arXiv:2301.05288v1 [cs.MA] 12 Jan 2023

具有非对称信息和隐藏行动的随机动态博弈方法

Yi Ouyang[1*]、Hamidreza Tavafoghi[2] 和 Demosthenis Teneketzis[3]

[1]Preferred Networks America, Inc.，美国加利福尼亚州伯林盖姆。

[2]Google，美国加利福尼亚州山景城。

[3]电气工程与计算机科学系，密歇根大学，美国密歇根州安娜堡。

*通讯作者。电子邮箱：ouyangyii@gmail.com；贡献作者：hamidreza.tavafoghi@gmail.com；teneket@umich.edu；

摘要

我们在离散时间下，考虑一类具有以下特征的一般序贯随机不对称信息动态博弈。底层系统具有由代理人的联合行动控制的马尔可夫动态。每个代理人的即时效用取决于当前系统状态以及代理人的联合行动。在每个时刻，每个代理人都会对当前系统状态以及代理人在上一时刻的行动进行私人噪声观测。此外，在每个时刻，所有代理人都会对当前系统状态以及其在上一时刻的行动进行公共噪声观测。每个代理人的行动都是其私人信息的一部分。目标是确定基于代理人信息压缩版本且能够按序计算的贝叶斯纳什均衡（BNE）策略组合；此类贝叶斯纳什均衡策略组合并不总是存在。我们提出一种实现上述目标的方法/方法论，并给出一个博弈实例，其中存在具有上述特征的贝叶斯纳什均衡策略组合。我们表明，在代理人没有公共观测的情况下，该方法论同样适用。

关键词：动态博弈、非对称信息、隐藏行动、公共信息、信息压缩、序贯分解

1 引言

我们研究离散时间下的一般类不对称信息序贯随机动态博弈。我们考虑这样一种情形：底层系统具有马尔可夫动态，并由代理人的联合行动控制。每个代理人的即时效用取决于代理人的联合行动和系统状态。在每个时刻，每个代理人都会获得一个私有噪声观测，该观测取决于当前系统状态以及代理人在前一时刻采取的行动。此外，在每个时刻，所有代理人都可能获得关于系统状态及其前一时刻行动的公共噪声观测。代理人的行动是隐藏的，即其他代理人无法直接观测到某个代理人的行动。因此，在每个时刻，代理人对博弈历史都拥有不对称且不完全的信息。具有上述特征的动态博弈广泛存在于工程领域（网络安全、交通运输、能源市场）、经济学领域（产业组织）以及社会技术应用中。

正如Tang 等人（2022年）所指出的，研究不对称信息动态博弈面临的关键挑战有：(i) 随着代理人不断获取信息，其策略的定义域会随时刻推移而扩大。因此，代理人策略的计算复杂度也会随时刻推移而增加。(ii) 由于信号传递[1]（Ho，1980），在许多情况下，代理人对时刻 t博弈状态的评估，因此也就是其在时刻 t采取的策略，取决于此前行动的代理人的策略。因此，我们无法像标准动态规划算法所提供的那样进行标准的序贯分解（依次确定均衡策略组合的各个组成部分）；在标准动态规划算法中，代理人在任意时刻 t 的最优策略不依赖于过去的策略（Kumar 和 Varaiya，1986年，第 6.5 章）。

为应对这些挑战，我们可以寻找基于代理人信息压缩版本且能够按序计算的均衡策略组合。然而，这样的均衡策略组合可能并不存在。

本文提出了一种方法，详见第 3 节，旨在应对上述挑战。根据该方法，我们首先对各时刻代理人的私人信息和公共信息进行压缩。然后，我们基于压缩后的信息定义策略，并证明：如果这一反向过程的每一步都有解，那么基于这些策略的贝叶斯纳什均衡就可以按时间顺序从后向前逐步确定。最后，我们给出一个例子，说明基于压缩信息的贝叶斯纳什均衡策略组合是存在的。

我们证明，所提出的方法适用于代理人没有共同观测且其行动是隐藏的情况。

[1]博弈中的信号传递比团队中的信号传递更为复杂，因为代理人的激励各不相同，而且他们的策略属于各自的私人信息。

1.1 相关文献

关于不对称信息动态博弈，文献已在重复折扣博弈的背景下进行了广泛研究；参见 Zamir（1992）；Forges（1992）；Aumann 等（1995）；Mailath 和 Samuelson（2006年）及其中所引文献。这类博弈的关键特征是不存在动态系统。此外，重复博弈研究主要关注其时域无限且代理人足够耐心（即折扣因子接近于 1）时的渐近性质。在重复博弈中，代理人随时间反复进行阶段（静态）博弈。这一系列文献的主要目标是探讨代理人能够形成自我执行的惩罚／奖励机制的情形，从而创造额外的均衡，使其收益高于长期仅仅进行阶段博弈均衡所能获得的收益。近期研究（参见 H¨orner 等（2011年）；Escobar 和 Toikka（2013年）；Sugaya（2012年））采用了与重复博弈研究相似的方法，考察存在底层动态马尔可夫系统时的不对称信息无限期动态博弈。在满足关于系统动态和信息结构的某些条件时，H¨orner 等（2011年）、Escobar 和 Toikka（2013年）及 Sugaya（2012年）的作者刻画了代理人足够耐心时所达到的一组渐近均衡。

本文研究的问题与Zamir(1992)；Forges(1992)；Aumann 等 (1995)；Mailath 和 Samuelson (2006年)；H¨orner 等 (2011年)；Escobar 和 Toikka (2013年)；Sugaya(2012年)所研究的问题在两个方面有所不同。首先，我们考虑一类动态博弈，其底层系统具有一般马尔可夫动态和一般信息结构；当时域无限且代理人足够耐心时，我们不将关注范围局限于渐近行为。其次，我们研究这样的情形：若不存在与其他代理人的策略互动，每个代理人面临的决策问题都是一个部分可观测马尔可夫决策过程（POMDP），而这本身就是一个复杂的问题。因此，要达到（并计算）一组将代理人之间策略互动纳入考量的均衡策略，是一项极具挑战性的任务。由此看来，代理人寻求通过形成自我执行的惩罚／奖励机制来实现均衡并不太现实；这类机制类似于无限重复博弈中使用的机制。我们认为，我们的结果为理解复杂动态环境中的策略性代理人的行为提供了新的见解，并补充了重复博弈文献中的现有研究成果。

非对称信息下的随机动态零和博弈已由 Renault (2006年)；Cardaliaguet 等 (2015年)；Gensbittel 和 Renault (2015年)；Li 等 (2017年)；Kartik 和 Nayyar (2021年)；Zheng 和 Casta˜n´on (2013年)；Li 和 Shamma (2014年) 进行研究。Renault (2006年)；Cardaliaguet 等 (2015年)；Zheng 和 Casta˜n´on (2013年)；Li 和 Shamma (2014年) 的作者研究具有马尔可夫动态且单方信息缺失的零和博弈（即一方为知情主体，另一方为不知情主体）。Gensbittel 和 Renault (2015年)；Li 等 (2017年)；Kartik 和 Nayyar (2021年) 的作者研究具有马尔可夫动态且双方均存在信息缺失的零和

博弈。Renault (2006年)；Cardaliaguet 等 (2015年)；Gensbittel 和 Renault (2015年)；Li 等 (2017年)；Kartik 和 Nayyar (2021年)；Zheng 和 Casta˜n´on (2013年)；Li 和 Shamma (2014年) 的研究均考虑特定的信息结构。具体而言：双方代理人的行动均为公开观测；在 Renault (2006年)；Cardaliaguet 等 (2015年)；Zheng 和 Casta˜n´on (2013年)；Li 和 Shamma (2014年) 中，知情主体能够完全观测动态系统的状态，而另一方无法直接观测系统状态；在 Gensbittel 和 Renault (2015年)；Li 等 (2017年) 中，每个代理人都能完全观测系统状态的一部分，且双方观测到的状态相互独立或条件独立（给定观测到的行动）。Kartik 和 Nayyar (2021年) 的作者考虑了一般的信息结构，其中每个代理人都拥有一些私人信息，且代理人共享有关动态系统状态及其行动的一些信息。Renault (2006年)；Cardaliaguet 等 (2015年)；Gensbittel 和 Renault (2015年)；Li 等 (2017年)；Kartik 和 Nayyar (2021年)；Zheng 和 Casta˜n´on (2013年)；Li 和 Shamma (2014年) 的作者通过利用零和博弈的一些性质推导结果，例如均衡策略的可互换性以及博弈值的唯一性。这些性质无法推广到非零和博弈。我们研究一类一般的随机动态博弈，其中包括作为特例的非对称信息随机动态零和博弈。与 Renault (2006年)；Cardaliaguet 等 (2015年)；Gensbittel 和 Renault (2015年)；Li 等 (2017年)；Zheng 和 Casta˜n´on (2013年)；Li 和 Shamma (2014年) 不同，我们考虑底层系统的一般马尔可夫动态；上述文献中的系统具有前文所述的特殊结构。我们考虑一种一般的信息结构，能够描述包含不可观测行动和不完全观测的情形，而这些情形并未涵盖在 Renault (2006年)；Cardaliaguet 等 (2015年)；Gensbittel 和 Renault (2015年)；Li 等 (2017年)；Zheng 和 Casta˜n´on (2013年)；Li 和 Shamma (2014年) 中。

Tang 等人（2022年）；Nayyar 等人（2014年）；Gupta 等人（2014年）；Ouyang 等人（2015年，2017年）；Vasal 和 Anastasopoulos（2016年）；Sinha 和 Anastasopoulos（2016年）；Gupta 等人（2016年）；Nayyar 等人（2013a年）所研究的问题与我们的问题最为相关。Nayyar 等人（2014年）；Gupta 等人（2014年，2016年）；Nayyar 等人（2013a年）的作者研究了一类动态博弈，其中智能体基于公共信息的信念（定义见 Nayyar 等人（2014年））独立于其策略，也就是说，智能体之间不存在信号传递。这一性质使他们能够运用 Nayyar 等人（2011年，2013b年）提出的公共信息方法，并定义一个等价的动态博弈，其中虚拟参与者之间的信息是对称的。因此，他们刻画了一类称为基于共同信息的马尔可夫完美均衡的动态博弈均衡。

我们的结果与 Nayyar 等人 (2014年);Gupta 等人(2014年, 2016年); Nayyar 等人 (2013a年)在两个方面有所不同。首先，我们考虑一类一般的动态博弈，其中代理人的 CIB 信念依赖于策略，因此存在信号传递。其次，Nayyar 等人

（2014年）；Gupta 等人（2014年、2016年）；Nayyar 等人（2013a年）要求代理人持续跟踪其所有私人信息。我们提出一种方法，以有效压缩代理人的私人信息，从而减少代理人形成 CIB 信念所需的变量数量。

Tang 等人等人的作者（2022年）；Ouyang 等人（2015年、2017年）；Vasal 和 Anastasopoulos（2016年）；Sinha 和 Anastasopoulos（2016年）研究了一类存在信号传递的不对称信息动态博弈。当时域有限时，Ouyang 等人（2015年、2017年）的作者提出了基于共同信息的完美贝叶斯均衡这一概念，并给出了博弈随时间演进的序贯分解。Vasal 和 Anastasopoulos（2016年）；Sinha 和 Anastasopoulos（2016年）将Ouyang 等人（2015年、2017年）的结果分别推广到有限时域线性二次高斯（LQG）动态博弈和无限期动态博弈。

Tang 等人（2022年）的研究将 Ouyang 等人（2017年）的模型扩展到了代理人团队之间的博弈。每位代理人都有自己的私人信息，并会延迟 d 与本团队成员分享这些信息；团队也拥有公共信息。Tang 等人（2022年）的作者考虑了两类策略：基于充分私有信息（SPIB）的策略，这类策略只压缩私人信息；以及基于充分私有与共同信息（SPCIB）的策略，这类策略会同时压缩公共信息和私人信息。他们证明，基于 SPIB 策略的贝叶斯纳什均衡（BNE）存在，且此类均衡的收益向量集合与所有 BNE 的集合相同。他们提出了一种反向归纳序贯程序；若该程序有解，则其解构成一个 SPCIB 贝叶斯纳什均衡，并指出了能够保证 SPCIB 贝叶斯纳什均衡存在的情形。Tang 等人（2022年）、Ouyang 等人（2015年、2017年）、Vasal 和 Anastasopoulos（2016年）以及 Sinha 和 Anastasopoulos（2016年）所研究的动态博弈类别满足以下假设：（i）代理人的行动可观测；（ii）每位代理人都能完全观测自己的局部状态/类型；（iii）给定代理人的行动后，各局部状态的演化相互独立。我们放宽 Tang 等人（2022年）、Ouyang 等人（2015年、2017年）、Vasal 和 Anastasopoulos（2016年）以及 Sinha 和 Anastasopoulos（2016年）的假设（i）—（iii），并研究一类具有非对称信息、隐藏行动、不完全观测以及受控耦合动态的一般动态博弈。

1.2 研究贡献

我们研究/分析离散时间下的一般序贯随机动态博弈，此类博弈具有非对称信息，其中底层系统是动态的，信息结构是非经典的；在每个时刻，代理人都拥有私有信息和公共信息，且其行动是隐藏的（一个代理人的行动无法被其他代理人直接观察到）。我们的核心贡献是提出一种方法论，用于发现基于代理人压缩后的私有信息和

公共信息，并且只要这一逆向过程的每一步都有解，就可以按时间顺序从后往前逐步确定。我们给出了一个存在此类贝叶斯纳什均衡策略组合的例子。我们还表明，当代理人之间没有共同观测且其行动被隐藏时，我们的方法论同样适用。

1.3 结构安排

本文其余部分安排如下：第2节介绍博弈模型及其均衡概念。第3节阐述我们的目标，并介绍实现该目标的方法论。第4 节首先介绍代理人私人信息和公共信息的压缩形式，这些信息足以用于决策；然后定义基于代理人压缩信息的基于充分信息（SIB）的策略。第5 节首先介绍基于充分信息的贝叶斯纳什均衡（SIB-BNE）；然后给出博弈的序贯分解，即一种反向归纳过程，只要该过程的每一步都有解，就能确定 SIB-BNE。第6 节给出一个例子，说明我们的方法论，并展示 SIB-BNE 的存在。第7 节表明，当代理人之间没有共同观测，且每位代理人的行动都属于其私人信息时，我们的求解方法适用于随机动态博弈。附录中比较了本文与Tavafoghi 等人(2022年)对压缩私人信息的定义，并介绍了与第6 节示例中 SIB-BNE 存在性相关的一些技术细节。

2 模型

下面我们介绍具有策略性主体的动态决策问题（动态博弈）模型；该模型类似于Tavafoghi 等人(2022年)针对非策略性主体动态决策问题提出的模型。

2.1 系统动态

有 N 位策略性主体，他们生活在一个动态的马尔可夫世界中，时域为 $\mathcal{T}:=\{1,2,\ldots,T\}$, $T<\infty$。令 $X_t \in \mathcal{X}_t$ 表示 $t \in \mathcal{T}$时的世界状态。在 t时刻，每位代理人（以 $i \in \mathcal{N}:=\{1,2,\ldots,N\}$为索引）选择一个行动 $a_t^i \in \mathcal{A}_t^i$，其中 $\mathcal{A}_t^i$ 表示其在 t时刻可用行动的集合。给定集体行动组合 $A_t:=(A_t^1,\ldots,A_t^N)$，世界状态按照以下随机动态方程演变，

$$X_{t+1} = f_t(X_t, A_t, W_t^x), \tag{1}$$

其中 $W_{1:\,T-1}^x$ 是一列独立随机变量。初始状态X_1 是一个随机变量，其概率分布为 $\mu_0 \in \Delta(\mathcal{X}_1)$。

在每个时刻 $t \in \mathcal{T}$，代理人 i 在采取行动之前，会收到关于当前世界状态 X_t 和行动组合A_{t-1}的带噪私有观测 $Y_t^i \in \mathcal{Y}_t^i$，表示为

$$Y_t^i = O_t^i(X_t, A_{t-1}, W_t^i), \tag{2}$$

其中，$W_{1:T}^i, i \in \mathcal{N}$是独立随机变量序列。此外，在每个 $t \in \mathcal{T}$，所有代理人都会收到关于当前世界状态 X_t 和行动组合 A_{t-1}的公共观测 $Z_t \in \mathcal{Z}_t$，其定义如下：

$$Z_t = O_t^c(X_t, A_{t-1}, W_t^c), \tag{3}$$

其中，$W_{1:T}^c$是独立随机变量序列。我们假设随机变量 X_1, $W_{1:T-1}^x$, $W_{1:T}^c$以及$W_{1:T}^i$, $i \in \mathcal{N}$相互独立。

为避免测度论方面的技术困难，并使表述清晰、简便，我们假设所有随机变量都取值于有限集。

假设 1.（有限博弈）集合 $\mathcal{N}$, $\mathcal{X}_t$, $\mathcal{Z}_t$, $\mathcal{Y}_t^i$, $\mathcal{A}_t^i$, $i \in \mathcal{N}$均为有限集。

2.2 信息结构

令 H_t 表示时刻 t 所有代理人的汇总信息。假设代理人具有完美回忆，则有 $H_t = \{Z_{1:t}, Y^{1}_{1:t}, \cdots, Y^{N}_{1:t}, A_{1:t-1}^{1:N}\}$，即 H_t 表示所有代理人过去和当前观测以及所有代理人过去行动的集合。代理人汇总信息的所有可能实现值构成的集合为 $\mathcal{H}_t := \prod_{\tau \le t} \mathcal{Z}_\tau \times \prod_{i \in \mathcal{N}} \prod_{\tau \le t} \mathcal{Y}_\tau^i \times \prod_{i \in \mathcal{N}} \prod_{\tau < t} \mathcal{A}_\tau^i$。

在时刻 $t \in \mathcal{T}$，所有代理人并不完全了解汇总信息 H_t。令 $C_t := \{Z_{1:t}\} \in \mathcal{C}_t$ 表示代理人关于 H_t 的共同信息，并令 $P_t^i := \{Y_{1:t}^i, A_{1:t-1}^i\} \backslash C_t \in \mathcal{P}_t^i$ 表示代理人 i 关于 H_t 的私人信息，其中 $\mathcal{P}_t^i$ 和 $\mathcal{C}_t$ 分别表示时刻 t 代理人 i 的私人信息和共同信息的所有可能实现值构成的集合。我们假设，观测 Y_τ^i、$\tau \in \{1, 2..., t\}$ 以及行动 A_τ^i、$\tau \in \{1, 2..., t-1\}$ 为代理人 i 所知，但不一定为时刻 $t \in \mathcal{T}$ 的所有其他代理人（记为 $-i$）所完全知晓。因此，对于所有 $i \in \mathcal{N}$，有 $P_t^i \subseteq \{Y_{1:t}^i, A_{1:t-1}^i\}$；对于所有 $t \in \mathcal{T}$，有 $H_t = \left(\bigcup_{i \in \mathcal{N}} P_t^i\right) \cup C_t$。由此，$\{C_t, P_t^i, i \in \mathcal{N}\}$ 在每个时刻 $t \in \mathcal{T}$ 都构成 $\mathcal{H}_t$ 的一个划分。第2节第2.5小节将讨论若干可视为本模型特殊情形的信息结构。

2.3 策略与效用：

令 $H_t^i := \{C_t, P_t^i\} \in \mathcal{H}_t^i$ 表示代理人 i 在 t时可获得的信息，其中 $\mathcal{H}_t^i$ 表示代理人 i 在t时信息的所有可能实现值的集合。代理人 i在 t时的行为策略记为 g_t^i，定义如下：

$$g_t^i : \mathcal{H}_t^i \to \Delta(\mathcal{A}_t^i) \tag{4}$$

其中，$\Delta(\mathcal{A}_t^i)$ 是 $\mathcal{A}_t^i$ 上的概率质量函数集合。我们用

$$g^i := (g_1^i, g_2^i, \ldots, g_T^i) \tag{5}$$

代理人 i；$g^i \in \mathcal{G}^i$，其中 $\mathcal{G}^i$ 是由 (4)-(5描述的可容许策略集合。我们将策略组合 g 记为

$$g := (g^1, g^2, \ldots, g^N) \tag{6}$$

$g \in \mathcal{G}$，其中 $\mathcal{G}$ 是由 (4)-(6描述的可容许策略组合集合。我们用

$$g^{-i} := (g^1, \ldots, g^{i-1}, g^{i+1}, \ldots, g^N) \tag{7}$$

代理人 i 在 t 时的瞬时效用取决于系统状态 X_t 和集体行动组合 A_t，表示为 $u_t^i(X_t, A_t)$。代理人 i 在时域 $\mathcal{T}$ 内的总效用表示为：

$$U^i(X_{1:T}, A_{1:T}) = \sum_{t \in \mathcal{T}} u_t^i(X_t, A_t). \tag{8}$$

2.4 均衡概念：

我们将贝叶斯纳什均衡（BNE）视为解概念（Fuden- berg 和 Tirole，1991年）。若对所有 $g^* = (g^{*1}, g^{*2}, \ldots, g^{*N})$，策略组合 $i \in \mathcal{N}$

$$\mathbb{E}^{g^*}\{U^i(X_{1:T}, A_{1:T})\} \geq \mathbb{E}^{g^{*-i}, \hat{g}^i}\{U^i(X_{1:T}, A_{1:T})\}, \quad \forall \hat{g}^i \in \mathcal{G}^i. \tag{9}$$

2.5 特殊情形

我们讨论若干不对称信息动态博弈的实例，它们是上述一般模型的特殊情形。

1）嵌套信息结构：考虑一个双人博弈，其中一名玩家是知情玩家，另一名玩家是不知情玩家，并具有一般马尔可夫动态。在每个时刻 $t \in \mathcal{T}$，知情玩家都会对状态 X_t 进行私人完全观测，即 $Y_t^1 = X_t$。不知情玩家无法观测状态 X_t。知情玩家和不知情玩家都能观测彼此的行动，即 $Z_t = \{A_{t-1}\}$。因此，对所有 $t \in \mathcal{T}$，我们有 $P_t^1 = \{X_{1:t}\}$、$P_t^2 = \emptyset$ 和 $C_t = \{A_{1:t-1}^1, A_{1:t-1}^2\}$。上述嵌套信息结构对应于 Renault（2006年）；Cardaliaguet 等（2015年）；Renault（2012年）；Li 和 Shamma（2014年，2017年）；Zheng 和 Castaˆn´on（2013年）所研究的动态博弈；其中在 Renault（2012年）和 Li 和 Shamma（2017年）的研究中，状态 X_t 是静态的。

2）延迟共享信息结构：考虑一个有 N 个博弈参与者且行动可观测的博弈，其中代理人以 d 步

的延迟观测彼此的观测。也就是说，$P_t^i = \{Y_{t-d+1:\ t}^i\}$ 且 $C_t = \{Y_{1:t-d}, A_{1:t-1}\}$。我们注意到，在我们的模型中，假设代理人在 t 时的公共观测 Z_t 仅是 X_t 和和 A_{t-1}的函数。因此，为了在我们的模型框架下描述具有延迟共享信息结构的博弈，我们需要扩充状态空间，将代理人最近的 d 个观测纳入增广状态的一部分。定义 $\tilde{X}_t := \{X_t, M_t^1, M_t^2, ..., M_t^d\}$为增广系统状态，其中 M_t^i: $= \{A_{t-i}, Y_{t-i}\} \in \mathcal{A}_{t-i} \times \mathcal{Y}_{t-i}, i \in \mathcal{N}$；也就是说， M_t^i充当代理人在 $t-i$ 时对观测 Y_{t-i} 的时间记忆。于是，我们有 $\tilde{X}_{t+1} = \{X_{t+1}, M_{t+1}^1, M_{t+1}^2, ..., M_{t+1}^d\} = \{f_t(X_t, A_t, W_t^x), (Y_t)\ M_t^1, ..., M_t^{d-1}\}$且 $Z_t = \{M_t^d, A_{t-1}\} = \{Y_{t-d}, A_{t-1}\}$。

上述环境呈现了对称信息结构与非对称信息结构之间的联系。代理人之间的信息不对称程度会随着 d 的增加而加剧。上述延迟共享信息结构对应于 Tavafoghi 等人(2016年)所研究的动态博弈。

3）隐藏行动下的完全控制动态：考虑一个有 N名博弈参与者的博弈，其中状态 $X_t := (X_t^1, X_t^2, ..., X_t^N)$包含 N个分量。代理人 i, $i \in \mathcal{N}$完全控制 X_t^i，即 $X_{t+1}^i = A_t^i$。代理人 i的行动 A_t^i, $t \in \mathcal{T}$对所有其他代理人 $-i$均不可观测。每位代理人 i, $i \in \mathcal{N}$都会在 $t \in \mathcal{T}$对系统状态进行带噪私有观测 $Y_i^t(X_t, W_t^i)$。因此，我们有 $P_t^i := \{A_{1:t}, Y_{1:\ t}^i\}$, $C_t = \emptyset$。

3 目标与方法论

3.1 目标

我们的目标有两个方面：(i) 确定基于代理人私人信息和公共信息压缩版本的贝叶斯纳什均衡策略组合。(ii) 通过对博弈进行序贯分解来计算上述策略组合，也就是说，采用反向归纳序贯程序，在该程序的每一步都有解时识别出一个均衡策略组合。

3.2 方法论

我们提出一种实现上述目标的方法论，具体步骤如下：

- 步骤 1。我们确定一种相互一致的代理人私人信息压缩方式，该方式足以满足决策需要（这种相互一致的压缩方式可能并不唯一）。基于这种压缩，我们引入基于充分私有信息（SPIB）的信念系统。

- 步骤 2. 基于步骤 1 的结果，我们通过定义基于公共信息（CIB）的信念系统，确定对决策而言足够的智能体的共同信息压缩。CIB信念系统确保在每个时刻，每个代理人的 CIB

信念都与其 SPIB信念保持一致，即使代理人偏离其均衡策略而采取任意策略也是如此。这种一致性意味着每个代理人都会形成自己的 CIB信念系统，并且每个代理人的 CIB信念系统都是所有代理人的共同知识。

- 步骤 3。基于对代理人私人信息与公共信息的压缩，我们为每个代理人引入基于充分信息（SIB）的策略（即在每个时刻依赖于代理人充分私人信息和 CIB 信念系统的策略）以及 SIB 贝叶斯纳什均衡。我们证明，SIB 策略满足最佳响应的一个关键闭合性性质。基于这一性质，我们给出博弈的序贯分解，也就是说，给出一个反向归纳序贯程序；如果该程序的每一步都有解，该程序便能确定一个 SIB 贝叶斯纳什均衡。

- 步骤 4。我们给出一个具有非对称信息和隐藏／不可观测行动的随机动态博弈例子，其中存在一个 SIB 贝叶斯纳什均衡。

4 私人信息与公共信息的压缩

在第 4.1 节中，我们刻画／确定所有代理人的私人信息的相互一致压缩，这些压缩足以用于决策。在第 4.2节中，我们引入基于公共信息的信念，即代理人公共信息的压缩版本，足以用于决策。

4.1 充分私人信息（步骤 1）

我们提出/考虑对代理人的私人信息进行压缩，并以相互一致的方式完成压缩，使压缩后的信息足以用于决策。

定义1（充分私人信息）。如果 $S_t^i, i = 1, \ldots, N$ 是代理人的充分私人信息，则称其满足

(i) S_t^i 是 H_t^i 的函数，使得对于某些共同已知的函数 $\zeta_t^i, i = 1, 2, \ldots, N$，有 $S_t^i = \zeta_t^i(H_t^i)$。(ii) S_t^i 可以使用某些共同已知的函数 $\phi_t^i, i = 1, 2, \ldots, N$，按顺序更新为 $S_t^i = \phi_t^i(S_{t-1}^i, Y_t^i, Z_t, A_{t-1}^i)$。(iii) 对于任意实现 $x_t, p_t^{-i}, p_t^i, c_t$，以及相应的 $s_t^{-i} = \zeta_t^{-i}(p_t^{-i}, c_t)$ 和 $s_t^i = \zeta_t^i(p_t^i, c_t)$，和任意策略组合 g，其中 $g_t^i \colon \mathcal{S}_t^i \times C_t \to \Delta(\mathcal{A}_t^i)\ \forall i, \forall t$，使得 $\mathbb{P}^g(p_t^i, c_t) > 0$，

$$\mathbb{P}^g(x_t, s_t^{-i} \mid s_t^i, c_t) = \mathbb{P}^g(x_t, s_t^{-i} \mid p_t^i, c_t) \tag{10}$$

注释1。关于动态团队充分私人信息的类似定义见（Tavafoghi 等人，2022年，定义2）。该定义略微

与上文的定义*1*不同，因为 Tavafoghi 等人*(*2022年）与本文的目标不同。在附录*.1* 中，我们证明，满足定义*1* 的充分私人信息可能违反 Tavafoghi 等人*(*2022年）中定义2的条件（ii）。在Tavafoghi 等人*(*2022年）中，对私人信息（和公共信息）的压缩不得造成性能损失，也就是说，我们必须能够确定基于压缩后的私人信息和公共信息的全局最优团队策略组合。本文的目标是确定基于压缩信息的贝叶斯纳什均衡策略组合，并按顺序计算这些策略组合（如果存在）。我们不关心压缩信息后可能失去的均衡；因此，我们不需要满足 Tavafoghi 等人*(*2022年）中定义2的条件（ii）。

定义 1 刻画了代理人私人信息的一组压缩。下文将证明，充分私人信息 $S_t^i, i \in \mathcal{N}$, $t \in \mathcal{N}$ 足以在具有正概率或零概率的实现值的信息集上构建信念系统。令 $\tilde{g}^i$ 表示一种均匀策略，即对代理人 $i \in \mathcal{N}$的每个行动赋予相同的概率。下文将证明，当代理人 i 的私人信息 p_t^i 被充分私人信息 s_t^i替代时，信念的策略无关性（Tavafoghi 等人，2022年，定理 1）仍然成立。也就是说，由 $(\tilde{g}^i, g^{-i})$ 构造的 $\mathbb{P}^{\tilde{g}^i, g^{-i}}(x_t, x_t^{-i} \mid s_t^i, c_t)$ 捕捉了代理人 i基于 h_t^i 的信念，即使他采用任意策略 $\hat{}g^i$，而不一定采用与 g^i 或 $\tilde{g}^i$相同的策略，只要代理人 $-i$ 采用 g^{-i}。

引理 1. 对于满足 h_t^i such that $\mathbb{P}^{\tilde{g}^i, g^{-i}}(h_t^i) > 0$ 的情形，我们有 $\mathbb{P}^{\tilde{g}^i, g^{-i}}(h_t^i) > 0$，并且

$$\mathbb{P}^{\hat{g}^i, g^{-i}}(x_t, s_t^{-i} \mid h_t^i) = \mathbb{P}^{\tilde{g}^i, g^{-i}}(x_t, s_t^{-i} \mid h_t^i) = \mathbb{P}^{\tilde{g}^i, g^{-i}}(x_t, s_t^{-i} \mid s_t^i, c_t). \tag{11}$$

*Proof*注意 $\mathbb{P}^{\tilde{g}^i}(a_t^i) = 1/|\mathcal{A}_t^i|$，因此 $\mathbb{P}^{\tilde{g}^i, g^{-i}}(h_t^i) > 0$鉴于 $\mathbb{P}^g(h_t^i) > 0$。根据充分私人信息定义的第 (i) 部分以及 Tavafoghi 等人 (2022年) 中定理 1 的第 (i) 部分，我们有

$$\begin{aligned}
\mathbb{P}^{\hat{g}^i, g^{-i}}(x_t, s_t^{-i} \mid h_t^i) &= \sum_{h_t^{-i}: \zeta_t^{-i}(h_t^{-i}) = s_t^{-i}} \mathbb{P}^{\hat{g}^i, g^{-i}}(x_t, h_t^{-i} \mid h_t^i) \\
&= \sum_{h_t^{-i}: \zeta_t^{-i}(h_t^{-i}) = s_t^{-i}} \mathbb{P}^{\tilde{g}^i, g^{-i}}(x_t, h_t^{-i} \mid h_t^i) \\
&= \mathbb{P}^{\tilde{g}^i, g^{-i}}(x_t, s_t^{-i} \mid h_t^i).
\end{aligned} \tag{12}$$

此外，根据充分私人信息定义中的条件（iii），我们有

$$\mathbb{P}^{\tilde{g}^i, g^{-i}}(x_t, s_t^{-i} \mid h_t^i) = \mathbb{P}^{\tilde{g}^i, g^{-i}}(x_t, s_t^{-i} \mid s_t^i, c_t). \tag{13}$$

□

4.2 CIB 信念系统（步骤 2）

给定压缩后的私人信息，接下来我们将代理人的公共信息压缩为信念系统的形式。我们将这种压缩后的信念系统称为基于公共信息（CIB）的信念系统。与Tang 等人（2022年</style）；Ouyang 等人（2017年）类似，如果所有代理人都将 CIB 信念系统视为共同知识，并且每位代理人都能利用 CIB 信念系统和自己的压缩后私人信息，计算其关于系统状态及其他代理人充分私人信息的信念，那么该 CIB 信念系统就足以支持决策。更具体地说，只要其他代理人遵循策略组合 g^{-i}，且代理人 i采取任意策略 $\hat{}g^i$，代理人 i 就应能利用 CIB 信念系统和自己的充分私人信息s_t^i 计算出$\mathbb{P}^{\hat{g}^i,\ g^{-i}}(x_t, s_t \mid h_t^i)$。

为确定满足上述充分性要求的 CIB 信念系统，我们按以下步骤进行。首先，我们为每位代理人定义 N CIB 信念系统 $\Pi^{\psi} :=\{\Pi^{\psi,1}, \Pi^{\psi,2}, \ldots, \Pi^{\psi,N}\}$（见下文定义 2）。每个信念系统 $\Pi^{\psi,i}$ 由一系列定义在 $\mathcal{X}_t \times \mathcal{S}_t$ 上的概率质量函数组成，并根据更新规则 $\psi = (\psi^1, \psi^2, \ldots, \psi^N)$依次更新；该更新规则是代理人之间的共同知识。对于代理人 i在 $t, \pi^{\psi}{}_t{}^i$ 时可获得的公共信息的每个实现 c_t，$\mathcal{X}_t \times \mathcal{S}_t$ 从代理人 i的角度描述基于 c_t 对 $\mathcal{X}_t \times \mathcal{S}_t$ 的信念。我们希望 $\pi^{\psi}{}_t{}^i$ 与 s_t^i相结合，使代理人 i 能够形成基于充分信息的私有信念，即关于博弈当前状态的信念 $\mathbb{P}^{\hat{g}^i, g^{*-i}}(x_t, s_t \mid s_t^i, c_t)$。此外，我们希望当代理人采用基于（$S_t, \Pi_t^{\psi}$）的策略时，CIB 信念系统能够刻画博弈的当前状态。为此，我们定义充分信息型（SIB）策略组合 $\sigma := (\sigma^i, i \in \mathcal{N})$，$\sigma^i := (\sigma_t^i, t \in \mathcal{T})$ $i \in \mathcal{N}$。σ 的每个组成部分 σ_t^i 都是 s_t^i以及代理人 i在 t时的充分私人信息和 $\pi_t^{\psi} = (\pi^{\psi}{}_t{}^i, i \in \mathcal{N})$的函数（见下文定义 3）。利用 NCIB 信念系统和 SIB 策略组合 σ，我们为每个 $\pi^{\psi}{}_t{}^i$ 定义更新方程，使每个 $\pi^{\psi}{}_t{}^i$ 都与 s_t^i 以及代理人 i基于充分私人信息的信念 $\mathbb{P}^{\hat{g}^i, g^{*-i}}(x_t, s_t \mid s_t^i, c_t)$一致；该信念在第 4.1 节（定义1）中定义，并且每个 $\pi^{\psi}{}_t{}^i$ 都是所有代理人的共同知识（见下文定义 4）。下面给出（正式的）定义。

定义2（基于公共信息（CIB）的信念系统）。给定一组更新函数 $\psi = \{\psi_t^i, i \in \mathcal{N}, t \in \mathcal{T}\}$，这些函数是 N 代理人之间的共同知识，依次定义

$$\Pi_t^{\psi,i} = \psi_t^i(\Pi_{t-1}^{\psi}, Z_t), i \in \mathcal{N}, t \in \mathcal{T} \tag{14}$$

其中

$$\Pi_t^{\psi} := \begin{bmatrix} \Pi_t^{\psi,1} \\ \vdots \\ \Pi_t^{\psi,N} \end{bmatrix}, t \in \mathcal{T} \tag{15}$$

$$\Pi_0^{\psi} := \begin{bmatrix} \mu_0 \\ \vdots \\ \mu_0 \end{bmatrix} \tag{16}$$

序列 $\Pi_{1:T}^{\psi} = (\Pi_1^{\psi}, \Pi_2^{\psi}, \ldots, \Pi_T^{\psi})$ 定义了一个 CIB信念系统； $\Pi_{t}^{\psi,\ i}$ 表示从代理人 i 的视角，基于 C_t 对 $\mathcal{X}_t \times \mathcal{S}_t$ 的 CIB 信念。

定义 3（SIB策略）。给定一个 CIB信念系统 $\Pi_{1:T}^{\psi}$，我们通过以下映射定义一个基于充分信息（SIB）的策略组合 $\sigma := (\sigma^1, \sigma^2, \ldots, \sigma^N)$, $\sigma^i := (\sigma_1^i, \sigma_2^i, \ldots, \sigma_T^i)$

$$\sigma_t^i : \mathcal{S}_t^i \times [\Delta(\mathcal{X}_t \times \mathcal{S}_t)]^N \to \Delta(\mathcal{A}_t), t = 1, 2, \ldots, i = 1, 2, \ldots, N. \tag{17}$$

根据定义 2 和3 ，我们提出一组条件，个体 CIB 信念系统 $(\Pi_{t}^{\psi,\ i}, t \in \mathcal{T})$ 必须满足这些条件，以确保每个代理人 i 都能利用 Π_t^{ψ} 和 S_t^i，在所有其他代理人 $-i$ 采用 SIB 策略 σ^{-i}时，根据 (X_t, S_t) 形成自己对博弈当前状态的（私人）信念。这组条件描述了 $\Pi_{t}^{\psi,\ i}$ 的序贯更新规则；该更新规则取决于新的公共观测在 t 处是否符合代理人所采用的策略。

定义 4（一致的CIB信念系统）。考虑一个 SIB策略组合 σ。令 $F_t^i(x_{t+1}, s_{t+1}, z_{t+1})(\pi_t^{\psi}; \sigma_t^{-i})$ 表示关于 $(x_{t+1}, s_{t+1}, z_{t+1})$ 的 CIB 信念，该信念通过递归地假设以下条件构造：(i) (x_t, s_t) 按照 $\pi^{\psi,\ i}_{t}$ 分布；(ii) 代理人 i 在 t 采用均匀策略 $\tilde{g}^i$ （即以相同概率选择每个行动 $a_t^i \in \mathcal{A}_t^i$ ）；(iii) 代理人 $-i$ 根据 σ_t^{-i} 行动。也就是说，

$$F_0^i(x_1, s_1, z_1) = \sum_{y_1} \left[\mathbb{P}\{z_1, y_1 \mid x_1\} \mu_0(x_1) \left(\prod_j \mathbb{I}\{s_1^j = \phi_1^j(z_1, y_1^j)\} \right) \right] \tag{18}$$

在 $t = 1$，且对于 $t \geq 1$。

$$\begin{aligned} & F_t^i(x_{t+1}, s_{t+1}, z_{t+1})(\pi_t^{\psi};\ \sigma_t^{-i}) \\ = & \sum_{y_{t+1}, x_t, s_t, a_t} \left[\mathbb{P}\{z_{t+1}, y_{t+1}, x_{t+1} \mid x_t, a_t\} \left(\prod_j \mathbb{I}\{s_{t+1}^j = \phi_{t+1}^j(s_t^j, z_{t+1}, y_{t+1}^j, a_t^j)\} \right) \right. \\ & \left. \left(\frac{1}{\mid A_t^i \mid} \prod_{j \neq i} \sigma_t^j(a_t^j)(\pi_t^{\psi}, s_t^j) \right) \pi_t^{\psi, i}(x_t, s_t) \right] \end{aligned} \tag{19}$$

我们定义更新规则 $\psi^{\sigma} = (\psi^{\sigma,\ i}_{t}, i \in \mathcal{N}, t \in \mathcal{T})$以及相应的 CIB信念系统 $\Pi_{1:T}^{\psi^{\sigma}}$ 如下。对于任意 t

(i) 如果 $\sum_{\hat{x}_{t+1},\ \hat{s}_{t+1}} F_t^i(\hat{x}_{t+1}, \hat{s}_{t+1}, z_{t+1})(\pi_t^{\psi^\sigma};\ \sigma_t^{-i}) > 0$（即从代理人 i的角度来看，新的公共观测 z_{t+1} 是可行的），那么 $\pi^{\psi^\sigma,}{}_{t+1}{}^{i}$ 可以递归地更新为

$$\pi_{t+1}^{\psi^\sigma,i}(x_{t+1}, s_{t+1}) = \frac{F_t^i(x_{t+1}, s_{t+1}, z_{t+1})(\pi_t^{\psi^\sigma};\ \sigma_t^{-i})}{\sum_{\hat{x}_{t+1},\hat{s}_{t+1}} F_t^i(\hat{x}_{t+1}, \hat{s}_{t+1}, z_{t+1})(\pi_t^{\psi^\sigma};\ \sigma_t^{-i})}, \tag{20}$$

通过贝叶斯法则。

(ii) 如果 $\sum_{\hat{x}_{t+1},\hat{s}_{t+1}} F_t^i(\hat{x}_{t+1}, \hat{s}_{t+1}, z_{t+1})(\pi_t^{\psi^\sigma};\ \sigma_t^{-i}) = 0$（即从代理人 i 的角度来看，新的公共观测 z_{t+1} 不可行），那么更新规则为

$$\pi_{t+1}^{\psi^\sigma,i}(x_{t+1}, s_{t+1}) = \frac{1}{|\mathcal{X}_{t+1} \times \mathcal{S}_{t+1}|}. \tag{21}$$

根据（20）和（21），我们可以写成

$$\Pi_{t+1}^{\psi^\sigma,i} = \psi_{t+1}^{\sigma,i}(\Pi_t^{\psi^\sigma}, Z_{t+1}). \tag{22}$$

$$\Pi_{t+1}^{\psi^\sigma} = \psi_{t+1}^{\sigma}(\Pi_t^{\psi^\sigma}, Z_{t+1}). \tag{23}$$

此外，对于所有 $i \in \mathcal{N}$，每个代理人都可以确定 $\sum_{\hat{x}_{t+1},\hat{s}_{t+1}} F_t^i(\hat{x}_{t+1}, \hat{s}_{t+1}, z_{t+1})(\pi_t^{\sigma^\psi};\ \sigma_t^{-i})$ 是正数还是零；因此，每个代理人都知道代理人 i 如何根据 σ_t^i、z_{t+1}、σ_t^{-i} 和 ψ^σ 计算 $\pi^{\psi^\sigma}{}_{,t+1}{}^{i}$。因此，$\pi^{\psi^\sigma,}{}_{t}{}^{i}$（从而 $\pi_t^{\psi^\sigma}$）是所有代理人的共同知识。我们将 $\Pi_{1:T}^{\psi^\sigma}$ 称为与 SIB策略组合 σ 一致的 CIB信念系统。

注 2. 由于充分私人信息是代理人的可用信息的函数，因此，SIB策略 σ_t^i 对应于策略 $g^{i,}{}_{t}{}^{\sigma}$ 由 $g^{i,}{}_{t}{}^{\sigma}(h_t^i) := \sigma_t^i(\zeta_t^i(h_t^i)\ \pi_t^{\psi^\sigma})$ 给出。因此，在本文余下部分中，我们采用以下约定：$\mathbb{P}^\sigma(\cdot) = \mathbb{P}^{g^\sigma}(\cdot)$ 和 $\mathbb{E}^\sigma[\cdot] = \mathbb{E}^{g^\sigma}[\cdot]$。

注释 3。对于由 (22)-(23) 定义的更新规则$\psi_t^\sigma, t \in \mathcal{T}$，有许多其他设定也能产生一致的CIB信念系统，即能够确保以下条件的CIB信念系统：(i) 在代理人 $-i$ 按照 σ^{-i}进行策略时，代理人 i 可以将其充分私人信息S_t^i 纳入其CIB信念 $\Pi^{\psi^\sigma}{}_{,t}{}^{i}$，从而形成其关于 (X_t, S_t^{-i}) 的私有信念；(ii) 即使代理人 i采用不同于 σ^i 的任意策略 $\hat{}g^i$，其按照 i形成的私有信念仍与在其完整历史 H_t^i 条件下关于 (X_t, S_t^{-i}) 的概率分布相同。一种这样的替代更新规则示例如下：当$\sum_{\hat{x}_{t+1},\hat{s}_{t+1}} F_t^i(\hat{x}_{t+1}, \hat{s}_{t+1}, z_{t+1})(\pi_t^{\psi^\sigma};\ \sigma_t^{-i}) > 0$时，由 (20)（贝叶斯法则）描述；当 $\sum_{\hat{x}_{t+1},\hat{s}_{t+1}} F_t^i(\hat{x}_{t+1}, \hat{s}_{t+1}, z_{t+1})(\pi_t^{\psi^\sigma};\ \sigma_t^{-i}) = 0$时，则在 $X_{t+1} \times S_{t+1}$ 上使用任意概率质量函数 $\pi^{\psi^\sigma}{}_{,t+1}{}^{i}(\cdot,\cdot)$。

定义 4 确保代理人 i 可以将其充分私人信息 S_t^i 纳入其CIB信念 Π^{ψ^σ}, $_t$ $_i$ ，从而形成其关于 (X_t, S_t^{-i}) 的信念，前提是代理人 $-i$ 按照 σ^{-i}进行策略。此外，即使该代理人采用不同于 σ^i 的任意策略 $\hat{}g^i$ ，这一信念仍足以计算在其完整历史 H_t^i 条件下关于 (X_t, S_t^{-i}) 的概率分布。下面我们使用记号$\mathbb{P}^{\hat{g}^i,\ \sigma^{-i},\ \psi^\sigma}(\cdot)$，将上述讨论形式化为引理 2 ： 该记号表示代理人 i 采用 $\hat{}g^i$ 、代理人 $-i$ 采用 $g^{-i,\ \sigma}(h_t^{-i}) = \sigma_t^{-i}(\zeta_t^{-i}(h_t^{-i})\ \pi_t^{\psi^\sigma})$，并使用更新规则 ψ^σ 时所得到的信念。

引理2. 考虑一个 SIB策略组合 σ，以及与之相关的一致 CIB信念系统 $\Pi_t^{\psi^\sigma}$ 。假设 $(x_t, h_t^i, h_t^{-i}$ 是在 $(\hat{}g^i, \sigma^{-i})$ 下以正概率实现的，其中 $\hat{}g^i$ 表示代理人i的任意策略。令 $s_t^i = \zeta_t^i(h_t^i)$ 和 $s_t^{-i} = \zeta_t^{-i}(h_t^{-i})$ 为相应的充分私人信息。那么，代理人 i 在时刻 t 的信念可利用 $\pi_t^{\psi^\sigma}$ 计算如下

$$\mathbb{P}^{\hat{g}^i,\sigma^{-i},\psi^\sigma}(x_t, s_t^{-i} \mid h_t^i) = \frac{\pi_t^{\psi^\sigma,i}(x_t, s_t)}{\sum_{s_t^{-i},x_t} \pi_t^{\psi^\sigma,i}(x_t, s_t^i, s_t^{-i})} \tag{24}$$

Proof 根据引理1 可得

$$\mathbb{P}^{\hat{g}^i,\sigma^{-i},\psi^\sigma}(x_t, s_t^{-i} \mid h_t^i) = \mathbb{P}^{\tilde{g}^i,\sigma^{-i},\psi^\sigma}(x_t, s_t^{-i} \mid h_t^i). = \mathbb{P}^{\tilde{g}^i,\sigma^{-i},\psi^\sigma}(x_t, s_t^{-i} \mid c_t, s_t^i). \tag{25}$$

根据贝叶斯法则可得

$$\mathbb{P}^{\tilde{g}^i,\sigma^{-i},\psi^\sigma}(x_t, s_t^{-i} \mid c_t, s_t^i) = \frac{\mathbb{P}^{\tilde{g}^i,\sigma^{-i},\psi^\sigma}(x_t, s_t \mid c_t)}{\mathbb{P}^{\tilde{g}^i,\sigma^{-i},\psi^\sigma}(s_t^i \mid c_t)} = \frac{\pi_t^{\psi^\sigma,i}(x_t, s_t)}{\sum_{s_t^{-i},x_t} \pi_t^{\psi^\sigma,i}(x_t, s_t^i, s_t^{-i})}. \tag{26}$$

(25) 和 (26) 的结合证明了引理 2。 □

注4. 假设 $X_t = (X_t^1, X_t^2, \ldots, X_t^N)$，且满足条件独立性质，即对于任意策略组合 g $\mathbb{P}^g(x_t, s_t \mid c_t) = \prod_i \mathbb{P}^{g^i}(x_t^i, s_t^i \mid c_t)$。那么可以证明，对于任意 i，

$$\pi_t^{\psi^\sigma,i}(x_t, s_t) = \prod_j \pi^{\psi^\sigma,i}(x_t^j, s_t^j) = \mathbb{P}^{\tilde{g}_t^i}(x_t^i, s_t^i \mid c_t) \prod_{j \neq i} \mathbb{P}^{\sigma^j}(x_t^j, s_t^j \mid c_t)$$

因此，对于 Tang 等人（2022）和 Ouyang 等人（2017）中具有条件独立性质的情形，可以使用简化信念 $\mathbb{P}^{\tilde{g}_t^i}(x_t^i, s_t^i \mid c_t)$ 和 $\mathbb{P}^{\sigma^j}(x_t^j, s_t^j \mid c_t)$ 作为压缩公共信息，来计算 CIB 信念 $\pi^{\psi^\sigma}{}_{,t}{}^{\ i}(x_t, s_t)$。当代理人的行动不可观测时，Tang 等人（2022）和 Ouyang 等人（2017）模型中系统组件之间的条件独立性可能会丧失。

5 序贯分解（步骤3）

本节将介绍博弈的序贯分解，即一种向后归纳的序贯程序；如果该程序的每一步都有解，它便能确定下文定义的充分信息基础贝叶斯纳什均衡（SIB-BNE）。具体安排如下：我们首先确立一项关键性质——最佳响应闭合性（第 5.1 节）；然后利用这一性质对博弈进行序贯分解（第 5.2 节）

定义 5（SIB-BNE）。考虑一个 SIB策略组合 $\sigma^* =(\sigma^{*1}, \sigma^{*2}, \ldots, \sigma^{*n})$ 及其对应的一致更新规则 ψ^{σ^*}。若 SIB策略组合 σ^* 是该动态博弈的贝叶斯纳什均衡，则称其为 SIB-BNE。也就是说，对所有 $i \in \mathcal{N}$，

$$\mathbb{E}^{\hat{g}^i, \sigma^{*-i}, \psi^{\sigma^*}}\{U^i(X_{1:T}, A_{1:T})\} \leq \mathbb{E}^{\sigma^*, \psi^{\sigma^*}}\{U^i(X_{1:T}, A_{1:T})\},$$
$$\textit{for all strategies (not necessarily SIB strategies) } \hat{g}^i. \tag{27}$$

5.1 最佳响应的闭性

本小节的核心结果将在以下定理中给出。

定理 1. 考虑一个固定且已知的 SIB策略组合 σ 及其对应的更新规则 ψ^σ。假设代理人 $-i$ 使用 σ^{-i}，且满足 ψ^σ。那么，存在一个使用 ψ^σ 的 SIB 策略 $\hat{}\sigma^i$，它在 ψ^σ 下是对 σ^{-i} 的最佳响应。

证明基于以下引理：3、4和5，我们将在下文陈述并证明这些引理。

引理 3. 考虑一个 SIB策略组合 σ、相应的更新规则 ψ^σ，以及一致的CIB信念系统 $\Pi_{1:T}^{\psi^\sigma}$。

如果代理人 $-i$ 按照 SIB 策略 σ^{-i} 行事，并使用更新规则 ψ^σ，那么代理人 i 的最佳响应问题是一个部分可观测马尔可夫决策过程，其状态和观测过程分别为

$$\tilde{X}_t = (S_t, \Pi_t^{\psi^\sigma}, X_t), t \in \mathcal{T} \tag{28}$$
$$\tilde{Y}_t = (Y_t^i, Z_t), t \in \mathcal{T} \tag{29}$$

，即时效用为

$$\tilde{u}_t^i(\tilde{X}_t, A_t^i) = \sum_{a_t^{-i}} \Big(\prod_{j \neq i} \sigma_t^j(a_t^j \mid S_t^j, \Pi_t^{\psi^\sigma})\Big) u_t^i(X_t, a_t^{-i}, A_t^i), t \in \mathcal{T} \tag{30}$$

引理3 的断言可直接由引理4 和5推出。

引理 4. 考虑 SIB策略组合 σ 及其对应的更新规则 ψ^σ。假设代理人 $-i$ 按照 SIB 策略 σ^{-i} 行事，使用 ψ^σ

而代理人 i 遵循任意策略 $\hat{}g^i$ （不一定是 SIB 策略）。那么

$$\mathbb{P}^{\hat{g}^i,\sigma^{-i},\psi^\sigma}(\tilde{x}_{t+1},\tilde{y}_{t+1}\mid\tilde{x}_{1:t},\tilde{y}_{1:t},a^i_{1:t})=\mathbb{P}^{\hat{g}^i\sigma^{-i},\psi^\sigma}(\tilde{x}_{t+1},\tilde{y}_{t+1}\mid\tilde{x}_t,a^i_t)\tag{31}$$

Proof 下一状态和观测 $\tilde{x}_{t+1}$、$\tilde{y}_{t+1}$ 的概率可通过以下方式计算

$$\begin{aligned}
&\mathbb{P}^{\hat{g}^i,\sigma^{-i},\psi^\sigma}(\tilde{x}_{t+1},\tilde{y}_{t+1}\mid\tilde{x}_{1:t},\tilde{y}_{1:t},a^i_{1:t})\\
=&\mathbb{P}^{\hat{g}^i,\sigma^{-i},\psi^\sigma}(x_{t+1},\pi^{\psi^\sigma}_{t+1},s_{t+1},y^i_{t+1},z_{t+1}\mid x_{1:t},\pi^{\psi^\sigma}_{1:t},s_{1:t},y^i_{1:t},z_{1:t},a^i_{1:t})\\
=&\sum_{y^{-i}_{t+1},a^{-i}_t}\mathbb{P}^{\hat{g}^i,\sigma^{-i},\psi^\sigma}(x_{t+1},\pi^{\psi^\sigma}_{t+1},s_{t+1},y_{t+1},z_{t+1},a^{-i}_t\mid x_{1:t},\pi^{\psi^\sigma}_{1:t},s_{1:t},y^i_{1:t},z_{1:t},a^i_{1:t})\\
=&\sum_{y^{-i}_{t+1},a^{-i}_t}\Big(\prod_j\mathbb{1}(s^j_{t+1}=\phi^j_{t+1}(s^j_t,y^j_{t+1},z_{t+1},a^j_t))\Big)\mathbb{P}\{z_{t+1},y_{t+1},x_{t+1}\mid x_t,a_t\}\\
&\qquad\mathbb{1}(\pi^{\psi^\sigma}_{t+1}=\psi^\sigma_{t+1}(\pi^{\psi^\sigma}_t,z_{t+1}))\Big(\prod_{j\neq i}\sigma^j_t(a^j_t\mid s^j_t,\pi^{\psi^\sigma}_t)\Big)
\end{aligned}\tag{32}$$

其中，最后一个等式由系统动态、定义 1 的定义 4 以及代理人 $-i$ 的 SIB 策略形式得出。由于（32）右端仅取决于（$\tilde{x}_t$、a^i_t），我们得出

$$\mathbb{P}^{\hat{g}^i,\sigma^{-i},\psi^\sigma}(\tilde{x}_{t+1},\tilde{y}_{t+1}\mid\tilde{x}_{1:t},\tilde{y}_{1:t},a^i_{1:t})=\mathbb{P}^{\hat{g}^i,\sigma^{-i},\psi^\sigma}(\tilde{x}_{t+1},\tilde{y}_{t+1}\mid\tilde{x}_t,a^i_t)\tag{33}$$

□

~ ~引理 4 表明，$\{X_t,Y_t,t\in\mathcal{T}\}$ 在给定$\{A^i_t,t\in\mathcal{T}\}$的条件下是一个马尔可夫过程

引理 5. 考虑一个 SIB策略组合 σ 及相应的更新规则 ψ^σ。假设代理人 $-i$使用 ψ^σ 遵循 SIB 策略 σ^{-i}，而代理人 i遵循任意策略 $\hat{}g^i$ （不一定是 SIB 策略）。则存在效用函数 $\tilde{u}^i_t$，使得 $\mathbb{E}^{\hat{g}^i,\sigma^{-i},\psi^\sigma}[\tilde{u}^i_t(\tilde{X}_t,A^i_t)]=\mathbb{E}^{\hat{g}^i,\sigma^{-i},\psi^\sigma}[u^i_t(X_t,A_t)]$对所有 $t\in\mathcal{T}$均成立。

Proof 回想一下，$\tilde{X}_t=(S_t,\Pi^{\psi^\sigma}_t,X_t)$。然后

$$\begin{aligned}
&\mathbb{E}^{\hat{g}^i,\sigma^{-i},\psi^\sigma}[u^i_t(X_t,A_t)]\\
=&\mathbb{E}^{\hat{g}^i,\sigma^{-i},\psi^\sigma}[u^i_t(X_t,A^{-i}_t,A^i_t)]\\
=&\mathbb{E}^{\hat{g}^i,\sigma^{-i},\psi^\sigma}\big[\mathbb{E}^{\hat{g}^i,\sigma^{-i},\psi^\sigma}[u^i_t(X_t,A^{-i}_t,A^i_t)\mid\tilde{X}_t,A^i_t]\big]\\
=&\mathbb{E}^{\hat{g}^i,\sigma^{-i},\psi^\sigma}\Big[\sum_{a^{-i}_t}\mathbb{P}^{\hat{g}^i,\sigma^{-i},\psi^\sigma}(a^{-i}_t\mid S_t,\Pi^{\psi^\sigma}_t,X_t,A^i_t)u^i_t(X_t,a^{-i}_t,A^i_t)\Big]\\
=&\mathbb{E}^{\hat{g}^i,\sigma^{-i},\psi^\sigma}\Big[\sum_{a^{-i}_t}\Big(\prod_{j\neq i}\sigma^j_t(a^j_t\mid S^j_t,\Pi^{\psi^\sigma}_t)\Big)u^i_t(X_t,a^{-i}_t,A^i_t)\Big]
\end{aligned}\tag{34}$$

因此，我们通过定义来证明该引理的断言

$$\tilde{u}_t^i(\tilde{X}_t, A_t^i) = \sum_{a_t^{-i}} \big(\prod_{j \neq i} \sigma_t^j(a_t^j \mid S_t^j, \Pi_t^{\psi^\sigma}) \big) u_t^i(X_t, a_t^{-i}, A_t^i)] \tag{35}$$

□

Proof of Theorem 1 根据引理 3 ，我们得出代理人 i 对 σ^{-i} 的最佳响应是一个状态为 $\tilde{X}_t$ 的部分可观测马尔可夫决策过程（Kumar 和 Varaiya， 1986 年，第 6 章）。根据部分可观测马尔可夫决策过程理论，我们知道：（i）在可用信息 $\tilde{X}_t = (S_t, \Pi_t^{\psi^\sigma}, X_t)$ 条件下关于状态 h_t^i 的信念是该代理人的信息状态；（ii）对于每个 $t \in \mathcal{T}$，都存在一个代理人 i 的最优策略，它是时刻 t的信息状态的函数。现在我们证明，S_t^i, $\Pi_t^{\psi^\sigma}$) 是代理人 i 在时刻 t， $t \in \mathcal{T}$的信息状态。
注意到 $S_{t+1}^i = \phi_t^i(S_t^i, Y_{t+1}^i, Z_{t+1}, A_t^i)$ 来自定义 1的第（ii）部分，而 $\Pi_{t+1}^{\psi^\sigma} = \psi_{t+1}^\sigma(\Pi_t^{\psi^\sigma}, Z_{t+1})$ 来自式（23）。

因此，我们只需证明，对于任意策略 $\hat{g}^i$ 和任意实现 h_t^i ，只要 $\mathbb{P}^{\hat{g}^i, \sigma^{-i}, \psi^\sigma}(h_t^i) > 0$ ，以下等式成立：

$$\mathbb{P}^{\hat{g}^i, \sigma^{-i}, \psi^\sigma}(s_t, \pi_t^{\psi^\sigma}, x_t \mid h_t^i) = \mathbb{P}^{\hat{g}^i, \sigma^{-i}, \psi^\sigma}(s_t, \pi_t^{\psi^\sigma}, x_t \mid s_t^i, \pi_t^{\psi^\sigma}) \tag{36}$$

为此，我们注意到 $s_t^i, \pi_t^{\psi^\sigma}$ 对代理人 i而言是完全已知的。此外，根据充分私人信息的定义和引理 2 ，我们有

$$\mathbb{P}^{\hat{g}^i, \sigma^{-i}, \psi^\sigma}(s_t^{-i}, x_t \mid h_t^i) = \frac{\pi_t^{\psi^\sigma, i}(s_t, x_t)}{\sum_{s_t^{-i}, x_t} \pi_t^{\psi^\sigma, i}(s_t^i, s_t^{-i}, x_t)}, \tag{37}$$

它是 $(s_t^i, \pi_t^{\psi^\sigma})$ 的函数。因此，

$$\mathbb{P}^{\hat{g}^i, \sigma^{-i}, \psi^\sigma}(s_t, \pi_t^{\psi^\sigma}, x_t \mid h_t^i) = \mathbb{1}(s_t^i = \zeta_t^i(h_t^i)) \mathbb{1}(\pi_t^{\psi^\sigma} = \gamma^{\psi^\sigma}(h_t^i)) \, \mathbb{P}^{\hat{g}^i, \sigma^{-i}, \psi^\sigma}(s_t^{-i}, x_t \mid p_t^i, c_t) \tag{38}$$

其中 $\gamma^{\psi^\sigma}(h_t^i) = \psi_t^\sigma(\psi_{t-1}^\sigma, \cdots)$ 是从 1 到 ψ^σ 的复合 t。然后，由于 (36) 和 (37) 以及 (38)，方程 $(S_t^i, \Pi_t^{\psi^\sigma})$ $t \in \mathcal{T}$是代理人 i 最佳响应问题的信息状态，定理 □1 的断言成立。

根据定理 1，与定义5等价的 SIB 贝叶斯纳什均衡定义如下

定义6（SIB 贝叶斯纳什均衡的等价定义）。考虑一个 SIB策略组合 $\sigma^* = (\sigma^{*1}, \sigma^{*2}, \ldots, \sigma^{*n})$ 及其对应的一致更新规则 ψ^{σ^*}。若对所有 $i \in \mathcal{N}$，SIB 策略组合 σ^* 都是 SIB 贝叶斯纳什均衡，

$$\mathbb{E}^{\sigma^i, \sigma^{*-i}, \psi^{\sigma^*}}\{U^i(X_{1:T}, A_{1:T})\} \leq \mathbb{E}^{\sigma^*, \psi^{\sigma^*}}\{U^i(X_{1:T}, A_{1:T})\} \tag{39}$$

对所有 $\sigma^i \in \Lambda^i$ ，其中 Λ^i 是代理人 i 的 SIB策略组合集合。

引理 3-5 和定理 1 的推论如下。考虑一个 SIB策略组合 σ、相应的更新规则 ψ^σ 以及一致的CIB信念系统 $\Pi_{1:T}^{\psi^\sigma}$；如果代理人 $-i$ 按照 σ^{-i} 行动，那么代理人 i 的最佳响应可由以下动态规划确定

$$\breve{V}_{T+1}^i(\cdot,\cdot) = 0 \text{ for all } i \tag{40}$$

$$\begin{aligned}\breve{V}_t^i(\pi_t^{\psi^\sigma}, s_t^i) = \max_{\tilde{\sigma}_t^i \in \Lambda_t^i} \mathbb{E}^{\tilde{\sigma}_t^i, \sigma_t^{-i}, \psi^\sigma}\{u_t^i(X_t, A_t) + \breve{V}_{t+1}^i(\psi_{t+1}^\sigma(\pi_t^{\psi^\sigma}, Z_{t+1}), S_{t+1}^i) \mid s_t^i\},\\ \forall \pi_t^{\psi^\sigma} \in \Delta(\mathcal{X}_t \times \mathcal{S}_t)^N, \forall s_t^i \in \mathcal{S}_s^i, t \in \mathcal{T}\end{aligned} \tag{41}$$

其中，Λ_t^i 是代理人 i 在时刻 t的 SIB策略集合。

5.2 序贯分解

给定一组价值函数 $V_{t+1} = \{V_{t+1}^i : \mathbf{\Pi}_{t+1} \times \mathcal{S}_{t+1}^i \to \mathbb{R},\ i \in \mathcal{N}\}$、一个 SIB策略组合 σ、由（23）定义的相应更新规则 ψ_{t+1}^σ，以及一致的 CIB 信念 $\pi_t^{\psi^\sigma}$，定义阶段博弈 $G_t(V_{t+1}, \pi_t^{\psi^\sigma})$ 如下。

(i) 有 N 个代理人。(ii) 系统状态为 X_t。(iii) 每个代理人 i观察私人信息 S_t^i 和公共信息 $\pi_t^{\psi^\sigma}$。(iv) 代理人 i关于状态 X_t 和其他代理人的私人信息 S_t^{-i}的信念由 $\pi_t^{\psi^\sigma, i}(x_t, s_t^{-i})$ 给出，即，

$$\pi_t^{\psi^\sigma, i}(x_t, s_t^{-i}) \in \Delta(\mathcal{X}_t \times \mathcal{S}_t^{-i}). \tag{42}$$

(v) 每位代理人 i 根据其可用信息选择行动 A_t^i；令 $\hat{\sigma}_t^i$表示代理人 i在此阶段博弈中的策略；然后，

$$\mathbb{P}^{\hat{\sigma}_t, \psi^\sigma}(A_t^i = a_t^i \mid s_t^i, \pi_t^{\psi^\sigma}) = \hat{\sigma}_t^i(a_t^i \mid s_t^i, \pi_t^{\psi^\sigma}). \tag{43}$$

(vi) 每位代理人 i 的效用为

$$U_{G_t(V_{t+1}, \pi_t^{\psi^\sigma})}^i = u_t^i(X_t, A_t) + V_{t+1}^i(\psi_{t+1}^\sigma(\pi_t^{\psi^\sigma}, Z_{t+1}), S_{t+1}^i) \tag{44}$$

其中，给定 (X_t, S_t, A_t) 时，(Z_{t+1}, S_{t+1}^i) 服从条件概率$\sum_{x_{t+1}, s_{t+1}^{-i}} \mathbb{P}(z_{t+1}, x_{t+1}, s_{t+1} \mid x_t, s_t, a_t)$，且条件概率$\mathbb{P}(z_{t+1}, x_{t+1}, s_{t+1} \mid x_t, s_t, a_t)$由下式给出

$$\begin{aligned}&\mathbb{P}(z_{t+1}, x_{t+1}, s_{t+1} \mid x_t, s_t, a_t)\\ =&\sum_{y_{t+1}} \mathbb{P}\{x_{t+1} \mid x_t, a_t\}\mathbb{P}\{z_{t+1}, y_{t+1} \mid x_{t+1}, a_t\}\\ &\left(\prod_j \mathbb{I}\{s_{t+1}^j = \phi_{t+1}^j(s_t^j, z_{t+1}, y_{t+1}^j, a_t^j)\}\right)\end{aligned} \tag{45}$$

(vii) 给定阶段博弈的一个策略组合 $\hat{}\sigma_t$，每个博弈参与者 i 的期望效用为

$$\begin{aligned}&\mathbb{E}^{\hat{\sigma}_t,\psi^\sigma}[U^i_{G_t(V_{t+1},\pi_t^{\psi^\sigma})} \mid s_t^i]\\ =&\sum_{x_t,s_t^{-i},a_t,z_{t+1},x_{t+1},s_{t+1}} \pi_t^{\psi^\sigma,i}(x_t,s_t^{-i})\prod_j \hat{\sigma}_t^j(a_t^i \mid s_t^i,\pi_t^{\psi^\sigma})\,\mathbb{P}(z_{t+1},x_{t+1},s_{t+1}\mid x_t,s_t,a_t)\\ &(u_t^i(x_t,a_t)+V_{t+1}^i(\psi_{t+1}^\sigma(\pi_t^{\psi^\sigma},z_{t+1}),s_{t+1}^i))\end{aligned} \tag{46}$$

请注意，阶段博弈 $G_t(V_{t+1},\pi_t^{\psi^\sigma})$ 中的所有随机变量不一定都与原始动态博弈中的对应变量相同，因为每个代理人 i 都可以选择任意的 SIB策略 $\hat{}\sigma_t^i$，而该策略可能不同于 SIB策略组合 σ所指定的策略。如果所有代理人都遵循 σ，则阶段博弈中的随机变量将与原始博弈中的对应变量一致。

定理 2 （序贯分解）。考虑一个 SIB策略组合 $\sigma=\{\sigma_t,t\in\mathcal{T}\}$ 以及相应的更新规则 $\psi^\sigma=\{\psi_t^\sigma,t\in\mathcal{T}\}$，其定义为(22)-(23)。定义

$$V_{T+1}^i(\cdot,\cdot)=0 \ \textit{for all}\ i \tag{47}$$

$$V_t^i(\pi^{\psi_t^\sigma},s_t^i)=\mathbb{E}^{\sigma_t,\psi^\sigma}[U^i_{G_t(V_{t+1},\pi_t^{\psi^\sigma})}\mid s_t^i] \tag{48}$$

其中，(48) 的右侧由 (46)给出。若对所有 $t\in\mathcal{T}$，都存在一个 SIB策略组合 $\hat{}\sigma_t$，使得 $\hat{}\sigma_t$ 是阶段博弈 $G_t(V_{t+1},\ \pi_t^{\psi^\sigma})$的贝叶斯纳什均衡，即

$$\mathbb{E}^{\hat{\sigma}_t^i,\hat{\sigma}_t^{-i},\psi^\sigma}[U^i_{G_t(V_{t+1},\pi_t^{\psi^\sigma})}\mid s_t^i]=\max_{\tilde{\sigma}_t^i\in\Lambda_t^i}\mathbb{E}^{\tilde{\sigma}_t^i,\hat{\sigma}_t^{-i},\psi^\sigma}[U^i_{G_t(V_{t+1},\pi_t^{\psi^\sigma})}\mid s_t^i] \tag{49}$$

f或对所有 $i\in\mathcal{N}$，其中 Λ_t^i 是代理人 i 在时刻 t 的 SIB 策略集合，一个 d

$$\hat{\sigma}_t=\sigma_t, \tag{50}$$

那么，SIB策略组合 σ 是原始动态博弈的 SIB-BNE。

Proof 假设对所有 $t\in\mathcal{T}$，都存在一个 SIB策略组合$\hat{\sigma}_t=(\hat{}\sigma_t^1,\hat{}\sigma_t^2,\ldots,\hat{}\sigma_t^N)$，它是阶段博弈 $G_t(V_{t+1},\pi_t^{\psi^\sigma})$ 的贝叶斯纳什均衡。那么对所有 $\pi_t^{\psi^\sigma}\in\Delta(\mathcal{X}_t\times\mathcal{S}_t)^N,s_t^i\in$

$\mathcal{S}_s^i$

$$\begin{aligned}&\mathbb{E}^{\hat{\sigma}_t^i,\hat{\sigma}_t^{-i},\psi^\sigma}[U^i_{G_t(V_{t+1},\pi_t^{\psi^\sigma})}\mid s_t^i]\\&=\max_{\tilde{\sigma}_t^i\in\Lambda_t^i}\mathbb{E}^{\tilde{\sigma}_t^i,\hat{\sigma}_t^{-i},\psi^\sigma}[u_t^i(X_t,A_t)+V_{t+1}^i(\psi_{t+1}^\sigma(\pi_t^{\psi^\sigma},Z_{t+1}),S_{t+1}^i)\mid s_t^i].\end{aligned}\tag{51}$$

方程（51）对所有 $t\in\mathcal{T}$在 $V_{T+1}^i(\cdot,\cdot)=0$ 以及所有 $i\in\mathcal{N}$均成立。当对所有 $\sigma_t=\sigma_t,\ t\in\mathcal{T}$时，方程（51）给出，对所有 $\pi_t^{\psi^\sigma}\in\Delta(\mathcal{X}_t\times\mathcal{S}_t)^N,\ s_t^i\in\mathcal{S}_s^i$,

$$\begin{aligned}V_t^i(\pi_t^{\psi^\sigma},s_t^i)&=\mathbb{E}^{\sigma_t^i,\sigma_t^{-i},\psi^\sigma}[U^i_{G_t(V_{t+1},\pi_t^{\psi^\sigma})}\mid s_t^i]\\&=\max_{\tilde{\sigma}_t^i\in\Lambda_t^i}\mathbb{E}^{\tilde{\sigma}_t^i,\sigma_t^{-i},\psi^\sigma}[u_t^i(X_t,A_t)+V_{t+1}^i(\psi_{t+1}^\sigma(\pi_t^{\psi^\sigma},Z_{t+1}),S_{t+1}^i)\mid s_t^i]\end{aligned}\tag{52}$$

对所有 $i\in\mathcal{N}$。

根据归纳法，(52)，以及更新规则 ψ^σ 与 σ 一致这一事实可知，对所有 $i\in\mathcal{N}$和 $t\in\mathcal{T}$,

$$\mathbb{E}^{\tilde{\sigma}_{t:T}^i,\sigma_{t:T}^{-i},\psi^\sigma}[\sum_{\tau=t}^T u_\tau^i(X_\tau,A_\tau)\mid s_\tau^i]\le\mathbb{E}^{\sigma_{t:T}^i,\sigma_{t:T}^{-i},\psi^\sigma}[\sum_{\tau=t}^T u_\tau^i(X_\tau,A_\tau)\mid s_\tau^i]\tag{53}$$

于是，时刻 $t=1$ 的式 (53) 给出

$$\mathbb{E}^{\tilde{\sigma}^i,\sigma^{-i},\psi^\sigma}\{U^i(X_{1:T},A_{1:T})\}\le\mathbb{E}^{\sigma,\psi^\sigma}\{U^i(X_{1:T},A_{1:T})\}\tag{54}$$

对所有 $\tilde{\sigma}^i\in\Lambda^i$ 对所有 $i\in\mathcal{N}$。因此，策略组合 σ 是原始动态博弈的一个 SIB-BNE（参见定义 6）。

□

注5. 请注意，即使阶段博弈 $G_t(V_{t+1},\pi_t^{\psi^\sigma})$ 存在贝叶斯纳什均衡 $\hat{}\sigma_t$，也可能出现 $\hat{}\sigma_t\neq\sigma_t$。因此，每个阶段博弈 $G_t(V_{t+1},\pi_t^{\psi^\sigma})$ 都存在贝叶斯纳什均衡，并不足以证明原始动态博弈存在贝叶斯纳什均衡。

注6. 在Tang 等人 (2022年) 的模型中，当每个团队仅由一名代理人组成时，SIB 贝叶斯纳什均衡与 Tang 等人(2022年) 提出的 SPCIB 贝叶斯纳什均衡相一致，只需按照注4中所述，对信息状态作适当映射。

注7. 对于(47)-(50)所描述的序贯分解方程中的价值函数集合，未必对所有 $i\in\mathcal{N}$ 和所有 $t\in\mathcal{T}$都存在解。

注8. 在定义4中，(21) 可以有不同的定义，而不同的(21) 会导致对 ψ的不同选择。并且，对于任意一种(21)的选择，定理2的结论仍然成立。

注9. 定理2（方程(47)-(50)）中对所有 $i\in\mathcal{N},\ t\in\mathcal{T}$）定义的序贯分解方程的价值函数在 CIB 信念$\Pi_t^{\psi^\sigma}$ 中可能不连续。

6 说明性示例（步骤 4）

在第 5 节中，我们指出（参见备注7），对所有 $i \in \mathcal{N}$， $t \in \mathcal{T}$，由（47）-（50）定义的序贯分解方程可能没有解，并且由（47）-（50）定义的价值函数关于 CIB 信念 $\Pi_t^{\psi^\sigma}$ 可能不连续（参见备注 9）。本节给出一个例子，以说明并突出上述备注。在这个例子中，两阶段随机动态博弈中代理人的效用取决于参数 c。我们证明：（i）相应序贯分解方程的价值函数关于 CIB 信念 $\Pi_t^{\psi^\sigma}$ 不连续；（ii）对于某些 c取值，存在 SIB-BNE。

6.1 模型

我们考虑以下两阶段随机动态博弈。博弈中有两位玩家/代理人，爱丽丝和 Bob。在第一阶段， $t = 1$， 系统状态 X_1服从定义在 $\{-1, 1\}$ 上的分布，其中 $\mu_0(-1) = \mathbb{P}(X_1 = -1) = 0.5$ 且 $\mu_1(1) = \mathbb{P}(X_1 = 1) = 0.5$。爱丽丝能够完美观测到 X_1，即 $Y_1^{Alice} = X_1$，并采取行动 $A_1^{Alice} \in \{-1, 1\}$；Bob 无法观测到 A_1^{Alice}，且 $Y_1^{Bob} = \emptyset$。Bob 在 $t = 1$不采取行动。在第二阶段， $t = 2$， 系统状态为 $X_2 = X_1 A_1^{Alice}$。爱丽丝和 Bob 有一个公共观测 $Z_2 = X_2 A_1^{Alice} W_1 = X_1 W_1$，其中$W_1 \in \{-1, 1\}$ 且 $\mathbb{P}(Z = i \mid X_1 = i) = 1 - p = 0.8, i \in \{-1, 1\}$，并且没有私人观测，即 $Y_2^{Alice} = Y_2^{Bob} = \emptyset$。这里 $p = 0.2 = \mathbb{P}(W_1 = -1)$。Bob 在 $t = 2$采取行动。爱丽丝在 $t = 2$不采取行动。Bob 的行动 $A_2^{Bob} \in \{-1, 1\}$。爱丽丝在 $t = 1$ 和 $t = 2$ 时的收益为

$$u_1^{Alice}(X_1, A_1) = \begin{cases} c & \text{if } A_1^{Alice} = 1 \\ 0 & \text{if } A_1^{Alice} = -1 \end{cases} \tag{55}$$

and

$$u_2^{Alice}(X_2, A_2) = \begin{cases} 2 & \text{if } X_2 = 1, A_2^{Bob} = 1 \\ 1 & \text{if } X_2 = -1, A_2^{Bob} = -1 \\ 0 & \text{otherwise} \end{cases} \tag{56}$$

分别。Bob 的收益为 $u_t^{Bob}(X_t, A_t) = -u_t^{Alice}(X_t, A_t)$ $t = 1, 2$。

该博弈的信息结构为

$$H_1^{Alice} = \{X_1\} \tag{57}$$
$$H_2^{Alice} = \{X_1, A_1^{Alice}, X_2, Z_2\} \tag{58}$$
$$H_1^{Bob} = \emptyset \tag{59}$$
$$H_2^{Bob} = \{Z_2\} \tag{60}$$

其中，H_t^{Alice}、H_t^{Bob}、 $t = 1$、2 分别描述了爱丽丝和 Bob 在第 1 阶段和第 2 阶段可获取的信息。

此例的动态过程和效用函数与 Tang 等人 (2022年) 中的例子 3 相同，但 Bob 不会像（Tang 等人，2022年，例子 3）中那样观察爱丽丝的行动。

6.2 序贯分解

由于爱丽丝在两个时刻都能完美观察到状态，即 $Y_1^{Alice} = X_1$ 且 $Y_2^{Alice} = X_2$，而 Bob 没有私人信息，$S_1^{Alice} = X_1$，$S_1^{Bob} = \emptyset$ 分别是阶段 $t = 1$ 爱丽丝和 Bob 的充分私人信息，并且 $S_2^{Alice} = X_2$，$S_2^{Bob} = \emptyset$ 分别是阶段 $t = 2$ 根据定义 1，爱丽丝和 Bob 的充分私人信息。

假设 $\sigma = (\sigma_1, \sigma_2) = (\sigma_1^{Alice}, \sigma_2^{Bob})$ 是一个 SIB策略，且 ψ^σ 是相应的更新规则。这里，σ 是一个均衡策略候选，用作爱丽丝和 Bob 的策略预测。注意，对于所有 $x_1 \in \mathcal{X}_1$，$\Pi^{\psi^\sigma, 1}{}^{Alice}(x_1) = \mu_0(x_1)$，以及 $\Pi^{\psi^\sigma, 1}{}^{Bob}(x_1) = \mu_0(x_1)$。

为利用定理 2的序贯分解求得贝叶斯纳什均衡，我们首先考虑时刻 2 的阶段博弈 $G_2(0, \pi_2^{\psi^\sigma})$。由于时刻 2 只有 Bob 这位代理人采取行动，且 $S_2^{Bob} = \emptyset$，任何贝叶斯纳什均衡 σ_2 都必须满足 $G_2(0, \pi_2^{\psi^\sigma})$

$$
\begin{aligned}
\hat{\sigma}_2^{Bob} &= \arg\max_{\tilde{\sigma}_2^{Bob}} \mathbb{E}^{\tilde{\sigma}_2^{Bob}, \psi^\sigma}[u_2^{Bob}(X_2, A_2)] \\
&= \arg\max_{\tilde{\sigma}_2^{Bob}} \Big(-2\,\mathbb{P}^{\tilde{\sigma}_2^{Bob}, \psi^\sigma}(X_2 = A_2^{Bob} = 1) - \mathbb{P}^{\tilde{\sigma}_2^{Bob}, \psi^\sigma}(X_2 = A_2^{Bob} = -1) \Big) \\
&= \arg\max_{\tilde{\sigma}_2^{Bob}} \Big(-2\pi_2^{\psi^\sigma, Bob}(1)\tilde{\sigma}_2^{\psi^\sigma, Bob}(1 \mid \pi_2^{\psi^\sigma}) \\
&\qquad - (1 - \pi_2^{\psi^\sigma, Bob}(1))(1 - \tilde{\sigma}_2^{\psi^\sigma, Bob}(1 \mid \pi_2^{\psi^\sigma})) \Big)
\end{aligned} \tag{61}
$$

由（61）可知，其中一个均衡 SIB 策略为

$$
\begin{aligned}
\sigma_2^{Bob}(\pi_2^{\psi^\sigma}) &= 1, \text{ if } \pi_2^{\psi^\sigma, Bob}(1) \le 1/3, \\
\sigma_2^{Bob}(\pi_2^{\psi^\sigma}) &= 0, \text{ if } \pi_2^{\psi^\sigma, Bob}(1) > 1/3,
\end{aligned}
$$

或者等价地

$$
\sigma_2^{Bob}(\pi_2^{\psi^\sigma}) = \mathbb{1}(\pi_2^{\psi^\sigma, Bob}(1) \le 1/3) \tag{62}
$$

注意，如果 $\pi^{\psi^\sigma, 2}{}^{Bob}(1) = 1/3$，那么 $\sigma_2^{Bob}(\pi_2^{\psi^\sigma})$ 可以在 $[0, 1]$ 中取任意值，且仍然是阶段博弈的贝叶斯纳什均衡。

爱丽丝在时刻 2 的充分私人信息为 $S_2^{Alice} = X_2$。根据（$\sigma_2^{Bob}(\pi_2)$ 给出的阶段博弈均衡 SIB策略（62），爱丽丝在 $t = 2$ 时的价值函数随后根据（48）给出如下：

$$
V_2^{Alice}(\pi_2^{\psi^\sigma}, x_2) = \mathbb{E}^{\sigma_2, \psi^\sigma}[u_2^{Alice}(X_2, A_2) \mid x_2]
$$

$$= \begin{cases} 2\mathbb{1}(\pi_2^{\psi^\sigma, Bob}(1) \le 1/3) & \text{if } x_2 = 1 \\ 1 - \mathbb{1}(\pi_2^{\psi^\sigma, Bob}(1) \le 1/3) & \text{if } x_2 = -1 \end{cases} \tag{63}$$

根据时刻 $t = 2$ 的上述价值函数，我们现在考虑时刻 $t = 1$ 的阶段博弈 $G_1(V_2, \pi_1^{\psi^\sigma})$。爱丽丝在该阶段博弈中的效用如下所示。

$$U^{Alice}_{G_1(V_2, \pi_1^{\psi^\sigma})} = u_1^{Alice}(X_1, A_1) + V_2^{Alice}(\psi_2^\sigma(\pi_1, Z), X_2) \tag{64}$$

如果爱丽丝采用 SIB策略 $\tilde{\sigma}_1^{Alice}$，则根据 (46) 可计算出该阶段博弈对于 $X_1 = -1$ 和 $X_1 = 1$的期望效用，计算方式为

$$\begin{aligned} &\mathbb{E}^{\tilde{\sigma}_1^{Alice}, \psi^\sigma}[U^{Alice}_{G_1(V_2, \pi_1^{\psi^\sigma})} \mid X_1 = -1] \\ =&c\tilde{\sigma}_1^{Alice}(1 \mid -1) + \mathbb{E}^{\tilde{\sigma}_1^{Alice}, \psi^\sigma}[V_2^A(\psi_2^\sigma(\pi_1^{\psi^\sigma}, X_1W_1), X_1A_1^{Alice}) \mid X_1 = -1] \\ =&(1+c)(1-\tilde{\alpha}_1) + (3\tilde{\alpha}_1 - 1)((1-p)\mathbb{1}(q_{-1} \le 1/3) + p\mathbb{1}(q_1 \le 1/3)) \\ =:&r_{-1}^A(\tilde{\alpha}_1, q) \end{aligned} \tag{65}$$

$$\begin{aligned} &\mathbb{E}^{\tilde{\sigma}_1^{Alice}, \psi^\sigma}[U^{Alice}_{G_1(V_2, \pi_1^{\psi^\sigma})} \mid X_1 = 1] \\ =&c\tilde{\sigma}_1^{Alice}(1 \mid 1) + \mathbb{E}^{\tilde{\sigma}_1^{Alice}, \psi^\sigma}[V_2^A(\psi_2^\sigma(\pi_1^{\psi^\sigma}, X_1W_1), X_1A_1^{Alice}) \mid X_1 = 1] \\ =&1 + (c-1)\tilde{\alpha}_2 + (3\tilde{\alpha}_2 - 1)((1-p)\mathbb{1}(q_1 \le 1/3) + p\mathbb{1}(q_{-1} \le 1/3)) \\ =:&r_1^A(\tilde{\alpha}_2, q) \end{aligned} \tag{66}$$

其中，$q = (q_{-1}, q_1)$，$q_{-1} = \psi^{\sigma,\ {}_2}{}^{Bob}(\pi_1^{\psi^\sigma}, -1)(1)$ 和 $q_1 = \psi^{\sigma,\ {}_2}{}^{Bob}(\pi_1^{\psi^\sigma}, 1)(1)$ 分别是 $\{X_2 = 1\}$ 在 $Z = -1$ 和 $Z = 1$ 时的 CIB 信念 $\pi^{\psi^\sigma,\ {}_2}{}^{Bob}(1)$，而 $\tilde{\alpha} = (\tilde{\alpha}_1, \tilde{\alpha}_2)$，$\tilde{\alpha}_1 = \tilde{\sigma}_1^{Alice}(-1 \mid -1)$ $\tilde{\alpha}_2 = \tilde{\sigma}_1^{Alice}(1 \mid 1)$ 表示爱丽丝的 SIB策略 $\tilde{\sigma}_1^{Alice}$。

注意，根据定义 4 中的贝叶斯法则，在 SIB策略 σ_1^{Alice}（表示为 $\alpha_1 = \sigma_1^{Alice}(-1 \mid -1)$ 和 $\alpha_2 = \sigma_1^{Alice}(1 \mid 1)$ 下，我们有

$$q_{-1} = \psi_2^{\psi^\sigma, Bob}(\pi_1^{\psi^\sigma}, -1)(1) = \frac{\mathbb{P}^\alpha(X_2 = 1, Z = -1)}{\mathbb{P}^\alpha(Z = -1)} = \alpha_2 p + \alpha_1(1-p) \tag{67}$$

$$q_1 = \psi_2^{\psi^\sigma, Bob}(\pi_1^{\psi^\sigma}, 1)(1) = \frac{\mathbb{P}^\alpha(X_2 = 1, Z = 1)}{\mathbb{P}^\alpha(Z = 1)} = \alpha_2(1-p) + \alpha_1 p \tag{68}$$

因此，由 $\hat{\alpha}_1 = \hat{\sigma}_1^{Alice}(-1 \mid -1)$ 和 $\hat{\alpha}_2 = \hat{\sigma}_1^{Alice}(1 \mid 1)$ 表示的 SIB策略 $\hat{\sigma}_1^{Alice}$，是时刻 $t = 1$ 的阶段博弈 $G_1(V_2, \pi_1^{\psi^\sigma})$ 的贝叶斯纳什均衡，当且仅当

$$\hat{\alpha}_1 \in \arg\max_{\tilde{\alpha}_1} r_{-1}^A(\tilde{\alpha}_1, (\alpha_2 p + \alpha_1(1-p), \alpha_2(1-p) + \alpha_1 p)) \tag{69}$$

$$\hat{\alpha}_2 \in \arg\max_{\tilde{\alpha}_2} r_1^A(\tilde{\alpha}_2, (\alpha_2 p + \alpha_1(1-p), \alpha_2(1-p) + \alpha_1 p)) \tag{70}$$

因此，SIB策略 σ_1^{Alice}，由 $\alpha_1 = \sigma_1^{Alice}(-1 \mid -1)$ 和 $\alpha_2 = \sigma_1^{Alice}(1 \mid 1)$ 表示，将满足序贯分解方程 (49)-(50)，当且仅当

$$\alpha_1 \in \arg\max_{\tilde{\alpha}_1} r_{-1}^A(\tilde{\alpha}_1, (\alpha_2 p + \alpha_1(1-p), \alpha_2(1-p) + \alpha_1 p)) \tag{71}$$

$$\alpha_2 \in \arg\max_{\tilde{\alpha}_2} r_1^A(\tilde{\alpha}_2, (\alpha_2 p + \alpha_1(1-p), \alpha_2(1-p) + \alpha_1 p)) \tag{72}$$

注10。请注意，函数 $r_{-1}^A(\tilde{\alpha}_1, q)$ 和 $r_1^A(\tilde{\alpha}_2, q)$ 在 q 处不连续。因此，无法通过依赖效用函数连续性的标准方法来证明均衡的存在，一般情况下也可能不存在均衡。

6.3 即时效用条件下 SIB-BNE 的存在性。

阶段博弈 $G_1(V_2, \pi_1^{\psi^\sigma})$ 是一个具有固定 σ_1 的标准式博弈。根据注5，$G_1(V_2, \pi_1^{\psi^\sigma})$ 的一个贝叶斯纳什均衡 $\hat{}\sigma$ 可能不同于 σ_1，而 $G_1(V_2, \pi_1^{\psi^\sigma})$ 的正则贝叶斯纳什均衡存在，并不足以保证在时刻 $t = 1$ 满足 (50)。为了将标准式博弈的均衡存在性结果应用于时刻 $t = 1$ 的序贯分解，我们引入代理人 0，由其选择 q-信念 $q = (q_{-1}, q_1)$，使得 (50) 得以满足。

$\hat{}$ 形式上，我们构造爱丽丝与代理人 0 之间的扩展阶段博弈 G_1。爱丽丝选择 $\tilde{\alpha} = (\tilde{\alpha}_1, \tilde{\alpha}_2)$，代理人 0 选择 $\tilde{q} = (\tilde{q}_{-1}, \tilde{q}_1)$。爱丽丝的效用为

$$\begin{aligned} r_1^A(\tilde{\alpha}, \tilde{q}) =& 0.5 r_{-1}^A(\tilde{\alpha}_1, \tilde{q}) + 0.5 r_1^A(\tilde{\alpha}_2, \tilde{q}) \\ =& 0.5c(1 - \tilde{\alpha}_1 + \tilde{\alpha}_2) + 0.5(2 - \tilde{\alpha}_1 - \tilde{\alpha}_2) \\ &+ 0.5(3(\tilde{\alpha}_2 p + \tilde{\alpha}_1(1-p)) - 1)\mathbb{1}(\tilde{q}_{-1} \le 1/3) \\ &+ 0.5(3(\tilde{\alpha}_2(1-p) + \tilde{\alpha}_1 p) - 1)\mathbb{1}(\tilde{q}_1 \le 1/3). \end{aligned} \tag{73}$$

代理人 0 的效用为

$$r_1^0(\tilde{\alpha}, \tilde{q}) = -(\tilde{q}_{-1} - \tilde{\alpha}_2 p - \tilde{\alpha}_1(1-p))^2 - (\tilde{q}_1 - \tilde{\alpha}_2(1-p) - \tilde{\alpha}_1 p)^2. \tag{74}$$

$\hat{}$ 爱丽丝和代理人 0 都是效用最大化者。博弈 G_1 的效用为 (74)-(73)，是一个策略为 $\tilde{\alpha} = (\tilde{\alpha}_1, \tilde{\alpha}_2)$ $\tilde{q} = (\tilde{q}_{-1}, \tilde{q}_1)$ 的标准式博弈。由于代理人 0 的效用 (74) 是一个二次函数，代理人 0 的任何最佳响应都必须满足 $\tilde{q}_{-1} = \tilde{\alpha}_2 p + \tilde{\alpha}_1(1-p)$ $\tilde{q}_1 = \tilde{\alpha}_2(1-p) + \tilde{\alpha}_1 p$。

注意，在扩展阶段博弈 $\hat{}\, G_1$中，效用函数 $r_1^A(\tilde{\alpha}, \tilde{q})\hat{}$ 在 $\tilde{q}$上并不连续。为证明 G_1的纳什均衡存在性，我们接下来应用Barelli 和 Meneghel (2013年)关于效用不连续博弈的存在性结果。

具体而言，Barelli 和 Meneghel (2013年) 的命题 2.4 保证了满足广义更优回应安全性质的博弈存在纳什均衡。根据Barelli 和 Meneghel (2013年) 中的定义 2.3，若对于任意非均衡的 $(\bar{\alpha},\ \bar{q}$，以下条件至少有一个成立，则该阶段博弈具有广义更优回应安全性质

- 我们可以找到一个 $\epsilon>0$ 和一个闭对应 $\phi^0(\tilde{\alpha},\tilde{q})$，使得

$$r_1^0(\tilde{\alpha},\phi^0(\tilde{\alpha},\tilde{q}))\geq r_1^0(\bar{\alpha},\bar{q})+\epsilon \tag{75}$$

对于所有 $\tilde{\alpha}_1\in(\bar{\alpha}_1-\epsilon,\bar{\alpha}_1+\epsilon)$，$\tilde{\alpha}_2\in(\bar{\alpha}_2-\epsilon,\bar{\alpha}_2+\epsilon)$，$\tilde{q}_{-1}\in(\bar{q}_{-1}-\epsilon,\bar{q}_{-1}+\epsilon)$，$\tilde{q}_1\in(\bar{q}_1-\epsilon,\bar{q}_1+\epsilon)$

- 我们可以找到一个 $\epsilon>0$ 和一个闭对应 $\phi^A(\tilde{\alpha},\tilde{q})$，使得 t

$$r_1^A(\phi^A(\tilde{\alpha},\tilde{q}),\tilde{q})\geq r_1^A(\bar{\alpha},\bar{q})+\epsilon \tag{76}$$

for all $\tilde{\alpha}_1\in(\bar{\alpha}_1-\epsilon,\bar{\alpha}_2+\epsilon)$, $\tilde{\alpha}_2\in(\bar{\alpha}_2-\epsilon,\bar{\alpha}_2+\epsilon)$, $\tilde{q}_{-1}\in(\bar{q}_{-1}-\epsilon,\bar{q}_{-1}+\epsilon)$, $\tilde{q}_1\in(\bar{q}_1-\epsilon,\bar{q}_1+\epsilon)$

在附录.2中，我们证明，当 $c>24$ 扩展阶段博弈 ˆG_1 是广义最佳响应安全的。因此，根据（Barelli 和 Meneghel，2013年，命题 2.4），扩展状态博弈 G_1 ˆ存在一个纳什均衡。

考虑任意纳什均衡（α，q） G_1。由于 q 是α 代理人 0 的最佳响应，根据代理人 0 的效用（74），我们有

$$q_{-1}=\alpha_2 p+\alpha_1(1-p) \tag{77}$$
$$q_1=\alpha_2(1-p)+\alpha_1 p \tag{78}$$

ˆ 此外，由于 α 是爱丽丝在 G_1中的 q 的最佳响应，

$$\begin{aligned}\alpha\in&\arg\max_{\tilde{\alpha}}\Big(0.5r_{-1}^A(\tilde{\alpha}_1,q)+0.5r_1^A(\tilde{\alpha}_2,q)\Big)\\=&\arg\max_{\tilde{\alpha}}\Big(0.5r_{-1}^A(\tilde{\alpha}_1,(\alpha_2 p+\alpha_1(1-p),\alpha_2(1-p)+\alpha_1 p))\\&\qquad+0.5r_1^A(\tilde{\alpha}_2,(\alpha_2 p+\alpha_1(1-p),\alpha_2(1-p)+\alpha_1 p))\Big)\\=&\Big(\arg\max_{\tilde{\alpha}_1}r_{-1}^A(\tilde{\alpha}_1,(\alpha_2 p+\alpha_1(1-p),\alpha_2(1-p)+\alpha_1 p)),\\&\qquad\arg\max_{\tilde{\alpha}_2}r_1^A(\tilde{\alpha}_2,(\alpha_2 p+\alpha_1(1-p),\alpha_2(1-p)+\alpha_1 p))\Big)\end{aligned} \tag{79}$$

因此，(71)-(72) 对 α成立，因此，序贯分解要求 (49)-(50) 在 $t=1$ 处由 SIB 策略 σ_1^{Alice}表示，α满足。由此，我们基于定理2建立了 SIB 均衡的存在性。

7 无共同观测的情形

我们考虑第 2 节中的模型，但假设代理人没有共同观测，也就是说，

$$Z_t = \emptyset \quad \forall t \in \mathcal{T}. \tag{80}$$

系统动态、代理人的私人观测、代理人策略的函数形式、其效用以及均衡概念（贝叶斯纳什均衡）均与第 2 节相同。

尽管在这一特殊情形下，代理人没有共同观测，我们仍可根据定义 3 定义 SIB 策略，并依据定义 4 构造一致的CIB信念系统，其中 $Z_t = \emptyset \, \forall t \in \mathcal{T}$。

由于没有共同观测，对于任意实现，我们始终有

$$\begin{aligned} &\sum_{\hat{x}_{t+1}, \hat{s}_{t+1}} F_t^i(\hat{x}_{t+1}, \hat{s}_{t+1}, z_{t+1})(\pi_t^{\psi^\sigma};\, \sigma_t^{-i}) \\ =&\sum_{\hat{x}_{t+1}, \hat{s}_{t+1}} F_t^i(\hat{x}_{t+1}, \hat{s}_{t+1})(\pi_t^{\psi^\sigma};\, \sigma_t^{-i}) = 1 > 0 \end{aligned} \tag{81}$$

因此，定义4 中的情形 (ii) 永远不会发生，而（20）可以简化为

$$\begin{aligned} &\pi_{t+1}^{\psi^\sigma, i}(x_{t+1}, s_{t+1}) \\ =&\frac{F_t^i(x_{t+1}, s_{t+1})(\pi_t^{\psi^\sigma};\, \sigma_t^{-i})}{\sum_{\hat{x}_{t+1}, \hat{s}_{t+1}} F_t^i(\hat{x}_{t+1}, \hat{s}_{t+1})(\pi_t^{\psi^\sigma};\, \sigma_t^{-i})} \\ =&F_t^i(x_{t+1}, s_{t+1})(\pi_t^{\psi^\sigma};\, \sigma_t^{-i}) \\ =&\sum_{y_{t+1}, x_t, s_t, a_t} \Bigg[\mathbb{P}\{y_{t+1}, x_{t+1} \mid x_t, a_t\} \left(\prod_j \mathbb{I}\{s_{t+1}^j = \phi_{t+1}^j(s_t^j, y_{t+1}^j, a_t^j)\} \right) \\ &\left(\frac{1}{|A_t^i|} \prod_{j \neq i} \sigma_t^j(a_t^j)(\pi_t^\psi, s_t^j) \right) \pi_t^{\psi, i}(x_t, s_t) \Bigg]. \end{aligned} \tag{82}$$

根据（82），我们可以写出

$$\Pi_{t+1}^{\psi^\sigma, i} = \psi_{t+1}^{\sigma, i}(\Pi_t^{\psi^\sigma}) \quad \forall i \in \mathcal{N}, \tag{83}$$

$$\Pi_{t+1}^{\psi^\sigma} = \psi_{t+1}^{\sigma}(\Pi_t^{\psi^\sigma}). \tag{84}$$

换言之，给定一个 SIB策略 σ，更新规则 ψ^σ 由（84）给出的确定性函数构成，相应的一致的CIB信念系统$\Pi_t^{\psi^\sigma}$，$t \in \mathcal{T}$以确定性方式演化。此外，由于定义 4 中的情形（ii）不会在没有共同观测的情况下发生，因此更新规则

s

$\psi^{\sigma}{}_{,\ t+1}^{\ i}$ 由（82）给出的结果恰好成为贝叶斯规则。因此，CIB 信念$\Pi^{\psi^{\sigma},\ t}{}^{\ i}$ 变成了一个常规的概率质量函数，由下式给出

$$\Pi_t^{\psi^\sigma,i}(x_t,s_t)=\mathbb{P}^{\tilde{g}^i,\sigma^{-i}}(x_t,s_t)\quad\forall i\in\mathcal{N} \tag{85}$$

其中，$\tilde{g}^i$ 表示均匀策略（即对所有 $t\in\mathcal{T}$，以相同概率选择每个行动 $a_t^i\in\mathcal{A}_t^i$ 的策略）。

注 11。 如果 N 代理人具有相同的效用，即我们面对的是一个动态团队问题，那么 $\Pi_t^{\psi^\sigma}$，$t\in\mathcal{T}$类似于*Witsenhausen*（*1973*）中出现的共同知识，该文分析了一个动态团队。Witsenhausen（1973）中的共同知识是关于系统历史 H_t，$t\in\mathcal{T}$的一系列（随时刻变化的）概率质量函数。这些概率质量函数以确定性方式演化，类似于本节模型中针对（82）的 $\Pi_t^{\psi^\sigma}$，$t\in\mathcal{T}$。

对于这个没有公共观测的特殊情形，定理 2 变为

推论1。 考虑一个 SIB策略组合 $\sigma=\{\sigma_t,t\in\mathcal{T}\}$ 以及由 $\psi^\sigma=\{\psi_t^\sigma,t\in\mathcal{T}\}$ 定义的相应更新规则 (83)-(84)，针对本节的模型。定义

$$V_{T+1}^i(\cdot,\cdot)=0\ \textit{for all}\ i \tag{86}$$

$$V_t^i(\pi^{\psi_t^\sigma},s_t^i)=\mathbb{E}^{\sigma_t,\psi^\sigma}[U^i_{G_t(V_{t+1},\pi^{\psi_t^\sigma})}\mid s_t^i] \tag{87}$$

其中 $U^i_{G_t(V_{t+1},\pi_t^{\psi^\sigma})}=\ u_t^i(X_t,A_t)\ +\ V_{t+1}^i(\psi_{t+1}^\sigma(\pi_t^{\psi^\sigma})\ S_{t+1}^i)$，并且在条件期望 $\mathbb{E}^{\sigma_t,\psi^\sigma}[\cdot]$中，给定 S_t^i 时 (X_t,S_t) 的分布为 $\pi^{\psi^\sigma}{}_{,t}^{\ \ i}(x_t,s_t^{-i})$，$A_t^i,i\in\mathcal{N}$ 由 $\sigma_t^i(a_t^i\ |s_t^i,\pi_t^{\psi^\sigma})$ 生成；给定 (X_t,S_t,A_t) 的条件概率 $\sum_{x_{t+1},s_{t+1}^{-i}}\mathbb{P}(x_{t+1},s_{t+1}\mid x_t,s_t,a_t)$ 为

$$\begin{aligned}&\mathbb{P}(x_{t+1},s_{t+1}\mid x_t,s_t,a_t)\\&=\sum_{y_{t+1}}\mathbb{P}\{x_{t+1}\mid x_t,a_t\}\mathbb{P}\{y_{t+1}\mid x_{t+1},a_t\}\left(\prod_j\mathbb{I}\{s_{t+1}^j=\phi_{t+1}^j(s_t^j,y_{t+1}^j,a_t^j)\}\right).\end{aligned} \tag{88}$$

如果对于所有 $t\in\mathcal{T}$，都存在一个 SIB策略组合 $\hat{}\sigma_t$，使得 $\hat{}\sigma_t$ 是阶段博弈 $G_t(V_{t+1},\pi^{\psi_t^\sigma})$ 的贝叶斯纳什均衡，即

$$\mathbb{E}^{\hat{\sigma}_t^i,\hat{\sigma}_t^{-i},\psi^\sigma}[U^i_{G_t(V_{t+1},\pi_t^{\psi^\sigma})}\mid s_t^i]=\max_{\tilde{\sigma}_t^i\in\Lambda_t^i}\mathbb{E}^{\tilde{\sigma}_t^i,\hat{\sigma}_t^{-i},\psi^\sigma}[U^i_{G_t(V_{t+1},\pi_t^{\psi^\sigma})}\mid s_t^i] \tag{89}$$

对所有 $i \in \mathcal{N}$，并且

$$\hat{\sigma}_t = \sigma_t, \tag{90}$$

那么， σ 这一SIB策略组合就是本节定义的无公共观测动态博弈的一个SIB-BNE。

注12。 由推论 $\{\sigma_t, t \in \mathcal{T}\}$ 中的序贯分解所确定的SIB-BNE策略组合 *1*以及信念 $\{\Pi_t^{\psi^\sigma}, t \in \mathcal{T}\}$ 也是完美贝叶斯均衡（PBE）Fudenberg和Tirole *(*1991年）。这是因为 $\{\sigma_t, t \in \mathcal{T}\}$ 满足序贯理性（式 (89)），且信念 $\{\Pi_t^{\psi^\sigma}, t \in \mathcal{T}\}$ 始终根据贝叶斯法则进行更新，因此满足一致性。

8 结论

我们研究了底层系统具有动态性、策略性主体的行动被隐藏（不可观测）且信息不对称的随机动态博弈。我们提出了一种计算贝叶斯纳什均衡策略组合的方法，该方法基于代理人信息的压缩版本，并且只要这一反向过程的每一步都有解，就可以按时刻从后向前依次确定这些策略组合。该方法强调：（i）公共信息／共同知识对于识别可按序计算的贝叶斯纳什均衡策略组合的重要性；（ii）博弈中的决策所需的充分公共信息与团队决策所需的充分公共信息之间的差异。这种差异源于：在博弈中，代理人有动机偏离其预测策略，而在团队中则没有这种动机。由于这种动机，在每个时刻，每个代理人都会基于公共信息对博弈状态形成自己的看法／信念，但所有这些不同的看法／信念都是所有代理人之间的共同知识。因此，CIB信念系统由以下序列描述 $\Pi_{1:T}^{\psi}$，该序列由定义 2指定。

我们的研究着重于确定所研究博弈的 SIB-BNE 策略组合。我们注意到，当代理人没有共同观测时（即第 7 节中的模型），由我们的方法论确定的 SIB-BNE 策略组合也是完美贝叶斯均衡（PBE）策略组合；但当代理人拥有共同观测时（第 2 节中的一般模型），情况并非如此。确定第 2 节一般模型的 PBE 策略组合是一个值得研究的有趣问题。

.1 充分信息

我们将定义1 的条件 (i)-(iii) 与 Tavafoghi 等人(2022年) 中定义2的条件进行比较；为便于阅读，我们在下文列出 Tavafoghi 等人(2022年) 的定义。

定义 7 (充分私人信息Tavafoghi 等人 (2022年))。 我们称$S_t^i = \zeta_t^i(P_t^i, C_t; g_{1:t-1})$, $i \in \mathcal{N}$, $t \in \mathcal{T}$是代理人的充分私人信息，当且仅当:

(i) 它可以递归地更新为

$$S_t^i = \phi_t^i(S_{t-1}^i, H_t^i \backslash H_{t-1}^i;\ g_{1:t-1}) \text{ for } t \in \mathcal{T}\backslash\{1\}, \tag{91}$$

(ii) 对于任意策略组合 g 以及所有具有正概率的实现值 $\{c_t, p_t, p_{t+1}, z_{t+1}, a_t\} \in \mathcal{C}_t \times \mathcal{P}_t \times \mathcal{P}_{t+1} \times \mathcal{Z}_{t+1}$,

$$\mathbb{P}^{g_{1:t}}\{s_{t+1}, z_{t+1} \mid p_t, c_t, a_t\} = \mathbb{P}^{g_{1:t}}\{s_{t+1}, z_{t+1} \mid s_t, c_t, a_t\}, \tag{92}$$

其中，对于 $\tau \in \mathcal{T}$， $s_\tau^{1:N} = \zeta_\tau^{1:N}(p_\tau^{1:N}, c_\tau;\ g_{1:\tau-1})$;

(iii) 对于每个形式为 $\tilde{g} := \{\tilde{g}_t^i \colon \mathcal{S}_t^i \times \mathcal{C}_t \to \Delta(\mathcal{A}_t^i)\ i \in \mathcal{N}, t \in \mathcal{T}\}$ and $a_t \in \mathcal{A}_t$, $t \in \mathcal{T}$ 的策略组合 $\tilde{g}$;

$$\mathbb{E}^{\tilde{g}_{1:t-1}}\{u_t^i(X_t, A_t) \mid c_t, p_t^i, a_t\} = \mathbb{E}^{\tilde{g}_{1:t-1}}\{u_t^i(X_t, A_t) \mid c_t, s_t^i, a_t\}, \tag{93}$$

对于所有正概率的实现值 $\{c_t, p_t^i\} \in \mathcal{C}_t \times \mathcal{P}_t^i$， 其中 $s_\tau^{1:N} = \zeta_\tau^{1:N}(p_\tau^{1:N}, c_\tau;\ \tilde{g}_{1:\tau-1})$ 对于 $\tau \in \mathcal{T}$;

(iv) 给定任意形式为 $\tilde{g} := \{\tilde{g}_t^i \colon \mathcal{S}_t^i \times \mathcal{C}_t \to \Delta(\mathcal{A}_t^i)\ i \in \mathcal{N}, t \in \mathcal{T}\}$, $i \in \mathcal{N}$ 的策略组合 $\tilde{g}$， 以及 $t \in \mathcal{T}$,

$$\mathbb{P}^{\tilde{g}_{1:t-1}}\{s_t^{-i} \mid p_t^i, c_t\} = \mathbb{P}^{\tilde{g}_{1:t-1}}\{s_t^{-i} \mid s_t^i, c_t\}, \tag{94}$$

对于所有正概率的实现值 $\{c_t, p_t^i\} \in \mathcal{C}_t \times \mathcal{P}_t^i$， 其中 $s_\tau^{1:N} = \zeta_\tau^{1:N}(p_\tau^{1:N}, c_\tau;\ \tilde{g}_{1:\tau-1})$ 对于 $\tau \in \mathcal{T}$。

定义1的条件(i)出现在定义7中 S_t^i 的定义里，而定义1中关于递归更新的条件(ii)与定义7中的条件(i)相同。定义1的条件(iii)直接推出定义7的条件(iii)和(iv)；条件(iii)中的效用 $u_t^i(X_t, A_t)$ 以及定义7的条件(iv)中的随机变量s_t^{-i}, 都是(x_t, s_t)的函数；在条件(iii)下，给定(p_t^i, c_t)时它们的分布与给定(s_t^i, c_t)时相同。

然而，对于满足定义 7 的充分私人信息，条件 (ii) 未必适用于满足定义 1的情况。考虑以下例子。假设$X_1 = Y_1^1$异或 Y_1^2，且 Y_1^1, Y_1^2以相同的概率取值于 $\{0, 1\}$。$Z_1 = \emptyset$ 和 $Z_2 = X_1$。那么， $S_1^1 = S_1^2 = \emptyset$ 满足定义 1 ，因为$\mathbb{P}(x_1, s_1^{-i} \mid p_1^i, c_1) = \mathbb{P}(x_1 \mid y_1^i) = 0.5 = \mathbb{P}(x_1, s_1^{-i} \mid s_1^i, c_1)$。然而，它们不满足定义 7 中的条件 (ii)，因为 $\mathbb{P}(z_2 \mid p_1, c_1, a_1) = \mathbb{P}(x_1 \mid y_1^1, y_1^2) = \mathbb{1}(x_1 = y_1^1$ 异或 $y_1^2) \neq \mathbb{P}(z_2 \mid s_1, c_1, a_1) = \mathbb{P}(x_1) = 0.5$。

.2 扩展阶段博弈的广义更优回应安全性质证明

ˆ 我们证明，当 $c > 24$ 扩展阶段博弈 G_1 第 6 节中的是广义更优回应安全的。为此，我们令 $\beta^*(q) = \mathbb{1}(q \leq 1/3)$，并考虑以下五种情况。

情形 (i) $r_1^0(\bar{\alpha}, \bar{q}) \neq 0$。在此情形下，贝叶斯法则在 $(\bar{\alpha}, \bar{q})$ 处不成立。我们聚焦于代理人 0，并选择信念使其满足贝叶斯法则，如下所示：

$$\phi^0(\tilde{\alpha}, \tilde{q}) = (\tilde{\alpha}_2 p + \tilde{\alpha}_1(1-p), \tilde{\alpha}_2(1-p) + \tilde{\alpha}_1 p) \tag{95}$$

那么，这个 ϕ^0 是一个闭对应。根据 ϕ^0 的这一构造，我们可以选取 $\epsilon > 0$，使得

$$r_1^0(\tilde{\alpha}, \phi^0(\tilde{\alpha}, \tilde{q})) = 0 > r_1^0(\bar{\alpha}, \bar{q}) + \epsilon$$

情形 (ii) $r_1^0(\bar{\alpha}, \bar{q}) = 0$，且 $\bar{\pi}_{-1} \neq 1/3$ 和 $\bar{\pi}_1 \neq 1/3$。

由于 $\beta^*(q) = 1$ 当 $q < 1/3$ 时，$\beta^*(q) = 0$ 当 $q > 1/3$ 时，$\beta^*(\cdot)$ 在 $q \neq 1/3$ 的点处连续。因此，我们可以找到 $\epsilon > 0$，使得对于所有 $\tilde{q}_{-1} \in (\bar{q}_{-1} - \epsilon, \bar{q}_{-1} + \epsilon)$，$\beta^*(\tilde{q}_{-1}) = \beta^*(\bar{q}_{-1})$；并且对于所有$\tilde{q}_1 \in (\bar{q}_1 - \epsilon, \bar{q}_1 + \epsilon)$，$\beta^*(\tilde{q}_1) = \beta^*(\bar{q}_1)$。在此区域内，我们有

$$r_1^A(\alpha, \tilde{q}) = r_1^A(\alpha, \bar{q}) \tag{96}$$

对于所有 α。令

$$\phi^A(\tilde{\alpha}, \tilde{q}) = \arg\max_{\alpha} r_1^A(\alpha, \tilde{q}) \tag{97}$$

由于 $r_1^A(\cdot)$ 在所考虑的区域内连续，根据伯格最大值定理，$\phi^A(\cdot)$ 具有闭图。注意，对于所有$\tilde{q}_1 \in (\bar{q}_1 - \epsilon, \bar{q}_1 + \epsilon)$，$\tilde{q}_1 \in (\bar{q}_1 - \epsilon, \bar{q}_1 + \epsilon)$

$$r_1^A(\phi^A(\tilde{\alpha}, \tilde{q}), \tilde{q}) = \max_{\alpha} r_1^A(\alpha, \tilde{q}) = \max_{\alpha} r_1^A(\alpha, \bar{q}) \tag{98}$$

如果 $\max_\alpha r_1^A(\alpha, \bar{q}) > r_1^A(\bar{\alpha}, \bar{q})$，我们可以找到 $\epsilon > 0$，使得对于 $\tilde{q}_1 \in (\bar{q}_1 - \epsilon, \bar{q}_1 + \epsilon)$，$\tilde{q}_1 \in (\bar{q}_1 - \epsilon, \bar{q}_1 + \epsilon)$，$r_1^A(\phi^A(\tilde{\alpha}, \tilde{q})\ \tilde{q}) = \max_\alpha r_1^A(\alpha, \tilde{q}) \geq r_1^A(\bar{\alpha}, \bar{q}) + \epsilon$。

如果 $\max_\alpha r_1^A(\alpha, \bar{q}) = r_1^A(\bar{\alpha}, \bar{q})$，那么爱丽丝没有有利可图的偏离。此外，由于 $r_1^0(\bar{\alpha}, \bar{q}) = 0$，代理人 0 没有有利可图的偏离。因此，如果 $max_\alpha r_1^A(\alpha, \bar{q}) = r_1^A(\bar{\alpha}, \bar{q})$，则 $(\bar{\alpha}, \bar{q})$ 是一个均衡。

情形（iii） $r_1^0(\bar{\alpha}, \bar{q}) = 0$, $\bar{\pi}_{-1} = 1/3$ 以及 $\bar{\pi}_1 \neq 1/3$。

注意，$\bar{q}_{-1} = 0.8\bar{\alpha}_1 + 0.2\bar{\alpha}_2 = 1/3$和 $\beta^*(\bar{q}_{-1}) = 1/3$。由于 $\bar{\pi}_1 \neq 1/3$，我们可以找到 $\epsilon > 0$，使得 $\beta^*(\tilde{q}_1) = \beta^*(\bar{q}_1)$ 对所有 $\tilde{q}_1 \in (\bar{q}_1 - \epsilon, \bar{q}_1 + \epsilon)$ 都成立。

因此，

$$r_1^A(\bar{\alpha},\bar{q}) = 0.5c(1-\bar{\alpha}_1+\bar{\alpha}_2) + 0.5(2-\bar{\alpha}_1-\bar{\alpha}_2) + 0.5(3\bar{q}_1-1)\beta^*(\bar{q}_1) \tag{99}$$

为爱丽丝选取

$$\phi^A(\tilde{\alpha},\tilde{q}) = (0,1) \tag{100}$$

对所有 $\tilde{\alpha}_i \in (\bar{\alpha}_i-\epsilon, \bar{\alpha}_i+\epsilon)$ $i=1,2$, $\tilde{q}_i \in (\bar{q}_i-\epsilon, \bar{q}_i+\epsilon)$ $i=-1,1$。我们得到

$$\begin{aligned} r_1^A(\phi^A(\tilde{\alpha},\tilde{q}),\tilde{q}) =& c+0.5+0.5(0.6-1)\beta^*(\tilde{q}_{-1}) + 0.5(2.4-1)\beta^*(\tilde{q}_{-1}) \\ =& c+0.5-0.2\beta^*(\tilde{q}_{-1}) + 0.7\beta^*(\bar{q}_{-1}) \end{aligned} \tag{101}$$

and

$$\begin{aligned} & r_1^A(\phi^A(\tilde{\alpha},\tilde{q}),\tilde{q}) - r_1^A(\bar{\alpha},\bar{q}) - \epsilon \\ =& 0.5c(1+\bar{\alpha}_1-\bar{\alpha}_2) - 0.5(1+\bar{\alpha}_1+\bar{\alpha}_2) \\ & - 0.2\beta^*(\tilde{q}_{-1}) + 0.5(2.4-3\bar{q}_1)\beta^*(\bar{q}_{-1}) - \epsilon \\ \geq& 0.5c(1+\bar{\alpha}_1-\bar{\alpha}_2) - 0.5*3 - 0.2 - 0.5*0.6 - \epsilon \end{aligned} \tag{102}$$

当 $\bar{q}_{-1}=1/3$ 时，则 $0.8\bar{\alpha}_1+0.2\bar{\alpha}_2 = 1/3 \Rightarrow \bar{\alpha}_1 = 5/12 - 3/12\bar{\alpha}_2$。因此，

$$1+\bar{\alpha}_1-\bar{\alpha}_2 = 17/12 - 15/12\bar{\alpha}_2 \geq 1/6 \tag{103}$$

其中，最小值在 $\bar{\alpha}_1=1/6$ 和 $\bar{\alpha}_2=1$ 处取得。

当 $c>24$ 时，则

$$0.5c(1+\bar{\alpha}_1-\bar{\alpha}_2) \geq c/12 > 2 \tag{104}$$

以及 $r_1^A(\phi^A(\tilde{\alpha},\tilde{q}),\tilde{q}) - r_1^A(\bar{\alpha},\bar{q}) - \epsilon > 0$。

情形（iv） $r_1^0(\bar{\alpha},\bar{q})=0$，且 $\bar{\pi}_1 = 1/3$ 和 $\bar{\pi}_{-1} \neq 1/3$。

该情形与情形（iii）类似。由于 $\bar{\pi}_{-1} \neq 1/3$，我们可以找到 $\epsilon>0$，使得 $\beta^*(\tilde{q}_{-1}) = \beta^*(\bar{q}_{-1})$ 对所有 $\tilde{q}_{-1} \in (\bar{q}_{-1}-\epsilon, \bar{q}_{-1}+\epsilon)$ 都成立。此外，

$$\begin{aligned} & r_1^A(\bar{\alpha},\bar{q}) \\ =& 0.5c(1-\bar{\alpha}_1+\bar{\alpha}_2) + 0.5(2-\bar{\alpha}_1-\bar{\alpha}_2) + 0.5(3\bar{q}_{-1}-1)\beta^*(\bar{q}_{-1}) \end{aligned} \tag{105}$$

为爱丽丝选择闭对应（如情形（iii）所示）

$$\phi^A(\tilde{\alpha},\tilde{q}) = (0,1) \tag{106}$$

对所有 $\tilde{\alpha}_i \in (\bar{\alpha}_i - \epsilon, \bar{\alpha}_i + \epsilon)$ $i = 1$, 2, $\tilde{q}_i \in (\bar{q}_i - \epsilon, \bar{q}_i + \epsilon)$ $i = -1$, 1。于是

$$\begin{aligned} & r_1^A(\phi^A(\tilde{\alpha}, \tilde{q}), \tilde{q}) \\ =& c + 0.5 - 0.2\beta^*(\bar{q}_{-1}) + 0.7\beta^*(\tilde{q}_{-1}) \end{aligned} \tag{107}$$

and

$$\begin{aligned} & r_1^A(\phi^A(\tilde{\alpha}, \tilde{q}), \tilde{q}) - r_1^A(\bar{\alpha}, \bar{q}) - \epsilon \\ =& 0.5c(1 + \bar{\alpha}_1 - \bar{\alpha}_2) - 0.5(1 + \bar{\alpha}_1 + \bar{\alpha}_2) \\ & + 0.5(0.6 - 3\bar{q}_{-1})\beta^*(\bar{q}_{-1}) + 0.7\beta^*(\tilde{q}_{-1}) - \epsilon \\ \geq& 0.5c(1 + \bar{\alpha}_1 - \bar{\alpha}_2) - 0.5 * 3 - 0.5 * 2.4 - \epsilon \end{aligned} \tag{108}$$

当 $\bar{q}_1 = 1/3$, $0.2\bar{\alpha}_1 + 0.8\bar{\alpha}_2 = 1/3 \Rightarrow \bar{\alpha}_2 = 5/12 - 3/12\bar{\alpha}_1$。因此,

$$1 + \bar{\alpha}_1 - \bar{\alpha}_2 = 7/12 + 15/12\bar{\alpha}_1 \geq 7/12. \tag{109}$$

当 $c > 24$ 时，则

$$0.5c(1 + \bar{\alpha}_1 - \bar{\alpha}_2) \geq 7/24c > 2.7 \tag{110}$$

以及 $r_1^A(\phi^A(\tilde{\alpha}, \tilde{q})\ \tilde{q}) - r_1^A(\bar{\alpha}, \bar{q})\ - \epsilon > 0$。

情形 (v) $r_1^0(\bar{\alpha}, \bar{q})\ = 0$, 且 $\bar{\pi}_1 = 1/3$ 并且 $\bar{\pi}_{-1} = 1/3$。

我们有

$$r_1^A(\bar{\alpha}, \bar{q}) = 0.5c(1 - \bar{\alpha}_1 + \bar{\alpha}_2) + 0.5(2 - \bar{\alpha}_1 - \bar{\alpha}_2) \tag{111}$$

为爱丽丝选取闭对应（如情形 (iii) 和 (iv) 中所示）

$$\phi^A(\tilde{\alpha}, \tilde{q}) = (0, 1) \tag{112}$$

对所有 $\tilde{\alpha}_i \in (\bar{\alpha}_i - \epsilon, \bar{\alpha}_i + \epsilon)$ $i = 1, 2$, $\tilde{q}_i \in (\bar{q}_i - \epsilon, \bar{q}_i + \epsilon)$ $i = -1, 1$。于是

$$\begin{aligned} & r_1^A(\phi^A(\tilde{\alpha}, \tilde{q}), \tilde{q}) - r_1^A(\bar{\alpha}, \bar{q}) - \epsilon \\ =& 0.5c(1 + \bar{\alpha}_1 - \bar{\alpha}_2) - 0.5(1 + \bar{\alpha}_1 + \bar{\alpha}_2) - 0.2\beta^*(\tilde{q}_{-1}) + 0.7\beta^*(\tilde{q}_{-1}) - \epsilon \\ \geq& 0.5c(1 + \bar{\alpha}_1 - \bar{\alpha}_2) - 0.5 * 3 - 0.2 - \epsilon \end{aligned} \tag{113}$$

然后，按照 (iv) 中的步骤，我们有 $r_1^A(\phi^A(\tilde{\alpha}, \tilde{q})\ \tilde{q})\ - r_1^A(\bar{\alpha}, \bar{q})\ - \epsilon > 0$。

参考文献

Aumann R、Maschler M、Stearns R（1995）不完全信息的重复博弈。麻省理工学院出版社

34 不对称信息与隐藏行动的动态博弈 Barelli P, Meneghel I（2013年）关于不连续博弈中均衡存在问题的注记。《计量经济学》81(2)：813–824 Cardaliaguet P, Rainer C, Rosenberg D 等（2015年）频繁行动和不完全信息下的马尔可夫博弈——极限情形。《运筹学数学》41(1)：49–71

Escobar J, Toikka J（2013年）马尔可夫型私人信息博弈中的效率。《计量经济学》81(5)：1887–1934

Forges F（1992）不完全信息的重复博弈：非零和。《博弈论手册》1：109–154

Fudenberg D, Tirole J（1991）博弈论。1991年。剑桥，马萨诸塞州 Gensbittel F, Renault J（2015年）双方均具有不完全信息的马尔可夫链博弈的价值。《运筹学数学》40(4)：820–841

Gupta A, Nayyar A, Langbort C, 等 (2014年) 面向非对称信息线性高斯博弈的基于共同信息的马尔可夫完美均衡。《SIAM 控制与优化杂志》52(5):3228–3260

Gupta A, Langbort C, Ba¸sarT (2016年) 非对称信息与资源受限参与者的动态博弈及其在信息物理系统安全中的应用。《IEEE网络系统控制汇刊》4(1):71–81

Ho Y (1980) 团队决策理论与信息结构。《IEEE学报》68(6):644–654

H¨orner J, SugayaT, Takahashi S 等（2011）折扣随机博弈中的递归方法：一种用于 $\delta \to 1$ 的算法及一项民间定理。《计量经济学》79(4):1277–1318 Kartik D, Nayyar A（2021）非对称信息下零和随机博弈的上值与下值。《动态博弈与应用》11(2):363–388 Kumar P, Varaiya P（1986）《随机系统：估计、辨识与自适应控制》。普伦蒂斯-霍尔公司。Li L, Shamma J（2014）非对称零和随机博弈的Lp表述。载于：第53届IEEE决策与控制会议，IEEE，第1930–1935页 Li L, Shamma J（2017）零和非对称重复博弈中的高效策略计算。arXiv 预印本 arXiv:170301952

Li L、Langbort C、Shamma J（2017年）求解双人零和重复贝叶斯博弈。arXiv 预印本 arXiv:170301957

Mailath G、Samuelson L（2006年）《重复博弈与声誉》。牛津大学出版社，牛津

Nayyar A、Mahajan A、Teneketzis D（2011年）延迟共享信息结构中的最优控制策略。IEEE 自动控制汇刊 56(7)：1606–1620

Nayyar A、Gupta A、Langbort C 等（2013a年）基于公共信息的不对称信息随机博弈马尔可夫完美均衡：有限博弈。IEEE 自动控制汇刊 59(3)：555–570

Nayyar A、Mahajan A、Teneketzis D（2013b年）具有部分历史共享的分散式随机控制：一种公共信息方法。IEEE 自动控制汇刊 58(7)：1644–1658

Nayyar A, Gupta A, Langbort C, 等（2014年）非对称信息随机博弈的基于公共信息的马尔可夫完美均衡：有限博弈。《IEEE 自动控制汇刊》59(3):555–570。https://doi.org/10.1109/TAC.2013.2283743

Ouyang Y, Tavafoghi H, Teneketzis D（2015年）具有私有马尔可夫动态的动态寡头博弈。载于：第54届IEEE决策与控制会议（决策与控制会议）

Ouyang Y, Tavafoghi H, Teneketzis D（2017年）不对称信息动态博弈：基于公共信息的完美贝叶斯均衡与序贯分解。《IEEE 自动控制汇刊》

Renault J（2006年）单方信息缺失的马尔可夫链博弈的价值。《运筹学数学》31(3):490–512

Renault J（2012年）具有知情控制者的重复博弈的价值。《运筹学数学》37(1):154–179

Sinha A, Anastasopoulos A（2016年）不对称信息无限时域动态博弈中的结构化完美贝叶斯均衡。美国控制会议

Sugaya T（2012年）不完全信息和私人信息马尔可夫博弈中的效率。工作论文

Tang D、Tavafoghi H、Subramanian V 等（2022年）团队之间具有延迟的团队内信息共享的动态博弈。《动态博弈与应用》，第1–59页

Tavafoghi H、Ouyang Y、Teneketzis D（2016年）关于具有延迟共享信息结构的随机动态博弈。载于：第55届 IEEE 决策与控制会议（CDC），第 7002–7009 页

Tavafoghi H、Ouyang Y、Teneketzis D（2022年）解决非对称信息动态决策问题的统一方法：非策略型代理人。《IEEE 自动控制汇刊》67(3)：1105–1119。https://doi.org/10.1109/TAC.2021.3060835

Vasal D、Anastasopoulos A（2016年）非对称信息动态线性二次高斯博弈的信号均衡。载于：第55届 IEEE 决策与控制会议（CDC），第 6901–6908 页

Witsenhausen HS（1973）序贯随机控制的一种标准形式。《数学系统理论》7(1)：5–11

Zamir S（1992）不完全信息的重复博弈：零和。《博弈论手册》1：109–154

Zheng J，Casta˜n´on DA（2013年）嵌套信息下马尔可夫零和博弈的分解技术。载于：第52届 IEEE 决策与控制会议，IEEE，第574–581页