (2009)和Friston, Daunizeau, et al. (2010)意义上的”主动推理”(active inference)。但请注意，只有当身体按照本体感觉预测改变而不是允许大脑改变其预测以符合当前本体感觉状态(这可能表明例如手臂当前正在桌子上休息)时，运动才会发生。在这种情况下，知觉的配方(改变感觉预测以匹配来自世界的信号)和动作的配方(改变身体/世界以匹配感觉预测)之间存在明显的紧张关系。

考虑到预测处理声称提供了一个有吸引力的感知和行动统一解释，这可能看起来令人惊讶。但事实上，正是这种统一性现在制造了麻烦。因为这个解释表明，行动是在感知控制之下的，至少在身体运动不是由独特的高级”运动指令”指定，而是隐含地——通过描述某些期望行动的本体感觉信号轨迹来指定的程度上是如此。因此，预测处理在这里表明，我们运动的形状是由关于本体感觉随着运动展开的流动的预测所决定的（见Friston, Daunizeau, et al., 2010; Edwards et al., 2012）。这些预测的本体感觉后果然后通过一系列嵌套的展开过程产生，最终形成简单的”反射弧”——流畅的例程，逐步将高级规范解析为适当的肌肉指令。

行动和感知之间的张力现在显现出来。因为消除本体感觉预测错误的另一种方式是通过改变预测以符合实际的感官输入（目前指定”手放在桌子上”的输入），而不是通过移动身体来实现预测的本体感觉流动。为了避免不动，智

能体需要确保基于与（比如）伸手取啤酒杯相关联的预测本体感觉状态的行动，胜过真实的感知（发出手目前静止不动的信号）。

有两种（功能等价的）方式可以实现这一点。要么需要降低与当前感官输入（指定手在桌子上静止不动）相关联的精度(precision)，要么需要增加与更高级别表征（指定到啤酒杯的轨迹）相关联的精度。只要这些之间的*平衡*是正确的，运动（和抓杯子）就会发生。在接下来的两节中，我们考虑如果这种平衡被改变或干扰会发生什么。

7.8 感觉更少，行动更多?

Brown, Adams, et al. (2013) 提供了一系列模拟，这些模拟（就像7.6节报告的实验一样）探索了改变感官输入和更高级别预测之间精度调节平衡的后果。这里的基本场景是给定的体感输入（例如，涉及触觉感觉的输入）是模糊生成的，可能是自我生成的力、外部施加的力或两者某种混合的结果。为了识别体感刺激的来源，系统（在这个简化模型中）必须使用本体感觉信息（关于肌肉张力、关节压力等的信息）来区分自我生成和外部生成的输入。来自更高处理级别的本体感觉预测（以主动推理的通常方式）被定位为产生运动。最后，感官预测错误的可变精度加权使系统能够在更大或更小程度上关注当前感官输入，灵活地平衡对输入的依赖（或信心）与对其自身更高级别预测的依赖（或信心）。

这样一个系统（完整实现见Brown, Adams, et al., 2013）能够在（且仅在）对当前感官输入的依赖与对更高级别预测的依赖之间的平衡正确时产生身体运动。在极限情况下，与更高级别本体感觉预测（指定期望轨迹）相关的错误将被赋予非常高的权重，而与当前本体感觉输入（指定肢体或效应器的当前位置）相关的错误将被低权重处理。这将极大地衰减当前感官信息，允许关于预测本体感觉信号的错误享有功能优先权，当系统移动以消除那些高权重错误时，成为自我实现的预言。

然而，随着感官衰减的减少，情况发生了戏剧性变化。在相反的极端情况下，当感官精度远高于与更高级别预测相关联的精度时，就没有当前感官输入的衰减，也不能发生运动。Brown等人使用许多不同的模拟运行探索了这种平衡，显示（正如现在所预期的）“随着先验精度相对于感官精度的增加，先验信念逐渐能够引发更自信的运动”（第11页）。那么，在主动推理设置中，“如果先验信念要在自我生成行为期间凌驾于感官证据之上，感官衰减是必要的”（第11页）。这已经是一个有趣的结果，因为它为7.7节中提到的各种感官衰减提供了根本原因，包括在自我生成运动期间，甚至对外部生成输入的衰减（这是最抵制基于标准前向模型解释的情况）。

重要的是，对感觉预测误差的信心降低意味着对此类误差原因的信念信心也会降低。Brown、Adams等人(2013, p. 11)将这种状态描述为由于”对感觉输入的注意力暂时暂停[回想一下，在PP中，注意力是通过增加预测误差的精度权重来实现的]“而导致的”短暂不确定性”状态。结果是，外部产生的感觉通常比内部产生的感觉记录得更加强烈。在自我产生的运动背景下，更高层次的预测能够影响运动，这只是由于当前感觉输入的衰减(精度降低)。因此，当体感状态由外部产生时，会比通过自我产生的行为产生的完全相同状态显得更加强烈(衰减更少)(见Cardoso-Leite等人, 2010)。如果然后要求受试者将外部产生的力量与内部产生的力量相匹配(如前面练习的力量匹配任务)，力量升级将立即跟随(关于使用上述主动推理设备对这种效应的一些令人信服的模拟研究，见Adams, Stephan等人, 2013)。

总之，行为(在主动推理(active inference)下)需要一种有针对性的分散注意力，其中当前感觉输入被衰减，以允许预测的感觉(本体感受)状态影响运动。乍一看，这是一个相当复杂(Heath Robinson / Rube Goldberg式)的机制[8]，涉及一种难以置信的自我欺骗。根据这个故事，只有通过淡化指定我们身体部位*实际上*当前在空间中如何排列的真实感觉信息，大脑才能”认真对待”决定运动的*预测*本体感受信息，允许这些预测(正如我们在第4章中看到的)直接作为运动命令发挥作用。在我看来，PP故事的这一部分是否正确，是这方面更大的开放性问题之一。然而，从积极的一面来看，这样的模型有助于理解熟悉的(但在其他方面令人困惑的)现象，如故意注意行为对流畅运动行为的损害(“窒息”)，以及各种躯体妄想和运动障碍，正如我们现在将看到的。

[7.9 干扰网络(感觉衰减)]

假设(为了论证)刚才叙述的故事是正确的，我们现在可以问自己，如果这种有针对性的分散注意力的能力受损或损坏会发生什么？这样的系统将无法衰减上升感觉预测误差的影响。衰减能力受损将倾向于(我们看到)阻止自我产生的运动。Edwards等人很好地总结了这种情况，他们指出：

如果高级表征的精度占主导地位，那么本体感受预测误差将通过经典反射弧得到解决，运动将随之而来。然而，如果本体感受精度更高，那么本体感受预测误差很可能通过改变自上而下的预测来解决，以适应没有感知到运动的事实。简而言之，精度不仅通过完全相同的机制决定感知中感觉证据和先验信念之间的微妙平衡，它还可以决定我们是否行动。(Edwards等人, 2012, p. 4)

如果感觉衰减受损，通常会导致运动的更高层次预测确实可能形成，但现在相对于感觉输入将享有降低的精度，使它们在功能上失效或(至少)严重受损。[9]

Brown、Adams等人(2013, p. 11)论证，这种效应模式也可能是从事某项运动或精细(但练习良好)身体活动时”窒息”日常体验的基础(见Maxwell等人, 2006)。在这种情况下，对运动的故意注意的部署似乎干扰了我们自己流畅轻松地产生运动的能力。问题可能是关注运动增加了当前感觉信息的精度，从而相应地降低了原本会影响流体运动的更高层次本体感受预测的影响。

在感觉衰减的正常过程出现高度损伤时，运动变得不可能，系统——尽管在生物力学上是健全的——却无法进行运动。Brown等人通过一个简单的仿真验证了这一点，该仿真中系统性的置信度或对感觉预测误差的确定性被改变。运动需要更高层次本体感觉预测的精确度相对于感觉证据的精确度要高。当情况相反时，运动就会被阻断。在这种条件下，恢复运动的唯一方法是人工提高更高层次状态的精确度（即增加更高层次预测误差的精确度）。被削弱的感觉衰减现在被克服，运动得以实现。这是因为更高层次的预测（它们展开为由某个目标动作所暗示的本体感觉状态轨迹）现在相对于（仍未衰减的）当前感觉状态享有增加的精确度。在某个点上（如Adams, Stephan等人2013年的仿真研究所证明的），这将允许运动发生，但同时消除力量匹配错觉（并且推测上能够使自我挠痒成为可能，如果这是仿真的一部分！）——这种组合是精神分裂症患者的特征。

然而，这种补救措施带来了代价。因为系统虽然现在能够自我产生运动，但变得容易出现各种”躯体妄想”。这是因为那些过度精确的（未衰减的）感觉预测误差仍需要得到解释。为了做到这一点，Adams, Stephan等人（2013年）以及Brown, Adams等人（2013年）研究的仿真代理推断出一个额外的外部力量——一个”隐藏的外部原因”来解释实际上纯粹是自我产生的感觉刺激模式。这个代理”相信当它用手指按压自己的手时，还有什么东西在推动它的手抵住它的手指”（Brown, Adams等人，2013年，第14页）。我们”期望”我们自己行动的感知后果相对于由外部力量引起的类似感觉后果会被衰减。但现在（尽管实际上是行动的发起者）仿真代理未能衰减这些感觉后果，释放出一股预测误差流，这招募了一个新的——但妄想性的——假设。这在精神分裂症中观察到的感觉衰减失败与关于代理的错误信念的出现之间建立了根本联系。

7.10”心因性疾病”和安慰剂效应

一个非常相似的推理干扰模式，同样是精确度加权的微妙经济变化的结果，可能解释某些形式的”功能性运动和感觉症状”。这指的是一系列所谓的”心因性”疾病，其中存在异常运动或感觉，但没有明显的”器质性”或生理原因。继Edwards等人（2012年）之后，我使用”功能性运动和感觉症状”这一术语来涵盖这类案例：这些案例有时被描述为”心因性”、“非器质性”、“无法解释的”，甚至（在较旧的说法中）“歇斯底里的”。接下来的建议同样适用于（尽管在那里它们与更积极的结果相关）理解”安慰剂效应”的效力和范围（见Büchel等人，2014年；Atlas & Wager 2012年；Anchisi and Zanon 2015年）。

功能性运动和感觉症状出人意料地常见，在大约16%的神经科患者中被诊断出来（Stone等人，2005年）。例子包括器质性无法解释的”麻醉、失明、耳聋、疼痛、疲劳的感觉运动方面、虚弱、失声、异常步态、震颤、肌张力障碍和癫痫发作”案例（Edwards等人，2012年，第2页）。令人震惊的是，困扰这些患者的问题轮廓往往遵循关于身体部位划分的”民间”概念（例如，瘫痪的手在哪里结束，未瘫痪的手臂在哪里开始）或视野的概念。另一个例子是：

“管状”视野缺陷，具有功能性中央视野缺失的患者报告相同直径的缺陷，无论是在他们附近还是远处绘制。这违反了光学定律，但可能符合关于视觉本质的（外行）信念。（Edwards等人，2012年，第5页）

类似地，机动车事故后所谓的”鞭打伤”在普通民众不了解这种伤害的预期”形态”的国家中原来是非常罕见的（Ferrari等人，2001年）。但在这种伤害被充分宣传的地方，Edwards等人（2012年，第6页）注意到”人口调查中对轻微交通事故医疗后果的期望反映了鞭打症状的发生率”。

期望和先验信念在这种（真正身体体验的）效应病因学中的作用通过其可操控性得到进一步证实。因此：

在澳大利亚一项关于轻微伤害后腰痛的相关研究中，一个全州范围的改变对此类伤害后果期望的运动导致了慢性背痛发生率和严重程度的持续显著减少（Buchbinder and Jolley, 2005年）。（Edwards等人，2012年，第6页）

一个先验信念可能影响感觉和运动表现的途径是通过注意力的分布。功能性运动和感觉症状在大量令人信服的文献中已经与身体注意力流动和分布的改变相关，更具体地说，与内省倾向和一种”身体聚焦的注意偏向”相关（见Robbins & Kirmayer, 1991，以及Brown, 2004和Kirmayer & Tailefer, 1997的综述）。试图将这一大量文献联系起来，Brown将”高级注意力对症状的重复重新分配”确定为主导线索。自然可以假设，在这种情况下，注意力的分配是结果而不是原因。然而，“追踪”民间感觉和运动生理学概念的倾向，以及用于识别功能性运动和感觉症状病例的诊断体征都反对这一观点。例如，“如果要求患有功能性腿部无力的患者弯曲其未受影响的髋部，其未被注意的’瘫痪’髋部将自动伸展；这被称为胡佛征(Hoover’s sign)（Ziv et al., 1998）”（Edwards et al., 2012，第6页）。在广泛的病例中，当受试者不注意受影响的部位时，功能性感觉和运动症状因此被”掩盖”。

当在预测处理框架内考虑时，功能性运动和感觉症状与注意力分配异常之间的这种联系特别具有启发性。在该框架内，正如我们在许多场合所注意到的，注意力对应于根据其估计精度（逆方差）对各个处理级别的预测误差信号的加权。这种加权决定了自上而下的期望和自下而上的感觉证据之间的平衡。如果我们一直在追求的模型类别是正确的，那么同样的平衡决定了感知什么以及如何行动。这开辟了一个空间来探索感觉和运动领域功能性症状病因学的统一模型。

[7.11 干扰网络（“心因性”效应）]

Edward等人认为，导致功能性运动和感觉症状的根本问题可能是感觉运动处理的（首先是）中间[12]级别精度加权机制的干扰。这种干扰（本身与任何其他生理功能障碍一样最终是生物性的）将包括在该中间级别对预测误差的过度加权，导致对该特定概率期望集的一种系统性过度信心，因此对任何似乎符合它们的自下而上的感觉输入也是如此。

假设这发生在某些显著的身体事件的背景下，如受伤或病毒感染。这些事件经常（但并非总是，见下文）在功能性运动或感觉症状发作之前出现(Stone et al., 2012)。在这种条件下：

来自这些诱发事件的显著感觉数据被赋予过度的精度（权重）…这在皮层层次结构的中间级别实例化了一个异常的先验信念，试图解释或预测这些感觉——该异常信念或期望通过在其形成过程中享有的异常高水平的精度（突触增益）而变得抗消退。（Edwards et al., 2012，第6页）

然而，诱发事件并不是功能性运动和感觉症状的必要条件（根据该模型）。因此，假设无论什么原因，形成了某种亚个人的、中间级别的对感觉或身体运动的期望（或同样地，对感觉或身体运动*缺乏*的某种期望）。由于该级别预测误差报告的干扰（夸大）精度，该中间级别预测现在享有增强的地位。现在，即使是随机噪声（正常范围内的波动）也可能被解释为信号，并且”检测”到刺激（或缺乏刺激）。这就是我们在文本中多次遇到的”白色圣诞节”效应。换句话说，从预测处理的角度来看，“躯体放大”和产生完全虚假的感知之间”可能只有量的——而非质的——差异”（Edwards et al., 2012，第7页）。

为了完整描述这幅图景，请注意，选择和”确认”中间层级假设的精确预测误差有助于强化中间层级先验(prior)，因此自我稳定了这种误导性的躯体推理模式。与此同时，更高层级的网络必须试图理解这些表面上得到确认但实际上是病理性自我产生的刺激模式（或缺乏刺激）。没有更高层级的解释可用，例如某些预期的感知或感觉，或者移动或不移动的系统性决定。可以说，更高层级并没有预测这种运动或感觉，尽管它源自系统本身。为了理解这一点，需要推断新的原因——比如基本疾病或神经损伤。简而言之，出现了Edwards et al. (2012，第14页）所描述的”代理归属错误(misattribution of agency)，即通常以自主方式产生的体验被感知为非自主的”。自我产生的感觉现在被归类为某种难以捉摸的生物功能障碍的症状。确实如此：但这种功能障碍也许也可以，也许更恰当地被认为是控制论的：即证据、推理和控制的复杂内在经济中的不平衡。

同样的广泛故事也适用于前面提到的所谓”安慰剂效应(placebo effects)“。近几十年来，人们越来越认识到这种效应的力量和范围（综述见Benedetti, 2013；Tracey, 2010）。期望值，相当普遍地，明显影响治疗的行为、生理和神经结果，无论是在惰性（经典安慰剂）治疗的背景下，还是在真实治疗的背景下（Bingel et al., 2011；Schenk et al., 2014）。在最近一篇关于”安慰剂镇痛(placebo analgesia)“的综述文章中（尽管作者更倾向于说”安慰剂痛觉减退(placebo hypoalgesia)“，从而强调基于期望的疼痛减轻而非疼痛消除），建议：

上行和下行疼痛系统类似于一个循环系统，允许实施预测编码(predictive coding)——意味着大脑不是被动地等待伤害性[疼痛性]刺激冲击它，而是基于先前经验和期望主动进行推理。（Buchel et al., 2014，第1223页）

建议是，疼痛缓解的自上而下预测在神经层级的多个层次上与自下而上的信号相结合，以一种由其估计精度调节的方式——分配给预测的确定性或可靠性。这为复杂仪式、明显复杂的干预措施以及患者对医生、从业者和治疗的信心的记录影响提供了非常自然的解释。

自闭症、噪音和信号

对证据、推理和期望的同一复杂经济的干扰可能（Pellicano & Burr, 2012）有助于解释所谓自闭症”非社交症状”的起源。这些是在感觉而非社交领域表现出来的症状。社交症状包括众所周知的识别其他代理人情感和意图的困难，以及对许多形式社交互动的厌恶。非社交症状包括对感觉——特别是意外感觉——刺激的过敏、重复行为以及高度规范化、受限的兴趣和活动。关于社交和非社交元素整个星座的概述，见Frith (2008)。

在感知领域的一个关键发现（Shah & Frith, 1983）是自闭症被试在当某个元素（如三角形）出现在更大有意义图形（如婴儿车图片）的背景中”隐藏”时，找到该元素的增强能力。在这种”嵌入图形”任务上始终优于神经典型被试的能力导致了这样的建议（Frith, 1989；Happé & Frith, 2006），即自闭症被试表现出”弱中央连贯性(weak central coherence)“，即一种突出部分和细节而牺牲对其发生的更大背景的轻松把握的处理风格。Pellicano和Burr指出，自闭症和神经典型人群之间存在显著感知处理差异的假设得到了进一步支持，研究显示自闭症被试对某些视觉错觉不太敏感（如Kanizsa三角形错觉、我们在第1章遇到的空心面具错觉，以及桌面错觉，见图7.1）。自闭症被试也更可能拥有绝对音高(absolute pitch)，在许多形式的视觉辨别方面表现更好（见Happé, 1996；Joseph et al., 2009；Miller, 1999；Plaisted et al., 1998a，b），并且对空心面具错觉不太敏感（Dima et al., 2009）。

图片

图7.1 **自闭症被试对使用先验知识解释模糊感觉信息的错觉不太敏感**

此类错觉的例子包括：(a) 卡尼扎三角形。三角形的边缘实际上并不存在，但在最可能的物理结构中会出现：一个白色三角形覆盖在三个规则圆形上。(b) 凹面错觉。对自然凸面的强烈偏见（或”先验”）抵消了竞争信息（如阴影），使人将凹面、空心面具（右）感知为正常的凸面（左）。(c) 谢泼德桌子错觉。平行四边形的2D图像实际上是相同的。然而，图像与许多3D形状一致，最可能的是倾斜约45°的真实桌子：为了与相同的2D图像保持一致，桌面需要具有非常不同的尺寸。

来源：Pellicano & Burr, 2012。

基于这些证据，一些作者（Mottron et al., 2006；Plaisted, 2001）探讨了自闭症涉及异常增强或加强的感觉体验的观点。这些观点被提出作为弱中央连贯性(weak central coherence)或削弱的自上而下期望影响概念的替代方案。然而，请注意，从广义贝叶斯视角来看，这种明显的对立失去了一些力量，因为真正重要的（见Brock, 2012）是自上而下和自下而上影响模式之间实现的平衡。

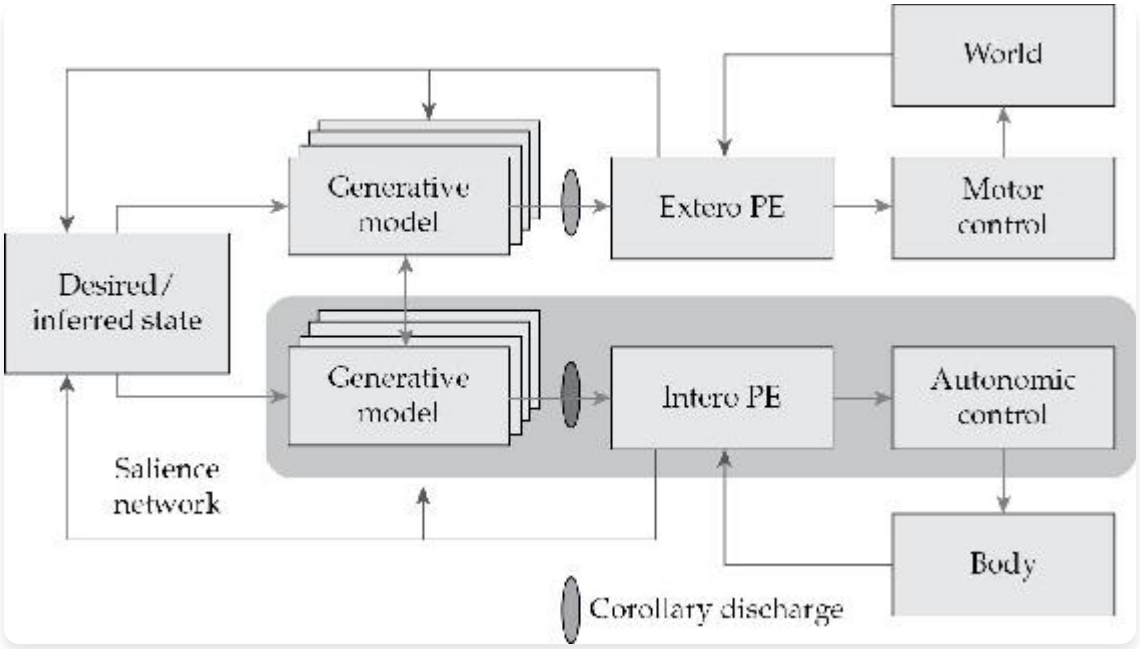
从贝叶斯视角出发，Pellicano和Burr将自闭症感知描述为由于先验知识影响减弱而导致的处理感觉不确定性系统性能力的干扰。[16] 这种削弱影响的结果是产生了一种积极的能力，即将更多传入信息视为信号而非噪音（导致发现隐藏图形和识别感觉数据真实轮廓的能力增强）。但这反过来意味着，大量传入信息被视为显著且值得关注，从而增加了费力的处理并产生显著的情感成本。例如，当神经正常的儿童学会在各种光照条件下识别物体，并能将物体投下的阴影用作有用信息来源时，这种情况对自闭症受试者来说是具有挑战性的（Becchio et al., 2010）。阴影不是作为与当前环境中某个物体存在相关的可预测感觉刺激模式而落入原位，而可能被视为需要进一步解释的感觉数据。换句话说，自上而下预测（先验）的影响通常可能起到——正如预测处理模型所建议的——剥离感觉信号大部分”新闻价值”的作用。这种影响的削弱（Pellicano和Burr称之为”低先验”[17]）会导致不断的信息轰炸，需要进一步处理，可能合理地产生严重的情感成本，并有助于出现涉及重复、隔离和注意力范围缩小的各种自我保护策略。

在我看来，这样的解释不仅作为适应自闭症”非社会症状”的手段具有前景，而且作为这些症状与流畅社会参与和人际理解障碍之间的潜在桥梁。人们可能合理地怀疑，领域越复杂，先验减弱对推理以及（因此）对表现和反应的影响越大。社会领域高度复杂（经常涉及对观点的观点的理解，比如当我们知道约翰怀疑玛丽没有说实话时）。此外，这是一个背景（正如每个肥皂剧粉丝都知道的）决定一切的领域，其中小的言语和非言语信号的含义必须在丰富的先验知识背景下进行解释。前面描述的信号/噪音失衡类型可能因此导致社会互动和（作为结果）社会学习的特别显著困难。正是在这个方向上，Van de Cruys et al. (2013) 建议：

自闭症中的繁重体验（参见感觉过载）可能源于一个持续发出预测错误信号的感知系统，表明总是还有东西需要学习，需要注意力资源。伴随的负面感受可能导致这些患者避免最多变或不可预测的情况，在这些情况下，依赖背景的高级预测比具体的感知细节更重要。这可能特别适用于社会互动。压倒性的预测错误导致这些患者（或他们的照顾者）通过日常活动中的精确例行程序和模式来外化和强化可预测性。（Van de Cruys et al., 2013，第96页）

然而，Van de Cruys等人建议，与其简单地从减弱先验(attenuated priors)的角度思考，不如专注于调节不同层级先验影响的机制，这可能更有成效。在预测处理框架内，这对应于根据任务和情境需求调节精度(precision)。为了支持这一提议，作者引用了各种证据，表明自闭症受试者可以构建和部署强先验，但可能在应用它们时遇到困难。如果这些先验被构建以适应一个信号，而这个信号从神经典型视角来看实际上包含大量噪音但被当作精确信号处理，那么这种情况就可能发生。然而，这种解释与Pellicano和Burr的更一般性描述之间并无深层冲突，因为在处理的各个层级为预测误差分配精度本身就需要精度估计。正是这些估计（技术上称为超先验(hyperpriors)）影响力的减弱，解释了上述讨论的各种效应范围（见Friston, Lawson, & Frith, 2013）。因此，自闭症和精神分裂症都可能涉及对这种复杂神经调节经济的（不同但相关的）干扰，影响体验、学习和情感反应。

总而言之，在（精度调节的）倾向性方面的变异——即将多少传入感官信息视为”新闻”，以及更广泛地说，在处理的各个阶段灵活修改自上而下和自下而上信息平衡的能力方面的变异——将在决定感知体验的性质和内容方面发挥重要作用。在一般人群中也可能存在这些维度上的一些变异，可能有助于学习风格和偏好环境的差异。因此，我们瞥见了一个丰富的多维空间，可以开始捕捉自闭症受试者之间以及神经典型人群内部所见的广泛变异。



image

图7.2 **Seth的内感受推理模型**

在该模型中，情绪反应依赖于对内感受输入原因的持续更新预测。从期望或推断的生理状态开始（这本身也会基于更高层级的动机和目标导向因素进行更新），生成模型(generative models)通过推论性放电(corollary discharge)预测内感受（和外感受）信号。应用主动推理(active inference)，预测误差(PEs)通过参与经典反射弧（运动控制）和自主反射（自主控制）转录为行动。由此产生的预测误差信号用于更新（功能耦合的）生成模型和生物体的推断/期望状态。（在高层级这些生成模型合并为单一多模态模型。）内感受预测被提议在以前脑岛皮层和前扣带皮层（AIC, ACC）为锚点的”显著性网络”（阴影区域）内生成、比较和更新，该网络将脑干区域作为内脏运动控制的目标和传入内感受信号的中继。来自AIC和ACC的交感和副交感流出是内感受预测形式，支配自主反射（如心率/呼吸频率、平滑肌行为），就像在运动控制的PP公式中本体感受预测支配经典运动反射一样。这个过程依赖于内感受（和本体感受）PE信号精度的暂时减弱。浅色/深色阴影箭头表示自上而下/自下而上连接。

来源：Seth, 2013。

[7.13 意识临场感]

刚才讨论的精神分裂症、自闭症以及功能性运动和感觉症状的解释在计算、神经科学和现象学描述之间无缝转换。我们已经看到，这种流畅的跨层级描述是预测处理模型群的标志之一。我们能否使用这套装置来阐明人类体验的其他方面？

其中一个方面是”意识临场感”(conscious presence)的感觉。[18]使用分层预测处理作为理论框架，Seth等人（2011）勾勒出对这种感觉的初步理论解释，这种感觉可以解释为真正存在于某个现实世界环境中的感觉。尽管这种解释是推测性的，但与大量已有理论和证据一致（总结见Seth等人，2011；重要发展见Seth，2014；综述见Seth，2013）。

世界现实感的改变或丧失被称为”现实解体”(derealization)，自我现实感的改变或丧失被称为”人格解体”(depersonalization)，出现其中任一或两种症状都被标记为人格解体障碍（DPD）（见Phillips等人，2001；Sierra & David，2011）。DPD患者可能描述世界似乎与他们隔离，仿佛他们是在镜子中或透过玻璃看到世界，DPD症状经常出现在精神分裂症等精神病的早期（前驱期）阶段，在妄想或幻觉等阳性症状出现之前，可能先出现普遍的”陌生感或非现实感”(Moller & Husby，2000)。

Seth等人建议，存在感的产生源于对内感受感觉信号的成功抑制（通过成功的自上而下预测）。内感受感觉信号是关于身体当前内部状态和条件的信号（顾名思义）——因此它们构成了一种”内在感知”形式，其目标包括内脏状态、血管舒缩系统、肌肉和供气系统等诸多方面。从主观角度来看，内感受意识表现为一系列不同的感觉，包括”疼痛、温度、瘙痒、感官触觉、肌肉和内脏感觉……饥饿、口渴和’空气饥饿’“（Craig, 2003，第500页）。因此，内感受系统主要关注疼痛、饥饿和各种内脏器官的状态，它不同于包括视觉、触觉和听觉的外感受系统，也不同于传递肢体相对位置、努力和力量信息的本体感受系统。最后，人们认为前岛叶皮质(AIC)在内感受信息的整合和使用中发挥特殊作用，在情感意识的构建中发挥更广泛的作用——也许是通过编码Craig (2003，第500页)所描述的”原始内感受活动的元表征”。

Seth等人提出了两个相互作用的子机制，一个涉及”能动性”并牵涉感觉运动系统，另一个涉及”存在感”并牵涉自主神经和动机系统（见图7.2）。这里的能动性组件从我们之前的一些讨论中会很熟悉，因为它基于精神分裂症中能动性感觉障碍的原始模型（Blakemore et al., 2000；另见Fletcher & Frith, 2009）。该解释（见4.2）将预测我们行为的感觉后果的能力减弱（精确度不足）作为产生异己控制感等感觉的主要组成部分。但该解释与最近的修正（7.7-7.9）兼容，这些修正表明主要病理实际上可能是未能减弱上行感觉预测误差的影响。兼容性（就目前目的而言）是有保证的，因为从功能角度来说，重要的是分配给下行预测和上行感觉信息的精确度之间的平衡——这种平衡可能因为高估某些较低层级信号的精确度（因此未能减弱当前感觉状态的影响）或低估相关较高层级预测的精确度而被破坏。

Seth等人建议，存在感的产生源于涉及解释外感受和本体感受误差的系统与涉及另一种预测类型的系统之间的相互作用：对我们自身复杂内感受状态的预测。Seth等人推测，这里的一个关键部位可能是AIC，因为这个区域（如上所述）被认为整合各种内感受和外感受信息（见Craig, 2002；Critchley et al., 2004；Gu et al., 2013）。AIC也已知参与对疼痛或情感负荷刺激的预测（见Lovero et al., 2009；Seymour et al., 2004；并回顾7.9中的讨论）。AIC还会因观看人们抓挠的电影而激活，其激活水平与观看者体验到的”瘙痒传染”程度相关（Holle et al., 2012），表明内感受推理可以是社会性的，也可以是生理性的（见Frith & Frith, 2012）。

Seth等人建议，通过对内感受状态起伏的成功自上而下预测来抑制AIC活动，会产生存在感（或至少是缺乏非存在感，见注释18）。而他们认为，人格解体障碍(DPD)源于病理性不精确的内感受预测。这种不精确（因此功能上被削弱的）下行预测将无法解释传入的内感受信息流，导致产生持续的（但没有根据的）预测误差爆发。这可能在主观

上表现为一种难以解释的奇异感，产生于外感受和内感受期望的交汇点。最终，在严重情况下，持续试图解释这种持续的误差信号可能导致新的但奇异的解释模式的出现（关于我们自身体现和能动性的妄想信念）。因此，Seth等人的解释在结构上与Fletcher和Frith (2009)的解释同构，如[第2章]所述。

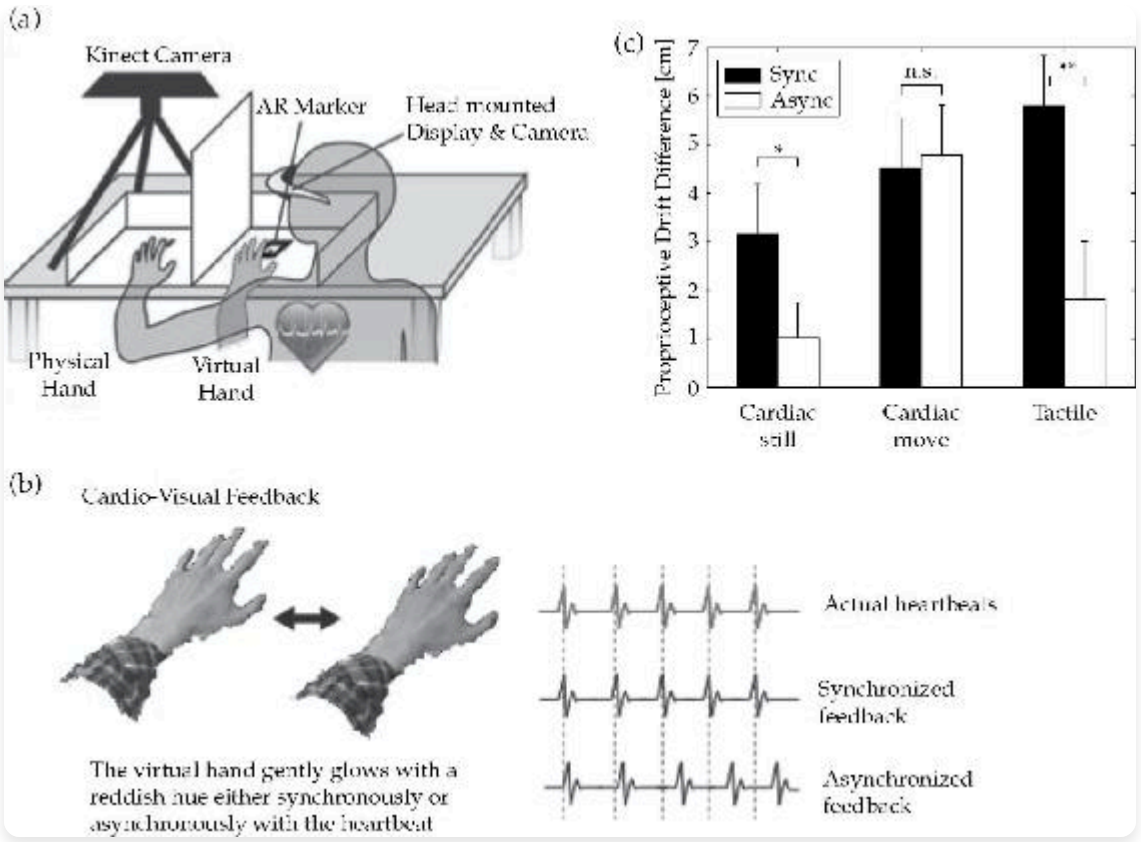
此外，越来越多的证据支持基于内感受推理的更一般性身体所有权体验(EBO)解释，这里的身体所有权体验是指”拥有并认同特定身体”的体验(Seth, 2013, p. 565)。这里的观点同样是，EBO可能是推理过程的结果，涉及”跨内感受和外感受域的自我相关信号的多感官整合”(Seth, 2013, pp. 565 – 566)。显然，我们自己的身体是世界中极其重要的一部分，如果我们要生存和繁荣，就必须对其保持某种掌控。我们必须建立并维持对自己身体位置(我们在哪里)、身体形态(当前形状和构成)以及内部生理状态(由饥饿、口渴、疼痛和唤醒状态所指示)的掌控。为了做到这一点，Seth (2013)论证说，我们必须学习和部署一个生成模型，该模型能够隔离”跨内感受和外感受域最可能是我的那些信号的原因”。这并不像听起来那么困难，因为当我们在世界中移动、感知和行动时，我们自己的身体处于独特的位置，能够生成各种时间锁定的多模态信号。这些包括外感受和内感受信号，它们以由我们自己的运动密切决定的方式一起变化。因此：

在世界上所有物理对象中，只有我们的身体才会唤起(即预测)这种多感官感觉——一种多感官输入的一致性，这种一致性被称为”自我指定性”(Botvinick, 2004)。(Limanowski & Blankenburg, 2013, p. 4)

身体感觉在这个建构过程中的重要性在许多著名的涉及所谓”橡胶手错觉”的研究中得到了证明(Botvinick & Cohen, 1998)，这在6.11节中提到过。回忆一下，在这些研究中，一只可见的人工手与被试的真手同步地被抚摸。注意视觉呈现的人工手会诱发对该手的短暂所有权感——当手突然受到锤子威胁时，这种所有权感会转化为真正的恐惧。所有这些都表明，我们对自己身体的体验是一个生成模型的持续产物，该模型从时间锁定的多模态感官信息洪流中推断身体位置和构成。考虑到在可见但非拥有的手上感受到精细定时的触摸序列在生态学上的不可能性，我们降级了信号的某些元素(那些指定实际手的精确空间位置的元素)，以得出最佳的整体假设——现在将橡胶手作为身体部分纳入其中。Seth指出，这种效应非常稳健，后来已扩展到面部感知和全身所有权(Ehrsson, 2007; Lenggenhager et al., 2007; Sforza et al., 2010)。

Suzuki et al. (2013)扩展了这些研究，纳入了内感受(而不仅仅是触觉)感官证据，他们使用虚拟现实头戴设备使显示的橡胶手与被试自己的心跳同步或不同步地”脉动”(通过改变颜色)。内感受心律与视觉脉冲之间的同步性增强了橡胶手所有权感(见图7.3)。这个令人兴奋的结果提供了第一个明确的证据表明：

内感受(例如心脏)和外感受(例如视觉)信号之间的统计相关性可以通过预测误差最小化导致自我相关信号预测模型的更新，就像在经典的RHI[橡胶手错觉]中纯粹的外感受多感官冲突可能发生的那样。(Seth, 2013, p. 6)



image

图7.3 **内感受橡胶手错觉**

a. 参与者面向桌子坐着，使他们的物理(左)手不在视线内。真实手的3D模型由Microsoft Kinect捕获，用于生成实时虚拟手，该虚拟手被投影到头戴式显示器(HMD)中增强现实(AR)标记的位置。被试佩戴连接到HMD的前置摄像头，因此他们看到摄像头图像叠加了虚拟手。他们还佩戴脉搏血氧计来测量心跳时间，并用右手做出行为反应。(b) 心脏-视觉反馈(左)通过在500毫秒内将虚拟手的颜色从自然颜色变为红色再变回来实现，要么与心跳同步，要么异步。触觉反馈(中)由画笔给出，该画笔被渲染到AR环境中。适用于AR环境的”本体感觉漂移”(PD)测试(右)通过实现虚拟测量器和光标客观测量感知的虚拟手位置。

PD测试通过要求参与者移动光标到估计位置来测量对真实（隐藏）手部感知位置 (c) 实验包含三个区块，每个区块四次试验。每次试验包含两个PD测试，中间穿插一个诱导期，在此期间提供心血管-视觉或触觉-视觉反馈（120秒）。每次试验以在HMD中呈现的问卷结束。(D) 在”心脏静止”（无手指运动）条件下，同步与异步心血管-视觉反馈的PD差异（PDD，诱导后减去诱导前）显著更大，但在”心脏运动”条件（有手指运动）下则无显著差异。同步与异步触觉-视觉反馈（“触觉”条件）的PDD也显著更大，重现了经典的RHI。每个柱状图显示跨参与者平均值和标准误差。(E) 探测所有权体验的主观问卷回应显示与PDD相同的模式，而控制问题显示心血管-视觉或触觉-视觉同步性无效应。

来源：改编自Seth (2013)，已获许可。

我们对自身体现性(embodiment)的持续感知，所有这些都表明，依赖于使用一个生成模型来容纳完整的（内感受和外感受）感觉冲击，该模型的维度关键地追踪我们自身的各个方面——我们的身体阵列、我们的空间位置，以及我们

自身的内在生理状态。

[7.14 情绪]

这些相同的资源可能被部署为一个有前景的情绪解释的起点。在这一点上，该提议借鉴了（带有转折的）著名的詹姆斯-兰格情绪状态模型，该模型认为情绪状态源于对我们自身对外部刺激和事件的身体反应的感知(James, 1884; Lange, 1885)。简而言之，那里的观点是，我们的情绪”感受”无非就是对我们自身变化的生理反应的感知。根据詹姆斯的观点：

*身体变化直接跟随对令人兴奋事实的感知，而…我们对这些变化发生时的感受就是情绪。*常识说，我们失去财富，感到悲伤并哭泣；我们遇到熊，感到害怕并逃跑；我们被对手侮辱，变得愤怒并攻击。这里要为之辩护的假设说这种顺序是不正确的…更理性的陈述是我们因为哭泣而感到悲伤，因为攻击而愤怒，因为颤抖而害怕…如果没有跟随感知的身体状态，后者在形式上将是纯粹认知的，苍白的、无色的、缺乏情感温暖的。那么我们可能会看到熊，并判断最好逃跑，受到侮辱并认为应该攻击，但我们不会真正感到害怕或愤怒。(James, 1890/1950, p. 449)

换句话说，对詹姆斯来说，正是我们对恐惧特征性身体变化（出汗、颤抖等）的内感受性感知构成了恐惧的真正感受，赋予它独特的心理特色。如果詹姆斯是对的，恐惧感本质上是对已经被暴露于威胁情境所诱发的生理特征的检测。

这样的解释是有前景的，但就其本身而言远非充分。因为它似乎需要不同情绪状态与独特的”原始生理”特征之间的一对一映射，并且似乎表明每当生理状态被诱发和检测时，相同的情绪感受就应该产生。这些含义都没有（见Critchley, 2005）得到观察和实验的证实。然而，基本的詹姆斯-兰格故事在重要工作中得到了扩展和完善，如Critchley (2005)、Craig (2002、2009)、Damasio (1999、2010)和Prinz (2005)。Seth (2013)和Pezzulo (2013)的最新工作继续了这一改进轨迹，添加了重要的”预测性转折”。Seth和Pezzulo各自建议，一个被忽视的核心组成部分是我们自身内感受状态的级联系列自上而下预测与（内感受）感觉预测误差中包含的前向流动信息之间的匹配（或不匹配）。这个故事表明，这种内感受预测”来自多个分层级别，高层级在制定下行预测时整合内感受、本体感受和外感受线索”(Seth 2013, p. 567)。

这些内感受、本体感受和外感受预测在不同情境中构建方式不同，每个都为其他预测提供持续指导。这里的单一推理过程整合所有这些信息来源，生成一个反映情境的融合体，被体验为情绪。因此，感受到的情绪整合基本信息（例如，关于身体唤起）与可能原因的高层级预测和可能行动的准备。通过这种方式：“内感受和外感受推理之间的密切相互作用意味着情绪反应不可避免地被认知和外感受情境所塑造，并且唤起内感受预测的感知场景总是会被情感性地着色”(Seth 2013, p. 563)。

前岛皮质(Anterior Insular Cortex)——如前所述——在这样的过程中发挥重要作用的位置极为有利。这个故事表明，情绪和主观感觉状态的产生，是多层次推理的结果，这些推理将感觉信号（内感受性、本体感受性和外感受性）与自上而下的预测相结合，从而产生对我们状况的感知以及我们可能即将要做什么的感觉。这种”行动准备状态”的感觉涵盖了我们的背景生理条件、对当前行动潜力的估计，以及对更广阔世界状态的感知。

这为容纳大量实验结果提供了一种新的自然方式，这些实验结果表明，我们情绪体验的特征既依赖于原始身体信号的内感受，也依赖于更高层次的”认知评估”（见Critchley & Harrison, 2013; Dolan, 2002; Gendron & Barrett, 2009; Prinz, 2004）。原始身体信号的一个例子是由——以Schachter and Singer (1962)的经典例子为例——肾上腺素注射引起的一般性唤醒。这些原始信号与情境诱导的”认知评估”相结合，使我们根据框架期望将同样的身体”证据”解释为兴奋、愤怒或欲望。这些实验很难重复，[20] 但更好的证据来自最近巧妙地操纵内感受反馈的研究——例如，显示虚假心脏反馈可以增强对情绪刺激主观评价的研究（见Valins 1966; Gray et al., 2007; 以及Seth 2013中的讨论）。

“预测性转向”因此允许我们将詹姆斯-兰格理论的核心洞察（内感受性自我监控是构建情绪体验的关键组成部分的观点）与其他因素（如情境和期望）作用的完全整合说明相结合。之前结合这些洞察的尝试采取了所谓”双因子”理论的形式，这些理论将主观感觉状态描述为本质上涉及两个组成部分的混合状态——身体感觉和”认知”解释。值得强调的是，新兴的情绪预测处理说明本身并不是”双因子”理论。相反，其主张是一个单一的、高度灵活的过程流畅地将自上而下的预测与各种自下而上的感觉信息相结合，主观感觉状态（连同外感受知觉体验的全部范围）由这一单一过程的持续展开所决定。[21]

这样的过程将涉及跨多个区域的分布式神经活动模式。这些模式本身将根据任务和情境，以及自上而下和自下而上影响之间的相对平衡而改变和变化（特别参见第2章和第5章）。重要的是，同样的过程不仅决定了知觉和情绪的流动，也决定了行动的流动（第4-6章）。因此，PP假设一个单一的、分布式的、不断自我重新配置的、预测驱动的机制作为知觉、情绪、推理、选择和行动的共同基础。在我看来，PP的情绪说明属于与所谓”行动主义”说明相同的大阵营（见Colombetti, 2014; Colombetti & Thompson, 2008; 以及第9章的讨论），这些说明拒绝任何根本的认知/情绪分歧，强调大脑、身体和世界之间持续的相互作用。

夜晚的恐惧

Pezzulo (2013)发展了一个在许多方面与Seth (2013)和Seth et al. (2011)互补的说明。Pezzulo的目标是看似非理性的”夜晚恐惧”体验。以下是Pezzulo开始其论述时使用的情景：

这是一个刮风的夜晚。你带着一点震惊上床睡觉，因为比如说，你发生了一个小车祸或刚看了一部鲨鱼袭击的恐怖电影。夜里，你听到窗户吱嘎作响。在正常情况下，你会把这种噪音归因于刮风的夜晚。但这个夜晚，小偷甚至杀手正在进入你房子的想法跳入你的脑海。通常你会立即排除这个假设，但现在它似乎相当可信，尽管你的镇上在过去几年里没有发生过盗窃案；你突然发现自己在期待一个小偷从阴影中出现。这怎么可能？（Pezzulo, 2013, p. 902）

Pezzulo论证，解释再次涉及内感受预测。因此，假设我们只考虑外感受感觉证据。考虑到我们的先验，风的假设然后提供了”解释”感觉数据的最佳方式。即使我们添加一些来自看到事故或观看电影的小的偏差或启动效应，仅此一点似乎不太可能改变那个结果。吱嘎作响的门声和移动阴影的景象肯定仍然最好由一个刮风但在其他方面安全正常的情况这一简单假设来解释。

然而，一旦我们将内感受预测(interoceptive prediction)的效应加入其中，情况就发生了变化。现在我们有两组需要解释的感觉证据。一组包括刚才提到的当前视觉和听觉信息。然而，另一组包括7.13中描述的复杂多维内感受信息（包括动机信息，以记录饥饿、口渴等的内感受状态形式）。让我们假设观看事故或恐怖电影——也许在睡前回忆起它——导致身体状态的改变，如心率加快和皮肤电反应增强，以及其他表示广泛唤醒的内部迹象。现在有两个同时发生的感觉证据流需要被”解释掉”。此外（在我看来，这是Pezzulo论述中的关键步骤），其中一个证据流——内感受流——通常以极大的确定性为人所知。揭示我们自身身体状态（如饥饿、口渴和广泛唤醒）的内感受证据流通常被赋予高可靠性，因此与这些状态相关的预测误差将享有高精度和强大的功能效力。[22]此时，Pezzulo论证，贝叶斯平衡更强烈地倾向于夜间小偷这一替代假设（最初看似不可信）。因为这个假设解释了两组数据，并且受到内感受数据——被估计为高度可靠的数据——的强烈影响。[23]

因此，这一解释——如同Seth (2013)的观点——为詹姆斯-兰格模型提供了一种贝叶斯阐释，根据该模型，感受情绪的各个方面涉及对我们自身身体（内脏、内感受）状态的感知。现在加入预测维度使我们能够将这一独立具有吸引力的提议与分层预测处理的完整解释机制联系起来。Seth和Pezzulo各自建议，感受情绪的相关方面取决于我们自身内感受和外感受期望以及传入的内感受和外感受感觉流的结合。此外，这所涉及的各种制衡本身由对以下方面的持续估计决定：(1)各种类型感觉信号的相对可靠性，以及(2)自上而下期望和自下而上感觉信息的相对可靠性。如果[第6章]的论述正确，所有这些都发生在一个根本上面向行动的经济体中，涉及对多种概率性可供性(affordances)的估计——

多种分级的行动和干预潜能。我们现在可以推测，这种可供性将部分由内感受信号选择和细化，使得Lowe and Ziemke (2011)所称的”行动倾向预测-反馈循环”成为可能：循环互动，其中情绪反应反映、选择和调节身体状态和行动。结果是一个极其复杂的认知-情绪-行动导向经济体，其基本指导原则简单而一致：多层次、多区域的预测流，在每个阶段都被我们自身不确定性的变化估计所影响。

[7.16 硬核内容的一瞥]

我相信，对不确定性、预测和行动的反思对于我们开始弥合人类生活体验世界与对心智和理性内部（和外部）机制的认知科学理解之间令人生畏的鸿沟是至关重要的。诚然，本章报告的关于精神分裂症、自闭症、功能性感觉和运动系统、人格解体障碍(DPD)、情绪和”夜间恐惧”的基于不确定性的讨论充其量是试探性和初步的。但它们开始在大致轮廓上暗示，通过计算和”系统层面”理论化，将我们的神经生理学理解与人类体验的形态和本质联系起来的方式。也许最重要的是，它们以一种开始汇集（也许是有史以来第一次）对各种神经心理学障碍的感知、运动、情绪和认知维度理解的方式做到这一切——表面上不同的元素现在在单一体系中结合在一起。

所呈现的图景与人类生活体验极其吻合。我认为，之所以如此，正是*因为*它将许多元素（感知、认知、情绪和运动）结合在一起，而以往的认知科学理论往往将这些元素分离开来。在我们的生活体验中，我们首先和最重要地遇到的是一个值得行动的世界：一个充满物体、事件和人的世界，呈现为适合参与的，充满情感、欲望和有意识与无意识期望的丰富网络。[24]要理解这个复杂经济体实际如何运作，如何应对各种干扰，以及如何支持甚至在”神经典型”体验内的巨大个体差异，一个关键工具是认识到这种复杂流在每个层面都受到估计不确定性的调节。在这里，PP表明，编码预测误差信号的估计精度或可靠性的系统发挥着关键且令人惊讶地统一的作用。[25]

这些模型留下了许多重要的未解之谜。例如，在精神分裂症患者的案例中，主要的病理机制真的是感觉预测误差影响的衰减失败，还是自上而下期望影响的因果性先行弱化？从贝叶斯的角度来看，这些结果是无法区分的，因为重要的是（正如我们现在经常提到的）自上而下期望和自下而上感知之间的平衡。但从临床角度来看，这些选择是截然不同的，涉及神经实现的不同方面。此外，还有人提出，低水平感觉衰减失败的一个结果实际上可能是人为地夸大了高水平的精确性，从而使运动成为可能（但代价是增加了对各种代理和控制错觉的暴露）。这种基于不确定性的微妙检查和平衡系统可能因此以许多不同的方式受到干扰，其中一些很难与不同的行为结果联系起来。

然而，探索改变或干扰这种微妙的检查和平衡系统的多种方式，为使用单一理论装置和单一桥接概念来解释各种各样的情况（包括”神经典型”反应的巨大变异）提供了黄金机会：分层面向行动的预测处理与不确定性估计的干扰。因此，我们可能正在进入（或至少在不太遥远的地平线上发现）“计算精神病学”的黄金时代（[Montague et al., 2012]），其中表面上不同的症状集可能通过对涉及感知、情感、推理和行动的核心机制的微妙不同干扰来解释。PP表明，这些干扰主要是对注意力和有针对性的非注意力的（多重和多样的）机制的干扰。突出各种形式的注意力暗示了与许多现有形式的治疗和干预的未来桥梁，从认知行为疗法到冥想，以及患者自身结果期望的惊人有效作用。

使用这种综合装置，我们是否可能逐寸逐现象地开始解决所谓的意识体验本身的”困难问题”——即为什么作为一个沉浸在视觉、声音和感觉世界中的人类代理者，感觉*像这样*（或者确实，像任何东西）的奥秘（[Chalmers, 1996]; [Levine, 1983]; [Nagel, 1974]）？现在说还为时过早，但感觉像是在进步。我们看到，这种进步很大程度上依赖于关于”内感受推理”作用的一系列最新经验知情的推测——大致是对我们自身内在身体状态的预测和适应。综合起来，并与关于预测、行动和想象的丰富PP描述自由混合，这些提供了一个惊人熟悉的愿景：一个生物自身的身体需求、状态和身体存在感构成了与结构化和内在有意义的外部世界进行认知、主动接触的支点。这种多层纹理，其中外部原因和有机体显著行动机会的世界以一种与对其自身身体状态的掌握不断交织的方式呈现给生物，可能就在我们称之为”意识体验”的那个永远难以捉摸而又永远熟悉的野兽的核心。

因此揭示的世界是一个为行动量身定制的世界，由内感受、本体感受和外感受期望的复杂多层模式构成，并由有针对性的注意力和估计的不确定性加以细化。这是一个意外缺失与真实存在同样显著（相对于我们最佳多层预测的新闻价值）的世界。这是一个结构和机会的世界，不断受到外部和内部（身体）语境的影响。通过将这个熟悉的世界重新带入视野，PP为理解代理性、体验和人类意义提供了一种独特而有前途的方法。

第三部分

支撑预测

第8章

懒惰的预测大脑

8.1 表面张力

“快速、廉价且失控”是埃罗尔·莫里斯1997年纪录片的名称，其中一部分致力于当时相当新的基于行为的机器人学学科的工作。这部电影的名字取自罗德尼·布鲁克斯和安妮塔·弗林1989年一篇著名论文的标题，该论文回顾了这一领域工作的许多新兴原则：旨在解决移动自主（或半自主）机器人面临的棘手问题的工作。这项新工作最引人注目的是它与许多关于适应性反应内在根源的根深蒂固假设的根本背离。特别是，布鲁克斯和其他人正在攻击可能被称为”符号化、模型重型”的方法，在这种方法中，成功的行为依赖于获取和部署关于操作环境性质的大量符号编码知识。相反，布鲁克斯的机器人使用了许多更简单的技巧、策略和计谋，它们的综合效果（在它们要操作的环境中）是支持快速、鲁棒、计算成本低廉的在线反应形式。

Brooks方法的最极端版本在本质上存在局限性，并且（也许并不令人意外）在面对真正复杂的多维问题空间时无法很好地扩展（相关讨论见Pfeifer & Scheier, 1999，第7章）。尽管如此，Brooks的工作是一系列极具生产力和重要性的工作浪潮的先锋，这些工作探讨了智能体可能利用自身身体形态、行动以及环境本身持续可操作结构所提供的众多机会的多种方式（见Clark, 1997; Clark, 2008; Pfeifer & Bongard, 2008）。

这产生了某种困惑。因为乍看之下，层次预测处理的工作可能看起来相当不同——它似乎强调的是存储知识日益增长的多层次复杂性，*而非*大脑、身体和世界的精妙、机会主义的舞蹈。这样的诊断是极其错误的。之所以错误，是因为所提供的首先是关于高效、自组织的适应性成功路径的故事。此外，这是一个故事，其中那些高效路径可能——并且经常确实——涉及利用身体和世界的复杂行动和干预模式。从正确的角度来看，PP因此成为一个新的强大工具，用于对具身心智工作所赞颂的高效问题解决策略的普遍性和力量进行有组织的（且在神经计算上合理的）理解。引人注目的是，PP提供了一种系统性手段来*结合*那些快速、廉价的反应模式与更昂贵、费力的策略，揭示这些仅是自组织动力学连续统一体上的极端两极。作为一种愉快的副作用，关注PP嵌入和阐明具身反应全谱的多种方式也有助于暴露对预测驱动大脑这一总体愿景的一些常见担忧（听起来不祥的”黑暗房间”反对意见）中的根本缺陷。

8.2 生产性懒惰

具身的、环境情境化心智工作中的一个反复出现的主题是”生产性懒惰”的价值。我将这个短语归功于Aaron Sloman，但这个一般想法至少可以追溯到Herbert Simon’s (1956)对经济而有效的策略和启发式的探索：这些问题解决方案在任何绝对意义上都不是最优的，也不能保证在所有条件下都有效，但”足够好”以满足需求，同时尊重时间和处理能力的限制。例如，我们很可能会选择一位值得信赖的朋友昨天提到的餐厅，而不是试图全面检查五英里半径内每家餐厅的评论和菜单。我们这样做时合理地相信它会足够好，从而节省了考虑进一步信息的时间和精力成本。

适应性有机体作为”满足者”而非绝对优化者的相关概念导致了”有界理性”(bounded rationality)领域的重要工作（Gigerenzer & Selton, 2002; Gigerenzer et al., 1999），探索简单启发式的意外效力，这些启发式有时可能会误导我们，但也能使用最少的处理资源快速做出判断。简单启发式在许多人类判断和反应的产生中的确定作用也在大量工作中得到了充分证明，这些工作显示了刻板场景和相关偏见在人类推理中有时具有扭曲作用（例如，Tversky & Kahneman, 1973，以及一个很好的综合性论述，Kahneman, 2011）。尽管如此，我们人类显然也能够进行更慢、更仔细的推理模式，至少在有限的时间内可以避免一些错误。为了适应这一点，一些理论家（例如Stanovich & West, 2000）提出了”双系统”观点，假设存在两种不同的认知模式，一种（“系统1”）与快速、自动、“习惯性”反应相关，另一种（“系统2”）与缓慢、费力、深思熟虑的推理相关。正如我们将看到的，PP视角提供了一种灵活的手段，在单一的总体处理机制内容纳这种多重模式和快速启发式策略的情境依赖使用。

8.3 生态平衡与棒球

然而，快速的、启发式引导的推理策略只是”生产性懒惰”丰富拼图的一部分。另一部分（我之前在这个领域的大部分工作的焦点，见Clark, 1997, 2008）涉及可以被认为是感知的生态高效使用，以及大脑、身体和世界之间的劳动分工。例如，正如Sloman (2013)指出的，在某些情况下，通过开放门的最佳方式是依靠简单的伺服控制或碰撞和转向机制。

或者考虑双足行走的任务。一些双足机器人（本田公司的旗舰产品”阿西莫”可能是最著名的例子）通过非常精确且耗能的关节角度控制系统来行走。相比之下，生物行走体最大限度地利用了整个肌肉骨骼系统和行走装置本身所具有的质量特性和生物力学耦合。因此，自然界的双足行走者广泛使用所谓的”被动动力学”(passive dynamics)，即物理装置本身固有的运动学和组织结构(McGeer, 1990)。正是这种被动动力学使得一些相当简单的玩具能够在没有任何板载电源的情况下，流畅地沿着缓坡行走。这些玩具具有最小的驱动装置且没有控制系统。它们的行走不是复杂关节运动规划和驱动的结果，而是其基本形态（身体形状、连接分布和组件重量等）的结果。正如Collins等人(2001，第608页)所巧妙指出的，行走因此是”有腿机制的自然运动，就像摆动是钟摆的自然运动一样”。

被动行走器（以及它们优雅的动力对应物，参见Collins等人，2001）符合Pfeifer和Bongard(2006)所描述的”生态平衡原则”(Principle of Ecological Balance)。该原则指出：

首先……给定某个任务环境，智能体的感觉、运动和神经系统的复杂性之间必须匹配……其次……形态、材料、控制和环境之间存在一定的平衡或任务分配。(Pfeifer & Bongard, 2006，第123页)

这一原则反映了当代机器人学的重要教训之一，即形态（可以包括传感器布置、身体结构，甚至基本建筑材料的选择等）和控制的协同进化为在大脑、身体和世界之间分散问题解决负荷提供了黄金机会。机器人学因此重新发现了J. J. Gibson和”生态心理学”(ecological psychology)持续传统中明确表达的许多思想（参见Gibson, 1979; Turvey & Carello, 1986; Warren, 2006）。因此，William Warren在评论Gibson(1979)的一段引言时建议：

生物学利用整个系统的规律性作为行为排序的手段。具体而言，环境的结构和物理学、身体的生物力学、关于智能体-环境系统状态的感知信息，以及任务的要求都有助于约束行为结果。(Warren, 2006，第358页)

另一个吉布森式主题涉及感知在行动中的作用。根据一个熟悉的（更古典的）观点，感知的作用是将解决问题所需的尽可能多的信息输入系统。这些是我们在第6章中遇到的”重构性”方法。例如，一个规划智能体可能扫描环境以建立一个关于外界事物及其位置的问题充分模型，此时推理引擎可以有效地抛弃世界，转而在内部模型上操作，规划然后执行响应（可能在执行过程中偶尔检查以确保没有发生变化）。替代方法（参见，例如，Beer, 2000, 2003; Chemero, 2009; Gibson, 1979; Lee & Reddish, 1981; Warren, 2005）将感知描述为产生性地耦合智能体和环境的通道，在可能的情况下避免将源于世界的信号转换为外部场景的持久内部模型的需要。

因此再次考虑第6章中描述的”外野手问题”。这是在棒球中跑动以接住”高飞球”的问题。给予感知其标准作用，我们可能已经假设视觉系统的工作是传导关于球当前位置的信息，以允许一个独特的”推理系统”预测其未来轨迹。然而，自然似乎找到了一个更优雅和高效的解决方案。该解决方案（其一个版本最初由Chapman(1968)提出）涉及以一种似乎保持球在视野中以恒定速度移动的方式跑动。只要外野手自己的运动抵消了球的光学加速度的任何明显变化，她就会到达球将要撞击地面的位置。这个解决方案，光学加速度抵消(Optical Acceleration Cancellation, OAC)，解释了为什么外野手在被要求站着不动并简单预测球将落在哪里时，通常表现相当糟糕。他们无法预测着陆点，因为OAC是一种通过关键涉及智能体自己运动的逐时刻自我修正来工作的策略。我们依赖这种策略的建议也得到了一些有趣的虚拟现实实验的证实，在这些实验中，球的轨迹在飞行中突然改变，以在现实世界中不可能发生的方式（参见Fink, Foo, & Warren, 2009）。OAC是快速、经济的问题解决的一个很好的例子。巧妙地使用光流中免费可得的数据使接球者能够避免部署丰富的内部模型来计算球的前向轨迹的需要。

这些策略（正如我们在第6章中所提到的，另见Maturana, 1980）暗示了感知耦合本身扮演着一个相当不同的角色。它们不是使用感知来获得足够的信息进入内部，越过视觉瓶颈，从而让推理系统”抛弃世界”并完全在内部解决问题，而是将传感器用作*一个开放的管道，允许环境量值对行为施加持续影响*。因此，感知被描述为开放一个通道，当这个通道中的活动保持在一定范围内时，就会出现成功的整体系统行为。在这种情况下，正如兰德尔·比尔所说，“焦点从准确表征环境转向持续与环境互动，借助身体来稳定适当的协调行为模式”（Beer, 2000，第97页）。

最后，具身智能体还能够以主动生成在认知和计算上有效的时间锁定感官刺激模式的方式作用于它们的世界。例如（更详细的讨论，见Clark, 2008），Fitzpatrick et al. (2003)（另见Metta & Fitzpatrick, 2003）展示了主动物体操作（推动和触摸视野中的物体）如何帮助生成关于物体边界的信息。他们的”婴儿机器人”通过戳刺和推动来学习边界。它使用运动检测来看到自己的手/臂在移动，但当手遇到（并推动）一个物体时，会突然出现运动活动的扩散。这种简单的特征从环境的其余部分中识别出物体。在人类婴儿中，抓握、戳刺、拉拽、吮吸和推动创造了丰富的时间锁定*多模态*感官刺激流。这种多模态输入流已被证明（Lungarella & Sporns, 2005）有助于类别学习和概念形成。所有这些能力的关键是机器人或婴儿维持与环境协调感觉运动互动的能力。这种工作表明，自生成的运动活动作为”神经信息处理的补充”（Lungarella & Sporns, 2005，第25页），因为：

智能体的控制架构（例如神经系统）关注并处理感官刺激流，最终生成运动动作序列，这些序列反过来指导感官信息的进一步产生和选择。[通过这种方式]“运动活动的信息结构化”和”神经系统的信息处理”通过感觉运动回路持续相互联系。（Lungarella & Sporns, 2005，第25页）

因此，机器人学和人工生命的一个主要研究方向强调问题解决负载在大脑、主动身体和局部环境可操作结构之间分布的重要性。这种分布使得”生产性懒惰”的大脑能够尽可能少地工作，同时仍然解决（或者更确切地说，当整个具身的、环境定位的系统解决）问题。

[8.4 具身流动]

具身认知的研究也质疑了存在一个顺序处理流程的观念，该流程的各个阶段整齐地对应于感知、思考和行动。当我们在日常行为中与世界互动时，我们通常不是首先被动地接收大量信息，然后制定完整计划，最后通过一系列运动

命令来实施计划。相反，感知、思考和行动相互配合、重叠，并开始融合，作为整个感知-运动系统与世界互动。

这种融合和交织的例子包括交互式视觉研究(Churchland et al., 1994)、动态场理论(Thelen et al., 2001)，以及”指示指针”(Ballard et al., 1997)(相关综述见Clark, 1997, 2008)。作为例证，我们来看看Ballard et al. (1997)研究的任务。在这个任务中，受试者被给定一个彩色积木的模型图案，要求通过从储备区域逐个移动相似的积木到新的工作区域来复制该图案。任务通过鼠标拖拽在显示器上完成，在执行过程中，眼动追踪技术精确监测你在解决问题时的注视位置和时间。Ballard等人发现，受试者*没有*做的是：看一眼目标，决定下一个要添加的积木的颜色和位置，然后通过从储备区域移动积木来执行他们的小计划。相反，在执行任务过程中使用了对模型的反复快速扫视——比你预期的扫视次数多得多。例如，在拿起积木的*前后都*会查看模型，这表明当瞥一眼模型时，受试者只存储一小片信息：要么是下一个要复制的积木的颜色，要么是位置，但不会同时存储两者。即使对同一位置进行反复扫视，看起来也只保留了极少的信息。相反，反复注视似乎是在”适时”提供特定的信息项以供使用。[4]对物理模型的反复扫视让受试者能够部署Ballard等人称之为”最小记忆策略”来解决问题。这个想法是，大脑创建程序时会最小化所需的工作记忆量，眼动在这里被招募来将新的信息片段放入记忆中。通过改变任务需求，Ballard等人还能够系统性地改变生物记忆和主动具身检索的特定组合，以解决问题的不同版本，得出结论：在这个任务中”眼动、头动和记忆负荷以灵活的方式相互权衡”(第732页)。这是具身认知的另一个现在已经熟悉(但仍然重要)的教训。这里的眼动让受试者能够在适当的时候将外部世界本身用作一种存储缓冲区(关于这类策略的更多内容，见Clark, 2008; Wilson, 2004)。

综合所有这些已经表明了一个更加整合的感知、认知和行动模型。感知在这里与行动的可能性纠缠在一起，并持续受到认知、情境和运动因素的影响。这也是之前Pfeifer et al.’s (2007)“信息流的自我结构化”概念所暗示的图景(8.3)。行动服务于”适时”传递信息片段以供使用，而这些信息指导行动，形成一个持续的循环因果拥抱。如此理解的感知不必产生一个丰富、详细、行动中性的内在模型来等待”中央认知”的服务以推断适当的行动。事实上，这些区分(感知、认知和行动之间的)现在似乎模糊了而不是阐明了真正的效应流。从某种意义上说，大脑被揭示为不是(主要)一个推理引擎或安静思考的器官，而是一个环境情境化行动控制的器官。廉价、快速、世界开发的行动，而不是对真理、最优性或演绎推理的追求，现在是关键的组织原则。具身的、情境化的智能体，所有这些都表明，是”软装配”的大师，构建、解散和重建临时组合体，利用任何可用的东西，创造轻松跨越大脑、身体和世界的流动问题解决整体。

[8.5 节俭的行动导向预测机器]

表面上，这些”来自具身的教训”似乎指向与预测驱动处理工作相当不同的方向。预测驱动处理通常被描述为结合证据(感觉输入)、先验知识(产生预测的生成模型)和不确定性估计(通过对预测误差的精度权重)来生成对世界如何的多尺度最佳猜测。但正如我们之前提到的，这是微妙的误导。因为真实世界的预测完全是关于选择和控制与世界接触的行动。只要这些智能体确实试图”猜测世界”，这种猜测总是而且处处以适合支持行动和干预循环的方式进行反映。在最基本的层面上，这仅仅是因为整个装置(基于预测的处理)存在只是为了帮助动物实现它们的目标，同时避免与世界的致命惊人遭遇。我们可以说，行动是预测橡胶与适应道路相遇的地方。一旦我们考虑预测在行动的产生和展开中的作用，图景就会发生戏剧性的改变。

新图景的轮廓在我们之前对Cisek的可供性竞争假设的讨论（6.5）中已经显现。预测性处理被证明实现了”可供性竞争”的强版本，其中大脑持续计算多个概率加权的行动可能性，并使用一种架构，在这种架构中感知、规划和行动持续交织，由高度重叠的资源支持，并使用相同的基本计算策略执行。预测性处理在这里导致了Cisek和Kalaska (2011)所称的”实用性”表征的创建和部署：这些表征专门用于产生良好的在线控制，而不是旨在丰富地镜像一个独立于行动的世界。这些表征同时服务于认识功能，以旨在测试我们假设并为行动控制本身产生更好信息的方式采样世界。结果是一幅神经处理从根本上面向行动的图景，将世界表征为一个不断演化的并行、部分计算的行动和干预可能性矩阵。因此，前一节中呈现的具身流动图景几乎逐点地与面向行动的预测性处理工作相呼应。

然而，为了完成这种和解，我们必须利用最后一个要素。这个要素是使用预测误差最小化和可变精度加权来塑造大脑内连接模式的能力，在各种时间尺度上选择能够可靠驱动目标行为的最简单回路。这也是我们之前遇到过的特征（见第5章）。但正如我们现在将要看到的，由此产生的回路巧妙地包含了Gibson、Beer、Warren等人建议的简单、节俭的”为耦合而感知”解决方案。更好的是，它们在一个流动的、可重新配置的内在经济的更大背景下容纳了这些”模型稀疏”的解决方案，在这种经济中，基于丰富知识的策略和快速、节俭的解决方案仅仅是单一尺度上的不同点。这些点反映了不同内外资源集合的招募：这些集合以由外部环境、当前需求和身体状态以及对我们自身不确定性的持续估计决定的方式形成和消解。这种招募过程本身不断地被调节，借助感知-运动反应的循环因果舞蹈，通过外部环境的演化状态。在那一点上（我将论证），来自具身性和情境性、世界利用行动工作的所有关键洞察都通过面向行动的预测性处理的独特装置得到了充分实现。

[8.6 混合匹配策略选择]

要在实践中看到这如何工作，从一个不同（但实际上相当密切相关）的文献中的一些例子开始会有所帮助。这就是关于选择和决策制定的大量文献。在那个文献中，通常区分”基于模型”和”无模型”的方法（例如，见Dayan, 2012; Dayan & Daw, 2008; Wolpert, Doya, & Kawato, 2003）。基于模型的策略如其名称所示，依赖于一个包含关于各种状态（世界情境）如何连接的信息的领域模型，从而允许对假定行动的价值进行某种有原则的估计（给定某个成本函数）。这种方法涉及获取和部署（在计算上具有挑战性）关于任务领域结构的相当丰富的信息体。相比之下，无模型策略据说”直接通过试错学习行动价值，而不构建环境的显式模型，因此不保留对状态转换概率的显式估计”（Gläscher等，2010，第585页）。这种方法实现预先计算的”策略”，将行动直接与奖励联系起来，通常利用简单的线索和规律性，同时仍然提供流畅、通常快速的反应。

无模型学习与选择和行动自动控制的”习惯性”系统有关，其神经基础包括中脑多巴胺系统及其对纹状体的投射，而基于模型的学习则更密切地与皮质（顶叶和额叶）区域的作用相关（见Gläscher等，2010）。这些系统中的学习被认为是由不同形式的预测误差信号驱动的——对无模型情况而言是情感显著的”奖励预测误差”（例如，见Hollerman & Schultz, 1998; Montague等，1996; Schultz, 1999; Schultz等，1997），对基于模型的情况而言是更情感中性的”状态预测误差”（例如，在腹内侧前额叶皮质中）。然而，这些相对粗糙的区别现在正在让位于一个更加整合的故事（例如，见Daw等，2011; Gershman & Daw, 2012），正如我们将要看到的。

我们应该如何理解PP与这种”无模型”学习之间的关系？一个有趣的可能性是，机载的可靠性估计过程可能会根据上下文选择策略。如果我们假设存在多个竞争的神经资源能够解决当前问题，就需要某种机制在它们之间进行仲裁。考虑到这一点，Daw et al. (2005)描述了一个广义贝叶斯”仲裁原则”，通过估计与不同”神经控制器”（例如，“基于模型”与”无模型”控制器）相关的相对不确定性，允许在当前情况下最准确的控制器来决定行动和选择。在PP框架内，这将使用熟悉的精度估计和精度加权机制来实现。每个资源都会计算一个行动方案，但只有最可靠的资源（在当前上下文中部署时不确定性最小的资源）才能确定驱动行动和选择所需的高精度预测误差。换句话说，一种元模型（富含精度期望的模型）将被用来确定和部署在当前情况下最佳的资源，在需要时在它们之间切换。

然而，这样的故事几乎肯定过于简化了。诚然，“基于模型/无模型”的区别是直观的，与习惯和理性之间、情感和分析评估之间的旧有（但日益失信的）二分法产生共鸣。但似乎并行的、功能独立的神经子系统这一概念可能经不起时间的考验。例如，最近的一项fMRI研究(Daw, Gershman, et al., 2011)表明，与其考虑不同的（功能隔离的）基于模型和无模型学习系统，我们可能需要假设一个单一的”更集成的计算架构”（第1204页），其中最常与基于模型和无模型学习相关的不同大脑区域（分别是前额皮质和背外侧纹状体）*各自*都进行无模型和基于模型的评估，并且以”与决定选择行为的比例相匹配”（第1209页）的比例进行。从PP视角思考这一点的一种方式是，将”无模型”反应与由感觉流主导的（“自下而上”）处理联系起来，而”基于模型”反应则是那些涉及更大和更广泛的”自上而下”影响的反应。[5]通过调整预测误差的精度加权来实现这两种信息源之间的上下文依赖平衡，然后允许任务和环境所要求的任何策略组合。

这种更集成的内在经济概念得到了一项决策任务的支持(Daw, Gershman et al., 2011)，在该任务中，实验者能够区分明显基于模型和明显无模型的对后续选择和行动的影响。这是可能的，因为无模型反应本质上是向后看的，将特定行动与之前遇到的奖励联系起来。仅表现出无模型反应的动物在这个意义上注定要重复过去，在环境要求时释放先前强化的行动。相比之下，基于模型的系统能够使用（如其名称所示）某种外部竞技场的内在替代物来评估潜在行动，在这个竞技场中将执行行动并做出选择——例如，这样的系统可能部署心理模拟来确定一个行动是否比另一个更可取。因此，部署基于模型系统的动物能够，用Seligman et al. (2013)的话来说，“导航到未来”而不是”被过去驱动”。

现在似乎很清楚，大多数动物都能够采用这两种形式的反应，并将密集的习惯性支持网络与偶发的真正前瞻性爆发相结合。根据标准图景，回想一下，存在着不同的神经价值评估系统和不同形式的预测误差信号来支持每种类型的学习和反应。使用顺序选择任务，Daw等人能够创造出这样的条件：其中一个或另一个神经价值评估系统的计算应该与行为分离，揭示出（在不同的、先前已识别的大脑区域中）无模型和基于模型系统对价值的独立计算存在。然而，他们发现在两个区域中都存在明显无模型和明显基于模型反应的神经相关性。引人注目的是，这意味着即使是纹状体计算的”奖励预测误差”也不能简单地反映使用真正无模型系统的学习。相反，纹状体中记录的活动”反映了无模型和基于模型评估的混合”（Daw et al., 2011，第1209页），“即使是与无模型强化学习最相关的信号，纹状体RPE（奖励预测误差），也反映了两种类型的评估，以与它们对选择行为的观察贡献相匹配的方式结合”（Daw et al., 2011，第1210页）。自上而下的信息，Daw et al. (2011)建议，可能在这里控制不同策略在不同情境中为行动和选择而结合的方式。基于模型和无模型评估之间更大的整合，他们推测，也可能来自某种混合学习例程的作用，其中基于模型的资源可能训练和调整（更快的、在情境中更高效的）无模型资源的反应。[6]

在更一般的层面上，这样的结果增加了日益增长的文献（综述见Gershman & Daw, 2012），该文献表明需要对标准决策理论模型进行深度重构。该模型假设效用和概率的不同表征，与或多或少独立的神经子系统的活动相关联，我们实际上可能面对的是一个更深度整合的架构，其中”感知、行动和效用被纠缠在一个复杂的网络中[涉及]感知和动机系统之间更丰富的动态交互集合”（Gershman & Daw, 2012，第308页）。本节中探索的更大图景在功能上是有意义的，允许”无模型”模式使用基于模型的方案来教授它们如何反应。在PP框架内，这导致（浅层）“无模型”反应在（更深层）基于模型的经济体中的层次嵌入。这有许多优势，因为基于模型的方案（上述第5章）具有深度的情境敏感性，而无模型或习惯性方案——一旦建立——就是固定的，绑定于先前成功行动情境的细节。通过在一个总体经济中精确地结合这两种模式，适应性智能体可以识别部署无模型（“习惯性”）方案的适当情境。如果这是正确的，“基于模型”和”无模型”的评估和反应模式，仅仅是沿着单一连续体的极端，并且可能以由手头任务决定的许多混合和组合出现。

8.7 平衡准确性和复杂性

我们现在可以将这些洞察置于一个更大的概率框架内。

Fitzgerald, Dolan, and Friston (2014，第1页)指出，“贝叶斯最优智能体既寻求最大化其预测的*准确性*，也寻求最小化它们用来生成这些预测的模型的*复杂性*”。最大化准确性对应于最大化模型对观察数据的预测能力。另一方面，最小化复杂性需要在与执行手头任务一致的情况下尽可能减少计算成本。形式上，这可以通过结合一个复杂性惩罚因子来实现——有时称为奥卡姆因子(Occam factor)，以十三世纪哲学家威廉·奥卡姆命名，他著名地告诫我们不要”超出必要性地增加实体”。总体”模型证据”然后是一种反映精确（和情境可变）准确性/复杂性权衡的复合量。Fitzgerald, Dolan, and Friston继续概述了一个具体方案（涉及”贝叶斯模型平均”[7]），其中”模型根据其证据[即，如刚才定义的其总体模型证据]而不是简单地根据其准确性进行加权或选择”（Fitzgerald et al., 2014，第7页）。在PP内，选择预测误差的精度加权变化提供了一种能够实现这种任务和情境敏感的不同模型之间竞争的机制，而突触修剪（见3.9和9.3）在更长的时间尺度上服务于复杂性减少。

所有这些都表明，对于广受欢迎的建议(8.2)可能需要重新思考，该建议认为人类推理涉及两个功能不同的系统的运作，一个用于快速、自动、“习惯性”反应，另一个专门用于缓慢、费力、深思熟虑的推理。与其说是真正二分的内在组织，我们可能会受益于一种更丰富的组织形式，其中快速、习惯性或基于启发式的反应模式通常是默认的，但在其中可能有大量的可能策略可供使用。这些策略之间的平衡则由可变的精度权重决定，因此（实际上）由各种形式的内源性和外源性注意力决定(Carrasco, 2011)。因此，人类和其他动物会部署多种——丰富的、节俭的，以及介于两者之间的所有策略——这些策略在根本统一的神经资源网络中得到定义（关于这种更加集成空间的一些初步探索，见Pezzulo et al., 2013，以及8.8）。

最后，即使在单个更加集成的系统内，可能发生的策略嵌套的复杂性也没有固定的限制。例如，我们可能使用某种快速粗糙的启发式策略来识别使用更丰富策略的情境，或者使用密集的模型探索策略来识别使用更简单策略就足够的情境。最高效的策略就是使整体复杂性成本最小化的（主动）推理。从这个新兴的有利位置来看，基于模型和无模型反应之间的区别（实际上系统1和系统2之间的区别，只要这些被视为不同的系统而不是模式[8]）看起来越来越浅显。现在这些只是对资源和影响力的不同混合的便利标签，每一种都以相同的一般方式根据环境需要被招募。

8.8 回到棒球

现在让我们回到前面描述的外野手问题。在这里，如果PP是正确的，已经活跃的神经预测和简单、快速处理的感知线索也必须共同工作，以确定不同预测误差信号的精度权重模式。这创造了（回顾第5章）一个瞬态的有效连接网络（一个临时的分布式回路），并且在该回路内，它设定了自上而下和自下而上影响模式之间的平衡。然而，在手头的情况下，效率要求选择一个回路，其中感知发挥第6章和8.3中描述的非重构性作用。在这种情况下，视觉感知的临时任务变成抵消飞球的光学加速度。这意味着对与抵消球的光学投影的垂直加速度相关的预测误差给予高权重，并且（直白地说）不太关注其他任何事情。

适当的精度权重因此选择了一个预先学习的、快速的、低成本的策略来解决问题。情境招募的精度权重模式在这里完成了一种集合选择或策略切换的形式。[9]这假设较慢的学习和适应性可塑性过程已经以使低成本策略可用的方式塑造了神经连接模式。但这是没有问题的。它可以通过最小化复杂性的驱动力在一般意义上得到激励（在合理的约束下，这与”满意化”的驱动力无法区分）。因此，所需的学习可以使用在许多时间尺度上运作的预测误差最小化来完成。这样的过程范围从儿童棒球运动员的缓慢学习，到职业运动员在比赛中考虑风况变化和对方击球手打法的更快在线适应。

结果是一个复杂但有回报的图景，其中预测学习的基础过程缓慢地安装包含精度期望的模型，允许有效连接模式”即时”构建和重建。这使得知识稀疏的快速反应模式能够根据当前情境被招募和调整。“有生产性地懒惰”和基于模型的方法的这种兼容性应该不足为奇。要看到这一点，我们只需要反思，支撑任何给定行为的模型或模型片段可以是一个简单的、易于计算的启发式（一个简化的”经验法则”），就像具有更复杂因果结构的东西一样容易。这种低成本模型在许多情况下将依赖于行动，利用循环因果商业模式（在感知输入和运动行动之间）来”及时”提供任务相关信息以供使用。

快速、自动、过度学习的行为特别适合由采用更启发式形式的模型控制。反映情境的精度分配的作用是选择和启用已被证明能够支持目标行为的低成本程序模型。这种低成本模型——OAC是一个很好的例子——在许多情况下将依赖于我们自己信息流的自我结构化，利用循环因果商业模式（在感知输入和运动行动之间）来”及时”提供任务相关信息以供使用。

更复杂的策略（直觉上更加”模型丰富”，尽管这现在只是连续体上的另一个位置）也可能涉及简化和近似。一个很好的例子是Battaglia等人（2013）关于”直觉物理学”的工作。人类智能体能够对普通物体的物理行为做出快速推理。这些推理可能包括发现书堆或洗涤用品堆叠不稳定并有倾倒风险，或者轻轻碰触的物体将要掉落并撞击其他物

体。Battaglia等人认为，这种能力的基础可能是一个概率场景模拟器（一个概率生成模型），能够基于部分、噪声信息快速做出判断。这种模拟器不依赖于命题规则，而是依赖于”物体几何形状、运动和力动力学的定量方面和不确定性”（第18327页）。Battaglia等人描述并测试了这样一个模型，显示它符合许多不同心理物理任务的数据。重要的是，Battaglia等人的模型通过使用对真实物体行为的近似来模拟物理世界，从而提供了鲁棒性和速度。通过这种方式，它”用精确性和真实性换取速度、通用性，以及做出对日常活动目的足够好的预测的能力”（Battaglia等人，2013，第18328页）。

这里的”直觉物理引擎”——或者任何其他名称的生成模型——产生简化的概率模拟，但仍然能够预测物理世界起伏变化的关键方面。这样的”直觉物理引擎”能够推断关于图8.1所示场景中物体可能行为的关键事实——例如在图1C中，哪个物体会首先掉落，朝什么方向，以及会产生什么样的连锁效应。对近似和不确定性估计的依赖也解释了错觉的存在（如图8.1F中的稳定性错觉）和关于物理世界推理中的错误。也就是说，我们日常的近似可能不容易”理解”使岩石塔保持稳定的精细平衡结构。一个能够做到这一点的模型在某些情况下会更准确，但会有一些时间成本（所以也许我们不会及时发现洗涤用品堆的不稳定性来防止灾难性的状态转换）。

image

图8.1 一些唤起物理直觉的场景

唤起强烈物理直觉的日常场景、活动和艺术。（a）一个杂乱的工作坊，展示了许多细致的物理属性。（b）场景A的基于3D物体的表示，可以支持基于模拟的物理推理。（c）一堆摇摇欲坠的盘子看起来像是等待发生的事故。（d）一个孩子通过堆叠积木锻炼他的物理推理。（e）积木游戏考验玩家的物理直觉。（f）“石头平衡”利用了我们强大的物理期望（照片和石头平衡由Heiko Brinkmann提供）。

来源：Battaglia等人，2013。

这些近似解决方案反映了Gershman和Daw（2012，第307页）所描述的一种”相对于其益处，维持[一个]完整表示的成本（例如，额外计算）的元优化”。Gershman和Daw（2012，第308页）认为，对感知、行动和效用的神经交织的最深层解释可能就在那里，在适应性压力中寻找和部署”将密度集中在高效用区域”的表示形式和统计近似。结果是一种元贝叶斯(meta-Bayesian)确定表示什么、何时表示以及如何表示的方法。因此，普遍粗暴优化的不合理含义被抛弃，转而采用提供效力、可靠性和能量效率某种组合的策略。这样的智能体将使用足够好完成工作的最有效策略，并且该策略在可重构流内的软组装中当前可用。

因此，处理复杂的时间压力世界需要使用许多策略，从非常简单的启发式到更复杂的相互作用近似结构。然而，这种多样化的景观可能构成基于不确定性的认知生态系统的一部分——在这个生态系统内，这些许多策略出现、消失和相互作用。在PP框架内，许多不同类型的策略可能通过不断变化的精度估计在每一刻被选择。这种估计改变了有效连接性的模式，使不同的内部（和外部，见下文）回路网络能够在不同时间控制行为。[10]

[8.9 扩展预测心智]

所有这些都暗示了一个非常自然的”扩展认知”模型（Clark，2008；Clark & Chalmers，1998），这里这简单地指生物外部结构和操作有时可能构成智能体认知例程的组成部分。据我所知，PP框架在实质上没有改变之前关于真正扩展认知系统可能性提出的支持和反对论证。[11]然而，PP确实提供了一个具体的、高度”扩展友好”的关于专门神经对认知成功贡献形状的提议。

要理解这一点，需要反思已知的外部（例如环境）操作通过部分构成提供了适合前几节描述的”基于元模型”选择的额外策略。这是因为参与和利用特定外部资源的行动现在将以与内部神经资源联盟本身相同的方式被选择。例如，在执行积木放置任务（Ballard et al., 1997）（在8.4节中描述）时，大脑必须对支撑各种行动的预测分配高精度，这些行动让我们在执行任务时能够”将世界用作其自身最佳模型”。这种与世界互动的行动反过来又由获得的估计决定，即可靠的、显著的（任务相关的）信息在某个位置和某个时间是可获得的。或者考虑这样的情况：通过使用某些生物外部设备（如笔记本电脑或智能手机）可以获得显著的高精度信息。选择行动以减少预测误差的核心程序现在将选择调用生物外部资源的行动。调用生物外部资源，以及移动我们自己的效应器和传感器来产生高质量的任务相关信息，在这里都是同一潜在策略的表达，反映了我们大脑对可靠的任务相关信息在何处何时可获得的最佳估计。因此选择的策略通常正如Ballard等人所建议的，是最小内部记忆策略，其成功条件既需要有机体行动也需要外部环境的配合。这些策略再次强调了跨越大脑、身体和世界的分布式资源网络的重要性。

作为一个简单的例证，考虑Pezzulo, Rigoli, 和 Chersi (2013)的工作。在这里，一个所谓的”混合工具控制器”(Mixed Instrumental Controller)决定是基于一组简单的、预计算的（“缓存的”）值来选择行动，还是通过运行心理模拟(mental simulation)来实现对实际执行该行动的可取性或其他方面的更灵活的、基于模型的评估。混合控制器计算”信息价值”(value of information)，只有当该价值足够高时才选择更有信息性（但成本更高）的基于模型的选项。在这些

情况下，心理模拟然后产生新的奖励预期，这些预期可以通过更新用于决定选择的价值来决定当前行动。我们可以将此视为一种机制，它逐时逐刻地决定（如前几节所讨论的）是利用简单的、已缓存的程序，还是使用某种形式的心理模拟来探索更丰富的可能性集合。很容易想象混合控制器的一个版本，它（基于过去的经验）确定它认为通过某种生物外部设备（如操作算盘、iPhone或物理模型）可获得的信息价值。因此，部署简单的缓存策略、更昂贵的心理模拟，或利用环境本身作为认知资源，都是适合使用PP装置进行上下文敏感招募的策略。

从这个角度来看，在以交互为主导的PP经济体中招募任务特定的内部神经联盟与招募任务特定的神经-身体-世界集合完全相当。扩展的（大脑-身体-世界）问题解决集合的形成和解散在这里遵循许多相同的基本规则和原则（平衡有效性和效率，反映对不确定性的复杂持续估计），就像招募由有效连接性结合的临时*内部*联盟一样。在每种情况下，被选择的都是一个临时的问题解决集合（一个”临时任务特定设备”，见Anderson, Richardson, & Chemero, 2012），作为上下文变化的不确定性估计的函数被招募。这只是我们在5.5节中描述的”瞬时组装的局部神经子系统”出现的具身的、环境嵌入的版本。

这些临时集合在我们描述为”具身流动”(embodied flow)的赋权上下文中（8.4节）出现和部署。在这些流动中，感知-运动程序提供新输入，招募新的瞬时资源集合。正是这些滚动循环最清楚地表征了野外的人类认知。在这些滚动循环中，任意复杂数量的”依赖世界”可能逐渐折叠进来，通过将工作从大脑卸载到（非神经的）身体，从有机体卸载到（物理的、社会的和技术的）世界，从而扩展我们的实际认知能力。PP使异常清楚的是，正是这些滚动循环是神经经济体持续（而不仅仅是在涉及心智扩展工具和技术的特殊情况下）服务的。随着这些循环展开，没有内在的小人监督所产生的分布式问题解决集合的反复软组装。相反，这些集合以由环境上下文中精确的、高质量的预测误差的渐进减少所决定的方式出现和消解。有机体显著的（高精度）预测误差因此可能是通过其在行动中的表达，将来自大脑、身体和世界的元素结合成临时问题解决整体的通用粘合剂。

[8.10 逃离黑暗房间]

预测误差最小化与适应性反应的极其广泛的策略是一致的。但在这种反应的丰富画卷中，有一个生动的线索似乎特别抗拒使用现有资源进行重构。这个生动的线索涉及游戏、探索和新奇事物的吸引力。有时人们担心，预测误差最小化的认知命令天生无法容纳这些现象，相反，它提供了一种寂静主义、故意认知削弱，甚至（也许）致命不活跃的处方！这种担忧认为，不幸的预测驱动有机体应该简单地寻找那些容易预测的状态，比如一个空荡荡的黑暗房间，在那里度过余生中日益饥饿、口渴和令人沮丧的日子。这就是所谓的”黑暗房间悖论”(Friston, Thornton, & Clark, 2012)。

这种担忧（尽管重要）在多个方面都是错误的。在最基本的生物学层面，它被呈现为对有机体完整性和持续性的威胁，这是错误的。在更加精细的”人类繁荣”层面，它被视为（见，例如，Froese & Ikegami, 2013）反对游戏、探索以及故意寻找新奇和新体验的行为，这也是错误的。在每一种情况下，解决这个悖论的方法是注意到进化和文化背景的重要作用，在这些背景下，逐时逐刻的预测误差最小化过程得以出现和展开。

基于预测误差的神经处理，正如我们所看到的，是多尺度自组织强有力配方的一部分。这种多尺度自组织并不是在真空中发生的。相反，它只能在进化的有机体（神经和整体身体）形式的背景下运作，以及（正如我们将在第9章中看到的）缓慢积累的物质结构和文化实践的同样变革性背景：一代又一代人类学习和经验的社会技术遗产。

为了开始将这个更大的图景聚焦，首先要注意的是，明确的、快时间尺度的预测误差最小化过程必须回应进化的、具身的和环境嵌入的代理人(agents)的需求和项目。这些代理人的存在本身（见Friston, 2011b, 2012c）因此已经暗示了巨大范围的结构隐含的生物特定”期望”。这些生物被构建来寻找配偶，避免饥饿和干渴，并且（即使在不饥饿或不干渴时）参与那种零星的环境探索，这将帮助它们为意外的环境变化、资源稀缺、新竞争者等做好准备。因此，在逐时逐刻的基础上，预测误差只有在这个复杂的生物定义”期望”集合的背景下才能被最小化。

引号标志着我认为在基于终生经验获得的期望与那些在某种程度上*结构隐含*的期望之间存在重要差异。我们被构建为通过肺部呼吸空气，因此我们体现了一种保持（主要）在水面上的结构”期望”——不像（比如说）章鱼。我们的一些行动倾向同样是内置的。触摸热盘子的反射反应是撤回。这种反射相当于一种避免组织损伤的基础”期望”。在这种减弱的意义上，每个具身代理人（甚至细菌）都已经是其环境的一种惊喜最小化模型，正如Friston (2012c)所声称的那样。因此我们读到：

生物系统可以从环境波动（如化学引诱剂或感官信号的浓度变化）中提取结构规律性，并将它们体现在其形式和内部动力学中。本质上，它们成为其局部环境中因果结构的模型，使它们能够预测接下来会发生什么，并对抗这些预测的令人惊讶的违反。(Friston, 2012c, p. 2101)

另一个简单的例子是我们物种特定的感官受体阵列及其在特定身体位置的放置所提供的信息。这（至少在夜视设备和其他感官增强设备发明之前）选择并限制了感官预测误差可以主动最小化的空间。但不仅仅如此。作为进化的生物，我们也”期望”（仍然带着我的引号）保持温暖、营养充足和健康，感知到与这些根深蒂固的规范的偏差将产生能够驱动短期反应和适应的预测误差。这样的代理人（当正常运作时）根本不会感受到黑暗房间的吸引力。这种生物定义的”期望”即使经过长期的生活经历（如忍受饥荒）也不会被修正。

面对暗房谜题时，首先要做的是从一个更加宏观的（长时间尺度）角度来看待事物。在这样的时间尺度上，任何形式的适应或变化都会减少”惊讶”（见1.7），其效果是帮助有机体抵抗解体和无序，在与环境的交换中最小化”自由能”（附录2）。在这些更长的时间尺度上考虑，我们可以说这相当于为它们提供了一种总体性的结构隐含”信念”或”期望”——仍然带着那些重要的引号——这些信念或期望调节和约束着我们时刻进行的显式预测误差最小化过程。相对于以这种方式定义进化智能体的完整”期望”集合，暗房通常[12]根本没有任何吸引力。正如Friston (2012c)所建议的，典型的进化智能体强烈”期望”不要在这种无回报的环境中待太长时间。这意味着暗房对我们这样的生物没有任何诱惑力。

然而，Friston表达这一重要事实的方式可能存在问题——因此我使用了所有这些引号。因为它有将各种最小化惊讶的方式混淆的威胁——例如，通过总体身体形态和神经解剖的细节，以及通过更显式的、基于生成模型的发出自上而下的概率预测来迎接传入的感官流。如果我的皮肤在割伤后愈合，说我以某种结构化、具身的方式”预测”了一个完整的膜是误导性的。然而，只有在这种牵强的意义上，例如，鱼的形状才能被说成体现了对海水流体动力学的期望。[13]也许我们应该允许在某种非常宽泛的意义上，鱼类的”惊讶”确实部分由这种形态学因素决定。然而，我们的关注点一直是由神经编码的生成模型发出的相互交织的预测套件——如果PP是正确的，这种过程是通过神经元处理的不对称双向级联中预测和预测误差信号的迭代交换来实现的。因此，我认为我们不应该恰当地（不带引号地）说所有这些基础适应状态和反应本身就相当于结构沉积的（Friston说是”具身的”）预测或期望。我认为，更好的说法是，愈合（连同一系列确保生存和成功的其他神经和身体机制）*为我们对世界的预测模型的形成奠定了基础*。预测误差最小化在这里作为众多过程中的一个出现——但它（我已经论证过）在允许像我们这样的智能体在感知和行动中遇到一个相互作用的远端原因的结构化世界方面发挥着非常特殊的作用，而不是简单地（像植物或非常简单的生命形式那样）运行保持我们生存的例程。

因此，“生物定义背景”最好被理解为——至少对我们的目的而言——为部署（有时，在某些动物中）更显式的预测误差最小化学习和反应策略奠定基础。尽管如此，生物定义背景是极其重要的，它既影响我们（在丰富意义上）预测什么，也关键地影响*我们在那种完整意义上不需要*预测什么——因为，例如，它已经被基本的生物力学特征照顾到了，如被动动力学(passive dynamics)和肌肉和肌腱的内置协同作用。只有在那个极具赋能的背景下，预测误差的在线计算才能解释我们复杂、流畅的行为成功形式。

[8.11 游戏、新奇性和自组织不稳定性]

暗房谜题还有另一个——稍微更微妙的——维度，在以下段落中得到了很好的体现：

如果我们的主要目标是最小化我们遇到的状态和结果的惊讶，这如何解释复杂的人类行为，如寻求新奇、探索，以及更高层次的愿望，如艺术、音乐、诗歌或幽默？根据这个原则，我们不应该更喜欢生活在一个高度可预测和无刺激的环境中，在那里我们可以最小化我们的长期惊讶吗？我们不应该厌恶新奇刺激吗？就目前而言，这似乎是极不可信的；新奇刺激有时是厌恶的，但通常恰恰相反。这里的挑战是调和作为自组织行为基础的基本要求与我们避免单调环境并积极探索以寻求新奇和刺激输入这一事实。(Schwartenbeck et al., 2013, p. 2)

正是在这个方向上，Froese and Ikegami (2013)建议最小化惊讶的好方法将包括”刻板自我刺激、紧张性从世界退缩和自闭性从他人退缩”。这里的担忧不是（完全）我们会寻找某个暗房死亡陷阱。关于基础结构和原生”期望”的观察已经处理了这种担忧。相反，担忧是PP故事对于新奇性和探索的更*积极*吸引力似乎奇怪地沉默。也就是说，它对于”为什么我们积极渴望（在一定程度上）新奇、复杂的状态”(Schwartenbeck et al., 2013)似乎奇怪地沉默。因此，它对例如娱乐、艺术和文化的巨大产业保持沉默。

这是一个庞大且具有挑战性的话题，我无法在几句简短的评论中希望解决。但解决方案的一部分可能本身就涉及（以某种自举方式）文化中介的终身学习形式，这些学习安装了全局策略(global policies)，积极支持日益复杂的新奇寻求和探索形式。在这种意义上，全局策略只是一个相当一般的行动选择规则——一个包含整个行动种类而不是单个行为的规则。与游戏和探索相关的最简单的此类策略是一个会降低状态价值的策略，占据该状态的时间越长，其价值越低。在资源在空间和时间上分布不均的世界中，这可能是一个适应性有价值的策略。

给这种策略一个动力学色彩可能会很有用。随着我们在空间和时间中的轨迹展开，潜在稳定的停止点（用动力系统语言称为吸引子(attractors)）不断出现和消解，通常受到我们自身不断演化的内在状态和行动的影响。然而，一些系统有破坏自己不动点的倾向，主动诱导不稳定性，其方式导致Friston, Breakspear, & Deco (2012)所称的”漫游或流动（wandering）动力学”。这样的系统似乎”为了其自身目的”而追求变化和新奇性。

一个有趣的例证涉及可能是发展机器人学的第一个”都市传说”。根据这个故事，一个机器人被设置为在玩具环境中最小化预测误差。但在做到这一点后，机器人并没有简单地停止行为，而是开始不断旋转，创造出各种光学伪影(optical artefacts)，然后继续对其进行建模和预测。这个故事（我在与Meeden et al., 2009的工作相关的情况下听到的）结果并不完全真实，尽管它基于一些确实实际展示的有趣机器人行为。[14]但是一个真正倾向于破坏自己不动点的生物，发现自己被困在一个高度受限的环境中，确实可能被驱使通过任何可用手段寻求新的视野。类似地，Lauwereyns (2012)报告了一项研究[15]显示：

被限制在黑暗房间中，刺激极少的人类会按按钮让彩色光点图案出现，偏好那些提供最多变化和不可预测性的图案序列。(Berlyne, 1966, p. 32, 引用于Lauwereyns, 2012, p. 28)

最近，Kidd et al. (2012)对7-8个月大的婴儿进行了一系列实验，测量对不同（且良好控制的）复杂性事件序列的注意力。他们发现，婴儿的注意力具有他们称为”金发姑娘效应(Goldilocks Effect)“的特征，专注于呈现中等程度可预测性的事件——既不太容易预测，也不太难预测。因此，当复杂性（计算为负对数概率）非常高或非常低时，婴儿看向别处的概率最大。Kidd等人建议，功能性结果是”婴儿隐式地寻求维持中等的信息吸收率，避免在过于简单或过于复杂的事件上浪费认知资源”(Kidd et al., 2012, p. 1)。

这种寻求”刚好足够新奇”情况的倾向是某种形式天赋规格(innate specification)的良好候选，因为它们会导致主动智能体(active agents)以理想适合其环境信息生成模型增量获取和调整的方式自我构建信息流。更一般地说，栖息在复杂、变化世界中的智能体将从各种驱使它们探索这些世界的策略中获益良多，即使没有看到即时收益或奖励。这样的智能体会主动扰动它们自己在空间和时间中的轨迹，以强制执行一定量的探索。[16]由此产生的”流动”轨迹(Friston, 2010; Friston et al., 2009)为新的学习和发现提供了适应性有价值的门户。[17]

扩展这一观点，Schwartenbeck et al. (2013)建议某些智能体可能获得积极重视访问许多新状态机会的策略。对于这样的智能体，某个当前状态的价值部分由它允许它们访问的其他可能状态的数量决定。艺术、文学和科学的复杂人造环境看起来像是结构化来支持和鼓励这种开放式探索和新奇寻求形式的领域的好例子。沉浸在这些设计师环境中的预测智能体将学会期待（因此要求并主动寻求）那些特有的新奇性和变化类型。在构建我们的文化和社会世界时，我们可能因此正在以促进日益稀有的探索和新奇寻求模式的方式构建*我们自己*。这种增量文化自我脚手架(incremental cultural self-scaffolding)（人类逃离黑暗房间的高潮）是下一章也是最后一章的主题。

[8.12 快速、廉价，也很灵活]

我们生活在变化且充满挑战的世界中。这样的世界要求使用多种策略，包括快速、高效的感知-行动耦合模式和较慢、费力的推理和心理模拟过程。为了在这样的世界中保持领先，我们必须运用我们所知道的来预测并主动塑造感官冲击。在这样做时，我们不是简单地与世界互动。相反，我们逐时逐刻地选择我们将要使用的策略（神经和超神经回路和活动）。这些策略范围从快速粗糙到缓慢准确，从那些由自下而上感官流主导的到更依赖自上而下情境调节的，以及介于两者之间的所有点和混合。它们也从高度探索性到深度保守性，当获得信息和经验的价值开始被获取这些信息和经验所涉及的成本和风险所超越时，能够实现流畅的切换。这些策略切换平衡了预期的时间、能量和计算成本与可能的收益。

像我们这样的生物因此被构建为持续主动、生产性懒惰，偶尔探索性和好玩的。我们被构建为在最小化努力（智力和体力）的同时最大化成功。我们在很大程度上通过部署基本面向行动的策略来做到这一点。像我们这样的心智不是在以某种被动、描述性的方式表征世界。相反，它们以复杂的滚动循环方式与世界互动，其中行动决定感知，感知选择行动，沿途唤起和利用各种环境结构和机会。

因此，关于预测处理组织可能过分强调计算昂贵、表征繁重的策略而非其他（更快、更粗糙、更”具身化”）策略的担忧得到了充分和令人满意的解决。永远活跃的预测大脑现在被揭示为一个懒惰的大脑：一个对任何以更少做更多的机会保持警觉的大脑。

第9章

成为人类

9.1 将预测置于适当位置

我们的神经经济存在是为了服务具身行动的需求。我们看到，它通过启动和维持复杂的循环因果流来做到这一点，其中行动和感知是共同决定和共同决定的。这些循环因果流制定了结构耦合，使有机体保持在其自身的可行性窗口内。这样，永远活跃的预测大脑的愿景与具身和情境心智的工作优雅地结合。我们已经看到了这方面的证据，当我们探索联合感知和行动的循环因果网络、充分利用身体和世界的低成本策略的使用，以及感知、决策和行动的复杂连续交织时。然而，为了完成这幅图画，我们现在必须探索社会和环境结构的嵌套网络如何告知并被这些展开的具身神经预测过程所告知的许多方式。

核心是多时间尺度的自组织过程。预测误差最小化为自组织提供了一个合理而强大的机制——一个能够产生极大复杂性的嵌套动力学制度的机制。但这种复杂性，在人类代理的相当特殊情况下，现在涉及一个有力且易变的社会文化外壳。我们人类——在地球自然秩序中独一无二——构建并反复重建社会、语言和技术世界，其规律性然后反映在进行预测的生成模型中。正是因为大脑本身是如此有力的无监督自组织器官，我们的社会文化沉浸才能如此有效。但只有在这两个基本力量之间的许多复杂且理解不足的相互作用中（在复杂的自组织神经动力学和不断演变的社会和物质影响漩涡之间），像我们这样的心智才从物质流中涌现。因此，我们必须在其适当的环境中面对永远活跃的预测大脑——与赋权的物质、语言和社会文化脚手架背景不可分割地交织在一起。接下来是对这一重大且重要任务的初步姿态。

9.2 重述：围绕预测误差的自组织

预测误差提供了一个有机体可计算的量，适合在许多方式和许多时间尺度上驱动神经自组织。我们在前面的章节中多次看到这一原理的作用，但值得停下来欣赏由此产生的有力的自组织扫描。在这个过程的核心是一个概率生成模型，它逐步改变以更好地预测冲击生物有机体或人工代理的感官数据的游戏。这导致学习能够分离出在不同空间和时间尺度上运作的相互作用的身体和环境原因。这种方法描述了一个用于自组织、基础、结构学习的有力机制。学习现在是有基础的，因为远端原因仅作为预测感官数据游戏的手段而被发现（这种游戏也反映了有机体自身对世界的行动和干预）。这种学习是结构揭示的，发掘了在不同空间和时间尺度上运作的原因之间的复杂相互依赖模式。所有这些提供了一种预测例程的调色板，可以以新颖的方式组合来处理新情况。

这样的系统是*自组织的*，因为它们不以任何特定的输入-输出映射为目标。相反，它们必须发现级联规律性模式，以最好地适应它们自己的（部分自我诱导的）感觉信息流。这是令人解放的，因为这意味着这样的系统可以提供不依赖于特定任务执行的认知方式（尽管要适应的感觉数据游戏本身受到人类活动广泛形式的约束）。

这样的系统也对上下文深度敏感。这是因为任何区域或任何级别的系统响应现在都要对上下文固定信息的完整向下（和横向）级联负责。这种非线性动力学图景进一步增加了复杂性，因为影响流本身是可重新配置的，因为不断变化的精度估计改变了有效连接的时刻模式。

通过围绕预测误差进行自组织，这些架构因此提供了对世界中有机体相关特征的多尺度把握——这种把握的特征是能够在感知和行动共同作用以消除高精度预测误差的持续循环中与世界互动的能力。

9.3 效率和”上帝的先验”

需要注意的一个重要点是，这样的系统并不是简单地因为感觉预测误差成功最小化就停止改变和变化。因为正如我们在第8章中看到的，还有另一个（较少被强调的）因素仍然可以驱动变化和学习。这个因素就是效率。效率（参见，例如，Barlow, 1961; Olshausen & Field, 1996）直观上是冗余和过度的反面。如果一个方案或策略只使用完成工作所需的最少资源，那么它就是高效的。一个足够丰富的生成模型，能够维持有机体对选择行为重要的规律性的把握，但使用最少的能量或表征资源（例如，使用少量参数）来做到这一点，在这个意义上是高效的。相比之下，使用大量参数来适应或响应相同数据的系统并不因此成为其世界的”更准确”建模者。相反，结果往往是对观察到的数据”过拟合”，其中一些数据只是信息信号周围的”噪声”或随机波动。

第8章中描述的光学加速度抵消过程是一个很好的例子，说明了一个将低复杂性（少量参数）与高行为杠杆作用相结合的模型。在最一般的层面上，追求效率简单地是最小化感觉预测误差总和这一整体要求的组成部分。这涉及找到成功参与感觉流的最简约模型。因为我一直认为，预测误差信号的深层功能作用不是招募关于世界的新的和更好的假设，而是利用感觉信息来指导与世界中与我们当前需求和项目相关的那些方面的流畅参与。

这一切在Feldman（2013，第15页）对”上帝的先验”的讨论中得到了很好的戏剧化呈现，这个名称颇为恶作剧地指称了一个误导性的想法，即”当最优贝叶斯观察者的先验与环境中客观存在的先验相匹配时，它就被正确调节了”。这种概念的深层问题一旦我们反思到主动代理(active agent)从根本上说不是简单地试图建模数据，而是想出在世界中适当行动的方法时就会出现。这将意味着将代理相关信号从噪声中分离出来，选择性地忽略感觉信号提供的大部分内容。此外，“将自己的先验过于紧密地拟合到关于世界如何表现的任何有限观察集合是不明智的，因为观察不可避免地是可靠和短暂因素的混合”（Feldman, 2013，第25页）。

不能保证在线预测学习将以这需要的高效方式正确地将信号从噪声中分离出来。但一切并非没有希望。因为即使在没有持续数据驱动学习的情况下，效率也可以提高（复杂性降低）。做到这一点的一种方法是”修剪”突触连接（也许，如第3章中推测的，在睡眠期间）通过移除弱的或冗余的连接。连接主义中的”骨骼化”算法（参见Mozer & Smolensky, 1990，以及Clark, 1993中的讨论）和恰当命名的唤醒-睡眠算法(wake-sleep algorithm)（Hinton et al., 1995）是这种程序的早期例子，每个都旨在在系统地减少表征过度的同时提供稳健的性能。这种修剪的主要好处是改善泛化——改善在广泛的表面上不同（但根本上相似）的情况下使用已知知识的能力。突触修剪为改善效率和降低模型复杂性提供了一个合理的机制——这种效果可能最经常发生在外感受感觉系统减弱或关闭时，如睡眠期间发生的那样（参见，例如，Gilestro, Tononi, & Cirelli, 2009; Tononi & Cirelli, 2006）。

9.4 混沌和自发皮层活动

突触修剪提供了一种内源性的手段来改善我们对世界的把握。它使我们能够通过消除虚假信息和关联来改善我们模型和策略的把握，从而避免——或至少修复——当系统使用宝贵资源来跟踪训练数据的偶然或不重要特征时发生的那种”过拟合”。这种突触修剪最好被视为改善我们已经在某种粗略意义上掌握的模型的机制。但我们通常做得比这更多。因为我们能够对自己的心理空间进行一种有意的想象性探索。这种能力的基础是免费获得的（正如我们在第3章中看到的），通过使用分层神经预测作为驱动学习、感知和行动的手段。部署这种策略的生物被证明是天然的想象者，能够”自上而下”地驱动自己的感觉运动系统。这样的生物可以从自动尊重生成模型所暗示的相互关联约束的心理模拟中受益。这种模拟提供了一种充分利用我们已经掌握的生成模型的手段，而突触修剪则有助于从内部改善该模型。

但所有这些仍然可能听起来有些保守，好像我们注定（直到新的经验被构建或介入）要大致保持在我们已获得的世界观的限制内。为了窥见更激进的内源性认知探索形式的可能性，回想一下6.6节中简要勾画的自发皮层活动的描述。根据该描述（见Berkes et al., 2011；另见Sporns, 2010第8章），这种自发活动并非”纯粹的神经噪声”。相反，它反映了生物对世界的整体模型。诱发活动（特定外部刺激产生的活动）则反映了该模型应用于特定感官输入时的情况。

这项工作和其他研究（见Sadaghiani et al., 2010）表明，自发皮层活动是感知和行动底层特定生成模型的表达（一种粗略的特征）。根据这种描述，“持续的活动模式反映了世界中因果动力学的历史信息内部模型（用于生成对未来感官输入的预测）”（Sadaghiani et al., 2010, p. 10）。将这种自发皮层活动的图景与关于自组织不稳定性的建议（8.11）结合起来，为认知空间的更激进探索开辟了一个有趣的可能性。

假设我们获得的世界模型是由一个从未完全稳定的动力学机制实现的，这很可能是由于（例如，见Van Leeuwen, 2008）各种混沌式效应。在这种条件下，模型本身（这不过是准备指导感知和行动的结构化神经活动的星座）不断地”闪烁”，探索其自身领域的边缘。这种活动的变化将决定对遇到的感官刺激的微妙不同反应。即使在没有令人信服的感官输入的情况下，这种活动也不会停止。相反，它将继续发生，产生围绕获得模型边缘的持续的刺激分离探索形式——我们可以推测，这些探索可能突然导致对一直占据我们注意力的问题或谜题的新的或更富想象力的（通常更简洁，见9.3）解决方案。此外，Coste et al. (2011)的研究表明，一些自发皮层活动与精度优化的波动有关。也许这种波动允许我们探索自己”元模型”的边缘——我们自己对上下文相关可靠性的估计。

这一切是否至少是关于新思想起源和创造性问题解决的深刻而持久谜题的部分解决方案？Sadaghiani et al. (2010)将他们的描述与机器学习和机器人学中的一些最新工作（Namikawa & Tani, 2010; Tsuda, 2001）联系起来，在这些工作中，这种心理”漫游”确实是新行为形式的主要来源。这种漫游本身可能由隐含的超先验(hyperpriors)强制执行，这些超先验将世界本身描述为变化和不稳定的，因此不适合那些在认知成就上停滞不前的系统。相反，我们会被持续驱动去探索自己知识空间的边缘，即使在没有新信息和经验的情况下，也会时刻微妙地改变我们的预测和期望（包括我们的精度期望）。

在这些主题的有趣扩展中，Namikawa et al. (2011)使用神经机器人模拟探索了复杂分层结构与自组织不稳定性（确定性混沌）之间的关系。在这项工作中，具有多时间尺度动力学的生成模型能够实现一组运动行为。在这些模拟中（就像它们在形式上密切相关的PP模型中一样）：

行动本身是符合……关节角度本体感觉预测的运动的结果，感知和行动都试图最小化整个层次结构中的预测误差，其中运动最小化本体感觉层面的预测误差。（Namikawa et al., 2011, p. 4）

在仿真中，影响较慢时间尺度（更高层次）网络动力学的确定性混沌被发现能够实现原始动作（智能体的基本行为库）之间新的自发转换。这种组织结构被证明既是涌现的又在功能上至关重要。它是涌现的，因为混沌动力学在高层网络中的集中是自然发生的，只要高层的时间常数显著大于其他区域的时间常数（数值细节见Namikawa et al., 2011，第3页）。这种混沌动力学的部分分离在功能上至关重要，因为通过将自组织混沌的影响限制在高层（较慢时间尺度）网络，机器人能够探索有用的、约束良好的可能动作序列空间，而不会同时破坏动作库本身的稳定、可重用元素。因此，它们能够产生新的自发动作转换，而不会干扰使其动作稳健且能够”按需”可靠重现的快速时间尺度动力学（在低层网络中）。

只有时间尺度动力学以这种方式充分分散的网络才能够显示”具有伴随行为原语自发转换的游走行为*和*意图性固定行为（可重复执行）[使用]相同的动力学机制”(Namikawa et al., 2011，第3页，斜体为后加）。相比之下，如果高层网络的时间尺度动力学被降低（变得更快，因此更接近低层网络的时间尺度），机器人行为就会变得不稳定和不可靠，导致作者得出结论：“分层时间尺度差异…对于实现以组合方式自由组合动作和在物理环境中稳定生成它们这两个功能是必需的”(Namikawa et al., 2011，第9页）。这些结果虽然是初步的，但开始暗示多时间尺度动力学的深层

功能作用——这种动力学作为分层预测处理的结果自然发生，并且通过具有不同时间响应特征的神经结构之间的分工合理地实现。

9.5 设计环境和文化实践

然而，在刚才探讨的改进和探索心理空间的任何机制中，没有什么是人类特有的。预测驱动学习、想象、有限形式的模拟，以及对多时间尺度动力学的巧妙利用，都可以在其他哺乳动物中找到，尽管程度不同。正如Roepstorff (2013，第224页）正确指出的，预测处理故事的最基本元素可能在许多类型的有机体和模型系统中找到。新皮质（容纳皮质柱的分层结构，为预测处理机制提供最令人信服的神经实现）在大小上显示出一些戏剧性的变化，但在所有哺乳动物中都很常见。PP模型的核心特征也可能在其他物种中使用其他结构得到支持（例如，昆虫大脑中发现的所谓”蘑菇体”被推测提供了一种实现用于预测的前向模型的方法，见Li & Strausfeld, 1999，以及Webb, 2004中的讨论）。

那么，是什么使我们（至少表面上）如此不同？是什么让我们——不像狗、黑猩猩或海豚——能够抓住远端原因，这些原因不仅包括食物、配偶和相对社会等级，还包括神经元、预测处理、希格斯玻色子和黑洞？一种可能性(Conway & Christiansen, 2001)是人类神经装置的适应以某种方式共同作用，在我们身上创造了一个比其他动物中发现的更复杂、更具情境灵活性的分层学习系统。就PP框架允许分布式系统内广泛的情境依赖影响而言，相同的基本操作原理可能（给定一些新的路由和影响机会）导致质量上新颖的行为和控制形式的出现。这种变化可能解释了为什么人类智能体显示出Spivey (2007，第169页）巧妙描述的对”*任何*时间依赖信号中的分层结构的异常敏感性”。

另一种（可能相关且肯定高度互补的）可能性涉及人类生活的一个强大的复合特征，特别是我们的时间协调社会互动能力（参见 Roepstorff, 2013）以及我们构建人工制品和设计环境的能力。其中一些要素也出现在其他物种中。但在人类的情况下，整个拼图在灵活的结构化符号语言（这是上述Conway和Christiansen处理的目标）和近乎强迫性的驱动力（Tomasello et al., 2005）来参与共享文化实践的影响下结合在一起。因此，我们能够在接触Roepstorff et al. (2010)所称的”模式化社会文化实践”的变革性背景下，反复重新部署我们的核心认知技能。这些包括使用符号铭文（作为”物质符号”遇到，参见Clark, 2006），这些符号嵌入在复杂的实践和社会例程中（Hutchins, 1995, 2014）。这样的环境和实践包括数学、阅读、[5]写作、结构化讨论和教育。这些设计环境的连续和调节构成了Sterelny (2003)描述为”增量下游认识论工程”的复杂而难以理解的过程。

这种堆叠和可传播的结构（设计环境和实践）对神经系统中预测驱动学习的潜在影响是什么？预测驱动学习例程使人类思维在多个空间和时间尺度上对行动就绪、有机体显著世界的统计结构具有渗透性，这反映在训练信号中。但这些训练信号现在作为复杂发展网络的一部分传递，该网络逐渐包括体现在我们所沉浸的符号和其他形式社会文化支架之间的统计关系网络中的所有复杂规律性。因此，我们自我构建了一种滚动的”认知生态位”(cognitive niche)，能够诱导获得生成模型，其范围和深度远远超过它们在与世界的简单感官接触中的表面基础。

要了解这是如何工作的，回想一下构建新想法或概念的方法（假设PP的资源）是遇到新的感官模式，导致高度加权（有机体显著）的预测误差。如果系统无法通过招募它已经掌握的某个模型来解释高度加权的误差，这些误差会导致可塑性增加，并且（如果一切顺利）获得关于负责令人惊讶的感官输入的远端原因的形状和性质的新知识。但我们人类也擅长故意操纵我们的物理和社会世界，使它们提供新的、越来越具有挑战性的模式来驱动新的学习。一个非常简单的例子是在某些文化中，通过故意使用算盘来支架学习心算的方式。对因此可获得的感官模式的体验有助于建立对许多复杂算术运算和关系的理解（有关讨论，参见Stigler et al., 1986）。具体的例子并不重要，但一般策略很重要。我们以使新知识和技能可获得的方式构建（并反复重构）我们的物理和社会环境（有关一些美妙的探索，参见Goldstone, Landy, & Brunel, 2011；Landy & Goldstone, 2005；以及信息理论转折，Salge, Glackin, & Polani, 2014）。渴望预测的大脑在具身行动过程中暴露于新颖的感官刺激模式，可能因此获得在基于物理操纵的生成模型重新调谐之前真正无法达到的知识形式。

这种重新调谐和增强现在由大量符号介导的循环到物质和社会文化中提供服务：这些循环涉及（参见Clark, 2003, 2008）笔记本、草图本、智能手机，以及（参见Pickering & Garrod, 2007）与其他代理的书面和口头对话。[6]这样的循环有效地启用了新形式的重入处理。它们采用”一阶”认知产品（如看到新紫色摩天大楼的视觉体验），用公共符号包装它（将其转化为书面或口头序列，“我今天看到了一座新的紫色摩天大楼”），并将其投入世界，以便它可以作为新类型的具体可感知物——书面或口头句子的感知（Clark 2006, 2008）——重新进入我们自己的认知系统和其他代理的认知系统。这些新的可感知物与其他语言形式的可感知物具有高度信息性的统计关系。一旦外化，想法或思想因此能够参与全新的更高阶和更抽象统计相关性网络。这种相关性的特征是词语预测其他词语的出现，数学符号和运算符的标记预测其他此类标记的出现，等等。

我们从兰道尔和同事关于”潜在语义分析”(LSA，latent semantic analysis)的精彩研究中，窥见了人类语言中复杂内在统计关系的力量。这项研究揭示了词汇与其出现的更大语境（句子和文本）之间统计关系（但是深层的，非一阶的）中所蕴含的大量信息（参见Landauer & Dumais, 1997；Landauer et al., 2007）。例如，这里展示的词汇间深层统计关系包含了能够帮助预测特定学科领域论文成绩的信息。更广泛地说（因为LSA是一种有严格限制的特定技术），我们人类沉浸其中的丰富符号世界现在可被证明*本身*就充满了关于意义关系的信息。这些意义关系反映在我们的使用模式中（因此也反映在出现模式中），它们可以被识别和利用，而不依赖于将词汇和符号与实际行动以及（我们其余的）感官世界联系起来的更基本的钩子。此外，其中一些意义关系存在于这样的领域中：其核心构念现在已经远远脱离了任何简单的感官特征，只有在量子理论、高等数学、哲学、艺术和政治（仅举几例）等深奥世界特有的内在关系中才能看到。

因此，我们对世界的最佳理解被赋予物质形式，并以这种新的形式作为公共可感知的对象——词汇、句子、方程式——变得可获得。这带来的一个重要副作用是，我们自己的思想和观念现在对我们自己和他人都变得可获得，成为刻意注意过程的潜在对象。这为一系列知识改进和知识检验技术打开了大门，从简单的询问理由的对话，到当代科学特有的测试、传播和同行评议的复杂实践。由于所有这些通过口语、书面文本、图表和图片进行的物质性公共载体，*我们*对世界的最佳预测模型（不像其他生物的模型）因此成为了适合公开批评和系统性、多主体、多代际测试与完善的稳定、可重新检查的对象。因此，我们对世界的最佳模型能够作为累积性、共同分布式推理的基础，而不仅仅是提供个体思考发生的*手段*。同样强大的预测处理机制，现在针对这些全新类型的统计上富有意义的”设计师输入”，然后能够发现和完善新的生成模型，锁定（有时主动创造）世界中越来越抽象的结构。结果是，人类建造的（物质和社会文化）环境成为新可传递结构的强大来源，这些结构训练、触发并反复转化渴求预测的生物大脑的活动。

总之，我们人类建造的世界不仅仅是我们生活、工作和娱乐的舞台。它们也构建了终生的统计沉浸，这些沉浸建立和重建生成模型，为每个主体的感知、行动和推理能力提供信息。通过构建一系列设计师环境，如人类建造的教育、结构化游戏、艺术和科学世界，我们反复重构自己的心智。这些设计师环境已经慢慢变得适合像我们这样的生物，它们”了解”我们就像我们了解它们一样。作为一个物种，我们一代又一代地不断完善它们。正是这种迭代的重构，而不是纯粹的处理能力、记忆、移动性，甚至学习算法本身，完成了人类心智的拼图。

9.6 白线

为了进一步理解这种文化中介重塑的力量和范围，回想第8章的主要寓意。寓意是预测大脑不注定要在要求苛刻和时间紧迫的世界中时时刻刻部署高成本、模型丰富的策略。相反，行动和环境结构都可以被调用来降低复杂性。在这种情况下，PP提供并部署充分利用身体、世界和行动的低成本策略。在跑去接飞球这个简单案例中，要解决的问题正是由那个球本身”提出”的，其飞行中的光学特性使低成本解决方案本身变得可获得。在这种情况下，我们不需要主动构建我们的世界以使低成本策略可获得，或提示其使用。然而，在其他情况下，文化雪球效应使我们能够以*既提示又帮助构成*通向行为或认知成功的低成本路径的方式来构建我们的世界。

一个最简单的例子是在蜿蜒的悬崖顶道路边缘涂白线。这种环境改变允许司机通过（部分地）预测各种更简单的光学特征和线索的起伏来解决保持汽车在道路上的复杂问题（例如，参见Land, 2001）。在这种情况下，我们正在建造一个更好的世界来进行预测，同时构建世界以在正确的时间提示该策略。换句话说，我们建造的世界提示更简单的策略，而这些策略只有因为我们首先改变世界的方式才变得可获得。其他例子包括在超市中使用标价（Satz & Ferejohn, 1994）、穿我们最喜爱的足球队的颜色，或展示我们所选亚文化的独特服装风格。

因此，预测误差最小化模型关于皮层处理最基本运作方式的全部潜力，可能只有当这个理论与对沉浸在大量多样化社会文化设计环境中所能产生的作用的理解相结合时才会显现（关于这个方向的一些早期步骤，请参见Roepstorff et al., 2010）。这样一种综合方法将实现一种”神经构建主义”版本（Mareschal et al., 2007），根据这种观点：

大脑的架构……以及环境的统计特性，都不是固定的。相反，大脑连接性受到广泛的输入、经验和活动依赖过程的影响，这些过程塑造并构建其模式和强度……这些变化反过来导致与环境的相互作用发生改变，对未来体验和感知的内容产生因果影响。（Sporns, 2007，第179页）

因此，人类思维和推理的独特之处可能最好通过Hutchins (2014，第35页)所描述的”在大时空尺度上运作的文化生态系统”的运作来解释。在这样的生态系统中，缓慢进化的文化传承实践塑造了神经预测误差最小化发生的世界本身。Hutchins推测，这些文化实践本身可能被有效地理解为在扩展的时空尺度上运作的熵（惊讶）最小化装置。行动和感知随后协同工作来减少预测误差，但这只是在文化分布式过程的更缓慢演化背景下进行的，这种过程产生了一系列实践和设计环境，它们对人类思维和推理的发展（例如，Smith & Gasser, 2005）和展开的影响很难被高估。

当然，也有不利的一面。不利的一面是，这些文化中介过程也可能产生各种路径依赖形式的成本（Arthur, 1994），在这种依赖中，后续解决方案建立在早期解决方案的基础上。次优的基于路径的特异性可能随后被冻结（也许就像备受讨论的QWERTY键盘或Betamax录像格式）到我们的物质人工制品、制度、记号法、测量工具和文化实践中。但这些成本是容易承受的。因为正是这些相同的轨迹敏感文化过程带来了巨大的认知收益，这些收益来自设计环境的缓慢、多代发展——这些环境帮助人类心智到达其他动物心智无法到达的地方。

9.7 为创新而创新

为这种社会文化-技术之火增添更多燃料，甚至可能如Heyes (2012)优雅论证的那样，我们许多*文化学习能力*本身就是*文化创新*，通过社会互动获得，而不是直接来自基本的生物适应。这里的观点是：

文化学习的专门特征——使文化学习特别擅长促进信息社会传播的特征——是在发展过程中通过社会互动获得的……它们既是文化进化的产物，也是文化进化的生产者。（Heyes, 2012，第2182页）

借用Heyes自己的类比，文化学习不仅仅是产生越来越多”谷物”（关于世界的可传播事实）的生产者，而且是”磨坊”的来源——“使我们能够从他人那里学习谷物的心理过程”（Heyes, 2012，第2182页）。

最明显的例子是读写，这是一对匹配的文化实践，它们似乎出现得太晚，不可能是遗传适应的结果。众所周知，阅读实践会导致人类神经组织的广泛变化（Dehaene et al., 2010; Paulesu et al., 2000）。由此产生的新组织利用了Anderson (2010)所描述的”神经重用”基本原则，其中预先存在的元素被招募和重新利用。通过这种方式：

学习阅读采用旧部分并将它们重塑为一个新系统。旧部分是计算过程和皮层区域，最初在遗传和文化上适应于物体识别和口语，但是个体发生的、文化的过程——识字训练——使它们成为专门用于文化学习的新系统。（Heyes, 2012，第2182页）

因此，阅读是文化遗传磨坊的一个很好的例子——一个文化进化的产物，它滋养和推动文化进化过程本身。Heyes论证，其他例子可能包括社会学习的关键机制（通过观察其他主体行动来学习）和模仿。如果Heyes是对的，那么*文化本身*可能对许多子机制负责，这些子机制为文化雪球提供了传递像我们这样的心智的手段和动力。

[9.8 词语作为操控精确性的工具]

词汇和短语享有双重生命。它们既是交流工具，似乎也在我们自己思想和观念的展开和发展中发挥作用。后一种角色有时被称为语言的超交流维度(supra-communicative dimension)（参见Clark, 1998; Dennett, 1991; Jackendoff, 1996）。这种超交流角色似乎相当神秘。仅仅因为用语言表达一个思想（你可能坚持认为她已经有了这个思想），一个智能体能获得什么认知优势？答案很可能是我们错误地描绘了这种情况。与其说仅仅表达我们已有的思想，这种行为必须以某种方式改变、影响或转换思维本身。但这到底是如何运作的呢？

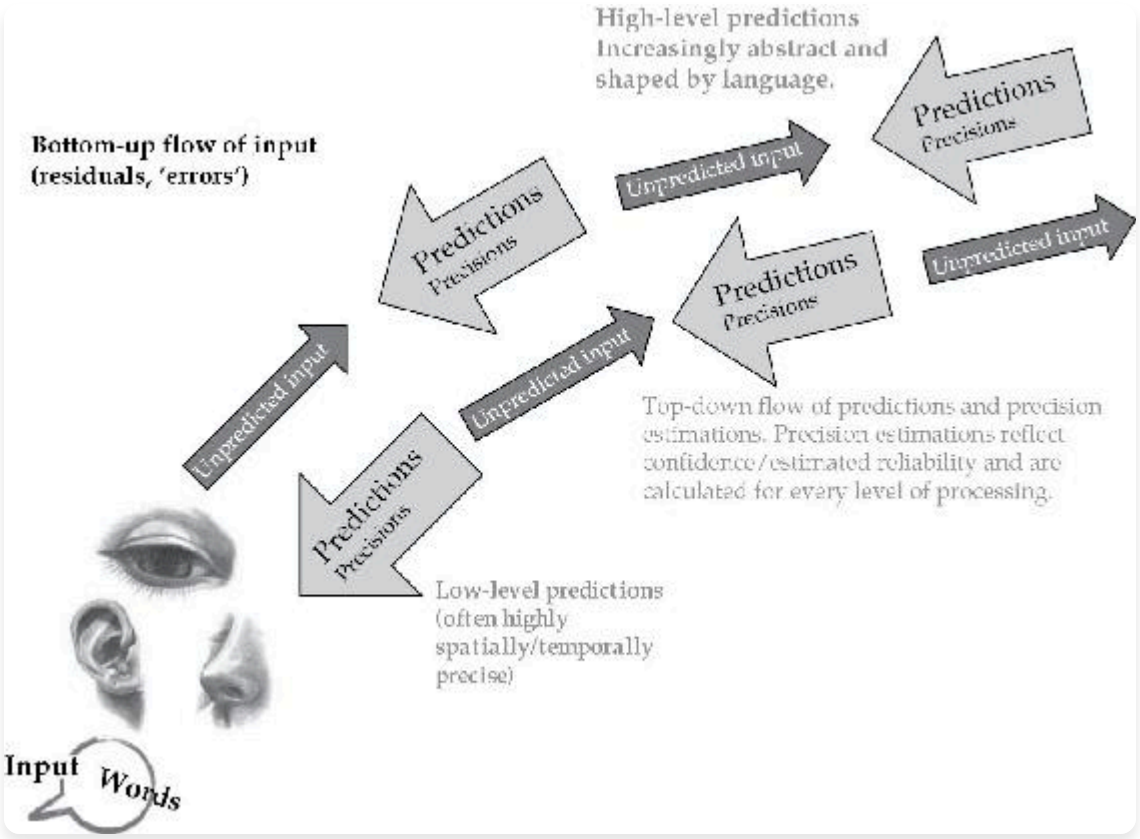
从预测处理(PP)的角度考虑，遇到的、自我产生的或对话共建的词汇或短语对个体处理和问题解决可能产生的影响。在这种情况下，我们要么独自一人，要么作为集体的一部分，创造”人工输入流”，这些输入流可能特别适合改变和细化内在处理流程，从而帮助决定感知、体验和行动。

在对这些想法的基础探索中，Lupyan and Ward (2013)使用连续闪烁抑制(Continuous Flash Suppression, CFS)技术进行了实验。[8]在CFS中，当变化的其他图像流呈现给另一只眼睛时，持续呈现给一只眼睛的图像会被抑制。这是双稳态感知(bi-stable perception)的另一个例子，与我们在第1章中探讨的双眼竞争案例相关。[9]Lupyan和Ward发现，被CFS掩盖意识的物体如果在试验开始前听到正确的词——如果被抑制的物体是斑马，就听到”斑马”这个词——可以被解除抑制（有意识地检测到）。听到正确的词提高了检测物体的”命中率”，也缩短了反应时间。作者认为，解释是”当与言语标签相关的信息与传入的（自下而上的）活动匹配时，语言为感知提供自上而下的增强，将原本不可见的图像推进意识”（Ward & Lupyan, 2013, p. 14196）。在这个实验中，言语增强是外部提供的。但是相关效应——不提供外部提示——也在有意识识别中得到了证明。因此，Melloni et al. (2011, 在3.6节中讨论）表明，形成可报告的有意识感知所需的起始时间根据恰当期望的存在与否而显著变化（约100毫秒），即使这些期望是在被试执行任务时自然产生的。将这两种效应结合起来表明，接触词汇的功能是改变或细化帮助构建我们持续体验的积极期望。

在某些方面，这似乎很明显。词汇会产生影响！但有新兴证据表明，接触词汇和短语所诱发的期望特别强烈、集中和有针对性。Lupyan and Thompson-Schill (2012)发现，听到”狗”这个词比简单地听到吠叫声更能有效改善相关辨别任务的表现。还有引人入胜的证据（参见Çukur et al., 2013，以及Kim & Kasstner, 2013中的讨论）表明，基于类别的注意（如在观看电影或视频片段时被告知暗中注意”车辆”或”人类”）会暂时改变神经元群体的调谐，使神经元或神经元集合的类别敏感性朝着被注意内容的方向转移。

这种指令诱导的皮层表征的瞬态改变可以通过改变特定预测误差信号精度权重的机制套件来实现。这很有启发性。结构化语言的一个强大特征是它能够廉价且非常灵活地针对我们自己理解或我们对另一个智能体理解的高度特定方面。在预测处理风格的认知架构背景下，这可能通过影响我们对精度的持续估计来发挥作用，因此影响分配给持续神经活动不同方面的相对不确定性。Yuval-Greenberg and Heeger (2013, p. 9365)的最新工作表明”CFS基于调节神经反应的增益，类似于降低目标对比度”。预测处理调节这种增益的机制当然是预测误差的精度权重。可以推测，语言提供了一种精细调谐的手段来人为操纵精度（因此暂时修改影响），在神经处理的不同层次上操纵预测误差。这种瞬态的、有针对性的、微妙的精度操纵可以选择性地增强或抑制我们自己或另一个智能体世界模型任何方面的影响。自我产生的（或心理排练的）语言随后会成为探索和利用我们自己获得的生成模型的全部潜力的有效手段，为操纵我们自己预测误差的精度权重提供一种*人工第二系统*——因此是一个”巧妙技巧”（Dennett, 1991），用于人为操纵我们对自己不确定性的估计，使我们能够充分灵活地利用我们所知道的。

我们可以说，词语（对我们这些语言使用者而言）是代谢成本低廉且灵活的”人工语境”来源(Lupyan & Clark, in press)。从预测处理(PP)的角度来看，词语串对神经处理的影响是灵活地修改自上而下信息的应用方式，以及它在每个处理层级上的影响程度（见图9.1）。这种用于有针对性自我操控的强大工具将为智力提供巨大提升，在远超单纯语言表现相关的方面改善表现。[10] 此外，这种工具的整体认知影响预计会与智能体语言技能库的精妙性和范围成正比。



image

图9.1 基本预测处理图式，语言作为附加输入

来源：改编自Lupyan & Clark, in press。

这确实是一个诱人的图景。但公共语言形式编码如何与PP假设的结构化概率知识表征相互作用，在很大程度上仍然未知。这种交互作用是前面描述的文化建构过程的核心，必须成为未来研究的重要目标。

[9.9 与他人共同预测]

在社会世界中，我们刚刚勾勒的许多技巧和策略在相互支持的混合中汇聚在一起。其他智能体（回想第5章）通常适合使用产生我们自己行动的同一生成模型进行预测。但其他智能体也提供了一种独特的”外部脚手架”形式，因为他们的行动和反应可以被利用来减少我们自己的个体处理负荷。最后，其他智能体本身也是预测者，这为互利（或有时是破坏性的——回想2.9）的”连续相互预测”过程开辟了一个有趣的空间。

Pickering和Garrod (2007, 2013)详细探讨的一个很好的例子是对话的共同构建。在对话中，Pickering和Garrod认为，每个人都使用自己的语言产生系统（因此是支撑自己行为的生成模型）来帮助预测他人的话语，同时也使用他人的输出作为自己正在进行的表达的一种外部脚手架。这些预测（正如PP所暗示的）是概率性的，跨越从音韵学到句法和语义的多个不同层级。随着对话的进行，多个预测因此与其相关概率持续共同计算（另见Cisek & Kalaska, 2011; Spivey, 2007）。在这种过程中，各方通常都在努力匹配或试图匹配自己的行为和期望与他人的行为和期望。随着对话的进行，词语、语法、语调、手势和眼球运动都可能被公开复制或暗中模仿（对语言和行为证据的便利回顾，见Pickering & Garrod, 2004）。特别是公开复制，有助于支持相互预测和相互理解，因为”如果B公开模仿A，那么A对B话语的理解就会通过A对自己先前话语的记忆得到促进”(Pickering & Garrod, 2007, p. 109)。结果是”预测和模仿可以共同解释为什么对话往往很容易，尽管它涉及持续的任务切换以及确定何时说话和说什么的需要”(p. 109)。

当然，这种搭便车并不局限于我们的对话互动。相反，在某种形式或另一种形式中，它似乎表征了人类联合行动的许多形式，从团队运动到与伴侣一起换床单(Sebanz & Knoblich, 2009)。个体智能体也可能主动约束自己的行为，以使自己更容易被其他智能体预测。因此，我们可能人为地稳定自己的公共人格，以鼓励他人与我们建立经济或情感安排(Ross, 2004)。在更宏大的尺度上，Colombo (in press)将社会规范(norm)（日常社会行为中大多未成文的”规则”，如在餐厅留小费）描述为其作用是通过创建结构或图式来减少相互不确定性的设备，在这些结构或图式中，行为变得更加相互可预测。Colombo论证，社会规范是熵最小化设备，表征为概率分布，服务于使社会行为可预测。因此，对我们自己行为的期望在本质上既是描述性的又是规范性的。

这种双重性质在个人叙事的认知作用中也很明显(Hirsh et al., 2013)：我们向自己和他人讲述的关于我们生活流程和意义的故事。这种叙事作为构建我们自己自我预测模型的高级元素发挥作用，因此影响我们自己未来的行动和选择。但个人叙事通常与他人共同构建，因此往往将社会的结构和期望反馈回来，使它们反映在个体用来理解自己行为和选择的模型中。因此，个人叙事可以被视为另一种共同的不确定性减少设备。

Roepstorff and Frith (2004) 指出，许多人类互动案例都涉及一种顶层的”脚本共享”，其中控制一个个体行动的最高层过程可能起源于另一个个体的大脑。他们详细研究的案例是构建对实验情境的充分理解，以允许被试参与特定的心理学实验。在这些案例中，人类个体通常能够通过一段言语指导来实现参与所需的（有时相当苛刻的）理解，在这个过程中，高层理解直接从实验者传达给被试。这就是Roepstorff and Frith 引人入胜地称为”行动的顶-顶控制”的案例，其中实验者的高层理解元素通过语言交流得以定位，从而控制另一个个体的反应模式。这种情况可以与训练猴子所需的漫长而艰苦的训练过程形成对比。Roepstorff and Frith 描述的特别具有挑战性的例子涉及执行威斯康星卡片分类任务的简化版本，在猴子能够充当合适的被试之前，需要进行一整年的操作性条件反射训练（参见Nakahara et al., 2002）。在这种训练之后，Nakahara 等人发现猴子和人类被试在执行任务时都出现了解剖学上相似的大脑激活，这表明按照计划，猴子确实学会了与人类被试相同的”认知集合”（相同的行动和选择指导方案）。但尽管有这种终点相似性，过程显然是根本不同的，因为：

人类参与者在”顶-顶”交流中直接从实验者那里接收这个脚本，而猴子必须仅通过提供给它的具体刺激和奖励来重建这个脚本。这是在猴子基于对情境的先前理解对实验者给予的奖励反应做出反应时发生的。（Roepstorff & Frith, 2004, p. 193）

在猴子的案例中，脚本同步需要通过真正自下而上的学习过程来重新创建实验者的顶层理解，而在人类被试的案例中，这种艰苦的路径可以通过明智地使用语言和预先存在的共同理解来避免。在已经拥有重要共同理解的人类被试中，语言因此提供了一种廉价的、易于获得的”顶-顶”行动控制路径。

循环的语言形式互动因此可以帮助创建Hasson et al. (2012, p. 114) 所描述的”脑-脑耦合”系统，其中”一个大脑的感知系统与另一个大脑的运动系统耦合”，以使新形式的联合行为得以涌现——例如，当一个个体在将大钢琴搬上楼梯时向另一个个体大声发出命令。

语言还为整个人类个体群体提供了集体协商复杂表征空间的手段。特别是，它提供了一种（参见Clark, 1998）驯服”路径依赖”学习的方法。路径依赖性，在其最熟悉的形式中，是结构化教育和训练的理论基础。这是必要的，因为某些想法只有在其他想法就位后才能被理解。这种”认知路径依赖性”可以很好地解释（参见例如Elman, 1993），将智力进步视为涉及在大型复杂空间中进行计算搜索过程的东西。先前的学习使系统倾向于尝试空间中的某些位置而不是其他位置。当先前学习是适当的时，发现某种新规律的工作变得可处理：先前学习充当要探索的选项空间的过滤器。我们一直在探索的基于预测的方法的层次性质使它们特别适合作为能够支持复杂路径依赖学习模式的内部机制，其中后期成就建立在早期成就的基础上。然而，与此同时，先前学习使某些其他规律更难（有时不可能）被发现。因此，先验知识总是既具有约束性又具有使能性。

当面对表现出路径依赖学习的个体时，语言允许想法被保存并（在某种意义上）在个体之间迁移这一平凡观察获得了新的力量。因为我们现在可以理解这种迁移如何可能允许极其精细和困难的智力轨迹和进程的共同构建。一个只有乔的经验才能提供的想法，但只有在玛丽大脑目前提供的智力生态位中才能繁荣并实现其全部潜力，现在可以通过在这些个体之间的旅行来实现其全部潜力。不同的个体（以及同一个体在不同时间）构成不同的”过滤器”，这样的个体群体使得任何单一个体都无法理解的学习和发现轨迹变得可用。因此，语言链接的社区内可用的智力生态位的多样性提供了一个令人惊叹的群体层面多个体轨迹矩阵。

总之，社会互动智能体从持续相互预测的嵌套式自我强化循环中获益。当互动的预测性智能体群体构建共享的社会世界时，这种联合叠加效应(joint piggy-backing)自然涌现，可能是人类互动低成本计算策略的基本来源。智能体间的交流创造了通往可能理解空间的新路径，使得交流智能体网络能够共同探索那些会迅速击败任何个体智能体的智力轨迹。

9.10 构建我们的世界

行动、文化学习、相互预测、巧妙运用语言以及多种形式的社会技术脚手架(scaffolding)的综合效应是变革性的。正是预测大脑与这整套相互支持的技巧和策略之间尚未被充分理解的炼金术，使我们具有了独特的人类特征。因此，我们更大故事的一个直接含义是，存在一种非常真实的意义，即人类智能体帮助构建了他们建模和居住的世界。

这种构建过程与听起来神秘的”构建世界”概念相当接近，至少如该概念在Varela等人(1991)中出现的那样。

Varela等人写道：

构建式感知方法(enactive approach)的总体关切不是确定如何恢复某个独立于感知者的世界；相反，它是确定感觉和运动系统之间的共同原理或规律性联系，这些联系解释了行动如何在依赖于感知者的世界中得到感知引导。(Varela等人，1991，第173页)

据Varela等人报告，这种感知方法在Merleau-Ponty(1945/1962)的工作中就有预示。在那里，Merleau-Ponty强调了感知本身在很大程度上由人类行动构建的重要性。因此，我们经常认为感知仅仅是信息来源，然后用于指导行动。但如果稍微扩展时间窗口，就会清楚地看到，我们同样可以将行动视为感知刺激本身的选择器。用Merleau-Ponty的话说：

由于生物体接收到的所有刺激反过来都只有通过其先前的运动才成为可能，这些运动最终将受体器官暴露于外部影响之下，人们也可以说行为是所有刺激的第一原因。(Merleau-Ponty，1945/1962，第13页)

在一个引人注目的比喻中，Merleau-Ponty然后将活跃的生物体比作一个移动的键盘，它移动自己以向”本身单调的外部锤子行动”提供不同的键(第13页)。世界”敲打给感知者”的信息因此在很大程度上是由感知者自身的性质和行动创造的(或者这个比喻所暗示的)：她将自己呈现给世界的方式。据Varela等人(1991，第174页)的观点，结果是”生物体和环境在相互规定和选择中结合在一起”。

Varela等人将这种关系描述为”结构耦合”(structural coupling)，其中”物种产生并规定了自己的问题域”(第198页)，并在这个意义上”构建”或产生(第205页)自己的世界。在讨论这些问题时，Varela等人还关注强调相关的结构耦合历史可能选择他们所描述的”非最优”特征、性状和行为：涉及”满意化”(satisficing)(见Simon，1956，以及第8章)的那些，这意味着满足于任何”足够好”的解决方案或结构”具有足够的完整性来持续存在”(Varela等人，1991，第196页)。预测处理(PP)具有兑现所有这些”构建主义”(enactivist)支票的资源，将生物体和生物体显著的世界描述为在相互规定过程中结合在一起，其中适合支持可行互动历史的最简单近似是被学习、选择和维持的。

启用预测处理的智能体可以说主动构建其世界的最简单方式是通过采样。这里的行动通过以旨在”提供”预测刺激模式的方式移动身体和感觉器官来服务于感知。特别是，它们旨在(第2章)提供高可靠性、任务相关信息的预测序列。在我看来，这是Merleau-Ponty想象的那种”主动键盘”效应的一个非常清楚的例子——生物体选择性地移动其身体和受体以尝试发现它所预测的刺激。通过这种方式，不同的生物体和个体可能以选择性采样的方式，既主动构建又持续确认不同”世界”的存在。正是在这个意义上，如Friston、Adams和Montague(2012，第22页)评论的那样，我们的隐式和显式模型可以说”创造了它们自己的数据”。

这样的过程在几个组织尺度上重复进行。因此，我们人类不仅仅是对某些自然环境进行采样。我们还以（正如我们刚刚看到的）各种强有力的、相互作用的、往往是累积性的方式来构建那个环境。我们通过建造物质人工制品（从住宅到高速公路）、创造文化实践和制度，以及交易各种符号化和记号化的道具、辅助工具和脚手架来做到这一点。我们的一些实践和制度也被设计来训练我们更有效地采样和利用我们人类建造的环境——例子包括体育练习、特定工具和软件使用的培训、学习快速阅读，以及许许多多其他的例子。最后，我们的一些技术基础设施现在以旨在减少预测性智能体负荷的方式进行自我改变，从我们过去的行为和搜索中学习，以便在正确的时间提供正确的选项。通过所有这些方式，以及在所有这些相互作用的空间和时间尺度上，我们构建并选择性地采样那些世界——在统计敏感交互的重复回合中——安装我们用来处理它们的生成模型。

在所有这些设置中，生成模型的任务是捕获将支持完成工作所需行动的最简单近似——那（正如我们在第8章中看到的）意味着考虑到生物的形态学、物理行动和社会技术环境所能完成的任何工作。因此，PP完全与强调节俭、满意化(satisficing)以及充分利用大脑、身体和世界的简单但充分解决方案普遍性的工作相协调。大脑、身体和部分自我构建的环境显示为”相互嵌入的系统”（Varela等，2001年，第423页），在情境成功的服务中共同工作。

[9.11 表征：坏的突破？]

然而，至少还有一个著名的棘手问题，PP和激活论(enactivism)（至少如果历史是任何指导的话）似乎注定要在这个问题上产生分歧。那就是”内部表征”的问题。因此Varela等明确表示，在激活论概念上，“认知不再被视为基于表征的问题解决”（第205页）。然而，PP大量处理内部模型——丰富的、节俭的，以及介于两者之间的所有点——其作用是通过预测感官数据的复杂表现来控制行动。激活论者可能担心，这就是我们关于神经处理的有前景故事”变坏”的地方。为什么不简单地抛弃关于内在模型和内部表征的讨论，坚持激活论美德的真正道路呢?

这个问题需要比我在这里（也许仁慈地）尝试的更多讨论。[14]尽管如此，PP和激活论之间的剩余距离可能不像那种露骨的对立所暗示的那么大。我们可以首先回忆起，尽管PP大量使用内在模型和表征的讨论，但它调用的表征是完全概率性的和面向行动的。这些表征（见[第5-8章]）在滚动感觉运动循环的背景下，*根本上*从事服务行动的业务。这样的表征旨在*参与*世界，而不是以某种行动中立的方式描绘世界。此外，它们仍然牢固植根于有机体-环境交互的模式中，这些模式提供了在成熟概率生成模型中反映的结构化感官刺激。该生成模型的作用是在多个竞争性行动可供性(affordances)的世界中提供高效的、上下文敏感的掌握。

这种掌握的形状被Itay Shani很好地捕捉到，他写道：

实际的感官系统本身不关心真理和准确性，而是关心行动和维持它们所嵌入的有机体功能稳定性的需要。它们不像理想化的科学观察者那样报告或记录什么在哪里，而是帮助有机体应对其外部和内部（躯体）环境中的变化条件。(Shani, 2006, 第90页)

如果PP是正确的，这正是指导感知和行动的多层概率生成模型所发挥的作用。[15]

这些多层次面向行动的概率生成模型所支配的状态包含什么*内容*？生成模型发出预测，估计各种可识别的世界状态（包括身体状态和其他智能体的心理状态）。但正如我们反复看到的，估计神经估计本身的情境变量可靠性（精度）也是必要的。其中一些精度加权的估计驱动行动，而正是行动随后对场景进行采样，提供感知从而选择更多行动。这种循环复杂性将使得用日常口语的术语和词汇来充分捕捉许多关键内在状态和过程的内容或认知角色变得困难（也许是不可能的）。这种词汇是为交流而”设计”的（尽管它也可能启用各种形式的认知自我刺激(cognitive self-stimulation)）。相比之下，概率生成模型是为了在滚动的、不确定性调制的感知和行动循环中与世界互动而设计的。因此构建的表征”不是世界中物体的实际再现或复制品，而是…对世界进行预测并基于与世界的互动修正其预测的不完整、抽象代码”（Lauwereyns, 2012，第74页）。在PP中，高级状态（生成模型的）针对空间和时间中大规模的、日益不变的模式。这些状态帮助我们跟踪特定的个体、属性和事件，尽管感觉刺激流中存在巨大的瞬间变化。通过下行预测级联展开，这些更高级别的状态同时为感知和行动提供信息，将它们锁定在连续的循环因果流中。这些模型不是简单地描述”世界是什么样的”，即使在”更高”更抽象的层次上考虑，也是为了接触那些对我们重要的世界方面而设计的。它们为*重要交互*提供对*重要模式*的把握。

这为内在表征争议中的和平提供了一个方案。Varela et al. (1991)强烈拒绝对”内在表征”的诉求。但对他们来说，这个概念暗示了对他们所称的”预给定世界”的”行动中性”捕捉。他们争论说，有机体和世界是通过结构耦合(structural coupling)的历史共同定义的：一种彼此的主动”拟合”，而不是被动的”镜像”。我试图表明，PP完全尊重这种直觉。它假设一个分层生成模型，通过使系统能够最小化预测误差，从而避免与环境的危害性（可能致命的）遭遇，来帮助维持系统的完整性和可行性。这个分布式内在模型本身是在多个时间尺度上运行的自组织动力学的结果，它选择性地发挥作用，使智能体暴露于它预测的刺激模式中。因此，生成模型的功能——正如行动主义者(enactivist)可能坚持的那样——是启用和维持服务于我们需求并保持我们可行性的结构耦合。

我们是否可以完全用非表征术语来讲述我们的故事，而根本不调用分层概率生成模型的概念？人们应该总是谨慎对待关于有朝一日在解释上可能实现什么的全面断言！但就目前情况而言，我根本看不出这如何能够实现。因为正是这种描述让我们理解这些循环动力学制度是如何产生并实现如此壮观结果的。这些制度产生并成功，是因为系统自组织以捕捉（部分自创的）输入流中的模式。这些模式指定了在不同空间和时间尺度上运作的身体和世界原因。减去这个指导愿景，剩下的只是跨越大脑、身体和世界的复杂循环动力学图景。这样的愿景当然是正确的，就其所及而言。但它不能解释（不能使之可理解）结构化的、适合感知、思维、想象和行动的有意义领域的出现。

然而，从PP提供的解释视角来考虑那些相同的循环动力学，许多事情自然就位了。有了这个图式在心中，我们理解感知、想象和基于模拟的推理是从单一认知架构中共同涌现的；我们看到这种架构如何同时支持感知和行动，将它们锁定在循环因果拥抱中；我们看到为什么，以及确切地如何，感知和行动本身是共同构建和共同决定的；我们看到在更长的时间尺度上，统计驱动的学习如何能够首先发掘相互作用的远端和身体原因，揭示一个充满人类规模行动机会的结构化世界；我们理解意外的遗漏和缺失如何能够与最具体的普通可感知物一样突出和在感知上引人注目。此外，我们从一个既容纳又统一机器学习、心理物理学、认知和计算神经科学以及（日益增长的）计算神经精神病学中令人印象深刻的大量工作的视角来欣赏这一切。这确实令人鼓舞。也许这个大致范围内的模型为我们提供了对具身心智的基础和统一科学形状的首次一瞥？

9.12 野外预测

我们的神经经济存在是为了服务于具身行动的需要。它通过启动和维持复杂的循环因果流来实现这一点，在这些因果流中，行动和知觉是相互决定和相互确定的。这些流制定了结构耦合，在服务我们需要的同时，让有机体保持在其自身专门的生存窗口内。所有这些都是由一个多层次生成模型协调的，正如我们的故事所暗示的，该模型被调整来预测当前感官信号的任务显著性方面。

这是一个充斥着表征、脱离世界的内在经济吗？绝对不是。这是一个面向行动的内在经济，旨在将具身智能体锁定到其世界中的机会上。从动力学角度来说，整个具身、主动系统在这里围绕有机体可计算的量”预测误差”进行自组织。这正是提供对不断演化的感官冲击的多层次、多区域把握的要素——这种把握必须跨越多个空间和时间尺度。这样的把握同时决定知觉和行动，并且选择（制定）正在进行的感官轰炸流本身。在这里发出感官预测的生成模型因此不过是对展开的感官流的多层次、多区域、多尺度、涉及身体和行动的把握。实现这种把握就是了解我们在体验和行动中遇到的结构化和有意义的世界。

在人类心智的某种特殊情况下，这种把握被层层叠叠的社会文化结构和实践进一步丰富和转化。浸润在这样的实践中，我们的预测性大脑被赋予了以新的、变革性的方式重新部署其基本技能的能力。理解文化、技术、行动和级联神经预测之间产生的相互作用无疑是二十一世纪认知科学面临的主要任务之一。

10

结论

预测的未来

记住所有模型都是错误的；实际问题是它们必须错误到什么程度才不再有用。

—[Box & Draper, 1987]

10.1 具身预测机器

预测处理(Predictive Processing, PP)提供了一个与具身和环境情境心智研究完美契合的大脑愿景（或者说我一直在论证的）。这是一种由行动以及连接行动和知觉的循环因果流锻造的契合。这种契合揭示了知觉、理解、推理和想象是共同涌现的，而不安定的游移动力学是具身心智的特征。在这种永远活跃、自组织的流动中，神经子集合以由相对不确定性的变化估计决定的方式形成和消解。这些临时电路招募并被变化的身体和身体外结构和资源网络招募。由此产生的瞬态整体是适应性和行为成功的真正（节俭、高效）载体。因此，预测性大脑与其说是”头脑中”的隔离推理引擎，不如说是面向行动的参与引擎，提供对任务显著机会的滚动把握。

雕刻和维持这种滚动把握的是向下（和横向）流动预测的深层认知引擎。正是这些深层认知引擎（多层次概率生成模型）使我们能够在知觉和行动中遇到一个为人类需要和目的而解析的世界。大量洒布着我们自身预测误差精度（方差或不确定性）的估计，这提供了一个强大的工具包来冲浪感官不确定性的波浪。如此装备的生物是自然界分离信号与噪音的专家，能够辨别构造其感官表面持续能量冲击的显著相互作用原因。

知觉和行动在这里作为单一计算硬币的两面出现。植根于多层次预测误差最小化程序，知觉和行动被锁定在复杂的循环因果流中。在该流动中，行动选择感官刺激，既测试又响应多层次级联识别的身体和环境原因。感知和行动配方在这里共同涌现，将运动处方与理解我们世界的持续努力结合起来。因此，知觉和行动是相似地、同时地构建的，并且密切交织。

这样的系统是知识驱动的，这要归功于编码在复杂多层次生成模型中的结构化概率诀窍。但它们也是快速和节俭的，能够使用这种诀窍来帮助为给定任务和情境选择最具成本效益的策略。许多这些成本效益策略将行动和身体及环境结构（和干预）的使用与昂贵的机载计算形式的使用进行权衡。与较慢的进化适应过程安装的全套策略和手段一起工作，这使得在任务和情境允许时能够灵活和智能地选择低成本高效策略。

因此，与具身和情境认知科学的契合得到了充分实现。知觉-行动循环是基础性的；低成本、表征高效的选项是首选的；而持续的误差最小化行动流允许任意复杂的外部资源套件的招募和使用——这些资源现在简单地被卷入正在进行的循环因果流中。

因此遇到的世界在很大程度上是由其呈现的行动可供性(affordances)所构建的。随着我们基于可供性的世界探索进行，内感受和外感受信息不断结合，环境原因得到识别，行为得到训练。这为体验、情感和情绪的解释提供了一个丰富的新入口：这些解释不会将认知和情感分割开来，而是将它们揭示为（至多）单一推理编织中的独特线索。在

这种密集、持续、多层次的交换中，内感受、本体感受和外感受信息不断协同工作，人类体验的流动在多样化系统期望和自结构感觉流的交汇点上呈现为一个连续的构建。

这里暗示了对感知中遭遇世界意味着什么的新理解。这将是一种理解，其中体验、期望、估计的不确定性和行动不可分割地交织在一起，共同在一个世界中提供把握和轨迹，这个世界的生物体相关特征作为我们自身活动的结果被持续揭示（在某些情况下，被持续创造）。如果这是正确的，那么心智与世界之间的契合不是通过某种形式的被动”恰当描述”而锻造的，而是通过行动本身：持续选择我们响应的刺激的行动。这种持续感觉运动流程的关键是使用精度加权(precision-weighting)来控制不仅是各个层次上先验知识（因此是预测）的相对影响，还有构成瞬态但统一的处理机制的大规模信息流，这些机制在我们从任务到任务、从情境到情境移动时形成和消散。

通过围绕预测误差进行自组织，通过学习生成式而非仅仅判别式（即模式分类）模型，这些方法实现了人工神经网络、机器人学、动力系统理论和经典认知科学中先前工作的许多梦想。它们使用多层架构执行无监督学习，并获得对领域内结构关系的令人满意的把握——这得益于其分层形式所实现的问题分解。此外，它们以牢固植根于构建学习的感觉运动体验模式的方式做到这一点，使用连续的、非语言形式的内部编码（概率密度函数和概率推理）。通过基于精度的有效连接模式重构，这些相同的方法将简单性嵌套在复杂性中，并根据任务和情境的要求充分（或较少）利用身体和世界。

10.2 问题、难题和陷阱

所有这些都暗示了一门真正基础且深度统一的具身心智科学的形态。在本文中，正是这个更大的图景我试图聚焦。这意味着专注于那个积极的、整合的故事——至少在预测心智科学仍处于起步阶段时，这种策略似乎是合理的。但仍然存在一系列需要解决的问题、陷阱和不足。在关键难题和问题中，有四个特别具有挑战性和重要性。

第一个涉及探索更大的近似空间和可能表征形式的迫切需要。复杂的现实世界问题需要使用对真正最优形式的概率推理的近似，神经元群体可能以多种方式表示概率，执行概率推理也有多种方式（见，例如，Beck et al., 2012; Kwisthout & van Rooij, 2013; Pouget, Beck, et al., 2013）。因此需要更多工作来发现大脑实际部署哪些近似。此外，这种首选近似的形式将与所使用的表征类型相互作用，并将反映可用于缩小搜索空间的领域知识的可用性。通过仿真（Thornton, ms）、使用行为指标（Houlsby et al., 2013）和探测生物大脑（Bastos et al., 2012; Egner & Summerfield, 2013; Iglesias, Mathys, et al., 2013; Penny, 2012）来解决这些问题，对于相对抽象的理论模型（如PP）得到测试并转化为人类认知的合理解释是至关重要的。

第二个问题涉及探索多种变体架构的需要。本文主要聚焦于一种可能的架构方案：该方案需要功能上不同的神经群体分别编码表征（预测）和预测误差，其中预测通过神经层级向后（和横向）流动，而残余预测误差的信息向前（和横向）流动。但该方案仅代表基于概率生成模型方法这一庞大复杂空间中的一个点，在一般范围内存在许多可能的架构，以及结合自上而下预测和自下而上感觉信息的可能方式。例如，Hinton及其同事在深度信念网络方面的奠基性工作（Hinton, Osindero, & Tey, 2006；Hinton & Salakhutdinov, 2006）尽管同样强调概率生成模型，但仍有所不同（第1章，注释13）；McClelland (2013)和Zorzi et al. (2013)将深度无监督学习的工作与上下文效应的贝叶斯模型以及如连接主义交互激活模型等神经网络模型相结合；Spratling (2010, 2011, 2014)提出了一种替代的预测编码模型PC/BC，代表预测编码/偏向竞争(biased competition)，使用不同的预测和误差流实现预测编码的关键原理，并由变体数学框架描述；Dura-Bernal, Wennekers et al. (2011, 2012)开发了Spratling PC/BC架构的变体，扩展了知名的HMAX（Riesenhuber & Poggio, 1999；Serre et al., 2007）物体识别前馈模型，在再现前馈方法许多计算效率的同时，适应了强烈的自上而下效应（如虚幻轮廓的感知）；Wacongne et al. (2012)开发了一个详细的神经元模型，使用分层的尖峰神经元网络实现预测编码（用于听觉皮层）；O’Reilly, Wyatte和Rohrlich（手稿；另见Kachergis et al., 2014）提供了预测学习的丰富神经计算解释，其中同一层在处理的不同时间阶段编码期望和结果，输入往往只能部分预测；Phillips and Silverstein (2013)发展了关于上下文敏感增益控制的更广泛、计算丰富的视角；den Ouden, Kok, and de Lange (2012)调查了大脑似乎编码预测误差信号的多种方式，以及这些信号在不同脑区的不同功能作用；Pickering and Garrod (2013)使用相互预测和前向模型的装置，提出了语言产生和理解的丰富认知心理学解释；机器人学家如Tani (2007)、Saegusa et al. (2008)、Park et al. (2012)、Pezzulo (2007)和Mohan, Morasso, et al. (2011)、Martius, Der, and Ay (2013)正在探索使用各种基于预测的学习程序，作为将高级认知功能建立在与世界感觉运动接触的坚实基础上的手段。

这种基于预测的学习和反应工作的显著繁荣既令人鼓舞又至关重要。因为只有通过考虑基于预测和生成模型的架构和策略的完整空间，我们才能开始向大脑和生物有机体提出真正有针对性的实验问题：这些问题有朝一日可能会支持其中一个模型（或者更可能的是，一个连贯的模型子集[2]）而非其他模型，或者可能揭示它们重要共同基础中的深层缺陷和失败。只有那时，我们才能发现，呼应Box和Draper值得赞扬的悲观情绪，这些模型实际上有多错误，或多有用。

第三组挑战涉及将这些理论扩展到直觉上”更高层次”的领域，包括长期规划、认知控制(cognitive control)、社会认知(social cognition)、意识体验和明确的、语言驱动的推理。在这里，本文至少提供了一些提示和建议——例如，将规划和社会认知与各种基于生成模型的模拟联系起来；将控制与上下文敏感的门控例程联系起来；将意识体验与内感受和外感受预测的精妙混合联系起来；将语言驱动的推理与精度权重的人工自我操作联系起来。但尽管有这些初步

的尝试，将这些理论扩展到此类领域仍然充其量是模糊的。（有关一些讨论，请参见 Fitzgerald, Dolan, & Friston, 2014; Harrison, Bestmann, Rosa, et al., 2011; Hobson & Friston, 2014; Hohwy, 2013, 第9-12章; Jiang, Heller, & Egner, 2014; King & Dehaene, 2014; Limanowski & Blankenburg, 2013; Moutoussis et al., 2014; Panichello, Cheung, & Bar, 2012; 以及 Seth, 2014）。也许最具挑战性的是，用预测、贝叶斯推理(Bayesian inference)和自估不确定性等更基本的过程来重构动机、价值和欲望（参见 Friston, Shiner, et al., 2012; Gershman & Daw 2012; Bach & Dolan 2012; Solway & Botvinick 2012; Schwartenbeck et al., 2014）。

第四组也是最后一组问题更具战略性和概念性。广泛依赖多层生成模型和自上而下预测的图景是否在某种程度上过度智力化了心智？我们是否正在倒退到一种过时的”笛卡尔式”观点，即心智是一个被隔离的内在舞台，充满了内在表征，却以某种方式与身体和世界提供的多种简化问题的机会疏远？正如我所论证的，没有什么比这更远离真相。相反，我们已经看到预测驱动学习和多层生成模型的使用如何直接服务于感知和行动这两大支柱，实现快速、流畅的上下文敏感反应形式。我试图表明，预测大脑是一个面向行动的*参与机器*，善于找到充分利用身体和世界的高效具身解决方案。大脑现在作为复杂节点出现在持续的双向流动中，其中内在（神经）组织对外在（身体和环境）因素和力量的持续重构开放，反之亦然。内在和外在在这里被锁定在持续的共同决定的交换模式中，因为预测代理持续选择它们接收的刺激。这种模式在更扩展的空间和时间尺度上重复，因为我们构建（并反复重构）缓慢但确实地构建我们的社会和物质世界。

因此揭示的大脑是一个不安分的、主动的器官，与身体和世界密切、持续地交换。如此装备的我们通过自我预测的感觉刺激的游戏，遇到一个充满意义、结构和机会的世界：一个为行动而解析的世界，充满未来，并被过去模式化。

[附录1]

[基础贝叶斯]

贝叶斯定理(Bayes theorem)提供了一种在响应新信息或证据时改变现有信念的最优方法。在感觉证据的情况下，它显示如何更新对某个假设的信念（例如，垫子上有只猫），作为感觉数据（例如，视网膜上的光线游戏，或更现实地说，从场景的主动探索中产生的展开光线游戏）被假设预测程度的函数。在这样做时，它假设某种先验信念状态，然后指定如何在新证据的光照下改变该信念。这允许随着越来越多的新证据到达，持续、理性地更新我们的背景模型（先验信念状态的来源）。

对于我们——诚然相当有限的——目的，这里的数学并不重要（尽管有一个很好的入门，请参见 Doya & Ishii, 2007；对于更非正式的内容，请参见 Bram, 2013，第3章；对于在人类认知中的应用讨论，请参见 Jacobs & Kruschke, 2010）。但数学所实现的是相当精彩的东西。它允许我们根据关于以下方面的背景信息来调整某些传入感觉数据的影响：（1）如果世界确实处于某种状态，获得该感觉数据的机会（即”给定假设的数据概率”，即数据被假设预测的程度）和（2）假设的先验概率(prior probability)（无论感觉数据如何，猫在垫子上的机会）。以正确的方式将所有这些结合在一起，会产生给定新感觉证据的假设修正（后验）概率的正确估计（基于你所知道的）。这里的好处是，一旦你将先验信念更新为后验信念，后验信念就可以作为下一次观察的先验。

这种运算很重要，因为未能考虑背景信息可能会严重扭曲对新证据影响的评估。经典例子包括如何在使用准确测试得到阳性结果时评估患某种疾病的概率，或在给定一些法医（例如DNA）证据时评估被告有罪的概率。在每种情况下，证据的恰当影响比我们直觉上想象的更强烈地取决于事先患病（或犯罪）的概率，而不依赖于那个特定证据。本质上，贝叶斯定理因此是一个将先验知识（例如基础比率）与Kahneman (2011)第154页所称的”证据的诊断性”——证据支持一个假设相对于另一个假设的程度——相结合的工具。一个直接含义是——正如宇宙学家卡尔·萨根著名地说过的——“非凡的声明需要非凡的证据”。

文中讨论的”预测处理”模型实现了正是这样一个”理性影响调整”过程。它们通过用一套基于系统对世界的了解以及对其自身感知和处理的情境变化可靠性的了解的自上而下概率预测来迎接传入的感官信号来做到这一点。贝叶斯定理有一个重新表述，特别清楚地表明了它与这种基于预测的感知模型的关系。这个重新表述（由Joyce, 2008翻译成散文）说：

以数据体为条件的两个假设的概率比等于它们的无条件[基线]概率比乘以第一个假设作为数据预测器超越第二个假设的程度。

你可能认为缺陷是，所有这些只有在你已经知道关于无条件概率（先验和”统计背景条件”）的所有那些信息时才会起作用。这就是所谓的”经验贝叶斯”（Robbins, 1956）能够做出特殊贡献的地方。因为在分层方案中，所需的先验本身可以从数据中估计，使用一个层次的估计来为下面的层次提供先验（背景信念）。在预测处理架构中，多层结构的存在以层次结构中一个层次对下面层次施加的约束形式诱导出这样的”经验先验”。这些约束可以使用标准梯度下降方法通过感官输入本身逐步调整。这种多层学习程序通过皮层的分层和相互连接的结构和布线看起来在神经上是可实现的（见Bastos et al., 2012; Friston, 2005; Lee & Mumford, 2003）。

然而需要一些谨慎的话。大脑执行某种形式的近似贝叶斯推理的概念越来越受欢迎。在其最广泛的形式中，这只需要意味着大脑使用生成模型（即体现关于任务统计结构的背景知识的模型）来计算给定当前感官证据时对世界的最

佳猜测（“后验分布”）。这样的方法不需要提供好的结果或最优推理。生成模型可能是错误的、粗糙的或不完整的，要么因为训练环境太有限或偏斜，要么因为真实分布太难学习或使用可用的神经回路无法实现。这样的条件迫使使用近似。寓意是”所有最优推理都是贝叶斯的，但不是所有贝叶斯推理都是最优的”（Ma, 2012，第513页）。换句话说，这里有一个很大的空间，需要仔细处理。有用的调查见Ma (2012)。

[附录2]

[自由能表述]

自由能表述起源于统计物理学，并在包括Hinton and von Camp (1993)、Hinton and Zemel (1994)、MacKay (1995)和Neal and Hinton (1998)在内的开创性处理中被引入机器学习文献。这样的表述可以说能够（例如，Friston, 2010）用来显示预测误差最小化策略*本身*是一个更基本命令的表现，即在系统与环境的交换中最小化热力学自由能的信息理论同构体。

热力学自由能是衡量可用于做有用功的能量的指标。转置到认知/信息领域，它表现为世界被表征（建模）的方式与世界实际存在方式之间的差异。不过这里需要小心，因为”世界被表征的方式”这个概念是一个滑溜的野兽，不能理解为暗示一种关于模型与世界之间隐含匹配的被动（“自然之镜”，见Rorty, 1979）故事。正如我们在本文中详细看到的，一个好模型的检验标准是它能多好地使有机体在滚动的行动循环中与世界互动，从而维持自己在可行性窗口内。互动越好，信息论自由能就越低（这是直观的，因为系统的更多资源被用于建模世界的”有效工作”）。预测误差报告了这种信息论自由能，它在数学上被构造为总是大于”惊讶度”（这里指的是给定世界模型下某些感觉状态的亚个人计算的不合理性，见Tribus, 1961）。熵，在这种信息论的表述中，是惊讶度的长期平均值，而减少信息论自由能相当于改进世界模型以减少预测误差，从而减少惊讶度（更好的模型做出更好的预测）。总体原理是，好的模型（根据定义）是那些帮助我们成功与世界互动的模型，因此帮助我们维持结构和组织，使我们在延长但有限的时间尺度上看起来抵抗熵的增加和（因此）热力学第二定律。

“自由能原理”本身然后声明”所有可以改变的量；即系统的组成部分，都将改变以最小化自由能”(Friston & Stephan, 2007, p. 427)。注意，这样表述的话，这是关于系统组织的所有要素（从总体形态到大脑的整个组织）的声明，而不仅仅是关于皮层信息处理。使用一系列优雅的数学公式，Friston (2009, 2010)建议，当这个原理应用于神经功能的各种要素时，会导致产生高效的内部表征方案，并揭示本文中探索的感知、推理、记忆、注意和行动之间联系背后的最深层原理。如果这个故事是正确的，形态学、行动倾向（包括环境生态位(niche)的主动结构化）和总体神经架构都是这个单一原理在不同时间尺度上运作的表达。

自由能解释具有很大的独立意义。它代表了关于感知和行动计算亲密性声明的一种”最大版本”，并且至少暗示了一个可能容纳对理解生命与心智之间复杂关系日益增长兴趣的一般框架（见，例如，Thompson, 2010）。本质上，这样的希望是阐明生物系统中自组织的可能性本身（见，例如，Friston, 2009, p. 293，以及第9章中的讨论）。

然而，对自由能原理及其在理解生命和心智方面的潜在应用的充分评估远远超出了本处理的范围。

注释

引言

1. Hermann von Helmholtz (1860)将感知描述为涉及概率推理。在Helmholtz门下学习的James (1890)也将感知描述为利用先验知识来帮助处理不完美或模糊的输入。这些洞察为心理学中的”分析-综合”范式提供了信息（见Gregory, 1980; MacKay, 1956; Neisser, 1967；综述见Yuille & Kersten, 2006）。Helmholtz的洞察也在一个重要的计算和神经科学工作体系中得到了追求（正如我们将在第1章中看到的）。对这一谱系至关重要的是机器学习中的开创性进展，始于反向传播学习的开创性连接主义工作（McClelland et al., 1986; Rumelhart et al., 1986），并延续到恰当命名的”Helmholz机器”的工作（Dayan et al., 1995; Dayan & Hinton, 1996; 另见Hinton & Zemel, 1994）。有用的综述见Brown & Brüne, 2012; Bubic et al., 2010; Kveraga et al., 2007。另见Bar et al., 2006; Churchland et al., 1994; Gilbert and Sigman, 2007; Grossberg, 1980; Raichle, 2010。

2. Barney当时正在与Dennett在课程软件工作室工作，该工作室是Dennett于1985年在塔夫茨大学共同创立的。项目顾问是塔夫茨大学地质学家Bert Reuss。

3. 这个评论是在Danny Scott的采访（“Racing to the Bottom of the World”）中做出的，发表在《星期日泰晤士报杂志》，2011年11月27日，第82页。

[4.] 这些情况有时被描述为”心因性”、“非器质性”、“无法解释”，甚至（在较早的说法中）“歇斯底里”。我从 Edwards et al. (2012) 借用了”功能性运动和感觉症状”这一术语。

第1章

[1.] 例子包括 Biederman, 1987; Hubel & Wiesel, 1965; 和 Marr, 1982。

[2.] 我有时会写道，作为一种简化说法，预测在这些系统中向下流动（从较高区域流向感觉外围）。这是正确的，但重要的是不完整的。它突出了这样一个事实：预测，至少在预测处理(PP)的标准实现中，从每个较高层次流向紧邻的下一层。但大量的预测信息也在层次内部横向传递。因此，向下流动的预测应该被理解为表示”向下和横向流动”的预测。感谢比尔·菲利普斯建议我在开始时澄清这一点。

[3.] 我看到这个格言被不同地归因于机器学习先驱马克斯·克洛斯，以及神经科学家鲁道夫·利纳斯和拉梅什·贾因。然而，“受控幻觉”的表述可能使我们对现实的感知把握看起来脆弱且令人不安地间接。相反，在我看来——尽管为此论证必须等到后面——这种感知观详细展示了感知（或更好地说，感知-行动回路）如何使我们与环境的突出方面建立真正的认知联系。因此，我稍后（第6章）会建议，将幻觉视为一种”不受控制的感知”形式可能更好。

[4.] 在实践中，从随机权重分配和非问题特定架构开始的简单反向传播网络往往学习非常缓慢，并且经常陷入所谓的局部最小值。

[5.] 这种诱惑的著名受害者包括乔姆斯基、福多，以及在较小程度上的平克（见他的1997年著作）。

[6.] 我指的是，以在输入和输出之间增加所谓”隐藏单元”层数的形式。关于这些困难的很好讨论，见 Hinton, 2007a。

[7.] 这里既有胡塞尔的回响，也有梅洛-庞蒂的回响。关于一些很好的讨论，见 Van de Cruys & Wagemans, 2011。

[8.] “分析通过综合”是一种处理策略，其中大脑不是通过自下而上地积累大量低级线索来构建其当前的世界原因模型。相反，大脑试图从其最佳的相互作用世界原因模型中预测当前的感觉线索套件（见 Chater & Manning, 2006; Neisser, 1967; Yuille & Kersten, 2006）。

[9.] 关于一个很好的论述，为基于预测的模型辩护，见 Poeppel & Monahan, 2011。

[10.] 这是对”最大似然学习”的一个计算上可处理的近似，如 Dempster et al. (1977) 的期望最大化(EM)算法中所使用的。

[11.] 不熟悉这个概念的读者可能想要重新审视引言中提出的SLICE的非正式例子。

[12.] 你可以在辛顿的网站上看到网络的运行：http://www.cs.toronto.edu/~hinton/digits.html。

[13.] 差异主要涉及允许和不允许的消息传递方案类型，以及在学习和在线性能期间使用和结合自上而下和自下而上影响的精确方式。例如，在辛顿的数字识别工作中，基于生成模型的预测在学习期间起关键作用。但这种学习提供的是用于快速在线判别的更简单（纯粹前馈）策略。我们稍后将看到（在第三部分中）预测处理系统如何流畅灵活地适应这种简单、低成本策略的使用。然而，所有这些方法共享的是，至少在学习期间，在多层次设置中使用生成模型的核心重点。关于辛顿和同事（通常更”工程驱动”）在”深度学习”系统上工作的有用介绍，见 Hinton, 2007a; Salakhutdinov & Hinton, 2009。

[14.] 关于这些失败的持续讨论，以及联结主义（和后联结主义）替代方案的吸引力，见 Bermúdez, 2005; Clark, 1989, 1997; Pfeifer & Bongard, 2006。

[15.] 见 Kveraga, Ghuman, & Bar, 2007。

[16.] 关于这个重要且更大空间的选择性指导，见第10章。

[17.] 据我所知，刚才复述的基本故事仍然被认为是正确的。但关于一些复杂性，见 Nirenberg et al., 2010。

[18.] 见 Alais & Blake, 2005 中的文章，以及 Leopold & Logothetis, 1999 的综述文章。关于一个优秀的介绍，见 Schwartz et al., 2012。

[19.] 这些方法使用它们自己的目标数据集来估计先验分布：一种利用表征层次模型的统计独立性的自举法。

[20.] 关于进一步讨论，见 Hohwy, 2013。

[21.] 这并不是 Pearl (1988) 意义上的”解释消除”，在那种情况下，该术语指的是确认一个原因会减少援引替代原因的需要的效应（当假设1增加其后验概率时，假设2会减少其后验概率，即使它们在证据到达之前是独立的）。这也不是与从本体论中消除不必要实体相关的”解释消除”。相反，这个想法很简单：被良好预测的感觉信号不被视为值得关注的新闻，因为它们的含义已经在系统响应中得到了解释。感谢 Jakob Hohwy 指出这一点。

[22.] 然而，请注意，前向处理中明显的效率在这里是以多层级生成机制本身为代价获得的：该机制的实现和操作需要一整套额外的连接来实现双向级联的向下扫动。

[23.] 选择性锐化和抑制的一致性也使得区分预测编码和”证据累积”解释（如 Gold and Shadlen (2001)）的经验含义变得更加困难——尽管并非不可能。相关综述见 Smith & Ratcliff, 2004。相关尝试见 Hesselmann et al., 2010。

[24.] 最简单的建议可能是这种分离本质上是时间性的，在不同的处理阶段涉及相同的单元。然而，这样的提议需要解决一系列技术问题，而这些问题并不影响更标准的建议（如 Friston 的建议）。

[25.] 最近也提出了实验测试（Maloney & Mamassian, 2009；Maloney & Zhang, 2010），旨在”操作化”目标系统（真正）使用贝叶斯方案计算其输出的主张，而不仅仅是表现得”仿佛”它这样做了。然而，这是一个值得大量进一步思考和研究的领域（相关精彩讨论见 Colombo & Seriès, 2012）。

[26.] Potter et al. (2014) 显示，理解视觉呈现场景的概念要旨（例如，“微笑的夫妇”或”野餐”或”有船的港口”）所需的最短观看时间可以短至13毫秒。这太快了，无法建立新的前馈-反馈循环（“再入循环”）。此外，结果并不依赖于受试者在看到图像*之前*被告知目标标签（从而控制了在图像呈现前可能已经存在的特定自上而下期望）。这表明这些类型的要旨确实可以（如 Barrett 和 Bar 所建议的）通过超快速前馈扫动提取。

[27.] 这意味着当我们从生态学上奇怪的实验室条件进行概括时需要非常小心，这些条件实际上剥夺了我们这种持续的背景信息。相关近期讨论见 Barrett & Bar, 2009；Fabre-Thorpe, 2011；Kveriga et al., 2007。

[28.] 这种效应在文献中早已为人所知，它们首先出现在感觉习惯化的研究中，最突出的是 Eugene Sokolov 关于定向反射的开创性研究（1960）。更多内容见第3章。

[29.] 关于这项近期工作的精彩讨论，见 de-Wit et al., 2010。

[30.] 可能的替代实现在 Spratling and Johnson (2006) 和 Engel et al. (2001) 中有讨论。

[31.] 这是一种相对神经活动（“大脑激活”）的测量，通过血流和血氧水平的变化来指示。假设是神经活动产生代谢成本，而这个信号反映了这种成本。人们普遍承认（例如，见 Heeger & Ross, 2002），与单细胞记录相比，这是一种相当间接的、基于假设的和”钝性”的测量。尽管如此，新形式的多变量模式分析能够克服使用这种技术的早期工作的一些局限性。

[32.] 作者指出，这可能是由于表征和错误检测之间处理成本的某些基本代谢差异，或者可能是由于其他原因，BOLD 信号更多地跟踪区域的自上而下输入而不是自下而上的输入（见 Egner at al., 2010, p. 16607）。

[33.] 这些观点强调”及时”主动引发任务相关信息以供使用。例子包括 Ballard, 1991；Ballard et al., 1997；Churchland et al., 1994。相关讨论见 Clark, 1997, 2008。我们将在第4到第9章回到这些主题。

[34.] 关于这一大量文献的窗口，见 Raichle, 2009；Raichle & Snyder, 2007；Sporns, 2010，第8章；Smith, Vidaurre, et al., 2013。另见第8章和第9章中关于自发神经活动的讨论。

[35.] 这种经济性和准备性在生物学上具有吸引力，并巧妙地回避了与更被动的信息流模型相关的许多处理瓶颈（见 Brooks, 1991）。在预测性处理(predictive processing)中，预测的向下流动承担了大部分计算”重负”，使得逐时刻处理只需关注由显著的（高精度的，见第2章）预测误差所标示的值得关注的偏离。我们将在第4章到第9章中讨论这些问题。

第2章

[1.] 想要另一个涉及运动且需要使用网络浏览器查看的新例子，请尝试 http://www.michaelbach.de/ot/cog-hiddenBird/index.html。

[2.] 这样构建的”信念”最好理解为由”概率密度函数”(probability density functions, PDFs)所诱导的，这些函数描述了某个连续随机变量取给定值的相对可能性。随机变量(random variable)简单地将数字分配给结果或状态，可以是离散的（如果可能的值位于不同点）或连续的（如果它们在区间上定义）。

[3.] 这种感觉不确定性的灵活计算是预测性处理和”卡尔曼滤波”(Kalman filtering)之间的共同基础（见Friston, 2002; Grush, 2004; Rao & Ballard, 1999）。

[4.] 回忆在这些模型中，误差单元从其自身层级的表示单元和上层的表示单元接收信号，而表示单元（有时称为”状态单元”）由同一层级和下层的误差单元驱动。见Friston, 2009，Box 3，第297页。

[5.] 所谓的”同步增益”(synchronous gain)（其中同步的突触前输入修改突触后增益；见Chawla, Lumer, & Friston, 1999）在这里也可能起作用。正如Friston (2012a)指出的，最近还有一些推测，涉及快速（伽马）和慢速（贝塔）频率振荡在分别传递自下而上的感觉信息和自上而下的预测方面的可能作用。这为浅层和深层锥体细胞的活动提供了很好的映射。对这些有趣的动态可能性的研究仍处于起步阶段，但见Bastos et al., 2012; Friston, Bastos, et al., 2015; Buffalo et al., 2011；以及Engel, Fries, & Singer, 2001中的更一般性讨论。另见Sedley & Cunningham, 2013，第9页。

[6.] Hohwy (2012)引用了Helmholtz 1860年《生理光学论文集》中的一段精彩引文，很好地捕捉了这种体验。引文（由Hohwy翻译）如下：

我们注意力的自然不受约束状态是游荡到不断新的事物上，所以当一个物体的兴趣被耗尽，当我们无法感知任何新东西时，注意力就会违背我们的意愿转向别的东西。……如果我们想让注意力停留在一个物体上，我们必须不断在其中发现新的东西，特别是当其他强烈感觉试图将其分离时。（Helmholtz, 1860，第770页；由JH翻译）

[7.] 在这里，这样的分配有时也可能误导，变化盲视(change blindness)和无意盲视(inattentional blindness)等现象的合理解释可以构建在不同类型和信息通道的精度加权变化基础上，见Hohwy (2012)。

[8.] 这些情况也可能涉及另一种与情绪相关的预测（在这种情况下是恐惧和唤醒）——关于我们自身生理状态的内感受预测(interoceptive predictions)。这些进一步的维度在第7章中讨论。

[9.] 我这里专注于视觉，但这些观点应该同样适用于其他感觉方式（例如，考虑我们通过触摸探索熟悉物体的方式）。

[10.] 尽管与PP风格方法有这种核心共性，但重要的差异（特别是关于奖励和强化学习的作用）将这些解释区分开来。见Friston, 2011a中的讨论；以及第4章。

[11.] 因此它不是Tatler, Hayhoe et al. (2011)所拒绝的那种基于”显眼性”(conspicuity)的地图。

[12.] 该模拟只捕捉了获胜假设对眼跳模式的直接影响。一个更复杂的模型需要包括精度神经估计的影响，驱动模拟智能体根据预期最可靠（不仅仅是最具区别性）感觉信息的位置来探测场景。

[13.] 例如，反映先前选择的购买建议可能导致新购买的循环（以及相应调整的推荐），逐步巩固和缩小我们自身兴趣的范围，使我们成为自己可预测性的受害者。关于一些讨论，见Clark, 2003，第7章。

[14.] 实际上，刺激对比度可以用作操纵精度的代理。这方面的经验证据是突触后增益的变化——在注意调节过程中看到的那种——由改变亮度对比度诱发。见Brown et al., 2013。

[15.] 类似的例子，见Friston, Adams, et al., 2012，第4页。

[16.] 现在需要更好地理解这些相互作用的多重机制（各种缓慢的神经调节剂可能与神经同步化复杂协调作用），同时彻底检验预测流动和预测误差权重（精确度）的调节效应可能表现的各种方式和层面（参见Corlett et al., 2010; Friston & Kiebel, 2009; Friston, Bastos et al., 2015）。另见Phillips & Silverstein (2013); Phillips, Clark, & Silverstein (2015)。

[17.] 有趣的是，作者们还能够将该模型应用于一种非药物干预：感觉剥夺。

[18.] Feldman and Friston (2010) 指出，精确度表现得仿佛它本身就是一种有限资源，因为提高某些预测误差单元的精确度需要降低其他单元的精确度。他们还有趣地评论道（同上第11页），“精确度表现得像资源的原因是，生成模型包含先验信念，即对数精确度以情境敏感的方式在感觉通道间重新分配，但在所有通道中保持守恒。”显然，这种’信念’绝不会被显式编码，而必须以某种方式内在于系统的基本结构中。这提出了重要问题（关于生成模型中什么是和什么不是显式编码的），我们将在第8章回到这个问题。

第3章

[1.] 另见DeWolf & Eliasmith, 2011，第3.2节。有趣的是，即使在身体运动被最小化时，关键的签名要素仍然保持，比如使用外科微型工具作为非标准工具来签名。经验丰富的显微外科医生能够一次成功，并显示出与正常书写相同的笔迹风格（除了在40倍放大下，他们可能会被纸张的单个纤维缠住！）；见Allard & Starkes, 1991，第148页。这样的壮举证明了我们使用现有元素和知识构建新技能表现的卓越能力。

[2.] Sokolov, 1960；另见Bindra, 1959；Pribram, 1980；和Sachs, 1967。

[3.] 见Sutton et al., 1965；Polich, 2003，2007。

[4.] 对于这个本质上功能性的层级或水平概念的正确神经元解释，我故意保持不确定的态度。但对于一些推测，见Bastos et al., 2012。

[5.] 当然，它们并不在这些描述下识别它们！

[6.] 我不试图具体说明哪些动物属于这个广泛的范畴，但双向新皮质回路的存在（或者——这显然更难确定——这种回路的功能类似物）应该提供一个好线索。

[7.] 体素(voxel)是”体积像素”。每个体素跟踪一个有限三维空间中的活动。单个fMRI体素通常代表相当大的体积（通常在3毫米×3毫米×3毫米立方体内为27立方毫米）。

[8.] 实际上，实验者的任务更困难，因为原始fMRI数据最多只能提供潜在神经活动本身的粗略阴影（由血动力学反应构建）。

[9.] 然而，请记住每个体素覆盖大量（约27立方毫米）的神经组织。

[10.] 感谢Bill Phillips的建议（个人交流）。

[11.] 这里可能也有空间进行一些关于”析取主义”(disjunctivism)的启发性讨论——粗略地说，这是真实知觉和幻觉不共享共同种类的观点。当然，很多取决于析取主义主张如何被阐释。关于可能表述的相当全面的样本，见Haddock & Macpherson, 2008和Byrne & Logue, 2009中的论文。

[12.] 这段话浓缩并稍微简化了在Hobson (2001)和Blackmore (2004)中发现的观点。另见Roberts, Robbins, & Weiskrantz, 1998；和Siegel, 2003。

[13.] 正如Hobson和Friston所指出的，这样的两阶段过程正是Hinton等人恰当命名的觉醒-睡眠算法(wake-sleep algorithm)的核心（见Hinton et al., 1995，以及第1章中的讨论）。

[14.] 见第2章以及第5章和第8章中的进一步讨论。

[15.] 这个二阶（“元记忆”(meta-memory)）组件对于解释熟悉感而无情节性回忆的感觉是必要的。（例如，想想识别一张脸为熟悉的，尽管没有与该个体任何特定遭遇的情节性记忆。）

[16.] 关于使用预测处理风格资源来处理记忆的另一次尝试，见Brigard, 2012。关于不同策略，见Gershman, Moore, et al., 2012。

[17.] 关于这种类型解释的另一个例子，见van Kesteren et al., 2012。

[18.] 例如，一些模型元素必须考虑伴随身体运动的感觉信号中的微小变化，这些信号有助于指定场景的视角，而其他元素则跟踪更加不变的项目，如感知对象的身份和类别。但正是这些层级之间精度调节的相互作用，通常由某种形式的持续运动行为所中介（见第二部分），现在成为智能、适应性反应的核心。

第4章

[1.] James, 1890, vol. 2, p. 527.

[2.] 这种感知到的对立在以下处理中显而易见：Anderson and Chemero (2013), Chemero (2009), Froese and Ikegami (2013)。更加包容的观点在以下处理中得到辩护：Clark (1997, 2008)。

[3.] 另见Weiskrantz, Elliot, and Darlington (1971)，他们在那里引入了被相当可爱地称为”标准挠痒设备”的装置，允许轻松地对自我和他人诱发的挠痒进行实验比较。

[4.] 见Decety, 1996; Jeannerod, 1997; Wolpert, 1997; Wolpert, Ghahramani, & Jordan, 1995; Wolpert, Miall, & Kawato, 1998。

[5.] 通过将标准解释吸收到基于预测处理的更一般框架中，可以为此类效应提供相关（但更简单和更一般）的解释。根据这种解释，我们之所以不会因为眼球运动就看到世界在移动，是因为由此产生的感觉流最好用一个内部模型来预测，该模型的最高层级描绘了一个可用于运动采样的稳定世界。

[6.] 在那个更完整的解释中，自发行为期间的下行精度期望会降低报告这些行为感觉后果的预测误差的音量或增益。实际上，我们因此暂时暂停对许多由我们自己造成的感觉扰动的注意（见Brown et al., 2013；以及第7章中的进一步讨论）。

[7.] 这些是解决身体动力学、系统噪声和所需结果准确性的成本函数(cost functions)。见Todorov, 2004; Todorov & Jordan, 2002。

[8.] 见Adams, Shipp, & Friston, 2012; Brown et al., 2011; Friston, Samothrakis, & Montague, 2012。

[9.] 完整的故事见Shipp et al., 2013。简而言之，“来自运动皮层的下行投射与视觉皮层中的自上而下或反向连接具有许多共同特征；例如，皮质脊髓投射起源于颗粒下层，高度发散，并且（连同下行皮质-皮质投射）靶向表达NMDA受体的细胞”（Shipp et al., 2013, p. 1）。

[10.] Anscombe的目标是欲望和信念之间的区别，但她关于适合方向的观察可以推广（正如Shea, 2013很好地指出的）到行动的情况，这里被理解为某些形式欲望的运动结果。

[11.] 因此，两个常被引用的成本是”噪声”和”努力”——流畅的行动似乎依赖于这些因素的最小化。见Harris & Wolpert, 1998; Faisal, Selen, & Wolpert, 2008; O’Sullivan, Burdet, & Diedrichsen, 2009。

[12.] 回想一下，这里使用的术语”信念”涵盖指导感知和行动的生成模型的任何内容。这样的信念不需要是主体可访问的，并且如前所述，通常被理解为概率密度函数(probability density functions)的表达，描述某个连续随机变量假设给定值的相对可能性。例如，在行动选择的背景下，PDF可能以连续的方式指定转换到某个后继状态的当前概率。

[13.] 最终，这一切归结为主动构建、选择性采样和选择性表征的结合。完整的图景直到我们文本的第三部分才会显现。但当所有这些因素共同作用时，认知主体锁定在周围（社会、物理和技术）环境的有机体相关结构上，在构建世界和被世界约束之间实现微妙的、维持生命的平衡行为。这与Varela、Thompson和Rosch在其经典著作（1991）《具身心智》中强调的平衡行为相同。

[14.] 我从我最喜欢的哲学家之一威拉德·范·奥曼·奎因那里借用了这个短语，他在评论一个相当臃肿的本体论时使用了它，写道：“怀曼人口过多的宇宙在许多方面都不可爱。它冒犯了我们这些对沙漠景观有品味的人的美感”（Quine, 1988, p. 4）。

[15.] 关于这种一般性策略的一些关键担忧，请参见Gershman & Daw, 2012。Gershman和Daw实际上担心，将成本和效用折叠为期望值提供了过于粗糙的工具，因为这使得很难看出意外事件（如中彩票）如何能被视为有价值的。这种担忧低估了预测和期望支持者可获得的资源。适应性智能体应该期望能够在某些时候从惊喜和环境变化中受益。这

表明两种方法在概念上可能并不像最初看起来那么不同。它们甚至可能是相互可翻译的，因为一方可能希望用效用和价值的话语表达的一切，另一方都可以用更高层次（更抽象和灵活）期望的话语来表达。关于这些及相关问题的更多内容，请参见第三部分。

[16.] 请参见Friston, 2011a；Mohan & Morasso, 2011。

[17.] 关于一些讨论，请参见Friston, 2009，第295页。

[18.] 这种观点的表面激进性主要取决于与将多个配对的前向和逆向模型假设为流畅、灵活运动行为基础的工作的对比。正如我们所看到的，另一些方法（如Feldman, 2009；Feldman & Levin, 2009）假设例如平衡点或参考点作为其核心组织原则，与基于本体感受预测的过程模型提供了非常自然的匹配（进一步讨论请参见Friston, 2011a；Pickering & Clark, 2014；Shipp et al., 2013）。

[19.] 在很大程度上（但请参见Mohan & Morasso, 2011；Mohan, Morasso, et al., 2013），认知发展机器人学(Cognitive Developmental Robotics, CDR)的工作保留了配对前向/逆向模型、传出副本(efference copy)以及价值和奖励信号的经典结构。然而，CDR实验是有启发性的，因为它们表明基于预测误差的编码可以以开启模仿学习大门的方式构建运动动作。

[20.] 我认为，预测处理(Predictive processing)在满足这一职责方面处于理想位置，因为它将强大的学习程序与涵盖感知和行动的完全统一的视角结合起来。

[21.] 在Park等人的工作中，这种早期学习是通过使用自组织特征图(Kohonen, 1989，2001)实现的。这与Hebbian学习密切相关，但它添加了一个”遗忘项”来限制学习关联的潜在爆炸。这种学习的另一种方法部署了修改的循环神经网络（“带参数偏置的循环神经网络”，参见Tani et al., 2004）。

[22.] 动态拟人化智能机器人—开放平台，来自弗吉尼亚理工机器人与机构实验室。请参见http://www.romela.org/main/DARwIn_OP:_Open_Platform_Humanoid_Robot_for_Research_and_Education。

第5章

[1.] 当然，它们也可能被简单地视为需要开发全新模型的进一步现象—这是一种效率较低的策略，可能在遇到陌生和异域事物时被要求采用，或者也许（参见Pellicano & Burr, 2012，以及Friston, Lawson, & Frith 2013的评论）由于上下文/预测机制本身的故障。另请参见[第7章]。

[2.] 请参见McClelland & Rumelhart, 1981，关于更新的”多项交互激活”模型，请参见Khaitan & McClelland, 2010；McClelland, Mirman, et al., 2014。

[3.] 这种形式的连接主义与PP范式中的工作之间存在深层联系。关于一些探索，请参见McClelland, 2013和Zorzi et al., 2013。

[4.] 关于另一种实现方式，尽管保留了PP模型的关键特征，并且也实现了功能层次结构，请参见Spratling (2010，2012)。

[5.] 回顾一下，经验先验(empirical priors)是与层次模型中中间层次相关的概率密度或”信念”。

[6.] 请参见Aertsen et al., 1987；Friston, 1995；Horwitz, 2003；Sporns, 2010。

[7.] Sporns (2010，第9页)建议结构连接在”秒到分钟的尺度”上保持稳定，而有效连接的变化可能在数百毫秒的量级上发生。

[8.] 关于这些方法的有用介绍，请参见Stephan, Harrison, et al., 2007和Sporns, 2010。

[9.] 反映可靠性和显著性上下文的期望因此以与关于其他世界或身体属性和事态的知识相同的方式获得。可靠性和显著性的估计本身由感官数据与整体生成模型的当前交互决定，导致”估计自己对可靠性估计的可靠性”这一日益棘手的问题（关于一些很好的讨论，请参见Hohwy, 2013，第5章）。

[10.] 正如通过各种形式的网络分析(network analysis)所实现的，例如能够揭示神经元之间以及神经元群体之间功能性和有效连接模式的分析方法（有关丰富而全面的综述，参见Sporns, 2002）。另见Colombo, 2013。

[11.] 因此，这样构建的大脑是”易变的”，包含”一个由有效连接以非线性方式耦合的功能专门化区域的集合”（Friston & Price, 2001, p. 277）。不过，在Friston和Price谈到”功能专门化”区域的地方，我认为最好遵循Anderson (2014, pp. 52 – 53)的观点，说成是”功能分化的”区域。

[12.] 这些描述与几个相关提议有着共同的核心承诺（综述见Arbib, Metta, et al., 2008, section 62.4, 和Oztop et al., 2013）。HMOSAIC模型（Wolpert et al., 2003）利用了同样类型的分层预测驱动多层模型作为支持动作模仿的手段，Wolpert等人进一步建议这种方法同样可能解决关于动作理解和”意图提取”的问题。

[13.] 为了便于阐述，我有时会说我们对不同领域掌握着不同的生成模型。然而，在数学上，这总是可以表述为对单一整体生成模型的进一步调节。这在Friston及其同事的各种处理中是清楚的，例如参见Friston, 2003。

[14.] 这里的建议并不是说我们的躯体感觉区域在动作观察期间本身变得不活跃。事实上，有相当多的证据（在Keysers et al., 2010中综述）表明这些区域在被动观察期间确实是活跃的。相反，它是说本体感觉预测误差的前向流动影响（相关方面）现在变得缓和，使得这种误差无法影响躯体感觉处理的更高水平。正是这样的模式（低层活动结合高层不活跃）被Keysers et al. (2010)发现。另见Friston, Mattout, et al., 2011, p. 156。

[15.] 正如Jakob Hohwy（个人通信）巧妙地指出的那样，这不是一个幸运或临时的技巧。它只是学习估计精度(precision)的最佳方式以最小化预测误差的另一种表现。

[16.] 关于这个主题有大量而复杂的文献，有用的综述见Vignemont & Fourneret, 2004，以及对基本概念的一些重要细化。另见Hohwy, 2007b。

[17.] 例如参见Friston, 2012b; Frith, 2005; Ford & Mathalon, 2012。

[18.] 参见Barsalou, 1999, 2009; Grush, 2004; Pezzulo, 2008; Pezzulo et al., 2013; 另见Colder, 2011; Hesslow, 2002。

[19.] 在这项工作中，一个模糊图形（著名的面孔-花瓶图形）以与梭状回面孔区域(FFA)中自发刺激前活动共变的方式被看作面孔或花瓶。然而，正如作者所指出的，“尚不清楚这里报告的FFA信号持续活动的变化是否类似于以静息状态网络形式描述的缓慢波动”（Hesselmann et al., 2008, p. 10986）。

[20.] 某些睡眠状态以及也许某些实践（如冥想和吟诵）可能是这个规则的例外。

[21.] 这些对我们自己不确定性的估计在各种最新的计算和神经科学工作中发挥着核心作用。例如参见Daw, Niv, & Dayan, 2005; Dayan & Daw, 2008; Knill & Pouget, 2004; Yu & Dayan, 2005; Daw, Niv, & Dayan 2005; Dayan & Daw, 2008; Ma, 2012; van den Berg et al., 2012; Yu & Dayan, 2005。

第6章

[1.] 如第1章所述，“感知作为受控幻觉”这个短语被各种归因于Ramesh Jain、Rodolfo Llinas和Max Clowes。Rick Grush在他关于”表征的模拟理论”的工作中也引用了它（例如，Grush, 2004）。

[2.] 鉴于预测处理故事的标准实现——关于替代实现，参见Spratling, 2008。

[3.] 虽然回想一下（在1.13中）超快前馈传递在揭示视觉呈现场景的概念要旨方面的惊人能力的讨论（Potter et al., 2013）。

[4.] 关于这方面的例子，参见6.9。

[5.] 可供性(affordance)的概念被广泛使用和广泛解释。在其创始人J. J. Gibson手中，可供性是远端环境中存在的与主体相关的行动可能性（尽管主体不一定认识到）（参见Gibson, 1977, 1979）。有关仔细、细致的讨论，参见Chemero, 2009，第7章。

[6.] 另见Kemp et al., 2007。

[7.] 这里的过程层次对应于Marr (1982)所描述的算法层次。

[8.] 经典批评是Fodor and Pylyshyn (1988)提出的，但相关观点也被更具包容性的理论家提出，如Smolensky (1988)，他后来关于最优性理论和和谐语法的工作(Smolensky & Legendre, 2006)同样兼容生成结构和统计学习。关于这一重要问题的进一步讨论，请参见Christiansen & Chater, 2003。

[9.] 关于层次贝叶斯方法这一吸引人特征的精彩讨论，请参见Tenenbaum et al., 2011。然而，仍需谨慎，因为多种知识表征形式*能够*在此类模型中共存这一事实，并不能详细地向我们展示这些不同形式如何在统一的问题解决过程中有效结合。

[10.] 综述请参见Tenenbaum et al., 2011。但对于应用于高级认知(higher-level cognition)的一些主张的重要批评，请参见Marcus & Davis, 2013。

[11.] 正如Karl Friston（个人通信）所建议的。

[12.] 例如，一个系统可能从一组所谓的”感知输入分析器”(Carey, 2009)开始，其作用是使一些输入特征在学习中更加突出。关于HBM学习和此类简单偏向的结合效应的讨论，请参见Goodman et al., in press。

[13.] 这种”行动导向”范式与来自发展心理学(Smith & Gasser, 2005; Thelen et al., 2001)、生态心理学(Chemero, 2009; Gibson, 1979; Turvey & Shaw, 1999)、动力系统理论(Beer, 2000)、认知哲学(Clark, 1997; Hurley, 1998; Wheeler, 2005)和现实世界机器人学(Pfeifer & Bongard, 2006)的洞察相关。关于这些主题的有用采样，请参见Núñez & Freeman, 1999中的文章。

[14.] Wiese (2014)对这个问题有很好的讨论。

[15.] 对强大（且往往相当抽象）的超先验(hyperpriors)的此类诉求显然将构成任何更大的、广义贝叶斯的关于人类经验形状的故事的重要组成部分。尽管如此，关于此类超先验的*存在*或*作用模式*都不需要特殊的解释。相反，它们在我们一直考虑的层次模型中很自然地出现，在那里它们可能是先天的（给它们一种几乎康德式的感觉）或以经验（层次）贝叶斯的方式获得。

[16.] 参见Ballard (1991)、Churchland et al., (1994)、Warren (2005)、Anderson (2014)，第163-172页。

[17.] 因此Hohwy et al. (2008)指出，“当涉及到基本感知推理时，’预测’和’假设’等术语听起来相当智识主义(intellectualist)。但其核心是，系统唯一的处理目标就是简单地最小化预测误差或自由能，事实上，关于假设和预测的谈论可以被翻译成这样一个不那么拟人化的框架[并且]使用相对简单的神经基础设施来实现”(Hohwy et al., 2008，第688-690页)。

[18.] 人们甚至可能否认恶魔式操作实际上欺骗了我们。相反，它们只是为关于外部现实的同样古老的真实知识建议了一个替代基础：一个由桌子、椅子、棒球比赛等构建的世界。关于这种回应，请参见Chalmers (2005)。

[19.] 通常，这些丰富的内部模型涉及符号编码，使用复杂的类语言知识结构来描述事态。PP并不暗示任何类似的东西。

[20.] 这种错置担忧的一个版本也出现在Froese & Ikegami, 2013中。

[21.] 感谢Michael Rescorla提供这个有用的术语建议。

[22.] 因此Friston在第1章中已经引用的一段话中建议，“现实世界的层次结构实际上被试图最小化预测误差的层次架构所’反映’，不仅在感觉输入层面，而且在层次的所有层面”(Friston, 2002，第238页)。

[23.] 这是一个弱意义，因为不能保证部署这种策略的存在的潜在思考将形成所谓的通用性约束(Generality Constraint)所要求的那种封闭集合（包含组成把握的所有可能组合）(Evans, 1982)。

[24.] 例如，参见Moore, 1903/1922; Harman, 1990。

[25.] 在这种情况下，我们在感知上被误导，尽管我们作为反思主体知道得更好。在更严重的情况下（妄想和幻觉的情况），我们可能在每个层面都被误导。

[26.] 例如，更广泛的生态学视角可能将橡胶手错觉及其同类揭示为作为一个易变的生物系统不可避免的代价，这个系统的部分可能生长、变化，并遭受各种形式的意外损伤和/或增强，这并非不可想象。

[27.] 关于这个错觉的精彩介绍及其配套图形，请参见 http://psychology.about.com/od/sensationandperception/ss/muller-lyer-illusion.htm。

[28.] 高级智能体(agent)可能会在这些模糊的元认知水域中管理一两个步骤。例如，参见 Daunizeau, den Ouden, et al., 2010a, b。

第7章

[1.] 关于元认知(metacognition)和预测误差的许多一般性问题的精彩讨论，请参见 Shea (2013)。该讨论以”奖励预测误差”为框架，但据我了解，所有关键观点都适用于本文讨论的更一般的感官预测误差情况。

[2.] 这个图解首次出现在 Frith and Friston (2012) 中。

[3.] 关于近期文献的回顾，请参见 Schütz et al., 2011。

[4.] 关于方程式，请参见 Adams et al., 2012，第8-9页。注意这些页面上的方程式首先指定了*产生*智能体(agent)接收的感官输入的过程（关于目标位置的外感受视网膜输入，在内在（视网膜）参考框架中，以及报告眼部角位移的本体感受输入），然后指定了智能体(agent)用来处理这些输入的生成模型(generative model)。本文只（非正式地）描述了后者。

[5.] 这里有一种双重否定，可能会有些令人困惑。图景是这样的：标准（神经典型）状态涉及对自我产生行为的感官后果的衰减（即减少）。当这种衰减本身被减少（消除或减轻）时，这些感官后果——至少在某种意义上——被更真实地体验。它们被体验为与外部产生的相同事件（比如手上的某种压力）具有相同的强度。但正如我们将看到的，这并不总是好事，也可能导致各种关于主体性(agency)和控制的错觉的出现。

[6.] 所有受试者都低估了自己施加的力量，但精神分裂症受试者的这种低估要少得多（参见 Shergill et al., 2005）。

[7.] 参见 Adams et al., 2012，第13-14页；Feldman & Friston, 2010；Seamans & Yang, 2004；另见 Braver et al., 1999。

[8.] Heath Robinson 是一位英国漫画家，他的作品与美国的 Rube Goldberg 类似，经常描绘为实现简单目标而设计的奇异复杂机械——例如，用一长串链条、滑轮、杠杆和布谷鸟钟来给面包片涂黄油或在预定时间送上一杯茶。

[9.] 值得重复的是，在所有这些情况下，真正重要的只是在各个层次计算的预测误差精度之间的相对平衡。从功能角度来说，高层精度是降低还是低层精度增加并不重要。这并不是说这样的差异不重要，因为它们涉及可能对病因学和治疗具有临床意义的因果模式。

[10.] “功能性运动和感觉症状”这个标签，与许多这些旧术语不同，被患者认为是可接受的（参见 Stone et al., 2002）。

[11.] 当被催眠的患者被告知他们的手瘫痪时，也注意到了民间生理学期望的影响。在这里，就像在所谓的”癔症性瘫痪”案例中一样，瘫痪的边界反映了我们关于手的边界的常识观念，而不是手作为可移动单元的真正生理学（参见 Boden, 1970）。

[12.] 例如，岛叶皮质(insular cortex)（在功能性疼痛症状的情况下）或前运动皮质(premotor cortex)或辅助运动区(supplementary motor area)（在功能性运动症状的情况下）。

[13.] 关于这种框架在幻肢痛特殊情况下的有趣应用，请参见 De Ridder, Vanneste, & Freeman, 2012。

[14.] 关于回顾，请参见 Benedetti, 2013；Enck et al., 2013；Price et al., 2008。Buchel 等人还推测（2014年，第1227-1229页）施用的阿片类药物(opioids)可能在调节自上而下预测的精度方面发挥类似的（某种程度上更直接和”机械性”的）作用。关于与预测处理(PP)模型广泛兼容的有趣进化论述，请参见 Humphrey, 2011。

[15.] 访问在线讨论论坛”wrongplanet.com“并搜索”hollow mask illusion”是很有启发性的。在一个帖子中，各种自闭症受试者报告最初未能看到空心面具错觉（而是看到它的真实样子，即作为一个倒置的凹面具而不是凸面形状）。但许多受试者能够学会以神经典型的方式看到它，而这些受试者随后发现自己无法将倒置面具体验为凹形。这与神经典型的正弦波语音体验非常相似（参见2.2）。

[16.] 这个故事有新兴的实证支持（参见 Skewes et al., 2014）。

[17.] 这可能不是这里最好的术语，因为它会邀请与超先验(hyper-prior)技术概念的错误类比（参见 Friston, Lawson, & Frith, 2013）。

[18.] 在探索所谓的”存在感”时，并不清楚体验目标是某种模糊但积极的存在感，还是简单地指正常缺乏非存在感、不真实感或断连感。Seth等人提出的模型在这个微妙但可能很有启发性的问题上故意保持不可知论。相反，他们的目标是在所谓的”解离性精神障碍”中受损的任何东西（存在感，或缺乏缺失感的感觉）：这些障碍中，世界或自我的现实感被改变或丧失。

[19.] 有时本体感觉被算作外感受的一部分。Seth等人自己也是这样划分的。本文的处理方式并不依赖于这种术语选择，但在查看图7.2时值得记住这一点。

[20.] 关于批评和复杂性，请参见Marshall and Zimbardo (1979)，Maslach (1979)，以及LeDoux (1995)。

[21.] 这里与Nick Humphrey复杂而精妙的工作产生了共鸣，他论证(Humphrey, 2006)感觉总是涉及某种主动的东西——用主动的（预测处理会说是”预期的”）反应来迎接传入信号的持续尝试。

[22.] 在他的论述接近尾声时，Pezzulo添加了进一步的（我认为重要的）转折，论证(2013, p. 18)“内感受信息是’风’、’小偷’和许多其他实体表征的组成部分”。这里的想法是，在特定外部事态存在时变得活跃的内部状态总是丰富地受到情境影响，而这种影响现在无缝地结合了”客观”和”主观”（例如，情感和身体相关的）元素。

[23.] 一个额外的因素(Pezzulo, 2013, p. 14)可能是单一模型（夜间小偷）相对于涉及两个同时发生但因果无关因素的复杂模型的更大简单性或简约性。我们将在第8章和第9章中对这种因素有更多说明。

[24.] 关于这种图景的精彩描述，请参见Barrett & Bar, 2009。

[25.] 与这一关键作用一致，影响不同神经群体的损伤或干扰，或选择性地影响强直性或阶段性多巴胺反应，通常会产生非常不同的行为和体验效果（见Friston, Shiner, et al. 2012）。

[26.] 当然，只有当你已经对文献中”困难问题”的典型表述有些怀疑时，这才会感觉像是进步。困难问题的真正信仰者会说，我们使用这些新奇资源能够取得进展的只是解释反应和判断模式的熟悉项目，而不是体验本身的存在。那些更乐观的人会认为，解释足够多的那些东西正是解释为什么会有体验的全部。关于这个项目在提供的广义贝叶斯框架内的形成，请参见Dennett, 2013作为开始。

第8章

[1.] 完整标题是”快速、廉价和失控：机器人入侵太阳系”——这暗指论文中提到向太空发送数百个微小、廉价机器人的想法。见Brooks & Flynn, 1989。

[2.] 例如，如果被要求判断一对城镇中哪个人口最多，我们通常会选择名字我们觉得最熟悉的那个。在许多情况下，这种熟悉度竞赛（或”识别启发法”）确实会追踪人口规模。关于平衡的评述，见Gigerenzer and Goldstein (2011)。

[3.] 有关于狗如何接飞盘的相关叙述，由于飞行路径偶尔的剧烈波动，这是一个更具挑战性的任务（见Shaffer et al., 2004）。

[4.] 为了测试这个假设，Ballard等人使用计算机程序在受试者看向别处时改变积木的颜色。对于大多数这些干预，受试者没有注意到变化，即使是对于之前多次访问过的积木和位置，或者是当前行动焦点的积木和位置。

[5.] 感谢Jakob Hohwy（个人交流）提供这个有用的建议。

[6.] 这可能被实现为所谓”演员-评论家”叙述（见Barto, 1995；Barto et al. 1983）的适应版本，但其中无模型的”演员”由基于模型的”评论家”训练（见Daw et al., 2011, p. 1210）。

[7.] 本质上，这涉及创建不同模型所做预测的加权平均值。相比之下，贝叶斯模型选择涉及选择一个模型而非其他模型（例如，见Stephan et al., 2009）。读者可能注意到，在本章的大部分内容中，我用模型选择来框架我自己的讨论。尽管如此，在某些（高敏感性）条件下，模型平均可以提供直接的模型选择。能够改变自己模型比较敏感性的智能体(agents)将受益于增加的灵活性，尽管尚不清楚哪种程序与行为和神经证据提供最佳拟合。进一步讨论见Fitzgerald, Dolan, & Friston, 2014。

[8.] 然而，“双系统”观点的较弱版本，如最近由Evans (2010)和Stanovich (2011)所辩护的那些，可能与更加整合的预测处理(PP)解释是一致的。另见Frankish，即将发表。

[9.] 这种效应已在一些简单的模拟中得到证实，这些模拟研究在多巴胺放电的强直水平变化影响下的提示到达行为——见Friston, Shiner, et al., (2012)。

[10.] 这种方法，即在分层贝叶斯系统的更大背景下对简单、高效的模型进行门控和启用，绝非预测处理独有。例如，它存在于Haruno, Wolpert, & Kawato, 2003提出的MOSAIC框架中。一般来说，这种策略在任何能够获得我们自身不确定性估计来门控和调节在线反应的地方都是可用的。

[11.] 关于正面论证的全面回顾，见Clark, 2008。关于批评，见Adams & Aizawa, 2008；Rupert, 2004，2009。关于正在进行的辩论的丰富样本，见Menary, 2010中的论文。

[12.] 除了睡觉时间！我使用”黑暗房间”一词是指这样的情景：我们退到一个黑暗的房间并待在那里，没有食物、水或娱乐直到死亡——这是批评者心中那种”可预测但致命的”情况。

[13.] 这个例子出现在Campbell, 1974中。

[14.] 这得到了Lisa Meeden（个人交流）的证实。Meeden认为这个故事可能源于一个学习机器人（Lee et al., 2009），该机器人被设计来寻找可学习的（但尚未学会的）感觉状态。其中一个这样的”好奇机器人”（或”内在动机控制器”），没有简单地首先学习最容易的东西，而是似乎通过尝试学习困难的东西和容易的东西来挑战自己。这种行为不是实验者预期的，导致机器人表现出来回摆动的观看例程。对于神话般的旋转行为，Meeden（个人交流）指出”看到一个基于好奇心的学习机器人被放置在一个相当简单的环境中很长一段时间会发生什么将会很有趣。它会开始通过尝试新的运动组合为自己创造新的体验吗？“关于机器人和内在动机的更多信息，见Oudeyer et al., 2007。

[15.] 最初的研究是Jones, Wilkinson, and Braden (1961)的研究。

[16.] 关于这种”探索/利用”权衡的更多信息，见Cohen, McClure, & Yu, 2007。

[17.] 在感知的情况下，这种”自组织不稳定性”的另一个基本收益可能是避免过度自信，从而总是为探索其他可能性留出空间。

第9章

[1.] 如前所述（4.5），我并不意味着暗示所有预测误差都可以通过行动来消除。然而，似乎正确的是，即使在误差本身无法通过行动解决的情况下，误差最小化例程仍然试图提供一种对世界的把握，这种把握适合于行动的选择。显著的未解释误差，即使在”纯感知”情况下，因此也是为了选择根本上设计来帮助我们选择行动的感知。

[2.] 这个例程（Mozer & Smolensky, 1990）从一个训练好的网络中移除最不必要的隐藏单元，从而提高了效率和泛化能力。类似地，在计算机图形学中，骨架化例程从图像中移除最不必要的像素。关于一些讨论，见Clark (1993)。

[3.] 感谢Jakob Hohwy让我注意到这项工作。

[4.] 在Namikawa等人的工作中，因此涉及的神经结构是前额皮质、辅助运动区和初级运动皮质（见Namikawa et al., 2011，第1-3页）。

[5.] 这种效应的力量和复杂性在Dehaene’s (2004)的”神经回收”解释中得到了很好的说明，该解释描述了神经前体、文化发展以及这些文化发展的神经效应之间的复杂相互作用，如在阅读和写作的关键认知领域中所表现的（见下文9.7）。

[6.] 关于在预测性、概率性背景下所有这些离散符号的最初出现的一些有趣推测，见König & Krüger, 2006。

[7.] 见，例如，Anderson, 2010；Griffiths & Gray, 2001；Dehaene et al., 2010；Hutchins, 2014；Oyama, 1999；Oyama et al., 2001；Sterelny, 2003；Stotz, 2010；Wheeler & Clark, 2008。关于有用的回顾，见Ansari, 2011。

[8.] 关于一些讨论，见Tsuchiya & Koch, 2005。

[9.] 然而，暴露于词汇在双眼竞争的情况下似乎不起作用，尽管用其他（图像一致的）声音进行启动确实有效，见Chen, Yeh, & Spence, 2012。

[10.] 这在某种程度上解释了为什么基于简单语言的测量指标（如词汇量）能够很好地预测非语言智力测试的表现。参见 Cunningham & Stanovich, 1997。另见 Baldo et al., 2010。

[11.] 现在有大量但并不完全统一的关于具身认知(enaction)的文献。然而，就我们的目的而言，仅考虑 Varela et al. (1991) 的经典论述就足够了。对具身认知和受具身认知启发的理论化更大领域的重要贡献包括 Froese and Di Paolo (2011)、Noë (2004、2010)，以及 Thompson (2010)。由 Stewart et al. (2010) 编辑的合集为理解这个更大领域提供了一个极好的窗口。

[12.] 这个图像的一部分是误导性的，因为它暗示外部世界仅仅是未分化扰动的来源（单调锤子的重复敲击）。看起来正确的是，主体通过让自己暴露于生成模型预测的各种刺激中，积极地对被采样的世界做出贡献。由于模型只需要适应和解释被采样的世界，这提供了一种非常真实的意义，即（在结构自我维持的总体约束下，即持续和生存）我们确实构建或”具身化”了我们个体和物种特有的世界。

[13.] 这种过程的变体可能存在于多个组织层面，也许可以深入到单细胞的树突层面。正是在这种情况下，Kiebel and Friston (2011) 建议人们可以将细胞内过程理解为最小化”自由能”（见[附录2]）。

[14.] 我在其他地方已经详细讨论了这些论证（见 Clark, 1989、1997、2008、2012）。关于*反对*内部表征的解释性诉求的持续论证，见 Chemero, 2009；Hutto & Myin, 2013；Ramsey, 2007。有用的讨论见 Gallagher, Hutto, Slaby, & Cole, 2013；Sprevak, 2010、2013。

[15.] 这也是 Engel et al. (2013) 所描述的”动态指令”(dynamic directives)更广泛发挥的作用——基于可能包括多个神经和身体结构的涌现集合的行动倾向。

[16.] 贝叶斯知觉和感觉运动心理学已经对这些可能是什么样的世界和身体状态有很多见解。例如，见 Körding & Wolpert, 2006；Rescorla, 2013，印刷中。

[17.] 关于一个精彩的尝试，见 Orlandi (2013)。Orlandi 具有挑衅性但论证严密的观点是，视觉不是一种认知活动，也不涉及内部表征的交易（尽管它有时可能产生心理表征作为一种最终产品）。然而，这个论证在范围上是受限的，因为它只针对在线视觉知觉所涉及的过程。

第10章

[1.] 这样的理解也可能具有社会和政治后果。因为在人类体验的核心，PP 表明，存在着我们自己（大多是无意识的）期望的大量层次。这意味着我们必须仔细考虑我们（和我们的孩子）所暴露的世界的形状。如果 PP 是正确的，我们的知觉可能被通过我们自己过去经验的统计透镜获得的无意识期望深深影响。因此，如果（仅举显而易见的例子）调节这些期望的世界是彻底性别歧视或种族主义的，那将构建积极构造我们未来知觉的地下预测机制——这是被污染的”证据”、不公正反应和自我实现负面预言的有力配方。

[2.] 其中一个子集当然是分层动态模型集合（见 Friston, 2008）。

[3.] 意识体验在这个集合中有些特殊，因为它很可能是动物认知的一个相当基本的特征。实际上，最近的工作（在第7章中回顾）将意识体验与内感受预测联系起来，这表明了这一点：对我们自己生理状态的预测。

附录1

[1.] 给定一些新数据（或感觉证据）的假设的后验概率与给定假设的数据概率乘以假设的先验概率成正比。

参考文献

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, *14*(1), 43 – 64.

Adams, R. A., Perrinet, L. U., & Friston, K. (2012). Smooth pursuit and visual occlusion: Active inference and oculomotor control in schizophrenia. *PLoS One*, *7*(10), e47502. doi:10.1371/journal.pone. 0047502.

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Struct. Funct.*, *218*(3), 611 – 643.

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013). The computational anatomy of psychosis. *Front. Psychiatry*, *30*(4), 47. 1 – 26.

Addis, D. R., Wong, A., & Schacter, D. L. (2007). 记忆过去与想象未来：事件构建和阐述过程中的共同和独特神经基质。*神经心理学*，*45*，1363 – 1377。

Addis, D. R., Wong, A., & Schacter, D. L. (2008). 未来事件情景模拟的年龄相关变化。*心理科学*，*19*，33 – 41。

Aertsen, A., Bonhöffer, T., & Krüger, J. (1987). 神经元群体中的协调活动：分析与解释。见 E. R. Caianiello（编），*认知过程物理学*（第1-3400 – 00页）。新加坡：世界科学出版社。

Aertsen, A., and Preißl, H. (1991). 生理神经网络中活动和连接的动力学。见 H. G. Schuster（编），*非线性动力学和神经网络*（第281 – 302页）。纽约：VCH出版社。

Alais, D., & Blake, R.（编）。(2005). *双眼竞争*。马萨诸塞州剑桥：MIT出版社。

Alais, D., & Burr, D. (2004). 腹语术效应源于近乎最优的双模态整合。*当代生物学*，*14*，257 – 262。

Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). 刺激可预测性降低初级视觉皮层的反应。*神经科学杂志*，*30*，2960 – 2966。

Allard, F., & Starkes, J. (1991). 体育、舞蹈和其他领域的运动技能专家。见 K. Ericsson & J. Smith（编），*迈向专业技能的一般理论：前景与局限*（第126 – 152页）。纽约：剑桥大学出版社。

Anchisi, D., & Zanon, M. (2015). 疼痛感知和安慰剂镇痛中感觉与认知整合的贝叶斯视角。*公共科学图书馆·综合*，*10*，e0117270。doi:10.1371/journal.pone.0117270。

Andersen, R. A., & Buneo, C. A. (2003). 后顶叶皮层的感觉运动整合。*神经病学进展*，*93*，159 – 177。

Anderson, M. L. (2010). 神经重用(Neural reuse)：大脑的基本组织原则。*行为与脑科学*，*33*，245 – 313。

Anderson, M. L. (2014). *颅相学之后：神经重用与交互式大脑*。马萨诸塞州剑桥：MIT出版社。

Anderson, M. L., & Chemero, A. (2013). 大脑GUTs的问题：对”预测”不同含义的混淆威胁形而上学灾难。*行为与脑科学*，*36*(3)，204 – 205。

Anderson, M. L., Richardson, M., & Chemero, A. (2012). 侵蚀认知边界：具身性的影响。*认知科学主题*，*4*(4)，717 – 730。

Ansari, D. (2011). 文化与教育：大脑可塑性的新前沿。*认知科学趋势*，*16*(2)，93 – 95。

Anscombe, G. E. M. (1957). *意图*。牛津：巴兹尔·布莱克威尔出版社。

Arbib, M., Metta, G., & Van der Smagt, P. (2008). 神经机器人学：从视觉到行动。见 B. Siciliano & K. Oussama（编），*机器人学手册*。纽约：施普林格出版社。

Arthur, B. (1994). *经济学中的递增收益与路径依赖*。安阿伯：密歇根大学出版社。

Asada, M., MacDorman, K., Ishiguro, H., & Kuniyoshi, Y. (2001). 认知发展机器人学作为人形机器人设计的新范式。*机器人与自主系统*，37，185 – 193。

Asada, M., MacDorman, K., Ishiguro, H., & Kuniyoshi, Y. (2009). 认知发展机器人学：综述。*IEEE自主心理发展汇刊*，*1*(1)，12 – 34。

Atlas, L. Y., & Wager, T. D. (2012). 期望如何塑造疼痛。*神经科学通讯*，*520*，140 – 148。doi:10.1016/j.neulet.2012.03.039。

Avila, M. T., Hong, L. E., Moates, A., Turano, K. A., & Thaker, G. K. (2006). 预期在精神分裂症相关追踪启动缺陷中的作用。*神经生理学杂志*，*95*(2)，593 – 601。

Bach, D. R., & Dolan, R. J. (2012). 知道自己不知道多少：不确定性估计的神经组织。*自然神经科学评论*，*13*(8)，572 – 586。doi: 10.1038/nrn3289。

Baess, P., Jacobsen, T., & Schroger, E. (2008). 用不可预测的自我启动音调抑制听觉N1事件相关电位成分：动态刺激内部前向模型的证据。*国际心理生理学杂志*，*70*，137 – 143。

Baldo, J. V., Bunge, S. A., Wilson, S. M., & Dronkers, N. F. (2010). 关系推理是否依赖于语言？基于体素的病灶症状映射研究。*大脑与语言*，*113*(2)，59 – 64。doi:10.1016/j.bandl.2010.01.004。

Ballard, D. (1991). 生动视觉。*人工智能*，*48*，57 – 86。

Ballard, D. H., & Hayhoe, M. M. (2009). 建模任务在注视控制中的作用。*视觉认知*，*17*，1185 – 1204。

Ballard, D., Hayhoe, M., Pook, P., & Rao, R. (1997). 认知具身的指示性编码。*行为与脑科学*，*20*，4。723 – 767。

Bar, M. (2004). 情境中的视觉对象。*自然神经科学评论*，5，617 – 629。

Bar, M. (2007). 主动大脑：使用类比和联想生成预测。*认知科学趋势*，*11*(7)，280 – 289。

Bar, M. (2009). 主动大脑：预测的记忆。主题专刊：大脑中的预测：使用我们的过去生成未来（M. Bar编）。*英国皇家学会哲学汇刊B*，*364*，1235 – 1243。

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmidt, A. M., Dale, A. M., Hamalainen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., & Halgren, E. (2006). 视觉识别的自上而下促进。*美国国家科学院院刊*，*103*(2)，449 – 454。

Bargh, J. A. (2006). 这些年我们一直在启动什么？关于无意识社会行为的发展、机制和生态学。*欧洲社会心理学杂志*，*36*，147 – 168。

Bargh, J. A., Chen, M., & Burrows, L. (1996). 社会行为的自动性：特质构念和刻板印象激活对行动的直接影响。*人格与社会心理学杂志*，*71*，230 – 244。

Barlow, H. B. (1961). 感觉信息的编码。在 W. H. Thorpe & O. L. Zangwill (编辑), *动物行为的当前问题* (pp. 330 – 360). 剑桥：剑桥大学出版社。

Barnes, G. R., & Bennett, S. J. (2003). 视觉目标瞬间消失期间的人类眼球追踪。*神经生理学杂志*, *90*(4), 2504 – 2520。

Barnes, G. R., & Bennett, S. J. (2004). 视觉目标瞬间消失期间的预测性平滑眼球追踪。*神经生理学杂志*, *92*(1), 578 – 590。

Barone, P., Batardiere, A., Knoblauch, K., & Kennedy, H. (2000). 投射到视觉区域V1和V4的纹外区域神经元的层状分布与层级等级相关，并表明距离规则的操作。*J. Neurosci.*, *20*, 3263 – 3281. pmid: 10777791。

Barrett, H. C., & Kurzban, R. (2006). 认知中的模块性：构建辩论框架。*心理学评论*, *113*(3), 628 – 647。

Barrett, L. F., & Bar, M. (2009). 用感觉看：人脑中的情感预测。*皇家学会哲学汇刊B*, *364*, 1325 – 1334。

Barsalou, L. (1999). 感知符号系统。*行为与脑科学*, *22*, 577 – 660。

Barsalou, L. (2003). 感知符号系统中的抽象。*伦敦皇家学会哲学汇刊：生物科学*, *358*, 1177 – 1187。

Barsalou, L. (2009). 仿真、情境概念化和预测。*伦敦皇家学会哲学汇刊：生物科学*, *364*, 1281 – 1289。

Barto, A., Sutton, R., & Anderson, C. (1983). 能够解决困难学习控制问题的神经元样自适应元素。*IEEE系统、人与控制论汇刊*, *13*, 834 – 846。

Barto, A. G. (1995). 自适应批评家和基底神经节。在 J. L. Davis, J. C. Houk, & D. G. Beiser (编辑), *基底神经节中的信息处理模型* (pp. 215 – 232). 剑桥，马萨诸塞州：MIT出版社。

Bastian, A. (2006). 学会预测未来：小脑适应前馈运动控制。*神经生物学当前观点*, *16*(6), 645 – 649。

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). 预测编码的典型微回路。*神经元*, *76*, 695 – 711。

Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J. R., De Weerd, P., Kennedy, H., and Fries, P. (2015). 视觉区域通过不同频率通道施加前馈和反馈影响。*神经元*, 1 – 12. doi:10.1016/j.neuron.2014.12.018。

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). 仿真作为物理场景理解的引擎。*美国国家科学院院刊*, *110*(45), 18327 – 18332. doi:10.1073/pnas.1306572110。

Becchio, C., et al. (2010). 自闭症谱系障碍儿童对阴影的感知。*PLoS One*, *5*, e10582. doi:10.1371/journal.pone.0010582。

Beck, D. M., & Kastner, S. (2005). 刺激情境调节人类纹外皮层的竞争。*自然神经科学*, *8*, 1110 – 1116。

Beck, D. M., & Kastner, S. (2008). 人脑中影响竞争的自上而下和自下而上机制。*视觉研究*, *49*, 1154 – 1165。

Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., & Pouget, A. (2012, April 12). 不是噪声，只是错误：次优推理在行为变异性中的作用。*神经元*, *74*(1), 30 – 39. doi:10.1016/j.neuron.2012.03.016。

Beer, R. (2000). 认知科学的动力学方法。*认知科学趋势*, *4*(3), 91 – 99。

Beer, R. (2003). 进化模型代理中主动分类感知的动力学。*适应行为*, *11*, 209 – 243。

Bell, C. C., Han, V., & Sawtell, N. B. (2008). 小脑样结构及其对小脑功能的意义。*神经科学年度评论*, *31*, 1 – 24。

Benedetti, F. (2013). 安慰剂(Placebo)和医患关系的新生理学。*生理学评论*, *93*, 1207 – 1246。

Bengio, Y. (2009). 为人工智能学习深度架构。*机器学习基础与趋势*, *2*(1):, 1 – 127。

Bengio, Y., & Le Cun, Y. (2007). 将学习算法扩展到人工智能。在 Bottou L., Chapelle O., DeCoste D. and Weston J. (编辑), *大规模核机器* (pp. 321 – 360). 剑桥，马萨诸塞州：MIT出版社。

Berkes, P., Orban, G., Lengyel, M., & Fiser, J. (2011). 自发的皮层活动揭示了环境最优内部模型的特征。*科学*, *331*, 83 – 87。

Berlyne, D. (1966). 好奇心和探索。*科学*, *153*(3731), 25 – 33。

Bermúdez, J. (2005). *心理学哲学：当代导论*. 纽约：Routledge出版社。

Berniker, M., & Körding, K. P. (2008). 估计运动错误的来源以进行适应和泛化。*自然神经科学*, *11*, 1454 – 1461。

Betsch, B. Y., Einhäuser, W., Körding, K. P., & König, P. (2004). 从猫的视角看世界：自然视频的统计特性。*生物控制论*, *90*, 41 – 50。

Biederman, I. (1987). 组件识别：人类图像理解理论。*心理学评论*, *94*, 115 – 147。

Bindra, D. (1959). 刺激变化、对新奇事物的反应和反应递减。*心理学评论*, *66*, 96 – 103。

Bingel, U., Wanigasekera, V., Wiech, K., Ni Mhuircheartaigh, R., Lee, M. C., Ploner, M., & Tracey, I. (2011). 治疗期望对药物疗效的影响：阿片类药物瑞芬太尼镇痛益处的成像研究。*科学转化医学*, *3*, 70ra14. 1 – 9 doi:10.1126/scitranslmed.3001244。

Bissom, T. (1991). 外星人/国度。*Omni. (科幻杂志)*

Blackmore, S. (2004). *意识：导论*. 纽约：牛津大学出版社。

Blakemore, S.-J., Frith, C. D., & Wolpert, D. W. (1999). 时空预测调节对自我产生刺激的感知。*认知神经科学杂志*, *11*(5), 551 – 559。

Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (1998). 自产挠痒感觉的中枢消除。*自然神经科学*, *1*(7), 635 – 640。

Blakemore S. J., Wolpert D., Frith C. (2000). 为什么你不能挠自己痒？*神经报告* 11, R11 – 16。

Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (2002). 动作意识的异常。*认知科学趋势*, *6*, 237 – 242。

Block, N., & Siegel, S. (2013). 注意和知觉适应。*行为与脑科学*, *36*(4), 205 – 206。

Boden, M. (1970). 意向性(intentionality)和物理系统。*科学哲学*, *37*, 200 – 214。

Botvinick, M. (2004). 探索身体所有权的神经基础。*科学*, *305*, 782 – 783。doi:10.1126/science. 1101836。

Botvinick, M., & Cohen, J. (1998). 橡胶手”感受”眼睛看到的触觉。*自然*, *391*, 756。doi:10. 1038/35784。

Bowman, H., Filetti, M., Wyble, B., & Olivers, C. (2013). 注意不仅仅是预测精度。*行为脑科学*., *36*(3), 233 – 253。

Box, G. P., & Draper, N. R. (1987). *经验模型构建与响应面*。新泽西州威利出版社。

Boynton, G. M. (2005). 注意和视觉知觉。*当前神经生物学观点*., *15*, 465 – 469。doi:10.1016/j.conb.2005. 06.009。

Brainard, D. (2009). 色觉的贝叶斯方法。在M. Gazzaniga（编），*视觉神经科学*（第4版）。马萨诸塞州剑桥：MIT出版社。

Bram, U. (2013). *统计思维*。旧金山：Capara图书。

Braver, T. S., Barch, D. M., & Cohen, J. D. (1999). 精神分裂症的认知和控制：多巴胺和前额叶功能的计算模型。*生物精神病学*, *46*(3), 312 – 328。

Brayanov, J. B., & Smith, M. A. (2010). 尺寸-重量错觉中动作和知觉感觉整合的贝叶斯和”反贝叶斯”偏差。*神经生理学杂志*, *103*(3), 1518 – 1531。

Brock, J. (2012, 11月2日). 自闭症知觉的替代贝叶斯解释：对Pellicano和Burr的评论。*认知科学趋势*, *16*(12), 573 – 574。doi:10.1016/j.tics.2012.10.005。

Brooks, R., & Flynn, A(1989). 快速、廉价且失控：太阳系的机器人入侵。*英国行星际学会杂志*., *42*(10), 478 – 485。

Brown, E. C., & Brüne, M. (2012). 预测在社会神经科学中的作用。*人类神经科学前沿*., *6*, 147。doi:10.3389/fnhum.2012.00147。

Brown, H., Adams, R. A., Parees, I., Edwards, M., & Friston, K. (2013). 主动推理(active inference)、感觉衰减和错觉。*认知过程*. 14 (4), 411 – 427。

Brown, H., & Friston, K. (2012). 自由能和错觉：Cornsweet效应。*心理学前沿*, *3*, 43。doi:10.3389/fpsyg.2012.00043。

Brown, H., Friston, K., & Bestmann, S. (2011). 主动推理(active inference)、注意和运动准备。*心理学前沿*, *2*, 218。doi:10.3389/fpsyg.2011.00218。

Brown, R. J. (2004). 医学无法解释症状的心理机制：整合概念模型。*心理学通报*., *130*, 793 – 812。

Bubic, A., von Cramon, D. Y., & Schubotz, R. I. (2010). 预测、认知和大脑。*人类神经科学前沿*., *4*, 25。doi:10.3389/fnhum.2010.00025。

Buchbinder, R., & Jolley, D. (2005). 媒体宣传对背部疼痛观念的影响在停止3年后仍然持续。*脊柱*, *30*, 1323 – 1330。

Büchel, C., Geuter, S., Sprenger, C., & Eippert, F., 等 (2014). 安慰剂镇痛：预测编码视角。*神经元*, 81(6), 1223 – 1239。

Buckingham, G., & Goodale, M. (2013). 当预测大脑严重出错时。*行为与脑科学*, 36(3), 208 – 209。

Buffalo, E. A., Fries, P., Landman, R., Buschman, T. J., & Desimone, R. (2011). 腹侧流中伽马和阿尔法相干性的层间差异。*美国国家科学院院刊*, *108*(27), 11262 – 11267。doi.org/10.1073/pnas.1011284108。

Burge, J., Fowlkes, C., & Banks, M. (2010). 自然场景统计预测凸性的图形-背景线索如何影响人类深度知觉。*神经科学杂志*, *30*(21), 7269 – 7280。doi:10.1523/JNEUROSCI.5551-09.2010。

Burge, T. (2010). *客观性的起源*。纽约：牛津大学出版社。

Byrne, A., & Logue, H.（编）。(2009). *析取主义(disjunctivism)：当代读本*。马萨诸塞州剑桥：MIT出版社。

Caligiore, D., Ferrauto, T., Parisi, D., Accornero, N., Capozza, M., & Baldassarre, G. (2008). 使用运动咿呀学语(motor babbling)和赫布规则建模带障碍物的达到和抓取发展。在*CogSys2008—认知系统国际会议*（E 1 – 8。马萨诸塞州剑桥：MIT出版社。

Campbell, D. T. (1974). 进化认识论。在P. A. Schlipp（编），*卡尔·波普尔的哲学*（pp. 413 – 463）。伊利诺伊州拉萨尔：开放法院。

Cardoso-Leite, P., Mamassian, P., Schütz-Bosbach, S., & Waszak, F. (2010). 感觉衰减的新视角：动作-效果预期影响敏感性，而非反应偏差。*心理科学*, *21*, 1740 – 1745。doi:10.1177/0956797610389187。

Carey, S. (2009). *概念的起源*。纽约：牛津大学出版社。

Carrasco, M. (2011). 视觉注意：过去25年。*视觉研究*., *51*(13), 1484 – 1525。doi:10.1016/j.visres.2011.04.012。

Carrasco, M., Ling, S., & Read, S. (2004). 注意改变外观。*自然神经科学*, *7*, 308 – 313。

Carruthers, P. (2009). 无脊椎动物概念面对一般性约束(generality constraint)（并获胜）。在Robert W. Lurz（编），*动物心智哲学*（pp. 89 – 107）。剑桥：剑桥大学出版社。

Castiello, U. (1999). 手部动作控制的选择机制。*认知科学趋势*, *3*(7), 264 – 271。

Chadwick, P. K. (1993). 通向不可能的阶梯：精神分裂情感性精神危机的第一手现象学记述。*心理健康杂志*, *2*, 239 – 250。

Chalmers, D. (1996). *意识心理*. 牛津：牛津大学出版社。

Chalmers, D. (2005). 黑客帝国作为形而上学。在 R. Grau (编), *哲学家探索黑客帝国*. 纽约：牛津大学出版社。

Chapman, S. (1968). 接住棒球。*美国物理学杂志*, *36*, 868 – 870。

Chater, N., & Manning, C. (2006). 语言处理和习得的概率模型。*认知科学趋势*, *10*(7), 335 – 344。

Chawla, D., Friston, K., & Lumer, E. (1999). 三个相互连接的皮层区域中的零滞后同步动力学。*神经网络*, *14*(6 – 7) 727 – 735。

Chemero, A. (2009). *激进具身认知科学*. 剑桥，马萨诸塞州：MIT出版社。

Chen, C. C., Henson, R. N., Stephan, K. E., Kilner, J. M., & Friston, K. J. (2009). 大脑中的前向和后向连接：功能不对称性的DCM研究。*神经影像*, *45*(2), 453 – 462。

Chen, Yi-Chuan, Su-Ling Yeh, & Spence, C. (2011). 人类感知意识的跨模态约束：双眼竞争的听觉语义调节。*心理学前沿*, *2*, 212. doi:10.3389/fpsyg.2011.00212。

Christiansen, M., & Chater, N. (2003). 语言中的成分性和递归性。在 M. A. Arbib (编), *大脑理论和神经网络手册* (pp. 267 – 271). 剑桥，马萨诸塞州：MIT出版社。

Churchland, P. M. (1989). *神经计算视角*. 剑桥，马萨诸塞州：MIT/Bradford图书。

Churchland, P. M. (2012). *柏拉图的相机：物理大脑如何捕捉抽象普遍性的景观*. 剑桥，马萨诸塞州：MIT出版社。

Churchland, P. S. (2013). *触碰神经：作为大脑的自我*. W. W. Norton。

Churchland, P. S., Ramachandran, V., & Sejnowski, T. (1994). 对纯视觉的批判。在 C. Koch & J. Davis (编), *大脑的大规模神经理论* (pp. 23 – 61). 剑桥，马萨诸塞州：MIT出版社。

Cisek, P. (2007). 动作选择的皮层机制：可供性竞争假说(affordance competition hypothesis)。*英国皇家学会哲学汇刊B*, *362*, 1585 – 1599。

Cisek, P., & Kalaska, J. F. (2005). 背侧前运动皮层中达成决策的神经相关性：多个方向选择的规格化和动作的最终选择。*神经元*, *45*(5), 801 – 814。

Cisek, P., & Kalaska, J. F. (2011). 与充满动作选择的世界互动的神经机制。*神经科学年度综述*, *33*, 269 – 298。

Clark, A. (1989). *微认知：哲学、认知科学和并行分布式处理*. 剑桥，马萨诸塞州：MIT出版社/Bradford图书。

Clark, A. (1993). *联想引擎：连接主义、概念和表征变化*. 剑桥，马萨诸塞州：MIT出版社，Bradford图书。

Clark, A. (1997). *在那里：重新将大脑、身体和世界结合在一起*. 剑桥，马萨诸塞州：MIT出版社。

Clark, A. (1998). 魔法词汇：语言如何增强人类计算。在 P. Carruthers & J. Boucher (编), *语言和思维：跨学科主题* (pp. 162 – 183). 剑桥：剑桥大学出版社。

Clark, A. (2003). *天生的赛博格：心智、技术和人类智能的未来*. 纽约：牛津大学出版社。

Clark, A. (2006). 语言、具身性和认知生态位(cognitive niche)。*认知科学趋势*, *10*(8), 370 – 374。

Clark, A. (2008). *超级大脑：动作、具身性和认知扩展* 纽约：牛津大学出版社。

Clark, A. (2012). 梦见整只猫：生成模型、预测处理和感知体验的能动概念。*心智*, *121*(483), 753 – 771. doi:10.1093/mind/fzs106。

Clark, A. (2013). 接下来是什么？预测大脑、情境代理和认知科学的未来。*行为与脑科学*, *36*(3), 181 – 204。

Clark, A. (2014). 作为预测的感知。在 M. Mohan, S. Biggs, & D. Stokes (编), *感知及其模态*. 纽约：牛津大学出版社，23 – 43。

Clark, A., & Chalmers, D. (1998). 扩展心智(extended mind)。*分析*, *58*(1), 7 – 19。

Cocchi, L., Zalesky, A., Fornito, A., & Mattingley, J. (2013). 认知控制过程中大脑系统的动态合作与竞争。*认知科学趋势*, *17*, 493 – 501。

Coe, B., Tomihara, K., Matsuzawa, M., & Hikosaka, O. (2002). 猴子在适应性决策任务中三个皮层眼区的视觉和预期偏见。*神经科学杂志.*, *22*(12), 5081 – 5090。

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). 我应该留下还是应该离开？人类大脑如何管理开发与探索之间的权衡。*伦敦皇家学会哲学汇刊B：生物科学*, *29*(362), 933 – 942. doi:10.1098/rstb.2007.2098。

Colder, B. (2011). 作为认知整合原则的仿真(emulation)。*人类神经科学前沿.*, *5*, 54. doi:10.3389/fnhum.2011.00054。

Cole, M. W., Anticevic, A., Repovs, G., and Barch, D. (2011). 精神分裂症中的可变全局断连和个体差异。*生物精神病学*, *70*, 43 – 50。

Collins, S. H., Wisse, M., & Ruina, A. (2001). 具有两条腿和膝盖的3-D被动动态行走机器人。*国际机器人研究杂志*, *20*(7), 607 – 615。

Colombetti, G. (2014). *感受身体*. 剑桥，马萨诸塞州：MIT出版社。

Colombetti, G., & Thompson, E. (2008). 感受身体：情感的能动方法。在 W. F. Overton, U. Muller, & J. L. Newman (编), *具身性和意识的发展视角* (pp. 45 – 68). 纽约：Lawrence Erlbaum Assoc。

Colombo, M. (2013). 向前推进（和超越）模块化辩论：网络视角。*科学哲学*, *80*, 356 – 377。

Colombo, M. (出版中). 解释社会规范遵从：对神经表征的呼吁。*现象学与认知科学*.

Coltheart, M. (2007). 认知神经心理学和妄想信念（第33届弗雷德里克·巴特利特爵士讲座）。*实验心理学季刊*, *60*(8), 1041 – 1062。

Conway, C., & Christiansen, M. (2001). 非人类灵长类动物的序列学习. *认知科学趋势*, *5*(12), 539 – 546.

Corlett, P. R., Frith, C. D., & Fletcher, P. C. (2009). 从药物到剥夺：理解精神病模型的贝叶斯框架. *精神药理学(柏林)*, *206*(4), 515 – 530.

Corlett, P. R., Krystal, J. K., Taylor, J. R., & Fletcher, P. C. (2009). 为什么妄想会持续存在？ *人类神经科学前沿*, *3*, 12. doi:10.3389/neuro.09.012.2009.

Corlett, P. R., Taylor, J. R., Wang, X. J., Fletcher, P. C., & Krystal, J. H. (2010). 走向妄想的神经生物学. *神经生物学进展*, *92*(3), 345 – 369.

Coste, C. P., Sadaghiani, S., Friston, K. J., & Kleinschmidt, A. (2011). 持续的大脑活动波动直接解释了Stroop任务表现中的试验间差异，间接解释了被试间差异. *大脑皮层*, *21*, 2612 – 2619.

Craig, A. D. (2002). 你感觉如何？内感受：身体生理状态的感觉. *自然神经科学评论*, *3*, 655 – 666.

Craig, A. D. (2003). 内感受：身体生理状态的感觉. *神经生物学当前观点*, *13*, 500 – 505.

Craig, A. D. (2009). 你现在感觉如何？前脑岛与人类意识. *自然神经科学评论*, *10*, 59 – 70.

Crane, T. (2005). 知觉问题是什么？ *综合哲学*, 40(2), 237 – 264.

Crick, F. (1984). 丘脑网状复合体的功能：探照灯假说. *美国国家科学院院刊*, *81*, 4586 – 4590.

Critchley, H. D. (2005). 自主神经、情感和认知整合的神经机制. *比较神经学杂志*, *493*(1), 154 – 166. doi:10.1002/cne.20749.

Critchley, H. D., & Harrison, N. A. (2013). 内脏对大脑和行为的影响. *神经元*, 77, 624 – 638.

Critchley, H. D., et al. (2004). 支持内感受意识的神经系统. *自然神经科学*, *7*, 189 – 195.

Crowe, S., Barot, J., Caldow, S., D’ Aspromonte, J., Dell’ Orso, J., Di Clemente, A., Hanson, K., Kellett, M., Makhlota, S., McIvor, B., McKenzie, L., Norman, R., Thiru, A., Twyerould, M., & Sapega, S. (2011). 咖啡因和压力对非临床样本中听觉幻觉的影响. *个性与个体差异*, *50*(5), 626 – 630.

Cui, X., Jeter, C. B., Yang, D., Montague, P. R., and Eagleman, D. M. (2007). 心理意象的生动性：个体差异可以客观测量. *视觉研究*, *47*, 474 – 478.

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). 自然视觉过程中的注意力扭曲了人脑中的语义表征. *自然神经科学*, *16*(6), 763 – 770. doi:10.1038/nn.3381.

Cunningham, A. E., & Stanovich, K. E. (1997). 早期阅读习得及其与10年后阅读经验和能力的关系. *发展心理学*, *33*(6), 934 – 945.

Damasio, A. (1999). *事件感受*. 纽约：哈考特，布雷斯出版社.

Damasio, A. (2010). *自我意识的产生：构建有意识的大脑*. 多伦多万神殿出版社.

Damasio, A., & Damasio, H. (1994). 检索具体知识的皮层系统：汇聚区框架. 收录于C. Koch (编), *大脑的大规模神经元理论* (第61 – 74页). 剑桥，马萨诸塞州：MIT出版社.

Daunizeau, J., Den Ouden, H., Pessiglione, M., Stephan, K., Kiebel, S., & Friston, K. (2010a). 观察观察者(I)：学习和决策的元贝叶斯模型. *公共科学图书馆 · 综合*, *5*(12), e15554.

Daunizeau, J., Den Ouden, H., Pessiglione, M., Stephan, K., Kiebel, S., & Friston, K. (2010b). 观察观察者(II)：决定何时决策. *公共科学图书馆 · 综合*, *5*(12), e15555.

Davis, M. H., & Johnsrude, I. S. (2007). 听到语音声音：听觉与语音感知界面上的自上而下影响. *听觉研究*, *229*(1 – 2), 132 – 147.

Daw, N., Niv, Y., & Dayan, P. (2005). 前额叶与背外侧纹状体系统在行为控制中基于不确定性的竞争. *自然神经科学*, *8*(12), 1704 – 1711.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). 基于模型的影响对人类选择和纹状体预测误差的作用. *神经元*, *69*, 1204 – 1215.

Dayan, P. (1997). 分层模型中的识别. 收录于F. Cucker & M. Shub (编), *计算数学基础* (第79 – 87页). 德国柏林：施普林格出版社.

Dayan, P. (2012). 如何设置这个东西的开关. *神经生物学当前观点*, *22*, 1068 – 1074.

Dayan, P., & Daw, N. D. (2008). 决策理论、强化学习(reinforcement learning)和大脑. *认知、情感与行为神经科学*, *8*, 429 – 453.

Dayan, P., & Hinton, G. (1996). 亥姆霍兹机器的变体. *神经网络*, *9*, 1385 – 1403.

Dayan, P., Hinton, G. E., & Neal, R. M. (1995). 亥姆霍兹机器. *神经计算*, *7*, 889 – 904.

De Brigard, F. (2012). 预测性记忆和令人惊讶的差距. *心理学前沿*, *3*, 420. doi:10.3389/fpsyg.2012.00420.

De Ridder, D., Vanneste, S., & Freeman, W. (2012). 贝叶斯大脑：幻觉感知解决感觉不确定性. *神经科学与生物行为评论*. http://dx.doi.org/10.1016/j.neubiorev.2012.04.001.

de Vignemont, F., & Fourneret, P. (2004). 能动感：Who系统的哲学和实证回顾. *意识与认知*, *13*, 1 – 19.

de-Wit, L., Machilsen, B., & Putzeys, T. (2010). 预测编码和对可预测刺激的神经反应. *神经科学杂志*, *30*, 8702 – 8703.

den Ouden, H. E. M., Daunizeau, J., Roiser, J., Friston, K., & Stephan, K. (2010). 纹状体预测误差调节皮层耦合. *神经科学杂志*, *30*, 3210 – 3219.

den Ouden, H. E. M., Kok, P. P., & de Lange, F. P. F. (2012). 预测误差如何塑造感知、注意和动机。*心理学前沿*，*3*，548。doi:10.3389/fpsyg.2012.00548。

Decety, J. (1996). 动作的神经表征。*神经科学评论*，*7*，285 – 297。

Dehaene, S. (2004). 人类阅读和算术皮层回路的进化：“神经元回收”假说。在 S. Dehaene, J. Duhamel, M. Hauser, & G. Rizzolatti (编)，*从猴脑到人脑* (第133 – 158页)。剑桥，马萨诸塞州：MIT出版社。

Dehaene, S., Pegado, F., Braga, L., Ventura, P., Nunes G., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2010). 学习阅读如何改变视觉和语言的皮层网络。*科学*，*330*(6009)，1359 – 1364。doi:10.1126/science.1194140。

Demiris, Y., & Meltzoff, A. (2008). 摇篮中的机器人：婴儿和机器人模仿技能的发展分析。*婴幼儿发展*，*17*，43 – 58。

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). 通过EM算法从不完整数据中进行最大似然估计。*皇家统计学会期刊B辑*，*39*，1 – 38。

Deneve, S. (2008). 贝叶斯脉冲神经元I：推理。*神经计算*，*20*，91 – 117。

Dennett, D. (1982). 超越信念。在 A. Woodfield (编)，*思想与对象* (第74 – 98页)。牛津：克拉伦登出版社。

Dennett, D. (1991). *意识的解释*。波士顿：小布朗出版社。

Dennett, D. (2013). 期待我们自己去期待：作为投影仪的贝叶斯大脑。*行为与脑科学*，*36*(3)，209 – 210。

Desantis, A., Hughes, G., & Waszak, F. (2012). 意向绑定是由动作的单纯存在而非运动预测所驱动的。*公共科学图书馆综合*，*7*，e29557。doi:10.1371/journal.pone.0029557。

Desimone, R. (1996). 视觉记忆的神经机制及其在注意中的作用。*美国国家科学院院刊*，*93*(24)，13494 – 13499。

Desimone, R., & Duncan, J. (1995). 选择性视觉注意的神经机制。*神经科学年度评论*，*18*，193 – 222。

Dewey, J. (1896). 心理学中的反射弧概念。*心理学评论*，*3*，357 – 370。

DeWolf, T., & Eliasmith, C. (2011). 运动控制的神经最优控制层次结构。*神经工程学报*，*8*(6)，21。doi:10.1088/1741-2560/8/6/065009。

DeYoe, E. A., & Van Essen, D. C. (1988). 猴子视觉皮层中的并行处理流。*神经科学趋势*，*11*，219 – 226。

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). 内侧颞叶中回忆和熟悉性的成像：三成分模型。*认知科学趋势*，*11*，379 – 386。

Dima, D., Roiser, J., Dietrich, D., Bonnemann, C., Lanfermann, H., Emrich, H., & Dillo, W. (2009). 使用动态因果建模理解精神分裂症患者为什么不能感知凹面面具错觉。*神经影像*，*46*(4)，1180 – 1186。

Dolan, R. J. (2002). 情绪、认知和行为。*科学*，*298*，1191 – 1194。

Dorris, M. C., & Glimcher, P. W. (2004). 后顶叶皮层的活动与动作的相对主观欲望性相关。*神经元*，*44*(2)，365 – 378。

Doya, K., & Ishii, S. (2007). 概率入门。在 K. Doya, S. Ishii, A. Pouget, & R. Rao (编)，*贝叶斯大脑：神经编码的概率方法*(第3 – 15页)。剑桥，马萨诸塞州：MIT出版社。

Doya, K., Ishii, S., Pouget, A., & Rao, R. (编). (2007). *贝叶斯大脑：神经编码的概率方法*。剑桥，马萨诸塞州：MIT出版社。

Dura-Bernal, S., Wennekers, T., & Denham, S. (2011). 反馈在物体感知层次模型中的作用。*实验医学与生物学进展*，*718*，165 – 179。doi:10.1007/978-1-4614-0164-3_14。

Dura-Bernal, S., Wennekers, T., & Denham, S. (2012). 基于层次贝叶斯网络和信念传播的类HMAX物体感知皮层模型中的自上而下反馈。*公共科学图书馆综合*，*7*(11)，e48216。doi:10.1371/journal.pone.0048216。

Edelman, G. (1987). *神经达尔文主义：神经元群选择理论*。纽约：基础图书出版社。

Edelman, G., & Mountcastle, V. (1978). *有意识的大脑：皮层组织和高级脑功能的群选择理论*。剑桥，马萨诸塞州：MIT出版社。

Edwards, M. J., Adams, R. A., Brown, H., Pareés, I., & Friston, K. (2012). “癔症”的贝叶斯解释。*大脑*，*135*(第11部分)，3495 – 3512。

Egner, T., Monti, J. M., & Summerfield, C. (2010). 期望和惊讶决定腹侧视觉流中的神经群体反应。*神经科学杂志*，*30*(49)，16601 – 16608。

Egner, T., & Summerfield, C. (2013). 将预测编码模型建立在实证神经科学研究基础上。*行为与脑科学*，*36*，210 – 211。

Ehrsson, H. H. (2007). 体外体验的实验诱导。*科学*，*317*，1048。

Einhäuser, W., Kruse, W., Hoffmann, K. P., & König, P. (2006). 自然条件下猴子和人类显性注意的差异。*视觉研究*，*46*，1194 – 1209。

Eliasmith, C. (2005). 表征问题的新视角。*认知科学杂志*，*6*，97 – 123。

Eliasmith, C. (2007). 如何构建大脑：从功能到实现。*综合*，*153*(3)，373 – 388。

Elman, J. L. (1993). 神经网络中的学习和发展：从小开始的重要性。*认知*，*48*，71 – 99。

Enck, P., Bingel, U., Schedlowski, M., & Rief, W. (2013). 医学中的安慰剂反应：最小化、最大化还是个性化？*自然药物发现评论*，*12*，191 – 204。

Engel, A., Maye, A., Kurthen, M., & König, P. (2013). 行动在哪里？认知科学中的实用主义转向。*认知科学趋势*，*17*(5)，202 – 209。doi:10.1016/j.tics.2013.03.006。

Engel, A. K., Fries, P., & Singer, W. (2001). 动态预测：自上而下处理中的振荡和同步。*自然评论*，*2*，704 – 716。

Ernst, M. O. (2010). 眼动：慢镜头中的错觉。*Current Biology*, *20*(8), R357 – R359.

Ernst, M. O., & Banks, M. S. (2002). 人类以统计学最优方式整合视觉和触觉信息。*Nature*, *415*, 429 – 433.

Evans, G. (1982). *指称的变种*. 牛津：牛津大学出版社。

Evans, J. St. B. T. (2010). *思考两次：一个大脑中的两种心智*. 牛津：牛津大学出版社。

Fabre-Thorpe, M. (2011). 快速视觉分类的特征和局限。*Frontiers in Psychology*, *2*(243). doi:10.3389/fpsyg.2011.00243.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). 言语听觉特异性地调节舌肌的兴奋性：一项TMS研究。*Eur. J. Neurosci.*, *15*(2), 399 – 402.

Fair, D. (1979). 因果关系和能量流动。*Erkenntnis*, *14*, 219 – 250.

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). 神经系统中的噪声。*Nature Rev. Neurosci.*, *9*, 292 – 303. doi:10.1038/nrn2258.

Fecteau, J. H., & Munoz, D. P. (2006). 显著性、相关性和放电：目标选择的优先级地图。*Trends in Cognitive Sciences*, *10*, 382 – 390.

Feldman, A. G. (2009). 动作-知觉耦合的新见解。*Experimental Brain Research*, *194*(1), 39 – 58.

Feldman, A. G., & Levin, M. F. (2009). 平衡点假说——过去、现在和未来。*Advances in Experimental Medicine and Biology*, *629*, 699 – 726.

Feldman, H., & Friston, K. (2010). 注意、不确定性和自由能。*Frontiers in Human Neuroscience*, *2*(4), 215. doi:10.3389/fnhum.2010.00215.

Feldman, J. (2010). 认知科学应该统一：对Griffiths等人和McClelland等人的评论。*Trends in Cognitive Sciences*, *14*(8), 341. doi:10.1016/j.tics.2010.05.008.

Feldman, J. (2013). 将你的先验调整到世界。*Topics in Cognitive Science*, *5*(1), 13 – 34. doi:10.1111/tops.12003.

Felleman, D. J., & Van Essen, D. C. (1991). 灵长类大脑皮层的分布式层次处理。*Cerebral Cortex*, *1*, 1 – 47.

Fernandes, H. L., Stevenson, I. H., Phillips, A. N., Segraves, M. A., & Kording, K. P. (2014). 自然场景搜索中额叶眼动区的显著性和扫视编码。*Cerebral Cortex*, *24*(12), 3232 – 3245.

Fernyhough, C. (2012). *光的碎片：记忆的新科学*. 伦敦：Profile Books.

Ferrari, PF 等人 (2003). 响应观察摄食和交流性口部动作的镜像神经元在腹侧前运动皮层中的作用。*European Journal of Neuroscience*, *17*(8), 1703 – 1714.

Ferrari, R., Obelieniene, D., Russell, A. S., Darlington, P., Gervais, R., & Green, P. (2001). 轻微头部外伤后的症状预期：加拿大和立陶宛的比较研究。*Clin. Neurol. Neurosurg.*, *103*, 184 – 190.

Fink, P. W., Foo, P. S., & Warren, W. H. (2009). 在虚拟现实中接住飞球：外野手问题的关键测试。*Journal of Vision*, *9*(13):14, 1 – 8.

FitzGerald, T., Dolan, R., & Friston, K. (2014). 模型平均、最优推理和习惯形成。*Frontiers in Human Neuroscience*, *8*, 1 – 11. doi:10.3389/fnhum.2014.00457.

Fitzhugh, R. (1958). 视神经信息的统计分析器。*Journal of General Physiology*, *41*, 675 – 692.

Fitzpatrick, P., Metta, G., Natale, L., Rao, S., & Sandini, G. (2003). 通过动作学习对象：迈向人工认知的初步步骤。*IEEE International Conference on Robotics and Automation (ICRA)*, 5月12 – 17日，台湾台北。

Flash, T., & Hogan, N. (1985). 手臂运动的协调：一个实验验证的数学模型。*J. Neurosci.*, *5*, 1688 – 1703.

Fletcher, P., & Frith, C. (2009). 感知即相信：解释精神分裂症阳性症状的贝叶斯方法。*Nature Reviews: Neuroscience*, *10*, 48 – 58.

Fodor, J. (1983). *心智的模块性*. 马萨诸塞州剑桥：MIT出版社。

Fodor, J. (1988). 对Churchland的”知觉可塑性和理论中性”的回复。*Philosophy of Science*, *55*, 188 – 198.

Fodor, J., & Pylyshyn, Z. (1988). 连接主义和认知架构：批判性分析。*Cognition*, *28*(1 – 2), 3 – 71.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). 顶叶：从动作组织到意图理解。*Science*, *308*, 662 – 667. doi:10.1126/science.1106138.

Fogassi, L., Gallese, V., Fadiga, L., & Rizzolatti, G. (1998). 在猕猴顶叶区PF(7b)中响应观察目标导向手/臂动作的神经元。*Soc. Neurosci. Abstr.*, *24*, 257.

Ford, J., & Mathalon, D. (2012). 预测未来：精神分裂症中的自动预测失败。*International Journal of Psychophysiology*, *83*, 232 – 239.

Fornito, A., Zalesky, A., Pantelis, C., & Bullmore, E. T. (2012). 精神分裂症、神经影像学和连接组学。*Neuroimage*, *62*, 2296 – 2314.

Frankish, K. (待出版). Dennett的推理双重过程理论(dual-process theory)。载于C. Muñoz-Suárez和F. De Brigard, F. (编)，*重新审视内容和意识*. 纽约：Springer。

Franklin, D. W., & Wolpert, D. M. (2011). 感觉运动控制的计算机制。*Neuron*, *72*, 425 – 442.

Freeman, T. C. A., Champion, R. A., & Warren, P. A. (2010). 平滑追踪眼动期间感知的头部中心速度的贝叶斯模型。*Current Biology*, *20*, 757 – 762.

Friston, K. (1995). 神经影像学中的功能性和有效连接：一个综合。*Hum. Brain Mapp.*, *2*, 56 – 78.

Friston, K. (2002). 超越颅相学：神经影像学能告诉我们关于分布式回路的什么？*Annu. Rev. Neurosci.*, *25*, 221 – 250.

Friston, K. (2003). 大脑中的学习和推理。*Neural Networks*, *16*(9), 1325 – 1352.

Friston, K. (2005). 皮层反应理论。*Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *360*(1456), 815 – 836.

Friston, K. (2008). 大脑中的分层模型(Hierarchical models). *PLoS Computational Biology*, *4*(11), e1000211.

Friston, K. (2009). 自由能原理(Free-energy principle)：大脑的粗略指南？*Trends in Cognitive Sciences*, *13*, 293 – 301.

Friston, K. (2010). 自由能原理(Free-energy principle)：统一的大脑理论？*Nat. Rev. Neurosci.*, *11*(2), 127 – 138.

Friston, K. (2011a). 运动控制的最优性是什么？*Neuron*, *72*, 488 – 498.

Friston, K. (2011b). 具身推理(Embodied inference)：或者说”我思故我在，如果我是我所思”。收录于 W. Tschacher & C. Bergomi (Eds.), *具身的含义（认知与交流）*(pp. 89 – 125). Exeter, UK: Imprint Academic.

Friston, K. (2011c). 功能性和有效性连接(Functional and effective connectivity)：综述。*Brain Connectivity*, *1*(1), 13 – 36. doi:10.1089/brain.2011.0008.

Friston, K. (2012a). 预测编码(Predictive coding)、精确性(precision)和同步性(synchrony)。*Cognitive Neuroscience*, 3(3 – 4), 238 – 239.

Friston, K. (2012b). 预测、感知和能动性(agency)。*Int. J. Psychophysiol.*, *83*, 248 – 252.

Friston, K. (2012c). 生物系统的自由能原理(Free energy principle)。*Entropy*, *14*, 2100 – 2121. doi:10.3390/e14112100.

Friston, K. (2013). 主动推理(Active inference)和自由能。*Behavioral and Brain Sciences*, *36*(3), 212 – 213.

Friston, K., Adams, R., & Montague, R. (2012). 什么是价值——累积奖励还是证据？*Frontiers in Neurorobotics*, *6*, 11. doi:10.3389/fnbot.2012.00011.

Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). 感知即假设(Perceptions as hypotheses)：扫视即实验(Saccades as experiments)。*Front. Psychol.*, *3*, 151. doi:10.3389/fpsyg.2012.00151.

Friston, K., & Ao, P. (2012). 自由能、价值和吸引子(attractors)。 *Computational and Mathematical Methods in Medicine*. Article ID 937860.

Friston, K. J., Bastos, A. M., Pinotsis, D., & Litvak, V. (2015). 局部场电位(LFP)和振荡——它们告诉我们什么？ *Current Opinion in Neurobiology*, *31*, 1 – 6. doi:10.1016/j.conb.2014.05.004

Friston, K., Breakspear, M., & Deco, G. (2012). 感知和自组织不稳定性(self-organized instability)。 *Front. Comput. Neurosci.*, *6*, 44. doi:10.3389/fncom.2012.00044.

Friston, K., Daunizeau, J., & Kiebel, S. J. (2009, July 29). 强化学习还是主动推理(Active inference)? *PLoS One*, *4*(7), e6421.

Friston, K., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). 行动和行为：自由能表述。 *Biol Cybern.*, *102*(3), 227 – 260.

Friston, K., Harrison, L., & Penny, W. (2003). 动态因果建模(Dynamic causal modelling)。 *Neuroimage*, *19*, 1273 – 1302.

Friston, K., & Kiebel, S. (2009). 感知推理的皮层回路。 *Neural Networks*, *22*, 1093 – 1104.

Friston, K., Lawson, R., & Frith, C. D. (2013). 关于超先验(hyperpriors)和次先验(hypopriors)：对Pellicano and Burr的评论。 *Trends in Cognitive. Sciences*, *17*, 1. doi:10.1016/j.tics.2012.11.003.

Friston, K., Mattout, J., & Kilner, J. (2011). 行动理解和主动推理。 *Biol. Cybern.*, *104*, 137 – 160.

Friston, K., & Penny, W. (2011). 事后贝叶斯模型选择(Post hoc Bayesian model selection)。 *Neuroimage*, *56*(4), 2089 – 2099.

Friston, K., & Price, C. J. (2001). 大脑功能的动态表征和生成模型。 *Brain Res. Bull.*, *54*(3), 275 – 285.

Friston, K., Samothrakis, S., & Montague, R. (2012). 主动推理和能动性：无成本函数的最优控制。 *Biol. Cybern.*, *106*(8 – 9), 523 – 541.

Friston, K., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., et al. (2012). 多巴胺、可供性(affordance)和主动推理。 *PLoS Comput. Biol.*, *8*(1), e1002327. doi:10.1371/journal.pcbi.1002327.

Friston, K., & Stephan, K. (2007). 自由能和大脑。 *Synthese*, *159*(3), 417 – 458.

Friston, K., Thornton, C., & Clark, A. (2012). 自由能最小化和暗室问题(dark-room problem)。 *Frontiers in Psychology*, *3*, 1 – 7. doi:10.3389/fpsyg.2012.00130.

Frith, C. (2005). 行动中的自我：从控制错觉(delusions of control)中学到的经验。 *Consciousness and Cognition*, *14*(4), 752 – 770.

Frith, C. (2007). *构建心智：大脑如何创造我们的心理世界*. Oxford: Blackwell.

Frith, C., & Friston, K. (2012). 错误感知和错误信念：理解精神分裂症。收录于 *神经科学与人类工作组：人类活动的新视角，教皇科学院*, 2012年11月8-10日, Casina PioIV.

Frith, C., & Frith, U. (2012). 社会认知的机制。 *Annu. Rev. Psychol.*, *63*, 287 – 313.

Frith, C., Perry, R., & Lumer, E. (1999). 意识体验的神经关联：实验框架。 *Trends in Cognitive Sciences*, *3*(3), 105 – 114.

Frith, U. (1989). *自闭症：解释这个谜团*. Oxford: Blackwell.

Frith, U. (2008). *自闭症：简明介绍*. New York: Oxford University Press.

Froese, T., & Di Paolo, E. A. (2011). 能动方法(enactive approach)：从细胞到社会的理论草图。*Pragmatics and Cognition*, *19*, 1 – 36.

Froese, T., & Ikegami, T. (2013). 大脑不是一个孤立的”黑盒子”，其目标也不是成为一个黑盒子。*Behavioral and Brain Sciences*, *36*(3), 33 – 34.

Gallagher, S., Hutto, D., Slaby, J., & Cole, J. (2013). 大脑作为能动系统(enactive system)的一部分。*Behavioral and Brain Sciences*, *36*(4), 421 – 422.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). 前运动皮层中的动作识别。*Brain*, *119*, 593 – 609.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). 社会认知基础的统一观点。*Trends in Cognitive Sciences*, *8*(9), 396 – 403.

Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). 视觉心理意象(visual mental imagery)和视觉感知的脑区：fMRI研究。*Brain Res. Cogn. Brain Res.*, *20*, 226 – 241.

Garrod, S., & Pickering, M. (2004). 为什么对话如此容易？*Trends in Cognitive Sciences*, *8*(1), 8 – 11.

Gazzola, V., & Keysers, C. (2009). 动作的观察和执行在所有测试对象中共享运动和体感体素：未平滑fMRI数据的单受试者分析。*大脑皮层*, *19*(6), 1239 – 1255。

Gendron, M., & Barrett, L. F. (2009). 重构过去：心理学中关于情绪的一个世纪观念。*情绪评论*, *1*, 316 – 339。

Gerrans, P. (2007). 疯狂的机制：没有进化心理学的进化精神病学。*生物学与哲学*, *22*, 35 – 56。

Gershman, S. J., & Daw, N. D. (2012). 感知、行动和效用：纠缠的线团。载于M. Rabinovich, M., K. Friston, & P. Varona (主编), *大脑动力学原理：全局状态相互作用* (pp. 293 – 312)。马萨诸塞州剑桥：MIT出版社。

Gershman, S. J., Moore, C. D., Todd, M. T., Norman, K. N., & Sederberg, P. B. (2012). 后继表征(successor representation)和时间情境。*神经计算*, *24*, 1 – 16。

Gibson, J. J. (1977). 可供性理论(affordances)。载于R. Shaw & J. Bransford (主编), *感知、行动和认知* (pp. 66 – 82)。新泽西州希尔斯代尔。

Gibson, J. J. (1979). *视觉感知的生态学方法*。波士顿：霍顿-米夫林出版社。

Gigernzer, G. & Goldstein, D. (2011). 识别启发式(recognition heuristic)：十年研究。*判断与决策制定* 6: 1: 100 – 121。

Gigerenzer, G., & Selten, R. (2002). *有限理性(bounded rationality)*。马萨诸塞州剑桥：MIT出版社。

Gigerenzer, G., Todd, P. M., & ABC研究小组. (1999). *让我们变聪明的简单启发式*。纽约：牛津大学出版社。

Gilbert, C., & Sigman, M. (2007). 大脑状态：感觉处理中的自上而下影响。*神经元*, *54*(5), 677 – 696。

Gilbert, D., & Wilson, T. (2009年5月12日). 为什么大脑会自言自语：情绪预测中的错误来源。*伦敦皇家学会哲学交易，B系列，生物科学*, *364*(1521), 1335 – 1341。doi:10.1098/rstb.2008.0305。

Gilestro, G. F., Tononi, G., & Cirelli, C. (2009). 果蝇睡眠和清醒功能中突触标记物的广泛变化。*科学*, *324*(5923), 109 – 112。

Glascher, J., Daw, N. D., Dayan, P., & O’Doherty, J. P. (2010). 状态对比奖励：基于模型和无模型强化学习中不同的神经预测错误信号。*神经元*, 66(4):585 – 95 doi:10.1016/j.neuron.2010.04.016。

Gold, J. N., & Shadlen, M. N. (2001). 支撑感觉刺激决策的神经计算。*认知科学趋势*, 5:1: pp. 10 – 16。

Goldstone, R. L. (1994). 分类对感知辨别的影响。*实验心理学杂志：一般*, *123*, 178 – 200。

Goldstone, R. L., & Hendrickson, A. T. (2010). 分类感知(categorical perception)。*跨学科评论：认知科学*, *1*, 65 – 78。

Goldstone, R. L., Landy, D., & Brunel, L. (2011) 改善感知以拉近远距离连接。*心理学前沿*, *2*, 385。doi:10.3389/fpsyg.2011.00385。

Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). 词汇分割的贝叶斯框架：探索语境效应。*认知*, *112*, 21 – 54。

Goodman, N., Ullman, T., & Tenenbaum, J. (2011). 学习因果理论。*心理学评论* 118.1: 110 – 119。

Gregory, R. (2001). 2001年梅达沃讲座：视觉的知识：知识的视觉。*哲学交易B*, 360. 1458: 1231 – 1251。

Gregory, R. (1980). 感知作为假设。*伦敦皇家学会哲学交易，B系列，生物科学*, *290*(1038), 181 – 197。

Griffiths, P. E., & Gray, R. D. (2001). 达尔文主义和发展系统。载于S. Oyama, P. E. Griffiths, & R. D. Gray (主编), *偶然性循环：发展系统与进化* (pp. 195 – 218)。马萨诸塞州剑桥：MIT出版社。

Griffiths, T., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). 认知的概率模型：探索表征和归纳偏见。*认知科学趋势*, *14*(8), 357 – 364。

Griffiths, T. L., Sanborn, A. N., Canini, K. R., & Navarro, D. J. (2008). 分类作为非参数贝叶斯密度估计。载于M. Oaksford & N. Chater (主编), *概率心智：认知理性模型的前景* 303-350。牛津：牛津大学出版社。

Grill-Spector, K., Henson, R., & Martin, A. (2006). 重复与大脑：刺激特异性效应的神经模型。*认知科学趋势*, *10*(1), 14 – 23。

Grossberg, S. (1980). 大脑如何构建认知代码？*心理学评论*, *87*(1), 1 – 51。

Grush, R. (2004). 表征的模拟理论：运动控制、想象和感知。*行为与脑科学*, *27*, 377 – 442。

Gu, X., et al. (2013). 前脑岛皮层和情绪意识。*比较神经学杂志*, *521*, 3371 – 3388。

Haddock, A., & Macpherson, F. (主编). (2008). *析取主义：感知、行动和知识*。牛津：牛津大学出版社。

Happe, F., & Frith, U. (2006). 弱连贯性解释：自闭症谱系障碍中的细节聚焦认知风格。*自闭症发展障碍杂志*, *36*, 5 – 25。

Happe, F. G. (1996). 在低水平研究弱中央连贯性：自闭症儿童不会屈服于视觉错觉。研究报告。*儿童心理学与精神病学杂志*, *37*, 873 – 877。

Harman, G. (1990). 体验的内在品质。载于J. Tomberlin (主编), *哲学视角4* (pp. 64 – 82)。加州阿塔斯卡德罗：山脊景出版社。

Harmelech, T., & Malach, R. 2013 神经认知偏见和人类皮层自发相关性模式。*认知科学趋势*, *17*(12), 606 – 615。

Harris, C. M., & Wolpert, D. M. (1998). 信号依赖噪声决定运动规划。*自然*, *394*, 780 – 784。doi:10.1038/29528。

Harris, C. M., & Wolpert, D. M. (2006). 扫视的主序列优化速度-准确性权衡。*生物控制论*, *95*(1), 21 – 29。

Harrison, L., Bestmann, S., Rosa, M., Penny, W., & Green, G. (2011). 人类大脑中的表征时间尺度：权衡过去信息以预测未来事件。*人类神经科学前沿*，*5*，37。doi:10.3389/fnhum.2011.00037。

Haruno, M., Wolpert, D. M., & Kawato, M. (2003). 运动生成的分层马赛克。*国际会议系列*，*1250*，575 – 590。

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). 海马体失忆症患者无法想象新体验。*美国国家科学院院刊*，*104*，1726 – 1731。doi:10.1073/pnas.0610561104。

Hassabis, D., & Maguire, E. A. (2009). 大脑的构建系统。*英国皇家学会哲学会刊B*，*364*，1263 – 1271。doi:10.1098/rstb.2008.0296。

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). 脑对脑耦合：创造和分享社会世界的机制。*认知科学趋势*，*16*(2)，114 – 121。

Hawkins, J., & Blakeslee, S. (2004). *论智能*。纽约：猫头鹰图书。

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). 腹侧颞叶皮层中面孔和物体的分布式重叠表征。*科学*，*293*，2425 – 2430。

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). 自然任务中的视觉记忆和运动规划。*视觉期刊*，*3*(1)，49 – 63。

Hebb, D. O. (1949). *行为的组织*。纽约：威利父子出版社。

Heeger, D., & Ress, D. (2002). fMRI告诉我们关于神经元活动的什么信息？*自然评论/神经科学*，*3*，142 – 151。

Helmholtz, H. (1860/1962). *生理光学手册*（J. P. C. Southall编，英译本，第3卷）。纽约：多佛出版社。

Henson, R. (2003). 启动的神经影像学研究。*神经生物学进展*，*70*，53 – 81。

Henson, R., & Gagnepain, P. (2010). 预测性、交互式多重记忆系统。*海马体*，*20*(11)，1315 – 1326。

Herzfeld, D., & Shadmehr, R. (2014). 小脑估计身体的感觉状态。*认知神经科学趋势*，*18*，66 – 67。

Hesselmann, G., Kell, C., Eger, E., & Kleinschmidt, A. (2008). 持续神经活动中的自发局部变异偏向知觉决策。*美国国家科学院院刊*，*105*(31)，10984 – 10989。

Hesselmann, G., Sadaghiani, S., Friston, K. J., & Kleinschmidt, A. (2010). 预测编码还是证据积累？错误推理和神经元波动。*公共科学图书馆综合*，*5*(3)，e9926。

Hesslow, G. (2002). 有意识思维作为行为和知觉的模拟。*认知科学趋势*，*6*(6)，242 – 247。

Heyes, C. (2001). 模仿的原因和后果。*认知科学趋势*，*5*，253 – 261。doi:10.1016/S1364-6613(00)01661-2。

Heyes, C. (2005). 通过联想进行模仿。见S. Hurley & N. Chater（编），*模仿的视角：从镜像神经元到模因*（第157 – 176页）。剑桥，马萨诸塞州：麻省理工学院出版社。

Heyes, C. (2010). 镜像神经元从何而来？*神经科学与生物行为评论*，*34*，575 – 583。

Heyes, C. (2012). 新思维：人类认知的进化。*英国皇家学会哲学会刊B* 367，2091 – 2096。

Hilgetag, C., Burns, G., O’Neill, M., Scannell, J., & Young, M. (2000). 解剖连接性定义了猕猴和猫皮层区域集群的组织。*伦敦皇家学会哲学会刊B生物科学*，*355*，91 – 110。doi:10.1098/rstb.2000.0551；pmid: 10703046。

Hilgetag, C., O’Neill, M., & Young, M. (1996). 视觉系统的不确定组织。*科学*，*271*，776 – 777。doi:10.1126/science.271.5250.776；pmid: 8628990。

Hinton, G. E. (1990). 将部分-整体层次结构映射到连接主义网络。*人工智能*，*46*，47 – 75。

Hinton, G. E. (2005). 大脑是什么样的图形模型？*国际人工智能联合会议*，爱丁堡。

Hinton, G. E. (2007a). 学习多层表征。*认知科学趋势*，*11*，428 – 434。

Hinton, G. E. (2007b). 要识别形状，首先学会生成图像。见P. Cisek, T. Drew & J. Kalaska（编），*计算神经科学：对大脑功能的理论洞察*（第535 – 548页）。阿姆斯特丹：爱思唯尔。

Hinton, G. E., Dayan, P., Frey, B. J., & Neal, R. M. (1995). 无监督神经网络的觉醒-睡眠算法。*科学*，*268*，1158 – 1160。

Hinton, G. E., & Ghahramani, Z. (1997). 发现稀疏分布式表征的生成模型。*英国皇家学会哲学会刊B*，*352*，1177 – 1190。

Hinton, G. E., & Nair, V. (2006). 从手写数字图像推断运动程序。见Y. Weiss（编），*神经信息处理系统进展*，18（第515 – 522页）。剑桥，马萨诸塞州：麻省理工学院出版社。

Hinton, G. E., Osindero, S., & Teh, Y. (2006). 深度信念网络的快速学习算法。*神经计算*，*18*，1527 – 1554。

Hinton, G. E., & Salakhutdinov, R. R. (2006). 用神经网络降低数据维度。*科学*，*313*(5786)，504 – 507。

Hinton, G. E., & von Camp, D. (1993). 通过最小化权重描述长度保持神经网络简单。*COLT-93会议录*，5 – 13。

Hinton, G. E., & Zemel, R. S. (1994). 自编码器、最小描述长度和亥姆霍兹自由能。见J. Cowan, G. Tesauro, & J. Alspector（编），*神经信息处理系统进展*，6。圣马特奥，加利福尼亚州：摩根考夫曼。

Hipp, J. F., Engel, A. K., & Siegel, M. (2011). 大规模皮层网络中的振荡同步预测知觉。*神经元*，*69*，387 – 396。

Hirschberg, L. (2003). 被叙事吸引。*纽约时报杂志*，11月9日。

Hobson, J., & Friston, K. (2012). 清醒和梦境意识：神经生物学和功能考虑。*神经生物学进展*，98(1)，82－98。

Hobson, J., & Friston, K. (2014). 意识、梦境和推理。《意识研究期刊》，21(1-2)，6-32。

Hobson, J. A. (2001).《梦境药店：化学改变的意识状态》。马萨诸塞州剑桥：麻省理工学院出版社。

Hochstein, S., & Ahissar, M. (2002). 从顶部观察：视觉系统中的层次结构和反向层次结构。《神经元》，36(5)，791-804。

Hohwy, J. (2007a). 功能整合与心智。《综合》，159(3)，315-328。

Hohwy, J. (2007b). 现象学中能动性和感知的自我感。《心理》，13(1) 1-20（Susanna Siegel编）。

Hohwy, J. (2012). 假设检验大脑中的注意力和意识感知。《心理学前沿》，3，96。doi:10.3389/fpsyg.2012.00096. 2012。

Hohwy, J. (2013).《预测性心智》。纽约：牛津大学出版社。

Hohwy, J. (2014). 自我证明大脑。《理性》。1-27 doi: 10.1111/nous.12062。

Hohwy, J., Roepstorff, A., & Friston, K. (2008). 预测编码解释双眼竞争：认识论回顾。《认知》，108(3)，687-701。

Holle, H., et al. (2012). 传染性瘙痒的神经基础以及为什么有些人更容易患病。《美国国家科学院院刊》，109，19816-19821。

Hollerman, J. R., & Schultz, W. (1998). 多巴胺神经元在学习过程中报告奖励时间预测误差。《自然神经科学》，1，304-309。

Hong, L. E., Avila, M. T., & Thaker, G. K. (2005). 精神分裂症患者在持续视觉跟踪过程中对意外目标变化的反应。《实验脑研究》，165，125-131。PubMed: 15883805。

Hong, L. E., Turano, K. A., O’Neill, J., Hao, L. I. W., McMahon, R. P., Elliott, A., & Thaker, G. K. (2008). 精神分裂症中预测性追踪内表型的细化。《生物精神病学》，63，458-464。PubMed: 17662963。

Horwitz, B. (2003). 难以捉摸的大脑连接性概念。《神经影像》，19，466-470。

Hoshi, E., & Tanji, J. (2007). 背侧和腹侧前运动区的区别：解剖连接性和功能特性。《当前神经生物学观点》，17(2)，234-242。

Hosoya, T., Baccus, S. A., & Meister, M. (2005). 视网膜的动态预测编码。《自然》，436(7)，71-77。

Houlsby, N. M. T., Huszár, F., Ghassemi, M. M., Orbán, G., Wolpert, D. M., & Lengyel, M. (2013). 认知断层扫描揭示复杂的、任务无关的心理表征。《当前生物学》，23，2169-2175。

Howe, C. Q., & Purves, D. (2005). 用图像-源关系统计解释米勒-莱尔错觉。《美国国家科学院院刊》，102(4)，1234-1239。doi:10.1073/pnas.0409314102。

Huang, Y., & Rao, R. (2011). 预测编码。《威利跨学科综述：认知科学》，2，580-593。

Hubel, D. H., & Wiesel, T. N. (1965). 猫的两个非纹状体视觉区域（18和19）的感受野和功能结构。《神经生理学期刊》，28，229-289。

Hughes, H. C., Darcey, T. M., Barkan, H. I., Williamson, P. D., Roberts, D. W., & Aslin, C. H. (2001). 人类听觉联合皮层对预期声学事件遗漏的反应。《神经影像》，13，1073-1089。

Humphrey, N. (2000). 如何解决心身问题。《意识研究期刊》，7，5-20。

Humphrey, N. (2006).《看到红色：意识研究》。马萨诸塞州剑桥：哈佛大学出版社。

Humphrey, N. (2011).《灵魂尘埃：意识的魔力》。新泽西州普林斯顿：普林斯顿大学出版社。

Humphreys, G. W., & Riddoch, J. M. (2000). 再来一杯咖啡：额叶损伤后的物体-动作组合、反应阻断和反应捕获。《实验脑研究》，133，81-93。

Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). 皮层反馈改善V1、V2和V3神经元对图形和背景的区分。《自然》，394，784-787。

Hurley, S. (1998).《行动中的意识》。马萨诸塞州剑桥：哈佛大学出版社。

Hutchins, E. (1995).《野生认知》。马萨诸塞州剑桥：麻省理工学院出版社。

Hutchins, E. (2014). 人类认知的文化生态系统。《哲学心理学》，27(1)，34-49。

Hutto, D. D., & Myin, E. (2013).《激进化身化主义：没有内容的基本心智》。

Iacoboni, M. (2009). 模仿、共情和镜像神经元。《心理学年度综述》，60，653-670。

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). 用自己的镜像神经元系统把握他人的意图。《公共科学图书馆生物学》，3，e79。

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). 人类模仿的皮层机制。《科学》，286(5449)，2526-2528。

Iglesias, S., Mathys, C., Brodersen, K. H., Kasper, L., Piccirelli, M., den Ouden, H. E., & Stephan, K. E. (2013). 感觉学习过程中中脑和基底前脑的层次预测误差。《神经元》，80(2)，519-530。doi:10.1016/j.neuron.2013.09.009。

Ingvar, D. H. (1985). “未来的记忆”：关于意识觉知时间组织的论文。《人类神经生物学》，4，127-136。

Ito, M., & Gilbert, C. D. (1999). 注意力调节警觉猴子初级视觉皮层的情境影响。《神经元》，22，593-604。

Jackendoff, R. (1996). 语言如何帮助我们思考。《语用学与认知》，4(1)，1-34。

Jackson, F. (1977).《感知：表征理论》。剑桥：剑桥大学出版社。

Jacob, B., Hirsh, R., Mar, A., & Peterson, J. B. (2013). 个人叙事作为认知整合的最高层次。《行为与脑科学》，36，216-217。doi:10.1017/S0140525X12002269。

Jacob, P., & Jeannerod, M. (2003).《观看方式：视觉认知的范围和局限》。牛津：牛津大学出版社。

Jacobs, R. A., & Kruschke, J. K. (2010). 应用于人类认知的贝叶斯学习理论。*威利跨学科评论：认知科学*, *2*, 8 – 21. doi:10.1002/wcs.80.

Jacoby, L. L., & Dallas, M. (1981). 自传体记忆与知觉学习的关系。*实验心理学杂志：综合*, *110*, 306 – 340.

James, W. (1890/1950). *心理学原理*，第一卷、第二卷。剑桥，马萨诸塞州：哈佛大学出版社。

Jeannerod, M. (1997). *动作的认知神经科学*。剑桥，马萨诸塞州：布莱克韦尔出版社。

Jeannerod, M. (2006). *运动认知：动作告诉自我什么*。牛津：牛津大学出版社。

Jehee, J. F. M., & Ballard, D. H. (2009). 预测性反馈可以解释外侧膝状体核的双相反应。*公共科学图书馆计算生物学*, *5*(5), e1000373.

Jiang, J., Heller, K., & Egner, T. (2014). 灵活认知控制的贝叶斯建模。*神经科学与生物行为评论*, 46, 30 – 34.

Johnson, J. D., McDuff, S. G. R., Rugg, M. D., & Norman, K. A. (2009). 回忆、熟悉感和皮质重现：多体素模式分析。*神经元*, *63*, 697 – 708.

Johnson-Laird, P. N. (1988). *计算机与心智：认知科学导论*。剑桥，马萨诸塞州：哈佛大学出版社。

Jones, A., Wilkinson, H. J., & Braden, I. (1961). 信息剥夺作为动机变量。*实验心理学杂志*, *62*, 310 – 311.

Joseph, R. M., et al. (2009). 为什么自闭症谱系障碍患者的视觉搜索能力更优越？*发展科学*, *12*, 1083 – 1096.

Joutsiniemi, S. L., & Hari, R. (1989) 听觉刺激的缺失可能激活额叶皮质。*欧洲神经科学杂志*, *1*, 524 – 528.

Jovancevic, J., Sullivan, B., & Hayhoe, M. (2006). 复杂环境中的注意力和凝视控制。*视觉杂志*, 6(12):9, 1431 – 1450. http://www.journalofvision.org/content/6/12/9, doi:10.1167/6.12.9.

Jovancevic-Misic, J., & Hayhoe, M. (2009). 自然环境中的适应性凝视控制。*神经科学杂志*, *29*, 6234 – 6238.

Joyce, J. (2008). 贝叶斯定理。*斯坦福哲学百科全书* (2008).

Kachergis, G., Wyatte, D., O’Reilly, R. C., de Kleijn, R., & Hommel, B. (2014). 连续动作的连续时间神经模型。英国皇家学会哲学汇刊B 369: 20130623.

Kahneman, D. (2011). *思考，快与慢*。伦敦：艾伦·莱恩出版社。

Kahneman, D., & Tversky, A. (1972). 主观概率：代表性判断。*认知心理学* **3** (3), 430 – 454. doi:10.1016/0010-0285(72)90016-3

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. A. (2006). 如果你更富有，你会更快乐吗？聚焦错觉。*科学*, *312*, 1908 – 1910.

Kaipa, K. N., Bongard, J. C., & Meltzoff, A. N. (2010). 自我发现使机器人社会认知成为可能：你是我的老师吗？*神经网络*, *23*, 1113 – 1124.

Kamitani, Y., & Tong, F. (2005). 解码人脑的视觉和主观内容。*自然神经科学*, *8*, 679 – 685.

Kanwisher, N. G., McDermott, J., & Chun, M. M. (1997). 梭状回面孔区：人类纹外皮质中专门用于面孔感知的模块。*神经科学杂志*, *17*, 4302 – 4311.

Kaplan, E. (2004). 灵长类视觉系统的M、P和K通路。在J. S. Werner & L. M. Chalupa (编辑), *视觉神经科学* (第481 – 493页). 剑桥，马萨诸塞州：麻省理工学院出版社。

Kawato, M. (1999). 运动控制和轨迹规划的内部模型。*神经生物学当前观点*, *9*, 718 – 727.

Kawato, M., Hayakama, H., & Inui, T. (1993). 视觉皮质区域间相互连接的前向-逆向光学模型。*网络*, *4*, 415 – 422.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). 从人脑活动中识别自然图像。*自然*, *452*, 352 – 355.

Keele, S. W. (1968). 熟练运动表现中的运动控制。*心理学通报*, *70*, 387 – 403. doi:10.1037/h0026739.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). 用分层贝叶斯模型学习超假设。*发展科学*, *10*(3), 307 – 321.

Keysers, C., Kaas, J. H., & Gazzola, V. (2010). 社会感知中的体感。*自然神经科学评论*, *11*(6), 417 – 428.

Khaitan, P., & McClelland, J. L. (2010). 在多项式交互激活模型中匹配精确后验概率。在S. Ohlsson & R. Catrambone (编辑), *第32届认知科学学会年会论文集* (第623页). 奥斯汀，德克萨斯州：认知科学学会。

Kidd, C., Piantadosi, S., & Aslin, R. (2012). 金发姑娘效应：人类婴儿将注意力分配给既不太简单也不太复杂的视觉序列。*公共科学图书馆综合*, *7*(5), e36399. doi:10.1371/journal.pone.0036399.

Kiebel, S. J., & Friston K. J. (2011) 自由能和树突自组织。*系统神经科学前沿* 5(80), 1 – 13.

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2009). 感知和分层动力学。*神经信息学前沿*, *3*, 20.

Kiebel, S. J., Garrido, M. I., Moran, R., Chen, C., & Friston, K. J. (2009). EEG和MEG的动态因果建模。*人脑映射*, *30*(6), 1866 – 1876.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). 预测编码：镜像神经元系统的解释。*认知过程*, *8*, 159 – 166.

Kim, J. G., & Kastner, S. (2013). 注意力灵活性改变对象类别的调节。*认知科学趋势*, *17*(8), 368 – 370.

King, J., & Dehaene, S. (2014). 主观报告和客观辨别作为广阔表征空间中分类决策的模型。*英国皇家学会哲学汇刊B*, 369.

Kirmayer, L. J., & Taillefer, S. (1997). 躯体形式障碍。在S. M. Turner & M. Hersen (编辑), *成人精神病理学与诊断* (第3版) (第333 – 383页). 纽约：威利出版社。

Kluzik, J., Diedrichsen, J., Shadmehr, R., & Bastian, A. J. (2008). 到达适应：是什么决定了我们学习工具的内部模型还是适应我们手臂的模型？*J. Neurophysiol.*, *100*, 1455 – 1464.

Knill, D., & Pouget, A. (2004). 贝叶斯大脑：不确定性在神经编码和计算中的作用。*Trends in Neurosciences*, *27*(12), 712 – 719.

Koch, C., & Poggio, T. (1999). 预测视觉世界：沉默是金。*Nature Neuroscience*, *2*(1), 79 – 87.

Koch, C., & Ullman, S. (1985). 选择性视觉注意的转换：探索潜在的神经回路。*Human Neurobiology*, *4*, 219 – 227.

Kohonen, T. (1989). *自组织和联想记忆*。柏林：Springer-Verlag出版社。

Kohonen, T. (2001). *自组织映射*（第3版，扩展版）。柏林：Springer出版社。

Kok, P., Brouwer, G. J., van Gerven, M. A. J., & de Lange, F. P. (2013). 先验期望偏向视觉皮层中的感觉表征。*Journal of Neuroscience.* 33(41): 16275 – 16284; doi: 10.1523/JNEUROSCI.0742-13.2013.

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). 少即是多：期望锐化初级视觉皮层中的表征。*Neuron*, *75*, 265 – 270.

Kolossa, A., Fingscheidt, T., Wessel, K., & Kopp, B. (2013). 基于模型的试验间P300幅度波动方法。*Frontiers in Human Neuroscience*, *6*(359), 1 – 18. doi:10.3389/fnhum.2012.00359.

Kolossa, A., Kopp, B., & Fingscheidt, T. (2015). 贝叶斯推理神经基础的计算分析。*NeuroImage*, *106*, 222 – 237. doi:10.1016/*j.neuroimage*.2014.11.007.

König, P., & Krüger, N. (2006). 符号作为特征提取和预测优化过程中的自发涌现实体。*Biological Cybernetics*, *94*(4), 325 – 334.

König, P., Wilming, N., Kaspar, K., Nagel, S. K., & Onat, S. (2013). 基于自身行动范围的预测作为一般计算原理。*Behavioral and Brain Sciences*, *36*, 219 – 220.

Körding, K., & Wolpert, D. (2006). 感觉运动控制中的贝叶斯决策理论。*Trends in Cognitive Sciences*, *10*(7), 319 – 326. doi:10.1016/j.tics.2006.05.003.

Körding, K. P., Tenenbaum, J. B., & Shadmehr, R. (2007). 记忆动态作为身体变化最优适应的结果。*Nature Neuroscience*, *10*, 779 – 786.

Kosslyn, S. M., Thompson, W. L., Kim, I. J., & Alpert, N. M. (1995). 初级视觉皮层中心理图像的拓扑表征。*Nature*, *378*, 496 – 498.

Koster-Hale, J., & Saxe, R. (2013). 心理理论(Theory of mind)：一个神经预测问题。*Neuron*, *79*(5), 836 – 848.

Kriegstein, K., & Giraud, A. (2006). 隐式多感官联想影响声音识别。*PLoS Biology*, *4*(10), e326.

Kuo, A. D. (2005). 人体姿势平衡中感觉整合的最优状态估计模型。*J. Neural Eng.*, *2*, S235 – S249.

Kveraga, K., Ghuman, A., & Bar, M. (2007). 认知大脑中的自上而下预测。*Brain and Cognition*, *65*, 145 – 168.

Kwisthout, J., & van Rooij, I. (2013). 弥合近似贝叶斯推理理论与实践之间的差距。*Cognitive Systems Research*, *24*, 2 – 8.

Laeng, B., & Sulutvedt, U. (2014). 眼瞳调节到想象的光线。*Psychol. Sci.*, *1*, 188 – 197. doi:10.1177/0956797613503556. 电子版2013年11月27日。

Land, M. (2001). 汽车驾驶是否涉及速度流场的感知？载于J. M. Zanker & J. Zeil（编），*运动视觉：计算、神经和生态约束*（第227 – 235页）。柏林：Springer Verlag出版社。

Land, M. F., & McLeod, P. (2000). 从眼球运动到行动：击球手如何击球。*Nature Neuroscience*, *3*, 1340 – 1345.

Land, M. F., Mennie, N., & Rusted, J. (1999). 视觉和眼球运动在日常生活活动控制中的作用。*Perception*, *28*, 1311 – 1328.

Land, M. F., & Tatler, B. W. (2009). *观看和行动：自然行为中的视觉和眼球运动*。牛津：牛津大学出版社。

Landauer, T. K., & Dumais, S. T. (1997). 柏拉图问题的解决方案：知识获取、归纳和表征的潜在语义分析理论。*Psychological Review*, *104*, 211 – 240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). 潜在语义分析(Latent Semantic Analysis)介绍。*Discourse Processes*, *25*, 259 – 284.

Landauer, T. K., McNamara, D. S., Dennis, S., & W. Kintsch（编）。（2007）。*潜在语义分析手册*。新泽西州马瓦：Lawrence Erlbaum Associates出版社。

Landy, D., & Goldstone, R. L. (2005). 我们如何学习我们尚未理解的事物。*Journal of Experimental and Theoretical Artificial Intelligence*, *17*, 343 – 369. doi:0.1080/09528130500283832.

Lange, C. G. (1885). *Om sindsbevaegelser: Et psyko-fysiologisk studie.*哥本哈根：Jacob Lunds出版社。

Langner, R., Kellermann, T., Boers, F., Sturm, W., Willmes, K., & Eickhoff, S. B. (2011). 模态特异性感知期望选择性调节听觉、体感和视觉皮层的基线活动。*Cerebral Cortex*, *21*(12), 2850 – 2862.

Lauwereyns, J. (2012). *大脑与凝视：论视觉的主动边界*。马萨诸塞州剑桥：MIT出版社。

LeDoux, J. E. (1995). 情感：来自大脑的线索。*Annual Review of Psychology*46, 209 – 235.

Lee, D., & Reddish, P. (1981). 俯冲的塘鹅：生态光学的范式。*Nature*, *293*, 293 – 294.

Lee, M. (2010). 贝叶斯模型中的涌现和结构化认知：对Griffiths等人和McClelland等人的评论。*Trends in Cognitive Sciences*, *14*(8), 345 – 346.

Lee, R., Walker, R., Meeden, L., & Marshall, J. (2009). 基于类别的内在动机。载于*第九届表观遗传机器人学国际会议论文集*。检索自http://www.cs.swarthmore.edu/~meeden/papers/meeden.epirob09.pdf。

Lee, S. H., Blake, R., & Heeger, D. J. (2005). 双眼竞争期间初级视觉皮层中的活动行进波. *Nature Neuroscience*, *8*(1), 22 – 23.

Lee, T. S., & Mumford, D. (2003). 视觉皮层中的分层贝叶斯推理. *Journal of Optical Society of America, A*, *20*(7), 1434 – 1448.

Lenggenhager, B., et al. (2007). 视频故我在：操控身体自我意识. *Science*, *317*(5841), 1096 – 1099.

Leopold, D., & Logothetis, N. (1999). 多稳态现象：感知中的变化观点. *Trends in Cognitive Sciences*, *3*, 254 – 264.

Levine, J. (1983). 唯物主义与感受性质(qualia)：解释鸿沟. *Pacific Philosophical Quarterly*, *64*, 354 – 361.

Levy, D. L., Sereno, A. B., Gooding, D. C., & O’Driscoll, G. A. (2010). 精神分裂症中的眼动追踪功能障碍：特征和病理生理学. *Current Topics in Behavioral Neuroscience*, *4*, 311 – 347.

Li, Y., & Strausfeld, N. J. (1999). 蟑螂蘑菇体内侧叶中的多模态传出和回归神经元. *J. Comp. Neurol.*, *409*, 603 – 625.

Limanowski, J., & Blankenburg, F. (2013). 最小自我模型和自由能原理. *Frontiers in Human Neuroscience*, *7*, 547. doi:10.3389/fnhum.2013.00547.

Littman, M.L., Majercik, S.M., and Pitassi, T. (2001). 随机布尔可满足性. *J. Autom. Reason.* 27, 251 – 296.

Lotze, H. (1852). *医学心理学或灵魂生理学*. 莱比锡，德国：Weidmannsche Buchhandlung.

Lovero, K. L., et al. (2009). 前脑岛皮层预期即将到来的刺激重要性. *Neuroimage*, *45*, 976 – 983.

Lowe, R., & Ziemke, T. (2011). 行动倾向的感觉：论目标导向行为的情感调节. *Frontiers in Psychology*, *2*, 346. doi:10.3389/fpsyg.2011.00346.

Lungarella, M., & Sporns, O. (2005). 信息自我结构化：学习和发展的关键原理. *Proceedings 2005 IEEE Intern. Conf. Development and Learning*, 25 – 30.

Lupyan, G. (2012a). 语言调节的感知和认知：标签反馈假说. *Frontiers in Cognition*, *3*, 54. doi:10.3389/fpsyg.2012.00054.

Lupyan, G. (2012b). 词语的作用是什么？走向语言增强思维理论. In B. H. Ross (Ed.), *学习与动机心理学*, 57 (pp. 255 – 297). 纽约：学术出版社.

Lupyan, G. (正在出版). 预测时代感知的认知渗透性：预测系统是可渗透系统. *Review of Philosophy and Psychology*.

Lupyan, G., & Bergen, B. (正在出版). 语言如何编程心智. *Topics in Cognitive Science*.

Lupyan, G., & Clark, A. (正在出版). 词语与世界：预测编码和语言-感知-认知接口. *Current Directions in Psychological Science*.

Lupyan, G., & Thompson-Schill, S. L. (2012). 词语的唤起力量：通过言语和非言语手段激活概念. *Journal of Experimental Psychology-General*, *141*(1), 170 – 186. doi:10.1037/a0024904.

Lupyan, G., & Ward, E. J. (2013). 语言可以将原本不可见的物体提升到视觉意识中. *Proceedings of the National Academy of Sciences*, *110*(35), 14196 – 14201. doi:10.1073/pnas.1303312110.

Ma, W. Ji (2012). 组织感知的概率模型. *Trends in Cognitive Sciences*, *16*(10), 511 – 518.

MacKay, D. (1956). 自动机的认识论问题. In C. E. Shannon & J. McCarthy (Eds.), *自动机研究* (pp. 235 – 251). 普林斯顿，新泽西州：普林斯顿大学出版社.

MacKay, D. J. C. (1995). 用于解码和密码分析的自由能最小化算法. *Electron Lett.*, *31*, 445 – 447.

Maher, B. (1988). 异常体验和妄想思维：解释的逻辑. In T. F. Oltmanns & B. A. Maher (Eds.), *妄想信念* (pp. 15 – 33). 奇切斯特：威利.

Maslach, C (1979). 未解释唤起的负性情感偏倚. *Journal of Personality and Social Psychology* 37: 953 – 969. doi:10.1037/0022-3514.37.6.953.

Maloney, L. T., & Mamassian, P. (2009). 贝叶斯决策理论作为视觉感知模型：测试贝叶斯迁移. *Visual Neuroscience*, *26*, 147 – 155.

Mamassian, P., Landy, M., & Maloney, L. (2002). 视觉感知的贝叶斯建模. In R. Rao, B. Olshausen, & M. Lewicki (Eds.), *大脑的概率模型* 13-36 剑桥，马萨诸塞州：麻省理工学院出版社.

Mansinghka, V. K., Kemp, C., Tenenbaum, J. B., & Griffiths, T. L. (2006). 结构学习的结构化先验. In *人工智能不确定性第22届会议论文集 (UAI)* 324 – 331 阿灵顿，弗吉尼亚州：AUAI出版社.

Marcus, G., & Davis, E. (2013). 高级认知的概率模型有多稳健？ *Psychol. Sci.*, *24*(12), 2351 – 2360. doi:10.1177/0956797613495418.

Mareschal, D., Johnson, M., Sirois, S., Spratling, M., Thomas, M., & Westermann, G. (2007). *神经构造主义—I：大脑如何构造认知*. 牛津，牛津大学出版社.

Markov, N. T., Ercsey-Ravasz, M., Van Essen, D. C., Knoblauch, K., Toroczkai, Z., & Kennedy, H. (2013). 皮层高密度逆流架构. *Science*, *342*(6158). doi:10.1126/science.1238406.

Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., Lamy, C., Misery, P., Giroud, P., Ullman, S., Barone, P., Dehay, C., Knoblauch, K., & Kennedy, H. (2014). 层级解剖学：猕猴视觉皮层中的前馈和反馈通路. *J. Comp. Neurol.*, 522(1): 225 – 259.

Marshall, G. D.; Zimbardo, P.G. (1979). 解释不充分的生理唤起的情感后果. *Journal of Personality and Social Psychology* 37: 970 – 988. doi:10.1037/0022-3514.37.6.970.

Marr, D. (1982). *视觉：计算方法*. 旧金山，加利福尼亚州：弗里曼公司.

Martius, G., Der, R., & Ay, N. (2013). 信息驱动的复杂机器人行为自组织. *PLoS One*, *8*(5), e63400.

Matarić, M. (1990). 使用老鼠大脑导航：机器人空间表示的神经生物学启发模型。收录于 J.-A. Meyer & S. Wilson (编), *会议录：从动物到仿生动物：首届自适应行为仿真国际会议(SAB-90)* (第169-175页)。马萨诸塞州剑桥：MIT出版社。

Matarić, M. (1992). 将表示整合到目标驱动的基于行为的机器人中。*IEEE机器人与自动化学报*, *8*(3), 304-312。

Maturana, H. (1980). 认知生物学。收录于 H. Maturana, R. Humberto, & F. Varela, *自创生与认知* (第2-62页)。多德雷赫特：Reidel出版社。

Maturana, H., & Varela, F. (1980). *自创生与认知：生命的实现*。马萨诸塞州波士顿：Reidel出版社。

Maxwell, J. P., Masters, R. S., & Poolton, J. M. (2006). 体育运动中的表现崩溃：再投入(reinvestment)和言语知识的作用。*运动研究季刊*, *77*(2), 271-276。

McBeath, M., Shaffer, D., & Kaiser, M. (1995). 棒球外野手如何确定跑向何处接住高飞球。*科学*, *268*, 569-573。

McClelland, J. L. (2013). 整合感知的概率模型与交互神经网络：历史与教程综述。*心理学前沿*, *4*, 503。

McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). 感知和认知中的交互激活与相互约束满足。*认知科学*, *6*, 1139-1189。doi:10.1111/cogs。

McClelland, J. L., & Rumelhart, D. (1981). 字母感知中上下文效应的交互激活模型：第1部分。基本发现的解释。*心理学评论*, *88*, 375-407。

McClelland, J. L., Rumelhart, D., & PDP研究小组 (1986). *并行分布式处理* (第二卷)。马萨诸塞州剑桥：MIT出版社。

McGeer, T. (1990). 被动动态行走。*国际机器人研究杂志*, *9*(2), 68-82。

Melloni, L., Schwiedrzik, C. M., Muller, N., Rodriguez, E., & Singer, W. (2011). 期望改变感知意识电生理相关性的特征和时间。*神经科学杂志*, *31*(4), 1386-1396。

Meltzoff, A. N. (2007a). ‘像我一样’：社会认知的基础。*发展科学*, *10*(1), 126-134。

Meltzoff, A. N. (2007b). 识别和成为意向性主体的’像我一样’框架。*心理学学报*, *124*(1), 26-43。

Meltzoff, A. N., & Moore, M. K. (1997). 解释面部模仿：理论模型。*早期发展与育儿*, *6*, 179-192。

Menary, R. (编). (2010). *延展心智*。马萨诸塞州剑桥：MIT出版社。

Merckelbach, H., & van de Ven, V. (2001). 另一个白色圣诞节：幻想倾向与本科生’幻觉体验’报告。*行为治疗与实验精神病学杂志*, *32*, 137-144。

Merleau-Ponty, M. (1945/1962). *感知现象学*。译者：Colin Smith。伦敦：Routledge和Kegan Paul出版社。

Mesulam, M. (1998). 从感觉到认知，*大脑*, *121*(6), 1013-1052。

Metta, G., & Fitzpatrick, P. (2003). 视觉与操作的早期整合。*自适应行为*, *11*(2), 109-128。

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *计划与行为结构*。纽约：霍尔特、莱因哈特与温斯顿公司。

Miller, L. K. (1999). 学者综合征：智力缺陷与异常技能。*心理学公报*, *125*, 31-46。

Millikan, R. G. (1996). 推拉表示。收录于 J. Tomberlin (编), *哲学视角*, 9 (第185-200页)。加利福尼亚州阿塔斯卡德罗：Ridgeview出版社。

Milner, D., & Goodale, M. (1995). *行动中的视觉大脑*。牛津：牛津大学出版社。

Milner, D., & Goodale, M. (2006). *行动中的视觉大脑* (第2版)。牛津：牛津大学出版社。

Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M. A., Morito, Y., Tanabe, H. C., Sadato, N., & Kamitani, Y. (2008). 使用多尺度局部图像解码器组合从人类大脑活动重构视觉图像。*神经元*, 60(5), 915-929。doi:10.1016/j.neuron.2008.11.004。

Mnih, A., & Hinton, G. E. (2007). 统计语言建模的三个新图形模型。*机器学习国际会议*，俄勒冈州科瓦利斯。可获取于：http://www.cs.toronto.edu/~hinton/papers.html#2007。

Mohan, V., & Morasso, P. (2011). 被动运动范式：最优控制的替代方案。*神经机器人学前沿*, *5*, 4。doi:10.3389/fnbot.2011.00004。

Mohan, V., Morasso, P., Sandini, G., & Kasderidis, S. (2013). 认知机器人中通过具身仿真进行推理。*认知计算*, *5*(3), 355-382。

Møller, P., & Husby, R. (2000). 精神分裂症的初期前驱期：寻找体验和行为的自然核心维度。*精神分裂症公报*, *26*, 217-232。

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). 基于预测性赫布学习的中脑多巴胺系统框架。*神经科学杂志*, *16*, 1936-1947。

Montague, P. R., Dolan, R. J., Friston, K., & Dayan, P. (2012). 计算精神病学。*认知科学趋势*, *16*(1), 72-80。doi:10.1016/j.tics.2011.11.018。

Moore, G. E. (1903/1922). 理想主义的反驳。重印于 G. E. Moore, *哲学研究*。伦敦：Routledge & Kegan Paul出版社，1922。

Moore, G. E. (1913/1922). 感觉数据的地位。*亚里士多德学会会议录*，1913。重印于 G. E. Moore, *哲学研究*。伦敦：Routledge & Kegan Paul出版社，1922。

Morrot, G., Brochet, F., & Dubourdieu, D. (2001). 气味的颜色。*大脑与语言*, *79*, 309-320。

Mottron, L., 等 (2006). 自闭症中的增强感知功能：更新与自闭症感知的八个原则。*自闭症与发育障碍杂志*, *36*, 27-43。

Moutoussis, M., Trujillo-Barreto, N., El-Deredy, W., Dolan, R. J., & Friston, K. (2014). 人际推理的形式化模型。*人类神经科学前沿*, *8*, 160. doi:10.3389/fnhum.2014.00160.

Mozer, M. C., & Smolensky, P. (1990). 骨架化：通过相关性评估从网络中去除冗余的技术。载于 D. S. Touretzky & M. Kaufmann (编辑), *神经信息处理进展*, 1 (第177 – 185页)。加利福尼亚州圣马特奥：摩根考夫曼出版社。

Muckli, L. (2010). 我们在这里遗漏了什么？初级视觉皮层v1中高级认知功能的脑成像证据。*国际成像系统技术杂志*, *20*, 131 – 139.

Muckli, L., Kohler, A., Kriegeskorte, N., & Singer, W. (2005). 初级视觉皮层沿表观运动轨迹的活动反映了错觉知觉。*公共科学图书馆生物学*, *13*, e265.

Mumford, D. (1992). 新皮层的计算架构。II. 皮层-皮层环路的作用。*生物控制论*, *66*, 241 – 251.

Mumford, D. (1994). 模式理论问题的神经元架构。载于 C. Koch & J. Davis (编辑), *皮层的大尺度理论* (第125 – 152页)。马萨诸塞州剑桥：麻省理工学院出版社。

Murray, S. O., Boyaci, H., & Kersten, D. (2006). 人类初级视觉皮层中感知角度大小的表征。*自然评论：神经科学*, *9*, 429 – 434.

Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). 形状知觉降低了人类初级视觉皮层的活动。*美国国家科学院院刊*, *99*(23), 15164 – 15169.

Murray, S. O., Schrater, P., & Kersten, D. (2004). 知觉分组和视觉皮层区域间的相互作用。*神经网络*, *17*(5 – 6), 695 – 705.

Musmann, H. (1979). 预测图像编码。载于 W. K. Pratt (编辑), *图像传输技术* 73-112 纽约：学术出版社。

Nagel, T. (1974). 成为蝙蝠是什么感觉？*哲学评论*, *83*, 435 – 456.

Nakahara, K., Hayashi, T., Konishi, S., & Miyashita, Y. (2002). 执行认知集合转换任务的猕猴的功能性磁共振成像。*科学*, *295*, 1532 – 1536.

Namikawa, J., Nishimoto, R., & Tani, J. (2011) 自发行为的神经动力学解释。*公共科学图书馆计算生物学*, *7*(10), e1002221.

Namikawa, J., & Tani, J. (2010). 通过混沌以组合方式学习模仿随机时间序列。*神经网络*, *23*, 625 – 638.

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). 从人类大脑活动中贝叶斯重建自然图像。*神经元*, *63*(6), 902 – 915.

Navalpakkam, V., & Itti, L. (2005). 建模任务对注意力的影响。*视觉研究*, *45*, 205 – 231.

Neal, R. M., & Hinton, G. (1998). EM算法的一个视角，为增量、稀疏和其他变体提供了理论依据。载于 M. I. Jordan (编辑), *图形模型中的学习* (第355 – 368页)。多德雷赫特：克卢维尔出版社。

Neisser, U. (1967). *认知心理学*。纽约：阿普尔顿-世纪-克罗夫茨出版社。

Newell, A., & Simon, H. A. (1972). *人类问题解决*。新泽西州恩格尔伍德克利夫斯：普伦蒂斯-霍尔出版社。

Nirenberg, S., Bomash, I., Pillow, J. W., & Victor, J. D. (2010). 视网膜神经节细胞的异质反应动力学：预测编码和适应的相互作用。*神经生理学杂志*, *103*(6), 3184 – 3194. doi:10.1152/jn.00878.2009.

Nkam, I., Bocca, M. L., Denise, P., Paoletti, X., Dollfus, S., Levillain, D., & Thibaut, F. (2010). 精神分裂症中受损的平滑追踪仅由预测障碍导致。*生物精神病学*, *67*(10), 992 – 997.

Noë, A. (2004). *知觉中的行动*。马萨诸塞州剑桥：麻省理工学院出版社。

Noë, A. (2010). *走出我们的头脑：为什么你不是你的大脑，以及从意识生物学中得出的其他教训*。纽约：法拉，斯特劳斯和吉鲁出版社。

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). 超越读心术：功能性磁共振成像数据的多体素模式分析。*认知科学趋势*, *10*, 424 – 430.

Núñez, R., & Freeman, W. J. (编辑). (1999). *重新认识认知：行动、意图和情感的首要地位*。俄亥俄州鲍林格林：印记学术出版社；纽约：霍顿-米夫林出版社。

O’Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). 注意力调节人类外侧膝状体核的反应。*自然神经科学*, *5*, 1203 – 1209.

O’Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A., & Savoy, R. L. (1997). 自主注意力调节人类MT-MST中的功能性磁共振成像活动。*神经元*, *18*, 591 – 598.

O’Regan, J. K., & Noë, A. (2001). 视觉和视觉意识的感觉运动方法。*行为与脑科学*, *24*(5), 883 – 975.

O’Reilly, R. C., Wyatte, D., & Rohrlich, J. (稿件). 丘脑皮层环路中的时间学习。预印本可在：http://arxiv.org/abs/1407.3432 获取。

O’Sullivan, I., Burdet, E., & Diedrichsen, J. (2009). 将变异性和努力作为协调的决定因素进行区分。*公共科学图书馆计算生物学*, *5*, e1000345. doi:10.1371/journal.pcbi.1000345.

Ogata, T., Yokoya, R., Tani, J., Komatani, K., & Okuno, H. G. (2009). 机器人通过重用自身前向-逆向模型预测和模仿他人动作。*ICRA ’09. IEEE国际机器人与自动化会议*, 4144 – 4149.

Okuda, J., 等 (2003). 思考未来和过去：额极和内侧颞叶的作用。*神经影像*, *19*, 1369 – 1380. doi:10.1016/S1053-8119(03)00179-4.

Olshausen, B. A., & Field, D. J. (1996). 通过学习自然图像的稀疏编码出现简单细胞感受野特性。*自然*, *381*, 607 – 609.

Orlandi, N. (2013). *纯真之眼：为什么视觉不是认知过程*。纽约：牛津大学出版社。

Oudeyer, P.-Y., Kaplan, F., & Hafner, V. (2007). 自主心理发展的内在动机系统。*IEEE进化计算学报*, *11*(2), 265 – 286.

Oyama, S. (1999). *进化之眼：生物学、文化和发展系统*. Durham, NC: Duke University Press.

Oyama, S., Griffiths, P. E., & Gray, R. D. (Eds.). (2001). *偶然性循环：发展系统与进化*. Cambridge, MA: MIT Press.

Oztop, E., Kawato, M., & Arbib, M. A. (2013). 镜像神经元：功能、机制和模型. *神经科学通讯*, *540*, 43 – 55.

Panichello, M., Cheung, O., & Bar, M. (2012). 预测性反馈和有意识的视觉体验. *心理学前沿*, *3*, 620. doi:10.3389/fpsyg.2012.00620.

Pariser, E. (2011). *过滤泡沫：互联网向你隐藏了什么*. New York: Penguin Press.

Park, H. J., & Friston, K. (2013). 结构和功能脑网络：从连接到认知. *科学*, *342*, 579 – 588.

Park, J. C., Lim, J. H., Choi, H., & Kim, D. S. (2012). 发展神经机器人学的预测编码策略. *心理学前沿*., *7*(3), 134. doi:10.3389/fpsyg.2012.00134.

Parr, W. V., White, K. G., & Heatherbell, D. (2003). 鼻子知道：颜色对葡萄酒香气感知的影响. *葡萄酒研究杂志*, *14*, 79 – 101.

Pascual-Leone, A., & Hamilton, R. (2001). 大脑的跨模态组织. *脑研究进展*, *134*, 427 – 445.

Paton, B., Skewes, J., Frith, C., & Hohwy, J. (2013). 颅骨束缚的感知和通过文化的精确度优化. *行为与脑科学*, *36*(3), p. 222.

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., et al. (2000). 文化对大脑功能的影响. *自然神经科学*, *3*(1), 91 – 96.

Pearle, J (1988) *智能系统中的概率推理*, Morgan-Kaufmann

Peelen, M. V. M., & Kastner, S. S. (2011). 人类枕颞皮层真实世界视觉搜索的神经基础. *美国国家科学院院刊. USA*, *108*, 12125 – 12130.

Pellicano, E., & Burr, D. (2012). 当世界变得过于真实：自闭症感知的贝叶斯解释. *认知科学趋势*, *16*, 504 – 510. doi:10.1016/j.tics.2012.08.009.

Penny, W. (2012). 大脑和行为的贝叶斯模型. *ISRN生物数学*, 1 – 19. doi:10.5402/2012/785791.

Perfors, A., Tenenbaum, J. B., & Regier, T. (2006). 刺激贫乏？一种理性方法. 在 *第28届认知科学学会年会论文集* 663-668 Mahwah, NJ: Lawrence Erlbaum Assoc.

Pezzulo, G. (2007). 自然和人工认知中的预期和面向未来的能力. 在 *人工智能50年，纪念文集，LNAI*, Berlin: Springer pp. 258 – 271.

Pezzulo, G. (2008). 与未来协调：表征的预期性质. *心智与机器*, *18*, 179 – 225.

Pezzulo, G. (2012). 认知控制的主动推理观点. *理论与哲学心理学前沿*. 3: 478 – 479. doi: 10.3389/fpsyg.2012.00478.

Pezzulo, G. (2013). 为什么你害怕怪物？感知推理的具身预测编码模型. *认知、情感和行为神经科学*, *14*(3), 902 – 911.

Pezzulo, G., Barsalou, L., Cangelosi, A., Fischer, M., McRae, K., & Spivey, M. (2013). 计算基础认知：基础认知与计算建模之间的新联盟. *心理学前沿*, *3*, 612. doi:10.3389/fpsyg.2012.00612.

Pezzulo, G., Rigoli, F., & Chersi F. (2013). 混合工具控制器：使用信息价值结合习惯性选择和心理模拟. *心理学前沿*, *4*, 92. doi:10.3389/fpsyg.2013.00092.

Pfeifer, R., & Bongard, J. (2006). *身体如何塑造我们的思维方式：智能的新观点*. Cambridge, MA: MIT Press.

Pfeifer, R., Lungarella, M., Sporns, O., & Kuniyoshi, Y. (2007). 论具身的信息论含义：原理和方法. 在 *计算机科学讲义 (LNCS)*, 4850 (pp. 76 – 86). Berlin: Heidelberg: Springer.

Pfeifer, R., & Scheier, C. (1999). *理解智能*. Cambridge, MA: MIT Press.

Philippides, A., Husbands, P., & O’Shea, M. (2000). 一氧化氮的四维神经元信号传导：计算分析. *神经科学杂志*, *20*, 1199 – 1207.

Philippides, A., Husbands, P., Smith, T., & O’Shea, M. (2005). 灵活耦合：扩散神经调节剂和自适应机器人技术. *人工生命*, *11*, 139 – 160.

Phillips, M. L., et al. (2001). 人格解体障碍：无感觉的思考. *精神病学研究*., *108*, 145 – 160.

Phillips W. A. & Singer W. (1997) 寻找皮层计算的共同基础. *行为与脑科学* 20, 657 – 722.

Phillips, W., Clark, A., & Silverstein, S. M. (2015). 论皮层内情境调节的功能、机制和功能障碍. *神经科学与生物行为评论 52*, 1 – 20.

Phillips, W., & Silverstein, S. (2013). 精神生活的连贯组织依赖于情境敏感的增益控制机制，这些机制在精神分裂症中受损. *心理学前沿*, *4*, 307. doi:10.3389/fpsyg.2013.00307.

Piaget, J. (1952). *儿童智力的起源*. New York: I. U. Press, Ed.

Pickering, M. J., & Clark, A. (2014). 超前思考：前向模型及其在认知架构中的位置. *认知科学趋势*, *18*(9), 451 – 456.

Pickering, M. J., & Garrod, S. (2007). 人们在理解过程中是否使用语言产生来进行预测？*认知科学趋势*, *11*(3), 105 – 110.

Pickering, M. J., & Garrod, S. (2013). 语言产生和理解的整合理论. *行为与脑科学*, *36*(04), 329 – 347.

Pinker, S. (1997). *心智如何工作*. London: Allen Lane.

Plaisted, K. (2001). 自闭症中的泛化能力降低：弱中央连贯性的替代理论. 在 J. Burack et al. (Eds.), *自闭症的发展：理论与研究的视角* (pp. 149 – 169). NJ Erlbaum.

Plaisted, K., et al. (1998a). 自闭症患者在结合目标视觉搜索中的增强表现：一项研究报告。*J. Child Psychol. Psychiatry*, *39*, 777 – 783.

Plaisted, K., et al. (1998b). 成年自闭症患者在感知学习任务中对新颖、高度相似刺激的增强辨别能力。*J. Child Psychol. Psychiatry*, *39*, 765 – 775.

Ploner, M., & Tracey, I. (2011). 治疗期望对药物疗效的影响：阿片类药物瑞芬太尼镇痛效果的成像研究。*Sci. Transl. Med.*, *3*, 70ra14.

Poeppel, D., & Monahan, P. J. (2011). 语音感知中的前馈和反馈：重新审视分析综合法。*Language and Cognitive Processes*, *26*(7), 935 – 951.

Polich, J. (2003). P3a和P3b概述。In J. Polich (Ed.), *变化检测：事件相关电位和fMRI发现* (pp. 83 – 98). Boston: Kluwer Academic Press.

Polich, J. (2007). P300更新：P3a和P3b的整合理论。*Clinical Neurophysiology*, *118*(10), 2128 – 2148.

Popper, K. (1963). *猜想和反驳：科学知识的增长*。London: Routledge.

Posner, M. (1980). 注意的定向。*Quarterly Journal of Experimental Psychology*, *32*(1), 3 – 25.

Potter, M. C., Wyble, B., Hagmann, C. E., & Mccourt, E. S. (2013). 在每张图片13毫秒的快速序列视觉呈现中检测意义。*Attention, Perception, & Psychophysics*, 1 – 10.

Pouget, A., Beck, J., Ma, Wei J., & Latham, P. (2013). 概率大脑：已知和未知。*Nature Neuroscience*, *16*(9), 1170 – 1178. doi:10.1038/nn.3495.

Pouget, A., Dayan, P., & Zemel, R. (2003). 群体编码的推理和计算。*Annual Review of Neuroscience*, *26*, 381 – 410.

Powell, K. D., & Goldberg, M. E. (2000). 在记忆引导性眼跳延迟期间闪现干扰物时外侧顶内区神经元的反应。*J. Neurophysiol.*, *84*(1), 301 – 310.

Press, C., Richardson, D., & Bird, G. (2010). 自闭症谱系障碍患者对情绪面部动作的完整模仿。*Neuropsychologia*, *48*, 3291 – 3297.

Press, C. M., Heyes, C. M., & Kilner, J. M (2011). 学习理解他人的行为。*Biology Letters*, *7*(3), 457 – 460. doi:10.1098/rsbl.2010.0850.

Pribram, K. H. (1980). 定向反应：大脑表征机制的关键。In H. D. Kimme (Ed.), *人类定向反射* (pp. 3 – 20). Hillsdale, NJ: Lawrence Erlbaum Assoc.

Price, C. J., & Devlin, J. T. (2011). 腹侧枕颞叶对阅读贡献的交互解释。*Trends in Cognitive Neuroscience*, *15*(6), 246 – 253.

Price, D. D., Finniss, D. G., & Benedetti, F. (2008). 安慰剂效应的全面回顾：最新进展和当前思考。*Annu. Rev. Psychol.*, *59*, 565 – 590.

Prinz, J. (2004). *直觉反应*。New York: Oxford University Press.

Prinz, J. (2005). 意识的神经功能理论。In A. Brook & K. Akins (Eds.), *认知与大脑：哲学与神经科学运动* (pp. 381 – 396). Cambridge: Cambridge University Press.

Purves, D., Shimpi, A., & Lotto, R. B. (1999). 康斯威特效应的经验解释。*Journal of Neuroscience*, *19*(19), 8542 – 8551.

Pylyshyn, Z. (1999). 视觉与认知是连续的吗？视觉感知认知不可渗透性的论证。*Behavioral and Brain Sciences*, *22*(3), 341 – 365.

Quine, W. V. O. (1988). 论存在的是什么。In *从逻辑观点看：九篇逻辑哲学论文* (pp. 1 – 20). Cambridge, MA: Harvard University Press.

Raichle M. E. (2009). 人类大脑成像简史。*Trends Neurosci.* 32:118 – 126.

Raichle, M. E. (2010). 大脑功能的两种观点。*Trends in Cognitive Sciences*, *14*(4), 180 – 190.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). 大脑功能的默认模式。*Proc. Natl. Acad. Sci. USA*, *98*, 676 – 682.

Raichle, M. E., & Snyder, A. Z. (2007). 大脑功能的默认模式：一个不断发展观念的简史。*Neuroimage*, *37*, 1083 – 1090.

Raij, T., McEvoy, L., Mäkelä, J. P., & Hari, R. (1997). 人类听觉皮层被听觉刺激的遗漏所激活。*Brain Res.*, *745*, 134 – 143.

Ramachandran, V. S., & Blakeslee, S. (1998). *大脑中的幻影：探索人类心智的奥秘*。New York: Morrow & Co.

Ramsey, W. M. (2007). *表征的重新思考*。Cambridge: Cambridge University.

Rao, R., & Ballard, D. (1999). 视觉皮层中的预测编码：对一些经典感受野外效应的功能解释。*Nature Neuroscience*, *2*(1), 79.

Rao, R., & Sejnowski, T. (2002). 预测编码、皮层反馈和尖峰时间依赖的皮层可塑性。In Rao, R. P. N., Olshausen, B. and Lewicki, M. (Eds.), *大脑的概率模型* (pp. 297 – 316). Cambridge, MA: MIT Press.

Rao, R., Shon, A., & Meltzoff, A. (2007). 婴儿和机器人模仿的贝叶斯模型。In C. L. Nehaniv & K. Dautenhahn (Eds.), *机器人、人类和动物的模仿与社会学习：行为、社会和交流维度* (pp. 217 – 247). New York: Cambridge University Press.

Reddy, L., Tsuchiya, N., & Serre, T. (2010). 解读心灵之眼：在心理想象过程中解码类别信息。*NeuroImage*, *50*(2), 818 – 825.

Reich, L., Szwed, M., Cohen, L., & Amedi, A. (2011). 独立于视觉经验的腹侧流阅读中心。*Current Biology*, *21*, 363 – 368.

Remez, R. E., & Rubin, P. E. (1984). 从正弦波句子感知语调。*Perception & Psychophysics*, *35*, 429 – 440.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). 没有传统语音线索的语音感知。*Science*, *212*, 947 – 949.

Rescorla, M. (2013). 贝叶斯感知心理学。In M. Matthen (Ed.), *牛津感知哲学手册*。Oxford University Press, NY.

Rescorla, M. (印刷中). 贝叶斯感觉运动心理学. *Mind and Language.*

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). 竞争机制支持猕猴V2和V4区域的注意. *Journal of Neuroscience*, *19*, 1736 – 1753.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1997). *Spikes: 探索神经编码*. Cambridge, MA: MIT Press.

Riesenhuber, M., & Poggio, T. (1999). 皮层物体识别的层次模型. *Nat. Neurosci.*, *2*, 1019 – 1025.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1996). 前运动皮层和运动动作识别. *Brain Res. Cogn. Brain Res.*, *3*, 131 – 141.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). 理解和模仿动作的神经生理机制. *Nat. Rev. Neurosci.*, *2*, 661 – 670.

Rizzolatti, G., & Sinigaglia, C. (2007). 镜像神经元和运动意向性. *Functional Neurology*, *22*(4), 205 – 210.

Rizzolatti, G., et al. (1988). 猕猴下区域6的功能组织：II. F5区和远端运动控制. *Exp. Brain Res.*, *71*, 491 – 507.

Robbins, H. (1956). 统计学的经验贝叶斯方法. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1: *Contributions to the Theory of Statistics*, 157 – 163.

Robbins, J. M., & Kirmayer, L. J. (1991). 躯体化的认知和社会因素. 在 L. J. Kirmayer & J. M. Robbins (编), *Current concepts of somatisation: Research and clinical perspectives* (pp. 107 – 141). Washington, DC: American Psychiatric Press.

Roberts, A., Robbins, T., & Weiskrantz, L. (1998). *前额皮层：执行和认知功能*. New York: Oxford University Press.

Rock, I. (1997). *间接知觉*. Cambridge, MA: MIT Press.

Roepstorff, A. (2013). 互动的人类：共享时间，构建物质性. *Behavioral and Brain Sciences*, *36*, 224 – 225. doi:10.1017/S0140525X12002427.

Roepstorff, A., & Frith, C. (2004). 动作自上而下控制的顶端是什么？认知实验中动作的脚本共享和”顶端的顶端”控制. *Psychological Research*, *68*(2 – 3), 189 – 198. doi:10.1007/s00426003-0155-4.

Roepstorff, A, Niewöhner, J, & Beck, S. (2010). 通过模式化实践培养大脑. *Neural Networks*, *23*, 1051 – 1059.

Romo, R., Hernandez, A., & Zainos, A. (2004). 腹侧前运动皮层知觉决策的神经元相关性. *Neuron*, *41*(1), 165 – 173.

Rorty, R. (1979). *哲学与自然之镜*. Princeton, NJ: Princeton University Press.

Ross, D. (2004). 社会主体间协调的元语言信号. *Language Sciences*, *26*, 621 – 642.

Roth, M., Synofzik, M., & Lindner, A. (2013). 小脑优化对外部感觉事件的知觉预测. *Current Biology*, *23*(10), 930 – 935. doi:10.1016/j.cub.2013.04.027.

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). 任务和情境决定你看向何处. *Journal of Vision*, *7*(14):16, 1 – 20.

Rumelhart, D., McClelland, J., & the PDP Research Group (1986). *并行分布式处理* (Vol. 1). Cambridge, MA: MIT Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986a). 通过误差传播学习内部表征. 在 D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (编), *Parallel distributed processing: Explorations in the microstructure of cognition*, Vol. 1: *Foundations* (pp. 318 – 362). Cambridge, MA: MIT Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986b). 通过反向传播误差学习表征. *Nature*, *323*, 533 – 536.

Rupert, R. (2004). 对扩展认知假设的挑战. *Journal of Philosophy*, *101*(8), 389 – 428.

Rupert, R. (2009). *认知系统与扩展心智*. New York: Oxford University Press.

Sachs, E. (1967). 大鼠学习的分离及其与人类分离状态的相似性. 在 J. Zubin & H. Hunt (编), *Comparative psychopathology: Animal and human* (pp. 249 – 304). New York: Grune and Stratton.

Sadaghiani, S., Hesselmann, G., Friston, K. J., & Kleinschmidt, A. (2010). 持续大脑活动、诱发神经反应和认知的关系. *Frontiers in Systems Neuroscience*, *4*, 20.

Saegusa, R., Sakka, S., Metta, G., & Sandini, G. (2008). 感觉预测学习: 如何建模自我和环境. *12th IMEKO TC1-TC7 Joint Symposium on Man Science and Measurement (IMEKO2008)*, Annecy, France, 269 – 275.

Salakhutdinov, R., & Hinton, G. (2009). 深度玻尔兹曼机. *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 5, 448 – 455.

Salakhutdinov, R. R., Mnih, A., & Hinton, G. E. (2007). 用于协同过滤的受限玻尔兹曼机. *International Conference on Machine Learning*, Corvallis, Oregon. 可访问: http://www.cs.toronto.edu/~hinton/papers.html#2007.

Salge, C., Glackin, C., & Polani, D. (2014). 基于赋权作为内在动机改变环境. *Entropy*, *16*(5), 2789 – 2819. doi:10.3390/e16052789.

Satz, D., & Ferejohn, J. (1994). 理性选择和社会理论来源. *Journal of Philosophy*, *91*(2), 71 – 87.

Sawtell, N. B., Williams, A., & Bell, C. C. (2005). 从火花到尖峰: 鱼类电感觉系统中的信息处理. *Current Opinion in Neurobiology*, *15*, 437 – 443.

Schachter, S., & Singer, J. (1962). 情绪状态的认知、社会和生理决定因素. *Psychological Review*, *69*, 379 – 399.

Schacter, D. L., & Addis, D. R. (2007a). 过去与未来的幽灵. *Nature*, *445*, 27. doi:10.1038/445027a.

Schacter, D. L., & Addis, D. R. (2007b). 建构性记忆的认知神经科学: 回忆过去与想象未来。*伦敦皇家学会哲学汇刊, B 系列*, *362*, 773 – 786。

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). 前瞻性大脑: 回忆过去以想象未来。*自然神经科学评论*, *8*, 657 – 661。

Schenk, L. A., Sprenger, C., Geuter, S., & Büchel, C. (2014). 期望需要治疗来增强疼痛缓解: 一项fMRI研究。*疼痛*, *155*, 150 – 157。

Schenk, T., & McIntosh, R. D. (2010). 我们是否有独立的感知和行动视觉流？*认知神经科学*，*1*(1)，52 – 62。doi:10.1080/17588920903388950。ISSN 1758-8928。

Scholl, B. (2005). 天赋性与（贝叶斯）视觉感知：调和先天论与发展。见P. Carruthers, S. Laurence & S. Stich (编)，*天赋心智：结构与内容*（第34 – 52页）。纽约：牛津大学出版社。

Schultz, W. (1999). 中脑多巴胺神经元的奖励信号。*生理学*，*14*(6)，249 – 255。

Schultz, W., Dayan, P., & Montague, P. R. (1997). 预测和奖励的神经基础。*科学*，*275*，1593 – 1599。

Schütz, A. C., Braun, D. I., & Gegenfurtner, K. (2011). 眼动与感知：选择性综述。*视觉杂志*，*11*(5)，1 – 30。doi:10.1167/11.5.9。

Schwartenbeck, P., Fitzgerald, T., Dolan, R. J., & Friston, K. (2013). 探索、新奇性、惊讶与自由能最小化。*心理学前沿*，*4*，710。doi:10.3389/fpsyg.2013.00710。

Schwartenbeck, P., FitzGerald, T. H. B., Mathys, C., Dolan, R., & Friston, K. (2014). 多巴胺中脑编码对期望结果的预期确定性。*大脑皮层*，bhu159。doi: 10.1093/cercor/bhu159

Schwartz, J., Grimault, N., Hupé, J., Moore, B., & Pressnitzer, D. (2012). 引言：感知中的多重稳定性：绑定感觉模态，概述。*皇家学会哲学汇刊B*，*367*，896 – 905。doi:10.1098/rstb.2011.0254。

Seamans, J. K., & Yang, C. R. (2004). 前额皮层多巴胺调节的主要特征和机制。*神经生物学进展*，*74*(1)，1 – 58。

Sebanz, N., & Knoblich, G. (2009). 联合行动中的预测：什么、何时和在哪里。*认知科学专题*，*1*，353 – 367。

Sedley, W., & Cunningham, M. (2013). 皮层伽马振荡是促进还是抑制感知？一个被问得太少但答案被过度假设的问题。*人类神经科学前沿*，*7*，595。

Selen, L. P. J., Shadlen, M. N., & Wolpert, D. M. (2012). 运动系统中的思考：反射增益追踪导向决策的演化证据。*神经科学杂志*，*32*(7)，2276 – 2286。

Seligman, M., Railton, P., Baumeister, R., & Sripada, C. (2013). 导航进入未来还是被过去驱动。*心理科学视角*，*8*，119 – 141。

Serre, T., Oliva, A., & Poggio, T. (2007). 前馈架构解释快速分类。*美国国家科学院院刊*，*104*(15)，6424 – 6429。

Seth, A. K. (2013). 内感受推理、情绪与具身自我。*认知科学趋势*，*17*(11)，565 – 573。doi:10.1016/j.tics.2013.09.007。

Seth, A. K. (2014). 感觉运动偶然性的预测处理理论：解释感知存在之谜及其在联觉中的缺失。*认知神经科学*，*5*(2)，97 – 118。doi:10.1080/17588928.2013.877880。

Seth, A. K., Suzuki, K., & Critchley, H. D. (2011). 意识存在的内感受预测编码模型。*心理学前沿*，*2*，395。doi:10.3389/fpsyg.2011.00395。

Seymour, B., et al. (2004). 时间差分模型描述人类的高阶学习。*自然*，*429*，664 – 667。

Sforza, A., et al. (2010). 我的脸在你的脸中：视触觉面部刺激影响身份感。*社会神经科学*，*5*，148 – 162。

Shaffer, D. M., Krauchunas, S. M., Eddy, M., & McBeath, M. K. (2004). 狗如何导航接住飞盘。*心理科学*，*15*，437 – 441。

Shafir, E., & Tversky, A. (1995). 决策制定。见E. E. Smith & D. N. Osherson (编)，*思维：认知科学邀请*（第77 – 100页）。马萨诸塞州剑桥：MIT出版社。

Shah, A., & Frith, U. (1983). 自闭症儿童的能力孤岛：研究报告。*儿童心理学与精神病学杂志*，*24*，613 – 620。

Shams, L., Ma, W. J., & Beierholm, U. (2005). 声诱发闪光错觉作为最优感知。*神经报告*，*16*(10)，1107 – 1110。

Shani, I. (2006). 意向性方向性。*控制论与人类认知*，*13*，87 – 110。

Shankar, M. U., Levitan, C., & Spence, C. (2010). 葡萄期望：认知影响在颜色-味觉交互中的作用。*意识与认知*，*19*，380 – 390。

Shea, N. (2013). 感知对比行动：计算可能相同但适应方向不同。*行为与脑科学*，*36*(3)，228 – 229。

Shergill, S., Bays, P. M., Frith, C. D., & Wolpert, D. M. (2003). 以眼还眼：力量升级的神经科学。*科学*，*301*(5630)，187。

Shergill, S., Samson, G., Bays, P. M., Frith, C. D., & Wolpert, D. M. (2005). 精神分裂症感觉预测缺陷的证据。*美国精神病学杂志*，*162*(12)，2384 – 2386。

Sherman, S. M., & Guillery, R. W. (1998). 一个神经细胞对另一个神经细胞可能产生的作用：区分”驱动者”与”调节者”。*美国国家科学院院刊*，*95*，7121 – 7126。

Shi, Yun Q., & Sun, H. (1999). *多媒体工程的图像和视频压缩：基础、算法和标准*。博卡拉顿CRC出版社。

Shipp, S. (2005). 无颗粒的重要性：视觉和运动皮层的比较论述。*皇家学会哲学汇刊B*，360，797 – 814。doi:10.1098/rstb.2005.1630。

Shipp, S., Adams, R. A., & Friston, K. J. (2013). 对无颗粒结构的反思：运动皮层中的预测编码。*神经科学趋势*，*36*，706 – 716。doi:10.1016/j.tins.2013.09.004。

Siegel, J. (2003). 我们为什么睡觉。*科学美国人*，11月，第92 – 97页。

Siegel, S. (2006). 直接实在论和知觉意识。*哲学与现象学研究*，*73*(2)，379 – 409。

Siegel, S. (2012). 认知渗透性和知觉辩护。*理性*，*46*(2)，201 – 222。

Sierra, M., & David, A. S. (2011). 人格解体：自我意识的选择性损害。*意识与认知*，*20*，99 – 108。

Simon, H. A. (1956). 理性选择和环境结构。*心理学评论*，*63*(2)，129 – 138。

Skewes, J. C., Jegindø, E.-M., & Gebauer, L. (2014). 知觉推理和自闭症特质。*自闭症* [提前在线发表]。10.1177/1362361313519872。

Sloman, A. (2013). 大脑还能做什么？*行为与脑科学*，*36*，230 – 231。doi:10.1017/S0140525X12002439。

Smith, A. D. (2002). *知觉问题*。马萨诸塞州剑桥：哈佛大学出版社。

Smith, F., & Muckli, L. (2010). 未受刺激的早期视觉区域携带周围环境信息。*美国国家科学院院刊*，107(46)，20099－20103。

Smith, L., & Gasser, M. (2005). 具身认知的发展：来自婴儿的六个教训。*人工生命*，*11*(1)，13－30。

Smith, P. L., & Ratcliff, R. (2004). 简单决策的心理学和神经生物学。*神经科学趋势*，*27*，161－168。

Smith, S.M., Vidaurre, D., Beckmann, C.F, Glasser, M.F., Jenkinson, M., Miller, K.L., Nichols, T., Robinson, E., Salimi-Khorshidi, G., Woolrich M.W., Ugurbil, K. & Van Essen D.C. (2013). 基于静息态功能磁共振成像的功能连接组学。*认知科学趋势*，17(12)，666-682。

Smolensky, P. (1988). 论连接主义的正确处理。*行为与脑科学*，*11*，1－23。

Smolensky, P., & Legendre, G. (2006). *和谐心智：从神经计算到最优性理论语法*，第1卷：*认知架构*，第2卷：*语言学和哲学含义*。马萨诸塞州剑桥：MIT出版社。

Sokolov, E. N. (1960). 神经元模型和定向反射。在M. A. B. Brazier (主编)，*中枢神经系统与行为* (第187－276页)。纽约：约西亚·梅西基金会。

Solway, A., & Botvinick, M. M. (2012). 目标导向决策作为概率推理：计算框架和潜在神经相关性。*心理学评论*，119(1)，120－154。doi: 10.1037/a0026435。

Sommer, M. A., & Wurtz, R. H. (2006). 丘脑对额叶皮层空间视觉处理的影响。*自然*，*444*，374－377。

Sommer, M. A., & Wurtz, R. H. (2008). 运动内部监控的大脑回路。*神经科学年鉴*，*31*，317－338。

Spelke, E. S. (1990). 物体知觉原理。*认知科学*，*14*，29－56。

Spence, C., & Shankar, M. U. (2010). 听觉线索对食物和饮料知觉及反应的影响。*感官研究杂志*，*25*，406－430。

Sperry, R. (1950). 视觉倒置产生的自发性视动性反应的神经基础。*比较生理心理学杂志*，*43*(6)，482－489。

Spivey, M. J. (2007). *心智的连续性*。纽约：牛津大学出版社。

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). 对语音竞争者的持续吸引。*美国国家科学院院刊*，*102*，10393－10398。

Spivey, M. J., Richardson, D., & Dale, R. (2008). 语言和认知中的眼手运动。在E. Morsella & J. Bargh (主编)，*行动心理学* (第225－249页)。纽约：牛津大学出版社。

Spivey, M., Richardson, D., & Fitneva, S. (2005). 思维超越大脑：语言和视觉信息的空间索引。在J. Henderson & F. Ferreira (主编)，*视觉语言与行动的接口* (第161－190页)。纽约：心理学出版社。

Sporns, O. (2002). 网络分析、复杂性和大脑功能。*复杂性*，*8*，56－60。

Sporns, O. (2007). 神经机器人模型能够教给我们关于神经和认知发展的什么。在D. Mareschal, S. Sirois, G. Westermann, & M. H. Johnson (主编)，*神经建构主义：观点与前景*，第2卷 (第179－204页)。牛津：牛津大学出版社。

Sporns, O. (2010). *大脑网络*。马萨诸塞州剑桥：MIT出版社。

Sprague, N., Ballard, D. H., & Robinson, A. (2007). 建模具身视觉行为。*ACM应用知觉期刊*，*4*，11。

Spratling, M. (2008). 预测编码作为视觉注意中偏向竞争的模型。*视觉研究*，*48*(12)，1391 – 1408。

Spratling, M. (2010). 预测编码作为皮层V1区反应特性的模型。*神经科学杂志*，*30*(9)，3531 – 3543。

Spratling, M. (2012). 在皮层功能预测编码模型中编码基本图像成分的生成权重和判别权重的无监督学习。*神经计算*，*24*(1)，60 – 103。

Spratling, M. (2013). 在大脑功能的预测编码解释中区分理论与实现 [对Clark的评论]。*行为与脑科学*，*36*(3)，231 – 232。

Spratling, M. (2014). 皮层中驱动器和调节器的单一功能模型。*计算神经科学杂志*，*36*(1)，97 – 118。

Spratling, M., & Johnson, M. (2006). 知觉学习和分类的反馈模型。*视觉认知*，13(2)，129 – 165。

Sprevak, M. (2010). 计算、个体化和表征的既有观点。*科学史与科学哲学研究*，*41*，260 – 270。

Sprevak, M. (2013). 关于神经表征的虚构主义。*一元论者*，*96*，539 – 560。

Stafford, T., & Webb, M. (2005). *心智黑客*。加利福尼亚：O’Reilly媒体出版社。

Stanovich, K. E. (2011). *理性与反思性心智*. 纽约: 牛津大学出版社.

Stanovich, K. E., & West, R. F. (2000). 推理中的个体差异：对理性辩论的启示. *行为与脑科学*, *23*, 645 – 665.

Stein, J. F. (1992). 后顶叶皮层中自我中心空间的表征. *行为脑科学*, *15*, 691 – 700.

Stephan, K. E., Harrison, L. M., Kiebel, S. J., David, O., Penny, W. D., & Friston, K. J. (2007). 神经系统动态的动态因果模型：当前状态与未来扩展. *生物科学杂志*, *32*, 129 – 144.

Stephan, K. E., Kasper, L., Harrison, L. M., Daunizeau, J., den Ouden, H. E., Breakspear, M., & Friston, K. J. (2008). fMRI的非线性动态因果模型. *神经影像*, *42*, 649 – 662.

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., and Friston, K. J. (2009). 群体研究的贝叶斯模型选择. 神经影像 46, 1004 – 1017. doi: 10.1016/j.neuroimage.2009.03.025.

Sterelny, K. (2003). *敌对世界中的思考：人类认知的进化*. 牛津: 布莱克威尔出版社.

Stewart, J., Gapenne, O., & Di Paolo, E. A. (主编). (2010). *行动：走向认知科学的新范式*. 剑桥, MA: MIT出版社.

Stigler, J., Chalip, L., & Miller, F. (1986). 技能的后果：台湾算盘训练案例. *美国教育杂志*, 94(4), 447 – 479.

Stokes, M., Thompson, R., Cusack, R., & Duncan, J. (2009). 心理想象过程中视觉皮层形状特异性群体编码的自上而下激活. *神经科学杂志*, *29*, 1565 – 1572.

Stone, J., Warlow, C., Carson, A., & Sharpe, M. (2005). 埃利奥特·斯莱特关于癔症不存在的神话. *皇家医学会杂志*, *98*, 547 – 548.

Stone, J., Warlow, C., & Sharpe, M. (2012a). 功能性无力：从发病性质看机制的线索. *神经病学、神经外科学与精神病学杂志*, *83*, 67 – 69.

Stone, J., Wojcik, W., Durrance, D., Carson, A., Lewis, S., MacKenzie, L., et al. (2002). 我们应该对有疾病无法解释症状的患者说什么？“需要冒犯的数量”. *英国医学杂志*, *325*, 1449 – 1450.

Stotz, K. (2010). 人性与认知-发展生态位构建. *现象学与认知科学*, *9*(4), 483 – 501.

Suddendorf, T., & Corballis, M. C. (1997). 心理时间旅行与人类心智的进化. *遗传学、社会学与普通心理学专著* 123, 133 – 167.

Suddendorf, T., & Corballis, M. C. (2007). 预见能力的进化：什么是心理时间旅行，它是人类独有的吗？ *行为脑科学*, *30*, 299 – 331.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). 神经重复抑制反映已实现的知觉期望. *自然神经科学*, *11*(9), 1004 – 1006.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). 刺激不确定性的诱发电位相关性. *科学*, *150*, 1187 – 1188.

Suzuki, K., et al. (2013). 内感受和外感受域间的多感觉整合调节橡胶手错觉中的自我体验. *神经心理学*.http://dx.doi.org/10.1016/j.neuropsychologia. 2013.08.014 20.

Szpunar, K. K. (2010). 情景未来思维：一个新兴概念. *心理科学展望*, *5*, 142 – 162. doi:10.1177/1745691610362350.

Szpunar, K. K., Watson, J. M., & McDermott, K. B. (2007). 设想未来的神经基础. *美国国家科学院院刊*, *104*, 642 – 647. doi:10.1073/pnas. 0610082104.

Tani, J. (2007). 关于自上而下预期与自下而上回归之间的相互作用. *神经机器人学前沿*, *1*, 2. doi:10.3389/neuro.12.002.2007.

Tani, J., Ito, M., & Sugita, Y. (2004). 镜像系统中分布式表征的多重行为图式的自组织：使用RNNPB的机器人实验综述. *神经网络*, *17*(8), 1273 – 1289.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). 自然视觉中的眼动引导：重新解释显著性. *视觉杂志*, *11*(5):5, 1 – 23.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). 如何培养心智：统计、结构与抽象. *科学*, *331*(6022), 1279 – 1285.

Teufel, C, Fletcher, P, & Davis, G (2010). 看见他人的心智：归因的心理状态影响知觉. *认知科学趋势*, *14*(2010), 376 – 382.

Thaker, G. K., Avila, M. T., Hong, L. E., Medoff, D. R., Ross, D. E., Adami, H. M. (2003). 与精神分裂症表型相关的平滑追踪眼动缺陷模型. *心理生理学*, *40*, 277 – 284. PubMed: 12820868.

Thaker, G. K., Ross, D. E., Buchanan, R. W., Adami, H. M., & Medoff, D. R. (1999). 对视网膜外运动信号的平滑追踪眼动：精神分裂症患者的缺陷. *精神病学研究*, *88*, 209 – 219. PubMed: 10622341.

Thelen, E., Schöner, G., Scheier, C., & Smith, L. (2001). 具身化的动力学：婴儿持续性伸手的场论. *行为与脑科学*, *24*, 1 – 33.

Thelen, E., & Smith, L. (1994). *认知与行动发展的动态系统方法*. 剑桥, MA: MIT出版社.

Thevarajah, D., Mikulic, A., & Dorris, M. C. (2009). 上丘在选择混合策略扫视中的作用. *神经科学杂志*, *29*(7), 1998 – 2008.

Thompson, E. (2010). *生命中的心智：生物学、现象学与心智科学*. 剑桥, MA: 哈佛大学出版社.

Thornton, C. (手稿). 稀疏编码预测处理的实验.

Todorov, E. (2004). 感觉运动控制中的最优性原理. *自然神经科学*, *7*, 907 – 915.

Todorov, E. (2008). 感觉与运动信息处理的相似性. 载于 M. Gazzaniga (主编), *认知神经科学* (第4版) (第613 – 624页). 剑桥, MA: MIT出版社.

Todorov, E. (2009). 最优行动的高效计算. *Proc. Natl. Acad. Sci. USA*, 106, 11478 – 11483.

Todorov, E., & Jordan, M. I. (2002). 最优反馈控制作为运动协调理论. *Nature Neuroscience*, *5*, 1226 – 1235.

Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). 先验预期介导听觉皮层对重复声音的神经适应：一项MEG研究. *J. Neurosci.*, *31*, 9118 – 9123.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). 理解和分享意图：文化认知的个体发生和系统发生. *Behavioral & Brain Sciences*, *28*(5), 675 – 691.

Tononi, G., & Cirelli, C. (2006). 睡眠功能和突触稳态. *Sleep Medicine Reviews*, *10*(1), 49 – 62.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). 现实世界场景中眼动和注意的情境引导：全局特征在物体搜索中的作用. *Psychological Review*, *113*, 766 – 786.

Toussaint, M. (2009). 概率推理作为计划行为模型. *Künstliche Intelligenz*, *3*(9), 23 – 29.

Tracey, I. (2010). 获得你所期望的疼痛：人类安慰剂、反安慰剂和重新评价效应的机制. *Nat. Med.*, *16*, 1277 – 1283.

Treue, S. (2001). 灵长类视觉皮层注意的神经相关性. *Trends Neurosci.*, *24*, 295 – 300. doi:10.1016/S0166-2236(00)01814-2.

Tribus, M. (1961). *热力学和热统计学：能量、信息和物质状态的介绍，及其工程应用*. New York: D. Van Nostrand.

Tsuchiya, N., & Koch, C. (2005). 连续闪烁抑制减少负后像. *Nature Neuroscience*, *8*(8), 1096 – 1101. doi:10.1038/nn1500.

Tsuda, I. (2001). 混沌大脑理论的合理性. *Behavioral and Brain Sciences*, *24*(5), 829 – 840.

Tulving, E. (1983). *情节记忆的要素*. New York: Oxford University Press.

Tulving, E., & Gazzaniga, M. S. (1995). 记忆的组织：走向何方？ In M. Gazzaniga (Ed.), *认知神经科学* (pp. 839 – 847). Cambridge, MA: MIT Press.

Turvey, M., & Carello, C. (1986). 知觉-行动的生态学方法：一篇图示文章. *Acta Psychologica*, *63*, 133 – 155.

Turvey, M. T., & Shaw, R. E. (1999). 认知的生态学基础：I. 动物-环境系统的对称性和特异性. *Journal of Consciousness Studies*, *6*, 85 – 110.

Tversky, A., & Kahneman, D. (1973). 可得性：判断频率和概率的启发式. *Cognitive Psychology*, *5*, 207 – 232.

Ungerleider, L. G., & Mishkin, M. (1982). 两个皮层视觉系统. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *视觉行为分析* (pp. 549 – 586). Cambridge, MA: MIT Press.

Uno, Y., Kawato, M., & Suzuki, R. (1989). 人类多关节臂运动中最优轨迹的形成和控制. *Biological Cybernetics*, *61*, 89 – 101.

Van de Cruys, S., de-Wit, L., Evers, K., Boets, B., & Wagemans, J. (2013). 自闭症中的弱先验与预测过拟合：对Pellicano和Burr (TICS, 2012)的回复. *I-Perception*, 4(2), 95 – 97. doi:10.1068/i0580ic.

Van de Cruys, S., & Wagemans, J. (2011). 在艺术中加入奖励：视觉艺术的预测误差假设性解释. *i-Perception*, special issue on Art & Perception, *2*(9), 1035 – 1062.

Van den Berg, R., Keshvari, S., & Ma, W. J. (2012). 编码精度变异下知觉中的概率计算. *PLoS One*, *7*(6), e40216.

Van den Heuvel, M. P., & Sporns, O. (2011). 人类连接组的富人俱乐部组织. *J. Neurosci.*, *31*, 15775 – 15786.

Van Essen, D. C., Anderson, C. H., & Olshausen, B. A. (1994). 感觉、运动和认知处理中的动态路由策略. In C. Koch & J. Davis (Eds.), *大脑的大规模神经理论* (pp. 271 – 299). Cambridge, MA: MIT Press.

van Gerven, M. A. J., de Lange, F. P., & Heskes, T. (2010). 使用层次生成模型的神经解码. *Neural Comput.*, *22*(12), 3127 – 3142.

van Kesteren, M. T., Ruiter, D. J., Fernandez, G., & Henson, R. N. (2012). 图式和新颖性如何增强记忆形成. *Trends in Neurosciences*, *35*, 211 – 219. doi:10.1016/j.tins.2012.02.001.

van Leeuwen, C. (2008). 混沌孕育自主性：偏见和过度照料之间的连接主义设计. *Cognitive Processing*, *9*, 83 – 92.

Varela, F., Thompson, E., & Rosch, E. (1991). *具身心智*. Cambridge, MA: MIT Press.

Vilares, I., & Körding, K. (2011). 贝叶斯模型：世界的结构、不确定性、行为和大脑. *Annals of the New York Acad. Sci.*, *1224*, 22 – 39.

Von Holst, E. (1954). 中枢神经系统与外周器官之间的关系. *British Journal of Animal Behaviour*, *2*(3), 89 – 94.

Voss, M., Ingram, J. N., Wolpert, D. M., & Haggard, P. (2008). 仅仅运动预期就会导致感觉信号衰减. *PLoS One*, *3*(8), e2866.

Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). 解释失配负波的预测编码神经模型. *Journal of Neuroscience*, *32*(11), 3665 – 3678.

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). 人类皮层中预测和预测误差层次的证据. *Proc. Natl. Acad. Sci. USA*, *108*, 20754 – 20759.

Warren, W. (2006). 行动和知觉的动力学. *Psychological Review113*(2), 358 – 389.

Webb, B. (2004). 预测的神经机制：昆虫有前向模型吗？ *Trends in Neurosciences*, *27*(5), 1 – 11.

Weber, C., Wermter, S., & Elshaw, M. (2006). 运动皮层的混合生成和预测模型. *Neural Networks*, *19*, 339 – 353.

Weiskrantz, L., Elliot, J., & Darlington, C. (1971). 自我挠痒的初步观察. *Nature*, *230*(5296), 598 – 599.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). 运动错觉作为最优感知。*自然神经科学*, *5*(6), 598 – 604.

Wheeler, M. (2005). *重构认知世界*。剑桥，马萨诸塞州：MIT 出版社。

Wheeler, M., & Clark, A. (2008). 文化、具身性和基因：解开三重螺旋。*皇家学会哲学会刊 B*, *363*(1509), 3563 – 3575.

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). 记忆的回声：生动的回忆重新激活感觉特异性皮层。*美国科学院院刊*, *97*, 11125 – 11129.

Wiese, W. (2014). Jakob Hohwy《预测心智》书评。*心智与机器*, *24*(2), 233 – 237.

Wilson, R. A. (2004). *心智的边界：脆弱科学中的个体——认知*。剑桥：剑桥大学出版社。

Wolpert, D. M. (1997). 运动控制的计算方法。*认知科学趋势*, *1*, 209 – 216.

Wolpert, D. M., Doya, K., & Kawato, M. (2003). 运动控制和社会互动的统一计算框架。*皇家学会哲学会刊*, *358*, 593 – 602.

Wolpert, D. M., & Flanagan, J. R. (2001). 运动预测。*当代生物学*, *18*, R729 – R732.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). 感觉运动整合的内部模型。*科学*, 269, 1880 – 1882.

Wolpert, D. M., & Kawato, M. (1998). 运动控制的多重配对前向和逆向模型。*神经网络*, *11*(7 – 8), 1317 – 1329.

Wolpert, D. M., Miall, C. M., & Kawato, M. (1998). 小脑中的内部模型。*认知科学趋势*, *2*(9), 338 – 347.

Wurtz, R. H., McAlonan, K., Cavanaugh, J., & Berman, R. A. (2011). 主动视觉的丘脑通路。*认知科学趋势*, *15*, 177 – 184.

Yabe, H., Tervaniemi, M., Reinikainen, K., & Naatanen, R. (1997). 声音省略的MMN揭示的整合时间窗口。*神经报告*, *8*, 1971 – 1974.

Yamashita, Y., & Tani, J. (2008). 多时间尺度神经网络模型中功能层次的涌现：类人机器人实验。*PLoS 计算生物学*, 4(11), e1000220.

Yu, A. J. (2007). 适应性行为：人类表现为贝叶斯学习者。*当代生物学*, *17*, R977 – R980.

Yu, A. J., & Dayan, P. (2005). 不确定性、神经调节和注意力。*神经元*, *46*, 681 – 692.

Yuille, A, & Kersten, D (2006). 视觉作为贝叶斯推理：综合分析？*认知科学趋势*, 10(7), 301 – 308.

Yuval-Greenberg, S., & Heeger, D. J. (2013). 连续闪光抑制调节早期视觉皮层的皮层活动。*神经科学杂志*, *33*(23), 9635 – 9643.

Zhu, Q., & Bingham, G. P. (2011). 人类投掷准备：在选择最佳投掷物体时，大小-重量错觉不是错觉。*进化与人类行为*, *32*(4), 288 – 293.

Ziv, I., Djaldetti, R., Zoldan, Y., Avraham, M., & Melamed, E. (1998). 通过新颖客观运动评估诊断”非器质性”肢体瘫痪：定量胡佛试验。*神经病学杂志*, *245*, 797 – 802.

Zorzi, M., Testolin, A., & Stoianov, I. (2013). 用深度无监督学习建模语言和认知：教程概述。*心理学前沿*, *4*, 515. doi:10.3389/fpsyg.2013.00515.

索引

注意

自闭症

贝叶斯大脑模型

期望

眼动追踪功能障碍(ETD)，[208] – [9]

面部

梭状回面部区域(FFA)对面部的反应，[48] – [49]

凹面错觉与，[49] – [51]，[114]，[207]，[224]

面部识别中的扫视，[71] – [72]，[74]

促进效应，[61] – [63]

熟悉记忆系统，[102] – 5

雪貂视角实验，[185] – [87]

Field, D.J.，[21]，[271]

图形-背景凸性线索，[198]

过滤泡沫场景，[75]

力量升级，[113] – [16]，[213]，[217]

Fornito, A.，[148]

前向模型

自由能公式，[264]，[305] – [6]

生成模型

要点(gist)

Gu, X., [228]

Guillery, R., [146]

习惯化

[89]

幻觉

手写

分层贝叶斯模型(HBMs)

超先验

低先验

想象

模仿学习

[135] – [36]

渐进下游认识论工程，[276]

推理

内感受

学习

记忆

预测处理(PP)模型

瘙痒

视觉皮层