网络博弈中的异质信念与多种群学习

Shuyue Hu[1], Harold Soh[2], 和 Georgios Piliouras[3]

[1]上海人工智能实验室[2]新加坡国立大学[3]新加坡科技设计大学

arXiv:2301.04929v1 [cs.MA] 12 Jan 2023

摘要

多智能体学习中的种群异质性效应具有实际意义，但我们对其仍远未充分理解。受此启发，我们提出一种多种群学习模型，该模型允许每个群体内部存在异质信念，并且智能体通过平滑虚拟博弈（SFP）对其信念作出响应。我们证明，系统状态——即信念上的概率分布——遵循一组偏微分方程演化。我们在涵盖网络竞争和网络协调的不同类别博弈中，证明了 SFP 收敛至量化反应均衡。我们还证明，在所有网络博弈中，信念最终都会趋于一致。尽管初始的信念异质性最终会消失，但我们表明，在协调博弈中，它对均衡选择至关重要，因为它有助于选出理想程度更高的均衡。相反，在网络竞争中，即使底层博弈存在多个不同的纳什均衡，最终的极限行为也与信念的初始化无关。

1 引言

平滑虚拟博弈（SFP）及其变体，可以说是人工智能和博弈论中研究最为广泛的学习模型之一 [2, 3, 21, 22, 9, 19, 36, 37, 42, 18, 17]。SFP 描述了一种基于信念的学习过程：智能体对对手的行动形成信念，并根据观察结果更新信念。通俗地说，智能体的信念可以理解为反映其对手采用各个策略的可能性。在博弈过程中，每个智能体都会针对自己的信念采取平滑最佳响应。大量关于 SFP 的文献都围绕齐次信念模型展开，在这类模型中，给定玩家角色中的所有智能体都具有相同的信念。这包括每个玩家角色各有一个智能体的模型 [3, 2, 39] ，以及只有一个群体但其中所有智能体都具有相同信念的模型 [21, 22]。已知 SFP 在许多类别的齐次信念模型中都会收敛（例如，大多数双人博弈 [9, 19, 3]）。然而，对于异质信念情形，即同一群体中的智能体具有不同信念的情形，SFP 的研究则相对较少。

异质信念（或更广义地说，种群异质性）的研究十分重要，也具有实际意义。从多智能体系统的角度来看，异质信念广泛存在于许多应用中，例如交通管理、在线交易和电子游戏。举例来说，人们普遍预期公众对自动驾驶车辆的看法会存在分歧，而且人们对出租车司机与非职业司机的行为也有不同的信念。从机器学习的角度来看，近期的实证进展表明，引入异质性可能会加速神经网络的基于种群的训练，并提升学习性能 [25, 29, 44]。从博弈论的角度来看，考虑信念的异质性有助于更好地解释一些人类实验的结果[10, 11]。

SFP的异质信念模型并非全新。开创性研究中， [12], Fudenberg 和 Takahashi 借助随机逼近理论中的技术，考察了双群体设定中的异质性问题。这种方法是 SFP 文献中的典型做法，它将每个个体的极限行为与常微分方程（ODE）联系起来，并为许多同质信念模型带来了重要见解 [3, 2, 19, 39]。然而，正如 Fudenberg 和 Takahashi 也指出的那样，这种方法“无法非常精确地估计系统初始条件的影响”。考虑这样一个例子：一个群体中的每个智能体都可以在两种纯策略之间进行选择

s_1 和 s_2。我们设想两种情况：(i) 群体中的所有智能体都持有相同的信念，认为对手以相同概率 0.5 选择 s_1 和 s_2 这两种混合策略；(ii) 一半智能体认为对手确定会选择纯策略 s_1，另一半则认为对手确定会选择纯策略 s_2。随机逼近方法通常会将这两种情况同等对待，因而无法充分揭示信念异质性及其对系统演化所产生的影响。这引出了我们提出的研究问题：

如何在 *SFP* 下演化？异质群体的演化程度如何，又需要满足哪些条件？ *es* 信念的异质性会在多大程度上影响其长期行为？

模型与解。本文研究允许信念存在异质性的广义多群体网络博弈中的 SFP 动态。在多群体网络博弈中，网络的每个顶点代表一个群体（连续体）的智能体，每条边代表两个相邻群体之间的一系列双人子博弈。请注意，多群体网络博弈涵盖了 [12] 中讨论的所有双群体博弈，并且是现实世界系统中图结构显而易见的若干子类的表示 [**?**]。我们假设，对于某个群体，个体智能体会针对每个邻居群体分别形成信念，并观察该群体的平均策略选择。我们采用一种不同于随机逼近的方法，将系统状态定义为信念空间上的概率测度，从而能够精确考察异质信念对系统演化的影响。该概率测度会随时间变化，以响应智能体的学习。因此，主要挑战在于分析该测度的演化，这通常需要发展新的技术。

首先，我们建立一个偏微分方程（PDE）系统，以追踪连续时间极限下测度的演变（命题1）。我们推导出的偏微分方程类似于物理学中常见的连续性方程[1]，且不允许存在一般解。借助矩闭合近似 [13], ，我们无需求解偏微分方程，便可直接分析均值与方差的动态（命题 2 和定理 1）。作为一项关键结果，我们证明，在所有网络博弈中，信念方差总是随时间以二次速度衰减（定理 1）。换言之，最终信念会趋于同质化，信念分布将收敛为一个点，而不受信念初始分布、智能体所参与的双人子博弈，以及群体和策略数量的影响。这一结果并非显而易见，甚至可能有些违背直觉。毕竟，人们或许会更自然地预期，信念分布会收敛到某个分布，而非一个点；近期关于 Q 学习和交叉学习 [23, 24, 27]的研究也印证了这一点。

从技术层面看，最终的信念同质化具有重要意义——它从直观上暗示，初始异质系统的渐近系统状态很可能与同质系统中的状态相同。我们证明，对于同质系统和初始异质系统，SFP 的不动点都对应于网络博弈中的量子响应均衡（QRE）[2]（定理 2）。作为主要结果，我们证明，在涵盖网络竞争和网络协调的不同类型博弈中，SFP 均收敛于 QRE，且这一结果与信念初始化无关。具体而言，对于竞争型网络博弈，我们首先通过李雅普诺夫论证证明，即使底层博弈具有多个不同的纳什均衡，SFP 在同质系统中仍收敛于唯一的 QRE（定理 3）。随后，我们利用初始异质系统的平均信念动态渐近自治 [31] 且其极限动态为同质系统的信念动态这一性质（引理 7），证明这一收敛结果同样适用于初始异质系统（定理 4）。对于协调网络博弈，我们也证明了在底层网络具有星形结构时，同质系统和初始异质系统均收敛于 QRE（定理 5）。

另一方面，最终的信念同质化可能会让人误以为信念异质性对系统演化影响甚微。我们以双群体鹿猎博弈为例，说明信念异质性实际上在均衡选择中发挥着至关重要的作用，尽管这种异质性最终会消失。如图1所示，即使初始信念的均值保持不变，改变初始信念的方差也会导致不同的极限行为；具体而言，较小的方差会导向较不理想的均衡 $(H,\ H)$，而较大的方差则会导向收益占优均衡$(S,\ S)$。因此，在网络协调的情形下，初始信念异质性有助于选择高度理想的均衡，并为均衡选择这一经典难题提供了有趣的见解 [26]。相反，在网络竞争的情形下，我们证明（竞争性网络博弈收敛到唯一量化反应均衡的定理3和定理4），并通过实验展示，即使底层博弈具有多个不同的纳什均衡，最终的极限行为也独立于信念的初始化。

[1]连续性方程是一种偏微分方程，用于描述物理系统中某种量（例如质量、能量、动量及其他守恒量）的输运现象。[2]量化反应均衡是一种有限理性下的博弈论解概念。本文所说的量化反应均衡是其典型形式，在文献中也称为逻辑斯蒂均衡或逻辑斯蒂量化响应均衡 [14]。

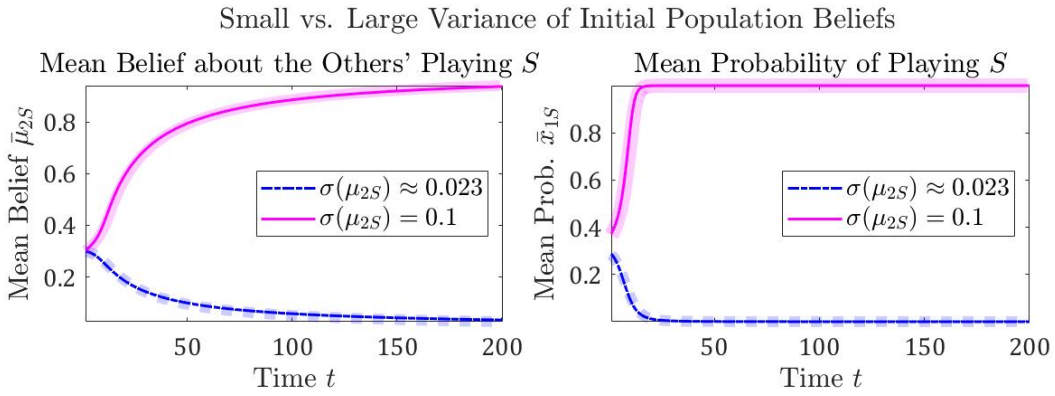


图1：初始信念的不同方差对系统动力学的影响（细线：偏微分方程模型的预测；宽阴影线：仿真结果）。$\bar{\mu}_{2S}$ 表示对群体 2 的平均信念，$\bar{x}_{1S}$ 表示群体 1 中选择策略 S 的平均概率。初始时，我们设定平均信念为 $\bar{\mu}_{2S} = \bar{\mu}_{1S} = 0.3$（设定详情见补充材料）。在初始平均信念相同的情况下，不同的初始方差 $\sigma^2(\mu_{2S})$ 会导致收敛到不同的信念（左图），甚至导致不同的策略选择（右图）。尤其是在鹿猎博弈中，较大的初始方差有助于选择收益占优均衡 $(S,\ S)$。

在竞争性网络博弈中收敛到唯一的量化反应均衡），并通过实验展示，即使底层博弈存在多个不同的纳什均衡，最终的极限行为也与信念的初始设定无关。

相关工作。 SFP及其变体近来在人工智能研究中备受关注[36, 37, 42, 18, 17]。已有大量文献分析了不同模型中的SFP [3, 7, 21, 19], ，而与我们工作最为相关的论文是 [12]。Fudenberg和Takahashi [12] 也研究了异质性问题，并预期在双群体设定下，信念最终会趋于同质化。本文考虑多群体网络博弈，这是对他们所研究设定的推广。[3] 此外，我们的方法更为基础，因为我们推导出的偏微分方程能够提供关于系统演化的丰富信息，从而精确估计异质性的时间效应，而这通常难以通过 [12]实现。因此，借助我们的方法，我们得以展示一个有趣的发现——初始异质性在均衡选择中起着至关重要的作用（图1）——而遗憾的是，使用 [12]中的方法无法展示这一点。最后但同样重要的是，据我们所知，我们的论文是首篇针对一般类别网络博弈中的平滑虚拟博弈进行系统研究的工作。

另一方面，网络化多智能体学习是当前人工智能和机器学习研究的前沿领域之一 [43, 30, 16]。关于网络博弈的最新理论进展给出了学习行为不发生混沌的条件 [6, 34], ，并研究了网络竞争情形下 Q 学习和连续时间虚拟博弈的收敛性 [7, 28]。然而， [7, 28] 这些研究认为每个顶点上只有一个智能体，因此其模型本质上针对的是同质系统。

Lahkar 和 Seymour [27] 以及 Hu 等人 [23, 24] 也将连续性方程作为工具，研究多智能体系统中的种群异质性；在这些系统中，单个智能体群体采用交叉学习或 Q 学习来参与对称博弈。他们通过证明或数值展示表明，异质性通常会持续存在。我们的结果补充了这些进展，指出在 SFP 下异质性会消失，并且异质性有助于选出非常理想的均衡。此外，在方法论上，我们借助渐近自治动力系统文献中的开创性结果（引理 1 和 2），为异质系统中学习动力学的收敛性建立了新的证明技巧；这些技巧本身也可能具有独立的研究价值。

[3]本文的分析涵盖所有一般的双群体网络博弈、所有一般的二部网络博弈（其中每条边上进行的博弈都相同），以及所有加权零和博弈；对于后者，图无需为二部图，且各条边上的博弈也无需相同。

2 预备知识

群体网络博弈。一个群体网络博弈（PNG） $\Gamma=(N,(V,E),(S_i,\omega_i)_{\forall i\in V},(\mathrm{A}ij)_{(i,j)\in E})$由分布在图 (V,E)上的多智能体系统 N组成，其中 $V=\{1,...,n\}$是顶点集合，每个顶点代表一个智能体群体（连续统），而 E 是由群体$i\neq j\in V$的配对构成的集合 (i,j)。对于每个群体 $i\in V$，该群体中的智能体拥有有限的纯策略（或行动）集合 S_i，其一般元素为 $s_i\in S_i$。智能体也可以采用混合策略（或选择分布）。对于群体 i中的任意智能体 k，其混合策略是一个向量 $\mathbf{x}_i(k)\in\Delta_i$，其中 Δ_i 是 $\mathbb{R}^{|S_i|}$中的单纯形，满足 $\sum_{s_i\in S_i}x_{is_i}(k)=1$和 $x_{is_i}(k)\geqslant 0$, $\forall s_i\in S_i$。每条边 $(i,j)\in E$定义了群体 i 和 j之间的一系列双人子博弈；在给定的时间步，每个群体 i 中的智能体都会随机与群体 j 中的另一名智能体配对，进行一场双人子博弈。我们分别用 $\mathbf{A}_{ij}\in\mathbb{R}^{|S_i|\times|S_j|}$ 和 $\mathbf{A}_{ji}\in\mathbb{R}^{|S_j|\times|S_i|}$表示这些双人子博弈中群体 i 和 j 的智能体所对应的收益矩阵。请注意，在给定的时间步，每个智能体都会选择一种（混合或纯）策略，并在所有双人子博弈中使用该策略。令 $\mathbf{x}=(\mathbf{x}_i,\{\mathbf{x}_j\}_{(i,j)\in E})$ 为一个混合策略剖面，其中 $\mathbf{x}_i$ （或 $\mathbf{x}j$）表示群体 i （或 j）中的一般混合策略。给定混合策略剖面 $\mathbf{x}$，在博弈 Γ 中采用 $\mathbf{x}_i$ 的期望收益为

$$r_i(\mathbf{x})=r_i(\mathbf{x}_i,\{\mathbf{x}_j\}_{(i,j)\in E}):=\sum_{(i,j)\in E}\mathbf{x}_i^\top\mathbf{A}_{ij}\mathbf{x}_j. \tag{1}$$

博弈 Γ是竞争性（或加权零和）的，当且仅当存在正数常量 $\omega_1,\ldots,\omega_n$，使得

$$\sum_{i\in V}\omega_i r_i(\mathbf{x})=\sum_{(i,j)\in E}\left(\omega_i\mathbf{x}_i^\top\mathbf{A}_{ij}\mathbf{x}_j+\omega_j\mathbf{x}_j^\top\mathbf{A}_{ji}\mathbf{x}_i\right)=0,\quad\forall\mathbf{x}\in\prod_{i\in V}\Delta_i. \tag{2}$$

另一方面，若对于每条边 $(i,j)\in E$，两人子博弈的收益矩阵满足$\mathbf{A}ij=\mathbf{A}_{ji}^\top$，则 Γ 是一个协调网络博弈

平滑虚拟博弈。SFP 是一种用于博弈学习的基于信念的模型。在 SFP 中，智能体对对手的行动形成信念，并通过平滑最优响应对这些信念作出响应。给定一个博弈 Γ，考虑群体 $i\in V$中的任意智能体 k。令 $V_i=\{j\in V\colon(i,j)\in E\}$ 为相邻群体的集合。智能体 k 为相邻群体 $j\in V_i$中的每种对手策略 $s_j\in S_j$ 维护一个权重 $\kappa^i_{js_j}(k)$。根据这些权重，智能体 k 对相邻群体 j形成信念，使得每种对手策略 s_j 以概率

$$\mu^i_{js_j}(k)=\frac{\kappa^i_{js_j}(k)}{\sum_{s'_j\in S_j}\kappa^i_{js'_j}(k)}. \tag{3}$$

令 $\boldsymbol{\mu}^i_j(k)$ 为信念向量，其第 s_j 个元素等于 $\mu^i_{js_j}(k)$。智能体 k 针对每个相邻群体分别形成信念，并对信念集合 $\{\boldsymbol{\mu}^i_j(k)\}_{j\in V_i}$作出平滑最优响应。给定博弈 Γ，智能体 k使用纯策略 $s_i\in S_i$ 的期望收益为

$$u_{is_i}(k)=r_i(\mathbf{e}_{s_i},\{\boldsymbol{\mu}^i_j(k,t)\}_{j\in V_i})=\sum_{j\in V_i}\mathbf{e}_{s_i}^\top\mathbf{A}_{ij}\boldsymbol{\mu}^i_j(k) \tag{4}$$

其中，$\mathbf{e}s_i$ 是一个单位向量，其第 s_i个元素为 $\mathbf{1}$。选择策略 s_i 的概率由下式给出

$$x_{is_i}(k)=\frac{\exp(\beta u_{is_i}(k))}{\sum_{s'_i\in S_i}\exp(\beta u_{is'_i}(k))} \tag{5}$$

其中，β 是温度（或理性程度）。我们假设智能体能够观察到每个邻居人群的平均混合策略。因此，在给定的时间步 t，智能体 k 按如下方式更新每种对手策略 $s_j\in S_j,\ j\in V_i$ 的权重:

$$\kappa^i_{js_j}(k,t+1)=\kappa^i_{js_j}(k,t)+\bar{x}_{js_j}(t) \tag{6}$$

其中，$\bar{x}_{js_j}$是群体 j中采取策略 s_j的平均概率，即 $\bar{x}_{js_j}=\frac{1}{n_j}\sum_{l\in\text{群体 }j}x_{js_j}(l)$，其智能体数量记为 n_j。为简便起见，我们假设系统 N中每个智能体的初始权重和$\sum_{s_j\in S_j}\kappa^i_{js_j}(k,\ 0)$相同，并将该初始权重和记为 λ。注意，公式 6 可以改写为

$$(\lambda+t+1)\mu^i_{js_j}(k,t+1)=(\lambda+t)\mu^i_{js_j}(k,t)+\bar{x}_{js_j}(t). \tag{7}$$

因此，尽管智能体 k 直接更新权重，但其个体状态可以用信念集合 $\{\boldsymbol{\mu}_j^i(k)\}_{j\in V_i}$来表征。为简化符号，以下通常省略括号中的时间索引 t 和智能体索引 k （具体省略哪个取决于上下文）。

3 种群网络博弈中的信念动态

注意，对于任意智能体 k，其信念 $\boldsymbol{\mu}_j^i(k)$ 位于单纯形 $\Delta_j = \{\boldsymbol{\mu}_j^i(k) \in \mathbb{R}^{|S_j|} | \sum_{s_j\in S_j} \mu_{js_j}^i(k) = 1, \mu_{js_j}^i(k) \geqslant 0, \forall s_j \in S_j\}$ 中。我们假设系统状态由定义在状态空间 $\Delta = \prod_{i\in V} \Delta_i$ 上的博雷尔概率测度P刻画。给定 $\boldsymbol{\mu}_i \in \Delta_i$，我们将边缘概率密度函数写作 $p(\boldsymbol{\mu}_i, t)$。注意， $p(\boldsymbol{\mu}_i, t)$ 是整个系统中对群体 i 的信念为 $\boldsymbol{\mu}_i$ 的智能体密度。定义 $\boldsymbol{\mu} = \{\boldsymbol{\mu}_i\}_{i\in V} \in \Delta$。由于智能体对不同邻居群体分别持有信念，联合概率密度函数 $p(\boldsymbol{\mu}, t)$ 可以因式分解，即 $p(\boldsymbol{\mu}, t) = \prod_{i\in V} p(\boldsymbol{\mu}_i, t)$。我们对初始边缘密度函数作出如下假设。

假设 1. 在时间 $t = 0$，对于每个群体 $i \in V$，边际密度函数 $p(\boldsymbol{\mu}_i, t)$ 连续可微，且在单纯形 Δ_i的边界处质量为零。

对于“良好”的概率分布而言，这一假设是标准且常见的。在这一温和条件下，我们使用与 [27, 23] 中类似的技术，通过以下命题确定系统状态 P 的演化。

命题1 (种群信念动态). 每个群体 $i \in V$的边际密度函数 $p(\boldsymbol{\mu}_i, t)$ 的连续时间动力学由以下偏微分方程支配

$$-\frac{\partial p(\boldsymbol{\mu}_i, t)}{\partial t} = \nabla \cdot \left(p(\boldsymbol{\mu}_i, t) \frac{\bar{\mathbf{x}}_i - \boldsymbol{\mu}_i}{\lambda + t + 1} \right) \tag{8}$$

其中， $\nabla\cdot$ 是散度算子，而 $\bar{\mathbf{x}}_i$ 是平均混合策略，其每个 s_i分量

$$\bar{x}_{is_i} = \int_{\prod_{j\in V_i} \Delta_j} \frac{\exp(\beta u_{is_i})}{\sum_{s_i'\in S_i} \exp(\beta u_{is_i'})} \prod_{j\in V_i} p(\boldsymbol{\mu}_j, t) \left(\prod_{j\in V_i} d\boldsymbol{\mu}_j \right) \tag{9}$$

其中 $u_{is_i} = \sum_{j\in V_i} \mathbf{e}_{s_i}^\top \mathbf{A}^{ij} \boldsymbol{\mu}_j$。

对于每个边际密度函数 $p(\boldsymbol{\mu}_i, t)$，总质量始终守恒（补充材料的推论 1）；此外，单纯形 Δ_i 边界处的质量始终为零，这表明智能体的信念永远不会走向极端（补充材料的推论 2）。

将系统状态的概念推广为信念分布，使我们能够探讨一个非常具体的问题——信念异质性对系统演化的影响。话虽如此，偏微分方程（公式 8）以难以求解而著称。在这里，我们根据分布的演化（公式 8）采用矩的演化方法。在下述命题中，我们表明，对信念异质性的刻画十分重要，因为平均系统状态的动态（或平均信念动态）确实会受到信念异质性的影响。

命题 2 (平均信念动态)。 每个群体$i \in V$的平均信念 $\bar{\mu}_i$ 动态由一个微分方程组所支配，对每个策略 s_i，

$$\frac{d\bar{\mu}_{is_i}}{dt} \approx \frac{f_{s_i}(\{\boldsymbol{\mu}_j\}_{j\in V_i}) - \bar{\mu}_{is_i}}{\lambda + t + 1} + \frac{\sum_{j\in V_i} \sum_{s_j\in S_j} \frac{\partial^2 f_{s_i}(\{\boldsymbol{\mu}_j\}_{j\in V_i})}{(\partial \mu_{js_j})^2} Var(\mu_{js_j})}{2(\lambda + t + 1)}. \tag{10}$$

其中， $f_{s_i}(\{\boldsymbol{\mu}\}_{j\in V_i})$ 是应用于策略 $s_i \in S_i$的 *logit* 选择函数（公式 *5*），*Var*(μ_{js_j}) 表示整个系统中信念 μ_{js_j} 的方差。

一般而言，平均信念动态同时受到信念分布的均值、方差以及无穷多个高阶矩的影响。为得到更具结论性的结果，我们采用矩闭合近似[4]，并假设三阶及更高阶矩的影响可以忽略不计。

现在，暂且假设系统信念是同质的——每个人的信念都相同。因此，平均信念动态实际上就是个体的信念动态。以下命题由公式 7 推出。

[4]矩闭合是用于估计群体模型矩的一种典型近似方法 [13, 15, 32]。使用矩闭合时，需要选定一个阶数，并将高于该阶数的所有累积量设为零。通常选择的阶数为 2，即将三阶及更高阶的累积量设为零。

命题 3 (**同质人群的信念动态**)。 对于同质系统，关于每个群体 $i \in V$ 的信念 $\boldsymbol{\mu}_i$ 动态由一个微分方程组支配，其中对每种策略 s_i,

$$\frac{d\mu_{is_i}}{dt} = \frac{x_{is_i} - \mu_{is_i}}{\lambda + t + 1} = \frac{f_{s_i}(\{\boldsymbol{\mu}_j\}_{j \in V_i}) - \mu_{is_i}}{\lambda + t + 1} \tag{11}$$

其中， μ_{is_i} 在每个邻居人群 $j \in V_i$中的所有智能体之间都相同。

直观而言，平均信念动态反映了系统中信念的变化趋势，而信念方差则反映了信念异质性。对比命题 2 和命题 3，可以清楚地看出，在异质系统中，信念方差（信念异质性）会影响平均信念动态（信念的变化趋势）。自然会产生这样的问题：信念异质性如何随时间演变？信念异质性对信念的变化趋势有多大影响？我们对这些问题的研究揭示了一个有趣的发现——信念方差渐近趋于零。

定理 1 (**人群信念方差的二次衰减**)。 每个群体 $i \in V$ 中信念 $\boldsymbol{\mu}_i$ 的动态由一个微分方程组支配，对于每种策略 s_i,

$$\frac{d\,Var(\mu_{is_i})}{dt} = -\frac{2\,Var(\mu_{is_i})}{\lambda + t + 1}. \tag{12}$$

在给定时刻 t，$Var(\mu_{is_i}) = \underbrace{\left(\frac{\lambda+1}{\lambda+t+1}\right)^2}\sigma^2(\mu_{is_i})$，其中 $\sigma^2(\mu_{is_i})$ 为初始方差。因此，方差 $Var(\mu_{is_i})$ 随时间以二次速度衰减至零。

无论智能体参与何种双人子博弈，也无论初始条件如何，方差都会以这种二次速度衰减。换言之，对于所有群体网络博弈，信念最终都会同质化。这一事实立即意味着极限下的系统状态。

推论 1. 随着时间 $t \to \infty$，每个群体 $i \in V$ 的密度函数 $p(\boldsymbol{\mu}_i, t)$ 都会演化为狄拉克 δ 函数，而每个群体 $i \in V$ 内选择分布的方差也会趋于零。

请注意，虽然每个群体内的选择分布会趋于同质，但不同群体之间的选择分布未必相同。这是因为每个群体的策略选择是对其自身相邻群体集合（通常彼此不同）作出的响应。

4 群体网络博弈中平滑虚拟博弈的收敛

关于信念同质化的发现并非显而易见，在技术上也十分重要。其一项推论是，初始信念异质的系统与初始信念齐次的系统具有相同的不动点。因此，根据齐次系统的信念动态（命题 3），可知系统的不动点具有以下性质。

定理 2 (**系统动态的不动点**)。 对于初始时具有同质或异质信念的任意系统，系统动力学的不动点是一对 $(\boldsymbol{\mu}^*, \mathbf{x}^*)$，满足 $\mathbf{x}_i^* = \boldsymbol{\mu}_i^*$ （对每个群体 $i \in V$ 均成立），且是以下方程组的解

$$x_{is_i}^* = \frac{\exp\left(\beta \sum_{j \in V_i} \mathbf{e}_{s_i}^\top \mathbf{A}_{ij} \mathbf{x}_j^*\right)}{\sum_{s_i' \in S_i} \exp\left(\beta \sum_{j \in V_i} \mathbf{e}_{s_i'}^\top \mathbf{A}_{ij} \mathbf{x}_j^*\right)} \tag{13}$$

对于每个策略 $s_i \in S_i$ 和群体 $i \in V$均成立。此类不动点始终存在，并且与种群网络博弈 Γ的量化反应均衡（*QRE*） *[33]* 一致。

请注意，上述定理适用于所有群体网络博弈。

我们研究在网络竞争和网络协调两种情形下，SFP 收敛到量化反应均衡的情况。受篇幅所限，以下主要关注网络竞争，并仅给出网络协调的主要结果。

4.1 网络竞争

考虑一个竞争性群体网络博弈 Γ。请注意，在竞争型网络博弈中，纳什均衡收益不一定唯一（这与双人情形明显不同），而且通常允许存在无穷多个纳什均衡。在以下定理中，我们聚焦于同质系统，证明信念动态会收敛到唯一的量化反应均衡，而不受底层博弈中纳什均衡数量的影响。

定理3 (**齐次网络竞争中的收敛**)。 给定一个竞争性 Γ，对于任何具有齐次信念的系统，信念动态（公式 *11*）都会收敛到唯一的量化反应均衡，该均衡是全局渐近稳定的。

证明概要。 我们通过证明 $\mathbf{x}_i$ 与 μ_i 之间的"距离"在达到量化反应均衡之前严格递减，来证明这一点。具体而言，我们用扰动收益来衡量距离，并构造一个严格的李雅普诺夫函数

$$L \coloneqq \sum_{i \in V} \omega_i \left[\pi_i \left(\mathbf{x}_i, \{\boldsymbol{\mu}_j\}_{j \in V_i} \right) - \pi_i \left(\boldsymbol{\mu}_i, \{\boldsymbol{\mu}_j\}_{j \in V_i} \right) \right] \tag{14}$$

其中， $\omega_1 \ldots \omega_n$ 是由 Γ给出的正权重， π_i 是定义为$\pi_i \left(\mathbf{x}_i, \ \{\boldsymbol{\mu}_j\}_{j \in V_i} \right) \coloneqq \mathbf{x}_i^\top \sum_{j \in V_i} A_{ij} \boldsymbol{\mu}_j - \frac{1}{\beta} \sum_{s_i \in S_i} x_{is_i} \ln(x_{is_i})$的扰动收益函数。 □

接下来，我们转向初始具有异质信念的系统。利用信念方差最终趋于零这一事实，我们建立以下引理。

引理 1. 对于初始信念存在异质性的系统，平均信念动态（方程 *10*）是渐近自治的 *[31]*，其极限方程为 $\frac{d\boldsymbol{\mu}_i}{dt} = \mathbf{x}_i - \boldsymbol{\mu}_i$；经过时间重参数化后，该方程等价于同质系统的信念动态（公式 *11* ）。

为便于说明，我们沿用惯例，分别用 ϕ 和 Θ表示渐近自治系统及其极限方程的解流。Thieme [40] 给出了以下奠基性结果，将 ϕ 和 Θ的极限行为联系起来。

引理 2 (Thieme [40] 定理 4.2)。 设 $(X$为度量空间， $d)$。假设 Θ的均衡是孤立的紧致 Θ-不变子集，且属于 X。任何预紧 Θ-轨道的 ω-Θ-极限集都包含一个Θ-均衡。点 $(s, x), s \geqslant t_0, x \in X$具有预紧的 ϕ-轨道。那么以下两种情况必有其一： *1)* 对某个 Θ-均衡 e， $\phi(t, s, x) \to e, t \to \infty$； *2)* (s, x) 的 ω-ϕ-极限集包含有限多个 Θ-均衡，它们以循环方式彼此相连。

结合以上结果，我们证明了初始异质系统的收敛性。

定理 4 (**初始异质网络竞争中的收敛性**)。 对于任意竞争性 Γ，任何初始具有异质信念的系统，其平均信念动态（方程 *10*）都会收敛到唯一的量化反应均衡。

以下推论可由信念同质化直接得出。

推论 2。 对于任意竞争性 Γ，在平滑虚拟博弈下，无论信念初始化情况如何，也无论 Γ 中纳什均衡的数量是多少，每个个体的选择分布和信念都会收敛到唯一的量化反应均衡（见定理 *2*）。

4.2 网络协调

我们将网络协调博弈的大部分结果放在补充材料中，这里仅总结主要结果。

定理 5 (**星形结构网络协调中的收敛**)。 给定一个协调Γ ，其网络结构由单个星形网络或多个不连通的星形结构组成，同质系统中信念动态（公式 *11*）的每个轨道，以及初始异质系统中平均信念动态（方程 *10*）的每个轨道，都会收敛到量化反应均衡集合。

请注意，由于只有两个顶点时，有无星形结构的网络博弈本质上相同，因此该定理适用于所有双群体协调博弈。我们还要指出，协调网络博弈中的纯纳什均衡或混合纳什均衡十分复杂；正如近期研究 [5, 4, 1], 所报告的，寻找纯纳什均衡是 PLS 完全的。因此，一般情况下的网络协调学习十分困难，通常需要满足某些条件才能进行理论分析 [34, 35]。

	H	S
H	(1, 1)	(2, 0)
S	(0, 2)	(4, 4)

表 1：鹿猎博弈。

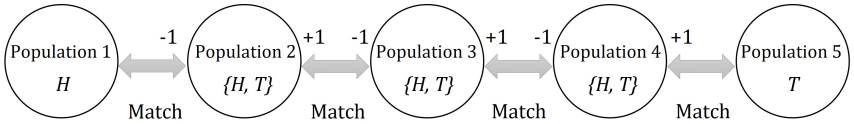


图 2：非对称匹配硬币博弈。

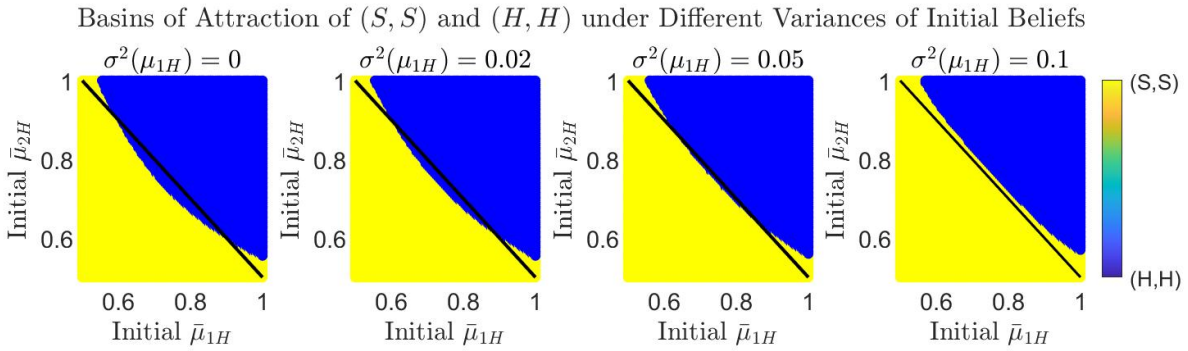


图 3：信念异质性有助于选择收益占优均衡 $(S,\ S)$（黄色：均衡$(S,\ S)$；蓝色：均衡 $(H,\ H)$）。随着初始信念的方差增大（从左图到右图），更多范围内的初始平均信念会在极限下近似达到均衡 $(S,\ S)$。对于每个图，两个人群的初始方差 $\sigma^2(\mu_{1H})$ 和 $\sigma^2(\mu_{2H})$ 相同。

5 实验：群体网络博弈中的均衡选择

本节将通过理论分析，并展示 SFP 在双群体协调（鹿猎）博弈和五群体零和（非对称匹配硬币）博弈中的实证研究，对我们的理论进行补充。重要的是，这两个博弈都存在多个纳什均衡，这自然引出了均衡选择问题。

5.1 双群体鹿猎博弈

我们在图1（引言部分）中展示了，在初始平均信念相同的情况下，改变初始信念的方差会导致不同的极限行为。接下来，我们将通过可视化不同均衡的吸引域如何受到影响，系统地研究初始信念异质性的作用。

博弈描述。我们考虑一个双群体鹿猎博弈，其中群体1和群体2中的每位玩家都有两个行动 $\{H,\ S\}$。如双矩阵收益（表1）所示，该博弈有两个纯策略纳什均衡： $(H,\ H)$ 和 $(S,\ S)$。虽然 $(H,\ H)$ 是风险占优均衡，但 $(S,\ S)$ 确实更理想，因为它既是收益占优均衡，也是帕累托最优均衡。

结果。 在这个博弈中，群体 1 对群体 2 形成信念，反之亦然。我们用一对数值 $(\bar{\mu}_{2H}, \bar{\mu}_{1H})$ 表示初始平均信念。给定初始信念的不同方差，我们对大量初始平均信念对应的平均信念动态进行数值求解。在图 3 中，对于每一对初始平均信念，我们根据系统最终收敛到哪一个量化反应均衡，为相应的数据点着色。我们观察到，随着初始信念的方差增大（从左图到右图），导致系统收敛到近似于收益占优均衡 (S, S) 的量化反应均衡的初始平均信念范围也随之扩大。换言之，较高的初始信念异质性程度会扩大 $(S,\ S)$ 的吸引域。因此，尽管信念异质性最终会消失，但它有助于实现均衡选择，因为它能帮助选出更理想的均衡。

5.2 五群体非对称匹配硬币博弈

我们已在推论 2 中证明，即使竞争性 Γ中存在多个纳什均衡，SFP 仍会收敛到唯一的量化反应均衡。接下来，我们将通过以下实证证据对此加以佐证：

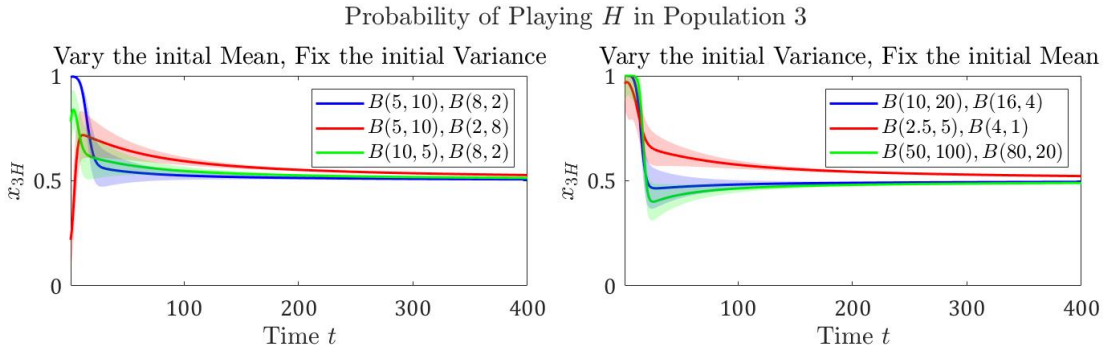


图 4：在不同的信念初始化下，SFP 选择一个唯一的均衡，在该均衡中，群体 3 中的所有智能体以概率 0.5 采取行动 H。对于每种初始化，我们运行 100 次仿真运行。细线表示平均混合策略（ H 的选择概率），阴影区域表示群体中混合策略的方差。在图例中， B 表示贝塔分布；两个贝塔分布分别对应于关于相邻群体 2 和 4 的初始信念。

采用不同信念初始化的基于智能体的仿真（仿真细节汇总于补充材料中）。

博弈描述。 考虑一个五群体非对称匹配硬币博弈 [28], ，其中网络结构呈线形（如图 2 所示）。每个智能体有两个行动 $\{H, T\}$。群体 1 和 5 中的智能体不进行学习；它们始终分别采取策略 H 和 T。对于群体 2 至 4 中的智能体，如果它们与下一群体中对手的策略匹配，则获得 $+1$ ；如果不匹配，则获得 -1 。相反，如果它们与上一群体中对手的策略不匹配，则获得 $+1$ ；如果匹配，则获得 -1。因此，该博弈有无穷多个纳什均衡，其形式为：群体 2 和 4 中的智能体采取策略 T，而群体 3 中的智能体在策略 H 和 T 之间无差异。

结果。 在这个博弈中，每个群体中的智能体都会形成两种信念（一种针对前一个群体，另一种针对下一个群体）。我们主要关注群体 3 的策略，因为纳什均衡在群体 3 的策略上有所不同。为进行验证，我们改变群体 3 对相邻群体 2 和 4 的信念，并固定群体 3 对其他群体的信念。如图 4 所示，在信念以不同方式初始化的情况下，群体 3 中的智能体会收敛到同一个均衡，其中它们都以概率 0.5 采取策略 H。因此，即使底层零和博弈存在多个纳什均衡，具有不同初始信念异质性的 SFP 仍会选出唯一均衡，从而解决均衡选择问题。

6 结论

我们研究了网络博弈中 SFP 的异质信念模型。通过用信念分布表示系统状态，我们证明，在所有网络博弈中，信念最终都会变得同质。我们证明，在一般竞争型网络博弈以及具有星形结构的协调网络博弈中，SFP 会收敛到量化反应均衡。实验结果表明，尽管初始信念异质性最终会消失，但它在均衡选择中发挥着关键作用，并有助于选出高度理想的均衡。

附录 A：第 3 节省略的推论与证明

命题1的证明

由正文中的公式 7 可知， $\boldsymbol{\mu}_j^i(k,t)$ 在两个离散时间步之间的变化为

$$\boldsymbol{\mu}_j^i(k,t+1) = \boldsymbol{\mu}_j^i(k,t) + \frac{\bar{\mathbf{x}}_j(t) - \boldsymbol{\mu}_j^i(k,t)}{\lambda + t + 1}. \tag{15}$$

引理 3. 在正文中的假设 *1* 下，对于群体 i中的任意智能体 k，其关于邻居人群 j 的信念$\boldsymbol{\mu}_j^i(k,t)$ 永远不会达到极端信念（即单纯形Δ_i的边界）。

证明。 假设 1 确保 $\bar{\mathbf{x}}_j(0)$ 位于单纯形 Δ_j的内部。此外，对于有限温度 β，logit 选择函数（正文中的公式 5）也确保此后 $\bar{\mathbf{x}}_j(t)$ 始终位于 Δ_j 的内部。因此，由公式 15 可知，对于每个时间步 t， $\boldsymbol{\mu}_j^i(k,t)$ 都将位于 Δ_j的内部。 □

下文中，为方便记号起见，我们有时会根据上下文省略智能体索引 k 和时间索引t 。考虑一个群体 i。我们将关于该群体的信念变化改写如下。

$$\boldsymbol{\mu}_i(t+1) = \boldsymbol{\mu}_i(t) + \frac{\bar{\mathbf{x}}_i(t) - \boldsymbol{\mu}_i(t)}{\lambda + t + 1}. \tag{16}$$

假设相邻两个时间步之间经过的时间为 $\delta \in (0,1]$。我们将上面的方程改写为

$$\boldsymbol{\mu}_i(t+\delta) = \boldsymbol{\mu}_i(t) + \delta\frac{\bar{\mathbf{x}}_i(t) - \boldsymbol{\mu}_i(t)}{\lambda + t + 1}. \tag{17}$$

接下来，我们考虑一个测试函数 $\theta(\boldsymbol{\mu}_i)$。定义

$$Y = \frac{\mathbb{E}[\theta(\boldsymbol{\mu}_i(t+\delta))] - \mathbb{E}[\theta(\boldsymbol{\mu}_i(t))]}{\delta}. \tag{18}$$

在 $\boldsymbol{\mu}_i(t)$处对 $\theta(\boldsymbol{\mu}_i(t+\delta))$ 应用泰勒级数，得到

$$\begin{aligned}\theta(\boldsymbol{\mu}_i(t+\delta)) = \theta(\boldsymbol{\mu}_i(t)) &+ \frac{\delta}{\lambda+t+1}\partial_{\boldsymbol{\mu}_i}\theta(\boldsymbol{\mu}_i)\left[\bar{\mathbf{x}}_i(t) - \boldsymbol{\mu}_i(t)\right] \\ &+ \frac{\delta^2}{2(\lambda+t+1)^2}\left[\bar{\mathbf{x}}_i(t) - \boldsymbol{\mu}_i(t)\right]^\top \mathbf{H}\theta(\boldsymbol{\mu}_i)\left[\bar{\mathbf{x}}_i(t) - \boldsymbol{\mu}_i(t)\right] \\ &+ o\left(\left[\delta\frac{\bar{\mathbf{x}}_i(t) - \boldsymbol{\mu}_i(t)}{\lambda+t+1}\right]^2\right)\end{aligned} \tag{19}$$

其中$\mathbf{H}$ 表示海森矩阵。因此，期望 $\mathbb{E}[\theta(\boldsymbol{\mu}_i(t+\delta))]$ 为

$$\begin{aligned}\mathbb{E}[\theta(\boldsymbol{\mu}_i(t+\delta))] = \mathbb{E}[\theta(\boldsymbol{\mu}_i(t))] &+ \frac{\delta}{\lambda+t+1}\mathbb{E}[\partial_{\boldsymbol{\mu}_i}\theta(\boldsymbol{\mu}_i(t))(\bar{\mathbf{x}}_i(t) - \boldsymbol{\mu}_i(t))] \\ &+ \frac{\delta^2}{2(\lambda+t+1)^2}\mathbb{E}\left[\left[\bar{\mathbf{x}}_i(t) - \boldsymbol{\mu}_i(t)\right]^\top \mathbf{H}\theta(\boldsymbol{\mu}_i)\left[\bar{\mathbf{x}}_i(t) - \boldsymbol{\mu}_i(t)\right]\right] \\ &+ \frac{\delta^2}{2(\lambda+t+1)^2}\mathbb{E}[o([\bar{\mathbf{x}}_i(t) - \boldsymbol{\mu}_i(t)]^2)]\end{aligned} \tag{20}$$

将 $\mathbb{E}[\theta(\boldsymbol{\mu}_i(t))]$ 移至等式左侧，并将等式两边除以 δ，即可得到量Y，即

$$\begin{aligned}Y = &\frac{1}{\lambda+t+1}\mathbb{E}[\partial_{\boldsymbol{\mu}_i}\theta(\boldsymbol{\mu}_i(t))(\bar{\mathbf{x}}_i(t) - \boldsymbol{\mu}_i(t))] \\ &+ \frac{\delta}{2(\lambda+t+1)^2}\mathbb{E}[[\bar{\mathbf{x}}_i(t) - \boldsymbol{\mu}_i(t)]^\top \mathbf{H}\theta(\boldsymbol{\mu}_i(t))[\bar{\mathbf{x}}_i(t) - \boldsymbol{\mu}_i(t)] + o\left((\bar{\mathbf{x}}_i(t) - \boldsymbol{\mu}_i(t))^2\right)]\end{aligned} \tag{21}$$

令 Y趋于 $\delta \to 0$，右侧第二项的贡献消失，从而得到

$$\lim_{\delta\to 0} Y = \frac{1}{\lambda+t+1}\mathbb{E}[\partial_{\boldsymbol{\mu}_i}\theta(\boldsymbol{\mu}_i(t))(\bar{\mathbf{x}}_i(t) - \boldsymbol{\mu}_i(t))] \tag{22}$$

$$= \frac{1}{\lambda+t+1}\int p(\boldsymbol{\mu}_i(t), t)\left[\partial_{\boldsymbol{\mu}_i}\theta(\boldsymbol{\mu}_i(t))(\bar{\mathbf{x}}_i(t) - \boldsymbol{\mu}_i(t))\right] d\boldsymbol{\mu}_i(t). \tag{23}$$

应用分部积分。我们得到

$$\lim_{\delta\to 0} Y = 0 - \frac{1}{\lambda+t+1}\int \theta(\boldsymbol{\mu}_i(t))\nabla\cdot\left[p(\boldsymbol{\mu}_i(t), t)(\bar{\mathbf{x}}_i(t) - \boldsymbol{\mu}_i(t))\right] d\boldsymbol{\mu}_i(t) \tag{24}$$

其中我们利用了引理 1 的结论：边界 $\partial\Delta_i$ 处的概率质量 $p(\boldsymbol{\mu}_i,t)$ 仍为零。另一方面，根据 Y 的定义，

$$\lim_{\delta\to 0} Y = \lim_{\delta\to 0}\int \theta(\boldsymbol{\mu}_i(t))\frac{p(\boldsymbol{\mu}_i,t+\delta)-p(\boldsymbol{\mu}_i,t)}{\delta}d\boldsymbol{\mu}_i = \int \theta(\boldsymbol{\mu}_i(t))\partial_t p(\boldsymbol{\mu}_i,t)d\boldsymbol{\mu}_i. \tag{25}$$

因此，我们得到等式

$$\int \theta(\boldsymbol{\mu}_i(t))\partial_t p(\boldsymbol{\mu}_i,t)d\boldsymbol{\mu}_i = -\frac{1}{\lambda+t+1}\int \theta(\boldsymbol{\mu}_i(t))\nabla\cdot\left[p(\boldsymbol{\mu}_i(t),t)(\bar{\mathbf{x}}_i(t)-\boldsymbol{\mu}_i(t))\right]d\boldsymbol{\mu}_i(t). \tag{26}$$

由于 θ 是一个测试函数，因此可得

$$\partial_t p(\boldsymbol{\mu}_i,t) = -\frac{1}{\lambda+t+1}\nabla\cdot\left[p(\boldsymbol{\mu}_i(t),t)(\bar{\mathbf{x}}_i(t)-\boldsymbol{\mu}_i(t))\right]. \tag{27}$$

重新排列各项，即可得到正文中的公式 8。根据概率分布下期望的定义，很容易得到正文中的方程 9。证毕。

备注： 我们推导出的偏微分方程与物理学中研究守恒量时常见的连续性方程类似。连续性方程描述物理系统中的输运现象（例如质量或能量的输运）。这为我们的偏微分方程模型提供了一种物理解释：在 *SFP* 下，异质系统的信念动态类似于智能体质量在单纯形 $\Delta=\prod_{i\in V}\Delta_i$中的输运。

命题1的推论

推论3。 对于任意群体 $i\in V$，系统对该群体的信念绝不会走向极端。

证明。 这是引理 1 的直接推论。 □

推论 4. 对于任意群体 $i\in V$，总概率质量 $p(\boldsymbol{\mu}_i,t)$ 始终保持守恒 *d.*

证明。 考虑总概率质量的时间导数

$$\frac{d}{dt}\int p(\boldsymbol{\mu}_i,t)d\boldsymbol{\mu}_i. \tag{28}$$

应用莱布尼茨法则交换微分与积分，

$$\frac{d}{dt}\int p(\boldsymbol{\mu}_i,t)d\boldsymbol{\mu}_i = \int \frac{\partial p(\boldsymbol{\mu}_i,t)}{\partial t}d\boldsymbol{\mu}_i. \tag{29}$$

将正文中的~~公式 8~~ 代入 $\frac{\partial p(\boldsymbol{\mu}_i}{\partial t}\ t)$

$$\begin{aligned}
&\frac{d}{dt}\int p(\boldsymbol{\mu}_i,\ t)d\boldsymbol{\mu}_i \\
&= -\int \nabla\cdot\left(p(\boldsymbol{\mu}_i,\ t)\ \frac{\bar{\mathbf{x}}_i-\boldsymbol{\mu}_i}{\lambda+t+1}\right)d\boldsymbol{\mu}_i &(30)\\
&= -\int \sum_{s_i\in S_i}\partial_{\mu_{is_i}}\left(p(\boldsymbol{\mu}_i,\ t)\frac{\bar{x}_{is_i}-\mu_{is_i}}{\lambda+t+1}\right)d\boldsymbol{\mu}_i &(31)\\
&= -\frac{1}{\lambda+t+1}\left[\int \sum_{s_i\in S_i}\partial_{\mu_{is_i}}p(\boldsymbol{\mu}_i,\ t)\left(\bar{x}_{is_i}-\mu_{is_i}\right)d\boldsymbol{\mu}_i + \int p(\boldsymbol{\mu}_i,\ t)\sum_{s_i\in S_i}\partial_{\mu_{is_i}}\left(\bar{x}_{is_i}-\mu_{is_i}\right)d\boldsymbol{\mu}_i\right] &(32)
\end{aligned}$$

应用分部积分，

$$\int \sum_{s_i\in S_i}\partial_{\mu_{is_i}}p(\boldsymbol{\mu}_i,t)\left(\bar{x}_{is_i}-\mu_{is_i}\right)d\boldsymbol{\mu}_i = 0 - \int p(\boldsymbol{\mu}_i,t)\sum_{s_i\in S_i}\partial_{\mu_{is_i}}\left(\bar{x}_{is_i}-\mu_{is_i}\right)d\boldsymbol{\mu}_i. \tag{33}$$

这里我们利用了边界 $\partial\Delta_i$ 处的概率质量 $p(\boldsymbol{\mu}_i,t)$ 始终为零这一点。因此，公式 32 中括号内的各项相互抵消，并且

$$\frac{d}{dt}\int p(\boldsymbol{\mu}_i,t)d\boldsymbol{\mu}_i = 0. \tag{34}$$

□

命题 2 的证明

引理 4. 关于各群体 $i \in V$ 的平均信念 $\bar{\boldsymbol{\mu}}_i$ 的动态由以下微分方程支配

$$\frac{d\bar{\mu}_{is_i}}{dt} = \frac{\bar{x}_{is_i} - \bar{\mu}_{is_i}}{\lambda + t + 1}, \qquad \forall s_i \in S_i. \tag{35}$$

证明。 关于策略 s_i 的平均信念的时间导数为

$$\frac{d\bar{\mu}_{is_i}}{dt} = \frac{d}{dt}\int \mu_{is_i} p(\boldsymbol{\mu}_i, t) d\boldsymbol{\mu}_i. \tag{36}$$

我们应用莱布尼茨法则交换微分与积分，然后代入正文中的公式 8 $\frac{\partial p(\boldsymbol{\mu}_i,\ t)}{\partial t}$。

$$\frac{d}{dt}\int \mu_{is_i} p(\boldsymbol{\mu}_i, t) d\boldsymbol{\mu}_i \tag{37}$$

$$= \int \mu_{is_i} \frac{\partial p(\boldsymbol{\mu}_i,\ t)}{\partial t} d\boldsymbol{\mu}_i \tag{38}$$

$$= -\int \mu_{is_i} \nabla \cdot \left(p(\boldsymbol{\mu}_i, t)\ \frac{\bar{\mathbf{x}}_{i} - \boldsymbol{\mu}_i}{\lambda + t + 1} \right) d\boldsymbol{\mu}_i \tag{39}$$

$$= -\int \mu_{is_i} \sum_{s_i \in S_i} \partial_{\mu_{is_i}} \left(p(\boldsymbol{\mu}_i, t) \frac{\bar{x}_{is_i} - \mu_{is_i}}{\lambda + t + 1} \right) d\boldsymbol{\mu}_i \tag{40}$$

$$= \gamma \left[\int \mu_{is_i} \sum_{s_i \in S_i} \left(\partial_{\mu_{is_i}} p(\boldsymbol{\mu}_i, t) \right) \left(\bar{x}_{is_i} - \mu_{is_i} \right) d\boldsymbol{\mu}_i + \int \mu_{is_i} p(\boldsymbol{\mu}_i, t) \sum_{s_i \in S_i} \partial_{\mu_{is_i}} \left(\bar{x}_{is_i} - \mu_{is_i} \right) d\boldsymbol{\mu}_i \right] \tag{41}$$

其中 $\gamma := -\frac{1}{\lambda+t+1}$。 对方程 41 中的第一项应用分部积分。

$$\begin{aligned} &\int \mu_{is_i} \sum_{s_i \in S_i} \left(\partial_{\mu_{is_i}} p(\boldsymbol{\mu}_i, t) \right) \left(\bar{x}_{is_i} - \mu_{is_i} \right) d\boldsymbol{\mu}_i \\ &= -\int \mu_{is_i} p(\boldsymbol{\mu}_i, t) \left[\sum_{s_i' \in S_i} \partial_{\mu_{is_i'}} (\bar{x}_{is_i'} - \mu_{is_i'}) \right] + p(\boldsymbol{\mu}_i, t) \partial_{\mu_{is_i}} \left[\mu_{is_i} (\bar{x}_{is_i} - \mu_{is_i}) \right] d\boldsymbol{\mu}_i \end{aligned} \tag{42}$$

这里我们利用了边界处的概率质量始终为零这一点。因此，由方程 41 可得

$$\frac{d}{dt}\int \mu_{is_i} p(\boldsymbol{\mu}_i, t) d\boldsymbol{\mu}_i \tag{43}$$

$$\begin{aligned} &= -\gamma \int \mu_{is_i} p(\boldsymbol{\mu}_i, t) \sum_{s_i' \in S_i} \partial_{\mu_{is_i'}} (\bar{x}_{is_i'} - \mu_{is_i'}) d\boldsymbol{\mu}_i - \gamma \int p(\boldsymbol{\mu}_i, t) \partial_{\mu_{is_i}} \left[\mu_{is_i} (\bar{x}_{is_i} - \mu_{is_i}) \right] d\boldsymbol{\mu}_i \\ &\quad + \gamma \int \mu_{is_i} p(\boldsymbol{\mu}_i, t) \sum_{s_i \in S_i} \partial_{\mu_{is_i}} \left(\bar{x}_{is_i} - \mu_{is_i} \right) d\boldsymbol{\mu}_i \end{aligned} \tag{44}$$

$$= \gamma \int p(\boldsymbol{\mu}_i, t) \left[\mu_{is_i} \partial_{\mu_{is_i}} \left(\bar{x}_{is_i} - \mu_{is_i} \right) - \partial_{\mu_{is_i}} \left[\mu_{is_i} (\bar{x}_{is_i} - \mu_{is_i}) \right] \right] d\boldsymbol{\mu}_i \tag{45}$$

$$= \gamma \int p(\boldsymbol{\mu}_i, t) \mu_{is_i} d\boldsymbol{\mu}_i - \int p(\boldsymbol{\mu}_i, t) \bar{x}_{is_i} d\boldsymbol{\mu}_i \tag{46}$$

$$= \frac{\bar{x}_{is_i} - \bar{\mu}_{is_i}}{\lambda + t + 1} \tag{47}$$

□

我们再次给出正文方程 9 中的平均概率 $\bar{x}_{is_i}$，如下所示:

$$\bar{x}_{is_i} = \int \frac{\exp\left(\beta u_{is_i}\right)}{\sum_{s_i' \in S_i} \exp\left(\beta u_{is_i'}\right)} \prod_{j \in V_i} p(\boldsymbol{\mu}_j, t) \left(\prod_{j \in V_i} d\boldsymbol{\mu}_j \right) \tag{48}$$

其中 $u_{is_i} = \sum_{j\in V_i} \mathbf{e}_{s_i}^\top \mathbf{A}^{ij} \boldsymbol{\mu}_j$。定义 $\bar{\boldsymbol{\mu}} := \{\bar{\boldsymbol{\mu}}_j\}_{j\in V_i}$ 和

$$f_{s_i}(\{\boldsymbol{\mu}_j\}_{j\in V_i}) := \frac{\exp\left(\beta \sum_{j\in V_i} \mathbf{e}_{s_i}^\top \mathbf{A}_{ij} \boldsymbol{\mu}_j\right)}{\sum_{s_i'\in S_i} \exp\left(\beta \sum_{j\in V_i} \mathbf{e}_{s_i'}^\top \mathbf{A}_{ij} \boldsymbol{\mu}_j\right)}. \tag{49}$$

对函数在平均信念 $\bar{\boldsymbol{\mu}}$ 处应用泰勒展开进行近似，我们得到 ve

$$f_{s_i}(\{\boldsymbol{\mu}_j\}_{j\in V_i}) \approx f_{s_i}(\bar{\boldsymbol{\mu}}) + \nabla f_{s_i}(\bar{\boldsymbol{\mu}}) \cdot (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}) + \frac{1}{2}(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}})^\top \mathbf{H} f_{s_i}(\bar{\boldsymbol{\mu}})(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}) + O(||\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}||^3) \tag{50}$$

其中 $\mathbf{H}$ 表示海森矩阵。因此，我们可以将公式 48 重写为

$$\begin{aligned}
\bar{x}_{is_i} &= \int f_{s_i}(\{\boldsymbol{\mu}_j\}_{j\in V_i}) \prod_{j\in V_i} p(\boldsymbol{\mu}_j, t) \left(\prod_{j\in V_i} d\boldsymbol{\mu}_j\right) \qquad (51)\\
&\approx f_{s_i}(\bar{\boldsymbol{\mu}}) + \int \nabla f_{s_i}(\bar{\boldsymbol{\mu}}) \cdot \boldsymbol{\mu} \prod_{j\in V_i} p(\boldsymbol{\mu}_j, t) \left(\prod_{j\in V_i} d\boldsymbol{\mu}_j\right) - \nabla f_{s_i}(\bar{\boldsymbol{\mu}}) \cdot \bar{\boldsymbol{\mu}} \\
&\quad + \int \frac{1}{2}(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}})^\top \mathbf{H} f_{s_i}(\bar{\boldsymbol{\mu}})(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}) \prod_{j\in V_i} p(\boldsymbol{\mu}_j, t) \left(\prod_{j\in V_i} d\boldsymbol{\mu}_j\right) \\
&\quad + \int O(||\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}||)^3 \prod_{j\in V_i} p(\boldsymbol{\mu}_j, t) \left(\prod_{j\in V_i} d\boldsymbol{\mu}_j\right) \qquad (52)
\end{aligned}$$

注意，在公式 52 中，第二项和第三项可以相互抵消。此外，对于任意两个相邻群体 $j,\ k \in V_i$，关于这两个群体的信念 $\boldsymbol{\mu}_j,\ \boldsymbol{\mu}_k$ 彼此独立。因此，这些信念的协方差为零。我们应用二阶矩闭合近似[32, 13]，得到

$$\bar{x}_{is_i} \approx f_{s_i}(\bar{\boldsymbol{\mu}}) + \frac{1}{2} \sum_{j\in V_i} \sum_{s_j\in S_j} \frac{\partial^2 f_{s_i}(\bar{\boldsymbol{\mu}})}{(\partial \mu_{js_j})^2} \mathrm{Var}(\mu_{js_j}). \tag{53}$$

因此，将 上述近似代入引理 4 中的 $\bar{x}_{is_i}$，即可得到平均信念的动态方程 s

$$\frac{d\bar{\mu}_{is_i}}{dt} \approx \frac{f_{s_i}(\bar{\boldsymbol{\mu}}) - \bar{\mu}_{is_i}}{\lambda + t + 1} + \frac{\sum_{j\in V_i} \sum_{s_j\in S_j} \frac{\partial^2 f_{s_i}(\bar{\boldsymbol{\mu}})}{(\partial \mu_{js_j})^2} \mathrm{Var}(\mu_{js_j})}{2(\lambda + t + 1)}. \tag{54}$$

证毕。

备注：使用矩闭合近似（仅考虑一阶矩和二阶矩）是为了获得更具结论性的结果。严格来说，平均信念动态也取决于三阶及更高阶矩。然而，我们在实验中观察到，这些矩通常对平均信念动态影响甚微。具体而言，在初始平均信念相同的情况下，初始信念的方差有时会改变系统的极限行为，但我们没有观察到三阶及更高阶矩也会出现类似现象。

命题 3 的证明

考虑 一个群体 i。由正文中的公式 7 可知，该群体中信念的变化可写为如下形式。 fs
关于这个 群体可以写为如下形式。

$$\boldsymbol{\mu}_i(t+1) = \boldsymbol{\mu}_i(t) + \frac{\mathbf{x}_i(t) - \boldsymbol{\mu}_i(t)}{\lambda + t + 1}. \tag{55}$$

假设两个连续时间步之间经过的时间为 $\delta \in (0,\ 1]$。我们将上述方程改写为

$$\boldsymbol{\mu}_i(t+\delta) = \boldsymbol{\mu}_i(t) + \delta \frac{\mathbf{x}_i(t) - \boldsymbol{\mu}_i(t)}{\lambda + t + 1}. \tag{56}$$

将项 $\boldsymbol{\mu}_i(t)$ 移到右边，并将等式两边除以 δ，

$$\frac{\boldsymbol{\mu}_i(t+\delta) - \boldsymbol{\mu}_i(t)}{\delta} = \frac{\mathbf{x}_i(t) - \boldsymbol{\mu}_i(t)}{\lambda + t + 1}. \tag{57}$$

假设相邻两个时间步之间的时间间隔 δ 趋于零。我们有 e

$$\frac{d\boldsymbol{\mu}_i}{dt} = \lim_{\delta \to 0} \frac{\boldsymbol{\mu}_i(t+\delta) - \boldsymbol{\mu}_i(t)}{\delta} = \frac{\mathbf{x}_i(t) - \boldsymbol{\mu}_i(t)}{\lambda + t + 1}. \tag{58}$$

注意，对于连续时间动力学，我们通常省略括号中的时间索引，从而得到命题 3 中的信念动态（公式 11）。证毕。

定理 1 的证明

不失一般性，我们考虑群体 i 对策略 s_i 的信念 μ_{is_i} 的方差。注意，

$$\mathrm{Var}(\mu_{is_i}) = \mathbb{E}[(\mu_{is_i})^2] - (\bar{\mu}_{is_i})^2. \tag{59}$$

因此，我们有

$$\frac{d\mathrm{Var}(\mu_{is_i})}{dt} = \frac{d\mathbb{E}[(\mu_{is_i})^2]}{dt} - 2\bar{\mu}_{is_i}\frac{d\bar{\mu}_{is_i}}{dt}. \tag{60}$$

考虑右侧的第一项。我们应用莱布尼茨法则交换微分与积分，然后代入正文中的公式 8 $\frac{\partial p(\boldsymbol{\mu}_i,\ t)}{\partial t}$。

$$\begin{aligned}
&\frac{d\mathbb{E}[(\mu_{is_i})^2]}{dt} \\
&= \int (\mu_{is_i})^2 \frac{\partial p(\boldsymbol{\mu}_i,\ t)}{\partial t} d\boldsymbol{\mu}_i && (61)\\
&= -\int (\mu_{is_i})^2 \nabla \cdot \left(p(\boldsymbol{\mu}_i,\ t)\ \frac{\bar{\mathbf{x}}_i - \boldsymbol{\mu}_i}{\lambda + t + 1} \right) d\boldsymbol{\mu}_i && (62)\\
&= -\int (\mu_{is_i})^2 \sum_{s_i \in S_i} \partial_{\mu_{is_i}} \left(p(\boldsymbol{\mu}_i, t) \frac{\bar{x}_{is_i} - \mu_{is_i}}{\lambda + t + 1} \right) d\boldsymbol{\mu}_i && (63)\\
&= \gamma \int (\mu_{is_i})^2 \sum_{s_i \in S_i} \partial_{\mu_{is_i}} p(\boldsymbol{\mu}_i, t) \left(\bar{x}_{is_i} - \mu_{is_i} \right) d\boldsymbol{\mu}_i + \gamma \int (\mu_{is_i})^2 p(\boldsymbol{\mu}_i, t) \sum_{s_i \in S_i} \partial_{\mu_{is_i}} \left(\bar{x}_{is_i} - \mu_{is_i} \right) d\boldsymbol{\mu}_i && (64)
\end{aligned}$$

其中 $\gamma := -\frac{1}{\lambda+t+1}$。对公式 64 中的第一项应用分部积分，得到

$$\begin{aligned}
&\int (\mu_{is_i})^2 \sum_{s_i \in S_i} \partial_{\mu_{is_i}} p(\boldsymbol{\mu}_i, t) \left(\bar{x}_{is_i} - \mu_{is_i} \right) d\boldsymbol{\mu}_i \\
&= -\int (\mu_{is_i})^2 p(\boldsymbol{\mu}_i, t) \left[\sum_{s_i' \in S_i} \partial_{\mu_{is_i'}} (\bar{x}_{is_i'} - \mu_{is_i'}) \right] + p(\boldsymbol{\mu}_i, t) \partial_{\mu_{is_i}} \left[(\mu_{is_i})^2 (\bar{x}_{is_i} - \mu_{is_i}) \right] d\boldsymbol{\mu}_i && (65)
\end{aligned}$$

其中，我们利用了边界处的概率质量始终为零这一事实。结合上述两个方程，我们得到

$$\begin{aligned}
&\frac{d\mathbb{E}[(\mu_{is_i})^2]}{dt} \\
&= -\gamma \int (\mu_{is_i})^2 p(\boldsymbol{\mu}_i, t) \left[\sum_{s_i' \in S_i} \partial_{\mu_{is_i'}} (\bar{x}_{is_i'} - \mu_{is_i'}) \right] + p(\boldsymbol{\mu}_i, t) \partial_{\mu_{is_i}} \left[(\mu_{is_i})^2 (\bar{x}_{is_i} - \mu_{is_i}) \right] d\boldsymbol{\mu}_i \\
&\quad + \gamma \int (\mu_{is_i})^2 p(\boldsymbol{\mu}_i, t) \sum_{s_i \in S_i} \partial_{\mu_{is_i}} \left(\bar{x}_{is_i} - \mu_{is_i} \right) d\boldsymbol{\mu}_i && (66)\\
&= \gamma \int \left[-p(\boldsymbol{\mu}_i, t) \partial_{\mu_{is_i}} \left[(\mu_{is_i})^2 (\bar{x}_{is_i} - \mu_{is_i}) \right] \right] + (\mu_{is_i})^2 p(\boldsymbol{\mu}_i, t) \partial_{\mu_{is_i}} \left(\bar{x}_{is_i} - \mu_{is_i} \right) d\boldsymbol{\mu}_i && (67)\\
&= \gamma \int 2(\mu_{is_i})^2 p(\boldsymbol{\mu}_i, t) d\boldsymbol{\mu}_i - \gamma \int 2\bar{x}_{is_i} \mu_{is_i} p(\boldsymbol{\mu}_i, t) d\boldsymbol{\mu}_i && (68)\\
&= -\frac{2\mathbb{E}[(\mu_{is_i})^2] - 2\bar{x}_{is_i}\bar{\mu}_{is_i}}{\lambda + t + 1}. && (69)
\end{aligned}$$

接下来，我们考虑公式 60 中的第二项。根据引理 4，我们有

$$2\bar{\mu}_{is_i}\frac{d\bar{\mu}_{is_i}}{dt} = \frac{2\bar{\mu}_{is_i}(\bar{x}_{is_i} - \bar{\mu}_{is_i})}{\lambda + t + 1}. \tag{70}$$

结合公式 69 和 70，方差的动态为

$$\frac{d\mathrm{Var}(\mu_{is_i})}{dt} = -\frac{2\mathbb{E}[(\mu_{is_i})^2] - 2\bar{x}_{is_i}\bar{\mu}_{is_i}}{\lambda + t + 1} - \frac{2\bar{\mu}_{is_i}(\bar{x}_{is_i} - \bar{\mu}_{is_i})}{\lambda + t + 1} \tag{71}$$

$$= \frac{2(\bar{\mu}_{is_i})^2 - 2\mathbb{E}[(\mu_{is_i})^2]}{\lambda + t + 1} \tag{72}$$

$$= -\frac{2\mathrm{Var}(\mu_{is_i})}{\lambda + t + 1}. \tag{73}$$

证毕。

备注： 我们认为，这种现象背后的原因有两个：*1*）智能体采用平滑虚拟博弈；*2*）智能体响应的是其他群体的平均策略选择，而不是某些固定智能体的策略选择。关于前者，我们注意到，在类似的设置下，如果智能体采用其他学习方法，例如 *Q* 学习和交叉学习，则群体同质化可能不会发生。关于后者，设想智能体根据某些固定智能体的策略调整其信念。例如，考虑两个群体；其中一个群体包含智能体 *A* 和 *C*，另一个群体包含智能体 *B* 和 *D*。假设智能体 *A* 和 *B* 组成一个固定配对，只根据彼此的策略调整信念；智能体 *C* 和 *D* 也同样如此。信念同质化可能不会发生。

附录 B：第 4.1 节中省略的证明

定理 2 的证明

信念同质化意味着，初始具有异质信念的系统的不动点与具有齐次信念的系统中的不动点相同。因此，我们聚焦于齐次系统来分析不动点。不难看出

$$\frac{d\boldsymbol{\mu}_i}{dt} = \frac{\mathbf{x}_i - \boldsymbol{\mu}_i}{\lambda + t + 1} = 0 \implies \mathbf{x}_i = \boldsymbol{\mu}_i. \tag{74}$$

将满足上述公式的系统动力学不动点记为 $(\mathbf{x}_i^*, \boldsymbol{\mu}_i^*)$ 对每个群体 i 而言。根据 logit 选择函数（正文中的公式 5），我们有

$$x_{is_i}^* = \frac{\exp(\beta u_{is_i})}{\sum_{s_i' \in S_i}\exp(\beta u_{is_i'})} = \frac{\exp(\beta \sum_{j\in V_i}\mathbf{e}_{s_i}^\top \mathbf{A}_{ij}\boldsymbol{\mu}_j^*)}{\sum_{s_i'\in S_i}\exp(\beta\sum_{j\in V_i}\mathbf{e}_{s_i'}^\top \mathbf{A}_{ij}\boldsymbol{\mu}_j^*)}. \tag{75}$$

利用不动点处的 $\mathbf{x}_i^* = \boldsymbol{\mu}_i^*, \forall i \in V$，我们可以将 $\boldsymbol{\mu}_j^*$ 替换为 $\mathbf{x}_j^*$。证毕。

定理3的证明

考虑一个群体 i。相邻群体的集合为 V_i，关于相邻群体的信念集合为 $\{\boldsymbol{\mu}_j\}_{j\in V_i}$，选择分布为 $\mathbf{x}_i$。给定一个种群网络博弈 Γ，期望收益为 $\mathbf{x}_i^\top \sum_{(i,j)\in E} A_{ij}\boldsymbol{\mu}_j$。定义扰动收益函数

$$\pi_i\left(\mathbf{x}_i, \{\boldsymbol{\mu}_j\}_{j\in V_i}\right) \coloneqq \mathbf{x}_i^\top \sum_{j\in V_i} A_{ij}\boldsymbol{\mu}_j + v(\mathbf{x}_i) \tag{76}$$

其中 $v(\mathbf{x}_i) = -\frac{1}{\beta}\sum_{s_i\in S_i} x_{is_i}\ln(x_{is_i})$。在这种形式的 $v(\mathbf{x}_i)$ 下，对 π_i 进行最大化，可得 logit 选择函数 [8] 所对应的选择分布 $\mathbf{x}_i$。基于此，我们提出以下引理。

引理 5. 对于种群网络博弈中 *SFP* 的选择分布 $\mathbf{x}_i$，

$$\partial_{\mathbf{x}_i}\pi_i\left(\mathbf{x}_i, \{\boldsymbol{\mu}_j\}_{j\in V_i}\right) = \mathbf{0} \quad \textit{and} \quad \sum_{j\in V_i}\left(A_{ij}\boldsymbol{\mu}_j\right)^\top = -\partial_{\mathbf{x}_i} v(\mathbf{x}_i). \tag{77}$$

证明。 该引理可直接由以下事实得出：对 π_i 进行最大化将通过 logit 选择函数 [8] 得到选择分布 $\mathbf{x}_i$。 □

所述　同质群体的 lief 动态在时间重参数化后可以简化　　.

引理 6。 给定 $\tau = \ln \frac{\lambda+t+1}{\lambda+1}$, 同质系统的信念动态（正文中的公式 *11*）等价于

$$\frac{d\boldsymbol{\mu}_i}{d\tau} = \mathbf{x}_i - \boldsymbol{\mu}_i. \tag{78}$$

证明。 由 $\tau = \ln \frac{\lambda+t+1}{\lambda+1}$, 我们有

$$t = (\lambda+1)(\exp(\tau) - 1). \tag{79}$$

根据链式法则，对于每个维度 s_i,

$$\begin{aligned}
\frac{d\mu_{is_i}}{d\tau} &= \frac{d\mu_{is_i}}{dt}\frac{dt}{d\tau} && (80)\\
&= \frac{x_{is_i} - \mu_{is_i}}{\lambda+t+1}\frac{d\left((\lambda+1)(\exp(\tau)-1)\right)}{d\tau} && (81)\\
&= \frac{x_{is_i} - \mu_{is_i}}{\lambda + (\lambda+1)(\exp(\tau)-1)+1}(\lambda+1)\exp(\tau) && (82)\\
&= x_{is_i} - \mu_{is_i}. && (83)
\end{aligned}$$

□

接下来，我们将李雅普诺夫函数 L 定义为

$$L \coloneqq \sum_{i\in V}\omega_i L_i \quad \text{s.t.} \quad L_i \coloneqq \pi_i\left(\mathbf{x}_i, \{\boldsymbol{\mu}_j\}_{j\in V_i}\right) - \pi_i\left(\boldsymbol{\mu}_i, \{\boldsymbol{\mu}_j\}_{j\in V_i}\right). \tag{84}$$

其中， $\{\omega_i\}_{i\in V}$ 是加权零和 Γ中定义的正权重集合。函数 L 是非负的，因为对于每个 $i \in V$，$\mathbf{x}_i$ 使函数 π_i达到最大值。当对于每个 $i \in V$，$\mathbf{x}_i = \boldsymbol{\mu}_i$时，函数 L 达到最小值 0。

将 L改写为

$$L = \sum_{i\in V}\left[\omega_i\pi_i\left(\mathbf{x}_i, \{\boldsymbol{\mu}_j\}_{j\in V_i}\right) - \omega_i\boldsymbol{\mu}_i^\top\sum_{j\in V_i}A_{ij}\boldsymbol{\mu}_j - \omega_i v(\boldsymbol{\mu}_i)\right]. \tag{85}$$

我们注意到，根据 Danskin 定理，$\pi_i\left(\mathbf{x}_i, \{\boldsymbol{\mu}_j\}_{j\in V_i}\right)$ 关于 $\boldsymbol{\mu}_j$, $j \in V_i$ 是凸的，而 $-v(\boldsymbol{\mu}_i)$ 关于 $\boldsymbol{\mu}_i$是严格凸的。此外，根据正文中公式 2 给出的加权零和性质，我们有

$$\sum_{i\in V}\left(\omega_i\boldsymbol{\mu}_i^\top\sum_{j\in V_i}A_{ij}\boldsymbol{\mu}_j\right) = 0 \tag{86}$$

因为对于每个 i, $j \in V$. 都有 $\mu_i \in \Delta_i$, $\mu_j \in \Delta_j$，因此，函数 L 是严格凸函数，并在唯一一点 $\mathbf{x}_i = \boldsymbol{\mu}_i, \forall i \in V$处取得其最小值 0

考虑函数 L_i。其时间导数为

$$\begin{aligned}
\dot{L}_i = {} & \partial_{\mathbf{x}_i}\pi_i\left(\mathbf{x}_i, \{\boldsymbol{\mu}_j\}_{j\in V_i}\right)\dot{\mathbf{x}}_i + \sum_{j\in V_i}\left[\partial_{\boldsymbol{\mu}_j}\pi_i\left(\mathbf{x}_i, \{\boldsymbol{\mu}_j\}_{j\in V_i}\right)\dot{\boldsymbol{\mu}}_j\right] \\
& - \partial_{\boldsymbol{\mu}_i}\pi_i\left(\boldsymbol{\mu}_i, \{\boldsymbol{\mu}_j\}_{j\in V_i}\right)\dot{\boldsymbol{\mu}}_i - \sum_{j\in V_i}\left[\partial_{\boldsymbol{\mu}_j}\pi_i\left(\boldsymbol{\mu}_i, \{\boldsymbol{\mu}_j\}_{j\in V_i}\right)\dot{\boldsymbol{\mu}}_j\right].
\end{aligned} \tag{87}$$

注意，根据引理 5，偏导数 $\partial_{\mathbf{x}_i}\pi_i$ 等于$\mathbf{0}$ 。因此，我们可以将其改写为

$$\begin{aligned}
\dot{L}_i &= \partial_{\boldsymbol{\mu}_i}\pi_i\left(\boldsymbol{\mu}_i, \{\boldsymbol{\mu}_j\}_{j\in V_i}\right)\dot{\boldsymbol{\mu}}_i + \sum_{j\in V_i}\left[\partial_{\boldsymbol{\mu}_j}\pi_i\left(\mathbf{x}_i, \{\boldsymbol{\mu}_j\}_{j\in V_i}\right) - \partial_{\boldsymbol{\mu}_j}\pi_i\left(\boldsymbol{\mu}_i, \{\boldsymbol{\mu}_j\}_{j\in V_i}\right)\right]\dot{\boldsymbol{\mu}}_j && (88)\\
&= -\left[\sum_{j\in V_i}\left(A_{ij}\boldsymbol{\mu}_j\right)^\top + \partial_{\mu_i}v(\boldsymbol{\mu}_i)\right](\mathbf{x}_i - \boldsymbol{\mu}_i) + \sum_{j\in V_i}\left(\mathbf{x}_i^\top A_{ij} - \boldsymbol{\mu}_i^\top A_{ij}\right)(\mathbf{x}_j - \boldsymbol{\mu}_j) && (89)\\
&= \left[\partial_{\mathbf{x}_i}v(\mathbf{x}_i) - \partial_{\mu_i}v(\boldsymbol{\mu}_i)\right](\mathbf{x}_i - \boldsymbol{\mu}_i) + \sum_{j\in V_i}\left(\mathbf{x}_i^\top A_{ij}\mathbf{x}_j - \boldsymbol{\mu}_i^\top A_{ij}\mathbf{x}_j - \mathbf{x}_i^\top A_{ij}\boldsymbol{\mu}_j + \boldsymbol{\mu}_i^\top A_{ij}\boldsymbol{\mu}_j\right). && (90)
\end{aligned}$$

其中，从方程 89 到 90，我们应用引理 5，将 $\sum_{j\in V_i}\left(A_{ij}\boldsymbol{\mu}_j\right)^\top$ 替换为 $-\partial_{\mathbf{x}_i}v(\mathbf{x}_i)$。因此，对所有群体求和后， L 的时间导数为

$$\begin{aligned}\dot{L}=&\sum_{i\in V}\omega_i\left[\partial_{\mathbf{x}_i}v(\mathbf{x}_i)-\partial_{\boldsymbol{\mu}_i}v(\boldsymbol{\mu}_i)\right](\mathbf{x}_i-\boldsymbol{\mu}_i)\\&+\sum_{i\in V}\sum_{j\in V_i}\omega_i\left(\mathbf{x}_i^\top A_{ij}\mathbf{x}_j-\boldsymbol{\mu}_i^\top A_{ij}\mathbf{x}_j-\mathbf{x}_i^\top A_{ij}\boldsymbol{\mu}_j+\boldsymbol{\mu}_i^\top A_{ij}\boldsymbol{\mu}_j\right).\end{aligned}\tag{91}$$

第二行中的求和等价于

$$\sum_{(i,j)\in E}(\omega_i\mathbf{x}_i^\top A_{ij}\mathbf{x}_j+\omega_j\mathbf{x}_j^\top A_{ji}\mathbf{x}_i)-(\omega_i\boldsymbol{\mu}_i^\top A_{ij}\mathbf{x}_j+\omega_j\mathbf{x}_j^\top A_{ji}\boldsymbol{\mu}_i)\tag{92}$$

$$-(\omega_i\mathbf{x}_i^\top A_{ij}\boldsymbol{\mu}_j+\omega_j\boldsymbol{\mu}_j^\top A_{ji}\mathbf{x}_i)+(\omega_i\boldsymbol{\mu}_i^\top A_{ij}\boldsymbol{\mu}_j+\omega_j\boldsymbol{\mu}_j^\top A_{ji}\boldsymbol{\mu}_i).\tag{93}$$

根据正文中公式 2 给出的加权零和性质，该求和等于 0，因此得到

$$\dot{L}=\sum_{i\in V}\omega_i\left[\partial_{\mathbf{x}_i}v(\mathbf{x}_i)-\partial_{\boldsymbol{\mu}_i}v(\boldsymbol{\mu}_i)\right](\mathbf{x}_i-\boldsymbol{\mu}_i).\tag{94}$$

注意，函数 v 是严格凹的，因此其二阶导数负定。根据这一性质， $L\leqslant 0$，且仅当$\mathbf{x}_i=\boldsymbol{\mu}_i$， $\forall i\in V$时取等号，这对应于量化反应均衡。因此， L 是一个严格李雅普诺夫函数，量化反应均衡的全局渐近稳定性由此得证。证毕。

备注：直观而言，上述定义的李雅普诺夫函数度量了量化反应均衡与给定信念集之间的距离。以熵正则化收益来度量距离的想法，源自文献 *[19]* 的开创性工作。然而，与本文研究的网络博弈不同，*Hofbauer* 和 *Hopkins* *[19]* 研究的是双人博弈中的 *SFP*。据我们所知，迄今尚无关于网络博弈中 *SFP* 的系统性研究。

定理 4 的证明

定理 4 的证明借鉴了渐近自治动力系统的重要成果[31, 40, 41]，这类系统通常定义如下。

定义*1. </style}* R^n中的非自治微分方程组

$$x'=f(t,x)\tag{95}$$

若其极限方程为

$$y'=g(y),\tag{96}$$

若 $f(t、\ x)\to g(x)$、$t\to\infty$成立，其中收敛在 R^n的每个紧子集上都是一致的。通常，方程*95*的解流称为渐近自治半流（记为 ϕ），方程*96*的解流称为极限半流（记为 Θ）。

基于此定义，我们在正文中证明了引理 1，现重述如下。

引理 7. 对于初始时具有异质信念的系统，平均信念动态渐近自治 *[31]*，其极限方程为

$$\frac{d\boldsymbol{\mu}_i}{dt}=\mathbf{x}_i-\boldsymbol{\mu}_i\tag{97}$$

经过时间重参数化后，该方程等价于同质系统的信念动态。

证明。我们 首次对异质系统的平均信念动态进行时间重参数化。假设 τ $=$

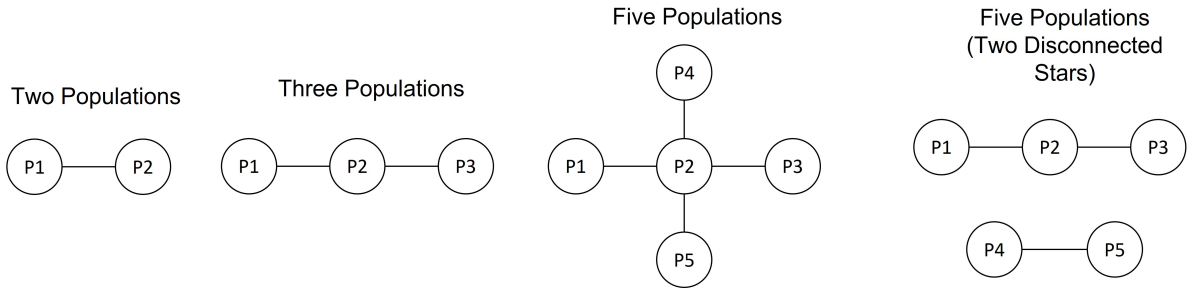


图5：底层网络由星形结构构成的群体网络博弈。

$\ln \frac{\lambda+t+1}{\lambda+1}$。根据链式法则和方程54，对于每个维度 s_i,

$$\frac{d\bar{\mu}_{is_i}}{d\tau} = \frac{d\bar{\mu}_{is_i}}{dt}\frac{dt}{d\tau} \tag{98}$$

$$= \left[\frac{f_{s_i}(\bar{\boldsymbol{\mu}}) - \bar{\mu}_{is_i}}{\lambda+t+1} + \frac{\sum_{j\in V_i}\sum_{s_j\in S_j}\frac{\partial^2 f_{s_i}(\bar{\boldsymbol{\mu}})}{(\partial\mu_{js_j})^2}\mathrm{Var}(\mu_{js_j})}{2(\lambda+t+1)}\right]\frac{d\left((\lambda+1)(\exp(\tau)-1)\right)}{d\tau} \tag{99}$$

$$= \frac{f_{s_i}(\bar{\boldsymbol{\mu}}) - \bar{\mu}_{is_i} + \frac{1}{2}\sum_{j\in V_i}\sum_{s_j\in S_j}\frac{\partial^2 f_{s_i}(\bar{\boldsymbol{\mu}})}{(\partial\mu_{js_j})^2}\left(\frac{\lambda+1}{\lambda+t+1}\right)^2\sigma^2(\mu_{js_j})}{\lambda+(\lambda+1)(\exp(\tau)-1)+1}(\lambda+1)\exp(\tau) \tag{100}$$

$$= f_{s_i}(\bar{\boldsymbol{\mu}}) - \bar{\mu}_{is_i} + \frac{1}{2}\sum_{j\in V_i}\sum_{s_j\in S_j}\frac{\partial^2 f_{s_i}(\bar{\boldsymbol{\mu}})}{(\partial\mu_{js_j})^2}\sigma^2(\mu_{js_j})\exp(-2\tau). \tag{101}$$

注意到 $\exp(-2\tau)$ 指数速度衰减至零，且对于单纯形 $\prod_{j\in V_i}\Delta_j$ 中的任意 $\boldsymbol{\mu}$，$\sigma^2(\mu_{js_j})$ 和 $\frac{\partial^2 f_{s_i}(\bar{\boldsymbol{\mu}})}{(\partial\mu_{js_j})^2}$ 均有界。因此，方程101局部一致收敛到以下方程：

$$\frac{d\bar{\mu}_{is_i}}{d\tau} = f_{s_i}(\bar{\boldsymbol{\mu}}) - \bar{\mu}_{is_i}. \tag{102}$$

注意，对于同质系统，$x_{is_i} = f_{s_i}(\bar{\boldsymbol{\mu}})$，且上述方程在代数上等价于方程97。因此，根据定义1，方程101是渐近自治的，其极限方程为方程97。

□

根据上述引理，我们可以将初始异质系统的极限行为与同质系统的极限行为正式联系起来。回顾正文中的定理3：在加权零和网络博弈 Γ 中，在 SFP 下，信念动态存在唯一的静止点（量化反应均衡）；这排除了存在有限多个彼此相连的均衡的情形。因此，结合正文中的引理 2，我们证明了初始异质系统的平均信念动态收敛到唯一的量化反应均衡。证毕。

附录C：第4.2节省略的结果与证明

对于网络协调情形，出于技术原因，我们考虑由一个星形网络或多个不连通的星形结构组成的网络。图1展示了所考虑的具有不同数量节点（群体）的网络结构示例。

在以下定理中，我们聚焦于同质系统，并证明信念动态收敛至量化反应均衡集合。

定理6 (**星形结构网络协调中的同质系统收敛性**)。给定一个协调 Γ，其网络结构由单个星形网络或多个不连通的星形结构组成，则同质系统的每条信念动态轨道都收敛至量化反应均衡集合。

证明：考虑星形结构中的根种群 j。其叶种群（相邻群体）集合为 V_j，关于叶群体的信念集合为 $\{\boldsymbol{\mu}_i\}_{i\in V_j}$，选择分布为 $\mathbf{x}_j$。给定博弈 Γ,

期望收益为 $\mathbf{x}_j^\top \sum_{i\in V_j} A_{ji}\boldsymbol{\mu}_i$。定义一个扰动收益函数

$$\pi_j\left(\mathbf{x}_j, \{\boldsymbol{\mu}_i\}_{i\in V_j}\right) \coloneqq \mathbf{x}_j^\top \sum_{i\in V_j} A_{ji}\boldsymbol{\mu}_i + v(\mathbf{x}_j) \tag{103}$$

其中 $v(\mathbf{x}_j) = -\frac{1}{\beta}\sum_{s_j\in S_j} x_{js_j}\ln(x_{js_j})$。在这种形式的 $v(\mathbf{x}_j)$下，π_j的最大化会通过 logit 选择函数 [8]得到选择分布 $\mathbf{x}^j$。

考虑根种群 j的一个叶种群 i。它只有一个邻居人群，即种群 j。因此，给定博弈 Γ，期望收益为 $\mathbf{x}_i^\top A_{ij}\boldsymbol{\mu}_j$。定义一个扰动收益函数

$$\pi_i\left(\mathbf{x}_i, \boldsymbol{\mu}_j\right) \coloneqq \mathbf{x}_i^\top A_{ij}\boldsymbol{\mu}_j + v(\mathbf{x}_i) \tag{104}$$

其中 $v(\mathbf{x}_i) = -\frac{1}{\beta}\sum_{s_i\in S_i} x_{is_i}\ln(x_{is_i})$。类似地，$\pi_i$ 的最大化会通过 logit 选择函数 [8]得到选择分布 $\mathbf{x}_{i\circ}$ 基于此，我们建立如下引理。

L引理 8. 具有起始结构的种群网络博弈中 *SFP* 的选择分布 e,

$$\partial_{\mathbf{x}_j}\pi_j\left(\mathbf{x}_j, \{\boldsymbol{\mu}_i\}_{i\in V_j}\right) = \mathbf{0} \quad \textit{and} \quad \sum_{i\in V_j}\left(A_{ji}\boldsymbol{\mu}_i\right)^\top = -\partial_{\mathbf{x}_j} v(\mathbf{x}_j) \qquad \textit{if } j \textit{ is a root population,} \tag{105}$$

$$\partial_{\mathbf{x}_i}\pi_i\left(\mathbf{x}_i, \boldsymbol{\mu}_j\right) = \mathbf{0} \quad \textit{and} \quad \left(A_{ij}\boldsymbol{\mu}_j\right)^\top = -\partial_{\mathbf{x}_i} v(\mathbf{x}_i) \qquad \textit{if } i \textit{ is a leaf population.} \tag{106}$$

证明。该引理直接源于以下事实：分别对 π_j 和 π_i 进行最大化，会通过 logit 选择函数 [8]得到选择分布 $\mathbf{x}^j$ 和 $\mathbf{x}_i$。□

为便于阅读，我们将附录 B 中引理 4 已证明的同质种群经时间重参数化后的信念动态重述如下：

$$\frac{d\boldsymbol{\mu}_i}{d\tau} = \mathbf{x}_i - \boldsymbol{\mu}_i. \tag{107}$$

令 $\mathcal{R}\subset V$为所有根种群的集合。定义

$$L \coloneqq \sum_{j\in\mathcal{R}} L_j \quad \text{s.t.} \quad L_j \coloneqq \boldsymbol{\mu}_j^\top \sum_{i\in V_j} A_{ji}\boldsymbol{\mu}_i + v(\boldsymbol{\mu}_j) + \sum_{i\in V_j} v(\boldsymbol{\mu}_i). \tag{108}$$

考虑函数 L_j。其时间导数 $\dot{L}_j$ 为

$$\dot{L}_j = \left[\partial_{\boldsymbol{\mu}_j}(\boldsymbol{\mu}_j^\top \sum_{i\in V_j} A_{ji}\boldsymbol{\mu}_i)\dot{\boldsymbol{\mu}}_j + \sum_{i\in V_j}\partial_{\boldsymbol{\mu}_i}(\boldsymbol{\mu}_j^\top \sum_{i\in V_j} A_{ji}\boldsymbol{\mu}_i)\dot{\boldsymbol{\mu}}_i\right] + \partial_{\boldsymbol{\mu}_j} v(\boldsymbol{\mu}_j)\dot{\boldsymbol{\mu}}_j + \sum_{i\in V_j}\partial_{\boldsymbol{\mu}_i} v(\boldsymbol{\mu}_i)\dot{\boldsymbol{\mu}}_i \tag{109}$$

$$= \sum_{i\in V_j}(A_{ji}\boldsymbol{\mu}_i)^\top(\mathbf{x}_j - \boldsymbol{\mu}_j) + \left[\sum_{i\in V_j}\boldsymbol{\mu}_j^\top A_{ji}(\mathbf{x}_i - \boldsymbol{\mu}_i)\right] + \partial_{\boldsymbol{\mu}_j} v(\boldsymbol{\mu}_j)(\mathbf{x}_j - \boldsymbol{\mu}_j) + \sum_{i\in V_j}\partial_{\boldsymbol{\mu}_i} v(\boldsymbol{\mu}_i)(\mathbf{x}_i - \boldsymbol{\mu}_i). \tag{110}$$

由于 Γ 是协调博弈，我们有 $\left(A_{ij}\boldsymbol{\mu}_j\right)^\top = \boldsymbol{\mu}_j^\top A_{ij}^\top = \boldsymbol{\mu}_j^\top A_{ji}$。因此，应用引理 8，可将 $\sum_{i\in V_j}(A_{ji}\boldsymbol{\mu}_i)^\top$ 替换为 $-v'(\mathbf{x}_j)$，并将 $\boldsymbol{\mu}_j^\top A_{ji}$ 替换为 $-v'(\mathbf{x}_i)$，得到

$$\dot{L}_j = -\partial_{\mathbf{x}_j} v(\mathbf{x}_j)(\mathbf{x}_j - \boldsymbol{\mu}_j) + \left[\sum_{i\in V_j}(-\partial_{\mathbf{x}_i} v(\mathbf{x}_i))(\mathbf{x}_i - \boldsymbol{\mu}_i)\right] + \partial_{\boldsymbol{\mu}_j} v(\boldsymbol{\mu}_j)(\mathbf{x}_j - \boldsymbol{\mu}_j) + \sum_{i\in V_j}\partial_{\boldsymbol{\mu}_i} v(\boldsymbol{\mu}_i)(\mathbf{x}_i - \boldsymbol{\mu}_i) \tag{111}$$

$$= (\partial_{\boldsymbol{\mu}_j} v(\boldsymbol{\mu}_j) - \partial_{\mathbf{x}_j} v(\mathbf{x}_j))(\mathbf{x}_j - \boldsymbol{\mu}_j) + \sum_{i\in V_j}(\partial_{\boldsymbol{\mu}_i} v(\boldsymbol{\mu}_i) - \partial_{\mathbf{x}_i} v(\mathbf{x}_i))(\mathbf{x}_i - \boldsymbol{\mu}_i) \tag{112}$$

注意，函数 v 是严格凹函数，因此其二阶导数负定。由此性质可知，$\dot{L}_j \geqslant 0$，且仅当 $\mathbf{x}_i = \boldsymbol{\mu}_i$, $\forall i \in V_j$ 且 $\mathbf{x}^j = \boldsymbol{\mu}_j$时取等号。因此，$\dot{L}$函数 L的时间导数，即 $L = \sum_{j\in\mathcal{R}} L_j \geqslant 0$，且仅当 $\mathbf{x}_i = \boldsymbol{\mu}_i$, $\forall i \in V_j$, $\mathbf{x}^j = \boldsymbol{\mu}_j$, $\forall j \in \mathcal{R}$时取等号。□

We 在下述定理中，将收敛结果推广到初始异质系统 .

定理 7 (**具有星形网络结构的初始异质网络协调中的收敛**)。 给定一个协调 Γ，其网络结构由单个星形网络或多个不连通的星形结构组成，则初始异质系统的平均信念动态的每个轨道都收敛到量化反应均衡集合。

证明。证明方法与初始异质竞争网络博弈的证明方法类似。根据正文中的引理 1，我们证明初始异质系统的平均信念动态渐近自治于同质系统的信念动态。因此，根据正文中的引理 2，同质系统的收敛结果也适用于初始异质系统。

□

备注：双人情形 *[19]* 以及某些多人情形 *[20, 39]*下，协调博弈和势博弈中 ***SFP*** 的收敛性已经得到证明。我们的工作与以往研究有两点不同。首先，我们允许存在异质信念。此外，我们考虑智能体各自持有关于其他智能体的独立信念，而以往研究中的智能体不会区分其他智能体。因此，即使系统信念是同质的，我们的设定仍不同于（且复杂于）以往的设定。

附录 D：省略的实验细节

偏微分方程模型的数值方法。偏微分方程通常很难求解，只有有限类型的偏微分方程存在解析解。因此，与以往研究 [23], 类似，我们采用数值方法求解偏微分方程；具体而言，我们采用有限差分法 [38]。

基于智能体的仿真。 所呈现的仿真结果是对100次独立仿真运行取平均，以平滑随机性。每次仿真运行中，每个群体都有1,000个智能体。对于每个智能体，其初始信念均从给定的初始概率分布中抽样。

图1的详细实验设置。 在初始方差较小的情况下，初始信念 μ_{1H} 和 μ_{2H} 服从 Beta$(280, 120)$ 分布。相反，在初始方差较大的情况下，初始信念 μ_{1H} 和 μ_{2H} 服从 Beta$(14, 6)$ 分布。因此，这两种情况下的初始平均信念均为 $\bar{\mu}_{1H} = \bar{\mu}_{2H} = 0.7$和 $\bar{\mu}_{1S} = \bar{\mu}_{2S} = 0.3$。在这两种情况下，初始权重和 $\lambda = 10$ 以及温度 $\beta = 10$。

图 3 的详细实验设置。我们通过数值求解平均信念动态（正文中的方程 10），来可视化鹿猎博弈中不同均衡的吸引域。初始方差已在每个面板的标题中给出。在所有情况下，初始权重和 $\lambda = 0$ 以及温度 $\beta = 5$。

图 4 的详细实验设置。我们让关于群体 1、3 和 5 的初始信念在不同情形下保持不变，并改变关于群体 2 和 4 的初始信念。关于群体 1、3 和 5 的初始信念分别记为 μ_{1H}、 μ_{3H} 和 μ_{5H}，其分布分别为 Beta$(20,\ 10)$、Beta$(6,\ 4)$ 和 Beta$(10,\ 5)$。关于群体 2 和 4 的初始信念已在图 4 的图例中给出。在所有情况下，初始权重和 $\lambda = 10$ 以及温度 $\beta = 10$。请注意，对于所有群体 $i = 1$，2，3，4，5。

源代码和计算资源。我们附上了用于复现正文中主要实验的源代码。Matlab 脚本 ***finitedifference.m*** 通过数值方法求解正文命题1中给出的偏微分方程模型。Matlab 脚本 $regionofattraction.m$ 可视化图 3 所示鹿猎博弈中不同均衡的吸引域。Python 脚本$simulation(staghunt).py$ 和 $simulation(matchingpennies).py$ 分别对应双群体鹿猎博弈和五群体非对称匹配硬币博弈的基于智能体的仿真。所有实验均在笔记本电脑（CPU：AMD Ryzen 7 5800H）上运行。

参考文献

[1] Yakov Babichenko 和 Aviad Rubinstein。确定拥塞博弈中纳什均衡的复杂性。载于第 53 届 ACM SIGACT计算理论研讨会论文集，第 1426–1437 页，2021 年。[2] Michel Benaïm 和 Mathieu Faure。渐近光滑虚拟博弈的一致性。《运筹学数学》，38(3)：437–450，2013 年。[3] Michel Benaım 和 Morris W Hirsch。扰动博弈中的混合均衡与虚拟博弈所产生的动力系统。《博弈与经济行为》，29(1-2)：36–72，1999 年。[4] Shant Boodaghians、Rucha Kulkarni 和 Ruta Mehta。网络协调博弈的平滑高效算法与归约。arXiv 预印本 arXiv:1809.02280，2018 年。[5] Yang Cai 和 Constantinos Daskalakis。多人博弈的极小极大定理。载于第二十二届 ACM-SIAM离散算法研讨会论文集，第 217–234 页。SIAM，2011 年。[6] Aleksander Czechowski 和 Georgios Piliouras。多智能体学习中的 Poincar\'{e}-Bendixson 极限集。2022 年。[7] Christian Ewerhart 和 Kremena Valkanova。网络中的虚拟博弈。《博弈与经济行为》，123：182–206，2020 年。[8] Drew Fudenberg、Fudenberg Drew、David K Levine 和 David K Levine。《博弈学习理论》，第 2 卷。麻省理工学院出版社，1998 年。[9] Drew Fudenberg 和 David M Kreps。混合均衡学习。《博弈与经济行为》，5(3)：320–367，1993 年。[10] Drew Fudenberg 和 David K Levine。稳态学习与纳什均衡。《计量经济学》，《计量经济学会杂志》，第 547–573 页，1993 年。[11] Drew Fudenberg 和 David K Levine。衡量实验博弈中的玩家损失。《经济学季刊》，112(2)：507–536，1997 年。[12] Drew Fudenberg 和 Satoru Takahashi。随机虚拟博弈中的异质信念与局部信息。《博弈与经济行为》，71(1)：100–120，2011 年。[13] Colin S Gillespie。质量作用模型的矩闭合近似。《IET系统生物学》，3(1)：52–58，2009 年。[14] Jacob K Goeree、Charles A Holt 和 Thomas R Palfrey。量子响应均衡。载于《量子响应均衡》。普林斯顿大学出版社，2016 年。[15] Leo A Goodman。性别的种群增长。《生物统计学》，9(2)：212–225，1953 年。[16] Shubham Gupta、Rishi Hazra 和 Ambedkar Dukkipati。具有涌现通信的网络化多智能体强化学习。载于第 19 届自主智能体与多智能体系统国际会议论文集，第 1858–1860 页，2020 年。[17] Jiequn Han 和 Ruimeng Hu。用于求解多智能体博弈中马尔可夫纳什均衡的深度虚拟博弈。载于《数学与科学机器学习》，第 221–245 页。PMLR，2020 年。[18] Erik Hernández、Antonio Barrientos 和 Jaime Del Cerro。选择性平滑虚拟博弈：一种基于博弈论的多机器人系统基础设施巡逻方法。《应用专家系统》，41(6)：2897–2913，2014 年。[19] Josef Hofbauer 和 Ed Hopkins。扰动非对称博弈中的学习。《博弈与经济行为》，52(1)：133–152，2005 年。[20] Josef Hofbauer 和 William H Sandholm。随机虚拟博弈的全局收敛性。《计量经济学》，70(6)：2265–2294，2002 年。

[21] Josef Hofbauer 和 William H Sandholm。《随机扰动收益的博弈演化》。《经济理论期刊》，132(1):47–69，2007年。[22] Ed Hopkins。关于最佳响应动态的一则说明。《博弈与经济行为》，29(1-2):138–150，1999年。[23] Shuyue Hu、Chin-wing Leung 和 Ho-fung Leung。《重复对称博弈中多智能体Q学习的动态：一种平均场理论方法》。载于《神经信息处理系统进展》，第12102–12112页，2019年。[24] Shuyue Hu、Chin-Wing Leung、Ho-fung Leung 和 Harold Soh。《群体博弈中Q学习的动态：一种受物理学启发的连续性方程模型》。载于第*21*届自主智能体与多智能体系统国际会议论文集，第615–623页，2022年。[25] Max Jaderberg、Valentin Dalibard、Simon Osindero、Wojciech M Czarnecki、Jeff Donahue、Ali Razavi、Oriol Vinyals、Tim Green、Iain Dunning、Karen Simonyan 等。《神经网络的基于群体的训练》。*arXiv*预印本 *arXiv:1711.09846*，2017年。[26] KG Binmore、AP Kirman 等。《博弈论前沿》。麻省理工学院出版社，1993年。[27] Ratul Lahkar 和 Robert M Seymour。《群体博弈中的强化学习》。《博弈与经济行为》，80:10–38，2013年。[28] Stefanos Leonardos 和 Georgios Piliouras。《多智能体学习中的探索-利用：灾变理论与博弈论的交汇》。载于*AAAI*人工智能会议论文集，第35卷，第11263–11271页，2021年。[29] Chin-Wing Leung、Shuyue Hu 和 Ho-Fung Leung。《自我对弈还是群体练习：在多智能体系统中学习交替马尔可夫博弈》。载于*2020*年第*25*届国际模式识别会议（*ICPR*），第9234–9241页。IEEE，2021年。[30] Yiwei Liu、Jiamou Liu、Kaibin Wan、Zhan Qin、Zijian Zhang、Bakhadyr Khoussainov 和 Liehuang Zhu。《从局部到全局的规范涌现：借助增量式社会工具消解自我强化子结构》。载于国际机器学习会议，第6871–6881页。PMLR，2021年。[31] L Markus。《渐近自治微分系统》。载于《非线性振荡理论的贡献 III》（S. Lefschetz 编），第17–29页。数学研究年刊，第36卷，1956年。[32] Timothy I Matis 和 Ivan G Guardiola。《通过忽略累积量实现矩闭合》。《*Mathematica*期刊》，12:12–2，2010年。[33] Richard D McKelvey 和 Thomas R Palfrey。《标准形博弈的量化反应均衡》。《博弈与经济行为》，10(1):6–38，1995年。[34] Sai Ganesh Nagarajan、David Balduzzi 和 Georgios Piliouras。《从混沌到秩序：博弈动态中的对称性与守恒定律》。载于国际机器学习会议，第7186–7196页。PMLR，2020年。[35] Gerasimos Palaiopanos、Ioannis Panageas 和 Georgios Piliouras。《拥塞博弈中采用固定步长的乘法权重更新：收敛、极限环与混沌》。《神经信息处理系统进展》，第30卷，2017年。[36] Julien Perolat、Bilal Piot 和 Olivier Pietquin。《同时行动多阶段博弈中的演员-评论家虚拟博弈》。载于国际人工智能与统计会议，第919–928页。PMLR，2018年。[37] Sarah Perrin、Julien Pérolat、Mathieu Laurière、Matthieu Geist、Romuald Elie 和 Olivier Pietquin。《平均场博弈的虚拟博弈：连续时间分析与应用》。《神经信息处理系统进展》，33:13199–13213，2020年。[38] Gordon Dennis Smith。《偏微分方程的数值解：有限差分法》。克拉伦登出版社，1985年。

[39] Brian Swenson 和 H Vincent Poor。《n× 2 势博弈中的平滑虚拟博弈》。载于*2019*年第*53*届阿西洛马会议：信号、系统与计算机，第1739–1743页。IEEE，2019年。[40] Horst R Thieme。《渐近自治微分方程的收敛结果与庞加莱-本迪克森三分法》。《数学生物学杂志》，30(7):755–763，1992年。[41] Horst R Thieme。《平面上的渐近自治微分方程》。《落基山数学杂志》，第351–380页，1994年。[42] Qiaomin Xie、Zhuoran Yang、Zhaoran Wang 和 Andreea Minca。《一般平均场博弈中的可证明虚拟博弈》。*arXiv*预印本 *arXiv:2010.04211*，2020年。[43] Kaiqing Zhang、Zhuoran Yang、Han Liu、Tong Zhang 和 Tamer Basar。《基于网络化智能体的完全去中心化多智能体强化学习》。载于国际机器学习会议，第5872–5881页。PMLR，2018年。[44] Rui Zhao、Jinming Song、Hu Haifeng、Yang Gao、Yi Wu、Zhongqian Sun 和 Yang Wei。《面向零样本人机协同的最大熵群体训练》。*arXiv*预印本 *arXiv:2112.11701*，2021年。