主动推断与人工推理

Karl Friston[1,2], Lancelot Da Costa[2,3], Alexander Tschantz[2,4], Conor Heins[2,5], Christopher Buckley[2,4],Tim Verbelen[2], Thomas Parr[6]

[1] Queen Square 神经病学研究所，伦敦大学学院，英国[2] VERSES，美国加利福尼亚州洛杉矶，90016[3] 图宾根 ELLIS 研究所，德国图宾根[4] 信息学系，萨塞克斯大学，布莱顿，英国[5] 马克斯·普朗克动物行为研究所，集体行为系，德国康斯坦茨[6] 努菲尔德临床神经科学系，牛津大学，英国

电子邮箱：k.friston@ucl.ac.uk；lance.dacosta@verses.ai；alec.tschantz@verses.ai；conor.heins@verses.ai；christopher.buckley@verses.ai；tim.verbelen@verses.ai；thomas.parr@ndcn.ox.ac.uk

摘要

本技术简报探讨如何对结果进行采样，以便最大程度地获取有关底层世界模型结构的信息。这一推广为在一组合理的生成模型或假设下进行结构学习，提供了一种有原则的方法。在主动推断中，策略——即行动的组合——依据其期望自由能进行选择，而期望自由能由期望信息增益和价值构成。信息增益对应于行动后果存在与否时预测后验分布之间的KL散度。基于累积的模型参数后验信念，可以通过贝叶斯模型约简快速且高效地评估模型的后验分布。随后，可利用所得的信息增益选择行动，以区分备选模型，这与最优实验设计的思路一致。我们通过部分观测离散模型来展示这类主动选择或推理；具体而言，采用此前用于描述人工顿悟和“顿悟时刻”的“三球”范式，通过（合成）内省或睡眠实现这些现象。我们重点关注样本效率：通过寻求能够最大程度消除世界模型不确定性的结果来提升样本效率，而结果正是由该世界模型生成的。

关键词：主动推断；主动学习；贝叶斯模型选择；对称性；结构学习；推理。

引言

主动推断为基于生成模型或世界模型进行感知与决策提供了一种第一性原理方法。它在多个方面区别于（无模型）强化学习以及机器学习中的相关方法。首先，它明确区分了生成模型所蕴含的未知变量，即潜在状态、参数和结构，参见（Kemp 和 Tenenbaum，2008）。这要求分别对三类未知量进行推断，从而形成主动推断、学习和选择（Parr 等人，2022）。至关重要的是，潜在状态可以包含智能体的行动或策略，从而将规划视为主动推断，参见（Attias，2003；Botvinick 和 Toussaint，2012）。这体现了主动推断与学习状态-动作策略之间的另一项区别：规划依赖于关于状态的信念的自由能泛函，而不是状态本身的价值函数。这意味着主动推断会优化关于潜变量的（贝叶斯）信念。一个关键后果是，最优策略能够通过探索性、寻求信息的行动来消解不确定性——不确定性是贝叶斯信念的一种属性，而这类行动无法被表述为优化状态本身的（价值）函数。

随之而来的信息寻求行为，取决于在给定行动组合（即策略）下所预期的信息增益。关于状态的期望信息增益通常以显著性来指称，这与视觉搜索文献中的贝叶斯惊奇度概念及相应的显著性图相对应（Itti and Baldi, 2009; Parr and Friston, 2017）。在机器人学中，这类动机有时被称为内在动机——参见（Oudeyer and Kaplan, 2007）——并表现为人工好奇心（Schmidhuber, 2010）。当应用于生成模型的参数时，期望信息增益有时被称为新颖性；也就是说，通过消除对生成模型参数的不确定性所提供的认识论可供性（Barto et al., 2013; Schwartenbeck et al., 2019）。本文介绍第三种信息增益，它与生成模型的结构有关（Da Costa et al., 2025; Tsividis et al., 2021）。赋予智能体这种内在动机，符合最优贝叶斯设计的原则（Lindley, 1956; Mackay, 1992; Parr et al., 2024）；也就是说，设计实验以获取能够最大程度消除假设或生成模型歧义的数据。

选择能够最大限度降低不确定性或模糊性的观测，是自然智能和某些自然种类自组织的标志（Friston 等，2023）。可以认为，这类行为植根于推断——即最小化变分自由能的过程——而这一区别正是人工智能与自然智能的分野；有时也会用系统一与系统二思维来表述，例如（Kelly 和 Barron，2022）。还可以进一步认为，在主动推断中，与“思考”相对应的是规划（作为推断），即在合理的策略下评估反事实的未来。这种评估依据的是（路径积分）

期望自由能，它由期望信息增益和期望价值构成（Friston 等，2023）。在主动推断的框架下，价值对应于先验偏好结果的对数概率。期望信息增益或认知价值衡量的是智能体预期在每种策略下会发生何种程度的信念更新。传统学习方案无法企及期望信息增益，因为信息增益取决于后验信念，而后验信念按定义会随时间、情境和经验而变化。换言之，不存在从观测到期望自由能（并隐含地到策略）的可学习映射，因为这种映射本身会随学习而改变。简而言之，尽管人可以学会推断，例如（Mazzaglia 等，2022），却无法学会保持好奇。那么，带有认知意图的规划是否足以解释推理?

人们大概会回答不会。推理——尤其是当前大语言模型文献中的用法（Xu 等，2025）——意味着某种程度的递归内省；在主动推断文献中，这种情况更适合用贝叶斯模型约简来理解。贝叶斯模型约简是一种快速而高效的贝叶斯模型选择方法，它允许人们依据某个完整模型或父模型的先验分布和后验分布，评估备选模型的证据（即边际似然）（Dickey，1971；Friston 等，2018）。它通常以离线方式实现，作为一个最小化变分自由能的过程；有观点认为，这类过程会在内省或睡眠期间发生（Friston 等，2017；Hinton 等，1995）。

严格来说，变分自由能等于模型复杂度减去准确度，其中复杂度是后验分布与先验分布之间的KL散度（Penny，2012）。因此，主动推断与学习旨在寻求一个后验分布，使其与先验信念之间的散度最小，同时又能准确解释观测数据。然而，在所有数据都已观测之后，人们可以通过选择一个使其与后验分布之间的散度最小的先验，来最小化变分自由能。这种事后选择通常是在一组关于模型参数的先验中进行的，这些先验构成一个模型或假设空间。可以从以下意义上将这种选择视为科学过程的基础：科学家构建一个小型模型空间（例如，一个备选假设和一个零假设），在模型空间上的无信息先验下获取数据，然后找出最能解释其实验数据的假设（Corcoran 等，2023；Lindley，1956）。依照这种观点，最佳模型是使数据最有可能出现的模型；也就是说，它最大化模型证据或边际似然。

可以将这种事后最小化变分自由能的过程推广到任意数量的假设——或约简模型——这些假设或模型都可以用模型参数上的先验来表述。在经过一段适当的证据累积后，一旦确定了最佳假设或模型，智能体便可据此采用有充分依据且最优的模型结构。下文将把这种结构学习形式（Gershman 和 Niv，2010；Tenenbaum 等，2011）称为应用奥卡姆剃刀。贝叶斯模型约简已被用于通过三球范式模拟“顿悟时刻”和“抽象推理”：在该范式中，（模拟的）受试者必须

发现支配“正确”选择的规则（Friston 等，2017）。这里，我们将重新采用这一范式，以展示主动推理或约简；其重点是选择恰当的数据——例如，进行恰当的实验——以消除对所考虑模型或假设的不确定性。本文将“推理”理解为：在关于观测如何生成的相互竞争假设下，为观测选择最佳解释。主动推理则可理解为寻找能够最有效地区分相互竞争假设的观测，以供后续“推理”之用。

我们将看到，所需的期望信息增益可以通过贝叶斯模型约简快速而高效地计算出来，并用于补充支撑动作选择的期望自由能。将主动推断扩展到涵盖抽象或推理（Chollet，2019），很大程度上受到了 ARC-AGI-3 挑战（ARC-AGI-3）的启发；具体而言，这项挑战要求发现一条规则、一种对称性、一种不变性、一条定律或一种抽象，而对此没有任何先验信息。本文所述的程序通过描述一种高效的方式来查询世界或对世界采取行动，以发现其中起作用的规则或偶然关系，从而应对这项挑战的一部分。这只是对更大挑战的部分解决方案，因为我们假设智能体已经学习了当前谜题或游戏的生成模型，但不知道什么才算是正确的响应。然而，了解游戏中潜在原因的因果结构，就能得到一组符合某些对称性约束的假设或模型。具体而言，我们将考察从潜在原因映射到成功结果的假设性似然映射，其中具有某些同构。这使智能体能够通过主动约简或推理，逐步排除不太可能的映射，直到发现规则。由此形成的方案与概率归纳逻辑编程（Lake 等，2015；Riguzzi 等，2014）有着关键的共同主张；尤其是对参数学习与结构学习的区分。不过，该方案关注的是隐式归纳推理的样本效率。

本文分为三个部分。第一部分简要回顾后续数值研究所用的部分可观测马尔可夫决策过程中的信念更新与传播。本节重点讨论模型上的期望信息增益如何引发主动推理，以及如何利用贝叶斯模型约简简单、自动地评估期望自由能的这一组成部分。随后，将这一方案应用于第二部分所述的三球范式。该范式具有若干需要认知觅食的特征。尤其是，这个问题是部分可观测的：智能体只有在选取球后才能看到每个球的颜色（例如，注视每个球所在的位置）。因此，这需要序贯策略优化，而传统方法（如强化学习）寻求的是状态-动作策略，难以应对这种情况。本节模拟贝叶斯最优行为，兼顾贝叶斯决策理论和实验设计这两个意义上的最优性。第三部分重复进行模拟，分别纳入和排除期望自由能的不同组成部分，以刻画它们对样本效率和性能的贡献。本文最后简要讨论主动推断的这一应用涵盖了什么，以及没有涵盖什么。

主动推断、学习与选择

主动推断以可观测结果的生成模型 为基础。该模型用于根据对世界期望状态的推断，找出结果最可能的成因。这些状态及其转移或路径都是潜在的或隐藏的，因为它们只能通过观测加以推断。有些路径是可控的，即可以通过行动实现。因此，某些观测取决于行动或策略，这就要求生成模型对行动后果形成预期。通过最小化变分自由能来优化这些预期，而策略的先验概率则取决于其期望自由能（Parr 等人，2022）。

图1概述了不确定性条件下感知与决策通常使用的生成模型，即部分可观测马尔可夫决策过程（POMDP）。任意时刻的结果取决于隐藏状态，而隐藏状态之间的转移取决于路径。在该模型中，路径是依赖于行动的随机变量。这个部分可观测马尔可夫决策过程由一组张量指定。第一组**A**将隐藏状态映射到结果模态；例如，外感受的（如视觉的）或本体感受的（如眼位）模态。这些参数编码了给定隐藏状态时结果的似然。第二组**B**编码特定路径下某个因子的隐藏状态之间的转移。因子对应 不同类型的状态；例如物体的位置与类别。其余张量编码关于路径**C**的先验信念，以及初始条件**D**和 **E**，有时称为隐藏原因。这些张量通常以狄利克雷分布为参数化形式，其充分统计量是狄利克雷计数，用于统计推断出特定状态与结果组合的次数。我们将重点讨论由狄利克雷计数编码的似然模型学习**a**。

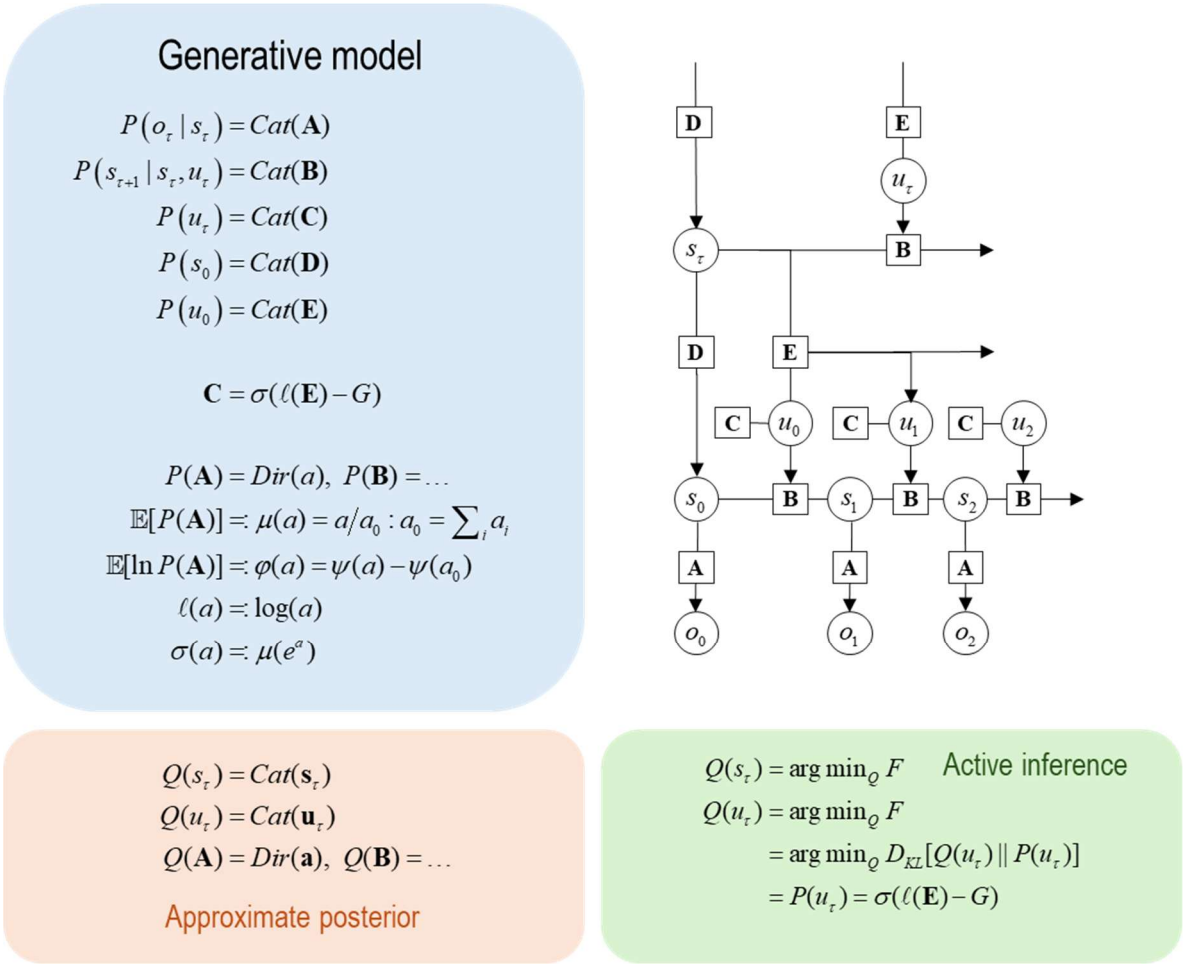

图1：**生成模型**。生成模型规定了可观察结果及其隐原因的联合概率。通常，模型以似然（给定原因时结果的概率）和先验（关于原因的先验）的形式表示。当先验取决于一个随机变量时，它称为经验先验。此处，似然由张量**A**规定，用于编码在状态（s）的每种组合下结果的概率。隐藏状态之间转移的先验**B**取决于路径（u），其转移概率由**C**编码；其中，**D**和**E** 分别规定初始状态和路径的先验。若某些路径能使其期望自由能（G）最小化，那么它们在先验上就更有可能出现。为清楚起见，我们假设此模型中只有一个因子和一条路径。蓝色面板以分类（Cat）分布的形式给出了生成模型的函数形式，这些分布由狄利克雷（Dir）分布参数化。下方的等式列出了图2中变分消息传递所需的各种算子。这些函数作用于其张量参数的每一列。$\psi(\cdot)$是双伽马函数。右侧面板将生成模型描绘为概率图模型，下方的面板列出了近似后验（Q）的函数形式及其优化。

图1中的生成模型意味着，结果是通过从期望自由能的softmax函数中选择一个策略而生成的。随后，使用所选路径组合（即策略）所指定的概率转移来生成隐藏状态序列。最后，隐藏状态在一个或多个模态中生成结果。 模型反演会更新后验信念（,,,）(,,,)()()()**sua**… Qs $uQsQuQ\ldots$ = … **AA**的充分统计量，这些信念在隐藏状态、路径和参数上进行因子分解。这种平均场分解有效地将信念更新划分为推断、规划和学习。

变分自由能与期望自由能

在变分（又称近似）贝叶斯推断中，模型反演意味着针对近似后验的充分统计量最小化变分自由能。为清晰起见，我们将只讨论单个因子，因此策略（即路径的组合）就成为路径。省略对初始状态和某些参数的依赖，对于模型 m，我们有：

$$
\begin{aligned}
F &= \mathbb{E}_Q[\ln \underbrace{Q(s_\tau, u_\tau, \mathbf{A})}_{\text{Posterior}} - \ln \underbrace{P(o_\tau \mid s_\tau, \mathbf{A})}_{\text{Likelihood}} - \ln \underbrace{P(s_\tau \mid s_{\tau-1}, u_\tau) P(u_\tau) P(\mathbf{A})}_{\text{Prior}}] \\
&= \underbrace{\mathbb{E}_Q[D_{KL}[Q(s_\tau, u_\tau, \mathbf{A}) \,||\, P(s_\tau, u_\tau, \mathbf{A} \mid s_{\tau-1}, o_\tau)]]}_{\text{Divergence}} - \underbrace{\ln P(o_\tau \mid m)}_{\text{Log evidence}} \\
&= \underbrace{\mathbb{E}_Q[D_{KL}[Q(s_\tau, u_\tau, \mathbf{A}) \,||\, P(s_\tau, u_\tau, \mathbf{A} \mid s_{\tau-1})]]}_{\text{Complexity}} - \underbrace{\mathbb{E}_Q[\ln P(o_\tau \mid s_\tau)]}_{\text{Accuracy}}
\end{aligned}
\tag{1}
$$

这里，期望是在 $Q(s_{\tau-1}, s_\tau, u_\tau, \mathbf{A}) = Q(s_{\tau-1}) Q(s_\tau \mid u_\tau) Q(u_\tau) Q(\mathbf{A})$ 下计算的。由于（KL）散度不可能小于零，倒数第二个等式意味着，当近似后验分布为真实后验分布时，自由能达到最小值。此时，自由能就变成生成模型的负对数证据（Beal，2003）。这意味着，最小化自由能等价于最大化模型证据。在主动推断下，通过对路径设置先验以最小化期望自由能，规划随之产生（Friston 等人，2015）。在一步前瞻策略的简单情形下（并将重点限制在似然参数上），我们有：

$$
\begin{aligned}
G(u) &= \mathbb{E}_Q[\ln Q(s_{\tau+1}, \mathbf{A} \mid u) - \ln Q(s_{\tau+1}, \mathbf{A} \mid o_{\tau+1}, u) - \ln P(o_{\tau+1} \mid c)] \\
&= -\underbrace{\mathbb{E}_Q[\ln Q(\mathbf{A} \mid s_{\tau+1}, o_{\tau+1}, u) - \ln Q(\mathbf{A} \mid s_{\tau+1}, u)]}_{\text{Expected information gain (Parameters)}} - \\
&\underbrace{\mathbb{E}_Q[\ln Q(s_{\tau+1} \mid o_{\tau+1}, u) - \ln Q(s_{\tau+1} \mid u)]}_{\text{Expected information gain (States)}} \underbrace{-\mathbb{E}_Q[\ln P(o_{\tau+1} \mid c)]}_{\text{Expected cost}} \\
&= -\underbrace{\mathbb{E}_Q[D_{KL}[Q(\mathbf{A} \mid s_{\tau+1}, o_{\tau+1}, u) \,||\, Q(\mathbf{A} \mid s_{\tau+1}, u)]]}_{\text{Novelty}} + \\
&\underbrace{D_{KL}[Q(o_{\tau+1} \mid u) \,||\, P(o_{\tau+1} \mid c)]}_{\text{Risk}} \underbrace{-\mathbb{E}_Q[\ln Q(o_{\tau+1} \mid s_{\tau+1}, u)]}_{\text{Ambiguity}}
\end{aligned}
\tag{2}
$$

这里，= τ+1 τ+1 | = τ+1 τ+1 | 0 ⋯ τ = τ+1 | τ+1 τ+1 | ：(, ,) (, , , , ,) (,) (,) *Q Qo s u P o s u o o P o s Q s u* **A A A A** 是在特定路径下，对下一时间步的参数、隐藏状态和结果的后验预测分布。 注意，期望是针对未来的观测而言的，因此称为期望自由能。这意味着，由*c*参数化的偏好结果属于先验信念，它们约束了作为推断的隐式规划：参见（Attias，2003；Botvinick 和 Toussaint，2012；Van Dijk 和 Polani，2013）。更长时间范围内的规划，只需将连续行动（或行动组合，即策略）上的期望自由能路径积分累积起来即可，具体如（Friston 等人，2021）所述。

期望自由能通过加入成本或约束来扩展期望信息增益。这些约束反映了这样一个事实：我们所处理的是具有特定结果的系统，而这些结果会根据其成本（ ）ln（ |）*o P o cC* =–或逆价值进行评分。从统计学角度而言，期望自由能以双重意义支撑贝叶斯最优行为，即贝叶斯决策理论（Berger，2011年）和最优实验设计（Lindley，1956年）。其函数形式源自自由能原理的路径积分表述：详情请参见（Friston 等，2023）。还可以用期望自由能来表示参数的先验，其中对路径进行边缘化后：

$$P(\mathbf{A}) = \sigma(-G(\mathbf{A}))$$

$$G(\mathbf{A}) = \mathbb{E}_Q[\ln P(s \mid \mathbf{A}) - \ln P(s \mid o, \mathbf{A}) - \ln P(o \mid c)]$$

$$= -\underbrace{\mathbb{E}_Q[\ln P(s \mid o, \mathbf{A}) - \ln P(s \mid \mathbf{A})]}_{\text{Expected information gain}} \underbrace{-\mathbb{E}_{Q_a}[\ln P(o \mid c)]}_{\text{Expected cost}}$$

$$= -\underbrace{\mathbb{E}_Q[D_{KL}[P(o,s \mid \mathbf{A}) \| P(o \mid \mathbf{A})P(s \mid \mathbf{A})]}_{\text{Mutual information}} \underbrace{-\mathbb{E}_{Q_a}[\ln P(o \mid c)]}_{\text{Expected cost}} \tag{3}$$

此处，：（= | | = | ,）( ) (,) *QPosaPsaPosa* 是结果与隐藏状态上的联合分布，由狄利克雷参数编码。请注意，狄利克雷参数编码了互信息，因为它们编码了结果与隐藏状态上的联合分布：对 *a* 张量的每一列进行归一化时，我们可以得到似然分布。然而，我们也可以对张量中的所有条目进行归一化，以得到联合分布。(3) 中的先验可以理解为实现了一种受约束的最大互信息原理或最小冗余原理（Ay 等人，2008；Barlow，1961；Linsker，1990；Olshausen 和 Field，1996）。在机器学习中，这类先验支撑了解缠结（Higgins 等人，2021；Sanchez 等人，2019），并且通常会产生稀疏表示（Gros，2009；Olshausen 和 Field，1996；Sakthivadivel，2022；Tipping，2001）。稍后在指定模型的先验时，我们将运用这一先验。

主动推断

在变分推断中，编码后验期望的充分统计量会更新，以最小化变分自由能。图2以变分消息传递的形式展示了这些更新（Dauwels，2007；Winn 和 Bishop，2005）。例如，对隐藏状态的期望是消息的 softmax函数，而这些消息是其他期望与观测的线性组合。

$$
\begin{aligned}
\mathbf{s}_{\tau}^{f} &= \sigma\left(\mu_{\uparrow \mathbf{A}}^{f} + \mu_{\rightarrow B}^{f} + \mu_{\leftarrow B}^{f} + \ldots\right) \\
\mu_{\uparrow \mathbf{A}}^{f} &= \sum\nolimits_{g \in ch(f)} \mu_{\uparrow \mathbf{A}}^{g,f} \\
\mu_{\uparrow \mathbf{A}}^{g,f} &= \mathbf{o}_{\tau}^{g} \odot \varphi(\mathbf{a}^{g}) \odot_{i \in pa(g) \setminus f} \mathbf{s}_{\tau}^{i}
\end{aligned}
\tag{4}
$$

在此，从似然因子传递的上行消息是期望状态与观测的线性混合[1]，其权重由参数化似然模型的狄利克雷计数的双伽马函数决定（参见连接权重）。图2中的表达式实际上就是变分自由能的不动点（即极小值）。这意味着消息传递对应一种不动点迭代方案（Beal，2003；Dauwels，2007；Winn 和 Bishop，2005）。[2]

[1] 该⊙ 记法表示和积算子；也就是说，它是点积或内积，会对数组或张量的一个维度求和。这些算子作用于向量**a**和张量**A**，其中**a** **A**⊙ 表示对乘积沿前导维度求和，而**A** **a**⊙ 表示沿尾随维度求和。例如，**1** **A**⊙表示对各行求和，而= $_0$**A** **1** **A**⊙表示对各列求和，其中**1**是全 1 向量，**A**是矩阵。若**A**是矩阵，则$_T$ = ·**a** **A** **a** **A**⊙。最后，$_{,i}$•**A**和$_{,i}$ •**A**分别表示**A**的第*i*列和第*i*行。这种记法取代了爱因斯坦求和记法，以避免视觉上的杂乱。

[2] 图2和公式（4）给出了随时间变化的前向消息和后向消息。在实践中（包括下文的示例），有时省略后向消息会更简单；此时，该方案就变成了贝叶斯滤波器，而非平滑器。还可以用和积信念传播方案中的消息替换变分消息，从而获得精确的边缘后验。

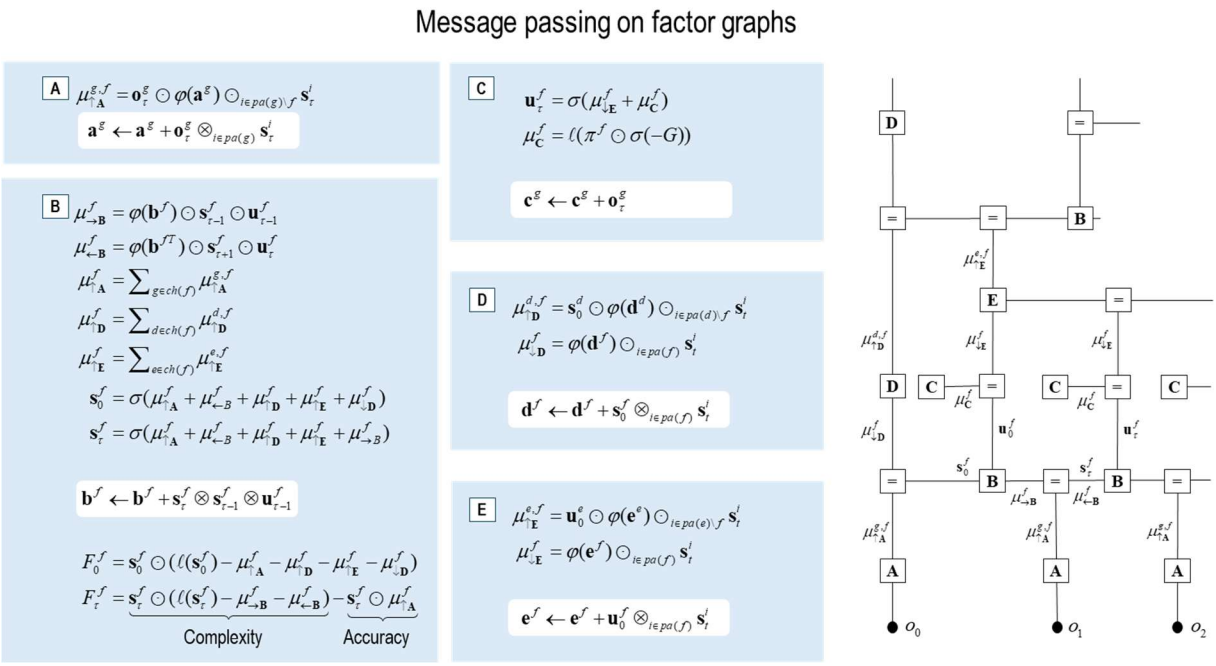

图2：**信念更新和变分消息传递**： 右侧面板将生成模型表示为Forney因子图，其中节点（方框）对应于生成模型的因子（标注有相关张量）。边连接共享随机变量依赖关系的因子。叶节点（实心圆）对应于已知变量，例如观测（o）。这种表示非常有用，因为它在因子图的边上搭建了消息传递的框架，而消息传递正是推断和规划的基础。这些消息的函数形式显示在 左侧面板中。例如，路径期望——面板**C**的第一个等式——是两个消息的softmax函数。第一个是来自**E** 的下降消息$_f\boldsymbol{\mu}_{\downarrow \mathbf{E}}$ ，它继承了关于上一级隐藏状态的期望。第二个是基于期望自由能的路径对数似然G。该消息取决于对偏好结果进行评分的狄利克雷计数——即模态g上的先验约束——这些约束编码于$_g\mathbf{c}$中：见方程（2）。矩阵π 编码了每个策略下路径的概率。下标表示时间，而上标表示不同的因子f和结果模态g。张量和矩阵用大写粗体表示，而后验期望用小写粗体表示。⊙ 符号表示广义内积（即点积）或张量缩并，而⊗ 表示外积。较浅面板中的更新对应于学习，即更新关于参数的贝叶斯信念。$ch(\cdot)$和$pa(\cdot)$分别返回潜变量的子节点和父节点。

主动推断中的信念更新在设计上具有样本高效性，因为它通过行动来最大化信息增益并最小化模糊性——期望自由能可用充分统计量表示如下：

$$G_h = \sum_g \underbrace{\mathbf{o}_{\tau+1}^{g,h} \odot \ell(\mathbf{o}_{\tau+1}^{g,h}) - \mathbf{o}_{\tau+1}^{g,h} \odot \varphi(\mathbf{c}^g)}_{\text{Risk}} \underbrace{-\mathbf{H}^g \odot \mathbf{s}_{\tau+1}^h}_{\text{Ambiguity}} \underbrace{-\mathbf{o}_{\tau+1}^{g,h} \cdot \mathbf{W}^g \cdot \mathbf{s}_{\tau+1}^h}_{\text{Novelty}} \tag{5}$$

$$\mathbf{o}_{\tau+1}^{g,h} = \mu(\mathbf{a}^g) \odot \mathbf{s}_{\tau+1}^h$$

$$\mathbf{s}_{\tau+1}^h = \mu(\mathbf{b} \odot \pi_h) \odot \mathbf{s}_\tau$$

$$\mathbf{H}^g = \mathbf{1} \odot (\mu(\mathbf{a}^g) \times \varphi(\mathbf{a}^g))$$

$$\mathbf{W}_{ij}^g = \frac{1}{\mathbf{a}_{ij}^g} - \frac{1}{\mathbf{a}_{0j}^g} + \ln \frac{\mathbf{a}_{0j}^g}{\mathbf{a}_{ij}^g} + \psi(\mathbf{a}_{ij}^g) - \psi(\mathbf{a}_{0j}^g) \approx \frac{1}{2\mathbf{a}_{ij}^g} - \frac{1}{2\mathbf{a}_{0j}^g}$$

这用策略 *h* 的风险、模糊性和新颖性来表示期望自由能；也就是说，*h* 是各因子上行动的第 h 种组合。

主动学习

在离散模型的情形下，学习对应于基于后验期望累积狄利克雷计数，从而更新模型参数。例如，对于似然张量，我们有：

$$\mathbf{a}^g \leftarrow \mathbf{a}^g + \mathbf{o}_\tau^g \otimes_{i \in pa(g)} \mathbf{s}_\tau^i \tag{6}$$

实际上，似然参数统计了结果及其潜在原因共同出现的次数。式（2）和（5）中对参数的期望信息增益，确保对结果进行采样，以尽量减少关于似然映射的不确定性。直观地说，这意味着要寻求能够提供似然张量中狄利克雷计数较少的列信息的结果。反之，当潜在状态和结果的组合已多次出现时，需要消除的不确定性就较少，因此这些结果提供的认识论可供性或期望信息增益也较低。

主动选择

与学习不同——学习通过优化参数的后验分布 来实现——贝叶斯模型选择或结构学习（Tenenbaum 等，2011；Tervo 等人，2016；Tomasello，2016）可以被表述为对模型参数先验的优化。从这一视角看，模型选择可以通过贝叶斯模型约简高效实现。贝叶斯模型约简是统计学中普遍存在的程序的一种推广，例如 Savage–Dickey 方法（Savage，1954）。通过将贝叶斯法则应用于完整模型和约简模型，可以轻易证明，对数边际似然的变化可以用后验狄利克雷计数**a**、先验计数*a*以及定义约简模型的先验计数来表示：$_m =$ $|a\ a\ m$。用 B 表示贝塔函数，我们有（Friston 等，2018）：

$$
\begin{aligned}
\mathcal{B}(\mathbf{a}, a_m, a) &= \ln P(o \mid a) - \ln P(o \mid a_m) \\
&= \ln \mathrm{B}(\mathbf{a})\mathrm{B}(a_m) - \ln \mathrm{B}(a)\mathrm{B}(\mathbf{a}_m) \\
\mathbf{a}_m &= \mathbf{a} + a_m - a
\end{aligned}
\tag{7}
$$

这里，$_m\mathbf{a}$对应于约简模型下得到的后验分布。

到目前为止，我们尚未考虑参数的先验。根据（3），我们可以假定一组模型$m \in \mathcal{M}$，在适当的约束下，这些模型都具有较高的互信息（潜在状态与观测之间）。在这种情况下，模型m成为一个随机变量，我们可以为其假定均匀先验。这就导出了一个参数先验，可将其视为贝叶斯模型平均：

$$
\begin{aligned}
&(\ )P \quad [\quad]_m\ {}_m a\ a \\
&= \mathbb{E}(\ )\ (\ )P\ Dir\ a = \\
&\mathbf{A}(\ )\ = 1 \mid \mathcal{M} \mid P\ m \\
&\quad /
\end{aligned}
\tag{8}
$$

这种先验设定使我们能够通过对过去累积了结果的狄利克雷参数进行贝叶斯模型约简，简单地评估模型的后验分布。应用贝叶斯法则，并略去不依赖于m的项，我们得到：

$$
\begin{aligned}
\ln P(m \mid o) &= \ln P(o \mid m) + \ln P(m) - \ln P(o) \\
&= \ln P(o \mid a) - \mathcal{B}(\mathbf{a}, a_m, a) - \ln \mid \mathcal{M} \mid - \ln P(o) \\
&\Rightarrow \\
P(m \mid o) &= \sigma(-\mathcal{B}) \\
\mathcal{B}_m &= \mathcal{B}(\mathbf{a}, a_m, a) = \ln P(o \mid a) - \ln P(o \mid m)
\end{aligned}
\tag{9}
$$

总之，如果我们从一组模型或假设空间 $\mathcal{M}$ 出发，其中各模型（此处为似然映射）同样合理，那么我们就有一种有原则的方法，可以通过贝叶斯模型平均来指定模型参数的先验。随后，可以不断累积支持各个模型的证据，直到我们对某个模型优于其余模型有足够的把握。然后，我们选择那个“最后的幸存者”，用于后续的推断和学习；也就是说，应用奥卡姆剃刀。这便是通过约简进行贝叶斯模型选择。关于这种自组织形式的仿生学示例，可参见（Kiebel 和 Friston，2011）：该示例将其应用于树突神经元过程以及对突触前输入的采样。那么，我们能否优化证据累积，以便后续进行模型选择？

利用（9）中后验模型的表达式，可以将期望自由能推广至涵盖模型方面的期望信息增益。这需要为智能体配备一个模型或假设库，使其能够主动搜寻消歧证据，并找出最

可信的模型，这可以通过应用奥卡姆剃刀来实现。鉴于狄利克雷参数化的完整模型或父模型（8）所累积的证据，期望信息增益相对容易评估。期望信息增益是指在特定策略下，包含结果与不包含结果时模型后验概率之间的预期KL散度。

$$
\begin{aligned}
G(u) = \\
&\underbrace{-\mathbb{E}_Q[\ln Q(m \mid \mathbf{A}, s_{\tau+1}, o_{\tau+1}, u) - \ln Q(m \mid \mathbf{A}, s_{\tau+1}, u)]}_{\text{Expected information gain (Models)}} \\
&\underbrace{-\mathbb{E}_Q[\ln Q(\mathbf{A} \mid s_{\tau+1}, o_{\tau+1}, u) - \ln Q(\mathbf{A} \mid s_{\tau+1}, u)]}_{\text{Expected information gain (Parameters)}} \\
&\underbrace{-\mathbb{E}_Q[\ln Q(s_{\tau+1} \mid o_{\tau+1}, u) - \ln Q(s_{\tau+1} \mid u)]}_{\text{Expected information gain (States)}} \\
&\underbrace{-\mathbb{E}_Q[\ln P(o_{\tau+1} \mid c)]}_{\text{Expected cost}}
\end{aligned} \tag{10}
$$

模型的期望信息增益所需的预测后验可由公式（5）、（6）和（9）推得：

$$
\begin{aligned}
Q\big(m \mid \mathbf{A}, s_{\tau+1}, o_{\tau+1}, u\big) &= \sigma(-\mathcal{B}) : \mathcal{B}_m = \mathcal{B}(\mathbf{a}^g + \Delta^g, a_m^g, a^g) \\
Q\big(m \mid \mathbf{A}, s_{\tau+1}, u\big) &= \sigma(-\mathcal{B}) : \mathcal{B}_m = \mathcal{B}(\mathbf{a}^g, a_m^g, a^g) \\
\Delta^g &= \mathbf{o}_{\tau+1}^{g,u} \otimes_{i \in pa(g)} \mathbf{s}_{\tau+1}^u
\end{aligned} \tag{11}
$$

请注意，此处的 Q 是预测后验，而不是模型上的近似后验。换言之，它是公式（8）中模型平均下的预测后验。这体现了当前 事后 应用贝叶斯模型约简（Friston 和 Penny，2011）与通过连续结构学习推断模型后验分布之间的区别。关于这一点的简要讨论，请参阅附录。

为清晰起见，我们仅假设到模态 g 的似然映射未知，并且只有一个具有可控路径的因子（即 $h = u$）。当多个映射未知时，只需将相应的期望信息增益加到期望自由能中即可。请注意，除非在未来连续的时间点更新参数，否则这一期望信息增益不适用于在动作序列上进行评估。因此，我们将模型上的期望信息增益的使用限制在下一个行动（或其组合）上，并在选出最佳模型后，对更长的时间范围进行规划。

纳入模型的期望信息增益，可以提供最能消除假设或模型歧义的证据或数据。这本身并不意味着要采用新模型。只有在以下情况下，才需要应用奥卡姆剃刀，用约简模型取代完整模型：

智能体充分确信约简模型是对迄今累积证据的最佳解释。我们可以用对数贝叶斯因子（即对数优势比）来量化应用奥卡姆剃刀的过程，评估最佳模型比任何其他模型更有可能的概率：

$$\mathcal{R} = \ln \frac{P(m^* \mid o)}{1 - P(m^* \mid o)} \tag{12}$$
$$m^* = \arg\max_m P(m \mid o)$$

当奥卡姆剃刀超过某个阈值时，智能体会用基于模型后验分布的贝叶斯模型平均，取代其参数后验分布。若将阈值设为 16 个自然单位，贝叶斯模型平均就会变成贝叶斯模型选择。这是因为，胜出的模型比备选模型至少可能性高 exp(16) 倍：

$$\mathbf{a} \leftarrow a_{m^*} = \mathbb{E}_{P(m|o)}[a_m] \Rightarrow Q(\mathbf{A}) = Dir(a_{m^*}) \tag{13}$$

在实践中，可以将所选模型中的狄利克雷计数乘以任意大的数（例如 512），以抑制新奇寻求。反过来，也可以让智能体更容易受影响，并直接使用约简模型（其狄利克雷计数通常在各列上的总和为一）。

随后的主动模型约简或推理，重现了科学家在消除对备择假设的不确定性时会采取的做法。她会在对假设采用无偏（平坦）先验的条件下，获取足够数量的数据，以形成预测后验分布。获取数据后，如果支持某个假设的证据相对于其他假设足够充分，她就会采纳该假设。这一类比强调了主动推断、学习和选择所伴随的时间尺度分离。简言之，推断是关于随时间变化的潜在状态进行信念更新的快速过程。学习是关于生成模型参数进行信念更新的缓慢过程，而选择则是积累了足够数据后才会进行的事后过程。

在这种事后方案中，通过学习狄利克雷参数来累积证据；这样便可仅依据迄今为止的观测进行贝叶斯模型选择。我们可以设想，针对选择模型前所需的数据点数量引入超先验。事实上，科学家通常需要进行功效计算，以估计分析前应收集多少观测。为简便起见，我们将仅使用公式（12）中的对数优势比，在其超过阈值时选择最佳模型——也就是应用奥卡姆剃刀。显然，这并不能保证所选模型就是最佳模型，因为智能体可能会过早得出结论。不过，只要阈值足够高，这类情况应该相对少见。

一个尚待解决的问题是如何构建备选模型，即如何规定模型空间。对 $\mathcal{M}$ 的必要规定，实际上为假设的规则、对称性、不变性等设定了超先验。用于说明上述方案应用的示例问题，展示了一种相当直接的自动规定模型空间的方法。我们将在更详细地描述该问题后，再回到这一规定。

三球范式

三球问题最初被用于引入主动学习和模型选择，作为洞察与「顿悟时刻」的数学对应物，这些现象体现了（人工）好奇心（Friston 等，2017）。在此，我们采用这一范式的一个细致版本，展示如何通过在期望自由能中加入针对模型的期望信息增益，主动采样数据，从而尽可能高效地实现贝叶斯模型约简。

简而言之，智能体面前有三个不同颜色的球，并受邀选择一种颜色。如果智能体做出选择（例如按下按钮），就会收到正确或错误的反馈。智能体的先验偏好 c, 设定为：无反馈、正确和错误这几种结果的相对对数概率分别为 0、2 和 -6 个自然单位。换言之，它偏好收到正确反馈，并认为错误结果极其出乎意料，因此会厌恶这种结果。智能体必须发现一条规则，才能确保获得偏好结果；这条规则取决于情境：如果中间的球是绿色，正确的颜色就是绿色。但如果中间的球是红色，正确的颜色就在左边；如果中间的球是蓝色，正确的颜色就在右边。这条规则被设定在似然映射中，该映射将表示三个球颜色及所选颜色的潜在状态映射到反馈模态。因此，智能体必须学习并选择能够编码构成该规则的偶然关系的似然张量。那么，智能体该如何发现这条规则呢?

如果智能体能够同时看到全部三个球，那么使用强化学习来学习该规则将相对简单；例如，参见（Silver 等，2021）。例如，球的颜色与响应的组合只有$3\times3\times3\times4 = 108$ 种（即三种颜色选择和一个“不响应”选项）。这意味着，在 500 次训练试次之后，至少看到每一种组合一次的概率至少为 $99\% = 1 - (1 - 1/108)^{500}$，因此几乎肯定能够学习到似然映射。然而，像三球范式这样的问题排除了可学习的状态-动作策略，因为智能体一次只能看到一个球，具体取决于它正在注视的位置。这会将马尔可夫决策问题转化为部分可观测问题，在这种问题中，智能体不仅必须选择正确的颜色，还必须选择注视的位置。这种无处不在的问题需要序贯策略优化。直观地说，知道左边有一个红色球并不能确定

正确的响应，因为中间的球可能是绿色或蓝色。这意味着，最优策略是先通过观察中间的球来消除关于情境的不确定性；如果它是绿色，就立即作出响应；如果它是蓝色或红色，则观察右边或左边。然后选择所观察到的颜色。接下来的问题现在可以用以下生成模型来描述。

该模型的潜在状态分为五类。其中三类状态（即因子）用于编码每个位置上球的颜色（我们将其称为位置因子）。其余可控因子分别是智能体的注视位置（起始位置、左、中或右）和智能体的选择（不选择、红色、蓝色或绿色）。这些隐藏状态足以生成两种结果模态：一种是单一的视觉模态（什么也看不见，或看到一个红色、绿色或蓝色的球），另一种是反馈模态（无反馈、正确、错误）。规则编码在似然张量 $\in \mathbb{R}^{3\times3\times3\times4\times4}$ g 中，它将隐藏状态映射到反馈模态。该模型的转移先验非常简单。在一次试次中，每个位置上球的颜色之间不会发生转移，因此转移张量为单位矩阵 3 3 f I = $\in \mathbb{R}^{\times}$ B。可控因子或动作依赖因子则取决于行动，4 4 4 f $\in \mathbb{R}^{\times\times}$B；例如，向左看会使位置状态转移到左侧，其他方向依此类推。选择因子也类似。每次试次中，球的颜色都从 $\mathbb{R}^{3}$ f∈D 中随机选取，同时智能体注视起始位置。策略依据期望自由能进行选择，而关于反馈模态的先验偏好由 ^{3} (=σ [0,2,–6] ∈) g c $\mathbb{R}$ 指定。智能体可以搜索并作出选择，持续 6 个时间点，之后开始新的试次，并随机选取球的颜色。

图4展示了配备上述生成模型且掌握正确规则的智能体的行为。该图以图像形式呈现了关于潜在状态和策略的后验密度所体现的信念更新与选择行为。在这个例子中，中间球的颜色是红色，这意味着正确选项是左边红色球的颜色。凭借信息性偏好和主动推断——向前考虑两步——智能体首先从起始位置观察中间的球，以消除对信息球颜色的不确定性。在消除对当前情境的不确定性后，它观察左边的球，以推断其颜色。随后，在剩余的步骤中，它都选择红色选项，从而使偏好的（即正确的）结果数量最大化。然而，在继续选择红色的同时，它也进一步消除了对右边那颗非信息球的不确定性。换言之，它在利用关于正确响应的后验信念的同时，继续进行视觉探索。我们现在要探讨的问题是：如果智能体不知道规则，会发生什么？在缺乏这方面知识时，它应该采用何种先验？

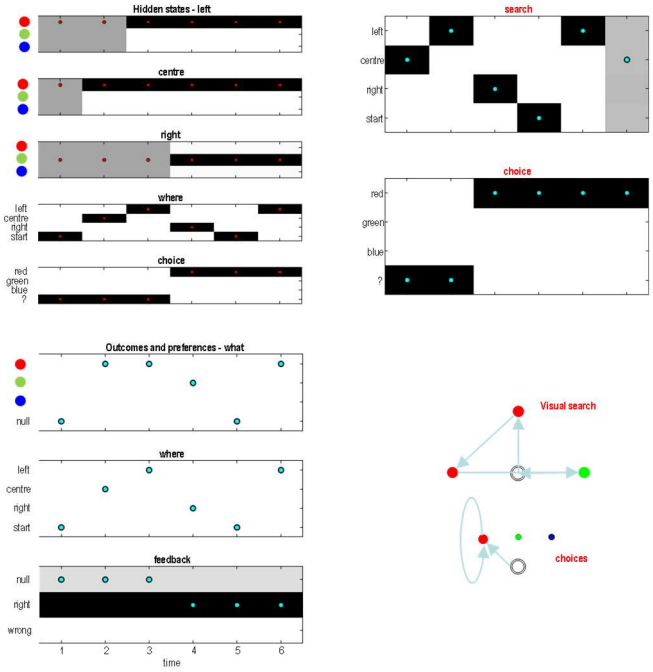

图3：选择行为。此图展示了三球范式下单次试验的信念更新。
左上方的面板以图像形式展示了五个潜在状态因子在六个时间点上的后验信念。其中两个潜在因子是可控的，并带有路径后验，每个路径后验有四个层级。这些后验分布显示在右上方面板中。左下方面板展示了智能体寻求的结果。在这些结果中，第三个（反馈）结果具有信息性先验偏好，因此智能体偏好正确，即判断为对，并且会对判断不正确（即错误）感到高度意外。颜色越深表示概率越高，圆点则代表实际状态和结果。右下方面板展示了智能体的响应，包括其注视位置（视觉搜索）和选择行为（选择）。在此示例中，智能体首先看向中间位置，并观察到颜色为红色。随后，它知道信息性（判据）位置在左侧。看向左侧后，它观察到一个红球，随后选择红色。接着，智能体看向右侧，以消除对右侧球颜色的剩余不确定性。一旦推断出正确的颜色，智能体便继续选择这种颜色，以最大化其偏好的反馈。

为简单起见，我们假设智能体已经学会了生成模型，但似然张量（即从潜在状态映射到反馈模态的映射）除外。正是这一似然映射包含了智能体必须发现的规则。那么，智能体还可以设想哪些其他似然映射?

在指定（或继承）生成模型参数的先验时，实际上也隐含地指定了一组约简模型，这些模型可以通过从完整先验中移除参数而得出。或者，也可以认为，初始先验实际上是对一组约简模型的混合——即贝叶斯模型平均——而这些模型都满足某些约束。但究竟是什么约束呢?现在可以转而考察式（3）中的超先验，并注意到，最大互信息的一个必要条件是

似然张量的每一列都应恰有一个非零项[3]。这也确保了稀疏性最大化，即零项的数量达到其上确界。

稀疏性或许是最简单的约束，其中所有模型都具有从每种潜在状态组合到某种响应的唯一映射；例如，正确或错误。这可以作为规则的最简定义。例如，如果我们将似然张量视为潜在状态与结果上的联合分布，就可以通过以正确结果为条件来指定先验，并考察相应子张量的结构：例如，$\in \mathbb{R}^{1\times3\times3\times4\times4}{}_{g\,\text{正确}}\mathbf{A}$。在当前示例中，有3×3×3×4×4 = 432 种潜在状态组合可以被赋予为一或零的浓度参数，具体取决于给定组合是否分别导致正确或错误的结果。这将产生 2^{432} 个模型或规则。显然，这是一种较弱的规则，因为规则应当将对观测的解释压缩为一个较小的集合，旨在最小化算法复杂度或描述长度：例如，（Hinton 和 Zemel，1993年；Ruffini，2017；Schmidhuber，2010；Wallace 和 Dowe，1999年）。简言之，稀疏性是最大化互信息的必要但非充分条件。要最大化互信息，某些隐藏状态组合必须生成正确结果，而另一些则不能。在此之外，互信息约束对于哪些规则是合理的，几乎无法提供更多信息。

然而，潜在状态具有因子结构，能够区分可控或可选择的状态与不可控或不可选择的状态。我们可以分别将它们称为选择状态和判据状态。可以考虑制定一条规则来实施以下约束：对于判据状态的任意组合，都存在一个唯一的选择状态，能够生成正确结果。这保证了最大互信息。例如，共有 3×3×3 = 27 种判据（即位置）状态组合，每种组合都可以与 4×4 = 16 个选择状态中的一个相关联。这将规则数量减少到 $16^{27} = 2^{31}$，但仍然大于阿伏伽德罗常数。那么，我们能否进一步施加更多对称性，以定义一个模型空间?

现在可以考虑因子之间的对称性，并假设在且仅在 选择状态与判据状态具有相同的支撑集或定义域时，两者之间存在同构映射。例如，此处的三个位置因子（左、中和右）与选择因子共享标注状态，即（红色、蓝色和绿色）。这种同构可能取决于情境，也可能不取决于情境；也就是说，取决于以下内容的组合

[3] 通过注意到以下事实，可以明确地证明这一点：在给定似然参数的情况下，状态与观测之间的互信息可以分解为状态的先验熵减去给定观测条件下状态的条件熵。只有后者取决于似然参数。这意味着，当对于任何合理的观测，状态的条件熵都为零时，即每个观测都只有一个合理状态时，可以获得最大互信息。将状态的先验向量相乘并进行重新归一化后，可以满足似然矩阵（或张量）的稀疏性约束。

判据状态。换句话说，对于每一种反馈结果，都存在似然张量的一个切片（选取方式为：一个维度取自选择因子，另一个维度取自判据因子），其形式为恒等映射。在当前示例中，这会产生 $3\times3\times3 = 27$ 个情境，它们可分别指定三个判据状态之一，并支持与某个选择状态形成同构。此约束产生 3^{27} 个模型，但数量仍然过多。那么，我们能否采用较弱的约束——即稀疏性约束?

较弱的对称性或同构涉及隐藏状态因子的大小。在这里，存在多个情境状态，可以唯一确定一个判据状态，这是因为每个情境状态的数量等于判据状态的数量。请回想，情境状态是判据状态的集合；在我们的设定中，有 3 个判据状态因子构成情境，每个情境状态因子都可能取 3 个值（红色、绿色或蓝色）。现在可以假设，一个情境状态唯一确定一个判据状态，从而满足稀疏性约束。在当前示例中，有三个可能的判据状态（左、中或右），每个判据状态都可以通过 3^3 种方式由三个情境状态（红色、绿色或蓝色）之一来确定（例如，左红色 → 左，左红色 → 中，.....，中红色 → 左，中红色 → 中，.....，右蓝色 → 右）。这会产生 $3\times3^3 = 81$ 条可能的规则[4]（指定的判据状态因子有 3 个可能取值，且有 3^3 种方式可将情境映射到判据）。这些对称性约束由此构成一个合理的假设空间，$\mathcal{M}$它 (i) 根据公式 (8) 联合确定完整先验，(ii) 提供通过公式 (10) 中的期望信息增益进行主动约简或推理的机会，以及 (iii) 支持随后根据公式 (12) 进行模型选择。

一般而言，可以通过识别一个与多个标注状态共享状态的（受控）选择因子和（非受控）判据因子，以狄利克雷张量的形式自动指定规则或似然模型。随后，识别候选情境因子，其各水平能够唯一指定判据因子。最后，对于每种由n个情境因子和m个判据因子构成的组合，只要选择与情境状态所指定的判据状态相符，就为一个正确结果分配狄利克雷分布，从而生成一个假定的似然映射。算法上，可以表示如下，其中，$_{f}{}_{i}S$表示因子f在第i种状态组合下的水平:

对于n个情境状态以及m个判据状态的m种指定

对于状态的每一种组合，i

[4] 其中一些规则是退化的，因为总共只有79条独特规则。

$$criterion = s_i^{context}$$

$$a^{feedback}_{correct, s_i^1, s_i^2, \ldots} = \begin{cases} 1 & s_i^{choice} = s_i^{criterion} \\ 0 & s_i^{choice} \neq s_i^{criterion} \end{cases} \quad (14)$$

$$a^{feedback}_{incorrect, s_i^1, s_i^2, \ldots} = \begin{cases} 0 & s_i^{choice} = s_i^{criterion} \\ 1 & s_i^{choice} \neq s_i^{criterion} \end{cases}$$

end

end

为简单起见，在生成模型空间时，我们省略了一些事务性细节。例如，有必要用落在同构之外的偶然关系来补充似然映射。在当前示例中，必须添加从无选择映射到无反馈 结果的狄利克雷参数。通过确保似然张量的每一列至少包含一个条目，即可轻松自动实现这一点。由此产生的模型空间会生成以下形式的规则：

“如果情境是情境状态，那么判据状态就是正确的选择”：

在这个例子中，规则 65 是用于生成反馈的规则：

如果中是红色，那么左状态就是正确的选择：如果中
是绿色，那么中状态就是正确的选择：如果中是蓝色，
那么右状态就是正确的选择：

在这种形式下，规则发现可视为概率归纳逻辑编程（PILP）（Lake 等，2015；Riguzzi 等，2014）。概率程序归纳逻辑将推断与一阶逻辑结合起来，从数据中学习概率逻辑程序。这意味着需要同时或作为独立子问题学习程序的结构（规则）和参数（概率）。在此，针对模型的期望信息增益使结构学习尽可能高效，而针对参数的期望信息增益则确保参数学习高效。

上文中，我们考察了映射到某种模态的似然映射，该模态提供相对稀疏且信息量不足的反馈。这类反馈通常出现在奖励学习范式中（其中，正确和错误分别可与奖励和惩罚相关联）。关键在于，稀疏奖励只有在“星星排列就绪”时才会给出。在这一观点下，规则对应于星星（即状态）可能排列就绪的不同方式（即以同构的方式排列）。思考如何为信息量更丰富的模态等价地生成模型空间，是很有意思的。例如，可以利用 *where* 因子的各个水平与三个球的位置之间的同构，再结合位置与结果颜色之间的同构，来指定到第一个（*visual*）模态的似然映射。在此例中，*location* 因子对应于情境状态，该状态规定三个判据状态中的哪一个会生成结果。在这里，存在

唯一一个满足这些同构约束的模型，因此无需进行模型选择。有趣的是，这种情境敏感的模型结构非常普遍，因为它建立在对“什么”（即 *criterion*）和“哪里”（即 *context*）进行因子分解的基础上，从而能够生成“如果我看向那里，我会看到什么？”这一问题的答案（Ungerleider，1994）。

显然，采用上述（同构）规则并不能提供一种通用的方法来规定模型空间。例如，我们可以认为，情境状态由另一个情境状态来规定，从而产生三阶假设。或者，我们可以增加一条约束，规定每个情境状态都指定一个唯一的判据状态，从而得到 $3\times3! = 18$ 种可能的规则。不过，这样就无法采用简单的一阶规则，即固定的判据状态。从情境规则的角度来看，一阶规则仅仅意味着：当情境状态处于第一层、第二层、第三层，依此类推时，判据状态都会被指定。例如：

如果中是红色或绿色或蓝色，那么左侧状态就是正确的选择：

如果先验看来合理，还可以考虑多选因素所带来的约束，以扩大模型空间。例如，我们可以考虑如下形式的规则：只有在查看起始位置时，才能获得正确的响应：

如果中是红色 且位置是起始位置，那么左侧状态就是正确的选择：如
果中是绿色 且位置是起始位置，那么……

总之，在正确反馈取决于情境敏感型同构这一假设下，三球范式提供了潜在状态与反馈模态之间 81 种可能的似然映射。这组似然模型包含真实规则，其形式为：如果某个情境因素的状态是这样，那么隐含判据因素的状态就决定满足该规则的控制因素的状态。在本例中，中央球的颜色决定了哪个球的位置与正确选择所对应的颜色相符。按照构造，这是一条相对简单的规则，涉及具有不同属性（例如颜色和位置）的物体。换言之，我们不必考虑物体的位置、姿态或颜色之间的同构；然而，原则上，类似的同构约束也可以应用于表达能力更强的生成模型。下文将使用随后的模型先验来模拟人工推理，其中智能体必须发现该规则。

人工推理

本节重复图3中的模拟，但起始时使用基于上述规则的贝叶斯模型平均结果得出的先验信念，这些信念涉及蕴含该规则的似然模型。图4展示了

64次试验的结果，每次试验包含六个时间点（即五次移动）；试验开始时，反馈似然采用狄利克雷先验，该先验基于公式（8）中的模型平均，并归一化至总和为1。起初，策略——即眼球运动与选择的组合——根据其期望自由能进行评估，且采用无信息的先验偏好。这意味着视觉搜索行为旨在消除对各位置球的颜色以及正在起作用的81条规则的不确定性。因此，智能体会查看每个位置，并选择不同颜色，以判断结果正确还是错误。随着试验进行，智能体在似然映射中累积狄利克雷计数，实质上是在不断积累证据，以揭示从潜在状态到结果的偶联关系映射。随后，利用这些证据，从初始先验或完整先验出发，通过贝叶斯模型约简评估各模型的期望信息增益。

随着证据不断累积，某些假设会被排除，而剩下的合理规则则会依次进行比较和筛除。当智能体足够确信最可能的模型优于所有其他模型时，便会应用奥卡姆剃刀：见公式（13）。换言之，它会用基于贝叶斯模型约简选出的映射，取代关于似然映射的后验信念。此时，智能体会采用信息性偏好，并以更大的信心规划未来两步。这一时刻可视为“顿悟时刻”，也可视为某一数据收集过程的结局，意味着智能体可以据此承诺采用某个特定假设。图4以智能体执行策略时的信心为依据，展示了这一转变。这里，信心就是关于策略的后验信念的精度或负熵。有趣的是，在人类神经生理学研究中，这种精度曾被用来解释多巴胺能放电（Friston等，2014；Schwartenbeck等，2015）。隐式规则发现或洞察伴随着行为的显著变化。图4的第二个面板从结果的角度展示了这一点。在认知觅食期间——也就是为了获取信息量最大的结果——智能体几乎每一步都会做出选择，而选择可能正确，也可能错误。这两种结果都会为关于规则或假设的后验信念提供信息。然而，一旦关于规则的不确定性消除，智能体便能获得最多的正确反馈，并完全避免令人厌恶的错误结果。请注意，正确结果的最大数量为三或四，具体取决于中央颜色是否为绿色：如果中央颜色为绿色，智能体就知道下一步应选择绿色；否则，它还必须额外进行一步认知探索，查看判定位置。

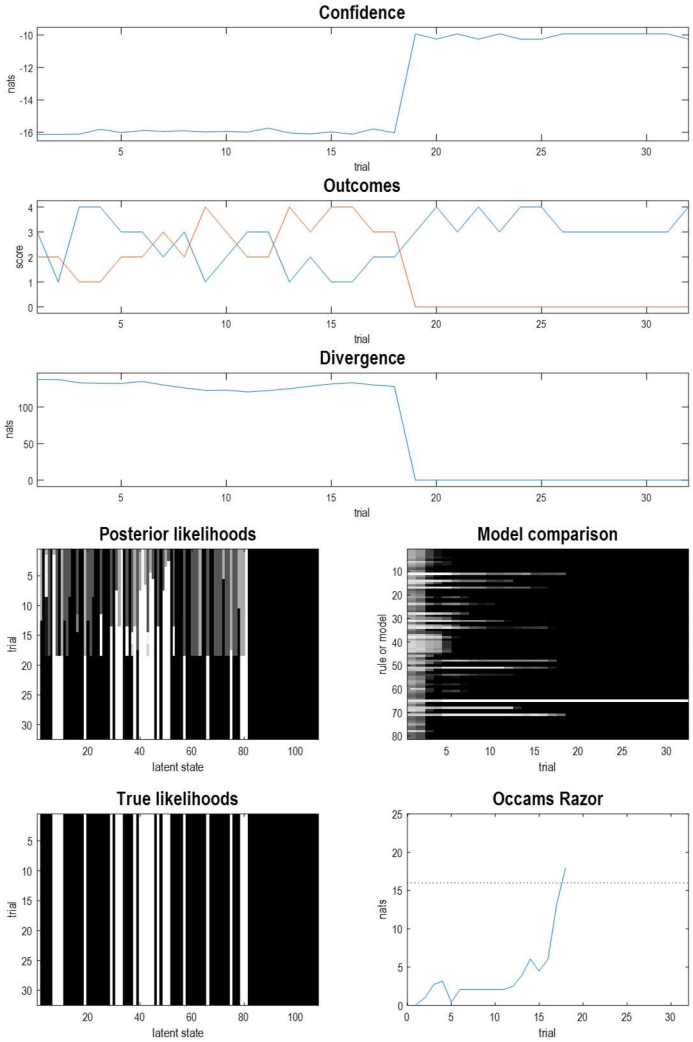

图4：主动推理。此图报告了图3所示游戏中，历经32次试验的信念更新及随后的模型选择。每次试验都从随机选取球的颜色开始，它们作为指示性或情境线索，帮助智能体消除对游戏性质的不确定性。具体而言，智能体最初有（81）种关于潜在状态与反馈模态之间似然映射的假设，并不断积累支持或反对这些假设的证据，直到选定其中一种（第18次试验）。上方面板报告了智能体对其响应的信心（即负熵，或对策略的信念精确度）。第二个面板显示结果。请注意，在模型选择之前，智能体会尽可能多地选择结果，不论这些结果是正确还是错误（分别以蓝线和红线表示）。然而，在它确定采用某个特定模型后（此处为正确模型65），便会避免寻求令人意外或令人厌恶的结果。它之所以能够做到这一点，是因为此时它已精确掌握了规则，如下方面板所示。后验似然和真实似然分别表示似然张量的一个切片，该切片报告了潜在状态所有组合下的正确结果。模型选择之后，后验似然会收敛于真实似然，第三个面板中的散度降至0。此时，模型分布（模型比较）变得明确，这反映在奥卡姆剃刀的阈值被跨越：该指标对获胜模型比其余模型更可能的对数优势比进行评分。

在前 18 次试验的证据累积过程中，关于似然映射的后验信念与最终选定的映射相去甚远；不过，随着狄利克雷计数不断累积，这种差距平均而言会缓慢缩小。下方各面板通过似然张量的一个切片总结了这些狄利克雷计数的累积情况；具体来说，该切片是公式（14）中将潜在状态的所有组合映射到正确结果的子张量。随着时间推移，隐含的概率分布变得越来越精确，直至选定（真实的）映射。此时，后验似然与真实似然相吻合，第三个面板中的 KL散度降为零。标题为“模型比较”的面板显示了基于贝叶斯模型约简得到的模型概率分布；也就是说，根据当前后验分布计算每个模型或规则的似然，如公式（9）所示。在此例中，真实规则是第 65 条，并且随着证据累积，其优势越来越明显。标题为“奥卡姆剃刀”的图表反映了这一点，其中报告了公式（12）中的对数优势比。在第 18 次试验时，该比值超过 16 纳特的阈值，于是实施奥卡姆剃刀。至此，主动推理的示例告一段落。在主动推理中，动作选择或选择行为本质上由期望信息增益驱动，该增益涉及潜在状态（即看不见的球的颜色）和似然映射（即决定响应是否正确的规则、偶然关系或模型）。

图5总结了使用随机选取的初始条件重复图4中的模拟64次所得的结果。上方面板报告后验似然映射与真实规则之间的KL散度。当发现真实规则时，该散度降至零；此时奥卡姆剃刀的阈值为16纳特（见第二个面板）。第三个面板以累积分数表示性能，即正确结果数减去错误结果数。请注意，在发现规则之前，性能分数往往会逐渐下降。这是因为出错的方式比正确的方式更多，但两者都具有信息价值。下方面板中的直方图展示了性能和发现时间。数值研究表明，发现规则所需的试验次数取决于每次试验的（随机）初始条件以及后续的认知觅食。在这个例子中，智能体有几次仅用九次试验就推断出了正确规则，但有时需要超过40次试验。发现时间的众数约为14次试验，累积分数约为150。不过，这种样本效率有多少可以归因于主动推理呢?

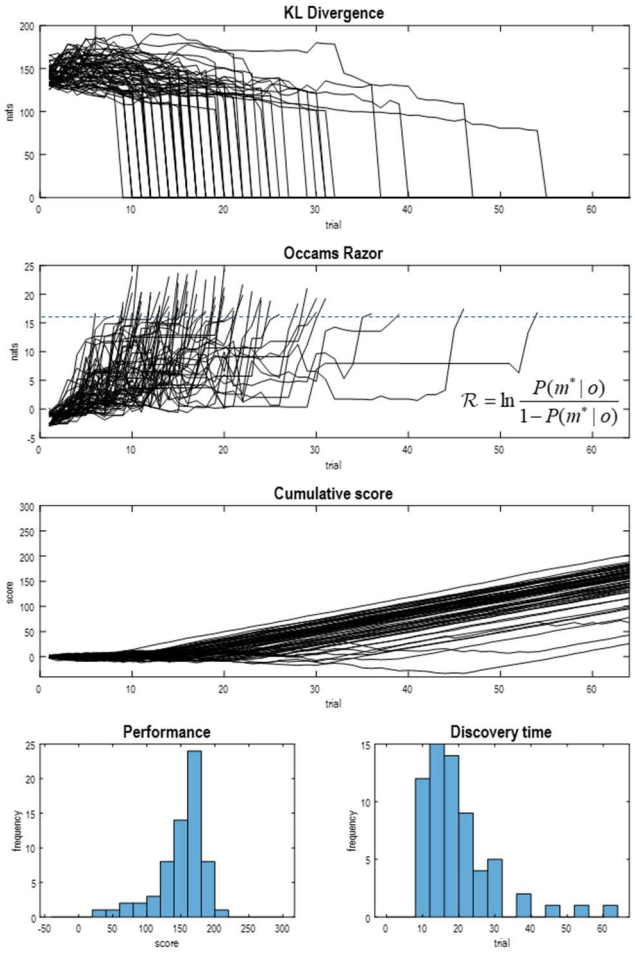

图5：假设检验。该图展示了64场游戏的模拟结果，每场游戏包含32次试验，采用前一幅图所述的设置。上方面板报告了反馈模态的似然映射后验分布与真实似然映射之间的 KL 散度。平均而言，散度逐渐减小，并在贝叶斯模型选择（即应用奥卡姆剃刀）时降至0。应用奥卡姆剃刀的依据是，对数优势比跨越阈值，该比值用于检验最可能模型相对于其他模型的概率。第三个面板报告累积分数，即正确与错误结果之差。直方图展示了累积分数的分布以及发现时间，即选出真实模型之前所经历的试验次数。请注意，在一场游戏中，奥卡姆剃刀过早地（在第六次试验时）被应用，导致智能体选错了模型。这意味着 KL 散度会增加，而发现时间则任意设为64。

贡献分析

图6直观展示了期望信息增益在规则发现或假设检验中的效能。左侧面板重现了图4中模型分布的演变过程——以及与之相伴的奥卡姆剃刀——其中期望信息增益涵盖模型、参数

和状态。对于完全相同的线索，随着从期望自由能中逐步移除信息增益项，其余面板给出了相应结果。当信息增益仅涉及参数和状态时，经过32次试验，主动模型选择仍无法在两个合理的假设之间作出区分。若只保留关于状态的期望信息增益，这一失效会更加明显：32次试验后，两个假设仍然同样合理。此外，经过15次试验后，智能体几乎选择了错误模型。在完全没有期望信息增益的情况下（即处于随机行为状态），第32次试验时奥卡姆剃刀几乎为零；这表明，从引出的结果中没有证据显示，获胜模型比其余模型更有可能。

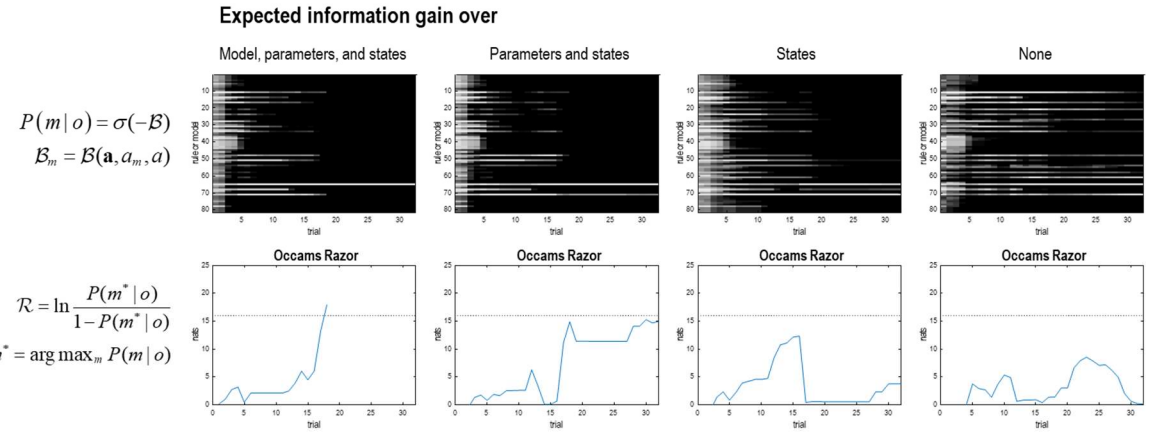

图6：期望信息增益与模型选择。此图重现了图4中的结果，但不包含关于模型、参数和状态的期望信息增益。上排展示了随着试验次数变化的模型后验分布；其中，关于模型、参数，最后是状态的期望信息增益依次被移除。下方各面板展示了移除期望信息增益的各个组成部分后，对奥卡姆剃刀的相应影响，以及由此导致的未能达到阈值的情况。

这些数值研究表明，尽管关于参数的期望信息增益足以排除许多不合理的模型，但在证据累积的后期，关于模型的期望信息增益会变得重要；也就是说，它有助于消除对少数几个同样合理的模型所存在的不确定性。接下来，我们将仔细考察不同类型信息增益的相对贡献。

我们重复了上述数值研究——每次试验使用相同的随机初始条件——但在计算期望自由能时，不考虑模型和参数方面的任何期望信息增益。信念更新和性能方面的差异总结于图7上排。第一幅图展示了有无模型参数方面的期望信息增益时的发现时间。每个点对应于在64次试验中，给定相同线索序列时，有无模型和参数方面的期望信息增益所对应的发现时间。平均而言，

发现时间有所不同。除五种情况外，模型和参数方面的期望信息增益均缩短了发现时间，使其从28.5次试验降至19.4次试验。发现时间较短意味着智能体能够利用其洞察，获得更多正确响应（并避免错误结果）。右上图显示，累积表现从89.2分提高至150分。

这种信息寻求行为的一个有趣之处在于，针对模型和参数的期望信息增益取决于能否消除对潜在状态的不确定性。这是因为，当智能体消除了对潜在状态的不确定性后，用于评估期望自由能的后验预测密度会更加精确。通过在有无期望信息增益的情况下重复模拟，可以说明由此产生的相互依赖关系。图7第二行的各面板展示了相应的发现时间和性能结果。在这种情况下，信息寻求带来的改善更为显著：约有一半模拟在64次试验后仍未能发现规则（有期望信息增益时唯一一次失败，是由于过早认定错误模型）。相应的性能平均得分从150分降至51分。这表明，消除对状态和模型参数的不确定性具有协同效应。

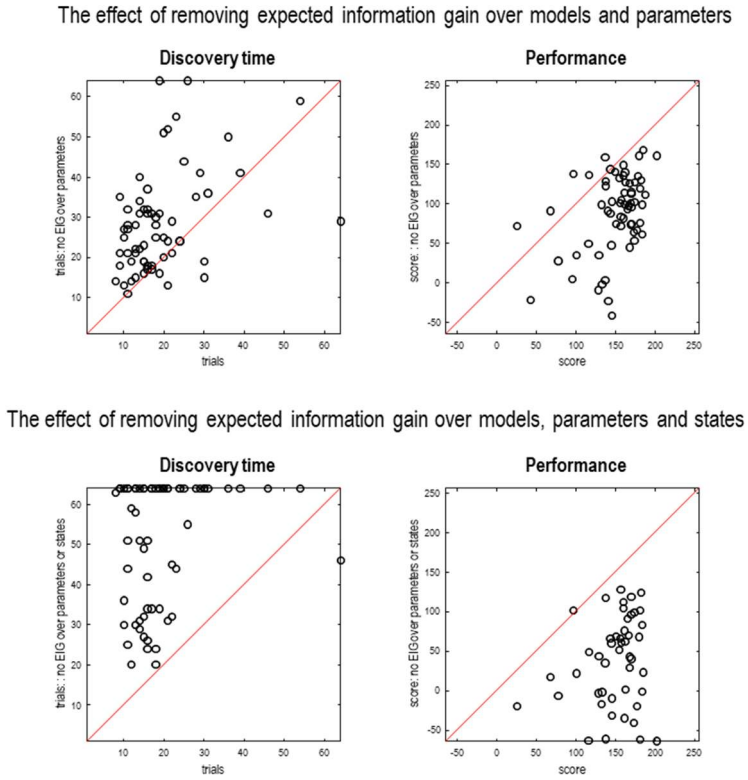

图7：好奇心的重要性。此图展示了在模型和参数（上排）以及模型、参数和状态（下排）上分别有无期望信息增益的情况下，重复图5中的模拟时的发现时间和性能。每个圆圈代表一场线索序列相同的游戏；因此，位于红线之上的圆圈构成

在缺乏期望信息增益的情况下，发现结果更差（性能结果则更好）。若64次试验后仍未发现规则，则任意将发现时间设为64。

结论

主动推断为有感知能力的行为提供了一种形式化解释：这类行为完全由最大化智能体与其感知到的世界之间交互的生成模型的证据所驱动。这有时被简洁地概括为自我证据化（Hohwy，2016）。从技术上说，这种基于第一性原理的自组织解释可以表述为一个自由能最小化过程，并跨越彼此可分离的时间尺度；其中，变分自由能为对数证据（亦称边际似然）提供上界。理解这一表述的一种有益视角是：通过最小化变分自由能，可以将一个难以处理的（精确）贝叶斯推断问题转化为优化问题（Da Costa 等人，2020；Feynman，1972；Hinton 和 Zemel，1993；Parr 和 Friston，2019）。

理解这种优化的一种有益视角，基于这样一个事实：这里所说的自由能是（近似）后验信念和先验信念的泛函；信念在此是指贝叶斯信念更新和信念传播意义上的信念（Bagaev 和 de Vries，2021；Winn 和 Bishop，2005；Yedidia 等人，2005）。这意味着，可以针对后验或——至关重要的是——先验来最小化变分自由能。如果将不同的模型结构与一组针对生成模型参数的先验约束关联起来，就能以直接的方式将结构学习纳入变分自由能的最小化过程。具体而言，可以利用贝叶斯模型约简，在特定模型下优化后验信念及其引发的行为，同时优化模型本身（Friston 和 Penny，2011）。本文探讨了应如何收集数据，以优化这类结构学习。

总之，从一组模型参数的先验出发，就可以对隐含的模型结构采用均匀先验，并随着证据（支持或反对不同模型）不断累积而更新后验信念。此外，如上所示，可以采样最能消除两个或多个合理模型之间歧义的数据。从技术上说，这只需扩展支撑动作选择的期望自由能，使其涵盖模型上的期望信息增益。我们通过贝叶斯模型约简来说明这一点：根据最优实验设计的原则（Lindley，1956），评估在特定策略下采取行动前后模型的后验分布，以计算期望信息增益。

这些程序通过解析推导和数值研究进行了展示，采用的是离散状态空间模型，即部分可观测马尔可夫决策过程。在这里，部分可观测性源于对感官世界采样的约束（例如，作为主动视觉，一次只能看一个地方）（Van de Maele 等人，2024）。在这种情境下使用贝叶斯模型约简，将

每个备选模型都视为均匀先验下贝叶斯模型平均的一个约简版本。换言之，每个模型都通过适当的收缩先验移除若干参数（或参数组合）来定义；参见（Efron 和 Morris，1973年；Friston 和 Penny，2011；Wipf 和 Rao，2007年）。在本研究采用的实现中，只需将狄利克雷计数设为零或较小的值（例如 1/32），即可应用收缩先验。这样便能快速、高效地对不同模型的似然进行评分，只需给定原始先验和一个可更新的后验分布即可。随着证据不断累积，关于约简模型的分布会越来越精确，直到某个模型脱颖而出，并可通过应用奥卡姆剃刀加以选定。

显然，无法保证所选择的模型能够解释未来遇到的数据。事实上，回看图4，64次模拟中有一次过早地采用了错误模型，仅经过六次试验便如此（即模型选择时 KL 散度上升的那次）。简言之，即使设定了严格的奥卡姆剃刀应用阈值，仍始终存在仓促下结论的可能。这种“仓促下结论”是一种特别值得关注的现象，关系到在精神分裂症研究中应用三球范式的可能性；计算表型分析表明，精神分裂症患者往往倾向于仓促下结论（Averbeck 等，2011年；Joyce 等，2013年；Moutoussis 等，2011年）。

如引言所述，这项工作——以及主动推理这一概念——在很大程度上得益于人工智能界对系统2思维的关注，尤其是 ARC-AGI-3 挑战。然而，本文所述的主动推理只涉及这一挑战的一小部分。首先，我们假设智能体知道未知规则所依托的生成模型；而不是必须通过对某种未分段的感觉输入（例如像素阵列）进行采样来学习生成模型。此外，我们将规则发现或抽象推理这一问题，简化为发现映射至规则定义结果的似然映射形式。这意味着，上文所展示的推理类型，以先学习世界的生成模型及对该世界采取行动所产生的后果为基础，然后才进行规则学习。

然而，如果生成模型的函数形式和结构已经学得，或已知，那么这种（因子）结构便可用来规定一类合规律的似然模型，这些模型实例化了一个未知的规则或程序。我们通过假设性的似然映射来介绍这种规定；这些映射具有某种上下文敏感对称性和同构，而它们是否适用于任何具有适当因子结构的生成模型尚未可知。无论上下文敏感的同构约束具有多大的普遍性，这些约束源自因子结构这一事实表明，我们进行推理所依据的规则，植根于对自然内在结构的剖析，其形式是条件独立因子，例如“是什么”和“在哪里”，或“物体”的属性。进而，这说明以对象为中心的表征和必要的因子模型至关重要，因为它们能够

应用各种归属于规则、对称性、等变性和定律的对称性或约束。

附录

区分贝叶斯模型平均下的推断，与推断支撑该平均的模型后验分布，这一点至关重要。在本文所讨论的事后方案中，推断和学习以最小化式（1）中的变分自由能为基础。

$$F = \mathbb{E}_{Q}[\ln \underbrace{Q(s_\tau)Q(u_\tau)Q(\mathbf{A})Q(m)}_{\text{后验分布}} - \ln \underbrace{P(o_\tau \mid s_\tau, \mathbf{A})}_{\text{似然}} - \ln \underbrace{P(s_\tau \mid s_{\tau-1}, u_\tau)P(u_\tau)P(\mathbf{A} \mid m)P(m)}_{\text{先验}}]$$

$$P(m) = a$$

$$a = \mathbb{E}_{P(m)}[a_m]$$

$$P(m) \propto a_m$$

这与推断模型本身不同。对模型进行推断，需要最小化包含模型近似后验的变分自由能：

$$\mathbf{F} = \mathbb{E}_{Q}[\ln \underbrace{Q(s_\tau)Q(u_\tau)Q(\mathbf{A})Q(m)}_{\text{Posterior}} - \ln \underbrace{P(o_\tau \mid s_\tau, \mathbf{A})}_{\text{Likelihood}} - \ln \underbrace{P(s_\tau \mid s_{\tau-1}, u_\tau)P(u_\tau)P(\mathbf{A} \mid m)P(m)}_{\text{Prior}}]$$

然而，在模型先验下进行推断和学习，即 $Q(m) = P(m)$，使得式 (1) 中的变分自由能成为模型推断所需变分自由能的下界近似。

$$\begin{aligned}
&= +\mathbb{E}[\quad - \quad] - \mathbb{E}[\quad | \quad - \ln(\,)\ln(\,)\ln(\,) \\
&\ln(\,)]\ Q\ Q\ F\ Q m\ P\ m\ P\ m\ P\ \mathbf{F}\ \mathbf{A}\ \mathbf{A}\ (\,)\ (\,)\ (\,) \\
&= -\mathbb{E}[\quad | \quad - \quad] \ln(\,)\ln(\,)(\,)([\quad])]\ Q\ Q\ Pm\ m\ Pm\ m \\
&= -\mathbb{E}\quad[\mathbb{E}\quad[\varphi\quad] - \varphi\ \mathbb{E} \\
&F\ P\ m\ P\ F\ a\ a_{\mathbf{A}}\ \mathbf{A}\ \mathbf{A}\ F \Rightarrow\ \geq \mathbf{F}
\end{aligned}$$

这基于詹森不等式，$\mathbb{E}[\varphi\quad] \leq \varphi\ \mathbb{E}(\,)(\quad[\quad])a\ \ a$，因为双伽马函数 ψ 在 $(0,\infty)$ 上严格凹，且 $0\ \ (\,)(\,)(\,)a\ \ a\ \ a\varphi\quad = \psi\quad - \psi$。从这一视角看，通过应用奥卡姆**'s**剃刀来更新模型的近似后验，可以视为对持续结构学习（即对模型进行推断）的一种离散（*post-hoc*）近似。有人或许会推测，在更新关于假设的后验信念之前，基于无信息模型先验不断累积证据，能够确保上述不等式成立。在实践中，*post hoc*方案意味着主动推断和学习期间无需同时考虑多个模型。

致谢

KF获得惠康信托基金资助（编号：226793/Z/22/Z）。TP获得NIHR学术临床奖学金（编号：ACF-2023-13-013）资助。

利益披露声明

作者没有需要披露的事项，也不存在利益冲突。

软件可用性声明

用于生成本文所报告模拟的 MATLAB 程序将通过 SPM 25 学术软件包公开提供（https://www.fil.ion.ucl.ac.uk/spm/），在正式发布前，读者可向作者索取。

参考文献

Attias, H., 2003.《基于概率推断的规划》，《人工智能与统计学》第9届国际研讨会论文集。Averbeck, B.B., Evans, S., Chouhan, V., Bristow, E., Shergill, S.S., 2011. 精神分裂症中的概率学习与推断。《精神分裂症研究》127, 115-122。Ay, N., Bertschinger, N., Der, R., Guttler, F., Olbrich, E., 2008. 自主机器人的预测信息与探索行为。《欧洲物理学杂志B》63, 329-339。Bagaev, D., de Vries, B., 2021.《用于可扩展贝叶斯推断的反应式消息传递》，第 arXiv:2112.13251 页。Barlow, H., 1961. 感觉信息转换背后可能遵循的原则，载于：Rosenblith, W.（编），《感觉通信》。麻省理工学院出版社，剑桥，马萨诸塞州，第217-234页。Barto, A., Mirolli, M., Baldassarre, G., 2013. 新颖性还是惊奇？《心理学前沿》4, 907。Beal, M.J., 2003.《近似贝叶斯推断的变分算法》。博士论文，伦敦大学学院。Berger, J.O., 2011.《统计决策理论与贝叶斯分析》。施普林格，纽约；伦敦。Botvinick, M., Toussaint, M., 2012. 作为推断的规划。《认知科学趋势》16, 485-488。Chollet, F., 2019.《论智能度量》，第 arXiv:1911.01547 页。Corcoran, A.W., Hohwy, J., Friston, K.J., 2023. 通过贝叶斯对抗性协作加速科学进步。《神经元》111, 3505-3516。Da Costa, L., Gavenciak, T., Hyland, D., Samiei, M., Dragos-Manta, C., Pattisapu, C., Razi, A., Friston, K., 2025. 对齐的结构学习智能体的可能原则。《神经计算》，1-47。Da Costa, L., Parr, T., Sajid, N., Veselic, S., Neacsu, V., Friston, K., 2020. 离散状态空间上的主动推断：综述。《数学心理学杂志》99, 102447。Dauwels, J., 2007. 因子图上的变分消息传递，2007年IEEE国际信息论研讨会，第2546-2550页。

Dickey, J.M., 1971.《加权似然比，关于正态位置参数的线性假设》。《数理统计年鉴》42, 204-&。Efron, B., Morris, C., 1973. Stein 估计规则及其竞争方法——一种经验贝叶斯方法。《美国统计学会杂志》68, 117-130。Feynman, R.P., 1972.《统计力学》。Benjamin，马萨诸塞州雷丁。Friston, K., Da Costa, L., Hafner, D., Hesp, C., Parr, T., 2021.《复杂推断》。《神经计算》33, 713-763。Friston, K., Da Costa, L., Sakthivadivel, D.A.R., Heins, C., Pavliotis, G.A., Ramstead, M., Parr, T., 2023. 路径积分、特定类型及奇异事物。《生命物理学评论》47, 35-62。Friston, K., Parr, T., Zeidman, P., 2018. 贝叶斯模型约简。arXiv预印本 arXiv:1805.07092。Friston, K., Penny, W., 2011. 事后贝叶斯模型选择。NeuroImage 56, 2089-2099。Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., Pezzulo, G., 2015. 主动推断与认知价值。《认知神经科学》，1-28。Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., Dolan, R.J., 2014. 选择的解剖学：多巴胺与决策。《伦敦皇家学会哲学汇刊》B辑：生物科学369。Friston, K.J., Lin, M., Frith, C.D., Pezzulo, G., Hobson, J.A., Ondobaka, S., 2017. 主动推断、好奇心与洞察。《神经计算》29, 2633-2683。Gershman, S.J., Niv, Y., 2010. 学习潜在结构：揭示自然的内在关联。《神经生物学观点》20, 251-256。Gros, C., 2009. 自主活动神经网络的认知计算：一个新兴领域。《认知计算》1, 77-90。Higgins, I., Chang, L., Langston, V., Hassabis, D., Summerfield, C., Tsao, D., Botvinick, M., 2021. 无监督深度学习揭示单个下颞叶面孔斑块神经元中的语义解缠。《自然·通讯》12, 6456。Hinton, G.E., Dayan, P., Frey, B.J., Neal, R.M., 1995. 无监督神经网络的“醒眠”算法。Science 268, 1158-1161。Hinton, G.E., Zemel, R.S., 1993. 自编码器、最小描述长度与亥姆霍兹自由能，《第6届国际神经信息处理系统会议论文集》。Morgan Kaufmann 出版公司，科罗拉多州丹佛，第3-10页。Hohwy, J., 2016.《自我证实的大脑》。《Nous》50, 259-285。Itti, L., Baldi, P., 2009. 贝叶斯惊奇度吸引人类注意力。《视觉研究》49, 1295-1306。Joyce, D.W., Averbeck, B.B., Frith, C.D., Shergill, S.S., 2013. 考察精神分裂症患者的信念与信心。《心理医学》43, 2327-2338。Kelly, M., Barron, A.B., 2022. 两个世界的优长：动物与人工智能中的双重推理系统。《认知》225, 105118。Kemp, C., Tenenbaum, J.B., 2008. 结构形式的发现。《美国国家科学院院刊》105, 10687-10692。Kiebel, S.J., Friston, K.J., 2011. 自由能与树突自组织。《系统神经科学前沿》5, 80。Lake, B.M., Salakhutdinov, R., Tenenbaum, J.B., 2015. 通过概率程序归纳实现人类水平的概念学习。Science 350, 1332-1338。Lindley, D.V., 1956. 关于实验所提供信息的度量。《数理统计年鉴》27, 986-1005。Linsker, R., 1990. 知觉神经组织——基于网络模型与信息论的一些方法。《神经科学年度评论》13, 257-281。Mackay, D.J.C., 1992. 用于主动数据选择的基于信息的目标函数。《神经计算》4, 590-604。

Mazzaglia, P., Verbelen, T., Catal, O., Dhoedt, B., 2022. 感知与行动的自由能原理：深度学习视角。Entropy（巴塞尔）24, 301. Moutoussis, M., Bentall, R.P., El-Deredy, W., Dayan, P., 2011. 妄想患者仓促下结论偏差的贝叶斯建模。Cogn Neuropsychiatry 16, 422-447. Olshausen, B.A., Field, D.J., 1996. 通过学习自然图像的稀疏编码，涌现简单细胞感受野特性。Nature 381, 607-609. Oudeyer, P.-Y., Kaplan, F., 2007. 什么是内在动机？计算方法类型学。Frontiers in Neurorobotics 1, 6. Parr, T., Friston, K., Zeidman, P., 2024. 主动数据选择与信息寻求。Algorithms 17, 118. Parr, T., Friston, K.J., 2017. 主动推断中的工作记忆、注意与显著性。Scientific Reports 7, 14678. Parr, T., Friston, K.J., 2019. 广义自由能与主动推断。Biological Cybernetics 113, 495-513. Parr, T., Pezzulo, G., Friston, K.J., 2022. 主动推断：心智、大脑与行为中的自由能原理。麻省理工学院出版社，剑桥。Penny, W.D., 2012. 使用赤池信息准则、贝叶斯信息准则和自由能比较动态因果模型。NeuroImage 59, 319-330. Riguzzi, F., Bellodi, E., Zese, R., 2014. 概率归纳逻辑编程史。Frontiers in Robotics and AI 第1卷 - 2014。Ruffini, G., 2017. 意识的算法信息论。Neuroscience of Consciousness 2017, nix019. Sakthivadivel, D.A.R., 2022. 高维稀疏耦合随机动力系统中的弱马尔可夫毯，p. arXiv:2207.07620. Sanchez, E.H., Serrurier, M., Ortner, M., 2019. 通过互信息估计学习解耦表示，p. arXiv:1912.03915. Savage, L.J., 1954. 统计学基础。Wiley，纽约。Schmidhuber, J., 2010. 创造力、乐趣与内在动机的形式理论（1990-2010）。IEEE自主心智发展汇刊 2, 230-247. Schwartenbeck, P., FitzGerald, T.H., Mathys, C., Dolan, R., Friston, K., 2015. 多巴胺能中脑编码对期望结果的预期确定性。Cereb Cortex 25, 3434-3445. Schwartenbeck, P., Passecker, J., Hauser, T.U., FitzGerald, T.H., Kronbichler, M., Friston, K.J., 2019. 好奇心与目标导向探索的计算机制。eLife 8, e41703. Silver, D., Singh, S., Precup, D., Sutton, R.S., 2021. 奖励已足够。人工智能 299, 103535. Tenenbaum, J.B., Kemp, C., Griffiths, T.L., Goodman, N.D., 2011. 如何发展心智：统计学、结构与抽象。Science 331, 1279-1285. Tervo, D.G.R., Tenenbaum, J.B., Gershman, S.J., 2016. 迈向结构学习的神经实现。《神经生物学观点》37, 99-105. Tipping, M.E., 2001. 稀疏贝叶斯学习与相关向量机。Journal of Machine Learning Research 1, 211-244. Tomasello, M., 2016. 文化学习再探。Child Development 87, 643-653. Tsividis, P., Loula, J., Burga, J., Arnaud, S., Foss, N., Campero, A., Pouncy, T., Gershman, S.J., Tenenbaum, J.B., 2021. 通过基于理论的建模、探索与规划实现人类水平强化学习。ArXiv abs/2107.12544. Ungerleider, L., 1994. 人脑中的“是什么”与“在哪里”。Current Opinion in Neurobiology 4, 157-165.

Van de Maele, T., Verbelen, T., Mazzaglia, P., Ferraro, S., Dhoedt, B., 2024. 基于主动推断的以对象为中心的场景表征。《神经计算》36, 677-704. Van Dijk, S.G., Polani, D., 2013. 信息约束驱动的目标导向行为组织。《复杂系统进展》16, 1350016. Wallace, C.S., Dowe, D.L., 1999. 最小消息长度与柯尔莫哥洛夫复杂度。《计算机杂志》42, 270-283. Winn, J., Bishop, C.M., 2005. 变分消息传递。Journal of Machine Learning Research 6, 661-694. Wipf, D.P., Rao, B.D., 2007. 解决同时稀疏逼近问题的经验贝叶斯策略。《IEEE信号处理汇刊》55, 3704-3716. Xu, F., Hao, Q., Shao, C., Zong, Z., Li, Y., Wang, J., Zhang, Y., Wang, J., Lan, X., Gong, J., Ouyang, T., Meng, F., Yan, Y., Yang, Q., Song, Y., Ren, S., Hu, X., Feng, J., Gao, C., Li, Y., 2025年. 迈向大型推理模型：大语言模型强化推理综述。《Patterns》（纽约）6, 101370. Yedidia, J.S., Freeman, W.T., Weiss, Y., 2005. 构建自由能近似与广义置信传播算法。《IEEE信息论汇刊》51, 2282-2312.