探索增强多智能体通信系统中的合作

Zhao Song、Chen Shen、Zhen Wang，*IEEE* 会士，以及 The Anh Han

arXiv:2603.01401v1 [cs.MA] 2 Mar 2026

摘要—**为多智能体系统设计能够促进合作的协议，仍是一项重大挑战。廉价交谈是指在正式采取行动之前进行的无成本、无约束力的通信，是解决这一问题的关键方案。然而，为了便于分析，现有理论框架往往排除随机探索或噪声，因此其对系统性能的实际影响在很大程度上尚未得到探究。为弥补这一不足，我们提出一种将信号传递与捐赠博弈相结合的两阶段演化博弈论模型，并将探索明确纳入决策过程。我们在不同拓扑结构上的基于智能体的模拟表明，存在一个能使全系统合作水平达到最高的普适最优探索率。从机制上看，适度探索会削弱背叛的稳定性，并催生自组织合作联盟，从而促成其周期性成功。此外，合作峰值的形成得益于振荡周期与放大效应之间的精妙平衡。我们的研究结果表明，与其追求确定性刚性，不如拥抱策略探索，将其作为一种人为设计的随机性；这对于维持合作并实现基于通信的智能系统的最优性能至关重要。**

索引词—**廉价交谈；随机探索；合作；多智能体系统；循环支配**

I. 引言

实现自主智能体之间的协调，仍然是多智能体系统（MAS）领域的一项核心挑战 [1], [2], [3], [4], [5]。在许多去中心化环境中，自利智能体往往面临两难：最大化局部效用与实现系统级最优性之间存在冲突，这种情形在数学上被形式化为社会困境。若缺乏有效的干预机制，此类系统很容易陷入不合作的“纳什均衡”，导致集体表现不佳 [6], [7]。过去几十年来，研究人员提出了各种结构性和算法性方案来打破这些僵局，涵盖声誉 [8], [9], [10], 标签中介 [11], [12], [13], 以及激励机制 [14], [15], [16], [17], [5]，以实现可持续的合作结果。此外，由人类和智能机器构成的混合系统日益受到关注，以确保合作关系 [18], [19], [20], [21], [22]。

在这些解决方案中，廉价交谈作为一种无成本、无约束力的信号传递协议，已成为一种关键的促进自适应协调的机制 [23], [24], [25]。不同于高成本的执行机制，廉价交谈能够在行动前促进“意图共享”，充当一种轻量级的信息交换渠道，几乎不会带来计算或传输开销 [26], [27], [28]。在工程应用中，此类协议广泛用于解决资源冲突和同步行为，典型例子包括自动化交易市场中的协商，以及自主车辆群中的避碰信号传递 [29], [30]。此外，理论模型和基于智能体的仿真均一致表明，实施此类博弈前通信层能够显著促进合作行为的涌现，在不改变底层物理约束的情况下，重塑系统的收益格局 [31], [32], [33]。

然而，为了便于分析，现有大多数研究廉价交谈的理论框架都建立在探索极少发生这一假设之上 [34], [35], [36]。在此语境下，探索指随机偏离既定或最优行为模式。相比之下，随机性是现实世界系统的内在属性，其表现形式多种多样，例如进化过程中的基因突变，或多智能体学习系统中的误差 [37], [38], [39]。一方面，理论研究一再表明，非零水平的探索对于防止进化系统陷入吸收态是必需的，或者能够显著改变均衡结果 [40], [41], [42]。另一方面，引入廉价交谈本身会拓展策略格局，使策略不确定性普遍存在。因此，一个关键的科学问题随之而来：在存在随机探索的情况下，廉价交谈如何维持合作？具体而言，这种随机性如何影响不同合作策略的进化成功？

为弥补这一关键空白，我们在演化博弈框架下建立了一个理论模型，将探索纳入两阶段廉价交谈博弈的决策过程。具体而言，我们的模型整合了一个博弈前通信阶段，在该阶段玩家决定是否发送合作信号；以及随后的博弈中决策阶段，在该阶段玩家参与捐赠博弈[43]。我们分析了由八种可能行为构成的完整策略空间，并将其划分为直觉策略（简单一致性）和审慎策略（条件推理），以反映不同的认知过程。此外，我们在群体更新过程中引入了探索，即行为误差：玩家可能随机采用任意策略，而不是模仿成功的榜样。需要明确的是，在本研究中，探索这一术语不同于

Z. Song 和 T.A. Han 任职于英国米德尔斯堡蒂赛德大学计算、工程与数字技术学院。（电子邮件：z.song@tees.ac.uk；t.han@tees.ac.uk）。C. Shen 任职于日本福冈 816-8580 九州大学工程科学学院。（电子邮件：steven shen91@hotmail.com）。Z. Wang 任职于中国西北工业大学网络安全学院和人工智能、光学与电子学院（iOPEN）。（电子邮件：w-zhen@nwpu.edu.cn）。

感知噪声或响应噪声，后两者通常出现在信号传输或决策中。我们的模型建立在以下研究所确立的框架之上： [36], 其中采用最小程度的探索来减轻有限尺寸效应；而本研究则将探索提升为核心策略变量，重点关注探索过程本身的影响。

基于这一理论模型，我们在结构化种群中开展基于智能体的仿真，重点关注含噪廉价交谈情境下合作的涌现。首先，我们证实，即便存在随机干扰，廉价交谈与网络互惠相结合，仍是维持合作的稳健途径。至关重要的是，我们发现存在一个非零的最优策略探索率，可使系统的整体合作水平达到最大。这一最优探索率有助于直觉型与审慎型合作策略稳定地取得成功。从机制上看，适度探索通过瓦解根深蒂固的背叛簇发挥作用，同时使合作策略能够组织成紧凑、具有韧性的空间簇。最后，我们指出，尽管探索是一种强有力的优化力量，但它并非万能之策；在严格条件下，例如极端困境强度，或廉价交谈失去效用的情形中，最优探索率的收益会减弱。

我们的主要贡献概括如下:

- 通过放宽“探索罕见”这一传统假设，我们提出一个两阶段演化博弈模型，其中纳入了一个刻画内在随机性的受控变体。这使得我们能够严格量化从确定性刚性到随机协调的转变。
- 我们从理论上证明，在非零探索率下，合作能够在多智能体通信系统中稳健地持续存在。这一结果弥合了理想化无噪声模型与现实世界复杂系统之间的差距，因为不确定性在现实系统中不可避免。
- 我们揭示了合作的循环成功机制。与没有探索时的静态均衡不同，这一由探索驱动的机制通过周期性自组织过程确保合作的韧性。我们表明，紧密合作集群的形成能够有效打破背叛者的支配地位，防止系统陷入非合作陷阱。
- 我们确定了一个能够使全系统合作水平最大化的最优探索率，这得益于周期性振荡与放大效应之间的微妙平衡。此外，我们的研究结果强调了多样性的关键作用，指出全局协调依赖于受控随机性，而非确定性稳定。

本文其余部分安排如下。在第II节，我们介绍预备知识。第III 节介绍模型与方法。在第 IV节，我们展示探索如何影响基于主体的仿真结果。在第V 节和VI节，我们讨论并总结本文的研究结果。

II. 预备知识

社会困境通常源于集体利益与个体动机之间的根本冲突。为定量分析这一冲突，研究人员经常采用双人双策略博弈作为基础理论模型。在此框架下，两名参与者同时选择合作（C）或背叛（D）。双方互动的结果由四个标准收益参数定义：相互合作时的奖励 R 、相互背叛时的惩罚 P 、背叛者利用合作者时的诱惑 T ，以及合作者被背叛者利用时的受骗者收益 S 。

基于 T， R， P, 与 S的相对排序，该框架涵盖四种基本博弈模型，每种模型均以不同的纳什均衡（NE） [44]为特征。在满足 $T > R > P > S$的囚徒困境博弈（PD）中，背叛是严格占优策略，从而导向位于（D， D）的唯一帕累托无效率纳什均衡。相反，满足 $T > R > S > P$ 的雪堆博弈（SDG）不存在占优策略，通常形成混合策略纳什均衡或稳定多态。满足 $R > T > P > S$ 的鹿猎博弈（SHG）呈现出协调困境，其特征是两个纯策略纳什均衡：收益占优的（C， C）和风险占优的（D， D）。最后，满足 $R > T > S > P$的和谐博弈（HG）促进无缝协调，其唯一纳什均衡自然收敛于相互合作（C， C）。

在这些模型中，捐赠博弈被广泛视为研究合作演化的标准范式，本质上是囚徒困境的一种对称且简化的形式。在捐赠博弈中，合作者付出成本 c，为对手提供收益 b（其中 $b > c$）；而背叛者既不付出成本，也不提供任何收益。这使收益矩阵简化为 $R = b-c$、 $T = b$、 $S = -c$ 和 $P = 0$。为进一步推广分析，引入困境强度 $r = c/(b-c)$。该参数有效刻画了合作的难度：较大的 r 表示困境强度较高，进化压力会抑制合作的存续；较小的 r 则表示困境强度较低，有利于合作涌现 [45][46]。

将 $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ 记为拓扑结构，其中 $\mathcal{V} = \{1, 2, \ldots, \mathcal{N}\}$ 指节点集， $\mathcal{E}$指连接集。将 $a_{ij} \in \mathbb{R}$ 记为邻接矩阵的元素，其中 $a_{ij} = 1$ 表示玩家 i 与玩家j相连，而 $a_{ij} = 0$ 表示两者不相连。此处我们考虑无向网络，每个节点的度为 $k_i = \sum_{j=1}^{N} a_{ij}$ $\mathcal{G}$ $k_i = k_j$ $\forall i$ $j \in \mathcal{V}$ 。若,,, 则为同质网络，否则为异质网络。具体而言，当 $k_i = N-1$, $\forall i \in \mathcal{V}$时， $\mathcal{G}$ 是一个充分混合网络，或完全图。

III. 模型与方法

A. 两阶段廉价交谈博弈

我们将这种互动建模为一个包含博弈前交流和博弈中决策的两阶段演化博弈，如图 1(a) 所示。第一阶段，玩家选择是否发送合作信号（A），或保持沉默（N）。这种信号属于廉价交谈：不具约束力，也不产生固有的传输成本。第二阶段，玩家参与捐赠博弈，根据第一阶段交换的信号，同时选择合作（C）或背叛（D）。

表 I：由三元组 (u,p,q) 定义的策略空间。 $u=1$ 表示主动发信号， $u=0$ 表示沉默。 p 和 q分别表示博弈中对信号或沉默作出的行动。

策略	信号 (u)	对信号的响应(p)	对沉默的响应(q)	策略描述	策略类型
ACC	1（A）	1（C）	1（C）	无条件合作	直觉策略
ACD	1（A）	1（C）	0（D）	条件合作	审慎策略
ADC	1（A）	0(D)	1(C)	悖论式合作	审慎策略
ADD	1(A)	0(D)	0（D）	战略背叛	审慎策略
NCC	0（N）	1（C）	1（C）	沉默合作	审慎策略
NCD	0（N）	1（C）	0（D）	有条件背叛	审慎策略
NDC	0（N）	0（D）	1（C）	虚假合作	审慎策略
NDD	0（N）	0（D）	0（D）	无条件背叛	直觉策略

形式上，玩家 i的策略可定义为一个三元组 $S_i=(u_i,p_i,q_i)\in\{0,1\}^3$，其中 u_i表示信号（1 表示 A，0 表示N）， p_i 表示共同参与者发出信号 A 时采取的行动（1 表示 C，0 表示D），而 q_i 表示共同参与者保持沉默 N 时采取的行动（1 表示 C，0 表示D）。这种表述产生了 $2^3=8$ 种不同的策略，列于表 I 中。

我们根据双重过程理论将这些策略分为两类，以体现推理的复杂性：

- **直觉策略($u_i=p_i=q_i$)**：这些玩家表现出稳定的、基于启发式的行为，无论对手发出什么信号，赛前信号都与其后续行动一致。此类别包括无条件合作（ACC）和无条件背叛（NDD）。我们假设，这种一致性不会产生额外的推理成本。

- **审慎策略($u_i\neq p_i$ 或 $p_i\neq q_i$)**：这些玩家会进行复杂的推理过程，例如采取条件行为（如 ACD，针对不同信号作出不同回应）或进行欺骗性信号传递（如 ADD 或 NDC，信号与行动相矛盾）。为模拟此类决策过程所需的投入，我们设定推理成本 $\gamma\in[0,1]$。

关键在于，当且仅当玩家采用审慎策略时，才会产生此项成本。成本函数 $C(S_i)$定义如下：

$$C(S_i)=\begin{cases}0, & \text{if } u_i=p_i=q_i\\ 1, & \text{otherwise.}\end{cases}\tag{1}$$

B. 收益公式

假设玩家 i 和 j 在一场博弈中配对。设 x为玩家 i 合作的概率， y 为玩家 j 合作的概率。在确定性情形下，这些概率取决于对手的信号：

$$x=u_jp_i+(1-u_j)q_i,\quad y=u_ip_j+(1-u_i)q_j.\tag{2}$$

玩家 i 对阵玩家 j 的期望收益由以下双线性形式给出：

$$\Pi(S_i,S_j)=\begin{pmatrix}x & 1-x\end{pmatrix}\mathbf{G}\begin{pmatrix}y\\1-y\end{pmatrix}^T-\gamma C(S_i).\tag{3}$$

C. 种群更新

我们分析了空间结构化种群中的演化动力学，该种群为局部交互提供了环境。为确保研究结果稳健，我们采用三种重要的正则网络来建模种群：格子网络、小世界网络和随机网络 [47]。在这些均质拓扑结构中，玩家平均拥有 $k=4$个邻居。此外，我们还考虑了无法提供网络互惠的充分混合网络，以及具有异质性特征的无标度网络，如图 1(b) 所示。

每位玩家 i 占据 $\mathcal{V}$ 中的一个固定节点，并通过连边$\mathcal{E}$仅与其一阶邻居集合Ω_i进行互动。最初，每个节点都以均匀随机的方式从八种可用类型中选择一种策略。每个时间步中，每位玩家 i通过与所有邻居进行博弈，累积总收益 ϕ_i:

$$\phi_i=\sum_{j\in\Omega_i}\Pi(S_i,S_j),$$

其中 $\Pi(S_i,\ S_j)$ 对应于方程（3）中推导出的两两期望收益。

D. 探索与模仿

策略的演化由结合随机探索的社会学习所支配。我们使用空间结构化种群上的异步蒙特卡洛模拟来建模种群动态。每个时间步中，随机选取一位焦点玩家 i，并通过两种机制之一更新其策略。以较小的概率 μ，玩家会从八种可用类型中均匀随机地选择一种策略并采用。这一机制对应于策略探索（或强化学习中的 ϵ-贪婪策略）。需要将这种探索与感知噪声（信号传输中的错误）区分开来；在此，我们严格关注前者，以探究策略探索的作用。以概率 $1-\mu$，玩家 i尝试向一位邻居学习。随机从 i的邻居 Ω_i 中选取一位榜样 j。玩家 i 以由费米函数 [48]决定的概率采用 j的策略：

$$W(S_j\to S_i)=\frac{1}{1+e^{\beta(\phi_i-\phi_j)}},\tag{4}$$

其中， ϕ_i 和 ϕ_j 分别是玩家i 和 j的累计总收益。参数 β 表示选择强度； β 越高，玩家就越可能模仿收益严格更高的邻居。

IV. 实验结果

A. 实现细节

异步蒙特卡洛模拟采用上一节所述的更新协议进行（见

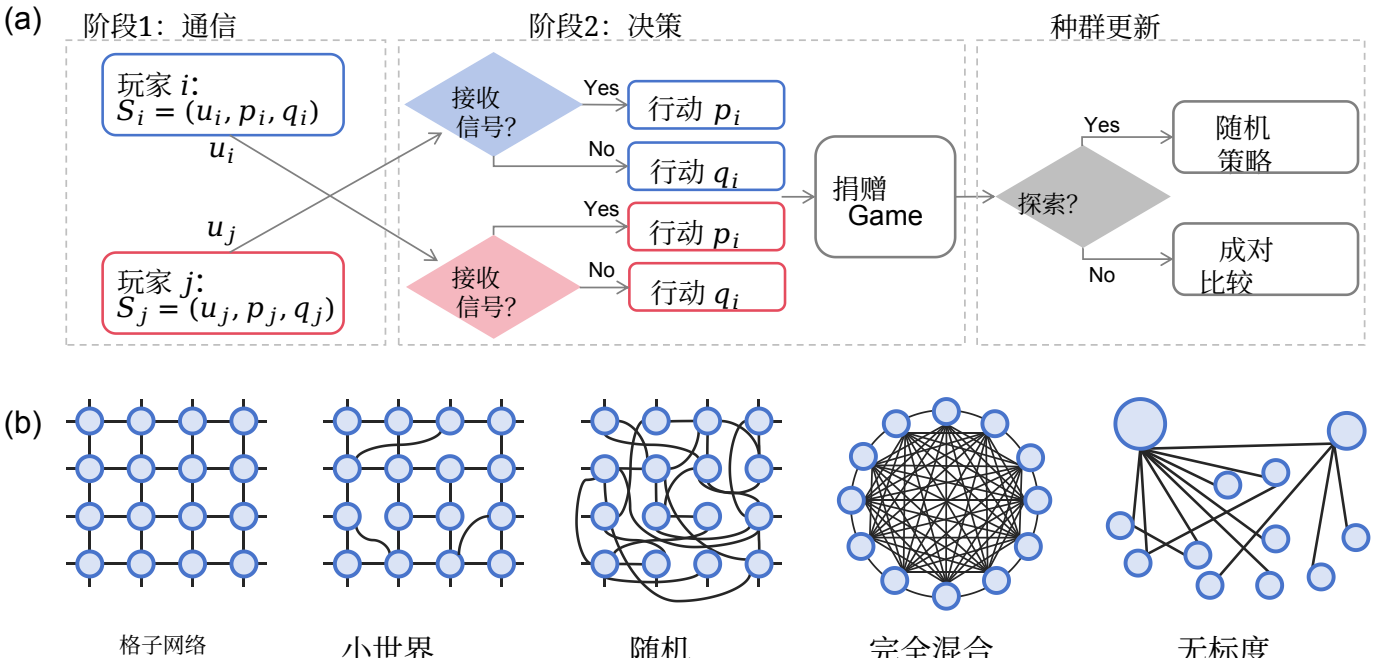

图 1：**两阶段廉价交谈博弈与种群结构**。(a) 博弈分为两个阶段：博弈前交流和博弈中的决策。在每个演化更新步骤中，焦点玩家要么探索一种新的随机策略，要么向随机榜样学习策略。(b) 所研究种群结构的示例：方格格子、小世界、随机、完全混合和无标度网络。

算法 1 基于主体的仿真

初始化： 将玩家安排在节点上，并为每位玩家分配一种随机策略。
对于 $t = 1$ 至 T_{max} **do**
  对于 $i = 1$ 至 $\mathcal{N}$ **执行** ▷一次蒙特卡洛步（MCS）
    从 $\mathcal{V}$ 中随机选择一个焦点玩家 i;
    通过与 Ω_i 交互，计算总收益 ϕ_i ;
    从 Ω_i 中随机选择一个邻居 j ;
    通过与 Ωj 交互，计算总收益 ϕ_j ;
    生成一个随机数 $r_1 \in [0, 1]$;
    若 $r_1 < \mu$ **则** ▷ 探索机制
      $s_i \leftarrow$ 从策略集中随机选择策略
    else ▷ 模仿机制
      计算费米概率:
      $W \leftarrow \frac{1}{1+\exp[\beta(\phi_i - \phi_j)]}$;
      生成一个随机数 $r_2 \in [0, 1]$;
      如果 $r_2 < W$ **则**
        $s_i \leftarrow s_j$;
      结束如果
    结束如果
  结束循环
  如果 $t > T_{max} - T_{avg}$ **则**
    记录当前合作者比例 $\rho_C(t)$;
  结束如果
结束循环

算法1）。一个蒙特卡洛步（蒙特卡洛步）包含 $\mathcal{N}$次这样的基本更新，从而确保平均每个蒙特卡洛步中每位玩家被选中一次以更新其策略。对于我们的模拟，我们将种群规模设置为 $\mathcal{N} = 50^2$。该选择强度设为 $\beta = 10$ 以模拟强选择。模拟运行 3×10^4 蒙特卡洛步，并对最后 5,000 步的数据取平均，以确保系统已达到统计稳态。

B. 廉价交谈仍是促进合作的有力工具

我们的结果表明，与网络互惠相结合，廉价交谈仍然是在存在策略探索的情况下维持合作的有力工具。我们将分析范围从规则晶格扩展到小世界网络和随机网络，并比较不存在探索与存在探索时平均合作频率的分布（图2的顶部一行）。在无探索情形下，三种拓扑结构中的分布均呈现两极化，主要聚集在完全背叛或完全合作两端。这种二元行为表明，无论网络结构如何，有限规模效应和吸收态都是有限种群中的普遍挑战。相比之下，引入策略探索后，所有拓扑结构中的分布都发生了显著变化：尽管仍然有所偏斜，但呈现出广泛且动态的分布，其中相当一部分结果维持了中等至较高水平的合作。具体而言，在各种拓扑结构中，合作频率处于0至0.8之间的情况显著增加，而合作频率为0的情况则大幅减少。这种一致性证实，策略探索是一种通用机制，能够防止系统陷入静态背叛，并使廉价交谈得以在多样化的结构环境中稳健地维持合作。

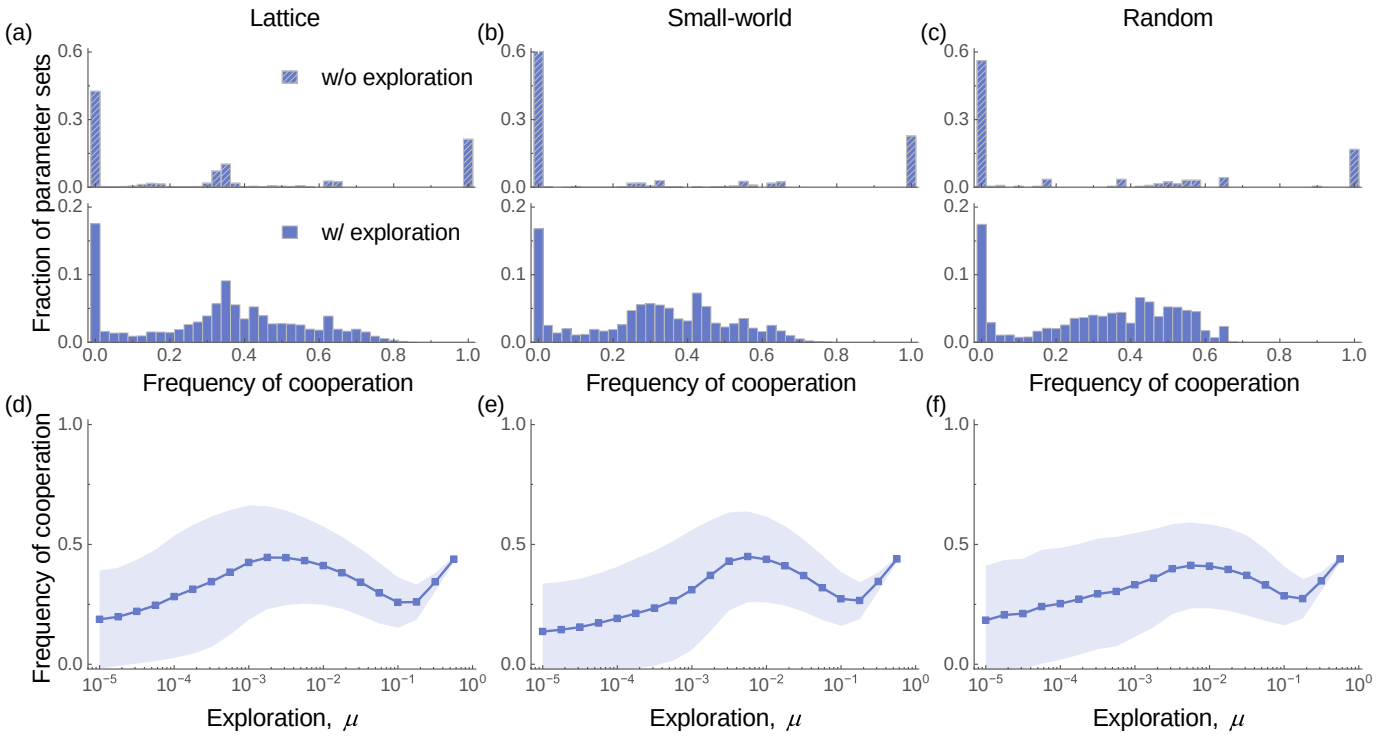

图 2：**在探索情景中，廉价交谈维持了合作，而最优探索使整体合作最大化**。面板 (a)-(c) 展示了不同参数集下无探索和有探索时的合作频率；面板 (d)-(f) 展示了合作平均频率及其标准差随探索变化的情况。图中分别展示了在格子网络、小世界网络和随机网络上进行的独立模拟结果，并均匀抽样了 $r \in [0, 0.3]$、$\gamma \in [0, 0.3]$ 和 $\mu \in [10^{-5}, 10^{-\frac{3}{4}}]$。

C. 最优探索提升整体合作

基于这一观察，我们将从定性分布分析转向定量研究，以探讨探索如何决定合作结果。

关键的是，系统存在一个能使合作结果最大化的最优探索率。如图2底行所示，在格子网络、小世界网络和规则随机网络中，合作平均频率与探索率之间的关系呈现出明显的非单调倒 U 形轨迹。探索率最低时（$\mu = 10^{-5}$），合作从基准水平开始，逐渐上升，并在$\mu = 10^{-3}$附近达到明显的峰值。超过这一最优值后，合作先下降，随后随着系统趋近随机行为而回升至接近 0.5。尽管峰值振幅会因拓扑性质（例如随机网络中较短的平均路径长度）而略有不同，但最优探索窗口的存在具有稳健性。这一普遍模式表明，适度的噪声会引发探索，而探索对于最大化系统的合作能力至关重要。

为进一步厘清策略探索如何影响廉价交谈的作用，我们在全混合网络和无标度网络上评估了该系统，从而揭示上述发现的功能边界。首先，在全混合种群中（图3(a- b)），没有探索时合作会彻底失败，而噪声只能使合作略有提升。这项对照实验证实，网络互惠是先决条件；探索只有在空间结构已能促进局部聚类时才会发挥催化作用。

其次，在无标度网络上（图3(c-d)），廉价交谈与网络本身强大的网络互惠相结合，能够确保较高水平的合作。在这些高度异质的拓扑结构中，高度连接的枢纽节点天然具备稳健的结构性互惠，从而确保近乎完美的合作。因此，引入随机探索并非起催化作用，而是会造成干扰；它可能削弱这些合作枢纽的主导地位，进而瓦解合作所需的结构凝聚力，导致表现持续下降。总之，最优探索效应只存在于分布式的同质拓扑结构中，包括格子网络、小世界网络和随机网络。在这些系统中，既没有无标度网络中占据压倒性优势的枢纽节点，又具备格子网络的局部结构，因此适度的探索率对于打破局部僵局、释放廉价交谈的全部潜力至关重要。

D. 探索促成合作策略的成功

为揭示所观察到的合作峰值的机制基础，我们从总体结果转向对策略组成的详细分析。

合作的提升源于对合作策略的选择性促进。图4 展示了三种网络拓扑中各策略的频率随探索率的变化情况，详细分布见

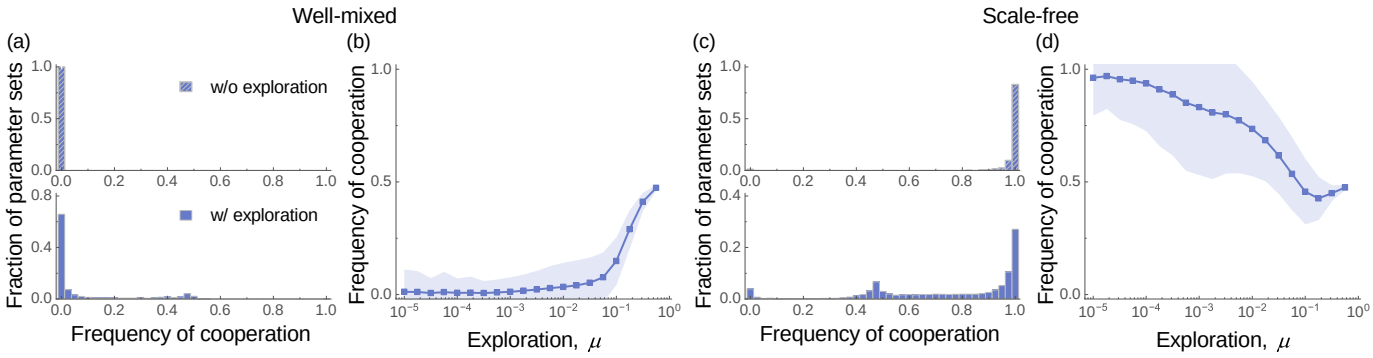

图 3：**在缺乏网络互惠时，廉价交谈无法维持合作；而在异质网络上，探索会扰乱合作**。面板 (a) 和 (c) 展示了无探索和有探索时不同参数集下的合作频率；面板 (b) 和 (d) 展示了合作平均频率和标准差随探索变化的情况。图中分别给出了完全混合网络和无标度网络上的独立模拟结果，以及均匀采样的 $r \in [0, 0.3]$、$\gamma \in [0, 0.3]$ 和 $\mu \in [10^{-5}, 10^{-\frac{3}{4}}]$。

图 S1、S4 和 S7。与图2中的总体合作峰值一致，最优探索率（约为 $\mu = 10^{-3}$）促成了三种关键合作策略的同时成功：直觉性的无条件合作（ACC）、审慎性的条件合作（ACD），以及寄生性的虚假合作（NDC）。在所有网络类型中，这些策略都在最优探索率附近协同达到峰值，而占主导地位的无条件背叛（NDD）则显著下降。这种差异化影响证实，适度的策略探索通过充分发挥信号传递与条件响应之间的协同作用来促进合作，而不仅仅是引入随机性。

E. 最优探索平衡振荡周期与放大效应。

为揭示这种合作的动态特性，我们考察策略频率随时间的演变，并确定了由最优探索驱动的周期性更替。图5展示了三种网络拓扑结构在不同探索率下，从全背叛状态开始的演化动力学。

适度的探索率能够平衡周期性成功的周期与振幅，从而最大限度地提高总体合作水平。在探索率极低时（$\mu = 10^{-5}$，第一列），由于探索不足，所有网络中的系统都困于以 NDD 占主导的初始吸收态。随着探索率上升（$\mu = 10^{-4}$，第二列），系统呈现出长周期、高振幅的振荡。在这一状态下，探索十分罕见。当合作突变体（ACD）终于入侵由背叛者占主导的群体时，由于后续入侵的时间尺度较慢，合作随之在整个系统中迅速激增。然而，虚假合作会逐渐累积，最终摧毁总体合作水平，随后系统便急剧崩溃。这些周期呈现出明显的相位滞后：NDD $\rightarrow$ ACD $\rightarrow$ ACC $\rightarrow$ NDC $\rightarrow$ NDD。尽管合作水平峰值很高，但波与波之间漫长的间隔时间会导致 NDD 长期占主导地位，使长期平均合作水平偏低——因此降低了长期平均合作水平。相反，在探索率较高时（$\mu = 10^{-2}$，10^{-1}，最右侧两列），系统转为短周期、低振幅的波动。过度探索如同一种破坏性噪声，会导致新策略同时出现在各处，从而使群体失去同步。因此，系统进入嘈杂的准稳态，策略频率在狭窄范围内快速波动。尽管合作持续存在，但由于噪声在合作增长至占主导地位之前便将其打散，合作水平受到抑制，仅维持在平庸水平。关键在于，适度的探索率（$\mu = 10^{-3}$，中间列）在振荡振幅与周期之间取得了临界平衡。在这一最优探索水平下，探索足够频繁，能够迅速打破 NDD 僵局，缩短低合作水平持续的时间；同时又足够适度，能够维护合作的空间完整性，保持较高的合作振幅。

最后，我们考察格点网络上策略的空间分布，揭示合作联盟这一合作的潜在路径。图6 展示了种群演化的快照。不同于高探索率下所呈现的无序混合状态（图6底行），最优探索率能够促成合作联盟的形成（图6顶行）。具体而言，条件合作者构成集群的保护性外壳，直接与背叛者对抗。在这些受保护的 enclaves 内部，无条件合作者和伪装合作者得以繁荣。这种空间组织形成了一道屏障：审慎策略（ACD）承担认知成本以筛除背叛者，从而构建出一个能够支持多样化直觉型和审慎型合作策略的生态位，直到 NDC 的内部剥削触发局部崩溃并重置循环。正是这种结构性自组织支撑了高水平的合作以及时间序列中观察到的循环动力学。

F. 探索并非万能良方

尽管探索通常有益，但它促进合作的能力是有条件的：当困境强度较高且推理成本较高时，探索无法带来最优的

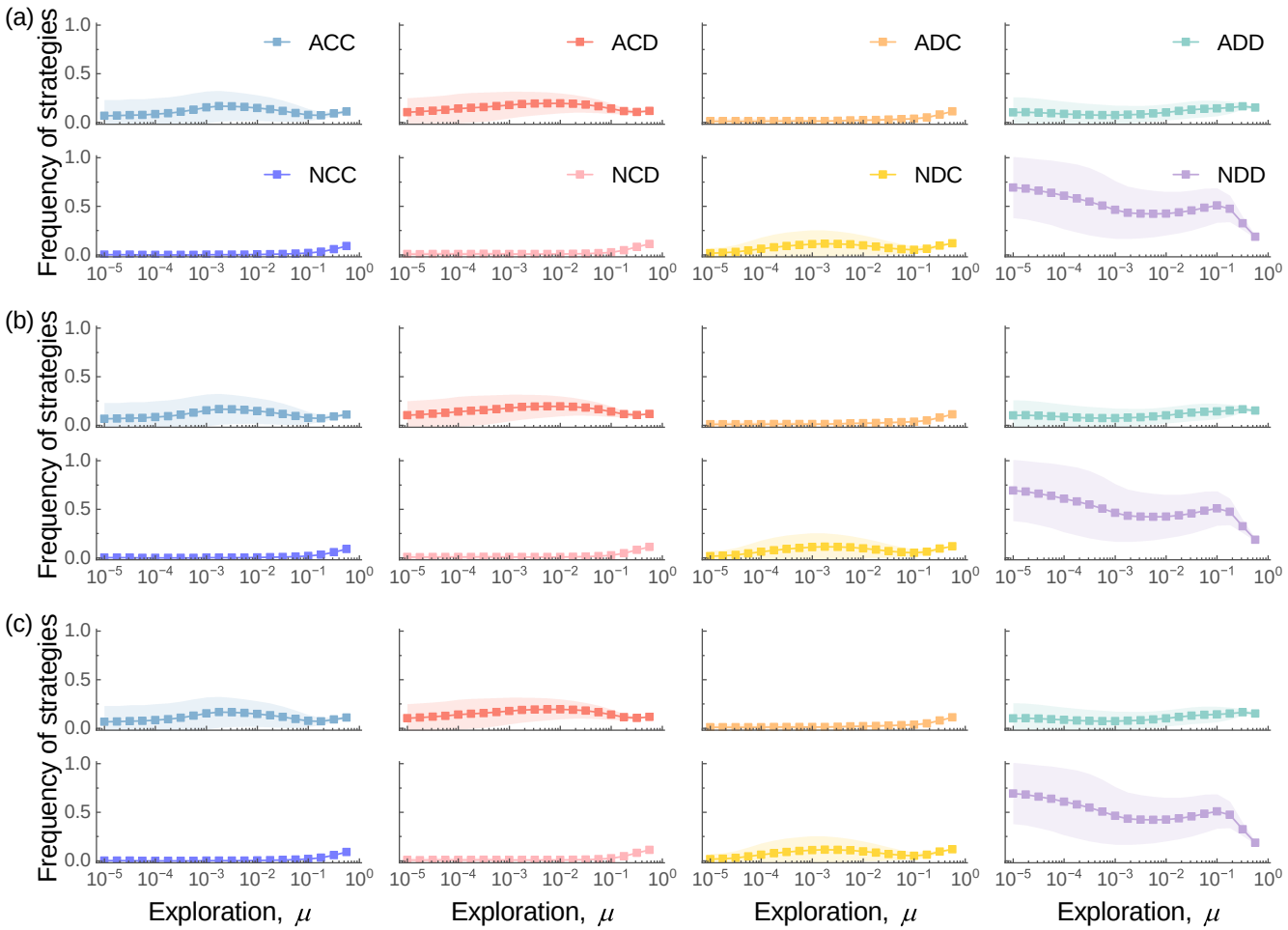

图 4：**探索使合作策略在直觉型和审慎型两类中均能达到最佳成功效果**。各面板从上到下依次展示格子网络、小世界网络和随机网络中，各策略的平均频率及标准差随探索程度的变化。结果基于 100,000 次模拟，并从 $\mu \in [10^{-5}, 10^{\frac{3}{4}}]$、$r \in [0, 0.3]$、$\gamma \in [0, 0.3]$ 的均匀分布中随机抽样得出。

合作率。以格点种群为例，首先，当推理成本较高时（图 7(a-b)(d)），探索带来的益处会受到抑制。高昂的成本会使采用条件逻辑的审慎策略灭绝，只留下无需成本的无条件策略（C 和 D）。因此，系统会回归传统的囚徒困境，而探索无法扭转这一局面。其次，当困境强度较高时（图 7(c)），探索的影响同样会受到抑制，因为背叛的诱因过于强烈。无论探索率如何，较高的困境强度都会阻碍可持续合作簇的形成。这些结果表明，只有当博弈的底层参数为条件策略的存续提供机会时，探索才能释放廉价交谈的合作潜力；若困境强度或认知成本过高，探索便无法建立合作。

V. 讨论

我们的研究结果与日益增多的理论文献相一致，这些文献强调了探索在演化

演化博弈中的建设性作用。一般而言，为了获得清晰的相变或简化演化动力学，探索往往被排除在外。然而，近期关于直接互惠的研究表明，探索可以通过防止系统冻结在由背叛主导的吸收态中，从而主动促进合作 [49]。同样，关于空间演化博弈的研究也表明，探索是缓解有限尺寸效应的重要机制，能够提高基于智能体的模型的可靠性 [50]。我们的工作通过在廉价交谈的情境中识别出一个特定的最优探索窗口，为这一理论叙事作出了贡献。与纯行动博弈不同——在纯行动博弈中，探索主要起到更新策略池的作用——在我们的基于通信的模型中，最优探索充当着精确的调节器：它提供足够的波动性以动摇僵化的背叛，同时又保持足够低的水平，以维护复杂的条件性信号传递策略的完整性。这强化了这样一种理论观点：策略噪声并非仅仅是需要最小化的干扰，而是驱动系统优化的基本参数。

除了重点关注策略探索外，我们的补充分析还揭示了一种有悖直觉的现象

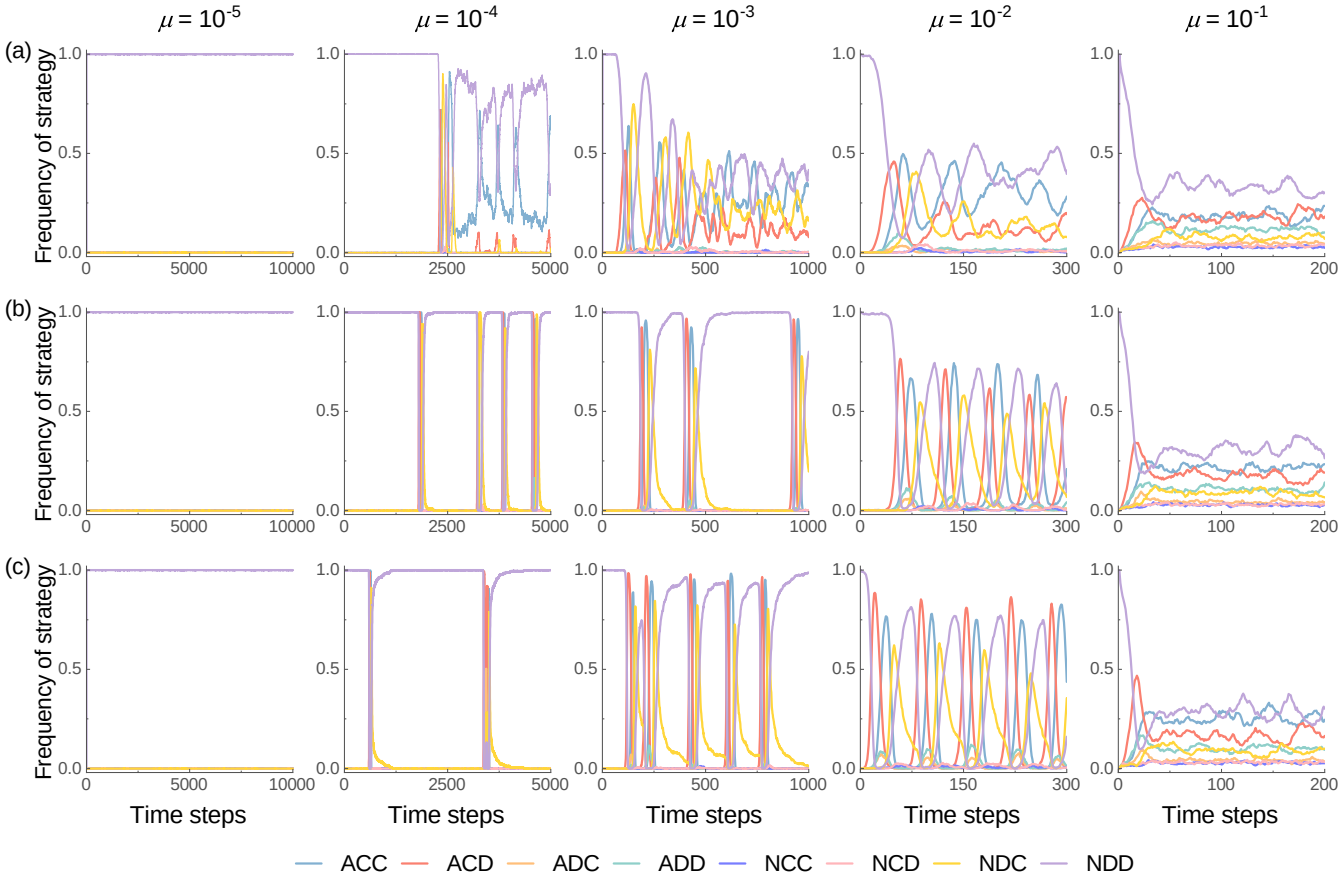

图5：**适度探索通过平衡周期性成功的振幅与周期来最大化合作**。图中展示了从完全由无条件背叛（NDD）组成的种群出发时策略频率的时间演化。探索使NDD（紫色）→ ACD（红色）→ ACC（蓝色）→ NDC（黄色）之间产生相位滞后。参数分别设为 $r = 0.02$、 $\gamma = 0.1$、$\mu = 10^{-5}$， 10^{-4}， 10^{-3}， 10^{-2}和 10^{-1} ，从左列到右列依次对应。

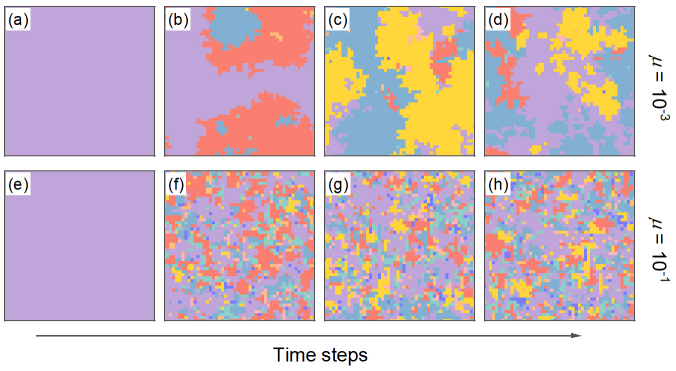

图6：**适度探索促成直觉策略与审慎策略的合作联盟**。各面板展示了格点网络上演化过程中的策略分布。参数设置为：顶行为 $r = 0.02$、 $\gamma = 0.1$和$\mu = 10^{-3}$，底行为 $\mu = 10^{-1}$ 。

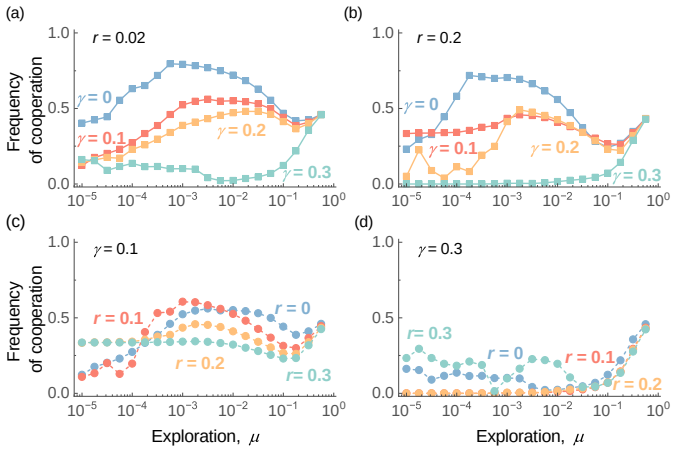

图7：**探索可以促进合作的存续，但它并非万能之策**。各面板展示了格点网络上合作频率随探索的变化。参数设置如下： $r = 0.02$ （a）， $r = 0.2$ （b）， $\gamma = 0.1$ （c），以及$\gamma = 0.3$ （d）。

与推理成本有关。尽管困境强度增加会单调地抑制合作（图 S2、S5、S8），但推理成本与直觉型合作者（ACC）的成功率之间呈现出独特的倒U形轨迹（图 S3、S6、S9）。这意味着，

严格的零成本环境并不一定有利于简单的无条件合作传播。从机

制层面来看，适度的推理成本可充当进化过滤器：它足够高，能够对复杂的寄生策略（例如发送信号却选择背叛的策略）施加选择压力，防止其利用通信渠道；同时又足够低，能够维持一定规模的条件合作者（ACD），以保护群体。在工程应用中，推理成本对应于复杂决策算法所需的计算开销（例如 CPU 周期、能耗或处理延迟）。因此，我们的研究结果提出了一项适用于资源受限系统的设计原则：与其追求零延迟处理，不如人为对复杂逻辑征收少量计算税；这或许看似矛盾，却能通过使剥削性策略在进化上变得得不偿失，从而稳定合作。未来的研究应验证这一“最优成本假说”，为系统设计提供有价值的指导。

从实践角度来看，最优探索率的存在为设计稳健的多智能体系统和去中心化协议提供了宝贵的见解。在许多工程系统中，例如自主集群机器人或点对点网络，设计者往往致力于实现确定性效率，试图消除错误或偏差 [51]，[52]。然而，我们的结果表明，强行施加严格的确定性可能适得其反，甚至可能导致整个系统陷入不合作的困境。相反，采用包含适度随机探索的协议——其功能类似于强化学习（RL）中的探索与利用权衡——可能会带来益处 [53]。正如强化学习智能体偶尔会偏离当前策略以寻找更优的全局解，我们的模型表明，“人为设计的随机性”能够让群体摆脱次优的局部最优。例如，在自动驾驶车辆的协调中，严格遵循确定性优化可能导致交通僵局；然而，允许智能体偶尔偏离标准策略，能够有效打破此类对称性，恢复合作的交通流 [54]。

尽管我们的框架提供了若干洞见，但它依赖于若干简化假设，从而为未来研究开辟了途径。首先，我们的模拟是在多种静态网络上进行的，以结合网络互惠的效应。由于现实世界中的社会与生物网络往往表现出时变特性，未来工作应研究动态交互拓扑如何影响最优探索原理 [55]。此外，我们的研究主要聚焦于度分布相对同质的去中心化网络。尽管我们观察到，探索可能通过动摇合作枢纽而破坏无标度网络中的合作，但关于探索如何具体影响枢纽与叶节点动态的细粒度分析，仍是未来研究有待解决的问题 [56]。其次，我们将通信渠道限制为二元信号。实际上，通信往往是多维的；将模型扩展以纳入更丰富多样的信号，或可揭示语义复杂性的提高与探索相结合是否会进一步增强合作稳定性 [26]。最后，我们的模型假设所有个体具有统一的全局探索率。然而，在自然系统中，探索倾向往往本身是一种演化性状，而非固定常数。未来工作可以探索探索率具有适应性的协同进化模型，类似于强化学习中的元学习参数，以理解系统如何自我调谐至最优探索水平 [57]。

VI. 结论

为探究探索的影响下廉价交谈如何维持合作，我们在空间结构化种群中构建了一个以两阶段廉价交谈博弈为特征的演化博弈框架。我们首先证实，即使存在探索，廉价交谈与网络互惠相结合仍然是维持合作的强大途径。至关重要的是，我们发现了一个能够最大化合作结果的最优探索率，该探索率促进了直觉型和审慎型合作策略的进化成功。从机制上看，适度的探索率通过削弱背叛的稳定性发挥作用，同时使合作策略能够形成紧密且富有韧性的空间簇，利用网络互惠抵御入侵。然而，策略探索并非万能药；在困境强度较高或推理成本过高的情况下，最优探索率会消失，合作仍受到抑制，此时廉价交谈的力量从根本上失效。

VII. 文章信息

致谢： 我们感谢英国工程与物理科学研究理事会（资助编号 EP/Y00857X/1）对 Z.S. 和 T.A.H. 提供的支持。

作者贡献： Z. S. 和 C. S. 提出了研究构想。Z. S. 进行了理论分析和演化模拟。所有合著者均讨论了研究结果并撰写了论文。

利益冲突： 作者声明不存在利益冲突。

参考文献

[1] Shaolin Tan、Yaonan Wang、Yao Chen 和 Zhen Wang。集体行为选择与漂移的演化动力学：集群、崩溃与振荡。*IEEE Transactions on Cybernetics*，47 1705，2016 7 1694( ): – 。[2] Panjun Sun、Shigen Shen、Zongda Wu、Haiping Zhou 和 Xiao-Zhi Gao。边缘服务中的信任合作激励：一种演化三方博弈。人工智能工程应用，116 105465，2022 : 。[3] Xin Zhou、Xiangyong Wen、Zhepei Wang、Yuman Gao、Haojia Li、Qianhao Wang、Tiankai Yang、Haojian Lu、Yanjun Cao、Chao Xu 等。野外的微型飞行机器人群。*Science Robotics*，7(66):eabm5954，2022。[4] Hao Guo、Zhao Song、Matja˘z Perc、Xuelong Li 和 Zhen Wang。多层网络中的合作第三方干预。*IEEETransactionsonSystems,Man, andCybernetics:Systems*，53(11):6646– 6657，2023。[5] The Anh Han、Zhao Song、Theodor Cimpeanu、Manh Hong Duong、Marcus Krellner、Valerio Capraro 和 Matjaz Perc。合作与社会福利。*Physics of Life Reviews*，56:33–60，2026。[6] 约翰·F·纳什。非合作博弈。载于价格理论基础 第*4*卷，第329–340页。Routledge，2024。[7] Xiao Fang、Guanghui Wen、Tingwen Huang、Zao Fu 和 Liang Hu。在马尔可夫切换通信网络中分布式寻求纳什均衡。*IEEETransactionson Cybernetics*，52 5355，2020 6 5343( ): – 。

[8] Feng Fu、Christoph Hauert、Martin A Nowak 和 Long Wang。基于声誉的伙伴选择促进社会网络中的合作。物理评论E——统计、非线性与软物质物理，78(2):026117，2008。[9] Hannelore Brandt、Christoph Hauert 和 Karl Sigmund。空间公共物品博弈中的惩罚与声誉。《伦敦皇家学会会刊》B辑：生物科学，270 1104，2003 1519 1099( ): – 。[10] Mei-huan Chen、Li Wang、Shi-wen Sun、Juan Wang 和 Cheng-yi Xia。具有自适应声誉分类的空间公共物品博弈中的合作演化。物理快报A，380(1-2):40–47，2016。[11] Tarik Hadzibeganovic、Pengbi Cui 和 Zhi-Xi Wu。合作者的不从众促进基于标签的多智能体网络系统中纯粹利他主义的出现。知识型系统，171:1–24，2019。[12] Naoki Masuda 和 Hisashi Ohtsuki。不完整社会信息下基于标签的间接互惠。《英国皇家学会会刊》B辑：生物科学，274(1610):689–695，2007。[13] Rick L Riolo、Michael D Cohen 和 Robert Axelrod。没有互惠的合作演化。自然，414(6862):441–443，2001。[14] Attila Szolnoki 和 Matjaz Perc。空间公共物品博弈中的奖励与合作。欧洲物理快报，92(3):38003，2010。[15] Theodor Cimpeanu、Cedric Perret 和 The Anh Han。促进最后通牒博弈公平性的成本高效干预。知识型系统，233:107545，2021。[16] The Anh Han。促进承诺性合作演化的制度性激励：确保参与与提升遵从同等重要。《英国皇家学会界面杂志》，19(188):20220036，2022。[17] Justin Semonsen、Christopher Griffin、Anna Squicciarini 和 Sarah Rajtmajer。一致性压力不断增加时的观点动力学。IEEE Transactions on Cybernetics，49(4):1270–1278，2018。[18] Ana Paiva、Fernando Santos 和 Francisco Santos。利用自主智能体推动亲社会行为。载于《AAAI人工智能会议论文集》，第32卷，2018。[19] Peter Andras、Lukas Esterle、Michael Guckert、The Anh Han、Peter R Lewis、Kristina Milanovic、Terry Payne、Cedric Perret、Jeremy Pitt、Simon T Powers 等。信任智能机器：深化社会技术系统中的信任。IEEE技术与社会杂志，37(4):76–83，2018。[20] Hao Guo、Chen Shen、Shuyue Hu、Junliang Xing、Pin Tao、Yuanchun Shi 和 Zhen Wang。通过自主智能体促进人-智能体混合群体中的合作。iScience，26(11)，2023。[21] Filippo Zimmaro、Manuel Miranda、Jos´e Mar´ía Ramos Fern´andez、Jes´us A Moreno L´opez、Max Reddel、Valeria Widler、Alberto Antonioni 等。通过歧视性人工智能和撒玛利亚人工智能，在一次性囚徒困境中实现合作的涌现。《英国皇家学会界面杂志》，21(218)，2024。[22] Tianjiao An、Yuexi Wang、Guangjun Liu、Yuanchun Li 和 Bo Dong。基于合作博弈的模块化机器人机械臂近似最优控制，用于人机协作。IEEE Transactions on Cybernetics，53(7):4691–4703，2023。[23] Joseph Farrell 和 Matthew Rabin。廉价交谈。《经济学展望杂志》，10(3):103–118，1996。[24] Brian Skyrms。信号：演化、学习与信息。牛津大学出版社，2010。[25] Vincent P Crawford 和 Joel Sobel。策略性信息传递。《计量经济学：计量经济学会杂志》，第1431 1451页，1982 – 。[26] Francisco C Santos、Jorge M Pacheco 和 Brian Skyrms。博弈前信号传递与合作的协同演化。《理论生物学杂志》，274(1):30–35，2011。[27] The Anh Han、Francisco C Santos、Tom Lenaerts 和 Luis Moniz Pereira。合作困境中意图识别与承诺之间的协同作用。科学报告，5(1):9312，2015。[28] Arthur J Robson。演化博弈中的效率：达尔文、纳什与秘密握手。《理论生物学杂志》，144 396，1990 3 379( ): – 。[29] Michael P Wellman、William E Walsh、Peter R Wurman 和 Jeffrey K MacKie-Mason。去中心化调度的拍卖协议。《博弈与经济行为》，35(1-2):271–303，2001。[30] Michael Wooldridge。多智能体系统导论。约翰·威立父子公司，2009。[31] David C Krakauer 和 Mark Pagel。空间结构与诚实无成本信号传递的演化。《伦敦皇家学会会刊》B辑：生物科学，260(1359):365–372，1995。[32] Colin J Torney、Andrew Berdahl 和 Iain D Couzin。动态环境中信号传递与合作觅食的演化。《PLoS计算生物学》，7(9):e1002194 ，2011。[33] Thomas Wiseman 和 Okan Yilankaya。囚徒困境中的合作、秘密握手与模仿。《博弈与经济行为》，37(1):216–242，2001。[34] Bin Wu、Chaitanya S Gokhale、Long Wang 和 Arne Traulsen。小突变率究竟有多小？《数学生物学杂志》，64(5):803–827，2012。[35] Laura Hindersin、Bin Wu、Arne Traulsen 和 Julian Garc´ıa。有限种群中进化博弈动力学的计算与模拟。《科学报告》，9(1):6946，2019。[36] Zhao Song、Chen Shen 和 The Anh Han。网络互惠将廉价交谈转化为促进合作的力量。《理论生物学杂志》，617:112303，2026。[37] Daniel Houser 和 Robert Kurzban。重新审视公共品实验中的善意与困惑。《美国经济评论》，92(4):1062–1069，2002。[38] Jelena Gruji´c、Carlos Gracia-L´azaro、Manfred Milinski、Dirk Semmann、Arne Traulsen、Jos´e A Cuesta、Yamir Moreno 和 Angel S´anchez。空间囚徒困境实验的比较分析：条件合作与收益无关性。《科学报告》，4 4615，2014 1( ): 。[39] George De Ath、Richard M Everson、Alma AM Rahat 和 Jonathan E Fieldsend。贪婪是好事：贝叶斯优化中的探索与利用权衡。《ACM进化学习与优化汇刊》，1(1):1–22，2021。[40] Luoer Chen、Churou Deng、Manh Hong Duong 和 The Anh Han。噪声社会困境中复制-突变动力学的均衡数量。《混沌、孤子与分形》，180:114565，2024。[41] Mauro Mobilia、Ivan T Georgiev 和 Uwe C T¨auber。随机格子网络洛特卡-沃尔泰拉模型中的相变与时空涨落。《统计物理杂志》，128(1):447–483，2007。[42] Arne Traulsen、Dirk Semmann、Ralf D Sommerfeld、Hans-J¨urgen Kram- beck 和 Manfred Milinski。进化博弈中的人类策略更新。《美国国家科学院院刊》，107(7):2962–2966，2010。[43] Karl Sigmund。《自私演算》。普林斯顿大学出版社，2010。[44] Yang Yang 和 Xiang Li。迈向用于网络顶点覆盖的雪堆博弈优化。《IEEE Transactions on Cybernetics》，43(3):948–956，2013。[45] Robert Axelrod 和 Douglas Dion。合作的进一步演化。《科学》，242(4884):1385–1390，1988。[46] Zhen Wang、Satoshi Kokubo、Marko Jusup 和 Jun Tanimoto。演化博弈中困境强度的普适标度。《Physics of Life Reviews》，14:1–30，2015。[47] Gy¨orgy Szab´o 和 Gabor Fath。图上的进化博弈。《物理报告》，446(4-6):97–216，2007。[48] Karl Sigmund、Hannelore De Silva、Arne Traulsen 和 Christoph Hauert。社会学习促进管理公地的制度。《自然》，466(7308):861–863，2010。[49] Josef Tkadlec、Christian Hilbe 和 Martin A Nowak。突变增强直接互惠中的合作。《美国国家科学院院刊》，120(20):e2221080120，2023。[50] Chen Shen、Zhixue He、Lei Shi 和 Jun Tanimoto。突变减轻空间演化博弈中的有限尺寸效应。《通讯物理学》，8(1):201，2025。[51] Manuele Brambilla、Eliseo Ferrante、Mauro Birattari 和 Marco Dorigo。群体机器人：从群体工程视角进行综述。《群体智能》，7(1):1–41，2013。[52] Eng Keong Lua、Jon Crowcroft、Marcelo Pias、Ravi Sharma 和 Steven Lim。对等覆盖网络方案的调查与比较。《IEEE通信综述与教程》，7(2):72–93，2005。[53] Joel Z Leibo、Vinicius Zambaldi、Marc Lanctot、Janusz Marecki 和 Thore Graepel。序贯社会困境中的多智能体强化学习。arXiv预印本 arXiv:1702.03037，2017。[54] Cathy Wu、Aboudy Kreidieh、Eugene Vinitsky 和 Alexandre M Bayen。混合自治交通中的涌现行为。载于《机器人学习会议》，第398–407页。PMLR，2017。[55] Aming Li、Lei Zhou、Qi Su、Sean P Cornelius、Yang-Yu Liu、Long Wang 和 Simon A Levin。时序网络上的合作演化。《自然通讯》，11(1):2259，2020。[56] Francisco C Santos 和 Jorge M Pacheco。无标度网络为合作的涌现提供了统一框架。《物理评论快报》，95(9):098104，2005。[57] Attila Szolnoki、Jeromos Vukov 和 Gy¨orgy Szab´o。空间社会困境中策略采纳噪声水平的选择。《物理评论E—统计、非线性与软物质物理》，80(5):056112，2009。