arXiv:2304.09978v1 [cs.GT] 19 Apr 2023

2×2标准式博弈的均衡不变嵌入、度量空间与基本集合

LukeMarris[1,2]、**Ian Gemp**[1] **和 Georgios Piliouras**[1]1DeepMind, [2]伦敦大学学院

标准式博弈的均衡解概念，例如纳什均衡、相关均衡和粗相关均衡，描述了这样一种联合策略组合：其中没有任何玩家有动机单方面偏离。这些概念在博弈论、经济学和多智能体系统中得到了广泛研究。均衡概念在收益的某些变换下保持不变。我们为所有标准式博弈构成的空间定义了一种受均衡启发的距离度量，并揭示了一种保距均衡不变嵌入。此外，我们提出了一种额外的变换，由此定义出一种更优响应不变的距离度量和嵌入。为展示这些度量空间，我们研究了 2×2 博弈。 2×2 博弈的均衡不变嵌入具有高效的双变量参数化（从八个变量缩减而来），其中每个变量都通过几何方式描述单位圆上的一个角度。从嵌入中可以在空间上推断出若干有趣的性质，包括：均衡支撑集、循环、竞争、协调、距离、最佳响应和对称性。考虑对称性后，2×2 博弈的最佳响应不变嵌入重新发现了一组包含 15 个博弈的集合及其各自的等价类。我们提出，这组博弈类别具有基础性，涵盖了 2×2 博弈中所有可能的有趣策略互动。我们为每个类别引入一种有向图表示并命名。最后，我们利用为2×2 博弈开发的工具，为大型标准式博弈和扩展式博弈开发博弈论可视化方法，旨在为其中发生的策略互动建立特征画像。

关键词： *2×2*博弈、不变性、嵌入、博弈度量空间、博弈拓扑、博弈分类、博弈周期表、博弈表示、博弈嵌入、纳什均衡、相关均衡、粗相关均衡、博弈论、一般和、混合动机

1. 引言

标准式博弈的均衡解，例如纳什均衡（NE）（Nash，1951年）、相关均衡（CE）（Aumann，1974年）以及粗相关均衡（CCE）（Hannan，1957年；Moulin andVial，1978年；Young，2004年），被广泛用于建模博弈中理性效用最大化玩家的策略行为。在某些类型的博弈中，例如双人零和博弈，纳什均衡这一解概念被视为基础性的（von Neumann and Morgenstern，1947年），因为在这类博弈中，它既无法被利用，也可互换。相关均衡和粗相关均衡在 n 人一般和博弈中十分重要，这类博弈可能需要借助相关装置来协调。博弈中的均衡集合对收益的某些变换保持不变（Morrisand Ui，2004年）。其中最广为人知的是仿射变换（对每位玩家的收益进行偏移并乘以正比例因子）。我们将此类变换称为均衡不变。此外，收益还存在某些对称性，这些对称性会使均衡集合呈现相应的等价对称性（例如策略或玩家的顺序）。我们将作用于对称性的变换称为均衡对称。最后，还有一种较弱的较优响应不变性概念：变换不会改变玩家对于针对其他玩家联合策略的响应的偏好。我们将此类变换称为更优响应不变。本文研究用于构建 n 人一般和标准式博弈度量空间及其嵌入的变换。为说明这些度量空间的重要性，我们随后聚焦于 2×2 博弈：探索其性质、可视化其结构，并重新发现一组由 15 个基础博弈组成的集合。

通讯作者：*marris@deepmind.com*

2×2 标准式博弈有两名玩家，每名玩家各有两种策略。玩家同时选择其中一种策略（也可以随机选择）。由此产生的联合策略决定每位玩家的收益。博弈只进行一次。收益表决定玩家的理性行为（即均衡集合）。常见博弈会根据其收益及由此产生的行为而得名。 2×2 博弈的应用十分广泛，其名称已进入大众文化（图 1），例如：胆小鬼博弈、囚徒困境、鹿猎博弈（Skyrms，2004年）、Bachor Stravinsky（性别之战）和匹配硬币[1]。对这类博弈的研究（Kelley 等，2002年）对于理解合作（Gauthier，1986年）、竞争、协调、《自然》（ Wilkinson，1984年）、激励结构（Sugden，2005年）、社会困境（Bruns 和 Kimmich，2021年）、功利主义行为、理性行为（Gintis，2014年）以及看似非理性的行为至关重要。博弈被用于为经济政策（Ostrom 等，1994年）、社会结构（Bicchieri，2005年；Binmore 和 Binmore，1994年；Skyrms，2004年）、外交政策（Schelling，1966年）、疫情应对和环境条约（Brânzei 等，2021年；Breton 等，2006年；Schosser，2022年）提供参考。

因此，人们投入了大量精力，为 2×2 博弈建立参数化、分类体系、拓扑和命名（Brams，1993年；Bruns，2015年；Kilgour 和 Fraser，1988年；Rapoport 和 Guyer，1966年；Rapoport 等，1976年；Robinson 和 Goforth，2005年；Robinson 等，2007年；Walliser，1988年）。其中最常见的方法侧重于序数博弈，这是一种等价关系，它承认博弈中产生的联合策略只有在各玩家收益元素的偏序意义上才存在有意义的差异。此类博弈共有726种（Fraser 和 Kilgour，1986年）。其他大多数研究（Böörs 等，2022年；Goforth 和 Robinson，2005年；Harris，1969年；Huertas-Rosero，2003年；Rapoport 和 Guyer，1966年）只考虑 2×2 博弈的子集（例如，对称博弈或严格序数博弈）。Borm（1987年）依据纳什均衡最佳回应，将所有 2×2 博弈划分为15个不同类别。Fishburn 和 Kilgour（1990年）后来证明，这些类别可以用二元博弈表示。这两项研究都没有为这些博弈提供接近程度的概念、距离度量或令人满意的参数化。

这项工作揭示了一般和 n 人标准式博弈的度量空间与嵌入。我们在 2×2 博弈（图2）中展示了这些嵌入。所有非平凡（定义3） 2×2 博弈均可转换为均衡不变嵌入，并且只需两个变量即可对其参数化；相比之下，表示原始收益需要八个变量（表1）。值得注意的是，这不会改变博弈的均衡集合。从几何上看，这两个变量各自描述圆上的一个角度，并具有空间意义：相似的博弈彼此相邻。博弈的性质，如均衡支撑集、零和不变性、共同收益不变性、是否为顺时针或逆时针循环、是否为协调或反协调，以及最佳响应动态，都可以从这一嵌入中轻易推断出来。利用对称性，可将均衡不变嵌入的区域缩小八倍，从而得到均衡对称嵌入。此外，这个空间还可进一步缩减为一个包含 15 个博弈的集合：最佳响应不变嵌入。由于许多均衡不变嵌入会映射到同一个最佳响应不变嵌入，因此它们也构成等价类。这些等价类与Borm (1987年）发现的相同，但推导过程采用了不同却相关的论证。我们通过将 Borm 的分类置于均衡不变嵌入中，揭示博弈之间的关系，从而改进了这一分类。我们还为这些等价类命名，并提供了优雅的图形化可视化。

最后，我们利用 2×2 中开发的工具，研究拥有更多玩家和策略的博弈，包括大型正则形式博弈和扩展式博弈。从历史上看，人们一直认为这类博弈难以可视化或概括。我们开发了博弈论可视化方法，用以刻画这些博弈中的策略互动。机器学习领域一直受益于对高维复杂数据进行简化可视化的方法。PCA（ Hotelling，1936；Pearson，1901年）等有理论依据的方法旨在降低维度，同时保留尽可能多的信息。其他理论依据较弱的方法，如 t-SNE（Hinton 和 Roweis，2003年；van der Maaten 和 Hinton，2008年），也非常流行。我们希望这些可视化方法能为博弈论从业者提供帮助；我们认为，这类可视化工具目前仍有待发展。

2. 预备知识与相关工作

博弈论 博弈论研究理性收益最大化主体的互动行为 s
在存在.....的情况下 其他智能体。智能体所处的环境称为博弈。我们将重点讨论

[1]在这项工作中，我们在博弈名称旁附上图形表示，用于描述序数博弈（如匹配硬币）中各玩家对联合收益的偏好，或最佳响应不变嵌入（如周期）中的最佳响应偏好。我们将在后续章节更详细地介绍这些表示。

	C	S
C	−9, −9	+1, −1
S	−1, +1	0, 0

(a) 胆小鬼博弈

	C	D
C	−2, −2	−9, +0
D	+0, −9	−5, −5

(b) 囚徒困境

	S	H
S	4, 4	1, 3
H	3, 1	2, 2

(c) 鹿猎博弈

	M	F
M	3, 2	0, 0
F	0, 0	2, 3

(d) 巴赫或斯特拉文斯基

	H	T
H	+1, −1	−1, +1
T	−1, +1	+1, −1

(e) 匹配硬币

图 1 | 常见 2×2 标准式博弈的收益表。玩家1选择一个行策略，玩家2选择一个列策略。每位玩家分别获得元组中的一个收益。每位玩家的联合收益偏好排序显示在各自表格的左上角。

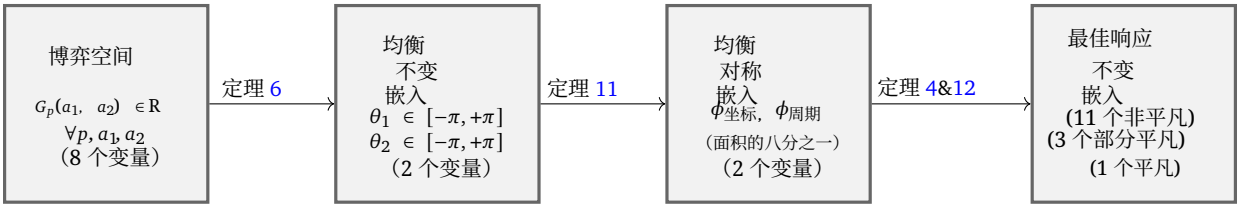


图 2 | 本工作的主要贡献概述，展示了如何将所有 2×2博弈转化为均衡不变嵌入、均衡对称嵌入和最佳响应不变嵌入。

一种特定类型的单轮同时行动博弈，称为标准式博弈。标准式博弈由 N 个玩家、每位玩家可选策略的集合 p（以 $a_p \in \mathrm{A}_p = \{a_p^A, a_p^B, ...\}$ 为参数），以及每位玩家在特定联合行动下的收益 $G_p(a)$ 构成，其中 $a = (a_1, ..., a_N) = (a_p, a_{-p}) \in \mathrm{A} = \otimes_p \mathrm{A}_p$。下标符号 $-p$ 表示“除玩家 p 以外的所有玩家”。复数形式“收益”用于描述所有玩家的收益，即 $G(a) = (G_1(a), ..., G_N(a))$。当收益不以联合策略 G_p 为参数时，它指的是形状为 $[|\mathrm{A}_1|, ..., |\mathrm{A}_N|]$ 的完整张量。收益的元素也可以直接用变量表示，$G_p(a_p^A, a_p^B) = g_p^{AB}$。玩家不必采取确定性行动（纯策略），也可以在其策略上随机化（混合策略），并且其行为可以与其他玩家相关联（联合混合策略）。玩家的行为由联合分布 $\sigma(a)$ 描述。每位玩家的目标是最大化其期望收益，$\sum_{a \in \mathrm{A}} \sigma(a) G_p(a)$。如果联合分布可以分解，即 $\sigma(a) = \otimes_p \sigma_p(a_p)$，玩家便可以独立行动，其中 $\sigma_p(a_p)$ 是边缘混合策略。然而，这种做法有所局限，因为它不允许玩家协调行动。可以采用一种称为相关装置的调解机制，让玩家执行任意的联合策略 $\sigma(a)$，而这些策略不一定能分解为各自的边缘分布。该调解机制会从公开已知的联合分布 $\sigma(a)$ 中抽样，并秘密地向每位玩家传达建议策略。博弈论在一类博弈中发展得最为成熟：这类博弈有两位玩家，且收益受到限制，即 $G_1(a_1, a_2) = -G_2(a_1, a_2)$，称为零和博弈。尤其是在 n 人一般和博弈中，并不存在公认的解概念。一种方法是考虑处于均衡状态的联合分布：任何玩家都没有动机单方面偏离的分布。博弈有时也以其形状来指称，例如：$|\mathrm{A}_1| \times ... \times |\mathrm{A}_N|$。本文主要关注一般和标准式博弈。

均衡解概念 广义而言，博弈的解是对具有自主性的理性自利玩家在博弈中可能如何行动的预测。解概念有很多种，不过最受欢迎的一类是均衡解，它描述了玩家策略的稳定分布：任何玩家都没有动机单方面偏离这些分布。最著名的是纳什均衡（NE）（Nash，1951年），它为每个玩家定义了一组稳定的边缘 $\sigma_p(a_p)$策略，这些策略可以独立执行。由于具有独立性（也称为可因子分解的联合策略属性），它通常用于纯竞争性（零和）情境。在双人零和博弈中，纳什均衡易于计算[2]，且具有可互换性[3]，也无法被利用[4]。在具有规范性的双人零和博弈中，纳什均衡被视为基础概念。有限博弈中的纳什均衡集合总是非空的。然而，一般而言，纳什均衡集合中的各个均衡可能彼此孤立。通常，如何选择最佳纳什均衡并不明确（即均衡选择问题（Harsanyi 和 Selten，1988年）），而寻找一个特定的纳什均衡是 NP 完全问题（Austrin 等人，2011）。寻找任意一个纳什均衡则

[2]可在多项式时间内求解。[3]如果存在任意两个均衡（$\sigma_1(a_1), \sigma_2(a_2)$）和（$\sigma'_1(a_1), \sigma'_2(a_2)$），那么（$\sigma_1(a_1), \sigma'_2(a_2)$）和（$\sigma'_1(a_1), \sigma_2(a_2)$）也都是均衡。[4]在均衡策略下，对手没有任何响应能够降低其收益。

PPAD 完全（Chen and Deng，2006年；Daskalakis 等人，2009年）。对于 n 人混合动机标准式博弈，其他中介均衡概念，例如相关均衡（CE）（Aumann，1974年）和粗相关均衡（CCE）（Hannan，1957年；Moulin and Vial，1978年）可能更为合适。这些概念描述了稳定联合策略分布所需满足的条件，$\sigma(a)$。（C）CE 允许更大的分布可行集，因此可能为玩家带来更高的收益。然而，它们无法独立执行：它们需要一个可信协调装置，向每位玩家推荐一个行动。如果环境允许协调，那么协调就是一种合理的属性。（C）CE 的主要优势在于，它们具有线性约束，并构成一个均衡凸多胞体。这意味着，对于 n 人一般和博弈，可以在多项式时间内计算出该集合。

充分支持相关均衡（WSCE）（Czumaj 等人，2014年；Goldbergand Papadimitriou，2006年）可以通过线性不等式约束来定义。玩家的偏离收益，$A\text{WSCE}_p(a'_p, a''_p, a_{-p})$，是指当玩家在被推荐策略 a''_p后偏离到策略 a'_p，而其他玩家采用 a_{-p}时，该玩家所获得的收益变化。若对于每一项推荐、偏离和玩家，a''_p、a'_p 和 p，联合策略 $\sigma(a)$ 的偏离收益均不超过某个常数 $\epsilon_p \leq \epsilon$，则该联合策略属于 ϵ-WSCE。标量 ϵ_p 描述了均衡的近似程度，同时通常也作为算法中的超参数，其取值一般设为零。偏离收益不等式可以表示为：

$$\sum_{a_{-p}\in\mathcal{A}_{-p}} \sigma(a_{-p}|a''_p)\left[A_p^{\text{WSCE}}(a'_p, a''_p, a_{-p}) := G_p(a'_p, a_{-p}) - G_p(a''_p, a_{-p})\right] \leq \epsilon_p \qquad \forall p \in [1, N], a''_p \neq a'_p \in \mathcal{A}_p. \tag{1}$$

<escape>条件分布较为繁琐，因此标准相关均衡（Aumann，1974年）改为通过联合分布来定义。当 $\epsilon_p = 0$时，WSCE 等价于相关均衡。</escape>

$$\sum_{a\in\mathcal{A}} \sigma(a)\left[A_p^{\text{CE}}(a'_p, a''_p, a) := \begin{cases} A_p^{\text{WSCE}}(a'_p, a''_p, a_{-p}) & a_p = a''_p \\ 0 & \text{otherwise} \end{cases}\right] \leq \epsilon_p \qquad \forall p \in [1, N], a''_p \neq a'_p \in \mathcal{A}_p \tag{2}$$

粗相关均衡（Hannan，1957年；Moulin andVial，1978年；Young，2004年）与相关均衡类似，但玩家只能在收到推荐策略之前考虑偏离。因此，粗相关均衡的偏离收益可以通过对所有可能的推荐策略求和，从相关均衡的定义中推导得出 a''_p：
$A_p^{\text{CCE}}(a'_p, a) = \sum_{a''_p} A_p^{\text{CE}}(a'_p, a''_p, a)$.

$$\sum_{a\in\mathcal{A}} \sigma(a)\left[A_p^{\text{CCE}}(a'_p, a) := G_p(a'_p, a_{-p}) - G_p(a)\right] \leq \epsilon_p \qquad \forall p \in [1, N], a'_p \in \mathcal{A}_p \tag{3}$$

纳什均衡（Nash，1951年）的定义与（相关）粗相关均衡类似，但增加了一个约束：联合分布必须能够因子分解$\otimes_p \sigma(a_p) = \sigma(a)$。无论是强制要求联合分布因子分解，还是直接优化边缘分布，都会导致非线性约束[5]。纳什均衡的定义也可以选择用粗相关均衡或相关均衡的偏离收益来表示。使用相关均衡的偏离收益，我们定义 ϵ-纳什均衡和 ϵ-弱粗相关均衡。

$$\sum_{a\in\mathcal{A}} \otimes_q \sigma(a_q) A_p^{\text{CE}}(a'_p, a''_p, a) \leq \epsilon_p \qquad \forall p \in [1, N], a''_p \neq a'_p \in \mathcal{A}_p \tag{4}$$

$$\sum_{a_{-p}\in\mathcal{A}_{-p}} \otimes_{q\neq p} \sigma(a_q) A_p^{\text{WSCE}}(a'_p, a''_p, a_{-p}) \leq \epsilon_p \qquad \forall p \in [1, N], a''_p \neq a'_p \in \mathcal{A}_p \tag{5}$$

均衡不变变换 某些对收益的变换，$G_p(a) \to \hat{G}_p(a)$，不会改变均衡集合，$\sigma^*(a) \to \hat{\sigma}^*(a) = \sigma^*(a)$。这些变换可用于减少博弈中的自由度。最常见的不变变换（Morris 和 Ui，2004年；Moulin 和 Vial，1978年；Ostrovski，2013年）是仿射变换（定理1），它由关于其他玩家策略的偏移量和一个正比例因子组成。Harsanyi 和 Selten（1988年）认为，不变变换和对称等价关系是均衡选择中的重要性质。Marris等人（2022a年）利用这两者，通过神经网络简化了（相关）粗相关均衡的计算。在机器学习中，不变性也得到了更广泛的应用，以便在丰富的拓扑空间上进行高效优化，例如半定规划和黎曼优化（Boumal等人，2014年）。本文利用现有的均衡不变变换，为一般的 n 玩家和 2×2 博弈推导出一种新的均衡不变嵌入，并提出一种新颖的同构关系（定理 4），借此重新发现一个基本的 2×2博弈类别集合。

[5]这就是为什么纳什均衡比（相关）粗相关均衡更难计算。

	A	B
A	g_1^{AA}, g_2^{AA}	g_1^{AB}, g_2^{AB}
B	g_1^{BA}, g_2^{BA}	g_1^{BB}, g_2^{BB}

(a) 完整标准式（8 个变量）

	A	B
A	g^{AA}	g^{AB}
B	g^{BA}	g^{BB}

(b) 对称、共同利益或零和游戏（4 个变量）

	A	B
A	0	g^{AB}
B	$-g^{AB}$	0

(c) 对称零和游戏（1 个变量）

表 1 | 描述 2×2标准式博弈及其更稀疏简化形式所需的变量数量。

双人双策略博弈2×2 博弈是可能的最小规模，但却具有显著的战略深度，并能解释现实世界中的许多交互行为。因此，已经积累了大量文献，致力于解释、可视化、分类、参数化和命名 2×2 博弈。直观来看，一个 2×2 博弈（表1a）需要 8 个变量来定义，这一数量过多，难以凭直觉把握。因此，许多方法试图通过利用不变性、对称性和等价关系，或考虑博弈的一个子集（例如对称博弈、共同收益博弈或零和博弈；表1b 和 1c），来降低这种复杂性。经过这种复杂性简化后，最终会得到一个有限的博弈集合或一个约化的博弈空间。该集合或空间可能具有某种结构，用以描述它与其他博弈之间的接近程度。仅将博弈划分到不同集合中的方法称为分类方法。那些还在博弈之间引入相似性或接近程度概念的方法则称为拓扑方法。那些为分类引入层次结构的方法称为分类体系。那些对博弈进行参数化的方法称为参数化方法。一种方法可能同时具有多种属性。

博弈可以根据其收益进行分类。Rapoport 和 Guyer（1966年）的“博弈分类”（以及他们的著作（Rapoport 等，1976年））利用了一个特定的等价类，在这个等价类中，只有每位玩家收益的顺序才重要。只要不改变顺序，收益大小的变化就会使博弈均衡发生可预测的缩放。此外，如果我们只考虑具有严格排序的博弈子集（严格序数博弈），就可以用序数集合 {1, 2, 3, 4} 的置换来表示收益。这样一来，每位玩家收益的严格排序方式有 4! = 24 种，由此得到 $24 \times 24 = 576$ 个博弈。为每位玩家采用策略置换[6]，可将严格序数博弈的数量减少到 144 个。再纳入玩家置换[7]，便可去除一半的非对称博弈，使严格序数博弈的数量减少到 78 个。这种方法的缺点是，它只能对收益顺序严格的博弈进行分类。Fraser 和 Kilgour（1986年）提出了对偏序序数博弈（偏序博弈）进行分类的方法，其中收益可以取相等的值。由此得到的偏序序数博弈总数为 1413 个（考虑策略对称性），若利用玩家对称性，则为 726 个。可以通过图形表示来直观呈现偏序博弈，其中展示了联合偏好的排序（图 1）。Goforth 和 Robinson（2005年）改进了严格序数博弈的分类方法，制作出一张“周期表”。144 个博弈被排列在一个 12×12 网格中，使相邻博弈具有相似的属性。Robinson 等人（2007年）将这一拓扑结构扩展到偏序博弈。Bruns（2015年）为所有序数博弈提出了一套公式化命名方案，因为此前只有一小部分博弈（通常是对称博弈）拥有公认的常用名称。

Harris（1969年）仅使用两个由收益函数决定的变量，对对称 2×2 博弈进行参数化分类（表 1b）：$r_3 = \frac{g^{BB}-g^{AB}}{g^{BA}-g^{AB}}$ 和 $r_4 = \frac{g^{BA}-g^{AA}}{g^{BA}-g^{AB}}$，并满足约束条件 ~~$g^{BA} > g^{AB}$~~。这~~定义了~~一个平面，其中不同区域对应着具有相似性质的博弈类别。Huertas-Rosero（2003年）也将对称 2×2 博弈划分为 8 个基本类别，并根据其纳什均衡进一步划分为总共 12 个类别。Böörs 等人（2022年）根据对称 2×2 博弈分解为共同收益部分和零和部分的方式，将其划分为 24 个类别。Germano（2006年）根据纳什均衡几何，将博弈划分为若干等价类别。Borm（1987年）根据最佳响应特征，将所有 2×2博弈划分为由 15 个博弈组成的集合。这些博弈还可以通过 4 个离散变量进行参数化。Borm 并未描述该集合中博弈之间的任何相似性或距离概念。我们的工作通过基于尺度的收益变换，重新发现了 Borm（1987年）的分类。此外，我们将这一分类扩展到度量空间，提供了一种效率高得多的二维嵌入方法来表示这些博弈，并为这些博弈命名。

[6]文献将其表述为“序关系图”等价关系。[7]文献中有时称之为“反射”。由于所有对称博弈都被保留，这种对称性并不能将空间完全减半。在 78 个严格序数博弈中，66 个是非对称的，12 个是对称的。文献较少利用这种对称性，因为它会互换玩家的角色。

3. 一般标准式博弈的度量空间

设 $G \in \mathrm{G} = (\mathrm{G}_1, ..., \mathrm{G}_N)$ 为具有特定数量玩家和策略的博弈空间。 $G_p \in \mathrm{G}_p = \mathrm{R}^{|\mathcal{A}_1| \times ... \times |\mathrm{A}^N|}$ 是玩家 p 的收益空间。我们可以为每个玩家引入近似参数（见公式 (1)） $\epsilon = (\epsilon_1, ..., \epsilon_N) \in \mathrm{R}^N$，并将其纳入博弈空间，从而得到一个新的元组空间 (G, ϵ)，我们称之为近似博弈空间。我们可以定义两个近似博弈 $d((G^A, \epsilon^A), (G^B, \epsilon^B))$ 之间的距离度量，使得近似标准式博弈的空间构成一个度量空间。令 $\sigma \in \Delta^{|\mathrm{A}|-1}$ 为概率单纯形。令 $\sigma^* \in (\mathrm{WS})(\mathrm{C})\mathrm{CE}(G, \epsilon) \subseteq \Delta^{|\mathrm{A}|-1}$ 为根据公式 (1)、(2)、(3) 或 (4) 处于均衡状态的联合分布子集。通常，我们选择 $\epsilon = 0$，因此可以省略与近似博弈相关的附加符号。

3.1. 均衡不变嵌入

我们关注的是研究博弈变换，$(G, \epsilon) \to (\hat{G}, \hat{\epsilon})$，这些变换不会改变近似均衡的集合，$(\mathrm{WS})(\mathrm{C})\mathrm{CE}(G, \epsilon) = (\mathrm{WS})(\mathrm{C})\mathrm{CE}(\hat{G}, \hat{\epsilon})$。我们将这种变换称为均衡不变变换。最常见的此类变换（Morris 和 Ui，2004年；Moulin 和 Vial，1978年；Ostrovski，2013年）是仿射变换，它由针对其他玩家策略的偏移量和一个正比例因子构成。

定理1 (仿射变换)。 *ϵ-NE*、*ϵ-WSNE*、*ϵ-CE*、*ϵ-WSCE*和*ϵ-CCE* 在每个玩家的收益进行仿射变换下均为均衡不变。具体而言，当

$$G_p(a) \to \hat{G}_p(a) = s_p G_p(a_p, a_{-p}) + b_p(a_{-p}), \quad and \quad \epsilon_p \to \hat{\epsilon}_p = s_p \epsilon_p, \tag{6}$$

原博弈中的一个 ϵ_p-均衡是变换后博弈中的一个 $s_p\epsilon_p$-均衡： $\sigma(a) \to \hat{\sigma}(a) = \sigma(a)$，其中$b_p(a_{-p})$ 是任意偏移量， s_p 是任意正标量。

证明。 考虑这些变换对偏离收益的影响。

$$A^{\mathrm{WSCE}}_p(a'_p, a''_p, a_{-p}) \to s_p G_p(a'_p, a_{-p}) + b_p(a_{-p}) - s_p G_p(a''_p, a_{-p}) - b_p(a_{-p}) = s_p A^{\mathrm{WSCE}}_p(a'_p, a''_p, a_{-p}) \tag{7a}$$

$$A^{\text{相关均衡}}_p(a'_p, a''_p, a) \to \begin{cases} s_p A^{\mathrm{WSCE}}_p(a'_p, a''_p, a_{-p}) & a_p = a''_p \\ 0 & \text{否则} \end{cases} = s_p A^{\mathrm{CE}}_p(a'_p, a''_p, a) \tag{7b}$$

$$A^{\mathrm{CCE}}_p(a'_p, a) \to s_p G_p(a'_p, a_{-p}) + b_p(a_{-p}) - s_p G_p(a) - b_p(a_{-p}) = s_p A^{\mathrm{CCE}}_p(a'_p, a) \tag{7c}$$

均衡完全由其不等式约束定义（方程（1）、（2）、（3）和（4））。仿射变换只会使不等式左侧乘以一个 s_p 尺度因子。如果我们将同一正比例因子应用于定义的右侧，不等式仍然成立。因此，未变换博弈中的一个 ϵ_p-均衡将是变换后博弈中的一个 $s_p\epsilon_p$-均衡。如果 $\epsilon_p = 0$ 均衡将保持不变。□

若 b_p 和 s_p 根据收益本身定义，则可通过仿射变换将每个玩家收益中的自由度减少 $|\mathrm{A}_{-p}| + 1$，同时不改变均衡。以均值 $b_p(a_{-p}) = -\frac{1}{|\mathcal{A}_p|}\sum_{a_p} G_p(a_p, a_{-p})$ 为偏移量，并以收益张量的弗罗贝尼乌斯范数 $s_p = 1/||G_p||_F$ 进行缩放，都是合理的选择。我们可以利用这一变换来设计距离度量 $d$equil，以及一种保距均衡不变嵌入。令 $\mathrm{G}^{\mathrm{equil}}$ 为通过保结构映射嵌入 G 的映射，使得 $\mathrm{G}^{\mathrm{equil}} \subset \mathrm{G}$。$\mathrm{G}^{\mathrm{equil}}$ 是一个流形和度量空间。

定义 1 （均衡不变嵌入）。

$$G^{equil}_p(a) = \frac{1}{Z}\left(G_p(a) - \frac{1}{|\mathcal{A}_p|}\sum_{a_p} G_p(a_p, a_{-p})\right) \qquad \epsilon^{equil}_p = \frac{1}{Z}\epsilon_p \qquad Z = \left\| G_p - \frac{1}{|\mathcal{A}_p|}\sum_{a_p} G_p(a_p, a_{-p}) \right\|_F \tag{8}$$

定义 2 （均衡不变距离度量）。

$$d^{equil}\left((G^A, \epsilon^A), (G^B, \epsilon^B)\right) = \sqrt{\sum_p \arccos\left(\sum_a G^{A,equil}_p(a) G^{B,equil}_p(a)\right)^2} + \sqrt{\sum_p (\epsilon^{A,equil}_p - \epsilon^{B,equil}_p)^2} \tag{9}$$

这个定义存在一个问题：当收益平凡 时，它的定义不明确，因为对于平凡博弈，归一化常数 Z 为零。在所有玩家的收益都平凡的博弈中，所有联合分布都是均衡。

定义 3（平凡收益）。

$$G_p(a_p, a_{-p}) = b_p(a_{-p}) \qquad \forall a_{-p} \in \mathcal{A}_{-p} \tag{10}$$

定义 4（非平凡收益）。

$$G_p(a_p, a_{-p}) \neq b_p(a_{-p}) \qquad \exists a_{-p} \in \mathcal{A}_{-p} \tag{11}$$

只需将零向量的范数定义为单位长度即可轻松解决这一问题，$\|\mathbf{0}\| = 1$；当收益为平凡收益时，就会出现这种情况。对于嵌入而言，这样定义很自然。对于距离，内积 $\sum_a G\text{A,equil}_p(a) G\text{B,equil}_p(a)$ 将为零，这等价于收益彼此正交。平凡收益被定义为与所有其他收益正交。这一界定并非不合理，但在我们的研究中，最常处理的是非平凡博弈，或者只比较具有相同平凡性结构的博弈；我们将在下文中采用这一假设。

均衡不变嵌入是一个斜流形（Trendafilov 和 Lippert，2002年）（由 N 个单位球面构成的乘积流形）。每个玩家的收益嵌入 $G\text{equil}_p$，都是其中一个球面表面上的点。两个博弈之间的距离，是这两个点在各个球面上的弧长所构成向量的范数。因此，两个 $\epsilon = 0$ 博弈之间的最大距离为 $\sqrt{N}\pi$。均衡不变嵌入将每个玩家收益的自由度减少了 $|\mathrm{A}_{-p}| + 1$。线性偏移分量贡献了 $|\mathrm{A}_{-p}|$ 这一部分，而非线性缩放则贡献了剩余的单位。事实证明，$|\mathrm{A}_{-p}|$是线性函数所能实现的最大降幅。

定理 2（线性偏移秩约简）。均衡不变嵌入的偏移量分量将收益的自由度减少了 $|\mathcal{A}_{-p}|$。这是在不改变均衡的情况下，通过线性变换所能减少的最大自由度。

证明。 偏离收益的计算 A（方程（1）、（2）和（3））是一种线性运算，因此可以表示为算子矩阵 T_p与玩家收益 G_p的矩阵乘法。我们使用收益和偏离收益的展平形式。

$$A_p^{\text{WSCE}}(a'_p \otimes a''_p, a_{-p}) = \sum_{a'''} T_p^{\text{WSCE}}(a'_p \otimes a''_p \otimes a_{-p}, a''') G_p(a''') \tag{12a}$$

$$A_p^{\text{CE}}(a'_p \otimes a''_p \otimes a) = \sum_{a'''} T_p^{\text{CE}}(a'_p \otimes a''_p \otimes a, a''') G_p(a''') \tag{12b}$$

$$A_p^{\text{CCE}}(a'_p \otimes a) = \sum_{a'''} T_p^{\text{CCE}}(a'_p \otimes a, a''') G_p(a''') \tag{12c}$$

通过考察算子矩阵的结构（第 A 节，引理1），我们可以计算其秩。

$$\operatorname{rank}\left(T_p^{\text{WSCE}}\right) = \operatorname{rank}\left(T_p^{\text{CE}}\right) = \operatorname{rank}\left(T_p^{\text{CCE}}\right) = |\mathcal{A}| - |\mathcal{A}_{-p}| \qquad \forall p \in [1, N] \tag{13}$$

一般而言，收益 G_p 可以是满秩矩阵，$\mathrm{R}^{|\mathcal{A}|}$。但是，与秩仅为 $|\mathrm{A}| - |\mathrm{A}_{-p}|$ 的算子矩阵相乘后，所得的偏离收益的秩最多为 $|\mathrm{A}| - |\mathrm{A}_{-p}|$。任何将博弈空间缩减超过 $|\mathrm{A}_{-p}|$ 的线性均衡不变变换，都意味着算子矩阵 T_p 的秩小于 $|\mathrm{A}| - |\mathrm{A}_{-p}|$，从而减少定义均衡的不等式约束的数量。因此，任何线性均衡不变变换最多只能将博弈空间缩减 $|\mathrm{A}_{-p}|$，而不会改变所有博弈的均衡集合。仿射变换的偏移量分量将自由度减少 $|\mathrm{A}_{-p}|$，并且根据定理1，不会改变偏离收益。□

可逆偏差收益一般而言，从收益到偏离收益的映射是不可逆的，因为它并非满秩线性运算：许多不同的博弈都可能产生相同的偏离收益。然而，在均衡不变嵌入中，这种映射是一一对应的，因此可以逆向执行这一过程，根据偏离收益求出不变嵌入。

算法1： 均衡不变嵌入采样

采样

数据： $|\mathcal{A}_1|$, ……, $|\mathcal{A}_N|$

结果： G均衡

$N \leftarrow$ 长度($|\mathcal{A}_1|$, ……, $|\mathcal{A}_N|$);

对于 $p \leftarrow [1, N]$，**执行 do**

assert $|\mathrm{A}_p| \geq 2$;

$G_p(a) \leftarrow \mathrm{N}(0, 1)$;

$G_p(a) \leftarrow G_p(a) - \frac{1}{|\mathrm{A}_p|} \sum_{a_p \in \mathcal{A}_p} G_p(a_p, a_{-p})$;

$G_p(a) \leftarrow \frac{G_p(a)}{||G_p(a)||_2}$;

G均衡 $\leftarrow$ 拼接(G_1, ……, G_N);

算法2： 平凡嵌入采样

数据： $|\mathcal{A}_1|$, ..., $|\mathcal{A}_N|$

结果： G平凡

$N \leftarrow$ 长度($|\mathcal{A}_1|$, ……, $|\mathcal{A}_N|$);

对于 $p \leftarrow [1, N]$，**执行**

断言 $|\mathrm{A}_p| \geq 2$;

$b_p(a_{-p}) \leftarrow \mathrm{N}(0, 1)$;

$G_p(a) \leftarrow b_p(a_{-p})$;

$G_p(a) \leftarrow \frac{G_p(a)}{||G_p(a)||_2}$;

Gtri $\leftarrow$ 拼接(G_1, ……, G_N);

定理3 (可逆偏差增益)。 均衡不变嵌入可以从偏离收益中恢复。

$$G_p^{equil}(a) = -\frac{1}{|\mathcal{A}_p|} \sum_{a'_p} A_p^{WSCE}(a'_p, a_p, a_{-p}) = -\frac{1}{|\mathcal{A}_p|} \sum_{a'_p, a''_p} A_p^{CE}(a'_p, a''_p, a) = -\frac{1}{|\mathcal{A}_p|} \sum_{a'_p} A_p^{CCE}(a'_p, a) \tag{14}$$

证明。 回顾 CCE 偏离收益的定义（公式（3））。对玩家自身的偏离策略取均值， $a'_p \in \mathcal{A}_p$，并重新整理。

$$A_p^{\mathrm{CCE}}(a'_p, a) = G_p(a'_p, a_{-p}) - G_p(a) \implies G_p(a) = \underbrace{\frac{1}{|\mathcal{A}_p|} \sum_{a'_p} G_p(a'_p, a_{-p})}_{\text{Zero when zero-mean}} - \frac{1}{|\mathcal{A}_p|} \sum_{a'_p} A_p^{\mathrm{CCE}}(a'_p, a) \tag{15}$$

当收益为不变嵌入时， Gequil$_p$ (a)，均值项为零，由此完成了粗相关均衡的证明。注意到 $A\mathrm{CCE}_p(a'_p, a) = \sum_{a''_p \in \mathcal{A}_p} A\mathrm{CE}_p(a'_p, a''_p, a)$，我们可以得到相关均衡的类似解。注意到$A\mathrm{WSCE}_p(a'_p, a''_p, a_{-p}) = \sum_{a_p \in \mathcal{A}_p} A\mathrm{CE}_p(a'_p, a''_p, a)$，我们可以得到 WSCE 的类似解。□

本文并未直接使用这一结果，但我们将其纳入文中，原因有以下几点。首先，它有助于加深我们对博弈空间以及收益与均衡定义及其约束之间关系的直观理解。其次，它为均衡不变嵌入的根本性质提供了进一步的证据。最后，它为逆博弈理论开辟了研究方向；逆博弈理论旨在根据观测到的均衡行为恢复收益。

均衡不变嵌入采样 在不变嵌入（算法 1）和平凡博弈（算法 2）上进行均匀采样很容易，其中 $\mathrm{N}(0, 1)$ 是均值为零、方差为一的正态分布。这种采样方法有坚实的理论依据，因为它所采样的博弈涵盖了所有有趣的策略互动。文献中常见一些理论依据较弱的采样方法（例如，对收益中的各个条目采用均匀分布， $G_p(a) = \mathrm{U}(0, 1)$），但这些方法可能无法均匀覆盖博弈的策略空间。我们建议使用均衡不变嵌入来评估均衡求解器，并生成训练和测试数据集。

3.2. 较优响应嵌入

此前我们证明了收益缩放（仿射博弈变换中的一种操作）具有均衡不变性；在此，我们证明一种新颖的逐策略缩放变换具有较优响应不变性[8]。该变换会使变换后的博弈中 ϵ-纳什均衡、 ϵ-弱粗相关均衡、 ϵ-相关均衡和 ϵ-WSCE 对应于其倒数缩放后的均衡。

[8]较优响应不变性意味着最佳响应不变性 nce.

定义 5（最佳响应不变）。

$$\arg\max_{a_p} G_p(a_p, a_{-p}) = \arg\max_{a_p} \hat{G}_p(a_p, a_{-p}) \quad \forall a_{-p} \in \mathcal{A}_{-p} \tag{16}$$

定义 6 (较优响应不变[9])。

$$\arg\operatorname{sort}_{a_p} G_p(a_p, a_{-p}) = \arg\operatorname{sort}_{a_p} \hat{G}_p(a_p, a_{-p}) \quad \forall a_{-p} \in \mathcal{A}_{-p} \tag{17}$$

定理 4 (逐策略缩放变换)。 ϵ-纳什均衡、 ϵ-弱粗相关均衡、 ϵ-相关均衡、 ϵ-*WSCE* 在对每个玩家的收益进行正的逐策略缩放后仍具有较优响应不变性，而均衡会相应地进行互为倒数的逐策略缩放（$s_p \to s_p(a_p)$）。具体而言，当

$$G_p(a) \to \hat{G}_p(a) = \left(\otimes_{q\in -p} s_q(a_q)\right) G_p(a),\ \textit{and}\ \epsilon_p^{WSCE} \to \hat{\epsilon}_p^{WSCE}(\sigma(a)) = \frac{\epsilon_p^{WSCE}}{Z_{-p}(a_p'')}\ \textit{or}\ \epsilon_p^{CE} \to \hat{\epsilon}_p^{CE}(\sigma(a)) = \frac{\epsilon_p^{CE}}{Zs_p(a_p'')}, \tag{18}$$

原博弈中的一个均衡有一个对应的均衡，

$$\sigma(a) \to \hat{\sigma}(a) = \frac{1}{Z\left(\otimes_p s_p(a_p)\right)}\sigma(a), \tag{19}$$

在变换后的博弈中，其中 $Z = \sum_{a\in\mathcal{A}} \sigma(a) \otimes_p s_p(a_p)$ 且 $Z_{-p}(a_p'') = \sum_{a_{-p}} {}^{\sigma(a_p'',\ a_{-p})} \otimes_{-p} s_p(a_p)$。

证明。 考虑这些变换对偏离收益的影响。

$$A_p^{\mathrm{WSCE}}(a_p', a_p'', a_{-p}) \to \qquad \hat{A}_p^{\mathrm{WSCE}}(a_p', a_p'', a_{-p}) = \left(\otimes_{q\in -p} s_q(a_q)\right) A_p^{\mathrm{WSCE}}(a_p', a_p'', a_{-p}) \tag{20a}$$

$$A_p^{\mathrm{CE}}(a_p', a_p'', a) \to \qquad \hat{A}_p^{\mathrm{CE}}(a_p', a_p'', a) = \left(\otimes_{q\in -p} s_q(a_q)\right) A_p^{\mathrm{CE}}(a_p', a_p'', a) \tag{20b}$$

将变换后的偏离收益和近似值代入相关均衡的定义（方程 2）。这可以证明等价于由未变换博弈构成的定义。

$$\sum_{a\in\mathcal{A}} \hat{\sigma}(a) \hat{A}_p^{\mathrm{CE}}(a_p', a_p'', a) \le \hat{\epsilon}_p \qquad \forall p \in [1, N], a_p'' \ne a_p' \in \mathcal{A}_p \tag{21a}$$

$$\sum_{a\in\mathcal{A}} \left[\frac{1}{Z\left(\otimes_p s_p(a_p)\right)}\sigma(a)\right] \left[\left(\otimes_{q\in -p} s_q(a_q)\right) A_p^{\mathrm{CE}}(a_p', a_p'', a)\right] \le \frac{1}{Zs_p(a_p'')}\epsilon_p \qquad \forall p \in [1, N], a_p'' \ne a_p' \in \mathcal{A}_p \tag{21b}$$

$$\sum_{a\in\mathcal{A}} \frac{1}{Zs_p(a_p)}\sigma(a) A_p^{\mathrm{CE}}(a_p', a_p'', a) \le \frac{1}{Zs_p(a_p'')}\epsilon_p \qquad \forall p \in [1, N], a_p'' \ne a_p' \in \mathcal{A}_p \tag{21c}$$

$$\sum_{a\in\mathcal{A}} \sigma(a) A_p^{\mathrm{CE}}(a_p', a_p'', a) \le \epsilon_p \qquad \forall p \in [1, N], a_p'' \ne a_p' \in \mathcal{A}_p \tag{21d}$$

通过在左侧将 a_p 变量代入 a_p''， $s_p(a_p)$ 和 $s_p(a_p'')$ 两项相互抵消；根据 $A\mathrm{CE}(a_p', a_p'', a)$ 的定义（公式 2），这种代入是允许的。因此，我们已证明：如果 $\sigma(a)$ 是未变换博弈中的一个均衡，那么 $\hat{}\sigma(a)$ 是变换后的博弈中的一个均衡，而且 $\hat{}\sigma(a)$ 可以直接由 $\sigma(a)$ 和 $s_p(a_p)$ 计算得出。现在考虑 WSCE 的定义（公式 1）。

$$\sum_{a_{-p}\in\mathcal{A}_{-p}} \hat{\sigma}(a_{-p}|a_p'') \hat{A}_p^{\mathrm{WSCE}}(a_p', a_p'', a_{-p}) \le \hat{\epsilon}_p \qquad \forall p \in [1, N], a_p'' \ne a_p' \in \mathcal{A}_p \tag{22a}$$

$$\sum_{a_{-p}\in\mathcal{A}_{-p}} \hat{\sigma}(a_p'', a_{-p}) \hat{A}_p^{\mathrm{WSCE}}(a_p', a_p'', a_{-p}) \le \hat{\sigma}(a_p'')\hat{\epsilon}_p \qquad \forall p \in [1, N], a_p'' \ne a_p' \in \mathcal{A}_p \tag{22b}$$

$$\sum_{a_{-p}\in\mathcal{A}_{-p}} \frac{1}{Zs_p(a_p'')}\sigma(a_p'', a_{-p}) A_p^{\mathrm{WSCE}}(a_p', a_p'', a_{-p}) \le \left(\sum_{a_{-p}} \frac{\sigma(a_p'', a_{-p})}{Zs_p(a_p'')\left(\otimes_{-p} s_p(a_p)\right)}\right)\hat{\epsilon}_p \qquad \forall p \in [1, N], a_p'' \ne a_p' \in \mathcal{A}_p \tag{22c}$$

$$\sum_{a_{-p}\in\mathcal{A}_{-p}} \sigma(a_p'', a_{-p}) A_p^{\mathrm{WSCE}}(a_p', a_p'', a_{-p}) \le \epsilon_p \qquad \forall p \in [1, N], a_p'' \ne a_p' \in \mathcal{A}_p \tag{22d}$$

[9]严格来说，我们要求参数排序按照行动索引来打破收益之间的平局。

NE 和 WSNE 的定义与 CE 和 WSCE 相同，但额外要求联合策略可分解为乘积形式： $\sigma(a) = \otimes_p \sigma_p(a_p)$。这一额外约束不会影响上述证明，因此该结果对 NE 和 WSNE 也成立。收益仅在其他玩家的策略上按正尺度缩放，这意味着具有最佳响应不变性。□

值得注意的是，定理4 对粗相关均衡并不成立。

定理 5 (逐策略缩放粗相关均衡反例)。 对每个玩家的收益进行逐策略缩放，并不会导致粗相关均衡发生相应的逐策略缩放。

证明。 考虑双人三策略博弈——缩放版石头剪刀布（G_1， G_2），该博弈在下文定义的 σ 处具有一个 CCE（以及其他均衡）。

$$G_1 = \begin{bmatrix} 0 & 4 & -1 \\ -2 & 0 & 1 \\ 2 & -4 & 0 \end{bmatrix} \quad G_2 = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix} \quad \sigma = \begin{bmatrix} 0 & \frac{1}{4} & \frac{1}{12} \\ \frac{1}{4} & 0 & \frac{1}{12} \\ 0 & 0 & \frac{1}{3} \end{bmatrix} \tag{23}$$

现在通过 $s_1 = [1, 1, 1]$ 和 $s_2 = [\frac{1}{2}, \frac{1}{4}, 1]$ 对该博弈进行尺度变换，从而得到熟悉的石头剪刀布博弈（$\hat{}G_1$， G_2）。

$$\hat{G}_1 = \begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix} \quad G_2 = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix} \quad \hat{\sigma} = \frac{1}{Z} \begin{bmatrix} 0 & 1 & \frac{1}{12} \\ \frac{1}{2} & 0 & \frac{1}{12} \\ 0 & 0 & \frac{1}{3} \end{bmatrix} \tag{24}$$

请注意，根据定理 σ， $\hat{}\sigma$计算得到的缩放版本并不是缩放后博弈的均衡。因此，通过反例可知，定理 4 一般情况下并不适用于粗相关均衡。□

幸运的是，对于双策略博弈，相关均衡与粗相关均衡等价（Monnot 和 Piliouras，2017）。在为 2×2博弈推导嵌入时，我们会利用这一性质。

注 1（双策略（相关）均衡等价关系）。对于每位玩家都有两个策略的 n 人博弈， ϵ-粗相关均衡和 $d\epsilon$-相关均衡等价 ϵ 和

n 人博弈的嵌入和距离度量可以通过逐策略缩放变换轻松定义。然而，对于双人博弈，使用零和游戏近似 $\epsilon = 0$，有一种自然的嵌入定义。

定义 7 （双人更优响应不变嵌入）。

$$G_1^{res}(a_1, a_2) = \frac{1}{Z(a_2)} \left(G_1(a_1, a_2) - \frac{1}{|\mathcal{A}_1|} \sum_{a_1} G_p(a_1, a_2) \right) \qquad Z(a_2) = \left\| G_1(:, a_2) - \frac{1}{|\mathcal{A}_1|} \sum_{a_1} G_1(a_1, a_2) \right\|_F \tag{25a}$$

$$G_2^{res}(a_1, a_2) = \frac{1}{Z(a_1)} \left(G_1(a_1, a_2) - \frac{1}{|\mathcal{A}_2|} \sum_{a_2} G_p(a_1, a_2) \right) \qquad Z(a_1) = \left\| G_1(a_1, :) - \frac{1}{|\mathcal{A}_2|} \sum_{a_2} G_1(a_1, a_2) \right\|_F \tag{25b}$$

定义 8 （双人更优响应不变距离度量）。

$$d^{res}\left(G^A, G^B\right) = \sqrt{\sum_{a_2} \arccos\left(\sum_{a_1} G_1^{A,res}(a_1, a_2) G_p^{B,res}(a_1, a_2)) \right)^2 + \sum_{a_1} \arccos\left(\sum_{a_2} G_1^{A,res}(a_1, a_2) G_p^{B,res}(a_1, a_2)) \right)^2} \tag{26}$$

因此，更优响应不变博弈嵌入是一个斜流形（由$|A_1| + |A_2|$ 个单位球面构成的乘积流形）。对于每个玩家的每种其他玩家策略所对应的玩家收益，嵌入的切片是这些球面之一表面上的一个点。两个博弈之间的距离是这些球面上对应两点之间弧长所构成向量的范数。因此，两个博弈之间的最大距离为$\sqrt{|A_1| + |A_2|}\pi$。

3.3. 对称嵌入

在标准式博弈中，策略的排列顺序是任意的，因此，可以将仅在策略置换上存在差异的博弈视为相等。此外，如果玩家的角色并不重要[10]，，那么也可以考虑在玩家置换意义下的等价性。如果我们定义策略的规范排序，就可以缩小需要考虑的均衡不变嵌入区域。以下是一种可能的排序定义。首先，对于每个玩家，p，独立地对所有其他玩家的策略所对应的元素进行排序，a_{-p}，然后按照字典序对该玩家自身的策略进行排序，a_p，从而得到一个顺序置换，$\tau_p^*(a_p)$。其次，为了确定玩家的顺序，先对每个玩家的完整收益进行排序，再按照字典序对玩家进行排序，从而得到一个玩家顺序置换，$\omega^*(p)$。

$$G'_p(a_p, a_{-p}) = \operatorname*{sort}_{a_{-p}} G_p(a_p, a_{-p}) \quad \forall p \qquad \tau_p^*(a_p) = \arg\operatorname*{lexsort}_{a_p} G'_p(a_p, a_{-p}) \quad \forall p \tag{27a}$$

$$G''_p(a) = \operatorname*{sort}_{a} G_p(a) \quad \forall p \qquad \omega^*(p) = \arg\operatorname*{lexsort}_{p} G''_p(a) \tag{27b}$$

当策略具有相同收益时，会出现偏序；因此，即使置换仅部分有序，所得收益仍是唯一的。这些置换可用于定义另一种嵌入：对称博弈嵌入。

定义 9（均衡对称博弈嵌入）。

$$G_p^{sym}(a) = G_{\omega^*(p)}(\tau^*_{\omega^*(p)}(a_{\omega^*(p)}), \ldots) \qquad \epsilon_p^{sym} = \hat{\epsilon}_{\omega^*(p)} \tag{28}$$

对称性不会减少自由度的数量，但可以通过利用博弈定义中的对称性来缩小博弈空间的范围。在一个标准式博弈中，存在 $\prod_p (|\mathsf{A}_p|!)$ 种这样的策略置换对称性，以及 $N!$ 种玩家置换对称性。这些对称性应与均衡不变嵌入或较优响应不变嵌入结合使用。

定义 10（均衡对称距离度量）。

$$d^{sym}\left((G^A, \epsilon^A), (G^B, \epsilon^B)\right) = \min_{\tau_p(a_p), \omega(p)} \sqrt{\sum_p \arccos\left(\sum_a G^{A,equil}_{\omega(p)}(\tau_{\omega(p)}(a_{\omega(p)}), \ldots) G_p^{B,equil}(a)\right)^2} + \sqrt{\sum_p \left(\epsilon^{A,equil}_{\omega(p)} - \epsilon_p^{B,equil}\right)^2} \tag{29}$$

4. 2×2均衡不变嵌入

我们现在聚焦于 2×2博弈的度量空间。

4.1. 推导 2×2均衡不变嵌入

2×2 博弈具有一种特别高效的均衡不变嵌入，仅由两个变量参数化。该嵌入的距离度量与上述定义的距离度量等价，因此在 2×2博弈上同样构成一个度量空间。

定理 6 (2×2 均衡不变嵌入)。所有非平凡 2×2 博弈收益矩阵 G_p，均可映射为仅由两个变量参数化的收益矩阵 $Gequil_p$，且不会改变博弈的均衡。该映射由方程 (30a) 和 (30b)给出，其中 $\theta_1 + \frac{\pi}{4} = arctan2(g_1^{AA} - g_1^{BA}, g_1^{AB} - g_1^{BB})$ 以及 $\theta_2 + \frac{\pi}{4} = arctan2(g_2^{AA} - g_2^{AB}, g_2^{BA} - g_2^{BB})$。

$$G_1 = \begin{bmatrix} g_1^{AA} & g_1^{AB} \\ g_1^{BA} & g_1^{BB} \end{bmatrix} \to G_1^{equil}(\theta_1) = \begin{bmatrix} \frac{1}{\sqrt{2}}\sin(\theta_1 + \frac{\pi}{4}) & \frac{1}{\sqrt{2}}\cos(\theta_1 + \frac{\pi}{4}) \\ -\frac{1}{\sqrt{2}}\sin(\theta_1 + \frac{\pi}{4}) & -\frac{1}{\sqrt{2}}\cos(\theta_1 + \frac{\pi}{4}) \end{bmatrix} \tag{30a}$$

$$G_2 = \begin{bmatrix} g_2^{AA} & g_2^{AB} \\ g_2^{BA} & g_2^{BB} \end{bmatrix} \to G_2^{equil}(\theta_2) = \begin{bmatrix} \frac{1}{\sqrt{2}}\sin(\theta_2 + \frac{\pi}{4}) & -\frac{1}{\sqrt{2}}\sin(\theta_2 + \frac{\pi}{4}) \\ \frac{1}{\sqrt{2}}\cos(\theta_2 + \frac{\pi}{4}) & -\frac{1}{\sqrt{2}}\cos(\theta_2 + \frac{\pi}{4}) \end{bmatrix} \tag{30b}$$

[10]如果每个玩家可使用不同的策略或不同数量的策略，角色可能会很重要。

证明。 考虑玩家1的收益 G_1。首先应用仿射不变变换的偏移量$G_1(a_1,\ a_2) \ \rightarrow \ \bar{G}_1(a_1,\ a_2) \ = \ G_1(a_1,\ a_2) \ + b_1(a_2)$，其中玩家1的收益矩阵的列（另一玩家的策略）被归一化为零均值偏移，$b_1(a_2) = -\frac{1}{|\mathcal{A}_1|}\sum_{a_1 \in \mathcal{A}_1} G_1(a_1,\ a_2)$，使得 $b_1 = [-\frac{1}{2}g_1^{AA} - \frac{1}{2}g_1^{BA},\ -\frac{1}{2}g_1^{AB} - \frac{1}{2}g_1^{BB}]$。然后，使用 $g^{A-B,\ A} = g_1^{AA} - g_1^{BA}$和 $g^{A-B,\ B} = g_1^{AB} - g_1^{BB}$进行变量代换，从而将描述收益所需的参数数量从 4 → 2减少。

$$\begin{bmatrix} g_1^{AA} & g_1^{AB} \\ g_1^{BA} & g_1^{BB} \end{bmatrix} \rightarrow \bar{G}_1 = \begin{bmatrix} \frac{1}{2}(g_1^{AA} - g_1^{BA}) & \frac{1}{2}(g_1^{AB} - g_1^{BB}) \\ \frac{1}{2}(g_1^{BA} - g_1^{AA}) & \frac{1}{2}(g_1^{BB} - g_1^{BA}) \end{bmatrix} = \begin{bmatrix} \frac{1}{2}g_1^{A-B,A} & \frac{1}{2}g_1^{A-B,B} \\ -\frac{1}{2}g_1^{A-B,A} & -\frac{1}{2}g_1^{A-B,B} \end{bmatrix} \tag{31}$$

现在应用仿射不变变换的尺度，即单位 L_2 归一化，$\bar{G}_1(a_1,\ a_2) \ \rightarrow G$均衡 $_1(a_1,\ a_2) = s_1\bar{G}_1(a_1,\ a_2)$，其中 $s_1 = \frac{1}{||\bar{G}_1||_2}$，这确保收益中的所有元素的范数均为 1。只要收益是非平凡的（零均值偏移后不为零），这种变换就是有效的。

$$\begin{bmatrix} \frac{1}{2}g_1^{A-B,A} & \frac{1}{2}g_1^{A-B,B} \\ -\frac{1}{2}g_1^{A-B,A} & -\frac{1}{2}g_1^{A-B,B} \end{bmatrix} \rightarrow \begin{bmatrix} \frac{1}{\sqrt{2}}\frac{g_1^{A-B,A}}{\sqrt{g_1^{A-B,A^2}+g_1^{A-B,B^2}}} & \frac{1}{\sqrt{2}}\frac{g_1^{A-B,B}}{\sqrt{g_1^{A-B,A^2}+g_1^{A-B,B^2}}} \\ -\frac{1}{\sqrt{2}}\frac{g_1^{A-B,A}}{\sqrt{g_1^{A-B,A^2}+g_1^{A-B,B^2}}} & -\frac{1}{\sqrt{2}}\frac{g_1^{A-B,B}}{\sqrt{g_1^{A-B,A^2}+g_1^{A-B,B^2}}} \end{bmatrix} \tag{32}$$

请注意，　收益中各元素的范数现在等于 1，因此我们现在有如下性质：

$$\left(\frac{g_1^{A-B,A}}{\sqrt{g_1^{A-B,A^2} + g_1^{A-B,B}}}\right)^2 + \left(\frac{g_1^{A-B,B}}{\sqrt{g_1^{A-B,A^2} + g_1^{A-B,B^2}}}\right)^2 = 1 \tag{33}$$

这是单位圆的方程，相对参数和相邻参数可以用单个角度参数 θ_1 表示，其中 $\frac{\pi}{4}$ 是为可视化而任意选取的偏移量。这进一步将描述玩家1收益所需的参数数量从 2 → 1 减少。

$$\tan\left(\theta_1 + \frac{\pi}{4}\right) = \frac{g_1^{A-B,A}}{g_1^{A-B,B}} \implies \theta_1 + \frac{\pi}{4} = \text{arctan2}(g_1^{A-B,A}, g_1^{A-B,B}) \tag{34}$$

函数 arctan2 定义为 Organick (1966年). 由于单位圆的半径为 1，因此我们可以直接从 θ_1 中恢复收益的各个元素。

$$G_1^{\text{equil}}(\theta_1) = \begin{bmatrix} \frac{1}{\sqrt{2}}\sin(\theta_1 + \frac{\pi}{4}) & \frac{1}{\sqrt{2}}\cos(\theta_1 + \frac{\pi}{4}) \\ -\frac{1}{\sqrt{2}}\sin(\theta_1 + \frac{\pi}{4}) & -\frac{1}{\sqrt{2}}\cos(\theta_1 + \frac{\pi}{4}) \end{bmatrix} \tag{35}$$

对于玩家2，也可以使用一组等价的约简，其中 $\theta_2 + \frac{\pi}{4} = \text{arctan2}(g^{A-B,\ A}_{\cdot\ 2},\ g^{A-B,\ B}_{\cdot\ 2})$。

$$G_2 = \begin{bmatrix} g_2^{AA} & g_2^{AB} \\ g_2^{BA} & g_2^{BB} \end{bmatrix} \rightarrow G_2^{\text{equil}}(\theta_2) = \begin{bmatrix} \frac{1}{\sqrt{2}}\sin(\theta_2 + \frac{\pi}{4}) & -\frac{1}{\sqrt{2}}\sin(\theta_2 + \frac{\pi}{4}) \\ \frac{1}{\sqrt{2}}\cos(\theta_2 + \frac{\pi}{4}) & -\frac{1}{\sqrt{2}}\cos(\theta_2 + \frac{\pi}{4}) \end{bmatrix} \tag{36}$$

□

因此，所有非平凡的 2×2 博弈均可使用两个变量进行参数化 (θ_1, θ_2)，这两个变量通过两个圆上的角度在几何上描述各个点 [11]。由此可以得出若干有趣的结构性质（图 3），这些性质将在接下来的章节中进行解释。类似地，部分平凡不变嵌入也可以在一维空间中进行可视化。对于部分平凡博弈，其中一名玩家对博弈不具有任何影响，可以使用单个参数进行描述，即 θ_1 或 θ_2，具体取决于哪名玩家对策略动态有所贡献。因此，部分平凡博弈可以映射到部分平凡均衡不变嵌入流形。对于没有任何玩家参与的平凡博弈，则将其映射到一个由零博弈构成的单点集。

[11]定义为两个圆的乘积流形。从拓扑学上看，这是一个环面。其他分类（Robinson和 Goforth，2005年）也具有类似的环面拓扑。

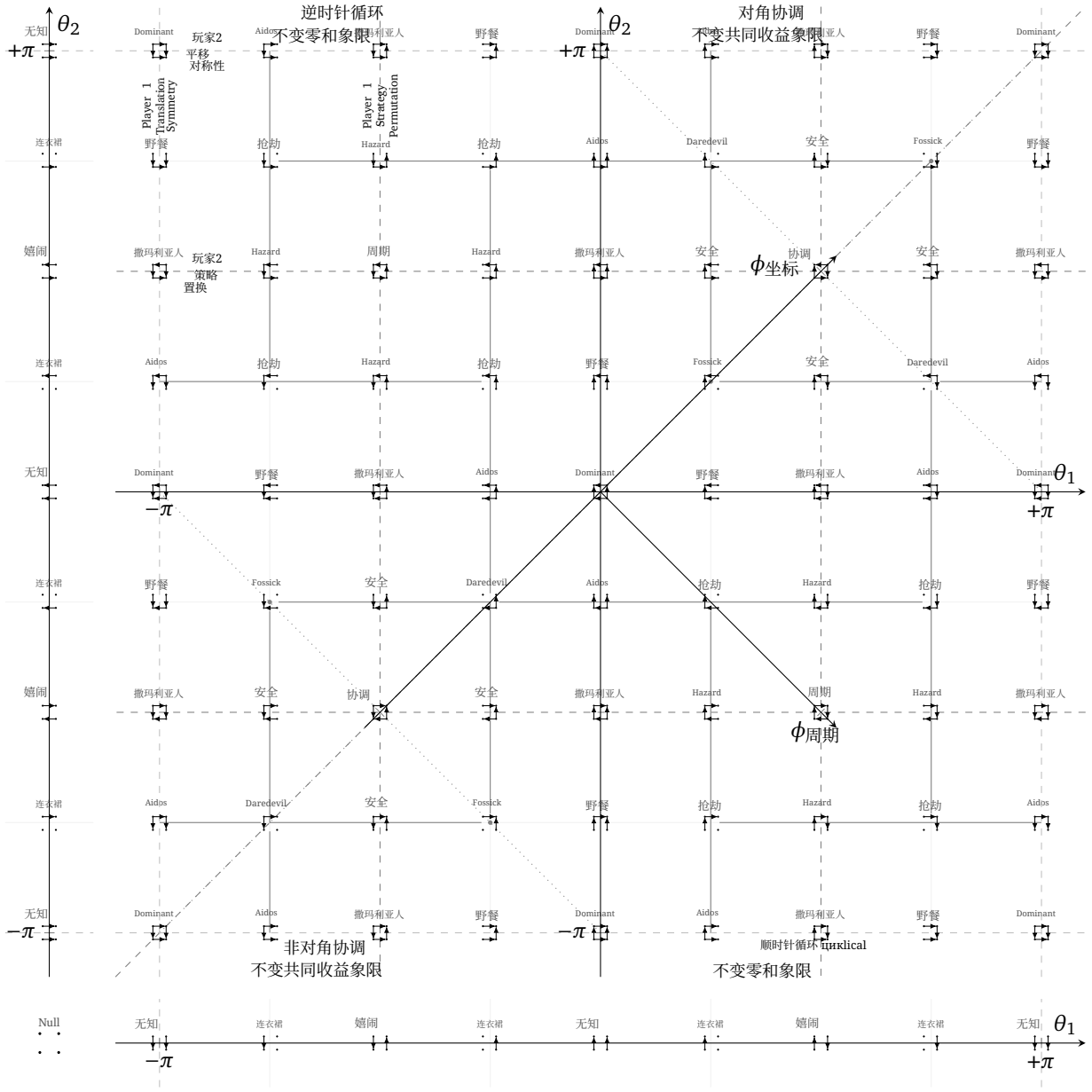


图 3 | 该 2×2均衡不变嵌入。博弈既可以按玩家进行参数化， θ_1 和 θ_2，也可以按博弈类型进行参数化，ϕ_{cycle} 和 ϕ_{coord}。对称性（表2）用虚线表示，根据所采用的对称性不同，可将区域面积最多缩减至原来的8分之一。策略置换均衡对称嵌入以西北方向的图案表示（）。策略与玩家置换均衡对称嵌入以东北方向的图案表示（）。带有零和假设的置换对称性以垂直方向的图案表示（）。各象限将空间划分为周期区域和协调区域。左上和右下象限为周期（Cycle）、不变零和区域。右上和左下象限为协调（Coordination）、不变共同收益区域。点线表示对称博弈，或策略经过置换的对称博弈。浅色实线表示等价类边界。灰色实线、点及其间的空间表示均衡支撑集的边界（另见图14）。Dominant、Picnic 和 Samaritan 具有纯均衡；方形区域为全支撑区域（周期型或协调型）；灰色线段边界具有两个（Aidos、Heist、Hazard）或三个（Safety、Daredevil）联合策略支撑集，而灰色点则表示对角或非对角支撑集（Fossick）。这11个非平凡最佳响应不变博弈集均以名称标注，并在名称后缀中注明均衡支撑集。Dominant 是唯一同时属于不变零和和不变共同收益的博弈，位于原点。 [注：某些 PDF 查看器无法完整呈现此图。]

4.2. 不变零和与不变共收益象限

双人零和博弈在文献中得到了深入研究，因为它们比混合动机博弈更容易求解。尤其对于纳什均衡（NE），该问题可以表述为一个极小极大优化问题。如果存在多个均衡，那么所有均衡的收益都相同，且它们可以互换（对手选择哪个均衡都无关紧要）。然而，这类博弈只占一般和博弈的一小部分。如果我们能通过不变变换，将更多博弈映射到零和博弈空间中，就能证明它们也具有这些有趣的性质。

文献中已经探讨过两种此类方法。首先，常和博弈是零和博弈经过标量偏移变换得到的，$G_p(a) \to \hat{G}_p(a) = G_p(a) + b_p$；众所周知，这类博弈具有零和博弈的性质。其次，严格竞争博弈（Adler 等，2009年）是零和博弈经过标量偏移和玩家尺度变换得到的，$G_p(a) \to \hat{G}_p(a) = s_p G_p(a) + b_p$。然而，还存在一个更大的博弈空间，其中的博弈都可以映射到零和博弈。

定义 11 (不变零和博弈)。 不变零和博弈是指那些可以通过均衡不变变换映射为零和博弈的博弈，$G_p(a) \to \hat{G}_p(a) = s_p G_p(a) + b_p(a_{-p})$ 使得 $\sum_p \hat{G}_p(a) = 0 \ \forall a \in \mathcal{A}$。

根据定义，不变零和博弈的均衡与其零和对应博弈的均衡相同。此外，变换后的博弈中的均衡仍可互换，但并非所有均衡都必然具有相同的收益。零和博弈的其他推广形式，包括 策略性零和(Moulin and Vial, 1978年）、最佳响应零和 (Rosenthal, 1974年)，以及序零和 (Shapley, 1963年，都包含不变零和博弈。可视化图中的对角象限（左上和右下）（图3）属于不变零和。

定理 7 (不变零和象限)。 当$\sin(\theta_1)\sin(\theta_2) < 0$ 满足方程(30a) 和(30b)时，相应的博弈($\hat{G}_1(\theta_1)$, $\hat{G}_2(\theta_2)$) 为不变零和。

证明。一个博弈要成为不变零和博弈，必须存在不变变换 $s_2 > 0$, $b_1(a_2)$，以及$b_2(a_1)$，使得 $G_1(a_1, a_2) + b_1(a_2) = -s_2 G_2(a_1, a_2) - b_2(a_1) \ \forall a_1, a_2$。考虑变换 $s_2 = -\sin(\theta_1)\sin(\theta_2)$, $b_1 = [-\frac{1}{2}\sin(\theta_1)\tan(\theta_2)$, $1\ 2\ \sin(\theta_1)\tan(\theta_2)]$，以及 $b_2 = [-\frac{1}{2}\cos(\theta_1), 1\ 2\ \cos(\theta_1)]$，，由此得到以下收益：

$$\hat{G}_1(\theta_1, \theta_2) = -\hat{G}_2(\theta_1, \theta_2) = \begin{bmatrix} \frac{1}{2}\frac{\sin(\theta_1)}{\tan(\theta_2)} + \frac{1}{\sqrt{2}}\sin(\theta_1 + \frac{\pi}{4}) & -\frac{1}{2}\frac{\sin(\theta_1)}{\tan(\theta_2)} + \frac{1}{\sqrt{2}}\sin(\theta_1 + \frac{\pi}{4}) \\ \frac{1}{2}\frac{\sin(\theta_1)}{\tan(\theta_2)} - \frac{1}{\sqrt{2}}\sin(\theta_1 + \frac{\pi}{4}) & -\frac{1}{2}\frac{\sin(\theta_1)}{\tan(\theta_2)} - \frac{1}{\sqrt{2}}\sin(\theta_1 + \frac{\pi}{4}) \end{bmatrix} \tag{37}$$

仅当 $s_2 > 0 \implies \sin(\theta_1)\sin(\theta_2) < 0 \implies \sin(\theta_1)\sin(\theta_2) < 0$时，这才是一个有效的变换。 □

还可以使用不变变换推导出一个类似的更大合作共收益博弈集合，$G_p(a) = G_q(a) \ \forall p, q \in [1, N]$。

定义 12 (不变共同收益博弈)。 不变共同收益博弈是指能够通过均衡不变变换映射为共同收益博弈的博弈，$G_p(a) \to \hat{G}_p(a) = s_p G_p(a) + b_p(a_{-p})$。

这些博弈恰好对应于 2×2加权势博弈（Monderer 和 Shapley, 1996），其中共同收益充当势函数。可视化中的非对角象限（右上和左下）为不变共同收益）。

定理 8 （不变共同收益象限）。 当正弦（θ_1）正弦（θ_2）> 0 满足方程（30a）和（30b）时，相应的博弈（$\hat{G}_1$（θ_1），$\hat{G}_2$（θ_2））是不变共同收益博弈。

证明。对于一个博弈要成为不变共同收益，必须存在不变变换 $s_2 > 0$、 $b_1(a_2)$，以及 $b_2(a_1)$，使得 $G_1(a_1, a_2) + b_1(a_2) = s_2 G_2(a_1, a_2) + b_2(a_1)$。考虑变换 $s_2 = \sin(\theta_1)\sin(\theta_2)$、 $b_1 = [\frac{1}{2}\sin(\theta_1)\tan(\theta_2)$、$-\frac{1}{2}\sin(\theta_1)\tan(\theta_2)]$，以及 $b_2 = [\frac{1}{2}\cos(\theta_1)$ 、 $-\frac{1}{2}\cos(\theta_1)]$，，其收益为：

$$\hat{G}_1(\theta_1, \theta_2) = \hat{G}_2(\theta_1, \theta_2) = \begin{bmatrix} \frac{1}{2}\frac{\sin(\theta_1)}{\tan(\theta_2)} + \frac{1}{\sqrt{2}}\sin(\theta_1 + \frac{\pi}{4}) & -\frac{1}{2}\frac{\sin(\theta_1)}{\tan(\theta_2)} + \frac{1}{\sqrt{2}}\sin(\theta_1 + \frac{\pi}{4}) \\ \frac{1}{2}\frac{\sin(\theta_1)}{\tan(\theta_2)} - \frac{1}{\sqrt{2}}\sin(\theta_1 + \frac{\pi}{4}) & -\frac{1}{2}\frac{\sin(\theta_1)}{\tan(\theta_2)} - \frac{1}{\sqrt{2}}\sin(\theta_1 + \frac{\pi}{4}) \end{bmatrix} \tag{38}$$

仅当 $s_2 > 0 \implies \underset{\text{正弦}(\theta_1)}{} \underset{\text{正弦}(\theta_2)}{} > 0 \implies \sin(\theta_1)\sin(\theta_2) > 0$ 时，这才是一种有效的变换。 □

原点处的博弈（我们将其命名为 Dominant）既是不变零和博弈，也是收益不变博弈。连同该博弈的置换版本，这是唯一具有这一性质的博弈。

定理 9 （Dominant 特殊情形）。 当 $\sin(\theta_1) = 0$ 且 $\sin(\theta_2) = 0$ 在方程 (30a)和 (30b)中成立时，相应的博弈 $(\hat{}G_1(\theta_1), \hat{}G_2(\theta_2))$ 既是不变零和博弈，也是不可变共同收益博弈。

证明。 考虑 $\theta_1 = \theta_2 = 0$的情况。当使用变换 $b_1 = [-\frac{1}{2}, \frac{1}{2}]$和 $b_2 = [-\frac{1}{2}, \frac{1}{2}]$时，博弈变为零和游戏（方程（39a））。此外，当使用变换 $b_1 = [\frac{1}{2}, -\frac{1}{2}]$和 $b_2 = [\frac{1}{2}, -\frac{1}{2}]$时，博弈变为共同收益博弈（方程-（39b））。

$$\hat{G}_1 = -\hat{G}_2 = \begin{bmatrix} 0 & \frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \tag{39a}$$

$$\hat{G}_1 = \hat{G}_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{39b}$$

□

因此，所有对称 2×2 博弈要么是不变零和、要么是不变共同收益，或两者兼具。各象限之间的边界既不是不变零和，也不是不变共同收益。Dominant是唯一同时具备不变零和和不变共同收益的博弈。

4.3. 2×2均衡不变距离度量

定理10 (2×2 均衡不变距离度量)。 两个 *2×2* 博弈之间的距离度量为：

$$d(\Theta^A, \Theta^B) = ||[\min(|\theta_1^A - \theta_1^B|, 2\pi - |\theta_1^A - \theta_1^B|), \min(|\theta_2^A - \theta_2^B|, 2\pi - |\theta_2^A - \theta_2^B|)]||_p \tag{40}$$

证明。 考虑两个参数化博弈 $\hat{}G_1^A(\theta_1^A)$和 $\hat{}G_1^B(\theta_1^B)$之间距离度量（公式（30a））中玩家1的预范数组成部分（公式（9））。

$$\begin{aligned} v_1 &= \arccos\left(\sin(\theta_1^A + \frac{\pi}{4})\sin(\theta_1^B + \frac{\pi}{4}) + \cos(\theta_1^A + \frac{\pi}{4})\cos(\theta_1^B + \frac{\pi}{4})\right) && (41a) \\ &= \text{反余弦}\ (\text{余弦}(\theta_1^A - \theta_1^B)) && (41b) \\ &= \min(|\theta_1^A - \theta_1^B|,\ \ 2\pi - |\theta_1^A - \theta_1^B|) && (41c) \end{aligned}$$

对玩家2也可以进行类似的计算。综合起来，得到方程（40）。 □

这种针对博弈的 2×2 均衡不变嵌入所定义的距离度量是自然的。正如定理6所确立的，具有相同均衡的博弈可以通过两个相互独立的单位圆上的点，或等价地通过它们的角度，即 $\Theta = (\theta_1, \theta_2)$，以简洁的方式表示为嵌入。沿单位圆的距离由弧长度量。因此，衡量两个博弈之间距离的一种自然方式，就是将它们在两个圆上对应嵌入之间的弧长相加。更一般地，我们可以考虑使用任意 p-范数来度量博弈表示之间的距离。总之，两个博弈之间的距离 d可以通过以下步骤求得：a）将博弈转换为均衡不变嵌入；b）计算两个嵌入中各对应分量之间的弧长；c）使用任意 p-范数计算距离，其中 $\Theta^A = (\theta_1^A, \theta_2^A)$和$\Theta^B = (\theta_1^B, \theta_2^B)$。

对于所有满足 $p < \infty$ 的范数，都存在一个博弈，它到其他任意博弈的距离均为最大。我们将这样的博弈称为相反博弈，它可以通过将表示平移 π 的常数来计算，方向为沿圆周。

$$\theta_1 \to \text{mod}(\theta_1 + 2\pi, 2\pi) - \pi \qquad \theta_2 \to \text{mod}(\theta_2 + 2\pi, 2\pi) - \pi \tag{42}$$

例如，顺时针周期博弈的对立博弈是逆时针周期博弈，而对角协调博弈的对立博弈是非对角（“反协调”）协调博弈。

对称性	变换描述	变换
玩家1策略置换	沿 θ_1平移，关于 $\theta_1 = 0$弃牌	$(\theta_1, \theta_2) \to (\theta_1 + \pi, -\theta_2)$
玩家2策略置换	在 $\theta_2 = 0$ 上弃牌，在 θ_2 上平移	$(\theta_1, \theta_2) \to (-\theta_1, \theta_2 + \pi)$
玩家置换	在 $\theta_1 = \theta_2$ 上弃牌	$(\theta_1, \theta_2) \to (\theta_2, \theta_1)$

表 2 | 用于推导均衡对称嵌入的标准式博弈中的 2×2 对称性。

算法 3： 均衡对称嵌入

数据： $\theta_1,\ \theta_2$

结果： θ^*_1, θ^*_2

$\theta^*_1,\ \theta^*_2 \leftarrow \theta_1,\ \theta_2$;

如果 $\theta^*_1 \le -\frac{\pi}{2}$ **或** $\frac{\pi}{2} < \theta^*_1$ **则**

  // Permute Player 1's Strategies

  $(\theta^*_1, \theta^*_2) \leftarrow (\theta^*_1 + \pi, -\theta^*_2)$;

如果 $\theta^*_2 \le -\frac{\pi}{2}$ **或** $\frac{\pi}{2} < \theta^*_2$ **则**

  // Permute Player 2's Strategies

  $(\theta^*_1, \theta^*_2) \leftarrow (-\theta^*_1, \theta^*_2 + \pi)$;

如果 $\theta^*_2 > \theta^*_1$ **那么**

  // Permute Players

  $(\theta^*_1, \theta^*_2) \leftarrow (\theta^*_2, \theta^*_1)$;

5. 2×2均衡对称嵌入

两种常见的对称性是策略置换和玩家置换。标准式博弈中策略和玩家的排列顺序是任意的，沿这些维度进行置换会得到策略上相同的博弈。考虑这些对称性可以缩小均衡不变嵌入的区域（表 2）。这个称为均衡对称嵌入的约化空间在可视化中以阴影标示（图3）。从均衡不变嵌入转换为均衡对称嵌入很容易（算法 3）。如果玩家的角色很重要，则只能使用策略置换，均衡对称嵌入的空间将缩小至原来的四分之一。通常可根据上下文判断所采用的对称性。

5.1. 推导 2×2 均衡对称嵌入

定理 11 (均衡对称嵌入)。 策略置换和玩家置换可以将均衡不变嵌入的面积缩小至原来的八分之一，但仅考虑规范排序的情况。

证明概要。 对称性（表 2），包括玩家置换以及每个玩家的策略置换，可用于缩减表示空间。每种缩减都会使表示区域缩小一半，将这些缩减组合起来，则可使表示区域缩小至原来的八分之一。 □

5.2. 2×2 均衡对称的玩家无关嵌入

我们可以定义均衡不变嵌入的一种新参数化，(ϕ周期,ϕ坐标)，它只是从 (θ_1, θ_2) 进行的一次基变换。在均衡对称嵌入中，$0 \le \phi$周期 ≤ 1 和 $-1 < \phi$坐标 ≤ 1，其中 ϕ周期 $\pm \phi$坐标 ≤ 1。这个基描述的是博弈的属性，而不是每位玩家的收益。从逐玩家基到博弈属性基的转换很简单。

$$\phi_{\text{coord}} = \frac{1}{\pi}(\theta_1 + \theta_2) \tag{43a}$$

$$\phi_{\text{cycle}} = \frac{1}{\pi}(\theta_1 - \theta_2) \tag{43b}$$

$$\theta_1 = \frac{\pi}{2}(\phi_{\text{coord}} + \phi_{\text{cycle}}) \tag{44a}$$

$$\theta_2 = \frac{\pi}{2}(\phi_{\text{coord}} - \phi_{\text{cycle}}) \tag{44b}$$

因此，博弈可以分别通过其协调性（“共同收益性”）和“循环性”（“零和性”）来描述。位于原点、协调性和循环性均为零的博弈具有传递性（即支配博弈）。

$$\hat{G}_1(\phi_{\text{coord}},\phi_{\text{cycle}})=\begin{bmatrix}\frac{1}{\sqrt{2}}\sin\left(\frac{\pi}{2}\left(\phi_{\text{coord}}+\phi_{\text{cycle}}\right)+\frac{\pi}{4}\right) & \frac{1}{\sqrt{2}}\cos\left(\frac{\pi}{2}\left(\phi_{\text{coord}}+\phi_{\text{cycle}}\right)+\frac{\pi}{4}\right)\\ -\frac{1}{\sqrt{2}}\sin\left(\frac{\pi}{2}\left(\phi_{\text{coord}}+\phi_{\text{cycle}}\right)+\frac{\pi}{4}\right) & -\frac{1}{\sqrt{2}}\cos\left(\frac{\pi}{2}\left(\phi_{\text{coord}}+\phi_{\text{cycle}}\right)+\frac{\pi}{4}\right)\end{bmatrix} \tag{45a}$$

$$\hat{G}_2(\phi_{\text{coord}},\phi_{\text{cycle}})=\begin{bmatrix}\frac{1}{\sqrt{2}}\sin\left(\frac{\pi}{2}\left(\phi_{\text{coord}}-\phi_{\text{cycle}}\right)+\frac{\pi}{4}\right) & -\frac{1}{\sqrt{2}}\sin\left(\frac{\pi}{2}\left(\phi_{\text{coord}}-\phi_{\text{cycle}}\right)+\frac{\pi}{4}\right)\\ \frac{1}{\sqrt{2}}\cos\left(\frac{\pi}{2}\left(\phi_{\text{coord}}-\phi_{\text{cycle}}\right)+\frac{\pi}{4}\right) & -\frac{1}{\sqrt{2}}\cos\left(\frac{\pi}{2}\left(\phi_{\text{coord}}-\phi_{\text{cycle}}\right)+\frac{\pi}{4}\right)\end{bmatrix} \tag{45b}$$

5.3. 2×2 均衡对称距离度量

对称嵌入中博弈之间的距离度量，也可以通过枚举所有（最多八个）等价的同构博弈，并求出最近一对之间的最小距离来定义。例如，对称嵌入中各点之间的部分距离见表4。

6. 2×2 最佳响应不变嵌入

利用较优响应不变嵌入（定义 7），我们可以推导出一组包含 15 个基本博弈的集合。对于双策略博弈，较优响应和最佳响应这两个概念是同义的。因此，我们将这组博弈称为 2×2最佳响应不变嵌入。与这些嵌入相对应的是 15 个等价类：许多均衡不变嵌入都映射到每一个最佳响应不变嵌入。这一区别微妙但十分重要。有时我们指的是最佳响应不变嵌入，有时指的是它们所属的等价类。最佳响应不变嵌入可以被视为每个等价类中博弈的规范（基本）示例。每个等价类中所有博弈的均衡（弱粗相关均衡、纳什均衡、WSCE、相关均衡和 CCE）均可通过对这些基本博弈的均衡进行简单缩放来计算。在这 15 个等价博弈类别中，11 个是非平凡的，3 个是部分平凡的，1 个是平凡的。这 15 个类别与 Borm (1987年) 提出的类别相同，并且对应于具有相同最佳响应动态的博弈。我们的工作改进了 Borm 的分类，因为在我们的推导中，最佳响应不变嵌入继承了均衡不变嵌入的距离度量。这使我们能够将这些博弈类别置于一个度量空间中，从而显著提高了嵌入的清晰度。

6.1. 推导最佳响应嵌入

定理 12 (2×2最佳响应嵌入)**.** *所有 2×2 博弈的收益都可以映射到一个由 81 个博弈组成的基本集合。经对称性约化后，该集合包含 11 个非平凡博弈、3 个部分平凡博弈和 1 个平凡博弈，共得到 15 个基本博弈。其中一种映射方式是：*

$$G_1=\begin{bmatrix}g_1^{AA} & g_1^{AB}\\ g_1^{BA} & g_1^{BB}\end{bmatrix}\to\left\{\begin{bmatrix}+1 & +1\\ -1 & -1\end{bmatrix},\begin{bmatrix}+1 & 0\\ -1 & 0\end{bmatrix},\begin{bmatrix}+1 & -1\\ -1 & +1\end{bmatrix},\begin{bmatrix}0 & +1\\ 0 & -1\end{bmatrix},\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix},\begin{bmatrix}0 & -1\\ 0 & +1\end{bmatrix},\begin{bmatrix}-1 & +1\\ +1 & -1\end{bmatrix},\begin{bmatrix}-1 & 0\\ +1 & 0\end{bmatrix},\begin{bmatrix}-1 & -1\\ +1 & +1\end{bmatrix}\right\} \tag{46a}$$

$$G_2=\begin{bmatrix}g_2^{AA} & g_2^{AB}\\ g_2^{BA} & g_2^{BB}\end{bmatrix}\to\left\{\begin{bmatrix}+1 & -1\\ +1 & -1\end{bmatrix},\begin{bmatrix}+1 & -1\\ 0 & 0\end{bmatrix},\begin{bmatrix}+1 & -1\\ -1 & +1\end{bmatrix},\begin{bmatrix}0 & 0\\ +1 & -1\end{bmatrix},\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix},\begin{bmatrix}0 & 0\\ -1 & +1\end{bmatrix},\begin{bmatrix}-1 & +1\\ +1 & -1\end{bmatrix},\begin{bmatrix}-1 & +1\\ 0 & 0\end{bmatrix},\begin{bmatrix}-1 & +1\\ -1 & +1\end{bmatrix}\right\}. \tag{46b}$$

证明。 使用逐策略尺度变换，并将 b_1 定义为通常的零均值偏移量和 $s_2=[|g_1^{A-B,A}|,|g_1^{A-B,B}|]$，我们可以推导出玩家1的玩家嵌入。

$$G_1=\begin{bmatrix}g_2^{AA} & g_2^{AB}\\ g_2^{BA} & g_2^{BB}\end{bmatrix}\to\begin{bmatrix}\frac{1}{2}g_1^{A-B,A} & \frac{1}{2}g_1^{A-B,B}\\ -\frac{1}{2}g_1^{A-B,A} & -\frac{1}{2}g_1^{A-B,B}\end{bmatrix}\to\begin{bmatrix}\frac{g_1^{A-B,A}}{|g_1^{A-B,A}|} & \frac{g_1^{A-B,B}}{|g_1^{A-B,B}|}\\ -\frac{g_1^{A-B,A}}{|g_1^{A-B,A}|} & -\frac{g_1^{A-B,B}}{|g_1^{A-B,B}|}\end{bmatrix} \tag{47a}$$

$$=\begin{bmatrix}\{+1,+1,+1,0,0,0,-1,-1,-1\} & \{+1,0,-1,+1,0,-1,+1,0,-1\}\\ \{-1,-1,-1,0,0,0,+1,+1,+1\} & \{-1,0,+1,-1,0,+1,-1,0,+1\}\end{bmatrix} \tag{47b}$$

将一个数除以其绝对值可得到它的符号，并采用约定（这不会造成影响） $\frac{0}{|0|} = 0$，由此得到 9 种收益可能性，其中一种是平凡的零收益。我们可以对玩家2进行类似的推导。因此，总共有 $9 \times 9 = 81$ 种博弈。其中，64 种是非平凡的，16 种是部分平凡的，1 种是平凡的。前面讨论的置换对称性有助于我们发现，在对称等价关系下，这些博弈中只有 15 种是不同的。□

<escape></escape>最佳响应不变嵌入集也可以使用二元收益来定义（<escape></escape>Fishburn 和 Kilgour<escape></escape>，<escape></escape>1990年<escape></escape>），因为我们只需要确定各策略之间的偏好。然而，我们选择使用三元收益（<escape>−1<escape></escape>，<escape></escape>0<escape></escape>，<escape></escape>+1<escape></escape>），因为这样可以更清晰地区分三种情况：偏好策略<escape> A<escape></escape>、偏好策略<escape> B<escape></escape>，或对两种策略无差异。这种偏好排序等价于每个玩家的最佳响应动态。对于双策略博弈，较优响应动态与最佳响应动态完全相同。最佳响应不变性与用于推导均衡不变嵌入的仿射变换密切相关（<escape></escape>Morris 和 Ui<escape></escape>，<escape></escape>2004年<escape></escape>；<escape></escape>Ostrovski<escape></escape>，<escape></escape>2013年<escape></escape>）。均衡不变性意味着较优响应不变性，而较优响应不变性又意味着最佳响应不变性。</escape>

6.2. 等价类、图表示与测度

最佳响应不变嵌入的等价类等同于考察玩家1收益中每一列的排序，以及玩家2收益中每一行的排序（最佳响应动态）。按照惯例，令左上角的联合策略为 AA，右下角为 BB，令玩家1为行玩家，玩家2为列玩家。这种排序可以简单地用有向图表示：玩家1对列的排序可以用竖直箭头表示（例如，表示玩家1在玩家2选择 A 时更偏好策略 B，而在玩家2选择 B时更偏好策略 A），玩家2对行的排序可以用水平箭头表示（例如 ）。没有偏好时，无需绘制边，。综合起来，我们可以将任何2×2 博弈可视化。例如，囚徒困境（图 1b）表示为 ，胆小鬼博弈（图 1a）表示为 ，匹配硬币（图1e）表示为 。这些有向图在图同构意义下共有 15 个等价类。




最佳响应不变嵌入中的大多数博弈，在均衡不变嵌入中都具有零测度等价类，这意味着随机抽样到这些博弈的概率为零。判断一个等价类的测度有一个简单的经验法则：a) 如果所有玩家在进行最佳响应时都不存在无差异情况，则该博弈具有正测度；b) 如果一名玩家在对某一行动进行最佳响应时处于无差异状态，则该类为零测度，并且是出现在两个正测度类之间的一个“边界类”；c) 如果两名玩家都对对手的某一行动处于无差异状态，则该类为一个出现在四个边界类之间的“点类”。通过目视检查（图 14），可以很容易地发现，只有4种博弈具有非零的抽样概率：撒玛利亚人、Dominant、协调和周期（表3）。这为零测度博弈为何很少出现在文献中提供了直观解释。我们的观察结果与 Borm (1987年) 所做的类似分析一致。

6.3. 最佳响应不变嵌入的命名

Borm (1987年）为 15 种基本博弈类别提供了数值分类。我们已经指定了参数化嵌入 $(\theta_1,\ \theta_2)$ 或$(\phi_{\text{坐标}},\ \phi_{\text{周期}})$以及图形表示。此外，这个集合规模足够小，适合采用标准化的命名方案。我们尽可能遵循既有的命名惯例，并从先前的工作中汲取灵感（Bruns，2015年）。对称博弈（即位于 $\theta_1 = \theta_2$ 或$\phi_{\text{周期}} = 0$上的博弈）研究得最为充分，也已有通用名称。只有匹配硬币和撒玛利亚人困境等少数非对称博弈得到过研究。我们谨慎避免以相似但略有不同的博弈为类别命名（例如，刻意避开了鹿猎博弈和胆小鬼博弈）。大多数博弈都存在玩家无差异的策略。这些博弈类别鲜少得到研究，因此没有通用名称。表3 列出了本研究为提出的 15 种基本博弈选定的名称。

Dominant 在这个对称博弈（图 4a）中，每个玩家都有一个严格占优策略。因此，该博弈只有一个纯策略纳什均衡和一个（相关）相关均衡。Dominant 嵌入位于均衡不变嵌入的原点处，与其他拓扑（Rapoport 等人，1976年）一样，并且是唯一一个同时满足 2×2 的博弈

博弈名称	N	PN	MN	C	B	θ	ϕ	S、M、T、I、其他名称
Dominant	Do	■		■	5	$0,\ 0$	$0,\ 0$	✓ $\frac{1}{4}$ N 0 囚徒困境
协调 Co		■ ■	■	■	14	$\frac{\pi}{2},\ \frac{\pi}{2}$	$1,\ 0$	✓ $\frac{1}{8}$ N 0 对战 , Hunt, 胆小鬼博弈
周期	Cy		■	■	15	$\frac{\pi}{2},-\frac{\pi}{2}$	$0,\ 1$	$\frac{1}{8}$ N 0 匹配硬币
撒玛利亚人	Sm	■		■	8	$\frac{\pi}{2},\ 0$	$\frac{1}{2},\ \frac{1}{2}$	$\frac{1}{2}$ N 0
Hazard	Hz	■	■	■	13	$\frac{\pi}{2},-\frac{\pi}{4}$	$\frac{1}{4},\ \frac{3}{4}$	– N 1
安全	Sf	■ ■	■	■	12	$\frac{\pi}{2},\ \frac{\pi}{4}$	$\frac{3}{4},\ \frac{1}{4}$	– N 1
Aidos	Ad	■ ■	■	■	7	$0,-\frac{\pi}{4}$	$-\frac{1}{4},-\frac{1}{4}$	– N 1
野餐	Pn	■		■	6	$\frac{\pi}{2},\ 0$	$\frac{1}{4},\ \frac{1}{4}$	– N 1
Daredevil	Dd	■ ■ ■	■ ■	■	11	$-\frac{\pi}{4},-\frac{\pi}{4}$	$-\frac{1}{2},\ 0$	✓ · N 2
Fossick	Fo	■		■	9	$\frac{\pi}{4},\ \frac{\pi}{4}$	$\frac{1}{2},\ 0$	✓ · N 2
抢劫	Hs	■ ■	■	■	10	$\frac{\pi}{4},-\frac{\pi}{4}$	$0,\ \frac{1}{2}$	· N 2
无知	Ig	■ ■	■	■	2			$\frac{1}{2}$ P 2
嬉闹	Hp	■ ■	■ ■	■	4			$\frac{1}{2}$ P 2
连衣裙	Dr	■ ■ ■	■ ■	■	3			– P 3 红色连衣裙
Null	Nu	■ ■ ■ ■	■	■	1			✓ 1 T 4 平凡 , Zero

表 3 | 最佳响应不变嵌入集合的命名方案及其性质 2×2 。目前只有 Dominant、协调、周期、撒玛利亚人和零博弈得到充分研究并有既定名称。其余博弈（存在无差异情形）要么完全未见于文献，要么鲜少出现。图例：图表示（G）、简称（N）、纯策略 Nash 均衡（PN）、混合策略 Nash 均衡（MN）、相关均衡支撑集（C）、Borm（1987年）的类别（B）、对称性（S）、测度（M）、平凡性（T）和无差异（I）。

	S	W
S	$+1,+1$	$+1,-1$
W	$-1,+1$	$-1,-1$

(a) Dominant

	A	B
A	$+1,+1$	$-1,-1$
B	$-1,-1$	$+1,+1$

(b) 协调

	H	T
H	$+1,-1$	$-1,+1$
T	$-1,+1$	$+1,-1$

(c) 周期

	H	N
R	$4,3$	$1,1$
W	$3,4$	$2,2$

(d) 撒玛利亚人

	R	C
N	$0,+1$	$0,+1$
I	$-2,+3$	$+2,+2$

(e) Hazard

	C	R
I	$+1,+4$	$-2,+3$
N	$0,+1$	$0,+1$

(f) 安全

	B	N
N	$+1,\ 0$	$+1,\ 0$
R	$-1,+1$	$-1,-1$

(g) Aidos

	P	T
A	$+1,+1$	$0,-1$
D	$-1,+1$	$0,-1$

(h) 野餐

	B	N
N	$0,\ 0$	$+1,\ 0$
R	$0,+1$	$-1,-1$

(i) Daredevil

	W	G
W	$+1,\ 0$	$+1,\ 0$
G	$-1,+1$	$-1,-1$

(j) Fossick

	P	R
N	$+1,\ 0$	$0,\ 0$
H	$-1,+1$	$0,-1$

(k) 抢劫

	A	B
W	$+1,\ 0$	$+1,\ 0$
L	$-1,\ 0$	$-1,\ 0$

(l) 无知

	B	N
N	$+1,\ 0$	$+1,\ 0$
R	$-1,+1$	$-1,-1$

(m) 嬉闹

	C	F
C	$-2,\ 0$	$+1,\ 0$
F	$-2,+1$	$-1,+1$

(n) 连衣裙

	A	B
A	$0,\ 0$	$0,\ 0$
B	$0,\ 0$	$0,\ 0$

(o) Null

图 4 | 15 个最佳响应不变博弈的叙事型收益表。

不变零和和不变共同收益。Dominant 的等价类中包含许多常见博弈，包括囚徒困境、和平博弈、僵局博弈、完全冲突博弈、和谐博弈和妥协博弈。尤其是，囚徒困境是经济学中研究最多的博弈之一。对均衡不变嵌入进行均匀抽样时，Dominant 等价类出现的概率为 $\frac{1}{4}$。此外，这个等价类最多样：其中的博弈可以是逆时针不变零和、顺时针不变零和、对角线不变共同收益、非对角线不变共同收益，或以上皆非。

协调 协调（图4b）是一种对称共同收益协调博弈，其中有两个同样理想的结果。玩家只需协调行动，确保实现其中一个结果。协调博弈有两个收益相同的纯策略纳什均衡（和），以及一个收益较低的混合纳什均衡。它是唯一具有正体积（相关）粗相关均衡的非平凡博弈。协调博弈的等价类中包含许多博弈，包括鹿猎博弈、胆小鬼博弈、保证博弈，以及巴赫或斯特拉文斯基。对均衡不变嵌入进行均匀抽样时，出现协调博弈等价类的概率为$\frac{1}{8}$。协调博弈等价类中的所有博弈都是不变共同收益博弈。

周期 这种博弈是反对称零和博弈（图4c），通常也称为匹配硬币。在这种博弈中，第一位玩家希望双方采取相同的策略（例如都选正面或都选反面），而第二位玩家则希望双方采取不同的策略。纯策略的最佳响应会导致循环动态。这种博弈有一个类似的三策略变体，具有相同的性质，即石头剪刀布，这是一种广受欢迎的儿童游戏。置换某位玩家的策略，会使博弈从顺时针周期变为逆时针周期。周期博弈没有纯策略纳什均衡，只有一个完全混合纳什均衡（概率均匀），以及（相关）粗相关均衡。周期博弈是唯一所有均衡都严格位于单纯形内部的博弈。周期博弈的最佳响应不变等价类中的所有博弈也都是循环的，但对其偏好策略的偏向不同。对均衡不变嵌入进行均匀抽样时，该等价类出现的概率为$\frac{1}{8}$。

撒玛利亚人 这个博弈（图 4d）有一个工人玩家和一个撒玛利亚人玩家。工人有两种策略：休息（R）或工作（W）。撒玛利亚人有两种策略：帮助（H）或不帮助（N）。撒玛利亚人总是更愿意提供帮助，但也更希望工人工作（而不是搭便车）。如果得到帮助，工人更愿意不工作；如果得不到帮助，工人则更愿意工作以维持生计。撒玛利亚人博弈只有一个纯策略纳什均衡和（相关）相关均衡。它得名于撒玛利亚人困境（Schmidtchen，2002年），这是 Buchanan （1975）提出的一个博弈。Samson（Brams，1993年）、Alibi（Robinson 和 Goforth，2005年）、Anticipation、Bully、Hamlet、非对称困境和 Called Bluff 都属于这个博弈的更优响应不变等价类。撒玛利亚人嵌入既非零和不变，也非共同收益不变。不过，其等价类中的其他博弈可能是零和不变、共同收益不变，或两者皆非。撒玛利亚人等价类很特别，因为在均衡不变嵌入上进行均匀抽样时，它出现的概率最高，为$\frac{1}{2}$。除了是最常见的等价类之外，它也是与所有其他博弈联系最紧密的等价类（第 6.4节）。

Hazard 在 Hazard（图4e）中，玩家1是一家保险公司，可以选择为玩家2提供健康保险（I），也可以拒绝承保（D）。玩家2喜欢在派对上喝酒，这属于风险行为（R），但也看重不喝酒带来的健康益处，这属于谨慎行为（C）。总体而言，在没有健康保险的情况下，玩家2对派对生活方式和健康生活方式并无偏好。然而，如果玩家1提供医疗保险，玩家2的风险容忍度就会提高，并会选择派对生活方式。如果不提供保险，玩家1不会赚钱；如果提供保险且玩家2选择健康生活方式，玩家1就会赚钱；如果为生活方式不健康的玩家2提供保险，玩家1则会亏钱。Bruns （2015年）简要定义了一个他称为道德风险的博弈，并给它起了一个公式化的名字：“Middle Hunt × Low Dilemma”。Hazard 有一个纯策略纳什均衡、一个混合纳什均衡，以及支撑集为 的（相关）粗相关均衡。Hazard 等价类存在无差异，因此在均衡不变嵌入中是零测度的。不过，它是一个边界类，与 Cycle 等价类相邻。在这些边界类中，若在均衡不变嵌入上进行均匀抽样，抽中它的概率为$\frac{1}{4}$。该等价类中的所有博弈都是不变零和的。

安全 该博弈（图 4f）也曾被 Bruns （2015年）提及，并称其为“中等狩猎 × 低和谐博弈”。它与 Hazard相似，不同之处在于，保险人现在通过提供免费健身房会员资格，激励玩家1采取健康生活方式。这解决了道德风险问题，如今玩家2也会因保险而受到激励，采取健康生活方式。安全博弈有两个纯策略纳什均衡（ 和 ）、一个混合纳什均衡以及（相关）粗相关均衡 。安全等价类具有无差异性，因此在均衡不变嵌入中是零测度的。然而，它是一个与协调等价类相邻的边界类。在这些边界类别中，从均衡不变嵌入中均匀抽样时，它出现的概率为$\frac{1}{4}$。该等价类中的所有博弈都是不变共同收益博弈。

Aidos 在 Aidos（图 4g）中，玩家1是一位神祇，可以选择显露自己的身份（R）或不显露（N），然而其性格羞怯，因此严格偏好不显露身份。玩家2是一名人类，希望相信事实。在缺乏任何一方的证据时，人类对神祇是否存在持无差异态度；然而，如果神祇显露身份，人类则更倾向于相信（B），而非不相信（N）。该博弈存在两个纯策略纳什均衡（ 和 ），它们的任意混合也都是混合纳什均衡和（相关）相关均衡。该博弈的主题灵感来源于 Revelation （ Brams，1993 年），这是一个具有类似动态特征的博弈。Aidos 等价类包含一个无差异关系，因此在均衡不变嵌入中具有零测度。然而，它是一个与之相邻的边界类

纯均衡策略发生变化的区域中的 Dominant 和撒玛利亚人等价类（与下文将介绍的野餐不同）。在边界类别中，若在均衡不变嵌入上进行均匀抽样，则其出现的概率为 1 4 。Aidos 等价类中的博弈可以是零和不变的、共同收益不变的，也可以两者皆非。该集合中唯一既非零和不变也非共同收益不变的博弈是 Aidos 嵌入。

野餐 在这个博弈中（图 4h），主人可以组织野餐（P）或订购外带餐（T），且严格偏好组织野餐。客人可以参加（A）或婉拒（D）。订购外带餐时，客人对是否参加无差异；但如果组织野餐，客人则更愿意参加。野餐只有一个纯策略纳什均衡。野餐等价类存在无差异情形，因此在均衡不变嵌入中是零测度的。不过，它是一个边界类，与纯均衡策略不变区域中的 Dominant 和撒玛利亚人等价类相邻（与 Aidos 不同）。在这些边界类中，若在均衡不变嵌入上均匀抽样，该类出现的概率为 1 4 。野餐嵌入既非零和不变，也非共同收益不变；不过，其等价类中的博弈可以是零和不变的，也可以是共同收益不变的。

Daredevil 在相关文献中，胆小鬼博弈（图 1a）有两个纯策略反协调纳什均衡（ 和 ）以及一个混合纳什均衡，该均衡将大部分概率质量置于最差的联合结果上。混合纳什均衡的存在意味着胆小鬼博弈具有全支持均衡，因此属于协调的等价类。然而，还有另一种有趣的对称、不变共同收益、零测度等价类博弈——Daredevil（图 4i），其性质与胆小鬼博弈相似。该博弈有相同的两个纯策略反协调纳什均衡（ 和 ），以及一个额外的纯策略纳什均衡：撞毁。它没有完全混合纳什均衡，但也有两个纯混合纳什均衡（ 和 ）。此外，它还具有（相关）相关均衡，这些均衡可以在任意纯策略纳什均衡之间任意混合。该博弈与胆小鬼博弈的不同之处在于，如果另一位玩家选择不转向，那么无论转向以避免损害，还是继续前进而撞毁，该玩家都无差异。它的动态与胆小鬼博弈相似，也可以视为胆小鬼博弈的一个极端边界情形。据我们所知，此前的文献尚未研究过这种博弈——考虑到它属于零测度等价类，这或许并不令人意外。Daredevil 属于点类，因为 Daredevil 嵌入是其等价类中唯一的博弈。它具有不变共同收益，并与 Aidos 和安全等价类相邻。在三个点类中，它出现的概率为 1 4。

Fossick 在文献中，鹿猎博弈（图 1c）是一种具有两个纯策略协调纳什均衡（ 和 ）的博弈，其中一个均衡优于另一个均衡，并且还具有一个完全混合纳什均衡 。鹿猎博弈也属于协调博弈的等价类。Fossick（图 4j）是鹿猎博弈的一种极端形式，它具有两个纯策略纳什均衡（ 和 ），但不存在混合纳什均衡。它还具有（相关）粗相关均衡，这些均衡只能在协调策略 上进行混合。在这个博弈中，玩家正处于一场淘金热中，他们可以选择寻找水源以维持自己在荒野中的生存，也可以选择淘金。如果另一位玩家寻找水源，那么该玩家对自己选择寻找什么无差异。然而，如果另一位玩家选择淘金，该玩家就会感到自己被冷落，并更愿意也选择淘金。据我们所知，这种博弈此前尚未在文献中得到研究。Fossick 属于一个零测度等价类，同时也是一个点类（Fossick 嵌入是该等价类中唯一的博弈）。在三个点类中，它出现的概率为 1 4。Fossick 具有不变共同收益性质。

抢劫 在抢劫博弈（图 4k）中，玩家1是一名紧张的劫匪，可以选择什么也不做（N），或实施抢劫（H）。玩家2是一名保安，可以选择巡逻（P）或休息（R）。如果劫匪不实施抢劫，保安对巡逻还是休息无差异。然而，如果发生抢劫，保安更倾向于巡逻。如果保安在休息，劫匪不确定抢劫是否值得冒险，因此无差异。然而，如果保安在巡逻，劫匪更倾向于不实施抢劫。这个博弈与追捕博弈（Brams，1993年）最为相似。抢劫博弈有两个纯策略纳什均衡（和），以及在两者之间混合的（相关）粗相关均衡。抢劫博弈属于一个零测度等价类，同时也是个点类。在三个点类中，抢劫博弈最为常见，出现概率为1 2。抢劫博弈是不变零和的。

无知无知（图4l）是一种部分平凡博弈，其中一位玩家完全没有偏好。在无知博弈中，知情玩家希望获胜（W）并避免失败（L），而无知玩家对其策略和结果无差异。无知博弈有两个纯策略纳什均衡（和），一个混合纳什均衡

以及（相关） 相关均衡。对部分平凡的均衡不变嵌入进行均匀抽样时，无知博弈的等价类出现的概率为1 2 。

嬉闹 嬉闹（图 4m）是一种部分平凡博弈，其中一位玩家完全没有偏好。嬉闹是成人和孩子之间的游戏。孩子可以举起双臂（A），也可以蜷缩成一团（B）；无论选择哪一种，也无论结果如何，孩子都无差异，只是乐于参与游戏。为方便起见，如果孩子举起双臂，成人更喜欢把孩子抛到空中（T）；如果孩子蜷缩成一团，成人更喜欢抓住孩子的脚把孩子抱起来（L）。嬉闹有两个纯策略纳什均衡（ 和 ）、三个混合纳什均衡（ 、 和 ），以及完全支撑集的（相关）粗相关均衡。均匀地从部分平凡的均衡不变嵌入中抽样时，嬉闹的等价类出现的概率为1 2 。

连衣裙 由 Simpson (2010年) 描述，连衣裙（图 4n）讲述的是两个人约会时的情形。玩家1自私地希望伴侣穿正式服装（F），只有在伴侣这样穿时，才会懒洋洋地更喜欢自己穿舒适的衣服（C）。玩家2也希望伴侣穿正式服装，但对自己穿什么无差异。由于玩家2的偏好完全取决于另一位玩家的策略，其收益是平凡的，因此连衣裙是一种部分平凡博弈。这个博弈有三个纯策略纳什均衡，以及这三个均衡的任意混合（ 、 和）两个混合纳什均衡（ 和 ）以及（相关）粗相关均衡 连衣裙的等价类是无知与嬉闹之间的一个边界类。在部分平凡的均衡不变嵌入中均匀抽样时，它的测度为零。

零博弈在零博弈中（图4o），玩家对策略没有偏好，这意味着任何纯策略或混合策略都是纳什均衡，任何联合分布都是（相关）相关均衡。形式为 $G_1(a_1,\ a_2) = b_1(a_2), G_2(a_1,\ a_2) = b_2(a_1)$ 的博弈属于零博弈等价类。零博弈是一个点博弈，也是平凡博弈中唯一的类。这种博弈有时称为零和游戏或平凡博弈。

6.4. 距离度量

最佳响应不变嵌入之间的距离可以使用均衡对称距离度量进行计算（表4）。周期是最孤立的博弈，其到其他博弈的平均距离最大。撒玛利亚人是最不孤立的博弈，其到其他博弈的平均距离最小。没有任何博弈距离撒玛利亚人超过两步[12] 。在部分平凡博弈中，连衣裙是无知与嬉闹之间的边界类。收益的微小扰动可能改变博弈，并导致均衡发生阶跃变化。距离度量（以及拓扑）展示了微小扰动可能导致转变的相邻博弈。此类分析有助于选择对置换稳健的均衡，或处理收益的不确定性。我们将在讨论部分进一步探讨这一点。

7. 博弈可视化

2×2 嵌入可用于生成以下内容的可视化图示： 2×2， $|\mathcal{A}_1| \times |\mathcal{A}_2|$，双人扩展式、n人多矩阵以及n人扩展式博弈（图5）。均衡和收益属性可以直接从这些可视化图示中推导出来。

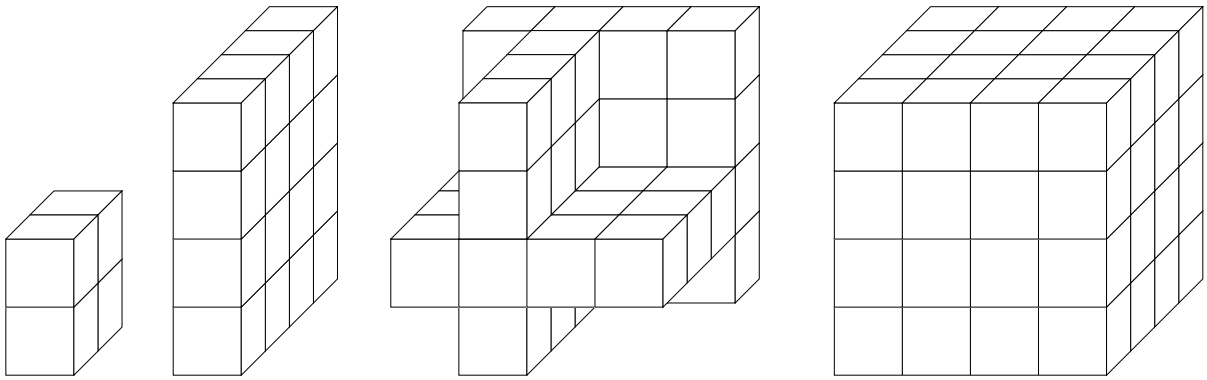
7.1. 2×2 博弈可视化

首先，考虑联合策略空间 $\sigma(a)$，它可以用一个扁平向量 $\boldsymbol{\sigma} = [\sigma(a^{AA}),\ \sigma(a^{AB}),\ \sigma(a^{BA}),\ \sigma(a^{BB}))]$, 表示，其中 $a^{IJ} = (a_1^I,\ a_2^J)$。概率分布的标准约束同样适用：概率非负，即 $\sigma(a) \geq 0\ \forall a \in \mathrm{A}$，且总和为 1，即 $\sum_{a\in\mathcal{A}} \sigma(a) = 1$。总和为 1 的约束意味着，对于一个 2×2 博弈的四种策略，可以仅用三个变量来表示其分布，因为在给定其余变量的情况下，其中一个变量是冗余的（例如 $\sigma(a^{BB}) = 1 - \sigma(a^{AA}) - \sigma(a^{AB}) - \sigma(a^{BA})$ ）。这意味着，通过忽略总和不为 1 的分布空间，可以仅在三维空间中可视化具有四个分量的联合分布。通常，这是通过指定一个四面体的四个顶点来实现的（三个】【。

[12]步数或“跳数”是按照图3所示的平面网格来测量的。

	Dominant	Coordination	Cycle	Samaritan	Hazard	Safety	Aidos	Picnic	Daredevil	Fossick	Heist	Ignorance	Horseplay	Dress	Null
Dominant	0	4	4	2	3	3	1	1	2	2	2				
协调	4	0	4	2	3	1	3	3	2	2	4				
周期	4	4	0	2	1	3	3	3	4	4	2				
撒玛利亚人	2	2	2	0	1	1	1	1	2	2	2				
Hazard	3	3	1	1	0	2	2	2	3	3	1				
安全	3	1	3	1	2	0	2	2	1	1	3				
Aidos	1	3	3	1	2	2	0	2	1	3	1				
野餐	1	3	3	1	2	2	2	0	3	1	1				
冒险者	2	2	4	2	3	1	1	3	0	2	2				
Fossick	2	2	4	2	3	1	3	1	2	0	2				
抢劫	2	4	2	2	1	3	1	1	2	2	0				
无知												0	2	1	
嬉闹												2	0	1	
连衣裙												1	1	0	
Zero															0

表 4 | 表：展示最佳响应不变嵌入中博弈之间的 L_1 距离。非平凡博弈与平凡博弈之间的距离未定义。



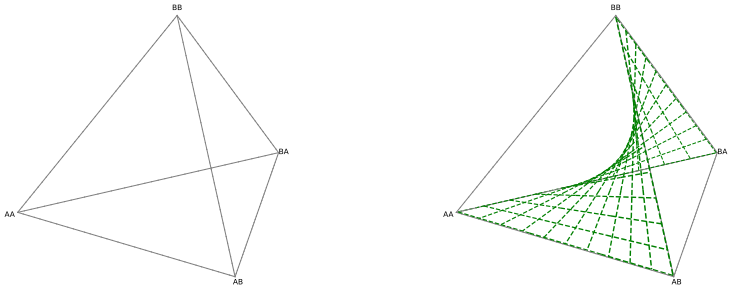
(a) (b) (c) (d)

图 5 | 可使用本文中的技术进行可视化的不同博弈示例：(a) 2×2博弈，(b) 双人博弈，(c) 多矩阵博弈。每个立方体代表博弈中的一种联合策略。超过两名玩家的标准式博弈 (d) 无法直接使用所述技术进行可视化，必须使用局部多矩阵近似，在某个联合策略附近进行近似。

维对象）。单纯形中的点随后用这四个顶点的混合来描述（称为重心坐标系 Möbius (1827)）。重心坐标 $\boldsymbol{\sigma}$，可通过线性变换 $\boldsymbol{x}$转换为笛卡尔坐标 $\boldsymbol{x} = T\boldsymbol{\sigma}$。矩阵的列 T 是正四面体的顶点。选择四面体顶点的方法有很多，其中一种见公式（48）。

$$\boldsymbol{x} = T\boldsymbol{\sigma} \qquad T = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & 0 & 0 \\ -\frac{\sqrt{3}}{6} & -\frac{\sqrt{3}}{6} & \frac{\sqrt{3}}{6} & 0 \\ -\frac{\sqrt{6}}{12} & -\frac{\sqrt{6}}{12} & -\frac{\sqrt{6}}{12} & \frac{\sqrt{6}}{4} \end{bmatrix} \tag{48}$$

每个非负不等式约束在平面方程中几何上对应一个法向量（例如 $[1,0,0,0]\boldsymbol{\sigma}^T \geq \mathbf{0}$）。这些平面将空间分成两半：其中可行的



(a) 概率单纯形　　(b) 纳什均衡流形

图 6 | 有效联合分布空间 $\sigma(a_1, a_2)$（四面体）与有效可分解联合分布空间 $\sigma(a_1, a_2) = \sigma(a_1)\sigma(a_2)$（流形）的可视化。四面体的顶点对应纯联合策略。四面体的内部对应混合联合策略。

分布以及不可行分布。这四个非负概率不等式约束共同形成一个具有四个面的凸多胞体，具体而言，在三维空间中可视化时，它是一个正四面体（图 6a）。 $\sigma(a_1, a_2)$ 对应于一个完整的联合分布。值得特别关注的是一类能够分解为其边缘分布的联合分布子集， $\sigma(a_1, a_2) = \sigma(a_1)\sigma(a_2)$，因为它们与纳什均衡（NEs）存在密切关系，而根据定义，纳什均衡必须能够分解。可因子化联合分布在四面体内部形成一个流形（图6b）。

现在考虑玩家收益 $G_p(a_1, a_2)$。我们知道，每位玩家的 2×2 均衡不变嵌入由圆上的一个角度参数化： $G\text{equil}_p(\theta_p)$。这个圆可以在三维空间中有意义地描绘出来： $x(\theta_p) = Tg\text{equil}_p(\theta_p)$。玩家的特定收益 $G\text{equil}_p$，可以用一支从原点指向圆上一点的箭头表示。这支箭头的方向具有明确含义：它是使联合策略下获得的均衡不变嵌入收益线性增加的方向，并指向均衡可能存在的区域。原始收益 $Tg_p(\theta_p)$ 可以是垂直于圆所在平面的任意向量。

各种均衡概念的偏离收益（方程（1）、（2）和（3））也可以进行可视化。偏离收益矩阵的每一行都对应一个半平面。所有行合在一起构成一个凸多胞体，也可以在三维空间中进行可视化。纳什均衡位于该多胞体与可因子分解的联合流形的交点处。

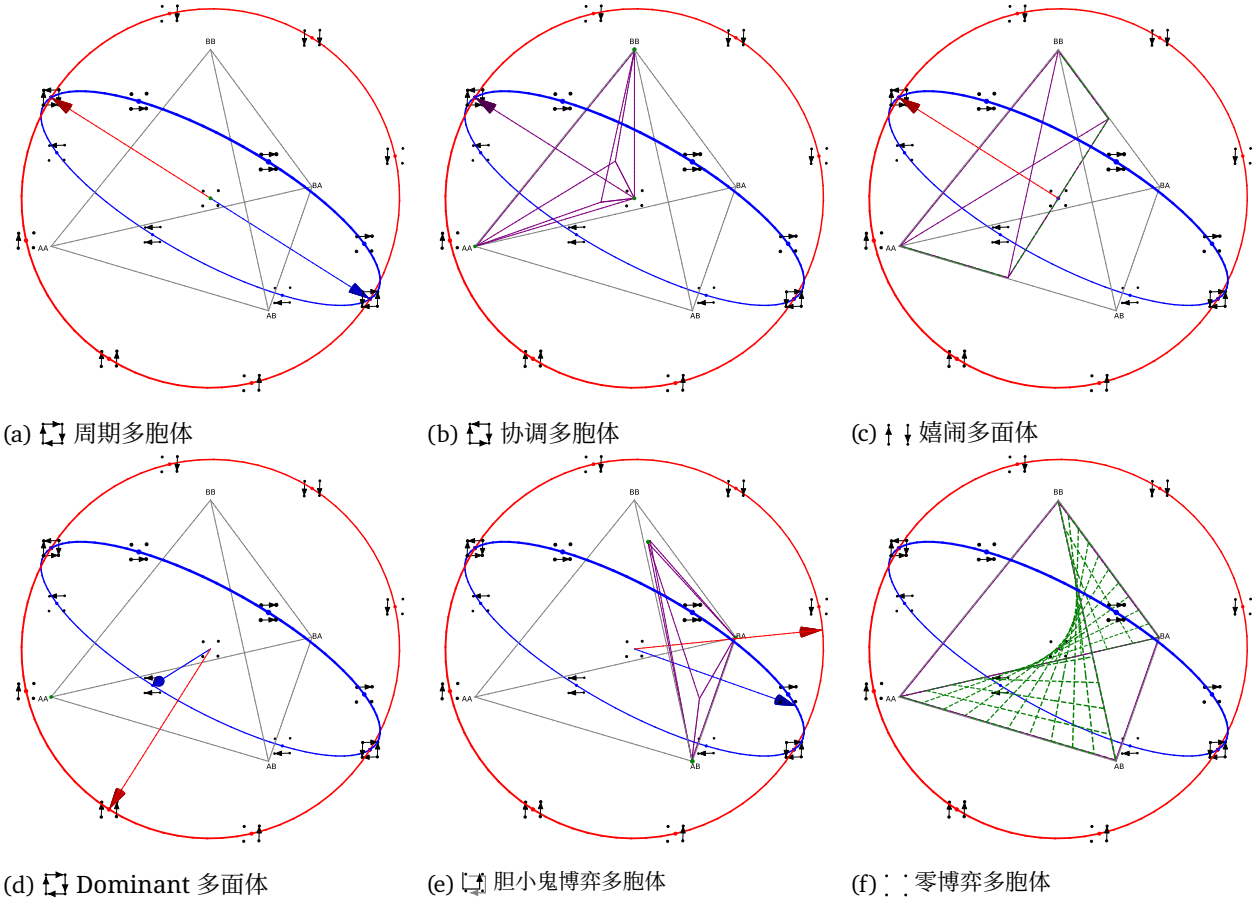
我们提出一种博弈可视化方案， 2×2其中包括博弈的均衡不变嵌入、最佳响应不变嵌入、（粗）相关均衡多胞体和纳什均衡集合（图7）。当箭头方向近乎相反时，该博弈是不变零和博弈（例如周期，图7a）。当箭头方向近乎一致时，该博弈是不变共同收益博弈（例如协调，图7b）。优势博弈的向量近乎正交。有时可能存在连续统纳什均衡，在这种情况下会用虚线表示（例如嬉闹，图7c）。

7.2. 双人博弈可视化

对于策略多于两种（$|A_1| \times |A_2|$）、策略众多（例如具有确定性策略的扩展式[13]博弈），甚至策略无限多（例如具有随机策略的扩展式博弈）的双人博弈，可以通过考察每个可能的 2×2 子博弈来概括，也可以通过对子博弈进行采样来近似， $\tilde{G}_p$。要生成一个点，需均匀采样策略 $\tilde{a}_1^A$、 $\tilde{a}_1^B$、 $\tilde{a}_2^A$和 $\tilde{a}_2^B$，从而得到采样的 2×2 收益。然后，求出均衡不变嵌入（$\tilde{\theta}_1$， $\tilde{\theta}_2$），并在适当时利用对称性。

$$\tilde{G}_p^{\text{global}}(\tilde{a}_1^A, \tilde{a}_1^B, \tilde{a}_2^A, \tilde{a}_2^B) = \begin{bmatrix} G_p(\tilde{a}_1^A, \tilde{a}_2^A) & G_p(\tilde{a}_1^A, \tilde{a}_2^B) \\ G_p(\tilde{a}_1^B, \tilde{a}_2^A) & G_p(\tilde{a}_1^B, \tilde{a}_2^B) \end{bmatrix} \to (\tilde{\theta}_1^{\text{global}}, \tilde{\theta}_2^{\text{global}}) \tag{49}$$

[13]所有扩展式博弈都有等价的标准式表示。然而，子博弈完美性无法用标准式表示。



(a) 周期多胞体 (b) 协调多胞体 (c) 嬉闹多面体

(d) Dominant 多面体 (e) 胆小鬼博弈多胞体 (f) 零博弈多胞体

图 7 | ： 2×2 博弈的可视化。θ_1（红色）和 θ_2（蓝色）参数显示在各自的单位圆上。箭头表示每个玩家收益的不变单位向量，因此其方向指示了能够使收益线性增加的联合分布。可行均衡的（粗）相关均衡多胞体以紫色显示。绿色表示纳什均衡。

大型双人博弈可以表示为由均衡不变嵌入采样得到的点云。我们提出，可以根据该图中嵌入的位置和密度推断博弈的性质。我们将这种采样方案称为全局采样，因为在所分析的整个标准式博弈中，所有策略都采用均匀采样。

扩展式博弈 我们对取自 OpenSpiel 库（8）的三个双人扩展式博弈进行了可视化（图 Lanctot 等人，2019年）。通过枚举每位玩家可采用的纯策略策略，可以将所有扩展式博弈转换为标准式博弈。为了生成可视化结果，我们不需要枚举所有策略；相反，只需对随机纯联合策略进行采样，即可近似实现可视化，从而将分析的尺度扩展到大型扩展式博弈。由于本次可视化所考虑的所有博弈均为非对称博弈，因此我们不利用玩家对称性。我们从收益中任意采样策略，因此策略的顺序也是任意的，我们可以利用策略对称性。因此，绘图覆盖的定义域为：$-\frac{\pi}{2} \leq \theta_p < \frac{\pi}{2}$。

库恩扑克（库恩，1950）是一种非对称、零和的简化扑克变体。首先，请注意，所有点都位于不变零和象限内，这正是理论对双人零和扩展式博弈的预测。有趣的是，出现了许多边界类别（Aidos、野餐和 Hazard）。这表明该博弈中存在玩家在不同策略之间无差异的情形。库恩扑克中的点高度集中于 Dominant 和撒玛利亚人类别，这确实表明，在大多数情况下，存在明显的策略，双方玩家或至少其中一方应当选择这些策略。此外，还存在数量可观的循环交互，这表明该博弈具有一定的战略深度。

双人合作迷你桥牌（Lockhart 等人，2020）是一种共同收益博弈。正如预期的那样，所有

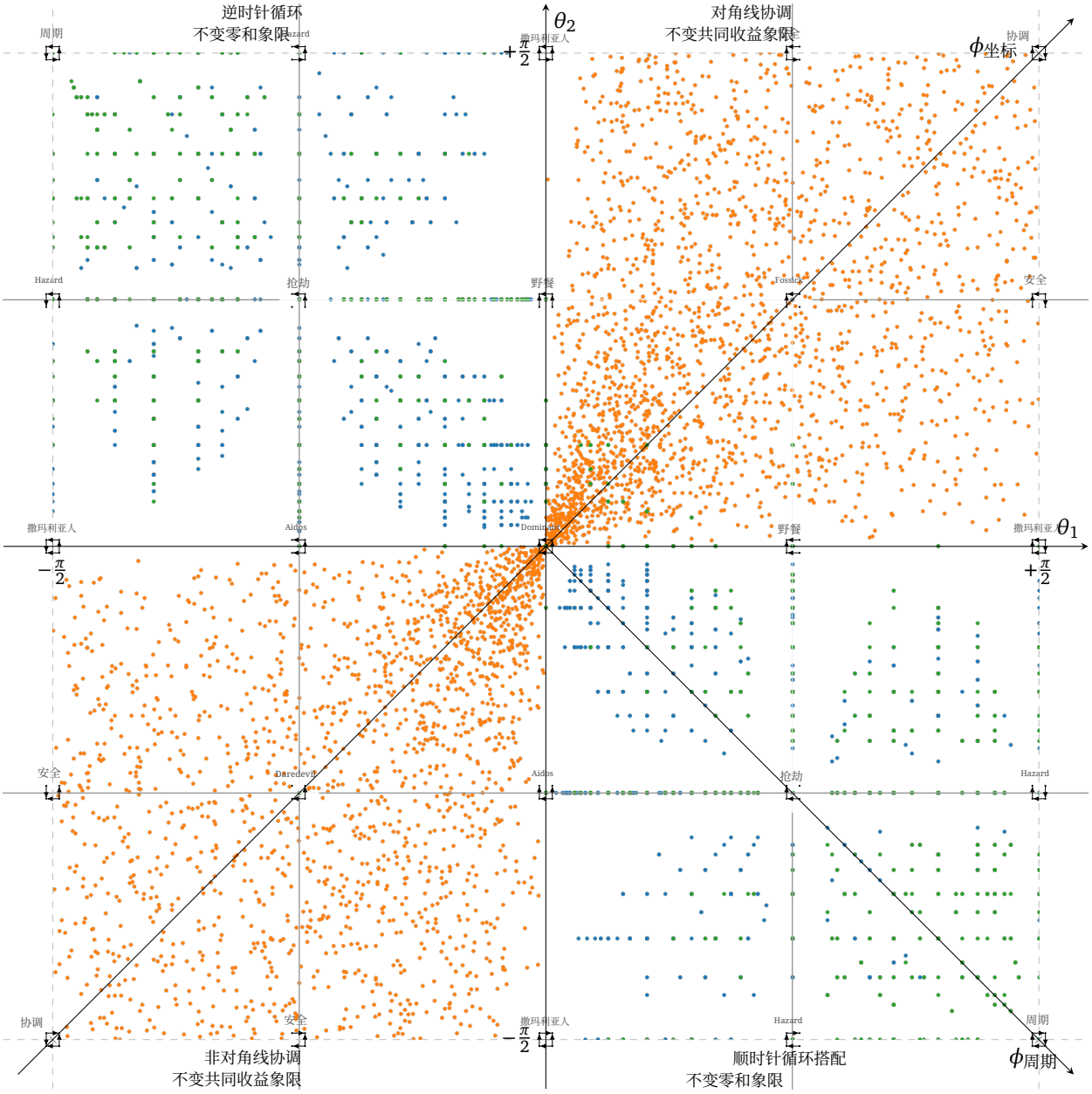


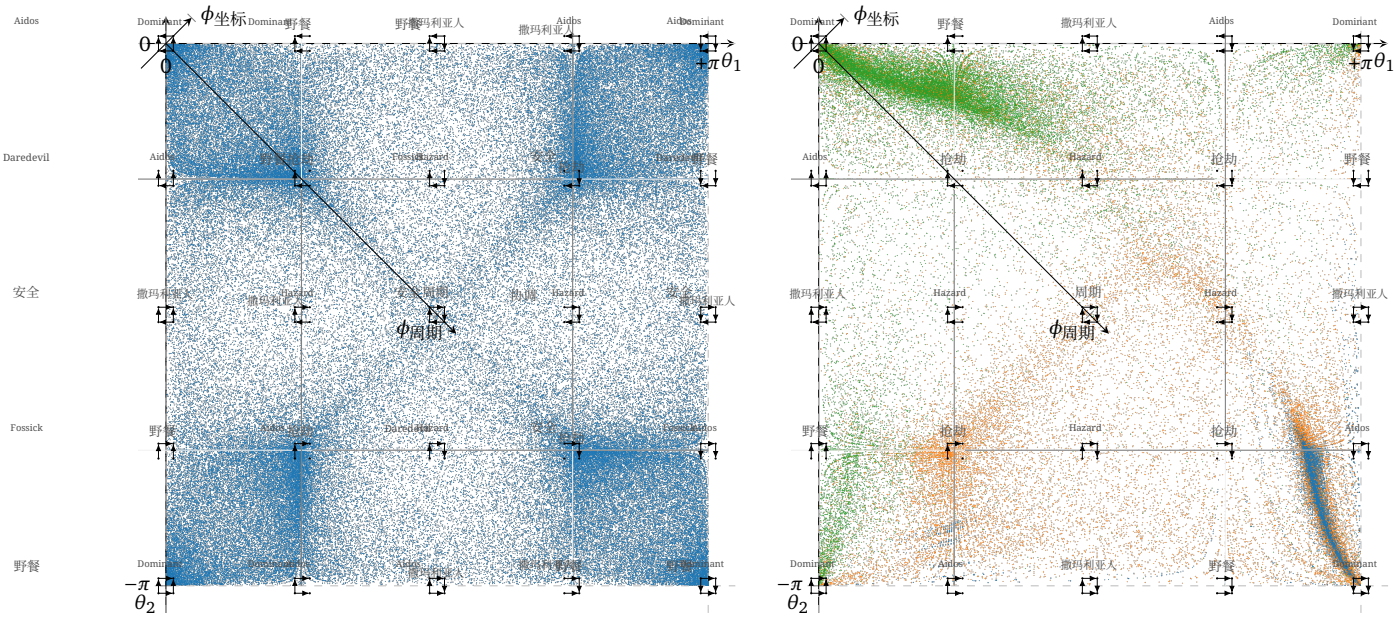
图 8 | 非对称双人扩展式博弈的策略空间可视化：零和昆恩扑克、共同收益 Tiny Bridge 和混合动机 Sheriff。零和博弈仅在对角象限中存在交互。共同收益博弈仅在非对角象限中存在交互。混合动机博弈可以在所有象限中存在交互。这些博弈均为非对称博弈，因此我们可以利用策略置换，而不能利用玩家置换，在约化不变嵌入上绘制这些交互。

安全 Fossick 野餐 抢劫 Hazard

这些点位于可视化的不变共同收益象限中。该博弈在原点附近也有较高的点位集中度，这表明大多数策略互动都属于 Dominant：双方玩家在策略之间都具有严格偏好。协调也占据了显著地位，这表明存在一些需要玩家通过协调来最大化收益的情形。Tiny Bridge 中似乎没有出现边界类别，这意味着玩家在策略选择上处于无差异状态的情况较为罕见。

Sheriff（Farina 等人，2019年）是一种非对称的混合动机博弈。该博弈的点云覆盖了不变共同收益象限和不变零和象限，但主要集中在不变零和象限，这表明该博弈的竞争性强于合作性。此外，边界类别（Aidos、野餐和 Hazard）再次出现，表明玩家处于无差异状态的情形。

撒玛利亚人 撒玛利亚人 野餐 Dominant Aidos



(a) 完整 888 × 888 AlphaStar 联赛（Vinyals 等人，2019年）的近似摘要。每种策略都对应于扩展式博弈中的一种学习到的策略，收益则通过经验计算得出。在定义域 $0 \le \theta_1 \le \frac{\pi}{2}$ 和 $-\frac{\pi}{2} \le \theta_2 \le 0$之外，该图是冗余的。

(b) AlphaStar 联赛的局部近似。对于两名玩家，第一种策略保持固定，仅对第二种策略进行采样。图例：早期训练、中期训练和后期训练。

图 9 | 大型经验正规形博弈分析。在训练早期和后期，局部策略空间主要呈现传递性（Dominant），而在训练中期，策略空间则呈现周期性（Cycle）。这与 Czarnecki 等人(2020）提出的关于技巧型博弈的“陀螺”假说相符。

AlphaStar 联赛 AlphaStar 联赛是一个对称的零和、 888×888、经验标准式博弈。它由 AlphaStar 学到的策略之间的收益构成（Vinyals 等人，2019年）。图9a展示了该博弈的可视化。该博弈是对称的，因此我们可以利用玩家对称性。该博弈是零和的，因此我们只需考虑零和象限。此图还具有策略不变性，因此我们也可以利用策略对称性。不过，为了更简单地与图 9b进行比较，我们不使用策略对称性。因此，此图覆盖的定义域为： $0 \le \theta_1 \le \pi$ 和 $-\pi \le \theta_2 \le 0$，尽管在 $0 \le \theta_1 \le \frac{\pi}{2}$ 和$-\frac{\pi}{2} \le \theta_2 \le 0$ 之外的部分是冗余的。这个经验博弈主要具有 Dominant 策略互动，但也包含显著的撒玛利亚人和周期成分。

Czarnecki 等人(2020) 假设，技能型博弈在训练早期的循环交互较少，因为策略可以轻松地实现传递性改进。在训练中期，存在大量能力较强且具有多样性的策略，从而产生有趣的循环交互。然而，在后期训练中，随着进行博弈所需的技能逐渐完善，策略之间的交互呈现出的循环性会降低。Czarnecki 等人(2020) 将这一现象描述为一种呈「陀螺」形状的循环分布。我们通过采用一种替代性的博弈采样方案，在可视化中检验这一假设：为双方玩家固定第一个策略 a^A，并为每位玩家随机采样第二个策略 $\tilde{a}_p^B$。

$$\tilde{G}_p^{\text{local}}(\tilde{a}_1^B, \tilde{a}_2^B) = \begin{bmatrix} G_p(a^A, a^A) & G_p(a^A, \tilde{a}_2^B) \\ G_p(\tilde{a}_1^B, a^A) & G_p(\tilde{a}_1^B, \tilde{a}_2^B) \end{bmatrix} \to (\tilde{\theta}_1^{\text{local}}, \tilde{\theta}_2^{\text{local}}) \tag{50}$$

AlphaStar 联赛博弈是对称的，因此对两个玩家使用相同的固定策略更为自然。请注意，一般来说，我们可以为每个玩家使用不同的固定策略。

$$\tilde{G}_p^{\text{local}}(\tilde{a}_1^B, \tilde{a}_2^B) = \begin{bmatrix} G_p(a_1^A, a_2^A) & G_p(a_1^A, \tilde{a}_2^B) \\ G_p(\tilde{a}_1^B, a_2^A) & G_p(\tilde{a}_1^B, \tilde{a}_2^B) \end{bmatrix} \to (\tilde{\theta}_1^{\text{local}}, \tilde{\theta}_2^{\text{local}}) \tag{51}$$

我们将这种抽样方案称为局部抽样，因为它是在固定背景策略 $a^{AA} = (a_1^A,\ a_2^A)$ 周围对偏离策略进行抽样。这种抽样策略将有助于揭示局部的策略动态。

围绕这些背景策略。

AlphaStar 联赛博弈中的策略对应于按照训练时间大致排序的策略。较晚出现的策略是在针对先前策略分布进行训练后形成的。因此，后期策略拥有更长的训练时间，也接受过针对更多样化对手的训练，因而更有可能更强。但在训练的各个阶段，博弈的策略格局究竟是什么样的？通过分析围绕对应于训练早期、中期和后期的背景策略展开的博弈，我们可以检验 Czarnecki 等人 (2020) 的假设（图 9b）。

毫不意外，早期训练主要集中在占优区域，在该区域中，玩家强烈倾向于偏离其训练不足的固定背景策略。相比之下，后期训练主要集中在相反的占优区域，在该区域中，玩家严格倾向于坚持其固定背景策略。中期训练的策略空间呈现出多样性，但与其他背景策略相比，其在周期区域中的分布占比更高。总体而言，这项分析支持“陀螺”假说。对训练时间线进行更深入的分析，或许能够揭示更多结构。

7.3. N人多矩阵博弈可视化

可以通过为所有玩家定义一个固定的背景策略 $a^{A\cdots A}=(a_1^A,\ldots,a_N^A)$，然后可视化所有玩家两两交互，从而将双人局部子博弈采样技术扩展到 n 人博弈。

$$\tilde{G}_{pq}^{\text{local}}(\tilde{a}_p^B,\tilde{a}_q^B)=\begin{bmatrix} G_p(a_p^A,a_q^A,a_{-p-q}^{\cdots}) & G_p(a_p^A,\tilde{a}_2^B,a_{-p-q}^{\cdots}) \\ G_p(\tilde{a}_p^B,a_q^A,a_{-p-q}^{\cdots}) & G_p(\tilde{a}_p^B,\tilde{a}_q^B,a_{-p-q}^{\cdots}) \end{bmatrix}\to(\tilde{\theta}_p^{\text{local}},\tilde{\theta}_q^{\text{local}}) \qquad \forall p\neq q\in[1,N] \tag{52}$$

为避免与原始 n 人博弈的玩家位置产生歧义，我们将表示参数从 θ_1 和 θ_2 重命名为 θ_r 和 θ_c，以表示行玩家和列玩家。由于在这种对 n 人博弈的成对近似中，玩家和策略的顺序十分重要，因此有必要使用完整不变空间：$-\pi\le\theta_r\le+\pi$ 和 $-\pi\le\theta_c\le+\pi$。无需绘制玩家对的两种置换：点云将保持相同，但会关于 $\theta_r=\theta_c$ 镜像对称。

仅考虑玩家之间的两两交互，是一种成熟且简洁的博弈表示，称为多矩阵近似（Janovskaja，1968）。这种表示由元组 $(G_{1,\,2}(a_1,\ a_2),\ G_{1,\,3}(a_1,\ a_3),\ \ldots,\ G_{N-1,\,N}(a_N,\ a_{N-1}))$ 描述。通过固定一个 n 人标准式博弈中的背景策略，我们可以围绕这些背景策略构建一个局部多矩阵近似。我们不必将自己局限于标准式博弈中的策略。相反，我们可以在扩展式博弈中，围绕固定的背景策略 $\pi^{A\cdots A}=(\pi_1^A,\ \ldots,\ \pi_N^A)$，对完整策略 $\tilde{\pi}_p^B$ 进行采样。可以采样随机策略或确定性策略（在我们的实验中，我们采样的是确定性策略）。因此，这种可视化方法可用于近似博弈在显著策略附近的策略动态，例如玩家当前在博弈中执行的策略，或者通过某种学习方法找到的策略。我们预计，围绕不同背景策略的局部多矩阵近似景观会有所不同，因为策略会影响博弈中的转移动态。

$$\tilde{G}_{pq}^{\text{local}}(\tilde{\pi}_p^B,\tilde{\pi}_q^B)=\begin{bmatrix} G_p(\pi_p^A,\pi_q^A,\pi_{-p-q}^{\cdots}) & G_p(\pi_p^A,\tilde{\pi}_2^B,\pi_{-p-q}^{\cdots}) \\ G_p(\tilde{\pi}_p^B,\pi_q^A,\pi_{-p-q}^{\cdots}) & G_p(\tilde{\pi}_p^B,\tilde{\pi}_q^B,\pi_{-p-q}^{\cdots}) \end{bmatrix}\to(\tilde{\theta}_p^{\text{local}},\tilde{\theta}_q^{\text{local}}) \qquad \forall p\neq q\in[1,N] \tag{53}$$

三人 Leduc 扑克 Leduc 扑克（Southey 等人，2005年），由 OpenSpiel（Lanctot 等人，2019年）实现，是一种简化的德州扑克实现，包含 N 名玩家、两种花色和 2（$N+1$）张牌。我们研究三人游戏（图10），其中每位玩家分别采用以下背景策略：总是加注、总是跟注和总是弃牌。如果某个信息状态下这些动作中的任何一个不可行，该策略将退回到跟注动作。不出所料，鉴于玩家必须支付盲注，总是弃牌是一种失败策略。在分析跟注对弃牌时，这一点最为明显，因为跟注几乎总是最佳策略。大多数点位于沿 θ_c 轴的博弈流域中：Dominant 和撒玛利亚人。这表明了两个性质。首先，无论玩家采用何种策略，跟注几乎总是行玩家的占优策略（在其他背景策略的情况下）。其次，弃牌策略的强弱取决于其他玩家的玩法。它有时是有利的（Dominant 和撒玛利亚人），有时则是不利的（Dominant 和撒玛利亚人）。在弃牌对加注的情况下，弃牌同样表现不佳。离弃牌有时是更优的策略（Dominant 和撒玛利亚人），但也存在一些情况下它

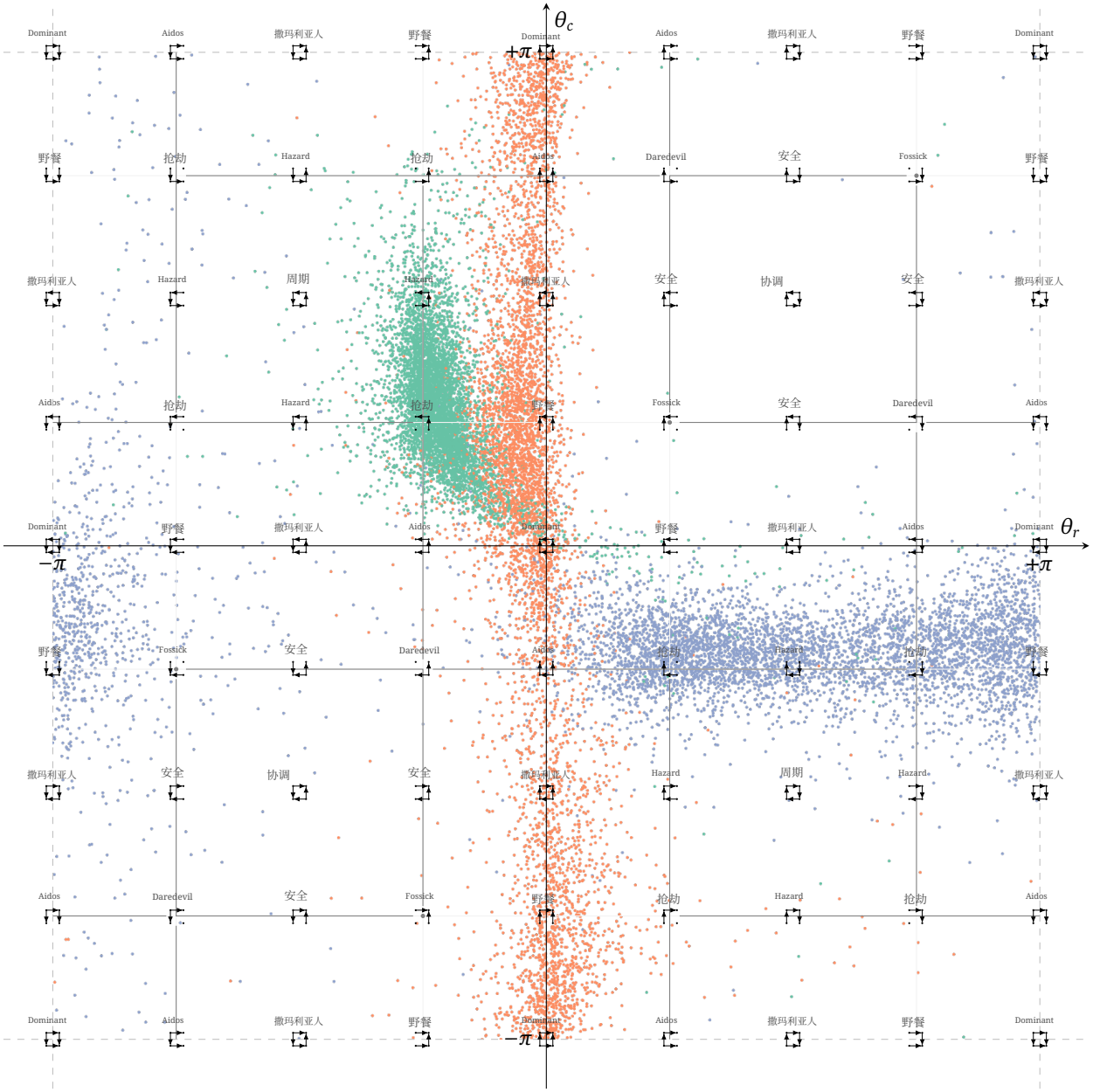
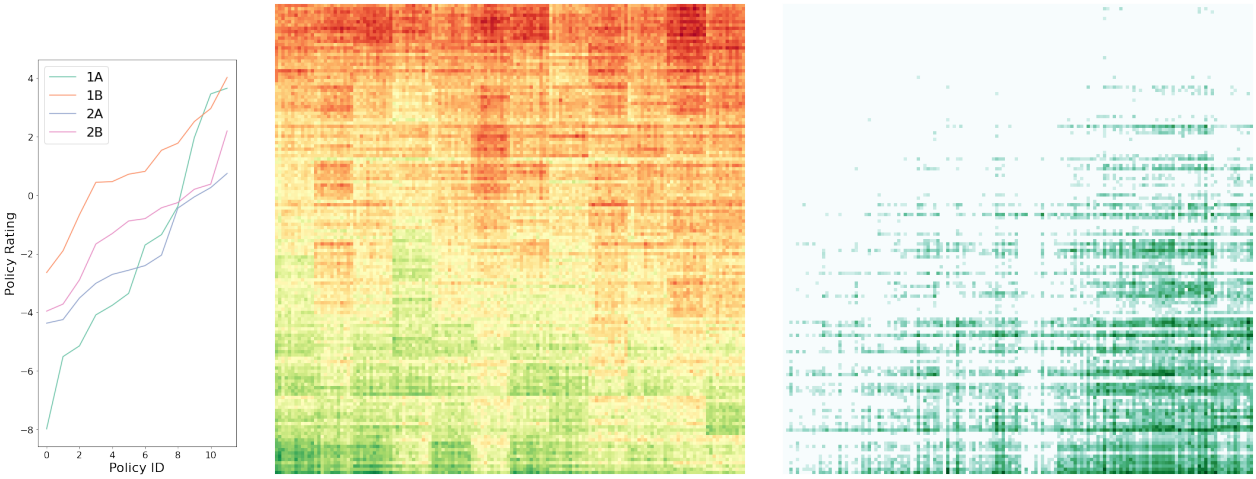


图 10 | 始终加注、始终跟注和始终弃牌这三种背景策略下，三人 Leduc 扑克的局部多矩阵近似可视化。图例：加注对跟注、跟注对弃牌、弃牌对加注。

非（Dominant 且 撒玛利亚人）。在面对其他背景策略时，弃牌很少是较差的策略，但在面对偏离策略（撒玛利亚人）时却是更受青睐的策略。一般而言，围绕 θ_c 轴聚类表明行玩家较强，而围绕 θ_r 轴聚类则表明列玩家较强。加注与跟注动态很有趣：大多数点位于逆时针不变零和区域。始终弃牌的玩家 3 采用了如此弱的策略，以至于其成为无关玩家，而局部的加注与跟注双人博弈几乎是完全零和的。跟注通常是应对加注的良好策略，而应对偏离加注的策略则较差；加注虽然是应对跟注的较差策略，但通常是应对偏离跟注的良好策略，从而形成一个周期。所有局部博弈中的大多数点都位于不变零和象限内，这并不令人意外，因为这是一个零和博弈。只有双人不变零和博弈才会完全位于不变零和象限内。



(a) 策略评分 (b) 队伍 1 对阵队伍 2 的收益 (c) 队伍 1 对阵团队 2 CCE

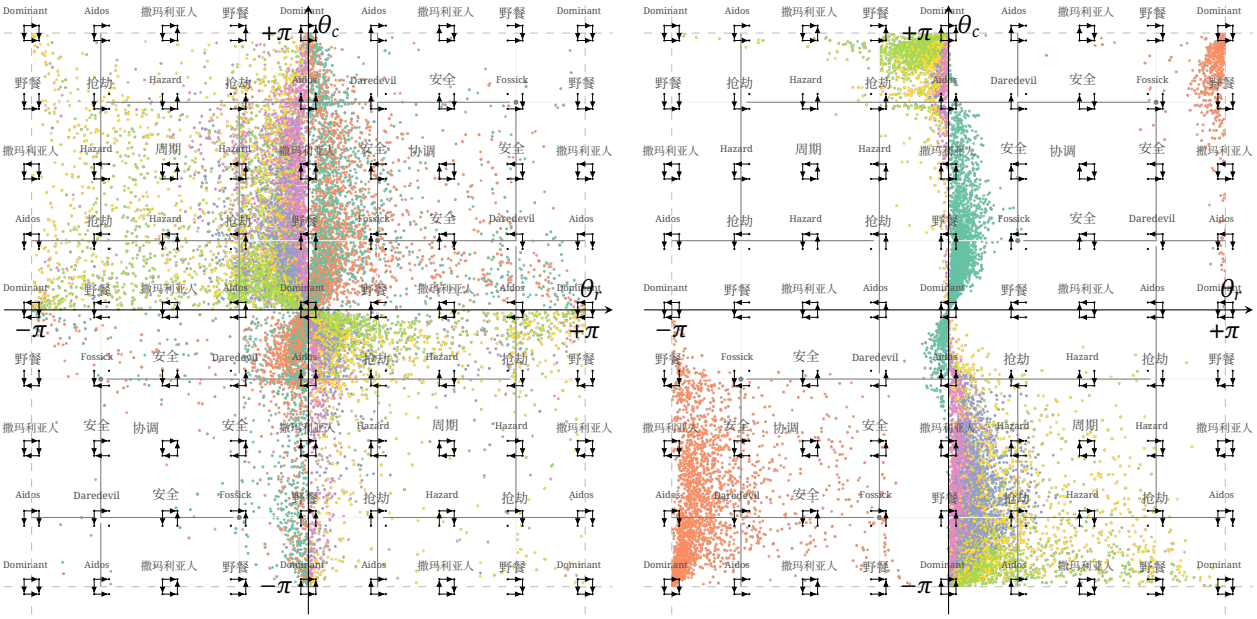
图 11 | Tiny Bridge 中四位玩家各自随机生成的十二种策略，采用基于 CCE 的博弈论评分方法进行评分（Marris等人，2022年b）。评分、收益和 CCE 已按评分从低到高重新排序。为方便起见，CCE 和收益以队伍对队伍的形式在二维平面中可视化；但它们仍然是四人博弈。绿色、黄色和红色分别表示队伍 1 收益高、为零和低。

Tiny Bridge 2对2 Tiny Bridge 是一种扩展式的双人对双人团队博弈（Lockhart等人，2020年），在 OpenSpiel（Lanctot等人，2019年）中实现，队伍之间采用零和动态，队伍内部则采用共同收益动态。该博弈是广受欢迎的纸牌游戏合约桥牌的简化版本，但仅包含两种花色，每种花色各有四张牌。我们使用上文介绍的多矩阵可视化工具来研究 Tiny Bridge 的动态。在此过程中，我们验证了该可视化方法能够：a) 捕捉队友与对手之间有趣的动态关系；b) 区分不同背景策略下的动态。为了生成在量化意义上存在差异的策略，我们为每位玩家生成十二种随机确定性策略，根据其预期回报计算经验标准式博弈（Wellman，2006年）（图 11b），并使用博弈论评分方案对这些策略进行评分（Marris等人，2022b）（图 11a）。该评分方案基于联合均衡分布下的预期收益计算（图 11c）。我们选择使用最大熵 CCE。我们将最强策略选为“最强对最强”的背景集合（图 12a），并将队伍 1 中最强的策略与队伍 2 中最弱的策略组合，作为“最强对最弱”的背景集合（图 12b）。使用不同颜色表示每组中的各个成对交互，从而对每个集合的局部多矩阵近似进行可视化。混合集合包含（相对而言）较强的队伍 1 策略：它们善于克制对手，也善于彼此协调。强策略集合则包含所有玩家的高水平策略。

两个团队分别标记为 1 和 2，每个团队各有两名玩家，分别标记为 A 和 B。对于这两组，正如预期的那样，团队内部的共同收益动态（1A vs 1B 和 2A vs 2B）仅出现在非对角象限中，而团队间的零和动态（1A vs 2A、1A vs 2B、1B vs 2A 和 1B vs 2B）则出现在对角象限中。聚焦于团队内部动态（1A vs 1B 和 2A vs 2B），对于最佳 vs 最佳背景集，点云集中在 Dominant 和撒玛利亚人附近。这表明两个团队都具有良好的团队内部合作。然而，对于最佳 vs 最差背景集，较差团队的策略（2A vs 2B）表现出较差的团队内部合作。点聚集在 Dominant 和撒玛利亚人附近，表明偏离背景策略是有利的。聚焦于团队间动态，最佳 vs 最差背景集表明，队伍 1 优于队伍 2（1A vs 2A、1A vs 2B、1B vs 2A 和 1B 对 2B）。这些点聚集在 θ_c 轴附近，表明行玩家更强。此外，这些点主要位于撒玛利亚人和 Dominant 的吸引域中：这说明行玩家更倾向于采用其背景策略，而非偏离；列玩家则更倾向于偏离，而非采用背景策略。玩家 1A 的动态（1A 对 2A 和 1A 对 2B）比玩家 1B 的动态（1B 对 2A 和 1B 对 2B）更接近 {v2}轴，这表明队伍 1 的两位玩家中，A 的竞争性更强。这并不令人意外，因为 1A 先出牌，在桥牌中自然占据优势，因为他们有机会先叫牌并传达

y

信息。在最佳对最佳背景策略下，团队间的动态更加均衡。



(a) 最佳队伍1对最佳队伍2　　(b) 最佳队伍 1 vs 最差队伍 2

图 12 | 使用两组不同的背景策略集合，对 2vs2 团队扩展式 Tiny Bridge 进行局部多矩阵近似的可视化。共有两个团队，分别为 1 和 2，每个团队有两名玩家，A 和 B。图例：1A 对 1B、2A 对 2B、1A 对 2A、1A 对 2B、1B 对 2A，以及 1B 对 2B。

8. 讨论

在这项工作中，我们为 n 人一般和标准式博弈推导了距离度量和嵌入，使其具有均衡不变性、均衡对称性和更优响应不变性。类似的度量和嵌入也可以很容易地推导出来，适用于其他紧凑型博弈表示形式，例如多矩阵博弈、对称博弈、零和博弈或共同收益博弈。

<escape>为探究均衡不变性，我们重点关注了最受欢迎的均衡解概念，包括纳什均衡、相关均衡和粗相关均衡。其他均衡，例如量化响应均衡（QRE）（McKelvey和 Palfrey，1995），也可能与我们在本研究中推导出的均衡不变嵌入相兼容。此外，这些概念还可能与进化稳定策略（ESS）存在有趣的联系。我们将其他均衡概念的验证留待未来工作。</escape>

定理3 阐明了偏离收益与均衡不变嵌入之间的直接联系。由于偏离收益直接描述了均衡空间，如果已知一个系统的均衡，就可以估计其偏离收益，进而估计均衡不变嵌入。因此，均衡不变嵌入可能成为逆博弈理论中的一种有用工具。

度量空间可用于执行扰动分析（如第6.4节所示）。从数据中估计收益并不罕见，例如在经验博弈论分析（EGTA）中（Walsh等人，2002年；Wellman，2006年）。这类收益具有不确定性，而收益的微小变化可能导致所得均衡发生巨大变化。博弈之间的距离这一概念有助于回答这样的问题：采用不同的估计值时，一个博弈的均衡可能如何变化。例如，它是否接近某个等价类边界，或者即将从不变零和转变为不变共同收益？

流行的博弈论评级方法，例如纳什平均（Balduzzi 等人，2018），被用于对标准式博弈中的策略进行评级和排名。这些排名在收益的仿射变换下保持不变（第 B 节）。因此，均衡不变嵌入保留了策略的博弈论排名，这进一步证明了其基础性。

此外，　　e，我们发现了一种新颖的逐策略尺度较优响应不变变换，它可以 b　　e
用于推导一个　2×2 最佳响应不变嵌入的集合。经过对称性处理后，这将产生 15 个等价项。　　e

类。这一相同的集合此前已通过研究 2×2博弈的最佳响应而得到识别（Borm，1987年）。最佳响应不变性与均衡不变性之间存在深刻的联系（Morris 和 Ui，2004年）。然而，这项工作还为最佳响应不变嵌入提供了均衡不变嵌入、距离度量、高效参数化、图表示和命名方案。

我们注意到，最佳响应不变嵌入中的大多数博弈都存在无差异现象（15 个中有 11 个）。无差异现象无法通过序数收益来刻画，因此需要采用部分序数收益。然而，针对 2×2 博弈的主流拓扑和分类方法都只关注序数博弈。我们认为，这种选择一方面限制了所研究的有趣博弈的范围，另一方面过分强调了不存在无差异现象的博弈，而这类博弈具有高度冗余性，在最佳响应不变等价类中仅占 15 个中的 4 个。

这些 2×2 最佳响应不变嵌入可用于快速计算任意 2×2博弈的均衡，只需将（粗）相关均衡多胞体的极点存储在一个查找表中，该表涵盖 15 种最佳响应不变嵌入。对于任意 2×2 博弈，均衡均可通过以下步骤计算：a) 计算均衡不变嵌入；b) 确定该博弈所属的等价类；c) 根据定理12对查找表中的均衡进行缩放。

对高维大型数据集进行总结和可视化，是机器学习领域的一项重要研究方向。PCA（Hotelling，1936；Pearson，1901年）等有理论依据的降维技术，在最大限度保留信息量的同时降低数据维度，已得到广泛应用。PCA 能够将数据可视化为较低维度的表示。其他理论依据相对较弱的技术，如 t-SNE（Hinton 和 Roweis，2003年；van der Maaten 和 Hinton，2008年），也是一种非常流行的数据集检查工具。尽管统计学和机器学习领域数十年来一直受益于此类工具，但博弈论领域却一直缺乏类似的分析工具，尽管已有一些尝试对博弈进行可视化（Czarnecki 等人，2020；Omidshafiei 等人，2020，2022年）。大型博弈极其复杂：它们具有众多可能的均衡，以及复杂的较优响应动态。本研究中的可视化工具建立在基本原理之上，有望成为开发此类大型博弈分析技术的重要第一步。

局限性

为了说明度量空间和嵌入，我们聚焦于 2×2 博弈。我们推导出了均衡不变嵌入的一种高效参数-化表示，该表示只需要两个变量。这是通过作出一个假设实现的：只有博弈的均衡，或者类似地，策略互动（最佳响应动态）是重要的。大多数解概念都基于均衡或最佳响应，而博弈研究的大部分工作都致力于寻找这些解。然而，这一假设确实带来了一个后果：联合策略 收益的偏好排序在均衡不变变换之后不一定能够保持。这意味着，诸如最大福利之类的均衡选择方法可能会在均衡不变嵌入[14]中选择不同的均衡。此外，以联合收益排序为基础进行论述的博弈叙事可能会瓦解。例如，研究最为广泛的博弈——囚徒困境——具有不会导致福利最大化均衡的支配策略。囚徒困境被转换为 Dominant，从保持均衡的角度来看这是直观的，但从福利或社会困境的角度来看则不那么直观。我们强调，这是本文所述方法的一个特性。如果只关注玩家最终的理性行为，那么将博弈分析简化为仅研究其均衡不变嵌入或最佳响应不变嵌入，是一种去除博弈冗余特征的有价值方式。

2×2 均衡不变嵌入可用于生成更大型博弈的可视化结果。通过观察点云的分布范围，以及点在最佳响应不变等价类中的密度分布，可以从可视化结果中读取收益结构、均衡性质及其他细节。尽管我们提供了证据，表明可视化结果能够揭示博弈的循环性，但它只能展示长度为二的循环，可能无法捕捉更长的循环。例如，使用 2×2 点云对石头剪刀布进行可视化时，只会显示 Dominant 点。

[14]不过，只要知道用于进行变换的均值和尺度，对线性规划的目标函数稍作修改，就能保留这种选择

9. 结论

我们研究了能够降低博弈空间自由度的收益变换。这使得博弈变得：a) 更容易理解，b) 更容易进行采样，以及 c) 更容易进行可视化。在本研究中，我们为博弈定义了一种受均衡启发的度量空间、一种均衡不变嵌入、一种均衡对称嵌入，以及一种较优响应不变嵌入。为了探索这些概念，我们将研究重点放在文献中已有充分研究的 2×2博弈上。我们发现了均衡不变嵌入的一种高效的双变量参数化方法 2×2。这两个变量在几何上表示单位圆上的角度，因此可以在二维空间中得到清晰的可视化。通过这种可视化，可以直观地读取均衡支撑集、循环性、竞争性、距离和对称性等多种性质，我们对此进行了深入探索。我们应用了一种新的均衡收益变换，进一步简化了问题，并重新发现了一组包含 15 个等价类的博弈。我们论证了这组博弈具有基础性，涵盖了所有有趣的 2×2策略互动。我们为这些类别命名，并探索了它们的性质。最后，我们开发了适用于 2×2 和 $|A_1|\times|A_2|$ 标准式博弈以及任意维度多矩阵博弈的可视化工具。由于所有扩展式博弈都有相应的标准式表示和局部多矩阵近似表示，因此我们探索了如何对大型博弈的策略空间进行可视化，而这类博弈迄今一直被认为无法进行可视化。我们希望本研究能够帮助建立对标准式博弈的直觉认识，彰显 15 2×2博弈类别集合的重要性，为博弈论研究者提供实用的可视化工具，并为博弈论算法开发者提供计算方面的见解。

致谢

感谢 Geor。 g Ostrovski、David Parkes、Joel Leibo、Thore Graepel、Karl Tuyls、Ed Hughes、Harshnira Patan i, 和 Kevin McKe 感谢提供有益的讨论和反馈。

参考文献

I. Adler、C. Daskalakis 和 C. H. Papadimitriou。《关于严格竞争博弈的一则短论》。载于 S. Leonardi（编），互联网与网络经济学，第 471–474 页，柏林、海德堡，2009年。Springer Berlin Heidelberg。ISBN 978-3-642-10841-9。R. Aumann。《随机策略中的主观性与相关性》。载于《数理经济学杂志》，1(1)：67–96，1974年。P. Austrin、M. Braverman 和 E. Chlamtac。《Nash 均衡的 NP 完全变体的不可近似性》。载于《近似、随机化与组合优化：算法与技术：第 *14* 届国际研讨会 *APPROX 2011* 和第 *15* 届国际研讨会 *RANDOM 2011*，*2011* 年 *8* 月 *17–19* 日，美国新泽西州普林斯顿。会议录》，第 13–25 页。Springer，2011年。D. Balduzzi、K. Tuyls、J. Perolat 和 T. Graepel。《重新评估评估》。载于《第 *32* 届神经信息处理系统国际会议论文集》，NIPS，第 3272–3283 页，美国纽约州雷德胡克，2018年。Curran Associates Inc。C. Bicchieri。《社会语法：社会规范的本质与动态》。剑桥大学出版社，2005年。doi：10.1017/CBO9780511616037。K. Binmore 和 E. Binmore。《博弈论与社会契约：只管玩》。经济学习与社会演化丛书。MIT Press，1994年。ISBN 9780262024440。M. Böörs、T. Wängberg、T. Everitt 和 M. Hutter。《通过分解进行分类：一种对称 2×2 博弈分类的新方法。《理论与决策》，93(3)：463–508，2022年10月。ISSN 1573-7187。doi：10.1007/s11238-021-09850-z。URL https://doi.org/10.1007/s11238-021-09850-z。P. Borm。《2×2 双矩阵博弈的分类》。载于《运筹研究中心研究手册》，29 (1-2)：69–84，1987年。ISSN 0008-9737。页数：16。N. Boumal、B. Mishra、P.-A. Absil 和 R. Sepulchre。《Manopt：用于流形优化的 Matlab 工具箱》。《机器学习研究杂志》，15(42)：1455–1459，2014年。URL https://www.manopt.org。

S. J. Brams。《引言》，第 1–18 页。剑桥大学出版社，1993年。doi：10.1017/CBO9780511558979.001。S. Brânzei、N. Devanur 和 Y. Rabani。《交换经济中的比例动态》。载于第 22 届 ACM 经济学与计算会议论文集，第 180–201 页，2021年。M. Breton、G. Zaccour 和 M. Zahaf。《环境项目联合实施的博弈论表述》。《欧洲运筹学杂志》，168(1)：221–239，2006年。B. Bruns 和 C. Kimmich。《原型博弈生成多样化的权力、冲突与合作模型》。《生态与社会》，26，2021年。B. R. Bruns。《博弈的名称：定位 2 × 2 博弈》。《博弈》，6(4)：495–520，2015年。ISSN 2073-4336。doi：10.3390/g6040495。URL https://www.mdpi.com/2073-4336/6/4/495。J. M. Buchanan。《撒玛利亚人困境》，第 71–86 页。罗素·塞奇基金会，1975年。ISBN 9780871546593。URL http://www.jstor.org/stable/10.7758/9781610446792.10。X. Chen 和 X. Deng。《确定双人纳什均衡问题的复杂性》。载于第 47 届 IEEE 计算机科学基础年会（FOCS'06），第 261–272 页。IEEE，2006年。W. M. Czarnecki、G. Gidel、B. Tracey、K. Tuyls、S. Omidshafiei、D. Balduzzi 和 M. Jaderberg。《现实世界中的博弈看起来像陀螺》。神经信息处理系统进展，33，2020年。A. Czumaj、M. Fasoulakis 和 M. Jurdziński。《对称双矩阵博弈中的近似良好支撑纳什均衡》，2014年。URL https://arxiv.org/abs/1407.3004。C. Daskalakis、P. Goldberg 和 C. Papadimitriou。《计算纳什均衡的复杂性》。《SIAM 计算杂志》，39：195–259，2009年2月。G. Farina、C. K. Ling、F. Fang 和 T. Sandholm。《扩展式博弈中的相关性：鞍点公式与基准》。载于神经信息处理系统会议（NeurIPS），2019年。P. C. Fishburn 和 D. M. Kilgour。《二元博弈》。《理论与决策》，29(3)：165–182，1990年11月。ISSN 1573-7187。doi：10.1007/BF00126800。URL https://doi.org/10.1007/BF00126800。N. M. Fraser 和 D. M. Kilgour。《非严格序数博弈：对 726 种可能性的全面计算机辅助分析》。《理论与决策》，20(2)：99–121，1986年3月。ISSN 1573-7187。doi：10.1007/BF00135087。URL https://doi.org/10.1007/BF00135087。D. Gauthier。《通过协议形成道德》。英国牛津：牛津大学出版社，1986年。F. Germano。《正规形博弈的若干几何性质与等价类》。《国际博弈论杂志》，34(4)：561–581，2006年11月。ISSN 1432-1270。doi：10.1007/s00182-006-0033-6。URL https://doi.org/10.1007/s00182-006-0033-6。H. Gintis。《理性之界：博弈论与行为科学的统一——修订版》。普林斯顿大学出版社，2014年。ISBN 9780691160849。D. Goforth 和 D. Robinson。《博弈周期表：用户参考手册》。2005年1月。P. W. Goldberg 和 C. H. Papadimitriou。《均衡问题之间的可归约性》。载于第 38 届 ACM 理论计算年会论文集，STOC '06，第 61–70 页，美国纽约州纽约，2006年。计算机协会。ISBN 1595931341。doi：10.1145/1132516.1132526。URL https://doi.org/10.1145/1132516.1132526。J. Hannan。《重复博弈中对贝叶斯风险的近似》。博弈论文集，3：97–139，1957年。R. J. Harris。《2×2 区间对称博弈的几何分类系统》。《行为科学》，14(2)：138，1969年3月1日。URL https://www.proquest.com/scholarly-journals/geometric-classification-system-2-x-interval/docview/1301271952/se-2。最后更新

-

2013-02-24。J. Harsanyi 和 R. Selten。《博弈中的均衡选择一般理论》，第1卷。麻省理工学院出版社，第1版，1988年。

G. Hinton 和 S. Roweis。随机邻域嵌入。神经信息处理系统进展，15:833–840，2003年。网址 http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.13.7959&rep=rep1&type=pdf。H. Hotelling。两组变量之间的关系。《生物统计学》，28(3/4):321–377，1936年。ISSN 00063444。网址 http://www.jstor.org/stable/2333955。A. F. Huertas-Rosero。2×2对称博弈制图。*CoRR*，cs.GT/0312005，2003年。网址 http://arxiv.org/abs/cs/0312005。E. Janovskaja。多矩阵博弈中的均衡点。《立陶宛数学杂志》，8(2):381–384，1968年。H. Kelley、J. Holmes、N. Kerr、H. Reis、C. Rusbult 和 P. Lange。人际情境图谱。2002年1月。doi：10.1017/CBO9780511499845。D. M. Kilgour 和 N. M. Fraser。所有序数 2 × 2博弈的分类学。《理论与决策》，24:99–117，1988年。H. W. Kuhn。一种简化的双人扑克。《博弈理论文集》，1:97–103，1950年。

M. Lanctot, E. Lockhart, J.-B. Lespiau, V. Zambaldi, S. Upadhyay, J. Pérolat, S. Srinivasan, F. Timbers, K. Tuyls, S. Omidshafiei, D. Hennes, D. Morrill, P. Muller, T. Ewalds, R. Faulkner, J. Kramár, B. D. Vylder, B. Saeta, J. Bradbury, D. Ding, S. Borgeaud, M. Lai, J. Schrittwieser, T. Anthony, E. Hughes, I. Danihelka, 和 J. Ryan- Davis。OpenSpiel：用于博弈中强化学习的框架。*CoRR*，2019年。

E. Lockhart, N. Burch, N. Bard, S. Borgeaud, T. Eccles, L. Smaira, and R. Smith. 2020年桥牌叫牌中的人机合作。L. Marris, I. Gemp, T. Anthony, A. Tacchetti, S. Liu, and K. Tuyls. 为解概念注入强劲动力：使用神经均衡求解器求解纳什均衡、相关均衡和粗相关均衡。*CoRR*，abs/2210.09257，2022a年。doi: 10.48550/ARXIV. 2210.09257。URL https://arxiv.org/abs/2210.09257。L. Marris, M. Lanctot, I. Gemp, S. Omidshafiei, S. McAleer, J. Connor, K. Tuyls, and T. Graepel. 2022b年均衡条件下n人一般和博弈的博弈论评分。URL https://arxiv.org/abs/2210.02205。R. D. McKelvey and T. R. Palfrey. 正规形博弈的量子响应均衡。博弈与经济行为，10(1):6–38，1995年。ISSN 0899-8256。doi: https://doi.org/10.1006/game.1995.1023。URL https://www.sciencedirect.com/science/article/pii/S0899825685710238。A. Möbius. 重心演算。J.A. Barth，1827年。D. Monderer and L. S. Shapley. 势博弈。博弈与经济行为，14(1):124–143，1996年。B. Monnot and G. Piliouras. 博弈中无悔学习的极限与局限。知识工程评论，32:e21，2017年。S. Morris and T. Ui. 最佳响应等价。博弈与经济行为，49(2):260–287，2004年。ISSN 0899-8256。doi: https://doi.org/10.1016/j.geb.2003.12.004。URL https://www.sciencedirect.com/science/article/pii/S0899825604000132。H. Moulin and J.-P. Vial. 策略零和博弈：其完全混合均衡无法进一步改善的博弈类别。《国际博弈论杂志》，7(3-4):201–221，1978年。J. Nash. 非合作博弈。数学年刊，54(2):286–295，1951年。S. Omidshafiei, K. Tuyls, W. M. Czarnecki, F. C. Santos, M. Rowland, J. Connor, D. Hennes, P. Muller, J. Péro- lat, B. D. Vylder, A. Gruslys, and R. Munos. 探索多人博弈的格局。《自然通讯》，11(1):5603，2020年11月。ISSN 2041-1723。doi: 10.1038/s41467-020-19244-4。URLhttps://doi.org/10.1038/s41467-020-19244-4。S. Omidshafiei, A. Kapishnikov, Y. Assogba, L. Dixon, and B. Kim. 超越奖励：离线多智能体行为分析的层次化视角，2022年。URL https://arxiv.org/abs/2206.09046。

E.I.Organick. *FORTRAN IV* 入门。 Addison-Wesley，1966年。E. Ostrom, R. Gardner, J. Walker, and J. Walker. 规则、博弈与公池资源。ACLS Humanities E-Book。密歇根大学出版社，1994年。ISBN 9780472065462。G. Ostrovski. 虚拟博弈动态引发的议题，2013年10月。URL http://wrap.warwick.ac.uk/58894/。K. Pearson. 关于空间中点系最佳拟合直线和平面。《伦敦、爱丁堡和都柏林哲学杂志与科学杂志》，2(11):559–572，1901年。doi: 10.1080/14786440109462720。URL https://doi.org/10.1080/14786440109462720。A. Rapoport and M. Guyer. 2×2博弈分类。《一般系统》，11:203–214，1966年。A. Rapoport, M. Guyer, and D. G. Gordon. *2×2*博弈。密歇根大学出版社，安娜堡，1976年。ISBN 0472087428。D. Robinson and D. Goforth. *2×2*博弈的拓扑：一张新的周期表。2005年1月。doi: 10.4324/ 9780203340271。D. Robinson, D. Goforth, and M. Cargill. 朝着非严格排序博弈的拓扑处理迈进 2×2 。2007年7月。R. W. Rosenthal. 若干类双人博弈中的相关均衡。《国际博弈论杂志》，3(3):119–128，1974年9月。ISSN 1432-1270。doi: 10.1007/BF01763252。URL https://doi.org/10.1007/BF01763252。T. Schelling.《武器与影响力》。耶鲁大学出版社，1966年。

D. Schmidtchen。帮助还是不帮助：重新审视撒玛利亚人困境，第470–484页。Springer Berlin Heidelberg，柏林，海德堡，2002年。ISBN 978-3-662-04810-8。doi:10.1007/978-3-662-04810-8_28。网址 https://doi.org/10.1007/978-3-662-04810-8_28。

J. Schosser. 大流行病期间有限资源使用的公平性。《*PLOS ONE*》, 17(6):e0270022, 2022年。

L. S. Shapley。《双人博弈中的若干主题》。1963年。

J. Simpson。《style id='1'>战略理性模拟。博士学位论文，阿尔伯塔大学，2010年。

B. Skyrms。《鹿猎博弈与社会结构的演化》。剑桥大学出版社，2004年。ISBN 9780521533928。

F. Southey、M. Bowling、B. Larson、C. Piccione、N. Burch、D. Billings 和 C. Rayner。贝叶斯诈唬：扑克中的对手建模。第 550–558 页，2005年。R. Sugden。《权利、合作与福利经济学》。帕尔格雷夫·麦克米伦，2005年。N. T. Trendafilov 和 R. A. Lippert。多模态普罗克鲁斯特问题。《线性代数及其应用》，349 (1-3)：245–264，2002年。L. van der Maaten 和 G. Hinton。使用 t-SNE 可视化数据。《机器学习研究杂志》，9：2579–2605，2008年。URL http://www.jmlr.org/papers/v9/vandermaaten08a.html。

O. Vinyals、I. Babuschkin、W. Czarnecki、M. Mathieu、A. Dudzik、J. Chung、D. Choi、R. Powell、T. E walds、P. Georgiev、J. Oh、D. Horgan、M. Kroiss、I. Danihelka、A. Huang、L. Sifre、T. Cai、J. Agapiou、M. Jaderberg 和 D. Silver。使用多智能体强化学习达到《星际争霸II》宗师级水平。《自然》，575，2019年11月。

J. von Neumann 和 O. Morgenstern。《博弈论与经济行为》。普林斯顿大学出版社，1947年。B. Walliser。 2 × 2 博弈的简化分类。《理论与决策》，25(2):163–191，1988年9月。ISSN 1573-7187。doi: 10.1007/BF00134158。URL https://doi.org/10.1007/BF00134158。

W. Walsh、R. Das、G. Tesauro 和 J. Kephart。在多智能体系统中分析复杂策略互动。2002年 01月。

M. P. Wellman. 经验博弈论分析的方法。载于 论文集：《第二十一届全国人工智能大会暨第十八届人工智能创新应用大会》，*2006*年*7*月*16*日至*20*日，美国马萨诸塞州波士顿，第1552–1556页。AAAI出版社，2006年。网址 http://www.aaai.org/Library/AAAI/2006/aaai06-248.php。

G. S. Wilkinson。吸血蝙蝠中的互惠性食物分享。《自然》，308(5955):181–184，1984年3月。ISSN 1476-4687。doi: 10.1038/308181a0。URL https://doi.org/10.1038/308181a0。H. P. Young。策略学习及其局限。牛津大学出版社，2004年。

A. 偏离收益作为线性算子

偏离收益（方程（1）、（2）和（3））可通过对每个玩家的收益进行线性变换得到。

引理 *1*。*WSCE*、相关均衡和 *CCE* 的偏离收益可以表示为对收益的线性运算：$AWSCE(a'_p, a''_p, a_{-p}) = \sum_{\tilde{a}} TCCE_p(a'_p, a''_p, a_{-p}, \tilde{a}) G_p(\tilde{a})$、$ACE(a'_p, a''_p, a) = \sum_{\tilde{a}} TCE_p(a'_p, a''_p, a, \tilde{a}) G_p(\tilde{a})$，以及 $ACCE(a'_p, a) = \sum_{\tilde{a}} TCCE_p(a'_p, a, \tilde{a}) G_p(\tilde{a})$。这些线性运算的秩为 $|A| - |A_{-p}|$。

证明。将收益展平为长度为 $|A|$ 的向量 $G_p(a''')$，将收益增益展平为长度分别为 $|A_p||A|$、$|A_p|^2|A_p|$ 和 $|A_p|^2|A|$ 的向量 $ACCE_p(a'_p \otimes a)$、$AWSCE_p(a'_p \otimes a''_p \otimes a_{-p})$ 和 $ACE_p(a'_p \otimes a''_p \otimes a)$，并使用形状分别为 $|A_p||A| \times |A|$、$|A_p|^2|A_{-p}| \times |A|$ 和 $|A_p|^2|A| \times |A|$ 的矩阵线性算子 $TCCE_p(a'_p \otimes a, a)$、$TWSCE_p(a'_p \otimes a''_p \otimes a_{-p}, a''')$ 和 $TCE_p(a'_p \otimes a''_p \otimes a, a''')$。检查 TCCE 1 的分块矩阵结构，其中 I 是形状为 $|A_{-1}| \times |A_{-1}|$ 的单位矩阵，它有 $|A_1|$ 个分块列和 $|A_1|^2$ 个分块行。注意，WSCE 的定义只是 CCE 的重排版本，即 $AWSCE_p(a'_p, a''_p, a_{-p}) = ACCE(a'_p, a''_p, a_{-p})$。类似地，可以检查 TCE 1，它有 $|A_1|$ 个分块列和 $|A_1|^3$ 个分块行。

$$\tilde{T}_1^{\text{WSCE}} = T_1^{\text{CCE}} = \left[\begin{array}{ccccc} 0 & 0 & \dots & 0 & 0 \\ I & -I & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ I & 0 & \dots & -I & 0 \\ I & 0 & \dots & 0 & -I \\ \hline -I & I & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & I & \dots & -I & 0 \\ 0 & I & \dots & 0 & -I \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline -I & 0 & \dots & 0 & I \\ 0 & -I & \dots & 0 & I \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -I & I \\ 0 & 0 & \dots & 0 & 0 \end{array}\right] \tag{54a}$$

$$T_1^{\text{CE}} = \left[\begin{array}{ccccc} 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \hline 0 & 0 & \dots & 0 & 0 \\ I & -I & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ I & 0 & \dots & 0 & -I \\ \hline -I & I & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \hline 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots \end{array}\right] \tag{54b}$$

第一个块列可以由其余块列的负和构造得到。因此，存在 $|A_{-1}|$ 个冗余列。其余列线性无关，因此秩为 $|A| - |A_{-1}|$。对于 $TCCE_p(a'_p \otimes a,\ a)$，也可以采用类似的构造方式。同样，其中一个块列，即 $|A_{-p}|$，不满足线性无关性，因此秩为 $|A| - |A_{-1}|$。□

B. 博弈论评分

博弈论评分是一种为标准式博弈中的策略评分的方法。设策略评分 $r_p(a_p)$ 是针对每位玩家的策略给出的数值标量，策略排名则是排序顺序 $w_p(a_p)$ = 参数排序 $r_p(a_p)$，其定义方式是：评分相同的策略具有相同的排名（例如，$[1.1, -0.2, 0.3, 0.3] \to [3, 0, 1, 1]$）。一种非常简单（但并非博弈论意义上的）策略评分方法，是考察它们的平均收益。

定义 13 （平均评分）。

$$r_p^{avg}(a_p) = \frac{1}{|\mathcal{A}_{-p}|} \sum_{a_{-p}} G_p(a_p, a_{-p}) \tag{55}$$

一种流行的博弈论评分方法——纳什平均（NA）（Balduzzi 等人，2018），通过对最大熵纳什均衡解进行混合来为收益加权。它最常应用于解唯一的双人零和博弈中。

定义 14 （纳什平均评分）。

$$r_p^{NA}(a_p) = \sum_{a_{-p}} G_p(a_p, a_{-p}) \left(\otimes_{q \in -p} \sigma_p^{MENE}(a_q) \right) \tag{56}$$

定理13 (排名不变性**)**。 纳什平均在仿射变换下具有秩不变性，$G_p(a) \to \hat{G}_p(a) = s_p G_p(a) + b_p(a_{-p})$，其中 $s_p > 0$ 和 $b_p(a_{-p})$ 是任意的。

证明。 考察仿射变换后的排名

$$\hat{r}_p^{\mathrm{NA}}(a_p) = \left(\sum_{a_{-p}} s_p G_p(a_p, a_{-p}) + b_p(a_{-p}) \right) \left(\otimes_{q \in -p} \sigma_p^{\mathrm{MENE}}(a_q) \right) \tag{57a}$$

$$= s_p r_p^{\mathrm{NA}}(a_p) + \underbrace{\sum_{a_{-p}} b_p(a_{-p}) \left(\otimes_{q \in -p} \sigma_p^{\mathrm{MENE}}(a_q) \right)}_{\text{常数：不依赖于 } a_p。} \tag{57b}$$

容易看出，$w_p(a_p) = \hat{w}_p(a_p)$。 □

因此，均衡不变嵌入保留了博弈论中的策略排名。此外，与 Balduzzi 等人 (2018) 最初提出的方法相比，均衡不变嵌入可能提供了更自然的收益归一化。

C. 补充图

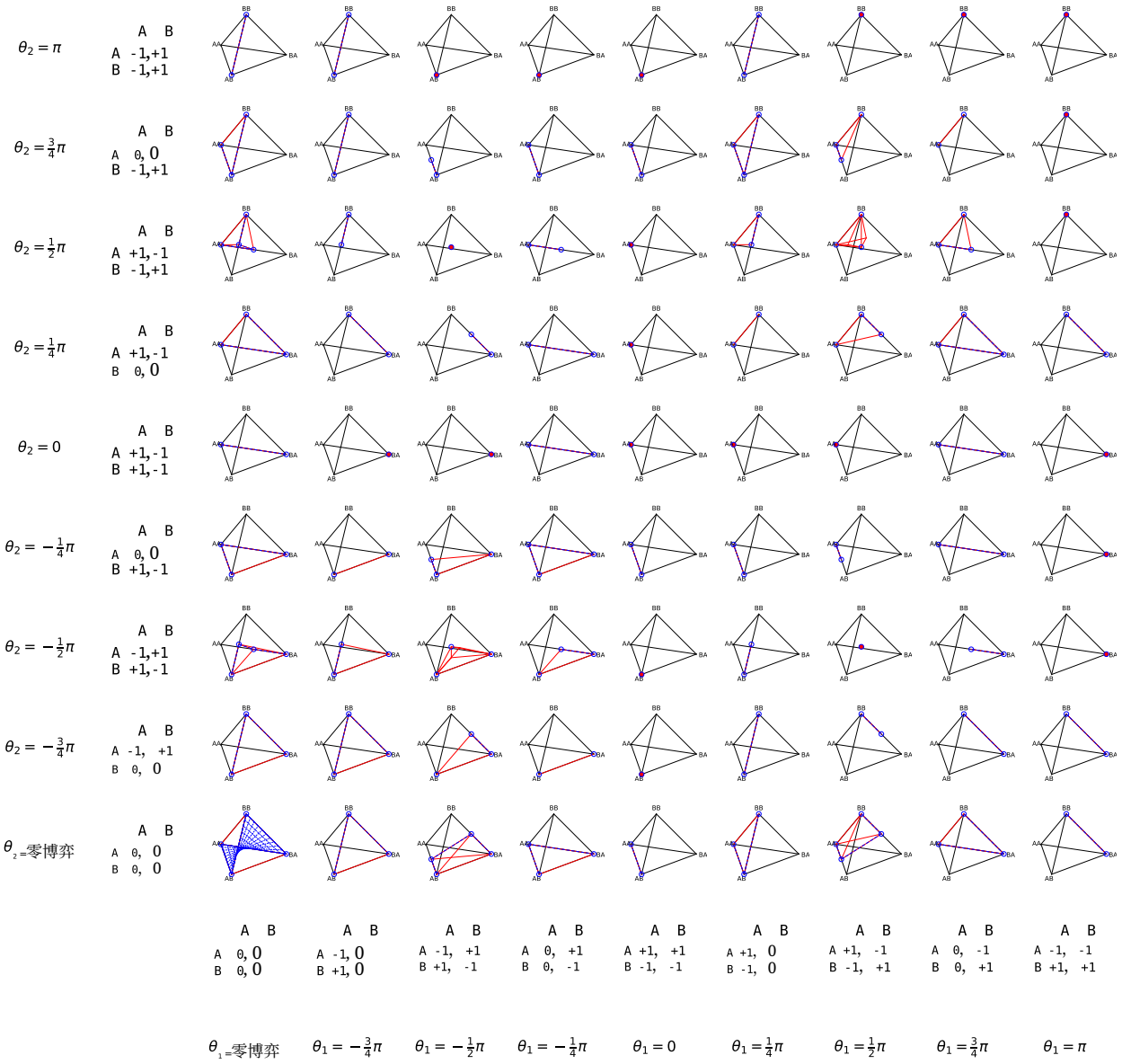


图 13 | 2×2 博弈的纳什均衡（蓝色虚线）与（相关）相关均衡（红色实心多面体）空间。对于双策略博弈，相关均衡和粗相关均衡相同。左列包含玩家1的收益为平凡收益的博弈。底行包含玩家2的收益为平凡收益的博弈。这些博弈中多面体的交集构成了非零博弈的多面体。在零博弈中，所有联合分布都是（相关）相关均衡，所有可因子化联合分布都是纳什均衡。对角象限中的纳什均衡具有凸（可互换）解，因为它们可以映射为零和博弈。非对角博弈不可互换，因为它们可以映射为共同收益博弈。

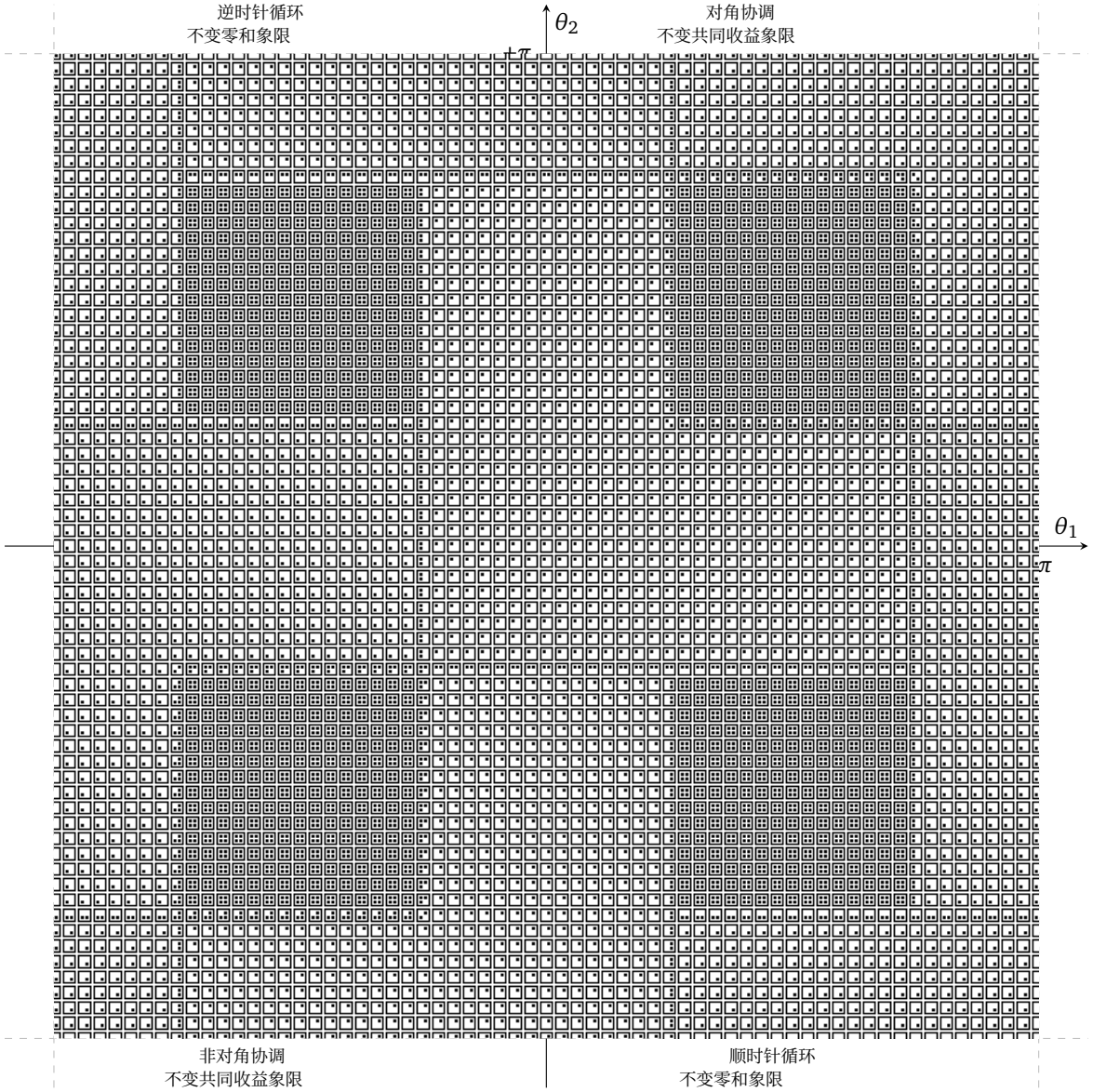


图 14 | 展示了在 2×2 博弈空间中，哪些联合策略可以在均衡中具有支撑集。平铺图表示一个联合分布（正方形）的支撑集，其中四个象限可以选择性地着色，分别对应四种联合策略。所有非零的支撑集组合 $2^4 - 1 = 15$ 均有可能。例如，Cycle 和 Coordination 具有全支撑集，Dominant 具有纯策略支撑集，Fossick 具有对角线支撑集，而 Daredevil 的支撑集涵盖除一个联合策略外的所有联合策略。空间中的大部分区域要么具有纯联合策略，要么允许全支撑混合均衡。其他均衡也可能存在，但它们的测度为零，且仅存在于纯策略博弈与全支撑博弈之间的边界上。

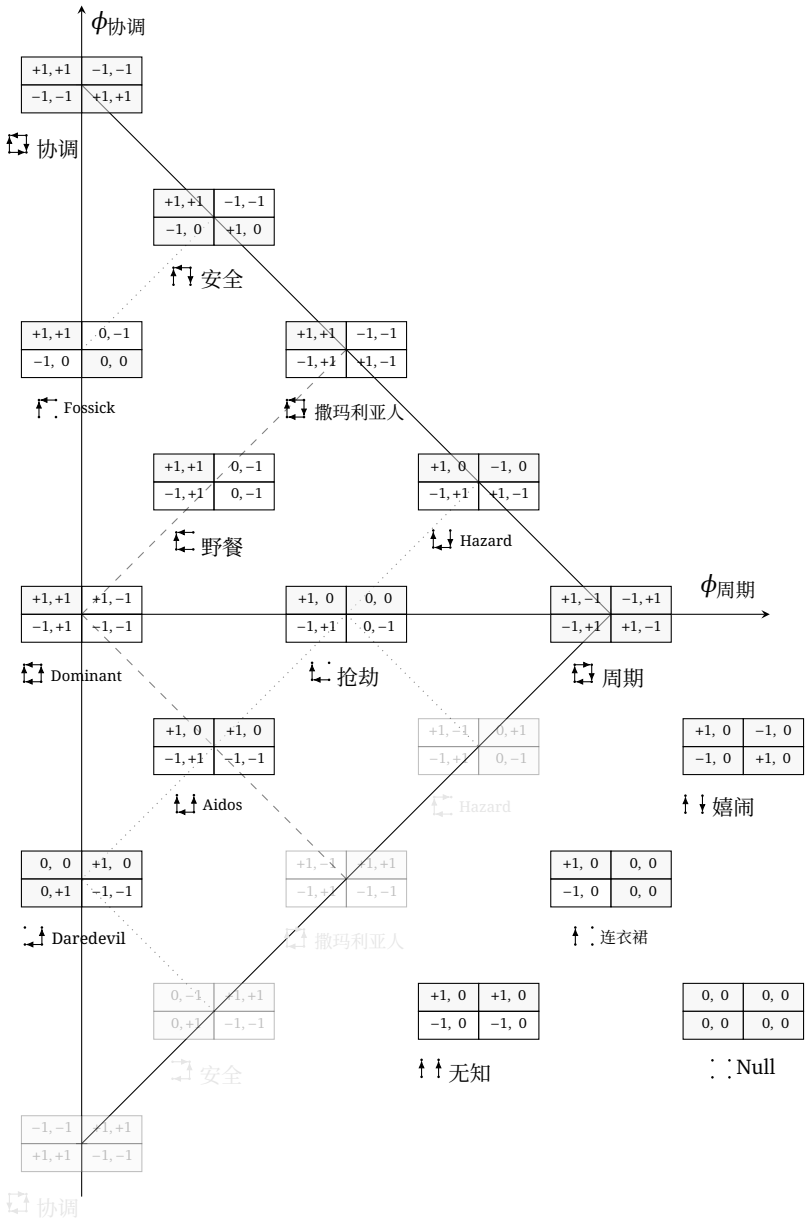


图 15 | 展示了15种基本的双人双策略最佳响应不变嵌入。收益显示在表格中，其中带阴影的联合策略出现在均衡中。褪色的博弈是其他博弈的对称置换。所有 2×2 博弈都将映射到这些博弈中的一个。灰色虚线显示了各个象限：右侧象限是顺时针循环的零和游戏，左侧半象限显示协调博弈。点线显示了均衡解区域，从顶部开始顺时针依次为对角线协调、左上Dominant、顺时针协调、右上Dominant和非对角线协调。均衡对称嵌入之外的博弈都是平凡的。