arXiv:2601.07830v1 [cs.LG] 12 Jan 2026

平衡努力与性能的最优学习率调度

瓦伦蒂娜·尼亚拉迪 *
伦敦大学学院盖茨比计算神经科学部
valentina.njaradi.23@ucl.ac.uk

罗德里g戈·卡拉斯科-戴维斯*
盖茨比计算神经科学部伦敦大学学院普林斯顿神经科学研究所
rodrigo.cd@princeton.edu

Peter Latham
伦敦大学学院 Gatsby 计算神经科学中心

Andrew Saxe
伦敦大学学院 Gatsby 计算神经科学中心 Sainsbury Wellcome 中心

摘要

如何高效地学习，是生物智能体面临的一项根本挑战，也日益成为人工智能体关注的问题。要有效学习，智能体必须调节学习速度，在快速进步的益处与努力、不稳定性或资源使用所带来的成本之间取得平衡。我们提出一个规范性框架，将这一问题形式化为最优控制过程：智能体在承担学习成本的同时，最大化累积表现。基于这一目标，我们推导出最优学习率的闭式解，其形式为一种闭环控制器，仅取决于智能体当前及预期的未来性能。在温和的假设条件下，该解可推广至不同任务和架构，并能在模拟中复现数值优化得到的学习率计划。对于简单的学习模型，我们可以从数学上分析智能体参数和任务参数如何影响学习率调度，并将其作为开环控制解来刻画。由于最优策略取决于对未来性能的预期，该框架预测了过度自信或信心不足如何影响投入和坚持性，并将学习速度控制与自我调节学习理论联系起来。我们进一步表明，一种简单的情景记忆机制可以通过回忆相似的过往学习经历，近似所需的性能预期，从而为实现近乎最优的行为提供一条在生物学上可信的途径。总体而言，这些结果为学习速度控制提供了一种规范性且具有生物学合理性的解释，并在一个统一且易于处理的数学框架中，将自我调节学习、努力分配和基于情景记忆的估计联系起来。

人类和其他动物始终面临着如何将有限的认知资源分配给学习的问题。这不仅意味着要决定什么值得学习，还要决定如何应对每项学习任务。例如，学习钢琴时，有人可能采用快速的学习策略，记住一串琴键以演奏简单的歌曲；另一些人则可能采用较慢的学习策略，学习乐理并掌握读乐谱的方法，以便日后将所学推广到更广泛的曲目。在这项工作中，我们聚焦于这一普遍问题的抽象版本：学习系统在权衡性能与快速学习的成本时，应如何调节学习速度。

从童年起，人类便会主动调节自己的学习（Taffoni 等人，2014；Raz 和 Saxe，2020；van Gog 等人，2020）。他们主动寻求信息，选择与自身能力相匹配的输入（Kidd 等人，2012），并权衡学习的潜在收益与成本（Bonawitz 等人，2018）。与被动学习相比，成年人在能够掌控学习进度和结构时，学习效果也更好（Castro 等人，2008；Markant 和 Gureckis，2014；Coenen 等人，2019；Tullis 和 Benjamin，2011）。个体自身的元认知控制至关重要：在比较集中学习与间隔学习的研究中，只有当间隔安排符合参与者自己选择的学习计划时，他们才能从中受益（Son，2010）。对于哪些内部状态会驱动这种控制，理论界意见不一：好奇心、无聊及其他动机

*共同贡献。

信号都被认为与之有关（Bazhydai 等人，2020；Hidi 和 Renninger，2019；Schwartenbeck 等人，2019；Geana 等人，2016；Agrawal 等人，2022）。但根本问题依然存在：哪些计算过程会引导人们决定如何学习?

许多因素都会影响学习计划。回到钢琴的例子，“正确”的策略取决于可用于练习的时间、努力程度、先验知识、任务难度、演奏乐器的长期价值，甚至个人的乐趣。证据表明这些因素都很重要，但孤立地研究它们可能会得出相互矛盾的结果。例如，人们有时偏好中等和高难度的任务（Baranes 等人，2014；Schulz 等人，2019），而其他研究则发现，人们偏好处于“恰到好处区间”的任务，即既不太容易也不太难（Kidd 等人，2012；Wilson 等人，2019；Ten 等人，2021；Cubit 等人，2021）。时间压力也很重要：时间紧迫时，学习者会优先选择简单任务；而有充足的学习时间时，他们则会选择更难的任务（Son 和 Metcalfe，2000）。学习者也会留意自己知识上的不足，并偏好尚未掌握的内容（Kidd 等人，2012；de Eccher 等人，2024）。学习所带来的未来奖励价值（Masís 等，2023；Obando 等人，2025）也会影响选择；研究表明，啮齿动物、猴子和人类都会权衡短期收益与学习的长期价值。

学习带来诸多益处，但也伴随着成本。有些是代谢方面的：人类研究表明，代谢与记忆和学习功能增强有关（Potter 等人，2010；Smith 等人，2011；Klug 等人，2022）；而在苍蝇和蝴蝶中，密集学习会降低繁殖适合度并缩短寿命（Mery 和 Kawecki，2004；Snell-Rood 等人，2011；Mery 和 Kawecki，2005）。另一些成本则源于认知控制本身所需的努力。安排和调节学习是一种认知控制（Masís 等人，2021，2023；Masis等人，2024；Carrasco- Davis 等人，2024），而认知控制本身也有成本，参与者在分配控制时会将这些成本考虑在内（Kool 和 Botvinick，2018；Kurzban 等人，2013；Shenhav 等人，2013，2017；Agrawal 等人，2022）。在学习情境中考虑控制成本时，研究表明，受试者往往会选择认知要求较高而非较低的任务（Jarvis 等人，2022），尤其是在这些任务提供学习机会时（Sayalı 等人，2023）。这表明，人类能够权衡参与具有挑战性的任务所产生的即时成本与学习预期带来的未来收益（Obando 等人，2025）。

这些发现表明，人类在决定如何学习时，会持续权衡成本与收益。从计算角度来表述，如何对智能体的学习进行最优调控，以提高学习速度或质量，是元学习领域长期以来广泛研究的问题（Hospedales 等，2020）。在神经网络中，元学习可以有多种形式：设计最优学习课程（Zhang 等，2021；Stergiadis 等，2021；Soviany 等，2022）、为网络寻找最优初始权重（Finn 等，2017）、持续学习（Parisi 等，2019；Wang 等，2024）、优化超参数（Franceschi 等，2018；Baik 等，2020），或确定整个训练过程中的最优学习率计划（Nakamura 等，2021；Mori 等，2025；Mignacco 和 Mori，2025）。然而，设计最优学习速度调度在计算上颇具挑战，因为在大多数情况下，这需要对整个学习轨迹中的性能进行积分和求导（Franceschi 等，2018；Carrasco-Davis 等，2024；Mori 等，2025；Mignacco 和 Mori，2025）。对于生物智能体而言，这种计算可能更具挑战性，因为它们进行最优控制计算的能力可能有限，而且所需的量也未必容易获得。尽管如此，管理学习是决定长期成功的关键因素。所有人类和动物每天都面临这一问题，这引出了一个关键问题：它们是如何做到的?

在本文中，我们采用规范性方法处理这一问题，将学习速度控制形式化为任务表现与努力之间的权衡（第 1 节）。我们推导出最优学习速度的闭环解析表达式，该表达式仅取决于智能体通过学习可能获得的量，例如其当前性能和预期最终性能，从而无需预测完整的学习轨迹（第 2.1 小节）。我们展示了该解可推广至不同任务和网络架构，并将其与次优策略进行比较，同时探讨生物智能体如何利用过去的学习经历所形成的记忆来学习并近似估计自身预期的最终性能（第 2.2 小节）。对于感知机等简单模型，我们推导出精确的开环解，从而直观说明任务难度和努力成本如何影响学习（第 2.3 小节）。我们将该框架扩展到包含未来奖励折扣的情形，为简单模型提出近似解，并通过模拟揭示折扣对最优学习策略更广泛的影响（第 2.4 小节）。我们将该框架的结果与动物实验进行比较，并提出进一步的实验预测（第 2.5 小节）。最后，我们讨论该理论对机器学习和认知神经科学的更广泛意义（第 3 节）。

1 平衡学习努力与性能

在我们的研究中，智能体在长度为 T 的学习回合中学习一项任务，并在整个回合期间收集奖励，同时承担学习成本。智能体的目标是在学习过程中最大化累积奖励，同时尽可能降低为此付出的努力。智能体可以通过调整其学习速度 $\mu(t)$来实现这一目标

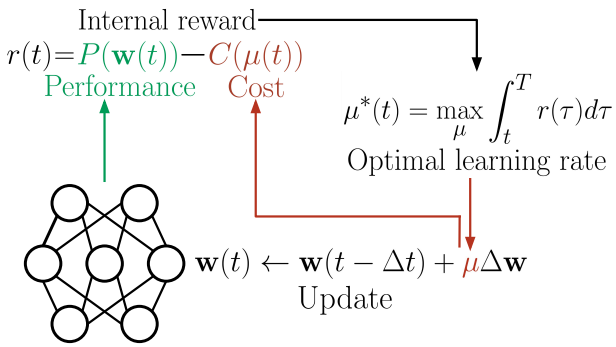


图1：**理论框架**：在神经网络训练过程中，计算权重更新的最优学习率，以最大化整个学习过程中的累积内部奖励。内部奖励是通过执行任务 $P(t)$ 所获得的奖励与以速率 μ 学习的成本 $C(\mu(t))$ 之差。

在整个回合期间（图 1）。智能体在学习回合中执行该任务的性能由可微函数 $P(t)$ 定义，该函数取决于智能体和任务的参数。要应用下文所述的解析方法，该函数必须是确定性的，并且关于智能体的参数连续可微。

我们将智能体建模为一个具有 N 个可学习参数（即权重）的人工神经网络，这些参数收集于向量$\mathbf{w} \in \mathbb{R}^N$中。在这种情况下， $P(t)$ 仅通过这些权重依赖于时间： $P(t) = P(\mathbf{w}(t))$。该网络的架构几乎不受限制，可以任意深、非线性、循环或包含注意机制，涵盖玩具模型和现实智能体。该网络通过连续时间极限下的梯度更新进行学习，这种机制称为梯度流（Saxe 等，2019；Boursier 等，2022；Miyagawa，2023）：

$$\frac{d\mathbf{w}}{dt} = \mu(t)\frac{dP(\mathbf{w}(t))}{d\mathbf{w}} \tag{1}$$

其中， $\mu(t)$ 是智能体在时间 t 的学习率。对网络唯一的条件是， $P(\mathbf{w})$ 关于 $\mathbf{w}$ 连续可微，且其梯度 ${}^{dP}{}_{(\mathbf{w}(t))}\, {}_{d\mathbf{w}}$ 关于 $\mathbf{w}$ 满足 Lipschitz 条件，对所有~~t 均~~成立。在实践中，这意味着模型中的所有激活函数都必须平滑，且一阶和二阶导数有界，因此不包括 ReLU 类激活函数。智能体只能通过调整学习率 $\mu(t)$ 来影响自身的学习，即加快学习、减慢学习或完全停止学习；其动态过程的其他方面（例如梯度流学习算法）则保持不变。

智能体收到的外部奖励 随任务表现 $P(\mathbf{w}(t))$ 而变化；随着任务在回合中逐渐被学会，任务表现通常会提高。同时，智能体以速度 $\mu(t)$ 进行学习时会产生成本 $C(\mu(t))$。内部奖励 率定义为这两者之差：

$$r(t) = P(\mathbf{w}(t)) - C(\mu(t)) \tag{2}$$

因此，平衡努力与表现对应于最大化内部奖励的折扣积分：

$$\mu^*(t) = \arg\max_{\mu(t)} \int_t^T d\tau\, \gamma^{\tau-t} \underbrace{(P(\mathbf{w}(\tau)) - C(\mu(\tau)))}_{\text{internal reward rate } r(t)} \tag{3}$$

其中， $\gamma \in [0,1]$ 是折扣因子，反映智能体对未来奖励的重视程度，而 T 表示回合时长。

尽管性能函数 $P(\mathbf{w})$可以采用任意连续形式，但解析可处理性取决于成本的选择。对于任意在 $\mu > 0$上为凸的函数 $C(\mu)$，都存在精确解；而只要问题 $\mu \frac{dC}{d\mu} - C(\mu) = \text{const}$，就存在解析解，并可得到 μ 的闭式表达式（见方法中的式 19）。在本研究中，我们采用关于学习率的二次成本，其中 β 用于设定学习成本：

$$C(\mu) = \beta\mu^2. \tag{4}$$

这种设定为我们提供了一种简单而通用的方式，将学习视为一个控制问题（Bertsekas，2000年；Boscain等人，2021年；Chacko 等，2024年），类似的设定也曾多次应用于认知控制文献（Shenhav 等，2013年；Masís 等人，2021年；Piray和 Daw，2021a年；Agrawal 等人，2022年）。值得注意的是，在

优化学习率时，我们推导出了闭环学习率调度的解析表达式（第 2.1 小节），并在较简单的设定中，给出了关于任务和智能体参数的开环表达式（第 2.3 小节）。这与以往研究形成对比：以往大多数解都是通过数值方法获得的（Franceschi 等，2018年；Carrasco- Davis 等，2024年；Mori 等，2025年；Mignacco 和 Mori，2025年）。通过适当定义性能指标，我们的表述可应用于任意任务。例如，在双选强迫选择任务中（Bogacz 等，2006；Masís 等，2023），性能可以对应于预测刺激—奖励关联的准确率；在下一词元预测任务中，$P(t)$可以是交叉熵损失的相反数；而在无监督学习问题中，$P(t)$可以表示重构误差的相反数或重构准确率。无论任务或模型架构如何，同一个问题依然存在：智能体应如何调度其学习率 $\mu(t)$，才能通过学习获得高额奖励，同时避免在努力上花费过多？换言之，什么时候值得加快学习，什么时候又最好放慢速度？

2 结果

乍看之下，要确定任意网络和任务的最优学习速度调度，似乎需要完整掌握学习轨迹，而这只有在简单模型中才能通过解析方法处理。本节的主要结果是，在没有折扣奖励的情况下（$\gamma = 1$），我们可以绕过这一障碍。我们推导出一个简单的解析表达式，用于计算能够使学习期内累积折扣内部奖励（式 3）最大化的最优学习率 $\mu^*(t)$，从而在无需掌握完整轨迹的情况下平衡性能与努力（第 2.1 小节）。值得注意的是，我们证明，在对学习动力学作出温和连续性假设的前提下，无论任务和网络架构如何，该解都适用。该表达式依赖于智能体在自然情境学习环境中可能获取的变量，例如其当前性能 $P(t)$ 和预期最终性能 $P(T)$。尽管学习者未必总能直接获取 $P(T)$，但我们将在后文说明，如何利用情景记忆模块估计未来性能，从而对其进行估算（第 2.2 小节）。在感知机等学习动力学可显式求解的特殊情况下，我们将展示这一通用闭环表达式如何进一步导出完整的开环计划（第 4.3 小节）。我们还将探讨任务参数如何影响学习率信号（第 2.4 小节），以及采用次优学习策略所带来的后果（第 2.1 小节、第 2.2 小节）。最后，我们将理论与现有实验进行比较（第 2.5 小节），并在讨论部分提出更多预测。

2.1 最优学习率计划的一般解

这个问题可以用标准的控制理论术语来表述，其中学习率就是控制信号。在连续时间控制问题中，Hamilton–Jacobi–Bellman（HJB）方程（Bellman，1958年；Bertsekas，2000年）是一种偏非线性微分方程，它为最优价值函数及相应的控制信号提供了条件。要获得该偏非线性微分方程的精确解析解，一个可行的条件是令折扣因子参数为 $\gamma = 1$，这对应于一个对所有未来奖励赋予同等价值的智能体。由此，我们可以得到最优学习率计划（推导过程见方法部分第4.1小节）。给定学习规则（式1）和控制成本 $C(\mu)$（式4），学习率的最优解为：

$$\mu^*(t) = \sqrt{\frac{1}{\beta}(P(T) - P(t))} \tag{5}$$

其中，β是努力成本，$P(t)$是时刻 t 的性能，而 $P(T)$ 是在学习回合结束时按照最优学习率计划 μ^*所达到的性能。

这一解所蕴含的最优策略直观易懂：一开始快速学习，并在早期投入大量努力；随后逐渐减少努力，同时继续享受早期进步带来的益处。这直接源于以下事实：随着 $P(t)$ 单调增加，$P(T) - P(t)$ 会随时间推移而下降。当性能快速提升时，最优学习速度也应同样快速下降。有趣的是，如果性能趋于平稳，预测结果并不是增加努力，而是维持相同的学习率并耐心等待。最后，一旦智能体掌握了任务，且 $P(T) - P(t) \to 0$，最优策略就是彻底停止学习。成本参数 β 调节性能与努力的权衡，因此，成本越高，整体上的最优学习率就越低。

为了展示这种行为，我们在不同任务上训练了深度非线性网络，并根据最优学习率（公式 5）设定其学习率。一个网络在教师–学生设置下训练（Goldt et al., 2019；Lee et al., 2022），另一个则在 MNIST 图像分类任务上训练（Deng, 2012）。教师–学生网络在梯度流极限下训练，我们展示了理论最优学习率（公式 5）与数值优化的学习率（图 2a,b）相吻合，符合此前描述的行为。尽管最优学习率计划是针对理想化的全批量梯度流学习动力学情形推导出来的，但现实中的智能体，无论是生物智能体还是人工智能体，往往通过批量更新或噪声梯度进行学习。在图像分类任务上训练的网络

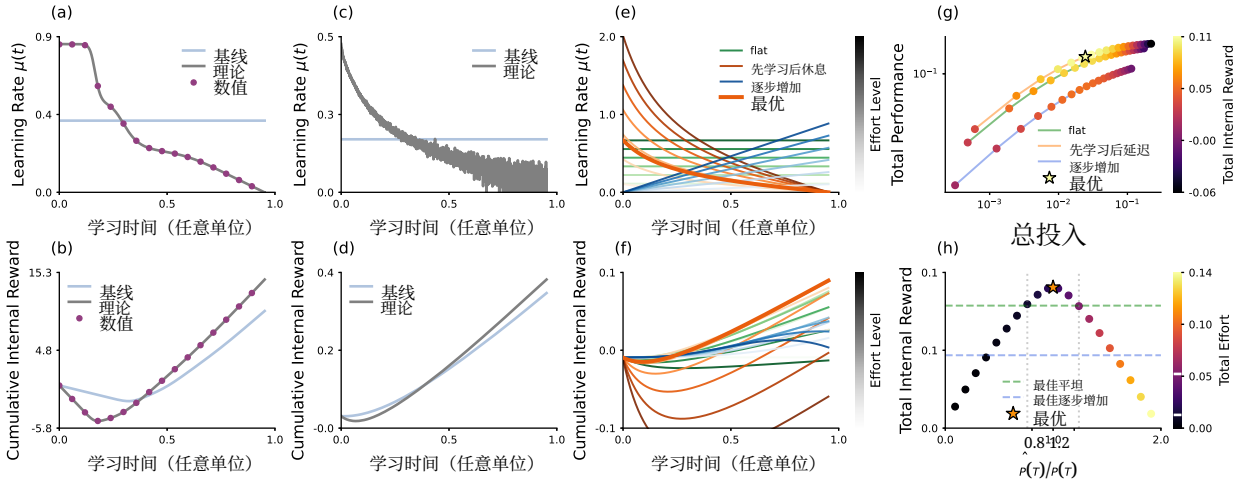


图 2：**闭环表达式可推广至任意任务和网络。（a、c）**：采用梯度流训练的教师-学生设置中的双层非线性网络**（a）**，以及采用随机梯度下降训练的数字分类/MNIST 任务**（c）**的最优学习率计划。**（b、d）**：5 中最优解（灰色）、采用性能最佳的固定学习率的基线智能体（实线蓝色）以及数值优化（紫色点）的累积内部奖励。**（e、f）**不同学习策略的学习率（e）和累积内部奖励（f）。颜色深浅表示策略投入的努力多少，颜色表示策略类型（紫色——固定速率，橙色——先学习后休息，蓝色——逐步增加，粗体橙色——最优）。**（g）**不同学习策略的性能/成本权衡图（每个点代表一种不同的学习率计划）。色条显示累积内部奖励率，即优化目标。黄色星号表示最优策略。各策略位于学习曲线对应的线上（颜色同（e、f））。**（h）**最优学习策略的累积奖励随最终性能相对估计误差 $\hat{P}(T)/P(T)$ 的变化情况，涵盖低估和高估状态（色条显示总投入）。彩色虚线表示每种学习曲线对应的性能最佳策略，星号标示真正的最优策略。最终性能估计误差不超过 20% 的最优曲线，其表现优于最佳平坦策略（由竖直灰色虚线标示）。相同的误差界限也用白线标示在投入色条上。实验细节见第 4.5 节。

采用随机梯度下降进行更新（图 2c,d）。理论学习率计划带有噪声，反映了性能信号的波动，但其表现始终优于固定基线策略。这表明，即使在噪声学习状态下，理论框架仍然有效；在这种状态下，学习必须依据不完美反馈进行在线调节（网络训练的详细信息见方法第 4.5 小节）。

相应的累积内部奖励信号（图 2b、d）凸显了学习策略的影响：智能体起初付出的成本高于获得的奖励，但这种前置努力会随着性能迅速提升而得到回报，带来更高的奖励率，从而弥补为加快学习而在早期努力中产生的成本。相比之下，采用平坦的小学习率的基线智能体最终也能学会任务，但其累积的总体内部奖励显著较少；相对于最优策略，它要么性能较差，要么付出了过多努力。

为了说明不同学习策略的展开过程，我们比较了三种在性质上截然不同的学习曲线，它们在努力如何随时间分配方面存在差异。第一种是固定学习曲线，即在整个学习回合中保持固定学习率和恒定努力。第二种是逐步增加曲线，即智能体从较慢的节奏开始，并逐渐增加努力。第三种是根据我们的理论推导出的最优“快速学习，之后休息”曲线。对于每种曲线，我们改变努力的最大水平，以形成不同的学习策略（图 2e），然后比较它们的性能、成本和总体目标（内部奖励总量）。三种策略的累积内部奖励率曲线存在显著差异，其中最优策略产生的总内部奖励最高（图 2f）。网络学习同样揭示了性能与努力之间的权衡：不出所料，更高的努力会带来更多外部奖励，但超过某一点后会出现收益递减（图 2g）。有趣的是，尽管多种策略可以产生相近的性能—努力比（例如固定策略和最优策略），但只有采用适当努力水平的最优形状才能最大化累积内部奖励。此外，正确设定学习策略的形式比精细调整确切的努力水平更为重要——即使对努力程度存在中等程度的估计偏差，遵循正确定性曲线的智能体仍能保持接近最优（图 2h）。

最优学习率的闭环解（公式 5）提供了一个规范性基准：它告诉我们，如果智能体能够同时获得其当前性能 $P(t)$ 和最终性能

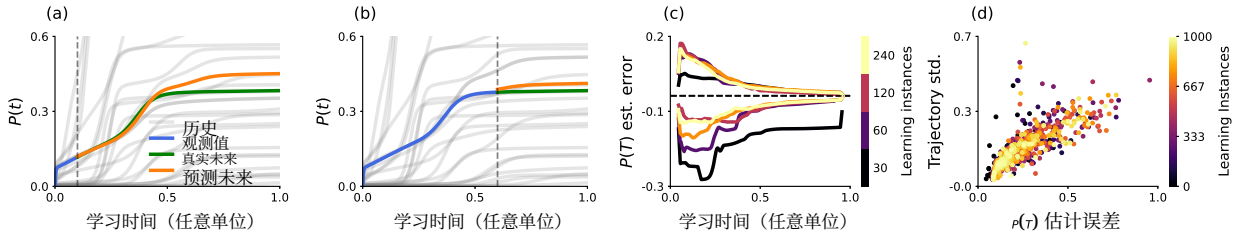


图 3：**通过情景记忆估计最终性能。（a、b）**：基于观测到的轨迹 $\mathcal{T}_o$（蓝色），利用其与记忆 $\mathcal{T}_h$（灰色）中存储的轨迹之间的相似性，估计整个学习过程中的未来轨迹（橙色）。真实未来轨迹以绿色显示。**（c）**：估计误差 $\hat{P}(T|t) - P(T)$ 随回合内时间 t 的变化。对于每种记忆大小（颜色；所用过往学习回合的数量），我们绘出各试次误差的第 25 和第 75 百分位数（平滑后）。随着记忆增加、当前轨迹被观测得更多，误差范围会缩小，且分布越来越对称，表明偏差和方差均有所降低。**（d）**：学习轨迹的变异性（以标准差衡量）与估计误差的关系，颜色表示学习实例在记忆中的顺序。变异性较高的轨迹会产生较大的估计误差。较晚的学习实例聚集在低变异性和接近零误差的区域，但即使在学习后期，轨迹变异性增加仍与可预测性降低相关。

$P(T)$，那么最优学习率会是什么样。在学习能够带来性能提升的情况下，最优学习率曲线呈现出“先快速学习，后期休息”的形态：开始时采用较高的学习率，随着性能接近其渐近值，再逐渐降低学习率。据我们所知，我们为学习率控制问题提供了一种新颖的闭环解析解，该方法可应用于此前讨论的其他任务和架构；而以往的研究主要依赖数值优化来计算最优学习率调度。在实践中，尽管 $P(t)$ 在大多数时候是可观测的（稀疏奖励或噪声反馈情况下除外），但 $P(T)$ 并不可观测，因为获取它需要了解学习系统未来的时间演化。这引出了一个自然的问题：真实智能体如何近似这一量？当估计不准确时又会发生什么？如前所述，轻微到中等程度的估计误差影响很小，但对 $P(T)$ 的严重误估计会导致显著差于最简单恒定学习策略的结果（图 2h）。这些考虑促成了下一节的内容：我们提出一种基于情景记忆来估计 $P(T)$ 的机制，并考察误估计对学习性能和努力的影响。

2.2 通过情景记忆估计最终性能

根据所提出的理论，最优学习要求在回合结束时，按照最优学习率计划 $\mu^*(\cdot)$，准确估计任务表现 $P(T)$。我们假设，对这一数值的估计构成了一种元学习：智能体基于先前的学习回合，学会预测自身的进步将如何演变。

我们提出了一种简单且在生物学上合理的模型，利用情景记忆估计未来性能。每条存储的记忆都对应一段过去的学习轨迹 $\mathcal{T}(t)$，该轨迹由随时间变化的性能 $\mathcal{T}(t) = \{P(0)\ P(dt) \ldots, P(t)\}$定义。面对新任务时，智能体会将当前正在经历的轨迹 $\mathcal{T}_o(t)$ 与 H 从记忆中检索出的 $\{\mathcal{T}_h$先前存储的轨迹 $t\}_{h \in H}$ 进行比较。智能体利用高斯核 $w_h(t) \propto k(\mathcal{T}_o(t)\ \mathcal{T}_h(t))$计算相似度权重 $\{w_h(t)\}_{h \in H}$，并根据截至当前时间 t 的信息，外推其在$T > t$时的未来性能：

$$\hat{P}(T \mid t) = \sum_{h \in H} w_h(t)\, P_h(T) \tag{6}$$

在每个任务（$t = 0$）开始时，由于尚未观测到任何轨迹，所有先前的经验对初始估计 $P(T \mid t = 0)$ 的贡献都相同。

我们在 1,000 个序贯学习回合中评估了该估计器。每个回合都涉及一个权重随机采样的非线性教师网络，而智能体网络每次都会随机初始化，并使用教师生成的数据进行训练，从而产生多样的学习动力学。随着每个回合的推进，智能体会根据观测到的轨迹（图 3a,b）持续优化 $\hat{P}(T \mid t)$，并据此计算不断变化的最优学习率 $\mu(t)$（见方法 4.2 ）。达到时间限制 $t = T$ 后，完整轨迹 $\mathcal{T}(T)$ 会被存储在记忆中，然后才开始下一个学习回合。

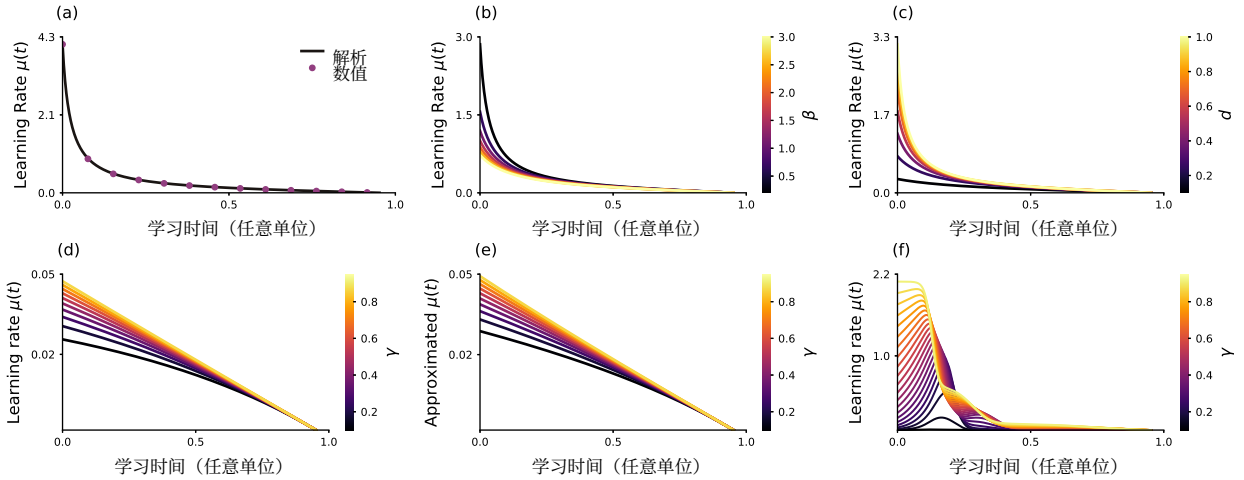


图4：**最优学习策略会随参数以可预测的方式变化。（a-c）**学习回归任务的线性感知机。**(a)** 最优学习率（点：数值优化结果；线：由公式7给出的解析表达式）。**(b)** 提高控制成本系数 β 会以非线性的方式降低学习率。**(c)** 任务难度定义为目标权重与初始权重之差的范数 $d = \|\mathbf{w}^* - \mathbf{w}_0\|^2$，任务难度越高，最优学习率越高。**(d,e)** 采用折扣时间学习线性回归任务的单个线性神经元的学习率，其中（e）为公式9给出的解析近似，（d）为数值计算结果。该近似保留了精确解的定性特征，表明折扣程度越高，学习率越低。**(f)** 学习教师-学生回归任务的两层线性网络，显示改变 γ 会对最优学习率产生非线性影响。总体而言，降低努力成本、提高任务难度或提高折扣因子，都会增大最优学习率。此外，当存在非线性学习动力学时，强折扣可能彻底改变学习计划的形状。实验细节见第4.5节。

随着当前轨迹中被观测到的部分不断增加，相似度计算会依赖更长的轨迹片段，从而改进对剩余性能的预测（图3c）。随着记忆中累积的学习轨迹越来越多，估计准确度的变异性和偏差都会降低（图 3c），进而能够更好地重建学习动力学，并减小 $\mu(t)$ 与理论最优值 $\mu^*(t)$ 之间的偏差。这表明了一种元学习形式：随着经历的学习回合增多，对学习的控制会更加出色。然而，估计准确度仍受到学习动力学内在变异性的强烈影响：变异性较高的轨迹通常更难预测，且会产生更大的估计误差，即使到了学习后期也是如此（图 3d）。

尽管通过情景记忆进行性能估计取决于任务分布的结构，但这些结果表明，一种简单的基于记忆的机制，只需将当前轨迹与先前经历过的轨迹进行比较，就能实现近乎最优的行为。这为生物智能体和人工智能体提供了一条可行途径，使它们能够在整个生命周期中进行元学习，在遇到未见任务时学会如何分配学习努力。

2.3 线性感知器的数学分析

虽然闭环解规定了在线最优学习率（公式 5），但它对由此产生的训练动力学所提供的洞见有限。在这里，我们推导出一个开环解，用以刻画完整的最优学习轨迹。闭环解非常适合需要实时适应的智能体，但难以据此详细分析完整轨迹。例如，在前文的非线性网络训练案例中（图 2a），学习率呈现出下降和平台期，与性能曲线的形状相呼应。我们可能会预期深层非线性网络中也会出现这类平台期，但无法提前预测其出现时间。同样，对于循环神经网络或 Transformer 等架构，只有在训练之后才能分析 $\mu(t)$ 的轨迹。相比之下，对于线性感知器这类简单模型，在对任务、数据分布和输出非线性作出某些假设后，其学习动力学是明确已知的（Horner，1993年；Saad 和 Solla，1995年），因此我们可以将其代入闭环表达式，推导出完整的开环解。

考虑一个权重为 $\mathbf{w}$ 的线性感知机，它在一个线性回归任务上进行训练，学习率成本（公式 4）和外部奖励定义为负均方误差（MSE）损失，$P(t) = -L(t)$（$\eta = 1$ 见公式 2）。训练输入从 $\mathbf{x} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ 中独立同分布地抽取，标签由目标权重 $y = \mathbf{x}^\mathrm{T}\mathbf{w}^*$ 生成。网络以权重 $\mathbf{w}_0$，初始化，任务的难度由平方距离 $d^2 = \|\mathbf{w}^* - \mathbf{w}_0\|^2$ 表示。

利用这一设定计算最优学习率（公式 5），可得到关于$\mu^*(t)$的自洽积分微分方程（见方法部分 4.3）。求解该方程可得以下开环表达式：

$$\mu(t) = \theta \tan\left(\sigma^2\theta(T-t)\right) \tag{7}$$

$$\theta = \frac{\sigma d}{\sqrt{2\beta}} \cos\left(\sigma^2\theta T\right) \tag{8}$$

其中，θ 是通过数值方法从自洽的公式 8 中得到的。

该解的开环形式保留了闭环形式的直观含义：由于 tan($T-t$) 是时间的递减函数，最优计划会先快速学习，然后逐渐放慢速度。我们验证了开环表达式与图 4a 中数值优化得到的学习计划一致。这个显式解有助于考察最优计划如何依赖于任务参数。尽管曲线的定性形态保持稳定，但较低的努力成本和较高的任务难度意味着应采用更高的学习率（图 4b、c）。

到目前为止，我们假设智能体对所有未来奖励的重视程度相同（$\gamma = 1$）。对于时间尺度很长的任务，这一假设可能不切实际；此时，考虑未来奖励折扣后最优学习策略会如何变化就变得十分重要。

2.4 时间折扣效应

引入折扣因子 $\gamma < 1$ 会显著改变控制问题，使近期改进优先于更遥远未来的奖励。折扣能够体现智能体对未来结果的不确定性，可用于考虑长时域预测所需的有限计算资源，并确保学习在随机环境中的稳定性。与无折扣情形不同，在无折扣情形下，学习努力可以分配到整个学习回合；现在，智能体会优先考虑近期结果，并逐步降低对更远期奖励的重视程度。

折扣的存在破坏了 Hamilton-Jacobi-Bellman 方程的解析可处理性。尽管如此，仍可使用同伦摄动法结合帕德近似，为线性模型获得近似解析解（Ganjefar 和 Rezaei，2016）；该方法已在控制理论文献中得到充分确立（Barari 等，2008；Abdulaziz 等，2008；Buhe 等，2023）（见方法部分 4.4）。这种方法生成一系列精度逐步提高的解析近似；在某些参数区间内，这些近似会收敛至真实解，其中前几个模态已经能够捕捉折扣因子 γ 的主要定性影响。

我们首先考虑使用均方误差损失训练单个神经元完成线性回归任务。首阶近似给出了一个简单且易于解释的最优学习率计划：

$$\mu^*(t) \approx \frac{1}{2\beta} \frac{\left(\frac{dP}{dw}\right)^2}{\frac{1}{T-t} - \frac{1}{2}\log\gamma} \tag{9}$$

该近似明确展示了折扣如何抑制学习率的幅度：降低折扣因子会减少总体努力分配，并降低学习率（$\log\gamma$ 项）。该近似还预测了学习回合接近结束时（$T-t$ 项）以及性能梯度达到饱和时（$\left(\frac{dP}{dw}\right)^2$ 项）控制的降低。尽管形式简单，这一首阶解仍与数值优化的学习率计划高度吻合（图 4d、e）。

对于更复杂的架构，解析近似的限制日益明显，因此我们分析了数值优化得到的解。在两层线性网络（图 4f）中，未折扣解（$\gamma = 1$）呈现出显著的下降和平台期，这反映了权重耦合所引发的分阶段学习动力学（Saxe etal., 2019; Braun et al., 2022）。随着折扣因子降低，这些特征逐渐消退，模型也变得越来越着眼于短期（$\gamma \ll 1$）。

有趣的是，当 γ 值足够小时，最优学习率计划会进入一个性质截然不同的状态（见图 4f）。学习率起初保持在接近零的水平，随后逐渐增加，之后再降低。这种行为反映了在强折扣和非线性学习动力学下对努力的理性分配。由于梯度较低，且剩余时间不足以将努力转化为性能提升，早期学习的成本相对较高；因此，最优策略是在初期节省学习率方面的努力，之后再投入，因为届时学习梯度更大，单位成本带来的提升也更高。这种状态不同于单神经元的情形，源于折扣与更深层网络特有的阶段性学习动力学之间的相互作用。

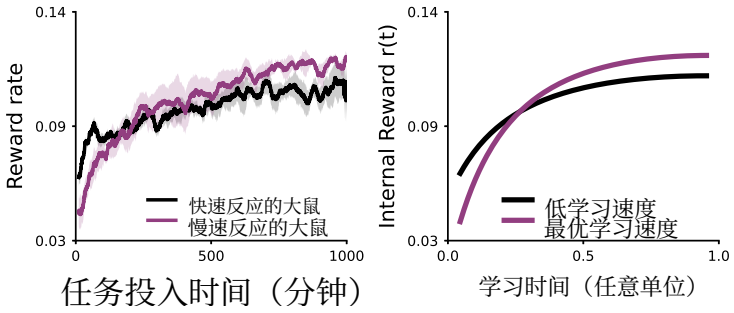


图5：**与数据的比较**。 左：二选一强迫选择任务中的瞬时奖励率随时间的变化。根据大鼠相对于此前实验中个体基线的反应时间对其进行分组（快速，黑色；缓慢，紫色）。数据改编自（Masís 等，2023），图7b。右：执行双高斯判别任务的神经网络模型。一个模型遵循最优学习策略（紫色），另一个模型使用相同的变化曲线，但学习率幅度较低（黑色）。动物表现出即时奖励率与学习速度之间的权衡：即时奖励较低（例如，控制程度更高导致反应时间更长）可能带来更快的学习速度和更高的累积奖励。我们的框架将这种权衡描述为控制成本与学习率之间的平衡。实验细节见第4.5节。

解析近似以及闭环和开环解共同表明，折扣因子越大、努力成本越低、任务难度越高，就越适合采用更大的学习率计划。相反，当未来奖励受到强烈折扣时，学习努力会逐渐变得不那么值得，即使仍有可能取得进步，学习也会变得更慢或延迟。

2.5 与动物行为的联系

接下来，我们考察我们的框架所作预测与动物在学习过程中的行为有何一致之处。据我们所知，目前尚无实验系统地检验关于学习速度控制和努力成本的理论，但我们的框架与认知控制文献中的观点有显著相似之处。具体而言，Masis 等人开展的一项实验（Masís 等，2023）旨在检验动物在决策任务中是否会考虑长期学习，并表明收集短期奖励与学习速度之间存在权衡。研究人员训练大鼠完成一项二元分类任务：在其视野的左侧和右侧各呈现一幅图像，只有其中一幅图像会带来奖励；大鼠通过舔舐左侧或右侧端口来表示选择（见（Masís 等，2023）中的图5）。研究人员根据学习回合开始时大鼠的反应时间对其进行分组。与训练开始时反应时间较短的大鼠相比，反应时间较长的大鼠单位时间内获得的奖励较少，但学习速度更快。此外，反应较慢的大鼠比快速反应的大鼠更早达到较高的瞬时奖励率（图5a）。当实验中限制大鼠只能快速或缓慢作出反应，而非让其自由反应时，这些结果依然成立，且效应量更大。（（Masís 等，2023），图5）。

我们将这一结果与我们的框架进行比较，并表明，学习率较低和较高的两个人工智能体所产生的奖励率曲线，在定性上与实验中的两组大鼠相似。在这一比较中，我们使用模型中的内部奖励率，以及行为数据中的瞬时奖励率；同时注意到，较长的反应时间实际上可能对应着较高的努力成本。也就是说，这些大鼠加快学习速度的成本，是反应率降低所造成的机会成本。我们无法断言两组大鼠的折扣因子或努力成本是多少，也无法排除大鼠采用了略有不同的学习策略、但仍能产生相似奖励率曲线的可能性。通过设计更精确、旨在检验本文所研究问题的实验，可以直接评估这些假设。

我们通过这项动物实验所强调的是，在分配给学习的控制中存在跨期选择，其中努力及其相关成本会影响最终的学习轨迹，进而影响累积奖励。我们的框架对这一基本问题进行了形式化处理；随后我们将论证并讨论，这类选择普遍存在于动物学习中，尤其是人类学习中，而长期规划和终身学习正是发展高级认知能力的核心。

3 讨论

在这项工作中，我们将学习速度调度问题形式化为一种平衡努力与表现的规范性优化，具体而言，是通过优化整个学习过程中的总内部奖励来实现的。内部奖励率会随着任务表现的提高而增加，而付出更多努力则会使其降低（第 1 节）。这一框架适用于人工智能体和生物智能体的学习，因为它可用于考察各种模型架构的学习过程，涵盖简单的玩具模型和现实的非线性网络；由于其抽象表述，它也适用于本文所呈现任务之外的其他任务。此外，我们的框架在数学上是可处理的，能够提供一个闭环解（第 2.1 小节），该解依赖于可在多个学习回合中估计的量（第 2.2 小节）。我们可以完整刻画学习轨迹，并为更简单的模型获得开环解（第 4.3 小节），以及针对时间折扣奖励获得近似解（第 2.4 小节）。这使我们能够得到明确的数学表达式，表明学习率会随着努力成本降低、折扣因子增加以及任务难度提高而增加。通过以最大化奖励这一规范性目标为基础，我们得以从第一性原理理解对学习的控制。

下文将进一步讨论我们这项工作的意义：首先考察其对动物行为的启示（第 2.5 小节），并联系心理学文献，具体涉及自我调节学习（Bandura，1977；Zimmerman，2000）以及学习情境中的认知努力（Shenhav 等，2013；Kool 和 Botvinick，2018；Masís 等，2023；Masis 等，2024）。随后，我们将探讨与神经科学的联系，因为根据我们的理论，计算最优学习努力所需的许多量都与大脑中的神经活动和神经递质动态相关（Doya，2002；Aston-Jones and Cohen，2005；Shenhav 等，2013）。最后，尽管我们的规范性目标旨在解释自然行为，我们的框架对机器学习也具有启示，因为它明确了一种优化学习的方法。机器学习领域已对学习率计划进行了广泛研究（Choi 等，2020；Schmidt 等，2021；Kashyap，2023），而鉴于大型架构的广泛应用，这一点尤为重要。

作为自我效能感的性能估计

最优学习率计划（公式 5）建议按感知到的学习收益分配努力，即根据预期未来性能与当前性能之间的差异来量化学习收益，$P(T) - P(t)$。这一观点与自我调节学习文献（Bandura，1977；Pajares 和 Miller，1994；Zimmerman，2000；Zimmerman 和 Schunk，2011）高度一致。该文献认为，自我效能感（学习者对自己有能力成功完成学习任务的信念）会影响他们在任务上投入多少努力以及坚持多久。

自我效能感不同于简单的行动结果预期。首先，个体可能明白什么行为会带来成功，却仍怀疑自己是否有能力做到。其次，自我效能感需要个体对自身学习能力以及努力对自身学习的影响具备一定程度的元认知。在我们的表述中，这种信念体现为个体当前能够达到的水平与其相信自己能够达到的水平之间的感知差距，$P(T) - P(t)$，这一差距决定了最优学习努力水平。与此一致，相较于问题本身的客观难度，自我效能感对问题解决任务中的性能具有更强的预测作用（Pajares 和 Miller，1994；Kontas 和 Özcan，2022）；这与我们的预测相呼应：无论任务难度如何，对 $P(T)$的估计不准确都会导致非最优的学习速度和较差的结果（第 2.1 节和第 2.2 节）。

因此，学会准确估计自身未来的性能，与自我信念和自我调节技能的发展直接相关；而这两者对于适应性学习和长期成功都至关重要（Zimmerman and Schunk，2011）。与此相关的是，我们的框架有一个局限：它考虑的是针对单一任务优化学习率，而原则上，控制信号可以是多维的。在更复杂的情境中，智能体可能需要通过将学习努力分配给多个同时进行的任务，以最大化这些任务的累积奖励，从而决定不仅学习多少，还要学习什么。过去已有研究在简化的学习情境中探讨过相关分析（Son andSethi，2006；Son，2010）。沿这一方向拓展我们的理论，有望为注意资源有限时如何将学习努力分配给多个任务提供一套有原则的解释。

学习努力与任务难度

最优学习率表达式在概念上也与学习努力成本相关。它预测，学习成本越高，投入的努力就应越少（较大的 β 意味着较小的 μ，公式 5），实际上是将学习率计划视为随时间变化的努力计划。对于极其困难的任务（初始性能 $P(0)$ 与最终性能 $P(T)$ 之间差距很大），该理论预测初始努力会非常高，这对于生物学习者而言可能不切实际。个体很可能存在一个内在的努力阈值，超过该阈值是他们不愿意或无法做到的。如果掌握一项任务需要的努力水平 $\beta\mu^*(t)^2$ 超过了这一阈值，学习者预计会停止投入，或放弃最优策略；非学习行为实验中也观察到了这种情况（Kool 等，2010；Kool 和 Botvinick，2018）。这一观点与有效课程的基于认知负荷的解释（Sweller，

1988；van Merriënboer 和 Sweller，2005）以及自我调节学习文献中的研究发现（Bandura 和 Schunk，1981；Schunk，1990）相契合。通过将复杂任务拆分为更小、更易于管理的部分——这种方法在自我调节学习文献中称为 近端子目标 ——学习者能够提升自我效能感、保持进展，并降低过早放弃学习的可能性（Stock 和 Cervone，1990）。

相反，任务难度突然增加可能会扰乱学习。实证研究表明，在关键的教育阶段过渡时期（例如从初中升入高中，或从高中升入大学），学生的学业表现往往会下降（Rice，2001；Neild，2009）。这些影响部分源于社会因素（Evans 等，2018），部分源于学业要求的变化（Hills，1965），还有部分源于心理因素（Lin 等，2023）。在我们的框架中，这类现象可能是学习者误判自身表现轨迹所致。例如，先前表现出色的学生可能会过度自信地进入新环境，高估自己的当前性能（高 $P(t)$），因而认为 $P(T) - P(t)$差距小于实际值。这会导致他们在学习上投入的努力不足，并对较低的考试分数感到意外。相反，低估当前能力却高估未来性能的学生（低 $P(t)$，高 $P(T)$）可能会在初期付出过多努力；若性能提升未能达到预期，努力与奖励之间由此产生的不匹配可能会导致倦怠（Salmela-Aro 等，2008；Barbosa 等，2016）。尽管这些对应关系仅是定性的，但这些模式说明，对$P(T)$的估计出现误差，或超出努力限制，可能会如何与现实中的学习轨迹相关联。

在线努力调节

我们的框架还指出，智能体原则上可以如何即时计算最优学习速度。在模型中，当前性能 $P(t)$ 表示任务成功与否的可观测指标，而预期最终性能 $P(T)$ 则必须经过估计，这一过程可能需要借助元学习，并利用先前的学习经历。在现实场景中，$P(t)$ 往往只能被稀疏地观测到（例如，反馈仅限于间歇性测试成绩），因此可能需要使用类似情景记忆的估计器，动态推断 $P(t)$ 和 $P(T)$。持续进行的估计可能产生预测误差，即预期性能与观测性能之间的差异；正如前文所述，这些误差可能驱动多种学习行为，例如脱离、过度自信和倦怠。自我调节学习领域的研究表明，频繁且准确的反馈能够提高投入（Schunk，1990；Schunk 和 Rice，1991）；在我们的框架中，这意味着需要对 $P(t)$ 作出精确及时的估计，以优化投入，而自我效能感则可能对应于对最终性能 $P(T)$ 的估计。

努力与神经调节

这一观点自然地联系到已知的、参与调节学习与努力的神经调制系统。多巴胺编码奖励预测误差（Glimcher, 2011; Schultz, 2016; Taira and Sharpe, 2025），并可能传递预期性能与观测值之间的偏差，例如 $P(t)$ 或 $P(T)$。血清素可能有助于长时间尺度调节和耐心（Miyazaki et al., 2014, 2018; Taira and Sharpe, 2025），并可能有助于估计 $P(T)$，以及抑制过早的努力分配。研究还发现，血清素可能通过降低感知努力成本，帮助个体克服费力行动（Meyniel et al., 2016），这表明它可能在调节有效成本参数 β 方面发挥作用。去甲肾上腺素（NE）及蓝斑核（LC）的活动编码努力成本，并动员认知控制（Varazzani et al., 2015; Aston-Jones and Cohen, 2005; Poe et al., 2020）；而作为 LC–NE 活动指标的瞳孔扩张则反映主观努力（Gilzenrat et al., 2010）。最后，背侧前扣带皮层整合奖励、成本和时间信息，以指导控制分配（Shenhav 等，2013），因而可能实现与确定最优学习率计划 $\mu^*(t)$相似的计算。基于这些实验观察，研究者在元学习或持续学习框架下，赋予神经调质更广泛的计算作用。人们认为，这些神经调质不仅参与学习本身，也参与学会学习，通过调节注意、感知努力、时间折扣及其他在任务学习外循环中起作用的变量来实现这一点（Doya, 2002; Lee et al., 2024）。我们的研究遵循类似的概念框架：学习率控制针对元目标进行优化（最大化时间积分奖励，并在多个学习回合中提高 $P(T)$估计的准确性），而非针对特定任务的即时性能进行优化。

我们认为，将这一框架扩展至其他学习系统，是一个很有前景的方向，有望使我们的抽象表述更接近生物学上合理的架构。这些系统可以采用梯度流以外的规则，例如赫布学习（Flesch 等，2022）、反馈对齐（Lillicrap and Kording, 2019）或高效编码原理（Song 等，2024）。此外，斯特鲁普网络等认知控制模型（Cohen 等，1990；Musslick 等，2020）会同时对多个神经元施加控制，在某些情况下，控制信号可能高度局部化（Aston-Jones and Cohen，2005；Shenhav 等，2013）。如前所述，目前我们的控制信号是一个标量，可在我们特定的设置中求解；该设置以学习率控制为核心，并将梯度流作为可塑性规则，但原则上仍可将其推广至广泛的学习系统。

由于我们的框架较为抽象，直接将其应用于具体的实验设置面临诸多挑战。要评估动物是否会实施类似于本文提出的、用于控制学习率的过程，需要精心设计心理学或神经科学实验，追踪相关变量，例如关于最优情况下最终性能的信念或估计、对任务表现的持续监测，以及对努力程度的主观感知。一个更为开放的问题是，这类最终性能估计可能如何形成。情景记忆是一种可能的机制，但基于模型的估计，或关于努力调节的学习动力学的内部模型等其他机制，也可能发挥作用。另一种可能是，学习率调度以基于模型的方式进行，借助一种动力系统来预测性能，以及学习率对该系统的影响（Pinon 等，2022；Dong 等，2025）。

范围与局限

我们的框架在学习过程中优化累积表现，同时对学习率施加明确的成本；这与标准机器学习实践形成对比，后者将学习率视为不产生成本的超参数，并主要以收敛时的表现作为衡量指标（Kashyap，2023；Choi 等，2020；Schmidt 等，2021）。因此，该框架并非旨在为常规深度学习优化提供指导，而是规范性地说明：当学习本身需要付出成本时，应如何调节学习速度。

从形式上看，这一问题可视为一项最优控制任务：学习率调节基于梯度的学习动力学的速度，而求解方案则通过求解哈密顿–雅可比–贝尔曼方程获得。尽管这一结构与经典控制方法（如线性二次调节器的表述形式）相似（Augustine，2023；Chacko 等，2024），但这些方法依赖线性动力学和二次成本，而对于动力学非线性地依赖自身表现的学习系统，这些条件并不成立。其他控制理论工具（如卡尔曼滤波）则在噪声条件下优化瞬时状态估计误差，所针对的是不同的目标（Piray 和 Daw，2021b；Si 等，2025）。

该框架在概念上也与强化学习中的价值函数，以及通过外循环优化超参数的元学习方法有关（Shenhav 等，2013；Carrasco-Davis 等，2024）。然而，与从非平稳经验中学习价值函数的强化学习不同，此处的价值函数是根据已知的学习动力学直接计算得出的；因此，若要将此方法扩展到强化学习智能体，就需要对策略学习过程中的价值函数动力学进行可处理的描述。

最后，学习率被视为一个连续控制变量，而生物系统可能受到模型未能涵盖的硬约束。尽管目前尚无实证证据明确控制学习所付出的成本的具体形式，但将这些成本明确纳入考量，能够以一种有原则的方式刻画学习速度、努力和性能之间的权衡，而不是对生物实现机制作出解释。

结束语

我们的研究包含若干部分：用于优化学习的规范性框架、基于第一性原理对其解的数学推导、基于仿真的验证，以及与认知控制和自我调节学习理论的联系。我们认为，通过规范性目标将最优学习率调度形式化，为理解动物如何在整个生命周期尺度上调节自身学习提供了一条有前景的途径。决定何时投入多少努力以提升自身性能的能力，是高级认知的核心，它塑造着动机、坚持性以及在变化环境中的适应能力。与此同时，我们的框架也有助于理解如何将这些原则推广到生物学习者之外。使智能体能够高效分配努力的相同机制，也可以为人工系统的设计提供启示，使其能够延长自身的学习过程，并将类似动机的控制机制融入改进本身的动力学之中。

4 方法

4.1 推导闭环解

最优学习率 $\mu^*(t)$ 是公式 3 右侧（RHS）的最大化值。事实证明，RHS 中的表达式是控制理论和连续强化学习中一个经过深入研究的对象，称为最优价值函数 $V(t=0,\ \mathbf{w}_0)$（Bellman，1958年）。对于任意给定的试次起始时间 t和任意初始权重向量 $\mathbf{w}$，该函数都定义了从 $(t,\ \mathbf{w})$ 开始，在最优控制及相应权重动力学下所获得的累积奖励：

$$V(t,\mathbf{w}) = \max_{t\le\mu(\tau)\le T} \int_t^T d\tau\ \gamma^{\tau-t}\ r(\tau) \tag{10}$$

其中，奖励率为 $r(\tau) = P(\mathbf{w}(\tau)) - C(\mu(\tau))$。通过运用贝尔曼最优性原理并对由此得到的方程进行推导，我们可以得出 Hamilton-Jacobi-Bellman 方程。该方程是贝尔曼方程的连续形式，用于描述价值函数（Yong 和 Zhou，1999；Bellman，1958年）：

$$\log\gamma\ V(t,\mathbf{w}) + \frac{\partial V}{\partial t} + \max_{\mu}\left\{ r(t) + \frac{\partial V}{\partial \mathbf{w}}^{\mathrm{T}} \frac{d\mathbf{w}}{dt} \right\} = 0 \tag{11}$$

让我们解释一下这个方程背后的直觉。前两项为负，大致对应于由于折扣因子 $\log\ \gamma\ V(t,\ \mathbf{w})$以及时间流逝 $\frac{\partial V}{\partial t}$ 导致的价值函数下降（即使状态没有变化，时间流逝也会减少可用价值）。后两项说明，通过施加控制μ，此刻可以获得多少奖励 $r(t)$；而 $\frac{\partial V}{\partial\ \mathbf{w}}^{\mathrm{T}}\frac{d\mathbf{w}}{dt}$ 则说明，施加该控制并根据由此产生的 $\frac{d\mathbf{w}}{dt}$ 更新状态后，价值将如何变化（变好或变差）。整个方程表明，对于最优价值函数，因奖励折扣和时间流逝而损失的价值，与因获得奖励以及将状态转移到更有价值区域而增加的价值之间存在平衡。

因此，最优控制选择就是使即时奖励与该控制所导致的价值变化之和最大化的控制。这个最大化项也称为哈密顿量（Yong and Zhou, 1999）；将方程 2 和 1 中的内部奖励率及学习动力学表达式代入后，我们还可以进一步将其具体表示为：

$$H(\mu,\mathbf{w},t) = P(\mathbf{w}) - C(\mu) + \mu\frac{\partial V}{\partial \mathbf{w}}^{\mathrm{T}}\frac{dP}{d\mathbf{w}} \tag{12}$$

最大化此类哈密顿量，得到的控制表达式仅取决于价值函数；但要找到价值函数，就必须完整求解 Hamilton-Jacobi-Bellman 方程（方程 11）。该方程是一个非线性偏微分方程，一般形式下没有闭式解。通常，求解它的唯一方法是采用数值方法，或进行冗长的解析近似（Beard et al., 1998；Saberi Nik et al., 2012；Ganjefar 和 Rezaei，2016）。然而，在某些简化条件下，可以求得精确的解析解，下面将对此进行说明。

我们可以用哈密顿量的最大化值来重写哈密顿量：

$$\frac{dH}{d\mu} = -\frac{dC}{d\mu} + \frac{\partial V}{\partial \mathbf{w}}^{\mathrm{T}}\frac{dP}{d\mathbf{w}} = 0 \tag{13}$$

$$H(\mu^*,\mathbf{w},t) = P(\mathbf{w}) - C(\mu^*) + \mu^* C'(\mu^*) \tag{14}$$

其中，$C'(\cdot)$ 表示其导数。代回方程 11 可得：

$$\log\gamma\ V(t,\mathbf{w}) + \frac{\partial V}{\partial t} + P(\mathbf{w}) - C(\mu^*) + \mu^* C'(\mu^*) = 0 \tag{15}$$

接着我们注意到，梯度流学习问题本身并不取决于时间，而只取决于初始权重 $\mathbf{w}_0$，、损失以及学习率的变化计划。这使得价值函数可以采用可分离形式：

$$V(t,\mathbf{w}(t)) = \Phi(\mathbf{w}) + \lambda(T-t) \tag{16}$$

$$V(T,\mathbf{w}(T)) = \Phi(\mathbf{w}(T)) = 0 \tag{17}$$

最后，我们考虑不含折扣奖励的情况，即 $\gamma = 1$，此时 11 中的 $\log\gamma V$ 项被移除，方程也大为简化：

$$-\lambda P(\mathbf{w}) - C(\mu^*) + \mu^* C'(\mu^*) = 0 \tag{18}$$

$$\Rightarrow \mu^* C'(\mu^*) - C(\mu^*) = \lambda - P(\mathbf{w}) \tag{19}$$

系数 λ 可由边界条件确定：回合结束时的学习率应使学习成本最小，因为任何更高的学习率只会增加成本，却不会增加奖励。因此，区间结束时的学习率满足 $\mu^*(T) = \arg\min\ C(\mu)$。上述方程给出了精确结果

最优学习率的结果取决于成本 $C(\mu)$ 的形式，它要么具有闭式解析解，要么可以通过数值方法求得。在本文中，我们考虑了 $C(\mu) = \beta\mu^2$，因此方程简化为$\beta\mu^{*2} = P(\mathbf{w}(T)) - P(\mathbf{w}(t))$，从而得到最终表达式：

$$\mu^* = \sqrt{\frac{1}{\beta}(P(\mathbf{w}(T)) - P(\mathbf{w}(t)))} \tag{20}$$

4.2 通过情景记忆估计最终性能

我们利用先前学习回合中的信息，为新任务在一个学习回合期间构建性能估计器。每条存储的记忆对应一条过去的学习轨迹，记为 $\mathcal{T}_h(t)$，其由随时间变化的性能定义：

$$\mathcal{T}_h(t) = \{P_h(0), P_h(dt), \ldots, P_h(t)\}. \tag{21}$$

面对新任务时，智能体会将其当前轨迹 $\mathcal{T}_o(t)$ 与从记忆中检索出的 H 先前存储的轨迹$\{\mathcal{T}_h(t)\}_{h\in H}$ 进行比较。基于相似度的权重 $\{w_h(t)\}_{h\in H}$ 通过高斯核计算得出：

$$k(\mathcal{T}_o(t), \mathcal{T}_h(t)) = \exp\left(-\frac{1}{2N_t\sigma_k^2}\sum_{i=0}^{N_t}\rho^{dt\cdot(N_t-i)}\left[P_o(dt\cdot i) - P_h(dt\cdot i)\right]^2\right), \tag{22}$$

其中， N_t 表示截至时间 t 两条轨迹中均存储的离散时间步数量，使得 $N_t \cdot dt = t$； ρ 表示记忆衰减常数，它为轨迹上较近点之间的距离分配更大的权重； σ_k表示核宽度。随后将相似度权重归一化为：

$$w_h(t) = \frac{k(\mathcal{T}_o(t), \mathcal{T}_h(t))}{\sum_{h'=1}^{H} k(\mathcal{T}_o(t), \mathcal{T}_{h'}(t))}. \tag{23}$$

在计算相似度之前，所有轨迹都会通过减去其均值并除以其标准差进行 白化 ，以避免性能尺度差异造成的影响。利用这些相似度权重，可以从记忆中外推时间 $t' > t$ 处的未来性能：

$$\hat{P}(t' \mid t) = \sum_{h=1}^{H} w_h(t)\, P_h(t'), \quad \forall\, t < t' \le T. \tag{24}$$

为了获得计算最优学习率所需的未来性能估计，该外推结果在 $t' = T$ 处求值，得到 $P(T \mid t)$。在观察新任务（$t = 0$）中的任何性能之前，所有相似度权重都设为 $1/H$，从而得到所有已存储记忆的平均轨迹。

4.3 计算感知机的开环解

我们考虑一个权重为 $\mathbf{w}(t) \in \mathbb{R}^d$ 的线性感知机，该感知机通过梯度流以时变学习率$\mu(t)$进行训练，

$$\frac{d\mathbf{w}}{dt} = -\mu(t)\nabla_{\mathbf{w}} L(\mathbf{w}), \tag{25}$$

用于具有平方误差损失的线性回归任务

$$L(\mathbf{w}) = \frac{1}{2}\left\langle (y - \mathbf{w}^\top \mathbf{x})^2 \right\rangle, \tag{26}$$

其中，输入来自各向同性高斯分布 $\mathbf{x} \sim \mathcal{N}(0, \sigma^2 I)$，目标值由教师 $\mathbf{y} = \mathbf{w}^{*\top}\mathbf{x}$生成。在这些假设下，梯度流动力学可以精确求解，得到

$$\mathbf{w}(t) = \mathbf{w}^* + (\mathbf{w}_0 - \mathbf{w}^*)e^{-2\sigma^2\int_0^t \mu(s)\, ds}. \tag{27}$$

定义平方初始化距离 $d^2 = \|\mathbf{w}_0 - \mathbf{w}^*\|^2$ 后，训练轨迹上的损失变为

$$L(t) = \frac{\sigma^2}{2} d^2 \exp\left(-2\sigma^2 \int_0^t \mu(s)\, ds\right). \tag{28}$$

开环目标是最大化累积内部奖励

$$R(\mu) = \int_0^T dt \left[L(t) + \beta \mu(t)^2 \right], \tag{29}$$

其中 $\beta > 0$ 会惩罚较大的学习率。对 R 关于 $\mu(\tau)$ 求泛函导数，可得积分微分最优性条件，

$$\mu(\tau) = \frac{\sigma^2 d^2}{2\beta} \int_\tau^T dt \, \exp\left(-2\sigma^2 \int_0^t \mu(s) \, ds \right). \tag{30}$$

为简化此表达式，我们引入累积学习率

$$M(t) = \int_0^t \mu(s) \, ds, \tag{31}$$

这会将最优性条件转化为一个二阶常微分方程 ，

$$M''(t) + \frac{\sigma^2 d^2}{2\beta} e^{-2\sigma^2 M(t)} = 0. \tag{32}$$

该方程有一个第一积分，

$$\mu(T)^2 - \mu(t)^2 = \frac{1}{\beta} \big(L(T) - L(t) \big), \tag{33}$$

它与前文推导的闭环解一致，并凸显了两种公式之间的一致性。

通过代换 $u(t) = e^{-\sigma^2 M(t)}$，可以求得该方程的闭式解，从而得到最优学习率计划的显式表达式，

$$\mu(t) = \theta \tan\big(\lambda \theta (T - t) \big), \tag{34}$$

其中， θ 由 $t = T$ 处的终端条件自洽地确定。关于 θ 的超越方程通过数值方法求解。

4.4 折扣因子的近似解

本节采用改进同伦摄动法（Saberi Nik 等，2012；Ganjefar 和 Rezaei，2016）求解式 11 中的 Hamilton-Jacobi-Bellman 方程。将内部奖励率、成本函数和梯度流学习动力学代入式 11 并对哈密顿量求最大值后，得到需要求解的非线性偏微分方程:

$$\log \gamma V(\mathbf{w}, t) + \frac{\partial V}{\partial t} + P(\mathbf{w}) + \frac{1}{4\beta} \left(\frac{\partial V}{\partial \mathbf{w}}^\top \frac{dP}{d\mathbf{w}} \right)^2 = 0 \tag{35}$$

$$V(\mathbf{w}(T), T) = 0 \tag{36}$$

$$\mu^*(t) = \frac{1}{2\beta} \frac{\partial V}{\partial \mathbf{w}}^\top \frac{dP}{d\mathbf{w}} \tag{37}$$

该偏微分方程目前没有已知的闭环解，但我们得到了一个解析近似解。我们构造一个含嵌入参数 $p \in [0, 1]$ 的同伦，将满足终端条件 $V_g(\mathbf{w}, T) = 0$、且解析上易于求解的初始猜测 $V_g(\mathbf{w}, t)$，连续变形为完整的非线性 HJB 方程（Saberi Nik 等，2012）。

假设价值函数可以进行形式摄动展开

$$V(\mathbf{w}, t; p) = \sum_{i=0}^{\infty} p^i V_i(\mathbf{w}, t), \tag{38}$$

将其代入同伦并收集关于 p 的同阶项后，可得到一组关于模态 V_i 的线性一阶偏微分方程。每个模态仅取决于低阶解，并在终端条件

$V_i(\mathbf{w}\ T) = 0$ 下求解，从而确保任意有限阶截断都满足边界条件。实践中，级数通常在低阶处截断，并在 $p = 1$ 处求值，以获得原始 HJB 方程的近似解。

为提高 $p = 1$ 处微扰解的收敛性，我们对截断级数应用帕德重求和，这是对（Ganjefar 和 Rezaei，2016）方法的改进。具体而言， p中的幂级数被有理帕德逼近式取代，其系数经过选择，使其与泰勒展开在给定阶数之前相匹配。通过在 $p = 1$ 处计算帕德逼近式，可得到价值函数和最优控制 $\mu^*(t)$ 的最终近似，从而大幅扩大近似能够保持良好性质的参数范围。

对于解析上可处理的情形，我们推导出了相应最优控制的闭式表达式，包括单神经元线性回归，以及在单一任务上训练的单层网络。将同伦帕德方法应用于单神经元模型时，其第一种模式见式 9，模拟结果见图 4d、e。

4.5 实现细节

所有网络均通过显式欧拉步（步长为 dt ）对梯度流进行近似，并采用全批量或小批量梯度下降进行训练。根据方程 5 得出的最优学习率，使用在线性能 $P(\mathbf{w}(t))$ 实现；真实最终性能 $P(\mathbf{w}(T))$ 则取自一个采用数值优化学习率进行训练的回合（MNIST 任务除外）。在此处展示的所有模拟中，性能均由损失函数定义，并经过平移，使性能始终为正 $P(\mathbf{w}) = L(\mathbf{w}(0)) - L(\mathbf{w})$。引入这一平移纯粹是为了便于解释，不会影响结果。

4.5.1 教师-学生

图 2 的 (a) 和 (b) 面板中使用的学生网络是一个两层前馈网络，包含 4 个输入、10 个带偏差的隐藏单元、6 个不带偏差的输出单元，隐藏层的激活函数为 tanh(·)。训练集由 2048 个样本组成，输入为 $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, I_4)$。标签由架构相同的教师网络生成，其权重取自均值为零、采用标准方差缩放的高斯分布。损失为均方误差损失$L(\mathbf{w}) = \frac{1}{N}\sum_{i=1}^{N}\|\mathbf{y}_i - f(\mathbf{x}_i)\|^2$。训练算法为全批量梯度下降，共运行 8000 步，步长按 $dt = 0.01$ 重新缩放。表现最佳的基线学习率为 0.413。学习成本设为 $\beta = 0.5$。

4.5.2 MNIST

图 2 的 (c) 和 (d) 面板中使用的网络是一个两层前馈网络，包含 196 个输入、100 个带偏差的隐藏单元、10 个不带偏差的输出单元，隐藏层激活函数为 softplus(·)，并在最后一层应用 softmax(·)。该网络在 MNIST 图像分类任务上进行训练（Deng，2012）。所有输入图像在输入网络前都按 2 倍比例下采样。损失为交叉熵损失$L(\mathbf{w}) = \sum_{i=1}^{N}\mathbf{y}_i^{\top}\log\ f(\mathbf{x}_i)$。训练采用小批量梯度下降，共运行 3000 步，每个类别的每批包含 100 个样本，并且每次更新均按 $dt = 0.001$ 重新缩放。最佳基准学习率为 0.172。真实最终性能通过求出在公式 5 的闭环表达式中使用时可使累积奖励最大化的 $P(T)$ 值来确定。结果发现 $P(T) = 0.067$。学习成本设为 $\beta = 0.8$。

4.5.3 双高斯判别

在图 2 的面板 (e,f,g,h) 中，为简化起见，所有智能体都是通过线性回归执行双高斯判别任务的单神经元模型。尽管任务很简单，学习率和权重动力学仍与复杂的 MNIST 分类任务相似。任务输出以相同的概率为 $y_i = 1$或 $y_i = -1$ ，输入则从相应的高斯分布 $x_i \sim \mathcal{N}(yd, \sigma^2)$ 中采样，其中 $d = 2, \sigma^2 = 4$。训练采用全批量梯度下降，并将更新按 $dt = 0.001$ 重新缩放，共运行 700 步。学习成本设为 $\beta = 0.5$。

4.5.4 图 3

一个具有 softplus 激活函数和六个隐藏神经元的非线性双层（学生）网络经过训练，将高斯输入 $\mathbf{x}_i \sim \mathcal{N}$（$\mathbf{0}$， I_3）的样本映射到由非线性双层（教师）网络生成的目标值 $y \in \mathbb{R}^2$。训练通过使用反向传播最小化均方误差来进行，并采用随机梯度下降更新、基础学习率 $\alpha = 10.0$，以及每次更新大小为 32 的随机批次。图中所示的 1000 个学习实例中的每一个，都对应于训练一个全新的学生–教师对。教师权重使用标准的 Xavier 初始化方案进行采样，而学生网络权重则从标准差为 0。001 的高斯分布中采样，以便在不同学习实例中诱导学生产生丰富的非线性学习动力学。为了在多个学习实例上初始化元训练，使用基础学习率来

生成初始学习轨迹。对于所有其他学习实例，我们使用公式 5 中闭环表达式的一个修改版本，其中基础学习率更新为

$$\frac{d\mathbf{w}}{dt} = -(\mu(t) + \alpha)\nabla_{\mathbf{w}} L(\mathbf{w}), \tag{39}$$

而最优学习率计划为

$$\mu^* = -\alpha + \sqrt{\frac{1}{\beta}\big(P(\mathbf{w}(T)) - P(\mathbf{w}(t))\big) + \alpha^2}, \tag{40}$$

其成本函数为 $C(\mu) = \beta\mu^2$。该表述用于对不产生成本的被动学习进行建模，对于 $\alpha = 0$ 也可获得定性相似的结果。为应用闭式解，我们使用第 2.2 小节和方法 4.2 中所述的情景记忆方法，估计每个学习实例的 $P(T)$ 值。每个训练实例包含 $N_t = 5000$ 个梯度步，记忆衰减为 $\rho = 0.7$，公式 22 中的核宽度为 $\sigma_k = 0.1$，$dt = 0.001$、$\beta = 0.01$ 和 $\gamma = 1.0$。图（a）展示了使用记忆中的 25 个学习实例时，内循环经过 500 步后预测的未来性能（垂直虚线）。图（b）展示了相同的分析，但预测内循环经过 3000 步后的未来ˆ性能。图（c）展示了对于记忆中不同ˆ数量的学习实例，$P(T \mid t) - P(T)$ 的第 25 和第 75 百分位数。图（d）展示了对于 1000 个学习实例中的每一个，$P(T \mid t) - P(T)$ 随时间变化的 $P(t)$ 标准差。

4.5.5 图4

面板 (a)：输入维度为 4 的线性感知机在教师-学生回归任务上进行训练，教师架构随机初始化且与学生架构匹配。训练共进行 1000 步，并按 $dt = 0.01$ 重新缩放，其中 $\beta = 0.1, \gamma = 1$。面板 (b、c) 展示模拟的解析表达式，其中 (b) 中努力成本 β 在 0.2 到 3.0 之间变化，(c) 中任务难度 d 在 0.1 到 1.0 之间变化。面板 (d、e) 展示线性神经元学习线性回归任务，训练共进行 500 步，并按 dt=0.001 重新缩放。面板 (d) 展示数值优化得到的学习率，面板 (e) 展示根据方程 9 中的表达式得到的学习率。折扣因子 γ 在 0.1 到 0.95 之间变化。努力成本为 $\beta = 5$。面板 (f)：学生-教师任务中，学生是一个具有 4 个输入、4 个隐藏神经元和 6 个输出的两层网络，以较小的权重初始化；教师是一个随机初始化的两层网络，具有 4 个输入、10 个隐藏神经元和 6 个输出，两者均采用 tanh(·) 非线性。训练进行 8000 步，其中 $dt = 0.01$，努力成本为 $\beta = 0.1$。折扣因子 γ 在 0.1 到 1.0 之间变化。

4.5.6 图5

左图改编自 Masís 等人（Masís 等，2023）的图 7b。右图展示了一个线性神经元在 $dt = 0.001$ 下经过 130 步训练的情况，任务为线性回归任务。努力成本为 $\beta = 0.05$。最优学习率由公式 5 得出；次优（较低）的学习速度具有与最优学习速度相同的形状，但缩小至其 0.6 倍。

4.6 数值优化

所有数值优化的学习率调度均采用改编自（Carrasco-Davis 等，2024）的元优化过程获得。通过最大化累积内部奖励 $R(\mu)$，直接优化学习率函数 $\mu(t)$。使用自动微分（通过 JAX 实现）沿完整训练动力学进行求导，从而计算 R 相对于 μ 的梯度。在每次元迭代中，通过函数空间中的梯度上升更新 μ，直至收敛到局部最优解。此过程无需依赖解析近似，即可得到数值优化的开环学习率调度。

致谢

感谢 Javier Masís 提供了有益的讨论。我们感谢以下资助来源：Gatsby 慈善基金会（GAT3850 和 GAT4058）资助 VN、RC、AS 和 PEL；Wellcome 提供的 Sainsbury Wellcome中心核心资助（219627/Z/19/Z）资助 AS；AS 获得 Schmidt Science博学奖。AS 是机器与大脑学习项目的 CIFAR Azrieli全球学者。RC 感谢普林斯顿总统奖学金。

参考文献

Fabrizio Taffoni、Eleonora Tamilia、Valentina Focaroli、Domenico Formica、Luca Ricci、Giovanni Di Pino、Gian- luca Baldassarre、Marco Mirolli、Eugenio Guglielmelli 和 Flavio Keller。内在动机引导的目标导向的行动选择的发展：一项针对儿童的实验。《实验脑研究》，232(7)：2167–2177，2014年7月。ISS N 1432-1106。doi:10.1007/s00221-014-3907-z。URL https://doi.org/10.1007/ s00221-014-3907-z。

Gal Raz 和 Rebecca Saxe。《婴儿期学习是主动的、由内在动机驱动的，并且依赖于前额叶皮层》。《发展心理学年度评论》，第 2 卷（2020 年第 2 卷）：247–268，2020年12月。ISSN 2640-7922。doi:10.1146/annurev-devpsych-121318-084841。URL https://www.annualreviews.org/content/ journals/10.1146/annurev-devpsych-121318-084841。出版方：Annual Reviews。

Tamara van Gog、Vincent Hoogerheide 和 Milou van Harsel。《心理努力在借助问题解决任务促进自我调节学习中的作用》。《教育心理学评论》，32(4)：1055–1072，2020年12月。ISSN 1573-336X。doi:10.1007/s10648-020-09544-y。URL https://doi.org/10.1007/s10648-020-09544-y。

Celeste Kidd、Steven T. Piantadosi 和 Richard N. Aslin。《金发姑娘效应：人类婴儿会将注意力分配给既不过于简单也不过于复杂的视觉序列》。《*PLOS ONE*》，7(5):e36399，2012年5月。ISSN 1932-6203。doi:10.1371/journal.pone.0036399。URL https://journals.plos.org/plosone/article?id=10. 1371/journal.pone.0036399。出版方：公共科学图书馆。

Elizabeth Bonawitz、Ilona Bass 和 Elizabeth Lapidow。《选择学习：幼儿期主动学习与教学中的证据评估》。载于 Megan M. Saylor 和 Patricia A. Ganea 编，《style id='1'>从婴儿期到童年期的主动学习：社会动机、认知与语言机制》，第 213–231 页。施普林格国际出版公司，沙姆，2018年。ISBN 978-3-319-77182-3。doi:10.1007/978-3-319-77182-3_12。URL https://doi.org/ 10.1007/978-3-319-77182-3_12。

Rui Castro、Charles Kalish、Robert Nowak、Ruichen Qian、Tim Rogers 和 Jerry Zhu。《人类主动学习》。载于《神经信息处理系统进展》，第 21 卷。Curran Associates 公司，2008 年。网址：https://papers. nips.cc/paper_ files/paper/2008/hash/fc49306d97602c8ed1be1dfbf0835ead-Abstract.html。

Douglas B. Markant 和 Todd M. Gureckis。选择还是接受更好？通过主动和被动假设检验进行学习。《实验心理学杂志：总论》，143(1):94–122，2014 年。ISSN 1939-2222。doi:10.1037/a0032108。地点：美国；出版方：美国心理学会。

Anna Coenen、Jonathan D. Nelson 和 Todd M. Gureckis。提出关于人类探究心理学的恰当问题：九项尚待解决的挑战。《心理学通报与评论》，26(5):1548–1587，2019 年 10 月。ISSN 1531-5320。doi:10.3758/s13423-018-1470-5。网址：https://doi.org/10.3758/s13423-018-1470-5。

Jonathan G. Tullis 和 Aaron S. Benjamin。《自定步调学习的有效性》。《记忆与语言杂志》，64(2):109–118，2011年2月。ISSN 0749-596X。doi:10.1016/j.jml.2010.11.002。URL https://www. sciencedirect.com/science/article/pii/S0749596X10000999。

Lisa K. Son。元认知控制与间隔效应。《实验心理学杂志：学习、记忆与认知》，36(1):255–262，2010年。ISSN 1939-1285, 0278-7393。doi:10.1037/a0017892。URL https://doi. apa.org/doi/10.1037/a0017892。

Marina Bazhydai、Katherine Twomey 和 Gert Westermann。好奇心与探索。载于 Janette B. Benson 编，《婴儿与幼儿发展百科全书（第二版）》，第370–378页。爱思唯尔，牛津，2020年1月。ISBN 978-0-12-816511-9。doi:10.1016/B978-0-12-809324-5.05804-1。URL https: //www.sciencedirect.com/science/article/pii/B9780128093245058041。

Suzanne E. Hidi 和 K. Ann Renninger。兴趣发展及其与好奇心的关系：所需的神经科学研究。《教育心理学评论》，31(4)：833–852，2019年12月。ISSN 1573-336X。doi:10.1007/s10648- 019-09491-3。URL https://doi.org/10.1007/s10648-019-09491-3。

Philipp Schwartenbeck、Johannes Passecker、Tobias U Hauser、Thomas HB FitzGerald、Martin Kronbichler 和 Karl J Friston。好奇心的计算机制与目标导向探索。*eLife*，8:e41703，2019年5月。ISSN 2050-084X。doi:10.7554/eLife.41703。URL https://doi.org/10.7554/eLife.41703。出版商：eLife Sciences Publications, Ltd。

Andra Geana、Robert C. Wilson、Nathaniel Daw 和 Jonathan D. Cohen。《无聊、信息寻求与探索》。载于认知科学学会第*38*届年会论文集，*CogSci 2016*，第1751–1756页。认知科学学会，2016年。网址：https://collaborate.princeton.edu/en/publications/ boredom-information-seeking-and-exploration。

Mayank Agrawal、Marcelo G. Mattar、Jonathan D. Cohen 和 Nathaniel D. Daw。机会成本的时间动态：认知疲劳与无聊的规范性解释。《心理学评论》，129(3):564–585，2022年4月。ISSN 1939-1471。doi:10.1037/rev0000309。Adrien F. Baranes、Pierre-Yves Oudeyer 和 Jacqueline Gottlieb。任务难度、新颖性和搜索空间大小对内在动机驱动的探索的影响。《神经科学前沿》，8，2014年10月。ISSN 1662-453X。doi:10.3389/fnins.2014.00317。网址：https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2014.00317/full。出版方：Frontiers。Eric Schulz、Rahul Bhui、Bradley C. Love、Bastien Brier、Michael T. Todd 和 Samuel J. Gershman。在现实世界消费者选择中进行结构化的不确定性驱动探索。《美国国家科学院院刊》，116(28):13903–13908，2019年7月。doi:10.1073/pnas.1821028116。网址：https://www.pnas.org/doi/10.1073/ pnas.1821028116。出版方：《美国国家科学院院刊》。Robert C. Wilson、Amitai Shenhav、Mark Straccia 和 Jonathan D. Cohen。最优学习的百分之八十五规则。《*Nature Communications*》，10(1):4646，2019年11月。ISSN 2041-1723。doi:10.1038/s41467-019-12552-4。网址：https://www.nature.com/articles/s41467-019-12552-4。出版方：自然出版集团。Alexandr Ten、Pramod Kaushik、Pierre-Yves Oudeyer 和 Jacqueline Gottlieb。人类在好奇心驱动的探索中监控学习进展。《*Nature Communications*》，12(1):5972，2021年10月。ISSN 2041-1723。doi:10.1038/s41467-021-26196-w。网址：https://www.nature.com/articles/s41467-021-26196-w。出版方：自然出版集团。Laura S. Cubit、Rebecca Canale、Rebecca Handsman、Celeste Kidd 和 Loisa Bennetto。幼儿对中等可预测性的视觉注意偏好。《儿童发展》，92(2):691–703，2021年3月。ISSN 1467-8624。doi:10.1111/cdev.13536。Lisa K. Son 和 Janet Metcalfe。学习时间分配中的元认知与控制策略。《实验心理学杂志：学习、记忆与认知》，26(1):204–221，2000年。ISSN 1939-1285。doi:10.1037/0278-7393.26.1.204。地点：美国。出版方：美国心理学会。Martina de Eccher、Roger Mundry 和 Nivedita Mani。儿童由主观不确定性驱动的抽样行为。《英国皇家学会开放科学》，11(4):231283，2024年4月。doi:10.1098/rsos.231283。网址：https: //royalsocietypublishing.org/doi/full/10.1098/rsos.231283。出版方：英国皇家学会。Javier Masís、Travis Chapman、Juliana Y Rhee、David D Cox 和 Andrew M Saxe。在知觉决策过程中策略性地管理学习。*eLife*，12:e64978，2023年2月。ISSN 2050-084X。doi:10.7554/eLife.64978。网址：https://doi.org/10.7554/eLife.64978。出版方：eLife Sciences Publications, Ltd. Javier Alejandro Masís Obando、Sebastian Musslick 和 Jonathan D. Cohen。学习预期塑造认知控制分配。《美国国家科学院院刊》，122(44):e2416720122，2025年11月。doi:10.1073/pnas.2416720122。网址：https://www.pnas.org/doi/10.1073/pnas.2416720122。出版方：《美国国家科学院院刊》。Wyatt B. Potter、Kenneth J. O'Riordan、David Barnett、Susan M. K. Osting、Matthew Wagoner、Corinna Burger 和 Avtar Roopra。能量传感器AMPK对神经元可塑性的代谢调节。《*PLOS ONE*》，5(2):e8996，2010年2月。ISSN 1932-6203。doi:10.1371/journal.pone.0008996。Michael A. Smith、Leigh M. Riby、J. Anke M. van Eekelen 和 Jonathan K. Foster。葡萄糖对人类记忆的增强作用：关于葡萄糖促进记忆效应的综合研究综述。《神经科学与生物行为评论》，35(3):770–783，2011年1月。ISSN 1873-7528。doi:10.1016/j.neubiorev.2010.09.008。Sebastian Klug、Godber M. Godbersen、Lucas Rischka、Wolfgang Wadsak、Verena Pichler、Manfred Klöbl、Marcus Hacker、Rupert Lanzenberger 和 Andreas Hahn。学习诱发代谢需求与功能性脑网络的协同神经元可塑性。《通讯·生物学》，5(1):428，2022年5月。ISSN 2399-3642。doi:10.1038/s42003-022-03362-4。网址：https://www.nature.com/articles/s42003-022-03362-4。出版方：自然出版集团。Frederic Mery 和 Tadeusz J. Kawecki。黑腹果蝇学习的运行成本。黑腹果蝇。《动物行为》，68(3):589–598，2004年9月。ISSN 0003-3472。doi:10.1016/j.anbehav.2003.12.005。网址：https://www.sciencedirect.com/science/article/pii/S0003347204001708。Emilie C. Snell-Rood、Goggy Davidowitz 和 Daniel R. Papaj。蝴蝶学习的繁殖权衡。《行为生态学》，22(2):291–302，2011年3月。ISSN 1045-2249。doi:10.1093/beheco/arq169。网址：https: //doi.org/10.1093/beheco/arq169。Frederic Mery 和 Tadeusz J. Kawecki。黑腹果蝇长期记忆的代价。《科学》，308(5725):1148–1148，2005年5月。doi:10.1126/science.1111331。网址：https://www.science.org/doi/10.1126/science.1111331。出版方：美国科学促进会。

Javier Alejandro Masís、Sebastian Musslick 和 Jonathan Cohen。学习的价值与认知控制分配。认知科学学会年会论文集，2021年。URL https: //escholarship.org/uc/item/7w0223v0。Javier Alejandro Masis、Sebastian Musslick 和 Jonathan D. Cohen。学习预期塑造认知控制分配，2024年8月。URL https://osf.io/d2cbg_v1。Rodrigo Carrasco-Davis、Javier Masís 和 Andrew M. Saxe。通过神经网络中的价值最大化实现元学习策略，2024年7月。URL http://arxiv.org/abs/2310.19919。arXiv:2310.19919 [cs]。Wouter Kool 和 Matthew Botvinick。心理劳动。《自然·人类行为》，2(12):899–908，2018年12月。ISSN 2397-3374。doi:10.1038/s41562-018-0401-9。URL https://www.nature.com/articles/ s41562-018-0401-9。出版方：自然出版集团。Robert Kurzban、Angela Duckworth、Joseph W. Kable 和 Justus Myers。主观努力与任务表现的机会成本模型。《行为与脑科学》，36(6):10.1017/S0140525X12003196，2013年12月。ISSN 0140-525X。doi:10.1017/S0140525X12003196。URL https://www.ncbi.nlm.nih.gov/pmc/articles/ PMC3856320/。Amitai Shenhav、Matthew M. Botvinick 和 Jonathan D. Cohen。控制的预期价值：前扣带皮层功能的整合理论。《神经元》，79(2):217–240，2013年7月。ISSN 0896- 6273。doi:10.1016/j.neuron.2013.07.007。URL https://www.sciencedirect.com/science/article/pii/S0896627313006077。Amitai Shenhav、Sebastian Musslick、Falk Lieder、Wouter Kool、Thomas L. Griffiths、Jonathan D. Cohen 和 Matthew M. Botvinick。迈向心理努力的理性机制解释。《神经科学年度评论》，40:99–124，2017年7月。ISSN 1545-4126。doi:10.1146/annurev-neuro-072116-031526。Huw Jarvis、Isabelle Stevenson、Amy Q. Huynh、Emily Babbage、James Coxon 和 Trevor T.-J. Chong。努力强化学习。《*Journal of Neuroscience*》，42(40):7648–7658，2022年10月。ISSN 0270-6474, 1529-2401。doi:10.1523/JNEUROSCI.2223-21.2022。URL https://www.jneurosci.org/content/42/40/7648。出版方：神经科学学会；栏目：研究文章。Ceyda Sayalı、Emma Heling 和 Roshan Cools。学习进展中介认知努力与任务投入之间的联系。《认知》，236:105418，2023年7月。ISSN 0010-0277。doi:10.1016/j.cognition.2023.105418。URL https://www.sciencedirect.com/science/article/pii/S0010027723000525。Timothy Hospedales、Antreas Antoniou、Paul Micaelli 和 Amos Storkey。神经网络中的元学习：综述，2020年11月。URL http://arxiv.org/abs/2004.05439。arXiv:2004.05439 [cs, stat]。Ji Zhang、Jingkuan Song、Yazhou Yao 和 Lianli Gao。基于课程的元学习。载于第*29*届 *ACM* 多媒体国际会议论文集，MM '21，第1838–1846页，美国纽约州纽约市，2021年10月。计算机协会。ISBN 978-1-4503-8651-7。doi:10.1145/3474085.3475335。URL https://doi.org/10.1145/3474085.3475335。Emmanouil Stergiadis、Priyanka Agrawal 和 Oliver Squire。面向小样本分类的课程元学习，2021年12月。URL http://arxiv.org/abs/2112.02913。arXiv:2112.02913 [cs]。Petru Soviany、Radu Tudor Ionescu、Paolo Rota 和 Nicu Sebe。课程学习：综述，2022年4月。URL http://arxiv.org/abs/2101.10382。arXiv:2101.10382 [cs]。Chelsea Finn、Pieter Abbeel 和 Sergey Levine。用于深度网络快速适应的模型无关元学习，2017年7月。URL http://arxiv.org/abs/1703.03400。arXiv:1703.03400 [cs]。German I. Parisi、Ronald Kemker、Jose L. Part、Christopher Kanan 和 Stefan Wermter。神经网络持续终身学习：综述。《神经网络》，113:54–71，2019年5月。ISSN 08936080。doi:10.1016/j.neunet.2019.01.012。URL http://arxiv.org/abs/1802.07569。arXiv:1802.07569 [cs]。Liyuan Wang、Xingxing Zhang、Hang Su 和 Jun Zhu。持续学习综合综述：理论、方法与应用。《*IEEE* 模式分析与机器智能汇刊》，46(8):5362–5383，2024年8月。ISSN 1939-3539。doi:10.1109/TPAMI.2024.3367329。URL https://ieeexplore.ieee.org/ document/10444954/?arnumber=10444954。会议名称：《IEEE 模式分析与机器智能汇刊》。Luca Franceschi、Paolo Frasconi、Saverio Salzo、Riccardo Grazzi 和 Massimilano Pontil。用于超参数优化与元学习的双层规划，2018年7月。URL http://arxiv.org/abs/1806.04910。arXiv:1806.04910 [cs, stat]。Sungyong Baik、Myungsub Choi、Janghoon Choi、Heewon Kim 和 Kyoung Mu Lee。自适应超参数元学习。载于神经信息处理系统进展，第33卷，第

第20755–20765页。Curran Associates 公司，2020年。URL https://proceedings.neurips.cc//paper/2020/hash/ ee89223a2b625b5152132ed77abbcc79-Abstract.html。Kensuke Nakamura、Bilel Derbel、Kyoung-Jae Won 和 Byung-Woo Hong。深度神经网络的学习率退火方法。*Electronics*，10(16):2029，2021年1月。ISSN 2079-9292。doi:10.3390/electronics10162029。URL https://www.mdpi.com/2079-9292/10/16/2029。期号：16；出版方：多学科数字出版研究所。Francesco Mori、Stefano Sarao Mannelli 和 Francesca Mignacco。基于统计物理与控制理论的持续学习最优方案，2025年5月。URL http://arxiv.org/abs/2409.18061。arXiv:2409.18061[cs]。Francesca Mignacco 和 Francesco Mori。最优学习的统计物理框架，2025年7月。URL http://arxiv.org/abs/2507.07907。arXiv:2507.07907 [cond-mat]。Andrew M. Saxe、James L. McClelland 和 Surya Ganguli。深度神经网络语义发展的数学理论。《美国国家科学院院刊》，116(23):11537–11546，2019年6月。ISSN 0027- 8424，1091-6490。doi:10.1073/pnas.1820226116。URL http://arxiv.org/abs/1810.10531。arXiv:1810.10531[cs, q-bio, stat]。Etienne Boursier、Loucas Pillaud-Vivien 和 Nicolas Flammarion。正交输入下浅层 ReLU 网络平方损失的梯度流动力学。神经信息处理系统进展，35: 20105–20118，2022年12月。URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/7eeb9af3eb1f48e29c05e8dd3342b286-Abstract-Conference.html。Taiki Miyagawa。迈向深度神经网络的运动方程：连续时间梯度下降与离散化误差分析，2023年2月。URL http://arxiv.org/abs/2210.15898。arXiv:2210.15898[cs]。Dimitri P. Bertsekas。动态规划与最优控制。Athena Scientific，第2版，2000年10月。ISBN 978-1-886529-09-0。U. Boscain、M. Sigalotti 和 D. Sugny。量子最优控制庞特里亚金极大值原理导论。*PRXQuantum*，2(3):030203，2021年9月。doi:10.1103/PRXQuantum.2.030203。URL https: // link.aps.org/doi/10.1103/PRXQuantum.2.030203。出版方：美国物理学会。Sanjay Joseph Chacko、Neeraj P.c. 和 Rajesh Joseph Abraham。线性二次调节控制器优化：比较研究。控制与优化研究成果，14:100387，2024年3月。ISSN 2666-7207。doi:10.1016/j.rico.2024.100387。URL https://www.sciencedirect.com/science/article/pii/S2666720724000171。Payam Piray 和 Nathaniel D. Daw。规划、网格场与认知控制中的线性强化学习。《*Nature Communications*》，12(1):4942，2021a年8月。ISSN 2041-1723。doi:10.1038/s41467-021-25123-3。URL https://www.nature.com/articles/s41467-021-25123-3。期号：1；出版方：自然出版集团。Rafal Bogacz、Eric Brown、Jeff Moehlis、Philip Holmes 和 Jonathan D. Cohen。最优决策的物理学：双选强迫选择任务中表现模型的形式化分析。《心理学评论》，113(4):700–765，2006。ISSN 1939-1471。doi:10.1037/0033-295X.113.4.700。出版地：美国；出版方：美国心理学会。Richard Bellman。动态规划与随机控制过程。信息与控制，1(3):228–239，1958年9月。ISSN 0019-9958。doi:10.1016/S0019-9958(58)80003-0。URL https://www.sciencedirect.com/science/article/pii/S0019995858800030。Sebastian Goldt、Madhu Advani、Andrew M Saxe、Florent Krzakala 和 Lenka Zdeborová。教师-学生设置中双层神经网络随机梯度下降的动力学。载于神经信息处理系统进展，第32卷。Curran Associates 公司，2019年。URL https://proceedings.neurips.cc/ paper_files/paper/2019/hash/cab070d53bd0d200746fb852a922064a-Abstract.html。Sebastian Lee、Stefano Sarao Mannelli、Claudia Clopath、Sebastian Goldt 和 Andrew Saxe。灾难性遗忘的马斯洛之锤：节点复用与节点激活，2022年5月。URL http://arxiv.org/abs/2205. 09029。arXiv:2205.09029。Li Deng。用于机器学习研究的 MNIST 手写数字图像数据库 [网络精选]。*IEEE* 信号处理杂志，29(6):141–142，2012年11月。ISSN 1558-0792。doi:10.1109/MSP.2012.2211477。URL https://ieeexplore.ieee.org/document/6296535/citations。Heinz Horner。受约束感知机中的学习与泛化动力学。*Physica A*：统计力学及其应用，200(1):552–562，1993年11月。ISSN 0378-4371。doi:10.1016/0378-4371(93)90560-Q。URL https://www.sciencedirect.com/science/article/pii/037843719390560Q。

David Saad 和 Sara A. Solla。软委员会机器中的在线学习。《物理评论*E*》，52(4):4225–4243，1995年10月。doi:10.1103/PhysRevE.52.4225。URL https://link.aps.org/doi/10.1103/PhysRevE.52. 4225。出版方：美国物理学会。

Soheil Ganjefar 和 Sara Rezaei。使用 Padé 逼近的最优控制问题改进同伦摄动法。《应用数学建模》，40(15):7062–7081，2016年8月。ISSN 0307-904X。doi:10.1016/j.apm.2016.02.039。URL https://www.sciencedirect.com/science/article/pii/ S0307904X16301147。

A. Barari、M. Omidvar、Abdoul R. Ghotbi 和 D. D. Ganji。将同伦摄动法和变分迭代法应用于非线性振子微分方程。《应用数学学报》，104 (2):161–171，2008年11月。ISSN 1572-9036。doi:10.1007/s10440-008-9248-9。URL https://doi.org/10. 1007/s10440-008-9248-9。

O. Abdulaziz、I. Hashim 和 S. Momani。《同伦摄动法在分数阶初值问题中的应用》。《计算与应用数学杂志》，216(2):574–584，2008年7月。ISSN 0377- 0427。doi:10.1016/j.cam.2007.06.010。URLhttps://www.sciencedirect.com/science/article/pii/ S0377042707003081。

Eerdun Buhe、Muhammad Rafiullah、Dure Jabeen 和 Naveed Anjum。《同伦摄动法在求解森林资源枯竭非线性数学模型中的应用》。《物理前沿》，11，2023年10月。ISSN 2296-424X。doi:10.3389/fphy.2023.1246884。URL https://www.frontiersin.org/journals/physics/articles/10.3389/fphy.2023.1246884/full。出版方：Frontiers。Lukas Braun、Clémentine Carla Juliette Dominé、James E. Fitzgerald 和 Andrew M. Saxe。《具有先验知识的深度线性网络的精确学习动力学》。载于《神经信息处理系统进展》，2022年10月。URL https://openreview.net/forum?id=lJx2vng-KiC。Albert Bandura。《自我效能感：迈向行为改变的统一理论》。《心理学评论》，84(2):191– 215，1977年。ISSN 1939-1471。doi:10.1037/0033-295X.84.2.191。地点：美国；出版方：美国心理学会。Barry J. Zimmerman。《实现自我调节》。载于《自我调节手册》，第13–39页。爱思唯尔，2000年。ISBN 978-0-12-109890-2。doi:10.1016/B978-012109890-2/50031-7。URL https://linkinghub.elsevier.com/retrieve/pii/B9780121098902500317。Kenji Doya。《元学习与神经调节》。《神经网络》，15(4):495–506，2002年6月。ISSN 0893- 6080。doi:10.1016/S0893-6080(02)00044-8。URL https://www.sciencedirect.com/science/article/ pii/S0893608002000448。Gary Aston-Jones 和 Jonathan D. Cohen。《蓝斑-去甲肾上腺素功能的整合理论：适应性增益与最优表现》。《神经科学年度评论》，28:403–450，2005年。ISSN 0147-006X。doi:10.1146/annurev.neuro.28.061604.135709。Dami Choi、Christopher J. Shallue、Zachary Nado、Jaehoon Lee、Chris J. Maddison 和 George E. Dahl。《深度学习优化器的实证比较》，2020年6月。URL http://arxiv.org/abs/1910.05446。arXiv:1910.05446 [cs]。Robin M. Schmidt、Frank Schneider 和 Philipp Hennig。《穿越拥挤山谷——深度学习优化器的基准测试》，2021年8月。URL http://arxiv.org/abs/2007.01547。arXiv:2007.01547 [cs]。Rohan Kashyap。《深度学习优化器综述——一阶和二阶方法》，2023年9月。URL http://arxiv.org/abs/2211.15596。arXiv:2211.15596 [cs]。Frank Pajares 和 M. David Miller。《自我效能感和自我概念信念在数学问题解决中的作用：路径分析》。《教育心理学杂志》，86(2):193–203，1994年。ISSN 1939-2176。doi:10.1037/0022- 0663.86.2.193。地点：美国；出版方：美国心理学会。Barry J. Zimmerman 和 Dale H. Schunk。《自我调节学习与表现：导论与概述》。载于《学习与表现的自我调节手册》。劳特利奇，2011年。ISBN 978-0-203-83901-0。页数：12。Hakki Kontas 和 Bahadir Özcan。《以自我效能来源及来自家庭、同伴和教师的成就期望解释中学生的数学素养》。《国际教育与读写研究杂志》，10(1):198–206，2022年。URL https://eric.ed.gov/?id=EJ1329200。出版方：澳大利亚国际学术中心有限公司；ERIC编号：EJ1329200。

Lisa K. Son 和 Rajiv Sethi。《元认知控制与最优学习》。《认知科学》，30(4):759–774，2006。ISSN 1551-6709。doi:10.1207/s15516709cog0000_74。URL

https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog0000_74。_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog0000_74。Wouter Kool、Joseph T. McGuire、Zev B. Rosen 和 Matthew M. Botvinick。决策与认知需求回避。《实验心理学杂志：总论》，139(4):665–682，2010。ISSN 1939-2222。doi:10.1037/a0020198。地点：美国；出版商：美国心理学会。John Sweller。问题解决过程中的认知负荷：对学习的影响。《认知科学》，12(2):257–285，1988。ISSN 1551-6709。doi:10.1207/s15516709cog1202_4。地点：荷兰；出版商：爱思唯尔科学。Jeroen J. G. van Merriënboer 和 John Sweller。认知负荷理论与复杂学习：最新进展与未来方向。《教育心理学评论》，17(2):147–177，2005。ISSN 1573-336X。doi:10.1007/s10648- 005-3951-0。地点：德国；出版商：施普林格。Albert Bandura 和 Dale H. Schunk。通过近端自我激励培养能力、自我效能感和内在兴趣。《人格与社会心理学杂志》，41(3):586–598，1981。ISSN 1939-1315。doi:10.1037/0022- 3514.41.3.586。地点：美国；出版商：美国心理学会。Dale H. Schunk。自我调节学习中的目标设定与自我效能感。《教育心理学家》，25(1):71–86，1990。ISSN 1532-6985。doi:10.1207/s15326985ep2501_6。地点：美国；出版商：Lawrence Erlbaum。Jennifer Stock 和 Daniel Cervone。近端目标设定与自我调节过程。《认知治疗与研究》，14(5):483–498，1990年10月。ISSN 1573-2819。doi:10.1007/BF01172969。URL https://doi.org/ 10.1007/BF01172969。Jennifer King Rice。解释从初中升入高中对数学和科学学生表现的负面影响。《教育管理季刊》，37(3):372–400，2001年8月。ISSN 0013-161X。doi:10.1177/00131610121969352。URL https://doi.org/10.1177/00131610121969352。出版商：SAGE Publications Inc. Ruth Curran Neild。升入高中期间脱离教育轨道：我们已知的情况及可采取的措施。《儿童的未来》，19(1):53–76，2009。ISSN 1054-8289。URL https://www.jstor.org/stable/27795035。出版商：普林斯顿大学。Danielle Evans、Giulia A. Borriello 和 Andy P. Field。升入中等教育的学业与心理影响综述。《心理学前沿》，9，2018年8月。ISSN 1664-1078。doi:10.3389/fpsyg.2018.01482。URL https://www.frontiersin.org/journals/psychology/articles/10. 3389/fpsyg.2018.01482/full。出版商：Frontiers。John R. Hills。转学冲击：初级学院转学生的学业表现。《实验教育学杂志》，33(3):201–215，1965年3月。ISSN 0022-0973, 1940-0683。doi:10.1080/00220973.1965.11010875。URL http://www.tandfonline.com/doi/abs/10.1080/00220973.1965.11010875。Shanyan Lin、Sofia Mastrokoukou、Claudio Longobardi、Paolo Bozzato、Francesca Giovanna Maria Gastaldi 和 Martina Berchiatti。学生升入高等教育：自我效能感、调节策略和学业成就的作用。《高等教育季刊》，77(1):121–137，2023。ISSN 1468- 2273。doi:10.1111/hequ.12374。URL https://onlinelibrary.wiley.com/doi/abs/10.1111/hequ.12374。_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/hequ.12374。Katariina Salmela-Aro、Noona Kiuru 和 Jari-Erik Nurmi。教育轨道在青少年学业倦怠中的作用：一项纵向研究。《英国教育心理学杂志》，78(4):663–689，2008。ISSN 2044-8279。doi:10.1348/000709908X281628。URL https://onlinelibrary.wiley.com/doi/abs/10.1348/ 000709908X281628。_eprint: https://bpspsychub.onlinelibrary.wiley.com/doi/pdf/10.1348/000709908X281628。Joselina Barbosa、Álvaro Silva、Maria Amélia Ferreira 和 Milton Severo。从中学升入医学院：自主学习与自我调节学习技能在新生倦怠中的作用。《葡萄牙医学学报》，29(12):803–808，2016年12月。ISSN 1646-0758。doi:10.20344/amp.8350。URL https: // www.actamedicaportuguesa.com/revista/index.php/amp/article/view/8350。Dale H. Schunk 和 Jo M. Rice。阅读理解教学中的学习目标与进度反馈。《阅读行为杂志》，23(3):351–364，1991。ISSN 0022-4111。地点：美国；出版商：Lawrence Erlbaum。Paul W. Glimcher。理解多巴胺与强化学习：多巴胺奖励预测误差假说。《美国国家科学院院刊》，108(supplement_3):15647–15654，2011年9月。doi:10.1073/pnas.1014269108。URL https://www.pnas.org/doi/10.1073/pnas.1014269108。出版商：《美国国家科学院院刊》。Wolfram Schultz。多巴胺奖励预测误差编码。《临床神经科学对话》，18(1):23–32，2016年3月。ISSN 1294-8322。doi:10.31887/DCNS.2016.18.1/wschultz。URL https://pmc.ncbi.nlm.nih.gov/ articles/PMC4826767/。

Masakazu Taira 和 Melissa J Sharpe。血清素和多巴胺在基于模型的学习中的互补作用。《*Current Opinion in Behavioral Sciences*》，61:101464，2025年2月。ISSN 2352-1546。doi:10.1016/j.cobeha.2024.101464。URL https://www.sciencedirect.com/science/article/pii/S2352154624001153。Kayoko W. Miyazaki、Katsuhiko Miyazaki、Kenji F. Tanaka、Akihiro Yamanaka、Aki Takahashi、Sawako Tabuchi 和 Kenji Doya。光遗传学激活背侧缝核血清素神经元可增强对未来奖励的耐心。《*Current biology*》*: CB*，24(17):2033–2040，2014年9月。ISSN 1879-0445。doi:10.1016/j.cub.2014.07.041。Katsuhiko Miyazaki、Kayoko W. Miyazaki、Akihiro Yamanaka、Tomoki Tokuda、Kenji F. Tanaka 和 Kenji Doya。奖励概率和时间不确定性会改变背侧缝核血清素神经元对耐心的影响。《*Nature Communications*》，9(1):2048，2018年6月。ISSN 2041-1723。doi:10.1038/s41467-018-04496-y。URL https: // www.nature.com/articles/s41467-018-04496-y。出版商：自然出版集团。Florent Meyniel、Guy M. Goodwin、Jf William Deakin、Corinna Klinge、Christine MacFadyen、Holly Milligan、Emma Mullings、Mathias Pessiglione 和 Raphaël Gaillard。血清素在克服努力成本方面的特定作用。*eLife*，5:e17282，2016年11月。ISSN 2050-084X。doi:10.7554/eLife.17282。Chiara Varazzani、Aurore San-Galli、Sophie Gilardeau 和 Sebastien Bouret。奖励/努力权衡中的去甲肾上腺素和多巴胺神经元：对行为中的猴子进行直接电生理比较。《*Journal of Neuroscience*》，35(20):7866–7877，2015年5月。ISSN 0270-6474, 1529-2401。doi:10.1523/JNEUROSCI.0454- 15.2015。URL https://www.jneurosci.org/content/35/20/7866。出版商：神经科学学会；栏目：文章。Gina R. Poe、Stephen Foote、Oxana Eschenko、Joshua P. Johansen、Sebastien Bouret、Gary Aston-Jones、Carolyn W. Harley、Denise Manahan-Vaughan、David Weinshenker、Rita Valentino、Craig Berridge、Daniel J. Chandler、Barry Waterhouse 和 Susan J. Sara。蓝斑核：重新认识蓝斑。《*Nature Reviews Neuroscience*》，21(11): 644–659，2020年11月。ISSN 1471-0048。doi:10.1038/s41583-020-0360-9。URL https://www.nature.com/ articles/s41583-020-0360-9。出版商：自然出版集团。Mark S. Gilzenrat、Sander Nieuwenhuis、Marieke Jepma 和 Jonathan D. Cohen。瞳孔直径反映了蓝斑核功能自适应增益理论所预测的控制状态变化。认知、情感与行为神经科学，10(2):252–269，2010年5月。ISSN 1531-135X。doi:10.3758/CABN.10.2.252。Sebastian Lee、Samuel Liebana Garcia、Claudia Clopath 和 Will Dabney。通过神经调节实现终身强化学习，2024年8月。URL http://arxiv.org/abs/2408.08446。arXiv:2408.08446 [cs]。Timo Flesch、David G. Nagy、Andrew Saxe 和 Christopher Summerfield。利用赫布式上下文门控和指数衰减的任务信号对人类持续学习进行建模，2022年3月。URL http://arxiv.org/abs/ 2203.11560。arXiv:2203.11560 [cs, q-bio]。Timothy P. Lillicrap 和 Konrad P. Kording。理解神经网络意味着什么？*arXiv:1907.06374 [cs,q-bio,stat]*, 2019年7月。URL http://arxiv.org/abs/1907.06374。arXiv: 1907.06374。Yuhang Song、Beren Millidge、Tommaso Salvatori、Thomas Lukasiewicz、Zhenghua Xu 和 Rafal Bogacz。将可塑性发生前的神经活动推断作为超越反向传播的学习基础。《*Nature Neuroscience*》，27(2): 348–358，2024年2月。ISSN 1546-1726。doi:10.1038/s41593-023-01514-1。URL https://www.nature.com/ articles/s41593-023-01514-1。出版商：自然出版集团。Jonathan D. Cohen、Kevin Dunbar 和 James L. McClelland。自动化过程的控制：对Stroop效应的并行分布式处理解释。《心理学评论》，97(3):332–361，1990年。ISSN 1939-1471。doi:10.1037/0033-295X.97.3.332。地点：美国；出版商：美国心理学会。Sebastian Musslick、Andrew Saxe、Abigail Hoskin、Yotam Sagiv、Daniel Reichman、Giovanni Petri 和 Jonathan Cohen。认知控制的理性有限性：共享表征与分离表征，2020年11月。URL https://osf.io/jkhdf_v1。Brieuc Pinon、Jean-Charles Delvenne 和 Raphaël Jungers。元强化学习的基于模型的方法：Transformer和树搜索，2022年8月。URL http://arxiv.org/abs/2208.11535。arXiv:2208.11535 [cs]。Hongyuan Dong、Dingkang Yang、Xiao Liang、Chao Feng 和 Jiao Ran。AdaLRS：面向高效基础模型预训练的损失引导自适应学习率搜索，2025年6月。URL http://arxiv.org/abs/2506.13274。arXiv:2506.13274 [cs]。Midhun T. Augustine。关于线性二次调节器和卡尔曼滤波器的简述，2023年8月。URL http://arxiv. org/abs/2308.15798。arXiv:2308.15798 [math]。Payam Piray 和 Nathaniel D. Daw。一种基于随机性与波动性联合估计的学习模型。《*Nature Communications*》，12(1):6587，2021年11月。ISSN 2041-1723。doi:10.1038/s41467-021-26731-9。URL https://www.nature.com/articles/s41467-021-26731-9。出版商：自然出版集团。

Jiaqian Si、Yanxiong Niu 和 Botao Wang。组合导航系统非线性滤波算法综述。传感器（瑞士巴塞尔），25(20):6462，2025年10月。ISSN 1424-8220。doi:10.3390/s25206462。URL https://pmc.ncbi.nlm.nih.gov/articles/PMC12567677/。Jiongmin Yong 和 Xun Yu Zhou。动态规划与HJB方程。载于 Jiongmin Yong 和 Xun Yu Zhou 编，《style id='3'>随机控制：哈密顿系统与HJB方程，第157–215页。施普林格，纽约州纽约市，1999年。ISBN 978-1-4612-1466-3。doi:10.1007/10.1007/978-1-4612-1466-3_4。URL https://doi.org/10.1007/ 978-1-4612-1466-3_4。R. W. Beard、G. N. Saridis 和 J. T. Wen。时不变哈密顿–雅可比–贝尔曼方程的近似解。《优化理论与应用杂志》，96(3):589–626，1998年3月。ISSN 1573-2878。doi:10.1023/A:1022664528457。URL https://doi.org/10.1023/A:1022664528457。H. Saberi Nik、S. Effati 和 M. Shirazian。通过同伦摄动法求解哈密顿 – 雅可比 – 贝尔曼方程的近似解析解。《应用数学建模》，36(11):5614–5623，2012年11月。ISSN 0307-904X。doi:10.1016/j.apm.2012.01.013。URL https://www.sciencedirect.com/science/ article/pii/S0307904X12000285。