通过影响增强局部模型中的SCHR ¨ODINGER桥实现潜在视角采择

Kevin Alcedo [1,2], Pedro U. Lima [1,2], Rachid Alami [3,4]
[1]系统与机器人研究所，[2]高等技术学院T´ecnico，里斯本大学（里斯本 ELLIS 单元），[3]LAAS-CNRS，[4]图卢兹人工与自然智能研究所(ANITI){
kevin.alcedo,pedro.lima}@tecnico.ulisboa.pt，rachid.alami@laas.fr

arXiv:2602.02857v1 [cs.RO] 2 Feb 2026

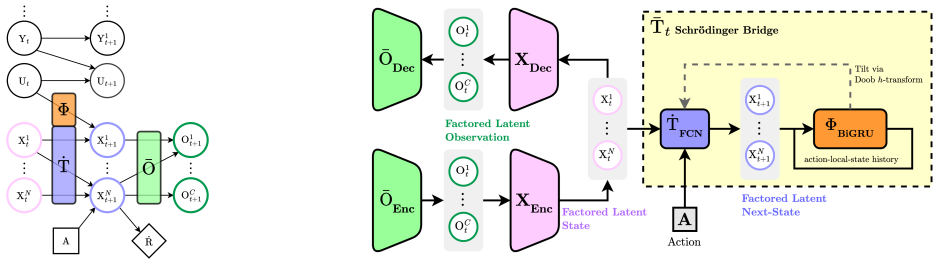
摘要

在与人类共同所处的环境中运行时，机器人需要在不确定性下做出决策。除外生动力学之外，它们还必须对他人隐藏的心理模型 和心理状态进行推理。尽管交互式部分可观测马尔可夫决策过程与贝叶斯心智理论的形式化具有原则性，精确的嵌套信念推断仍难以处理，而人工指定的模型在开放世界环境中也较为脆弱。我们通过学习结构化心理模型以及他人心理状态的估计器来同时解决这两个问题。在基于影响的抽象基础上，我们实例化一个影响增强局部模型，将具备社会意识的机器人任务分解为局部动力学、社会影响和外生因素。我们提出：（a）一个神经符号世界模型，用以实例化因子化的离散动态贝叶斯网络；以及（b）一个视角转换算子，将其建模为所学局部动力学上的摊销Schr¨odinger桥，用于将因子化的自我中心信念迁移为他者中心信念。我们表明，该架构使智能体能够在基于模型的强化学习中，经由决策时心理状态规划（信念空间中的Schr¨odinger桥）综合出社会感知策略，并在MiniGrid社会导航任务中给出了初步结果。

1 引言

社会感知机器人与人类和机器人并肩行动，在追求自身目标的同时，帮助他人实现目标，并照顾已知或推断出的需求。因此，它必须在部分可观测性条件下做出决策：既要应对环境的部分可观测性，也要应对他人的心理模型 和心理状态的部分可观测性。我们采用经典交互式部分可观测马尔可夫决策过程（I-POMDPs）（Gmytrasiewicz & Doshi, 2004）和贝叶斯心智理论（BToM）（Baker et al., 2011）的视角：心理模型 是（已知或推断出的）内部转移、观测、奖励和动作模型，而心理状态则是潜状态上的概率分布（信念）。知识表示领域通过本体/语义记忆（Milliez et al., 2014; Sarthou et al., 2019）、认知型人类感知任务规划器（Shekhar et al., 2025）以及动态认知逻辑（Bolander & Andersen, 2011; Bolander et al., 2021），将视角采择（Flavell, 1992）形式化，从而构建出可解释且可验证的系统。然而，在开放世界环境中，这些符号模型面对随机动力学时十分脆弱；精确的嵌套信念推理也很快变得难以处理：在第 0 层需要多项式空间，而在因子化动力学和指数级时域下，最坏情况下需要指数空间（Mundhenk et al., 2000）。相比之下，概率序贯决策天然能够处理不确定性，但通常假定模型已预先给定，仍受嵌套信念激增问题的困扰。

作为迈向统一符号知识表示与概率序贯决策的一步，我们提出一种数据驱动架构，用以学习（i）因子化离散心智模型，以及（ii）用于因子化潜在空间中视角采择的可处理信念到信念映射。其核心思想是将视角采择刻画为经由在已学习的局部参考动力学上定义的 Schr¨odinger 桥（SB）（Fortet, 1940）进行的信念传输，从而产生与局部动力学相容的他者中心信念。在基于影响的抽象（IBA）框架（Oliehoek 等人，2021年）的基础上，



(a) (b)

图 1：（a）影响增强局部模型；（b）神经符号世界模型架构。

我们实例化一个影响增强局部模型（IALM），将具备社会意识的机器人任务分解为局部动力学、社会影响和外生因素。我们借助神经符号世界模型，并将视角转换算子实现为在决策时刻在线运行的摊销 SB，从而使这一方案得以落地。受模拟理论（Goldman，1992年）启发，智能体植入一个心理状态，并复用其认知机制来模拟他人的心理状态动力学，从而产生一种“假如”他者中心信念，用于类似于心理状态空间中模型预测控制的决策。

在网格世界社会导航任务中的初步结果表明，基于薛定谔桥的视角采择学习速度更快，且取得的回报高于使用真实的无上下文当前信念。我们的贡献包括：（1）一种适用于社会任务的实用影响增强分解方法；（2）一种因子化的离散神经符号世界模型；（3）通过摊销薛定谔桥实现视角采择，从而能够在决策时刻进行基于模型的信念空间规划。

2 背景

2.1 影响增强局部模型

IBA 框架（Oliehoek 等，2021年）旨在应对结构化多智能体场景中对大型状态空间进行划分的挑战：它将全局模型因子化为规模更小、易于处理的局部模型，并识别能够简洁捕捉边界效应的影响点。在精确建模时，影响是一个充分统计量，使交互式主动学习模型（IALM）能够在固定外部策略的情况下计算出精确的最佳响应，同时避免对整个全局空间进行推理。在实践中，枚举所有影响源并更新其估计值难以处理，因此通常采用近似方法（Suau 等，2022年）。因子化表示明确呈现了概率依赖结构，并构成紧凑的两阶段动态贝叶斯网络（二位动态贝叶斯网络，2DBN）（Boutilier 等，1999；Poupart，2005）。

交互式主动学习模型将局部因子化部分可观测马尔可夫决策过程（FPOMDP）扩展为具有因子化全局状态 $\mathcal{S} = \{S_1, \ldots, S_k\}$ 的模型，该全局状态划分为以下变量：局部状态 $\mathcal{X}$、影响源 $\mathcal{U}$ 和非局部状态 $\mathcal{Y}$。设 $s_t \in S = \times_{i=1}^{k} S_i$, $x_t \in X = \times_{i \in \mathcal{X}} S_i$, $u_t \in U = \times_{i \in \mathcal{U}} S_i$ 和动作 $a_t \in \mathcal{A}$ 处于时间 t。Oliehoek 等人（2021年）的关键洞见是，最小的 d-分离集（Koller & Friedman，2009年） D_t，即动作-局部状态历史 $h_t = \langle x_1, a_1, \ldots, a_{t-1}, x_t \rangle$ 的子集，足以预测影响分布 $I(u_t \mid D_t)$。 D_{t+1} 由函数 $d(\cdot)$ 确定性地构造： $D_{t+1} = d(h_t)$。由此得到元组 $\mathcal{M}^{\text{IALM}} = \langle \bar{\mathcal{S}}, \mathcal{A}, \bar{T}, R, \mathcal{O}, \bar{O}, b_0 \rangle$，其中 $\bar{\mathcal{S}}$ 是增广局部状态空间 $\bar{s}_t = \langle x_t, D_{t+1} \rangle (x_{t+1} \mid \bar{s}_t, a_t)$ 是增广局部转移， $O(o_{t+1} \mid a_t, x_{t+1})$ 是带有 $o_{t+1} \in \mathcal{O}$ 的局部观测， $R(x_t, a_t, x_{t+1})$ 是局部奖励，而 b_0 是初始局部信念。影响通过边缘化调整局部转移：

$$\bar{T}(x_{t+1} \mid \bar{s}_t, a_t) = \sum_{u_{t+1} \in U} \Pr(x_{t+1} \mid x_t, a_t, u_{t+1}) \; I(u_{t+1} \mid D_{t+1}) \tag{1}$$

2.2 薛定谔桥

薛定谔桥（SB）最初旨在解释在假定的动力学模型下，运动粒子的起点与终点之间最可能的路径分布（即桥）。它可以表述为连接两个端点边缘分布的路径上的概率律（Schr¨odinger，1931年；

Chetrite 等，2021）；其目标是发现一种描述轨迹动力学的新概率律。薛定谔桥可视为最优传输的一个特例：基于参考动力学的动态熵正则最优传输（L´eonard，2014）；我们最小化相对于参考转移模型所诱导的轨迹分布的路径级 KL 散度。

给定参考局部动力学 $T(x_{t+1} \mid x_t,\ a_t)$、起点/终点边缘分布（信念） b_0 与 b_n，并以固定动作序列 $a_{0:n-1}$为条件，我们寻求一个桥$\bar{T} = \{\bar{T}_t\}_{t=0}^{n-1}$（一族随时间变化的转移），使其诱导的轨迹分布尽可能接近参考 T的轨迹分布。具体而言，我们计算 $\bar{T}$，使其诱导的轨迹分布 $\mathbb{P}_{\bar{T}}$在端点边缘分布（b_0，b_n）的约束下最小化KL$(\mathbb{P}_{\bar{T}} \,\|\, \mathbb{P}_{\dot{T}})$。由此得到一阶、时间非齐次马尔可夫链。此处我们关注时空 Sinkhorn 方法（Fortet，1940），其中前向势与后向势 $\phi_t,\ \psi_t \in \mathbb{R}_{>0}^{|\mathcal{X}|}$ 经由参考 $\dot{T}$传播并生成该桥。我们用上下文 c_t 对势进行参数化：$= D_{t+1} = d(h_t)$，其中 $h_t = \langle x_1,\ a_1,\ \ldots\ldots,\ a_{t-1},\ x_t \rangle$。Schrödinger系统变为：

$$\begin{aligned} \psi_t(x_t; c_t) &= \sum_{x_{t+1}} \dot{T}(x_{t+1} \mid x_t, a_t)\psi_{t+1}(x_{t+1}; c_{t+1}), \\ \phi_{t+1}(x_{t+1}; c_{t+1}) &= \sum_{x_t} \phi_t(x_t; c_t)\dot{T}(x_{t+1} \mid x_t, a_t), \\ p_t(x_t) &= \phi_t(x_t; c_t)\psi_t(x_t; c_t), \quad p_0 = b_0,\ p_n = b_n \end{aligned} \tag{2}$$

受控转移采用 Doob h-变换形式（Doob，1957）：

$$\bar{T}_t(x_{t+1} \mid \bar{s}_t, a_t) = \dot{T}(x_{t+1} \mid x_t, a_t)\frac{\psi_{t+1}(x_{t+1}; c_{t+1})}{\psi_t(x_t; c_t)}, \quad \text{where } \bar{s}_t := \langle x_t, D_{t+1} \rangle \tag{3}$$

诱导出的轨迹分布（初始状态分布为 μ_0）如下：

$$\mathbb{P}_{\bar{T}}(x_{0:n}) \propto \phi_0(x_0; c_0)\mu_0(x_0)\left[\prod_{t=0}^{n-1} \dot{T}(x_{t+1} \mid x_t, a_t)\right]\psi_n(x_n; c_n) \tag{4}$$

一种理解这一过程的方式是将其视为控制即推断（Todorov，2006），旨在以最小代价引导噪声被动动态过程。在第 3 节中，我们将交互式主动学习模型实例化为一个神经-符号世界模型，用于学习（$\mathcal{S}$、T、$\mathcal{O}$、O、D）。随后，我们通过对用于视角采择的 T 进行摊销式 Schrödinger 桥倾斜，将 $T = \{T_t\}$ 实现为式 1 中理想交互式主动学习模型转移的近似。势函数 ϕ 和 ψ 通过认知反事实进行离线训练。

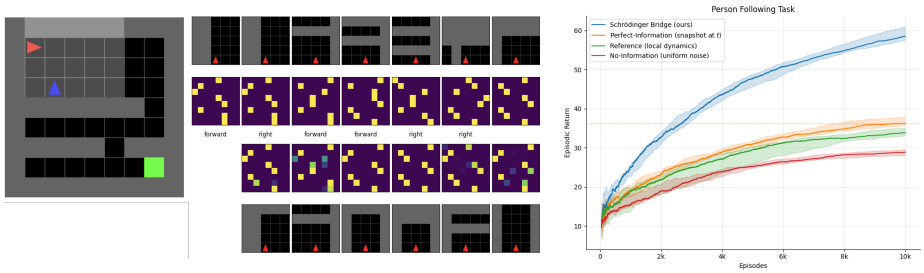
3 方法

3.1 神经符号世界模型

我们将交互式主动学习模型实例化为一个神经符号世界模型，该模型由一组层级结构的离散变分自编码器（离散变分自编码器）组成，采用 Gumbel-Softmax（Jang 等，2017；Maddison 等，2017），并在转移（o_t、a_t、o_{t+1}）上进行端到端训练，显式建模二维动态贝叶斯网络的结构组件（见图 1）。原始观测被编码为形状为（C、F）的潜变量，其中 C 是离散观测因子的数量，每个因子有 F 个类别。随后，这些变量被映射为因子化状态潜变量（N、K），其中有 N 个状态因子，每个因子有 K 个类别。局部、无影响的转移模型 T 将当前因子化状态和一个独热动作映射为下一个因子化潜状态。该转移模型使用 α-entmax（Peters 等，2019）交叉注意力处理输入状态因子，从而诱导出精确为零、由数据驱动的稀疏性，并利用潜在父子结构。

3.2 视角转换算子

扩展 Alcedo 等人（2025年）的工作，我们通过基于 SB 的视角转换算子估计他人的信念：ˆtor: $b_i = \Phi_{\mathrm{SB}}(b_0,\ \hat{x}_i,\ a_{0:n-1},\ T)$，其中 b_0 是自我中心信念，$\hat{x}_i$ 是对另一智能体 i 的局部观测相对状态，T 是学习得到的局部且不受影响的转移模型（参考动力学），$a_{0:n-1}$ 是为智能体 i 假设的候选动作序列，该序列由上游



(a) (b)

图 2：（a）视角转换表征。左：全局状态。右：真实观测（上）；因子化的离散潜状态（第二行）；动作（第三行）；下一步潜在状态（第四行）；重建的观测（最后一行）。（b）5 个随机种子上的回合回报中位数，阴影区域表示中位数的 95% 自助法置信区间；虚线：完美信息基线的最终中位数。

ˆ局部规划器生成，b_i 是估计得到的他者中心信念。我们通过使用带有 α-entmax 注意力的双向序列模型（双向门控循环单元，BiGRU）学习势函数（ϕ_t，ψ_t），从而实现 Φ_{SB}。该模型以通过使用参考模型 T进行滚动展开生成的动作-局部状态历史 $h_t = \langle x_1, a_1, \ldots\ldots, a_{t-1}, x_t \rangle$ 为输入。α-entmax 在 h_t 上诱导的稀疏性，作为确定性更新，用于根据动作-局部状态历史 $D_{t+1} = d(h_t)$计算近似的最小d集。训练时，我们从¯ ˙环境中的单智能体任务采样“认知反事实”（他者中心信念） b_n，以摊销能够通过对 T 应用 Doob h-变换来实现 T 的势函数，并在端点约束下最小化 $\mathrm{KL}\big(\mathbb{P}_{\bar{T}} \,\|\, \mathbb{P}_{\dot{T}}\big)$。推理时，我们使用由 c_t 参数化的摊销势函数（ˆϕ_t，ˆψ_t）ˆ对 $\dot{T}$ 进行倾斜：$= D_{t+1} = d(h_t)$，以获得对 b_i的估计。学习目标在逐因子层面上实现，使 $d(h_t)$能够通过屏蔽因子来上下文式地管理倾斜。目标函数为：$\mathcal{L} = \mathcal{L}_{\bar{O}} + \mathcal{L}_{\dot{T}} + \mathcal{L}_{\Phi}$。我们采用将在完整论文中讨论并进行消融实验的训练稳定化技术，包括：Gumbel-Softmax τ 退火、知识蒸馏置信度门控，以及两个推理层级上的社会加权认识论反事实目标。

4 初步结果、局限性与后续步骤

我们在多智能体、基于模型的强化学习环境中评估了用于影响感知策略学习的方法。**环境：** 部分可观测Minigrid（Chevalier-Boisvert 等人，2023年），具有离散观测和动作。**跟随人员任务：** 跟随一个固定策略的智能体前往未知目标，同时遵守社会规范（Alcedo 等，2025年）。环境会根据接近程度和可见性给予奖励，到达目标时给予大量终止奖励，对碰撞施加惩罚，并在跟随者长时间距离过远时截断回合。**影响感知策略：** $a_t = \pi_{\mathrm{DQN}}(b_t^0, \hat{b}_t^i)$。**基线：** （i）完美信息：对另一个智能体当前信念的真实估计，不含上下文或历史信息：ˆ $b_t^i = b_t^i$；（ii）无信息：ˆ $b_t^i \sim \mathrm{Uniform}(\Delta_X)$；以及（iii）参考（局部动态）多步滚动展开：ˆ$b_t^i = \prod^{n-1} \dot{T}(\cdot)$。**结果：** 薛定谔桥（SB）的学习速度更快，且回报高于所有基线（见图 2），这凸显了决策时刻进行上下文感知心理状态规划相较于无上下文估计的优势。目前，我们的方法假设其他智能体的感知模型和动作模型具有同质性；未来我们将通过异质认识论反事实调整信念到信念变换目标，以解决这一问题。

5 相关工作

潜在视角采择已通过元学习智能体嵌入（Rabinowitz 等，2018）、生成他者中心观测的潜在状态空间世界模型（Chen 等，2023），以及部分可观测条件下的潜在对手／队友建模（Labash 等，2020；Papoudakis 等，2021）加以研究。与此不同，我们将视角采择视为信念到信念传输，在学习得到的 IALM 中实现摊销 SB，以生成用于决策时心理状态规划的他者中心信念；我们扩展了 Alcedo 等（2025）的仅局部模型方法。Suau 等

（2022）在多智能体强化学习场景中大规模实例化交互式语言模型，并采用预先指定的影响源和学习得到的预测器；Suau 等（2025）则在像素空间中学习影响感知记忆和 d-set选择器（我们的方法则在潜在空间中运行）。最后，Wang 等（2023）提出了用于师生学习场景的广义信念传输方法，该方法通过静态非平衡熵最优传输进行建模；而我们使用薛定谔桥。

致谢

本研究得到 LARSyS 的 FCT 资助（DOI：10.54499/LA/P/0083/2020、10.54499/UIDP/50009/2020 和 10.54499/UIDB/50009/2020），以及通过项目 C645008882-00000055、负责任人工智能中心获得的葡萄牙恢复与韧性计划（PRR）资助；并部分得到法国国家研究署（法国国家研究署）AAPG-2024/Ostensive 项目的资助；此外，还得到欧洲学习与智能系统卓越计划（ELIAS，资助协议编号：101120237）的资助。

参考文献

Kevin Alcedo、Pedro U. Lima 和 Rachid Alami。《视角转换的神经符号世界模型：社会感知机器人导航框架》。发表于*IEEE*机器人与人类交互通信国际研讨会（*RO-MAN*）论文集，荷兰埃因霍温，2025年。已录用发表。

Chris Baker、Rebecca Saxe 和 Joshua Tenenbaum。贝叶斯心智理论：对联合信念—愿望归因进行建模。载于认知科学学会年会论文集，第 33 卷，2011年。

Thomas Bolander 和 Mikkel Birkegaard Andersen。单智能体和多智能体系统的认知规划。《应用非经典逻辑期刊》，21：34–9，2011年。

Thomas Bolander、Lasse Dissing 和 Nicolai Herrmann。用于人机协作的基于DEL的认知规划：理论与实现。载于第 *18* 届知识表示与推理原理国际会议论文集，第 120–129 页，2021年11月。DOI：10.24963/kr.2021/12。网址：https://doi.org/10.24963/kr.2021/12。

Craig Boutilier、Thomas Dean 和 Steve Hanks。决策论规划：结构假设与计算优势。《人工智能研究期刊》，11(1)：1–94，1999年7月。ISSN 1076-9757。

Kaiqi Chen、Jing Yu Lim、Kingsley Kuan 和 Harold Soh。《用于人机交互的潜变量发射增强视角采择》（LEAPT）。载于*2023*年 *IEEE/RSJ* 国际智能机器人与系统会议（*IROS*），第 8006–8013 页，2023年。DOI：10.1109/IROS55552.2023. 10342126。

Rapha¨el Chetrite、Paolo Muratore-Ginanneschi 和 Kay Schwieger。E. Schr¨odinger 的 1931年论文《论自然法则的逆转》。《欧洲物理学杂志H》，46:28，2021年。DOI：10.1140/epjh/s13129-021-00032-7。

Maxime Chevalier-Boisvert、Bolun Dai、Mark Towers、Rodrigo Perez-Vicente、Lucas Willems、Salem Lahlou、Suman Pal、Pablo Samuel Castro 和 J Terry。MiniGrid 与 MiniWorld：面向目标导向任务的模块化、可定制强化学习环境。载于《神经信息处理系统进展》，第 36 卷，第 73383–73394 页。Curran Associates, Inc.，2023年。

J.L. Doob。《调和函数的条件布朗运动与边界极限》。法国数学会公报´´，85:431–458，1957年。DOI：10.24033/bsmf.1494。网址：https://www.numdam.org/articles/10.24033/bsmf.1494/。

John H. Flavell. 关于观点采择的视角。，第107–139页。让·皮亚杰研讨会系列。Lawrence Erlbaum Associates, Inc，Hillsdale，新泽西州，美国，1992年。ISBN 0-8058-1050-1（精装）。

Robert Fortet. R´esolution（求解）d'un syst`eme（系统）d'´equations（方程），M. Schr¨odinger（薛定谔先生）。*Journal de Math*（《数学杂志》）´*ematiques Pures et Appliqu*（纯粹与应用）´*ees*，9e s´erie（第9辑），19(1-4):83–105，1940年。网址 https://www.numdam.org/ item/JMPA_1940_9_19_1-4_83_0/。

P.J. Gmytrasiewicz与P. Doshi。交互式部分可观测马尔可夫决策过程：性质与初步结果。载于第三届自主智能体与多智能体系统国际联合会议论文集，*2004*年。*AAMAS 2004*，第1374–1375页，2004年。

Alvin I. Goldman。为模拟理论辩护。《心智与语言》，7:104–119，1992年。DOI: https://doi.org/10.1111/j.1468-0017.1992.tb00200.x.

Eric Jang、Shixiang Gu与Ben Poole。基于Gumbel-Softmax的分类重参数化。载于第五届国际学习表征会议（*ICLR*）论文集，法国土伦，2017年。

Daphne Koller与Nir Friedman。概率图模型：原理与技术——自适应计算与机器学习。麻省理工学院出版社，2009年。ISBN 0262013193。

Aqeel Labash、Jaan Aru、Tambet Matiisen、Ardi Tampuu和Raul Vicente。《深度强化学习智能体中的视角采择》。《计算神经科学前沿》，第14卷（2020），2020年。ISSN 1662-5188。DOI：10.3389/fncom.2020.00069。URL https://www.frontiersin.org/journals/computational-neuroscience/articles/10.3389/fncom.2020.00069。Christian L´eonard。《Schr¨ödinger问题及其与最优输运若干联系的综述》。《离散与连续动力系统》，34(4):1533–1574，2014年。ISSN 1078-0947。DOI：10.3934/dcds.2014.34.1533。URL https://www.aimsciences.org/article/id/d5bcf817-901d-4104-b7da-eade7847c53e。Chris J. Maddison、Andriy Mnih和Yee Whye Teh。《Concrete分布：离散随机变量的一种连续松弛》。载于第五届国际学习表征会议（*ICLR*）论文集，法国土伦，2017年。Gr´égoire Milliez、Matthieu Warnier、Aur´élie Clodic和Rachid Alami。《赋予交互式机器人关于视角采择和信念管理的推理能力的框架》。载于第*23*届*IEEE*机器人与人类交互通信国际研讨会（*IEEE RO-MAN 2014*），英国爱丁堡，*2014*年*8*月*25*日至*29*日，第1103–1109页。IEEE，2014年。DOI：10.1109/ROMAN.2014.6926399。Martin Mundhenk、Judy Goldsmith、Christopher Lusena和Eric Allender。《有限时域马尔可夫决策过程问题的复杂性》。《美国计算机学会会刊》，47(4):681–720，2000年7月。ISSN 0004-5411。DOI：10.1145/347476.347480。URL https://doi.org/10.1145/347476. 347480。Frans Oliehoek、Stefan Witwicki和Leslie Kaelbling。《结构化多智能体环境中影响的充分统计量》。人工智能研究期刊，70:789–870，2021年5月。ISSN 1076-9757。DOI：10.1613/jair.1.12136。Georgios Papoudakis、Filippos Christianos和Stefano Albrecht。《部分可观测条件下用于深度强化学习的智能体建模》。载于M. Ranzato、A. Beygelzimer、Y. Dauphin、P.S. Liang和J. Wortman Vaughan（编），神经信息处理系统进展，第34卷，第19210–19222页。Curran Associates, Inc.，2021年。URL https://proceedings.neurips.cc/paper_files/paper/2021/file/a03caec56cd82478bf197475b48c05f9-Paper.pdf。Ben Peters、Vlad Niculae和Andr´é F. T. Martins。《稀疏序列到序列模型》。载于Anna Korhonen、David Traum和Llu´ís M`árquez（编），第*57*届计算语言学协会年会论文集，第1504–1519页，意大利佛罗伦萨，2019年7月。计算语言学协会。DOI：10.18653/v1/P19-1146。URL https: //aclanthology.org/P19-1146/。Pascal Poupart。利用结构高效求解大规模部分可观测马尔可夫决策过程。博士论文，多伦多大学，加拿大，2005年。AAINR02727。

Neil Rabinowitz、Frank Perbet、Francis Song、Chiyuan Zhang、S. M. Ali Eslami和Matthew Botvinick。《机器心智理论》。载于Jennifer Dy和Andreas Krause（编），第*35*届国际机器学习大会论文集，第80卷《机器学习研究论文集》，第4218–4227页。PMLR，2018年7月10日至15日。URL https://proceedings. mlr.press/v80/rabinowitz18a.html。

Guillaume Sarthou、Aur´elie Clodic 和 Rachid Alami。《Ontologenius：面向机器人智能体的长期语义记忆》。载于*2019* 年第 *28* 届 *IEEE* 国际机器人与人类交互通信会议（*RO-MAN*），第 1–8 页，2019 年。DOI：10.1109/RO-MAN46459.2019.8956305。

Erwi ¨n Schr¨odinger。《论自然定律的反演》。普鲁士科学院会议报告，物理数学部，第 144–153 页，1931年。

Shashank Shekhar、Anthony Favier 和 Rachid Alami。《一种能够预测人类信念和决策的认知型人类感知任务规划器》。载于《社会机器人学》：第 *16* 届国际会议 *ICSR + AI 2024*，欧登塞，丹麦，*2024*年*10*月*23*日至*26*日，会议论文集，第二部分，第 208–222 页，柏林、海德堡，2025年。施普林格出版社。ISBN 978-981-96-3518-4。DOI：10.1007/ 978-981-96-3519-1 20。URL https://doi.org/10.1007/978-981-96-3519-1_20。

Miguel Suau、Jinke He、Matthijs T. J. Spaan 和 Frans Oliehoek。影响增强型局部模拟器：为大型网络化系统中的快速深度强化学习提供可扩展的解决方案。载于第*39*届国际机器学习大会，第162卷《机器学习研究论文集》，第20604–20624页。PMLR，2022年7月17–23日。

Miguel Suau、Jinke He、Elena Congeduti、Rolf A. N. Starre、Aleksander Czechowski 和 Frans A. Oliehoek。面向部分可观测马尔可夫决策过程的深度强化学习的影响感知记忆架构。《神经计算与应用》，37(19):13145–13161，2025年7月。ISSN 1433-3058。DOI：10.1007/s00521-022-07691-7。URL https://doi.org/10.1007/ s00521-022-07691-7。

Emanuel Todorov. 线性可解马尔可夫决策问题。载于 B. Sch¨olkopf、J. Platt 和 T. Hoffman（编），《神经信息处理系统进展》，第 19 卷。麻省理工学院出版社，2006年。网址：https://proceedings.neurips.cc/paper_files/paper/ 2006年/file/d806ca13ca3449af72a1ea5aedbed26a-Paper.pdf。

Junqi Wang、Pei Wang 和 Patrick Shafto。广义信念传输。载于第*37*届神经信息处理系统国际会议论文集，NIPS ’23，雷德胡克，纽约州，美国，2023年。Curran Associates Inc.