第 23 卷第 6 期
2023 年 11 月

经 济 学（季 刊）
China Economic Quarterly

Vol. 23, No. 6
November, 2023

有限理性下的集体合作：理论与应用

胡　涛　刘　烁*

摘　要：本文研究有限理性下的集体合作博弈。主要的理论创新是将 Osborne and Rubinstein (1998) 提出的“抽样均衡”概念拓展到博弈参与者具有异质性的情况，并全面分析了均衡的存在性、唯一性和稳定性等问题。我们发现抽样均衡可以很好地解释为何在一次性互动环境下陌生群体之间也可能存在合作，并揭示了考虑合作成本异质性分布的重要性。本文的结果可以被应用到信任与合作对国民财富积累的影响、文化与经济行为的联系等现实问题，为相关的实证文献提供理论支持。

关键词：集体合作；有限理性；抽样均衡

DOI：10.13821/j.cnki.ceq.2023.06.09

一、引　言

人类与其他物种的一个重大区别是可在由无亲缘关系的个体组成的大规模群体中进行合作（参见如 Boyd et al.，2003；黄少安和张苏，2013）。毫无疑问，集体行动在人类社会中无处不在，而个体在行动中是否愿意信任他人、提供协作对于整个集体的福利有至关重要的影响。大量例子表明，当集体中有足够多的个体采取协作行动时，整体的福利将非常可观；反之，当协作的人数较少时，集体行动很可能会面临诸如公地悲剧（Hardin，1968；Ostrom，2008）、转型发展动力不足等问题，从而导致集体福利显著降低。因此，人类的集体合作现象吸引了大量来自不同学科学者的关注：世界著名期刊《科学》（*Science*）在 2005 年创刊 125 周年之际组织了一份调研（Kennedy and Norman，2005），向全世界最优秀的科学家征集了 125 个“驱动基础科学研究以及决定未来科学研究方向”的科学难题，并由时任主编从中进一步提炼出 25 个“大问题”。最终，《科学》杂志给出的 25 个“大问题”几乎全是自然科学问题，仅有两个社会科学问题——“人类的合作行为如何演进”便是其中之一。

经济学界长期对集体合作问题保持高度关注。一般认为，长期关系和声誉机制都能促进群体中的合作行为。然而，这些基于重复博弈的理论并不能很好地解释为什么人们在与大量的陌生人进行的一次性博弈中也往往会视情况提供不同程度的协作。关于非重复的集体合作博弈，经济学文献中最著名的两个代表可能是公共品博弈（public good

* 胡涛，北京大学经济学院；刘烁，北京大学光华管理学院。通信作者及地址：刘烁，北京市海淀区颐和园路 5 号北京大学光华管理学院，100871；电话：(010) 62747635；E-mail：shuo.liu@gsm.pku.edu.cn。刘烁感谢国家自然科学基金项目（72103006）和中国信息经济学会（E22100961）对本研究的资助。沃迈、朱景琛和三位匿名审稿人为本文的写作提供了宝贵意见。文责自负。

game）与协作博弈（coordination game）。其中，曼瑟尔·奥尔森（Mancur Olson）在他的经典著作《集体行动的逻辑》中指出，由于"搭便车"现象的存在，大规模的群体很难在公共品博弈中实现充分合作和有效供给（Olson，1965）。虽然高水平的群体合作确实有可能在协作博弈中实现，但由于相关模型往往假设参与者的行动选择具有互补性，很多时候博弈还会同时存在一个低合作水平的均衡。对此，托马斯·谢林（Thomas Schelling）在他的经典专著《冲突的策略》中进行了精彩描述，并提出了非常著名的"聚点"（focal point）理论来解释合作均衡的出现（Schelling，1960）。

无可否认，Olson（1965）、Schelling（1960）以及后续一系列关于公共品博弈和协作博弈的工作极大地加深了我们对大规模、非重复性的集体合作问题的认识。然而，我们注意到已有的工作大都建立在博弈者具有无限理性（unbounded rationality）的前提下，并通过求解模型的（贝叶斯）纳什均衡对集体合作的现象进行解读。这意味着，在均衡的状态下，博弈的参与者需要对他人的行动（以及理性程度和支付函数）具有绝对精确的信念预期（Aumann and Brandenburger，1995）。在博弈的参与者为众多彼此陌生的个体或交易不具有长期性的情况下，这一假设是否恰当值得商榷。研究集体行动的著名学者、2009年诺贝尔经济学奖获得者埃莉诺·奥斯特罗姆（Elinor Ostrom）就曾直接指出，大量与集体合作行为有关的实验经济学研究表明，人们即使在重复博弈中也没有表现出明显的采用纳什均衡策略的倾向（Ostrom，1998）。这意味着，经典的理性人框架可能并不足以让我们完全理解群体在非重复博弈中的合作表现。有鉴于此，奥斯特罗姆认为构建一个以更符合现实的行为假设为基础的集体行动理论框架是"政治科学的核心任务"，并呼吁学者重视来自行为经济学的理论创新和实验发现。

为了回应奥斯特罗姆的批判和呼吁，本文将研究有限理性下的集体合作博弈。特别地，我们借鉴了由 Osborne and Rubinstein（1998）开创性地建立的"过程理性"（procedurally rational）模型，并将它应用到集体合作博弈问题的研究中。在我们的模型中，每位博弈者需要选择是否与他人合作（或信任他人）。每位博弈者的最终收益将取决于他以及与他随机匹配的另一位博弈者的行动选择。关键地，每位博弈者在考虑要不要与他人协作时，可以通过随机抽样的方式观察（和估计）不同行动能给他带来的支付，并据此做出最终的行动选择。我们将 Osborne and Rubinstein（1998）提出的"抽样均衡"（sampling equilibrium）的概念拓展到博弈者具有异质性的情形（不同个体可能具有不同的合作成本），并以此作为本文模型的均衡解概念。与经典的纳什均衡相比，抽样均衡的最大优势在于它并不要求博弈的参与者对他人的行动选择和理性程度等因素具有绝对精确的信念预期——博弈者仅需依据抽样结果选择为他带来最高支付的行动。这一直观的决策过程与我们日常生活中的经验比较相符。此外，来自实验经济学的证据表明，纳什均衡关于博弈者信念的假设在现实情况下很难得到满足（Costa-Gomes and Weizsäcker，2008；Friedman and Ward，2019），而人们在博弈中的行为可能更符合抽样均衡的预测（Selten and Chmura，2008；Goerg and Selten，2009）。

将博弈参与者的异质性引入我们的均衡分析框架至少具有以下两点价值。一方面，这为以往多支考虑同质化参与者特殊情形的理论文献提供了重要补充。研究抽样均衡以及更广泛的过程理性的文献中关于异质性的讨论尚不多见，但逐渐引起了一些学者的关注。例如，Osborne and Rubinstein（2003）考虑了一个多选项的投票模型，并假设不同

类型的投票者对候选项具有不同的偏好。与本文以支付结果为抽样对象的设定不同，Osborne and Rubinstein（2003）考虑以行动本身为抽样对象的情形。他们对于过程理性的假设较适合于个体支付直接取决于所有参与者的行动的博弈（如投票），而本文更关注个体收益只取决于其匹配到的“对手”的行动的场景（如非熟人之间的合作）。Rowthorn and Sethi（2008）则在过程理性的框架下考虑博弈者的社会性偏好差异（例如完全自私或视情况互惠）。他们研究了一个独裁者博弈（dictator game）的扩展应用，其模型中接收转移支付的委托人将根据抽样结果决定是否对负责分配的代理人的行动加以限制（如设定转移下限），而自私程度不同的代理人对于委托人设定的限制会有不同的反应。与 Rowthorn and Sethi（2008）相比，本文保留了博弈参与者仅关注自身物质收益的经典设定，侧重讨论他们合作成本的异质性与过程理性的交互作用。近期，Salant and Cherry（2020）一文也把博弈者的异质性纳入模型的均衡分析。然而，该文对于博弈者支付异质性的考虑仅限于均匀分布，本文则是考察了一般化的分布函数。此外，该文以统计推断刻画抽样过程的设定方式也与本文抽样过程的设定有明显差异。

另一方面，考虑参与人偏好异质性的过程理性模型为集体合作问题的研究带来了新的理论发现和应用启示。我们发现在较为一般的异质性环境下，集体合作博弈存在唯一的抽样均衡（这与基于完全理性的纳什均衡解概念所面临的多重性问题形成鲜明的对比）。基于这一清晰的理论预测，我们进行了较为深入的比较静态分析和福利分析，发现虽然更高的合作收益一般情况下都会伴随着更高的社会信任水平，但如果人群中不存在特别热衷于合作的个体（或者单向信任他人带来的支付过低），那么无论相互合作的潜在收益有多大，群体中无任何人合作都有可能是唯一的均衡结果。这一新颖的理论发现具有深刻的政策启示。特别地，它意味着政策制定者在评估不同的措施在提升社会合作、信任水平方面的效果时，必须细致地考虑群体中合作成本的异质性分布。例如，新冠肺炎疫情发生以来，很多地方政府都出台了针对抗疫一线医务人员的保障政策。根据我们的理论结果，这些政策虽然不会直接增加整个社会的合作收益，但却可以降低那些最热衷于奉献和合作个体的合作成本，而这将对推动整个社会的合作抗疫起到非常关键的作用。相反地，如果政府忽视了对一线医护人员等自愿合作者的兜底保障，仅寄望于通过宣传抗疫成功的潜在收益来促进民众协作配合，其效果很可能会不太理想。

考虑到在很多现实情形中，博弈者可能会经过多次抽样后才做出决策，我们也进一步将模型拓展到多次抽样的环境。分析结果表明，对于任意的抽样次数和合作成本分布函数，模型都存在一个稳定的抽样均衡。此外，我们发现多次抽样可以提高合作现象在人群中出现的可能性。这是因为当人群中合作成本较低的博弈者占比较高时，平均而言多次抽样的经历将放大合作行动相对于非合作行动的支付优势。

最后，我们把模型的比较静态分析和福利分析结果应用到信任与合作对国民财富积累的影响、文化与经济行为的联系等现实问题，为相关的实证文献提供理论支持。例如，我们的理论可以用于解释为什么社会信任与国民财富增长之间具有正相关关系（参见如刘凤委等，2009；吕朝凤等，2019；Knack and Keefer，1997；Durlauf and Fafchamps，2005）。又如，同宗族同方言可促进信任的实证发现（如潘越等，2019；马双和

赵文博，2019）背后可能存在更深刻的机理："文化同质"很可能并非群体呈现更高信任水平和经济绩效最主要的原因，更关键的也许是该同源文化恰好能降低合作的成本投入。这些讨论在帮助我们更好地从理论上理解近期文献中的一系列实证发现的同时，还凝练出了更精确、细致的"可以被验证的含义"（testable implications），对后续的实证工作有一定的启示作用。

二、模型设定

我们考虑以下一个共有一个连续统单位（a continuum）的参与者的群体博弈。在这个博弈中，每一位博弈者需要作出一个行动选择 $a \in A \equiv \{0，1\}$。其中，行动 $a=1$ 的经济含义可以被解读为"合作""投入"或"信任"等。相应地，行动 $a=0$ 的经济含义可以被解读为"不合作""不投入"或"不信任"等。

在进行行动选择的同时，每一位博弈者将随机与人群中的另一位博弈者匹配。每位博弈者 $i \in [0，1]$最终获得的效用同时取决于他的行动选择a_i，以及与他匹配的博弈者 j 的行动选择a_j：

$$u_i(a_i,\ a_j)=\begin{cases} B-c_i & \text{若} a_i=a_j=1 \\ b-c_i & \text{若} a_i=1,\ a_j=0, \\ 0 & \text{若} a_i=0 \end{cases}$$

其中，$B>0$ 为两位博弈者均采取合作行动时的收益；$b \in \mathbb{R}$ 为当匹配到的另一方选择不合作时，博弈者单方选择合作所能获得收益（或额外承受的损失）；$c_i \in [\underline{c}，\bar{c}]$为博弈者 i 选择合作行为所需付出的成本。综上所述，i 与其匹配对手 j 之间博弈的支付矩阵如图 1 所示。我们假设 $b<B$，并将任意博弈者选择不合作时所获得的支付正规化为零。

		博弈者j	
		1	0
博弈者i	1	$(B-c_i,\ B-c_j)$	$(b-c_i,\ 0)$
	0	$(0,\ b-c_j)$	$(0,\ 0)$

图 1　博弈的支付矩阵

我们假设每位博弈者异质性合作成本是她的私人信息，并互相独立地根据相同的累计分布函数 F 抽取。为了讨论的方便，我们进一步假设 F 是绝对连续的，并有一个对应的密度函数为 f。值得注意的是，与 Schelling（1960）和 Olson（1965）等经典的集体行动理论相比（具体参见 Anauati et al.，2016；Medina，2007），我们在上述模型中加入个体异质性。这一设定是具有广泛实证支持的：不少研究（例如 Fischbacher et al.，2001；Fischbacher and Gächter，2006；Burlando and Guala，2005；周业安等，2013）表明，集体行动中个体在偏好方面有丰富的异质性。以信任行为为例，人们的信任感以及诚信观很可能来自家庭教育（Alesina et al.，2015），而不同的家庭在教育子女方面付出的努力成本可能不同（韦倩等，2019）。因此，不同的个体采取合作行动的成本存在异质性这一假设不仅增加了模型结论的一般性，而且也使得模型的设定更贴合现实情况。

正如我们在引言中提到的，研究集体行动的经典理论文献受到的一个重要批判是其采用的模型解概念对博弈参与者理性程度的要求过分苛刻。例如，经典的纳什均衡要求每个博弈者对其他博弈者的行动有绝对精确的信念。然而，对于一些博弈而言（比如本文所关注的多人同时博弈），人们很难不经任何交流就对其他众多陌生人的选择有准确的预期。作为对相关质疑的回应，本文借鉴 Osborne and Rubinstein（1998）的“过程理性”概念，求解模型的“抽样均衡”。简单而言，Osborne and Rubinstein（1998）提出人们的实际行动决策基于以下一个抽样过程：博弈者对自己的每一个可选行动 $a \in A$ 都进行随机抽样尝试，然后选取抽样结果中对应的支付最大的行动。在多人博弈中，人们若以这种“支付最大化”的抽样经历为处事方式，最后形成的稳态均衡就是抽样均衡。特别地，当每位博弈者对每一个可选行动均抽样 K 次时，我们称其对应的稳态均衡为 $S(K)$-抽样均衡。为了尽可能清晰地展示模型的理论直觉，我们在接下来（第二和第三部分）将首先把注意力集中到单次抽样的情形，围绕 $S(1)$-抽样均衡进行分析和讨论。在稍后的第四部分，我们将仔细考察更为一般化的多次抽样情形。

均衡的严格定义如下。记任意合作成本为 $c \in [\underline{c}, \overline{c}]$ 的博弈者选择行动 $a \in A$ 的概率为 $\alpha_c(a)$。那么，任意博弈者 i 的抽样经历将是一个概率分布：记 i 对行动 a_i 带来支付进行抽样时产生的随机变量为 $v_i(a_i, \alpha)$，含义是如果其遇到的对手 $-i$ 的类型为 c，则取值等于 $u_i(a_i, a_{-i})$ 的概率为 $\alpha_c(a_{-i})$，其中 a_{-i} 为 i 匹配到的对手 $-i$ 的行动选择。具体而言，任何博弈者在人群中遇到合作行为的概率为 $\alpha(1)=\int_{\underline{c}}^{\overline{c}} \alpha_c(1)\,\mathrm{d}F(c)$，遇到不合作行为的概率为 $\alpha(0)=1-\alpha(1)$。基于此，可知成本类型为 $c_i \in [\underline{c}, \overline{c}]$ 的博弈者选取的行动 $a \in A$ 为最优的概率等于：

$$
\begin{aligned}
w(a, \alpha, c_i) = {} & (\alpha(1))^2 1_{u_i(a,1)>u_i(a',1)} + (\alpha(0))^2 1_{u_i(a,0)>u_i(a',0)} \\
& + \alpha(1)\alpha(0)\left(1_{u_i(a,1)>u_i(a',0)} + 1_{u_i(a,0)>u_i(a',1)}\right),
\end{aligned}
$$

此处 1 为指示函数（indicator function）。

定义 1　当一个概率分布 $\alpha \in \Delta A$ 满足 $\alpha_{c_i}(a)=w(a,\alpha,c_i)$ $\forall a \in A$，$c_i \in [\underline{c}, \overline{c}]$ 时，我们称它为 $S(1)$-抽样均衡。

换言之，当任意行动 $a \in A$ 被随机抽中的概率（亦即选择该行动的博弈者在人群中的比例）恰好等于其为博弈者带来最高的抽样支付的概率时，我们的博弈便达到了 $S(1)$-抽样均衡。

三、模型分析

（一）模型求解

根据图 1 的支付矩阵，给定对手的行动选择，博弈者的最优行动选择依然取决于其合作成本的大小。因此，合作成本的异质性意味着刻画博弈者的抽样支付结果的（均衡）分布并不是一个简单的任务。幸运的是，在我们的模型中个体博弈者的抽样结果的预期分布仅取决于人群中选择合作行动的博弈者的整体比例。依赖于这一观察，我们对模型的 S(1)-抽样均衡进行以下全面的刻画。

定理 1　(i) 若 $F(B)-F(b)=1$，任意的概率分布 $\alpha \in \Delta A$ 均可构成一个 $S(1)$-抽

样均衡。(ii) 若 $F(B)-F(b)<1$，则存在一个唯一的 $S(1)$-抽样均衡 $\alpha^* \in \Delta A$，其中选择合作行动的博弈者的比例为

$$\alpha^*(1)=\frac{F(b)}{1-(F(B)-F(b))}. \tag{1}$$

证明：假设人群中合作行动的选择初始分布为 $\alpha \in \Delta A$。注意到当博弈者选择不合作时，他通过抽样获得的支付结果总是为零（$u_i(0,1)=u_i(0,0)=0$）。接下来，让我们考虑一个合作成本为 $c_i \in [\underline{c}, \overline{c}]$ 的博弈者 i，并进行以下分类讨论：

一方面，若合作成本符合 $c_i>B$（$c_i<b$），则无论博弈者 i 在对行动 $a_i=1$ 带来的支付进行抽样时匹配到怎样的对手，他都会得到严格为负（严格为正）的收益。因此，对于这一类别的博弈者，选择不合作（合作）时的抽样支付总是要大于选择合作（不合作）时的抽样支付。换言之，合作成本大于 B（小于 b）的博弈者总是会选择不合作（合作）。

另一方面，若合作成本符合 $c_i \in (b, B)$，博弈者 i 的行动选择将取决于具体的抽样结果。特别地，当他对行动 $a_i=1$ 的支付进行抽样时，他将以 $\alpha(1)$ 的概率匹配到一位选择合作的对手。因为 $B-c_i>0$，此时博弈者得到的抽样支付必然会高于行动 $a_i=0$ 对应的抽样支付。然而，该博弈者还有 $\alpha(0)=1-\alpha(1)$ 的概率将抽到一位选择不合作的对手。在这种情况下，由于 $b-c_i<0$，他得到的抽样支付必然会低于行动 $a_i=0$ 对应的抽样支付。因此，对于合作成本处于中间范围的博弈者（$b<c_i<B$），行动 $a_i=1$ 给他们带来最高的抽样收益的概率为 $\alpha(1)$，而 $a_i=0$ 给他们带来最高的抽样收益的概率为 $\alpha(0)$。[①]

综上所述，在事前的状态下，博弈者选择 $a_i=1$ 的抽样收益是最优的概率为

$$w(1,\alpha)=\int_{-\infty}^{b} \mathrm{d}F(c)+\int_{b}^{B}\alpha(1)\,\mathrm{d}F(c)=F(b)+(F(B)-F(b))\cdot\alpha(1).$$

将上式代入定义1中的均衡条件 $\alpha(a)=w(a,\alpha)$，我们得知任意的 $S(1)$-抽样均衡必须满足

$$\alpha(1)=F(b)+(F(B)-F(b))\cdot\alpha(1). \tag{2}$$

显而易见，如果 $F(B)-F(b)=1$，任意 $\alpha(1)\in[0,1]$ 均符合式（2），因此定理1中的命题（i）成立。另一方面，当条件 $F(B)-F(b)<1$ 满足时，式（2）有唯一解 $\alpha^*(1)=\dfrac{F(b)}{1-(F(B)-F(b))}$，该解对应的即为均衡时人群中选择合作的博弈者的比例。因此，定理1中的命题（ii）成立。

定理1表明了我们的模型具有非常明确的均衡预测和丰富的经济学含义。特别地，在较为一般化的参数条件下（$F(B)-F(b)<1$），我们的模型存在唯一的 $S(1)$-抽样均衡。例如，当收益参数满足 $B\geqslant\overline{c}$ 且 $b>\underline{c}$ 时，我们有 $F(B)=1$ 及 $F(b)>0$（假设 $f(c)>0\ \forall c\in[\underline{c},\overline{c}]$），从而得到 $\alpha^*(1)=1$。这与我们直觉的预期一致：如果个体无论如何都能从合作选择中获得足够高的收益，那么群体中的所有人都会选择合作。我们模型的另一个角点均衡预测则更为微妙：当 $b\leqslant\underline{c}$ 时，我们有 $F(b)=0$，则无论 $F(B)-F(b)<1$ 是否成立，模型中总是存在一个所有博弈者都选择不合作（即 $\alpha^*(1)=0$）

① 由于成本刚好为 $c_i\in\{b,B\}$ 的博弈者的测度为零，我们在分析的时候可以略过这部分人群。

的 $S(1)$-抽样均衡。直觉上，这是因为当单向合作带来的收益 b 太小（或带来的额外损失太大）时，博弈者抽样遇到选择合作的对手的概率总是小于合作带来的抽样收益是最优的概率（$\alpha(1)<w(1,\alpha)$）。因此，无论相互合作的潜在收益 B 有多大，群体中都可能不会出现任何合作（例如在 $B<\overline{c}=+\infty$ 的情况下）。我们可以从两个方面来解读这个可能不那么显然的结果。一方面，该发现可以说明某些由衷地甘于奉献和合作的个体（即 $c_i<b$ 的博弈者）的存在对于推动群体合作非常关键。因此，向关键个体提供“兜底”保障（如通过法律明确“见义勇为”者不会被过分苛刻地追责、落实对抗疫一线医护人员的生活保障等）将对社会合作和信任的提高至关重要。另一方面，我们的结果也说明了如果政策制定者忽视对自愿提供合作的个体的收益保证，仅仅寄望于通过宣传合作的潜在收益或重奖成功合作典型（即增加 B）的方式来提高社会合作、信任水平的话，其效果很可能不会太理想。

定理 2　当个体无论如何都能从合作选择中获得足够高的收益（$B\geqslant\overline{c}$，$b>\underline{c}$）时，群体中的所有人都选择合作（$\alpha^*(1)=1$）将是唯一的均衡结果。当人群中不存在特别热衷于合作的个体或人们不能确保自己合作获得的收益足够大（$b\leqslant\underline{c}$）、但却存在部分几乎没有团队协作意识的个体（$B<\overline{c}$）时，无论相互合作的潜在收益有多大，群体中无任何人合作（$\alpha^*(0)=1$）都是唯一的均衡。

（二）均衡稳定性分析

Sethi（2000）通过构建一系列的反例说明，虽然 $S(1)$-抽样均衡非常直观且具有良好的存在性质，但它有可能是不稳定的。显然，基于不稳定均衡所得出的结论在稳健性方面让人存疑。鉴于此，我们将在进一步分析定理 1 中 $S(1)$-抽样均衡的经济学含义之前，简要地讨论并验证它的稳定性。

与 Sethi（2000）类似，我们考虑以下“抽样动态”（sampling dynamics）：在人群中，每一期均有相同比率的博弈者进入和退出。未退出的往期博弈者会继续采取其之前的行动，而新进入的博弈者则会对可选行动各抽样尝试一次，然后选取抽样结果中对应的支付最大的行动。因此，人群在新一期的行动分布将由往期博弈者的行为模式和新进入者的抽样结果共同决定。

正式地，记在 t 期人群中的行动选择分布为 $\alpha_t\in\Delta A$。假设在每一个长度为 $h\geqslant0$ 的离散时间段内，有 $1-e^{-h}$ 比例的博弈者会从该人群中退出，同时又会有相同比例的新博弈者加入。对于新进入人群的博弈者，给定成本类型 $c\in[\underline{c},\overline{c}]$，与定理 1 推理类似，合作行动 $a=1$ 为其带来最高抽样支付的概率为 $\widetilde{w}(a,\alpha_t,c)=F(b)+(F(B)-F(b))\cdot\alpha_t(1)$。据此，事前状态下，新进入人群的博弈者选择合作行动的概率为 $w(1,\alpha_t)=\widetilde{w}(a,\alpha_t,c)$。因此，在经过一个长度为 h 的离散时间段后，人群中选择合作行动的博弈者比例约为

$$\alpha_{t+h}(1)=e^{-h}\cdot\alpha_t(1)+(1-e^{-h})\cdot w(1,\alpha_t).\tag{3}$$

对式（3）中的时间长度 h 取极限，我们得到以下关于人群中选择合作行动的博弈者比例的准确演化方程式：

$$\dot{\alpha}_t(1)\equiv\lim_{h\to0}\frac{\alpha_{t+h}(1)-\alpha_t(1)}{h}=w(1,\alpha_t)-\alpha_t(1).\tag{4}$$

显然，抽样均衡必然是上述抽样演化动态过程的静止点（rest point，即满足 $\dot{\alpha}_t(a)=0\ \forall a\in A$ 的概率分布）。参考 Hirsch and Smale（1974）关于动态演化系统稳定性的标准定义，我们对 $S(1)$-均衡稳定性的正式定义如下：

定义 2 如果一个 $S(1)$-抽样均衡 α^* 满足下面的条件：$\forall\delta>0$，$\alpha^*\in U_\delta$，都存在一个子邻域 $V\subset U_\delta$，使得如果 $\alpha_{t_0}\in V\cap\Delta A$，则有 $\alpha_t\in U_\delta\cap\Delta A\ \ \forall t>t_0$。我们称 α^* 为稳定的 $S(1)$ 均衡。

下面的定理表明，我们模型中的 $S(1)$-抽样均衡具有良好的稳定性。

定理 3 在任意的参数条件下，$S(1)$-抽样均衡都是稳定的。

证明：当模型的参数满足 $F(B)-F(b)=1$ 时，由定理 1 可知，任意的概率分布 $\alpha\in\Delta A$ 都可构成一个 $S(1)$-抽样均衡。与此同时，在上述参数条件下，对任意的 $\alpha_t(1)$，式（4）都将简化为 $\dot{\alpha}_t(1)=0$。这意味着式（4）对应的演化系统实质上是静态的。因此，对于任意的 α_{t_0}，显然都有 $\alpha_t=\alpha_{t_0}\ \forall t>t_0$。这意味着 Hirsch and Smale（1974）的稳定性条件必然成立。

接下来，我们考虑 $F(B)-F(b)<1$ 的情形。由定理 1 可知，此时存在唯一的 $S(1)$-抽样均衡 α^*，其中选择合作行动的博弈者的比例由式（2）给出。与此同时，代入 $w(1,\alpha_t)$ 的表达式后，演化方程式（4）可进一步简化为：

$$\dot{\alpha}_t(1)=F(b)-(1-(F(B)-F(b)))\cdot\alpha_t(1).$$

因此，$\dot{\alpha}_t(1)$ 的正负情况与 $\alpha_t(1)-\alpha^*(1)$ 完全一致。结合 $\dot{\alpha}_t(1)$ 是关于 $\alpha_t(1)$ 的线性函数，这意味着若起始点 α_{t0} 坐落在 $S(1)$-抽样均衡 α^* 的一个邻域，则随着时间的推移，它总是会收敛到 α^*。由此可知，均衡 α^* 是渐近稳定的（asymptotically stable），所以它必然也是稳定的（Hirsch and Smale，1974；Sethi，2000）。

（三）内点均衡：比较静态分析及与纳什均衡的比较

当 $b\in(\underline{c},\overline{c})$ 时，我们有 $F(b)\in(0,1)$，因此均衡时的合作人群比例将处于内点水平：$\alpha^*(1)\in(0,1)$。通过直接的求导分析，我们得到以下比较静态结论：

定理 4 假设 b，$B\in(\underline{c},\overline{c})$，且 $f(c)>0\ \forall c\in[\underline{c},\overline{c}]$。随着（单方或者相互）合作带来的收益的提高，人们的整体合作水平将提高。

证明：对式（1）求导得到以下两个比较静态结论

$$\frac{\partial\alpha^*(1)}{\partial B}=\frac{f(B)F(b)}{(1-(F(B)-F(b)))^2}>0,$$

$$\frac{\partial\alpha^*(1)}{\partial b}=\frac{f(b)(1-F(B))}{(1-(F(B)-F(b)))^2}>0.$$

此处值得注意的是，即使 b，$B\in(\underline{c},\overline{c})$ 的条件不满足，我们依然有 $\frac{\partial\alpha^*(1)}{\partial B}\geqslant0$ 及 $\frac{\partial\alpha^*(1)}{\partial b}\geqslant0$。

定理 4 从直觉上很容易理解：无论 B 还是 b 提高，合作行动为博弈者带来的支付在所有情况下都不会降低，而且在特定的抽样结果下会带来支付的提高。因此，合作行动为博弈者带来最高的抽样收益的概率增加，进而使得均衡时有更大比例的人群选择合

作。定理 4 的比较静态结论有较为重要的现实意义。人们一直以来都希望了解哪些因素可以促进陌生群体之间的集体合作，至少从本模型的结论来看，提高合作成功的收益与失败的收益都可促进合作与信任。

接下来，我们进一步讨论成本异质性的分布对均衡合作水平 $\alpha^*(1)$ 的影响。我们考察最简单的一阶随机占优分布：设 G 与 F 是两个不同的分布函数，其中 F 一阶随机占优 G，即有 $F(c)\leqslant G(c)\ \forall c\in\mathbb{R}$。据此，我们得到以下定理。

定理 5　假设 $F(B)-F(b)<1$，$G(B)-G(b)<1$，且群体 F 的合作成本分布函数一阶随机占优另外一个群体 G 的成本分布函数。那么，群体 F 的整体合作水平将低于群体 G。

证明：记群体 F 和 G 的均衡合作水平分别为 $\alpha^F(1)$ 和 $\alpha^G(1)$。根据定理 1，我们有

$$\alpha^F(1)=\frac{F(b)}{1-(F(B)-F(b))},\ \alpha^G(1)=\frac{G(b)}{1-(G(B)-G(b))}. \tag{5}$$

进行简单的并项调整后，我们得到

$$\alpha^F(1)\leqslant\alpha^G(1)\Leftrightarrow\frac{F(b)}{1-F(B)}\leqslant\frac{G(b)}{1-G(B)}.$$

由于 F 一阶随机占优 G 意味着 $F(b)\leqslant G(b)$ 及 $F(B)\leqslant G(B)$，上述不等式必然成立。

定理 5 的直觉与定理 4 类似：若群体 F 的合作成本分布一阶随机占优群体 G，则在群体 G 中合作行动为博弈者带来最高的抽样收益的概率必然会高于群体 F，因此其均衡的整体合作水平也会更高。值得注意的是，如果我们仅仅假设群体 F 的平均合作成本高于群体 G（即 $\mathbb{E}_F[c]\geqslant\mathbb{E}_G[c]$），那么定理 4 中的比较结论并不一定成立。例如，我们可以假设 $b\in(\underline{c},\ \overline{c})$ 并想象以下两个分布函数 F 和 G。其中，分布 F 的支撑集为（$b-\varepsilon$，b]，因此我们有 $\alpha^F(1)=1$。分布 G 的支撑集中的点主要在成本下确界 $\underline{c}$ 附近，但也有测度非零的点取值高于 B。容易证明，只要 $\underline{c}$ 附近的支撑集的概率测度足够大，则 $\mathbb{E}_F[c]\geqslant\mathbb{E}_G[c]$ 成立，但我们总是有 $\alpha^G(1)<1$（即群体 F 虽然平均合作成本更高，但其均衡时的整体合作水平还是高于群体 G）。

虽然定理 4 和定理 5 的结论较为直观，但它们却很难在传统的完全理性模型中得到如此清晰的体现。例如，我们可以考虑本文博弈的（贝叶斯）纳什均衡解。正式地，如果一组策略映射 s_i：$[\underline{c},\ \overline{c}]\rightarrow A$ 满足 $\mathbb{E}_F[u_i(s_i(c),s_{-i}(\tilde{c}))]\geqslant\mathbb{E}_F[u_i(a,s_{-i}(\tilde{c}))]\ \forall a\in A,c\in[\underline{c},\ \overline{c}]$ 及 $i\in[0,\ 1]$，那么我们就称它为一个纳什均衡。给定图 1 的支付矩阵，易知当模型参数满足 $b\leqslant\underline{c}$ 时博弈总是存在一个群体完全无合作的纳什均衡（$s_i(c)=0\ \forall c\in[\underline{c},\ \overline{c}]$ 及 $i\in[0,\ 1]$）；当 $B\geqslant\overline{c}$ 时，博弈则总是存在一个群体中所有人都合作的均衡（$s_i(c)=1\ \forall c\in[\underline{c},\ \overline{c}]$ 及 $i\in[0,\ 1]$）。最后，假设博弈还存在一个内点纳什均衡，其中合作者的比例为 $\lambda\in(0,\ 1)$，那么对于成本类型为 $c\in[\underline{c},\ \overline{c}]$ 的博弈者合作是最优行动的条件是 $\lambda(B-c)+(1-\lambda)(b-c)\geqslant0$。该条件等价于 $c\leqslant c^*\equiv\lambda(B-b)+b$，表明对应的内点均衡具有明显的临界值特征：所有成本低于 c^* 的博弈者将选择合作，高于 c^* 的博弈者将选择不合作。因为纳什均衡要求信念与行动完全相符，一个对应临界值为 $c^*\in(\underline{c},\ \overline{c})$ 的内点均衡存在的充要条件为等式 $F(c^*)(B-b)+b-c^*=0$ 成立。由于相关参数条件并不互斥，上述的角点和内点纳什均衡可以在博弈中同时存在。实际上，由

于 $Z(c)\equiv F(c)(B-b)+b-c$ 不一定是单调函数，博弈甚至可能同时存在多个内点纳什均衡。此时，如果我们不引入其他（外生于本文模型的）精炼均衡的条件（比如以动态稳定解为精炼标准），则很难得到与定理4和定理5类似的、非常清晰的比较静态分析。例如，在图2中，存在三个不同的纳什均衡，对应的群体合作水平分别为 $F(c_1^*)<F(c_2^*)<F(\bar{c})=1$。如果我们假设群体将共同选择合作程度处于中间水平的纳什均衡，那么由图可知，当相互合作的收益 B 受到外生冲击提高至 B' 时，群体的合作水平反而可能会下降($F(c'_2)<F(c_2^*)$)。这一结论明显很难和我们常识的认知联系起来。综上所述，我们认为对于集体合作这一重要议题，本文基于“过程理性”的抽样均衡模型可以比现有文献中基于“完全理性”的纳什均衡模型提供更多、更精确的可被检验的实证预测，从而也更能带来新颖和深刻的发现。[①]

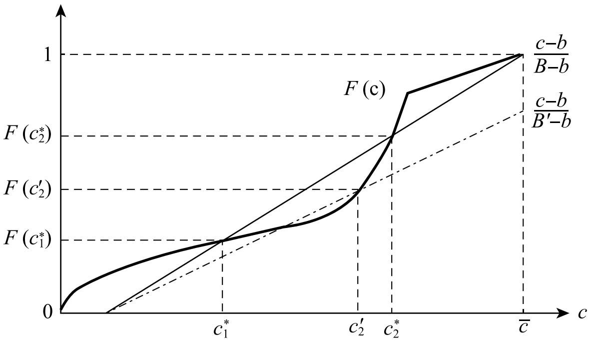

图2　理性模型中的多重纳什均衡情形

注：图中粗实线为分布函数 $F(c)$ 在其支撑集 $[\underline{c},\bar{c}]=[0,2]$ 的取值变化，细实线（点划线）为方程 $(c-b)/(B-b)$/ $((c-b)/(B'-b))$ 在这一区间的取值变化；此处我们设定 $B=2$，$B'=2.5$ 及 $b=0.2$。可以看到，粗实线和细实线在合作成本的支撑集上共有三个交点，此时博弈存在三个纳什均衡，对应的群体合作水平分别为 $F(c_1^*)<F(c_2^*)<F(\bar{c})=1$。当合作的收益从 B 提高至 B' 时，从粗实线和点划线的相交可见，中间位置纳什均衡的群体合作水平反而下降了（$F(c'_2)<F(c_2^*)$）。

需指出，上述比较结论依赖于定理4和定理5中的前提条件 $F(B)-F(b)<1$。反之，若有 $F(B)-F(b)=1$，则模型存在无穷多个抽样均衡，其数量很可能远远多于纳什均衡（注意当 c 均匀分布于 $[b,B]$ 时，博弈也存在无穷多个纳什均衡）。这说明了我们在应用本部分的分析结论时需注意适用的范围。例如，当群体中存在至少一部分由衷乐于奉献的个体（$b>\underline{c}$）或者很不合群、几乎没有团队协作意识的个体（$B<\bar{c}$）时，那么抽样均衡具有唯一性，其在解释和预测方面相对于纳什均衡的优势可能会较为明显。反之，若上述条件不满足，则考虑博弈参与者的过程理性并不一定能显著提升模型的表现。

① 在完全理性模型的框架下，全局博弈理论（global games）（Morris and Shin，2003）也曾对集体合作问题进行与本文类似的讨论。其中，Carlsson and Van Damme（1993）假设成功合作具有不确定的共同价值（common value），而每位博弈者只拥有关于这一共同价值的不完美的私人信息。他们证明了，在满足“极限占优”（limit dominance；其含义与我们的条件 $F(B)-F(b)<1$ 类似）等一系列标准条件下，上述博弈存在唯一的纳什均衡。与该文相比，我们模型中合作成功的共同价值部分是确定的（参数 B 为公共信息），但私人价值（private value）部分则存在不确定性和异质性。我们证明了在这种情况下，由于博弈参与者的有限理性，均衡结果也可能是唯一的。因此，本文采取的过程理性模型与全局博弈的经典模型具有一定的互补性。

四、拓展：多次抽样均衡

在不少现实场景（如家庭、宗族、工作团队）中，人们之间的互动很可能相当频繁，因此完全可以对不同行动进行多次抽样。比如，博弈者 A 可以通过博弈者 B 在过往多件事情上的表现（比如 B 是否每一期都固定偿还他向 A 借的小额贷款），才决定是否给予他完全信任。为此，本部分将抽样均衡的定义拓展到人群多次互动的环境，刻画其存在与稳定方面的性质，进而验证我们从单次抽样情形中得到的理论直觉的稳健性，并获得一些新的启示。

我们假设人群中每个人对每一个可选行动均抽样 $K \geqslant 1$ 次，并最终选取其抽样经历中效用总和最大的行动。与单次抽样的情形类似，无论抽样结果如何，成本类型为 $c<b$ 的所有博弈者必然选择合作，类型为 $c>B$ 的博弈者必然不合作。对于成本类型为 $c \in [b, B]$ 的博弈者，如果他在对合作行动的支付结果进行的 K 次抽样中有 $m \in \{0, 1, \cdots, K\}$ 次遇到合作者，有 $K-m$ 次遇到非合作者，那么对应的总抽样支付为：

$$m(B-c)+(K-m)(b-c).$$

由于选择不合作的行动的收益总是等于零，我们在简易计算后可以判定当且仅当 $m>K(c-b)/(B-b)$ 时，博弈者从合作中得到的总抽样支付才会大于不合作的总抽样支付。

接下来，对于所有 $m=0, 1, \cdots, K$，我们定义以下阈值：

$$c^m \equiv \frac{m}{K}(B-b)+b.$$

那么，给定人群中选择合作行动的博弈者比例为 $\alpha(1)$，对于成本类型为 $c \in (c^{m-1}, c^m)$ 的博弈者，合作行为的 K 次抽样总支付比不合作行为更高的概率为：

$$\sum_{k=m}^{K} C_K^k\ (\alpha(1))^k\ (1-\alpha(1))^{K-k},$$

其中 $C_K^k=\dfrac{K!}{k!(K-k)!}$ 为组合数公式。下一步，对于所有 $m=1, \cdots, K$，我们定义

$$\Phi_m(\alpha(1)) \equiv (F(c^m)-F(c^{m-1})) \cdot \Big(\sum_{k=m}^{K} C_K^k\ (\alpha(1))^k\ (1-\alpha(1))^{K-k}\Big).$$

那么，在事前状态下，K 次抽样中，合作行动 $a=1$ 为博弈者带来最高抽样总支付的概率为

$$w_K(1, \alpha) \equiv F(b)+\sum_{m=1}^{K} \Phi_m(\alpha(1)). \tag{6}$$

相应地，我们有 $w_K(0, \alpha)=1-w_K(1, \alpha)$。

定义 3　当一个概率分布 $\alpha \in \Delta \mathrm{A}$ 满足 $\alpha=w_K(a, \alpha)\ \forall a \in A$ 时，我们称它为 $S(K)$-抽样均衡。

显而易见，定义 1 中的 $S(1)$-抽样均衡是定义 3 的 $S(K)$-抽样均衡的一个特例。此外，与第三部分第（二）小节类似，我们可以考虑如下一个基于抽样动态的演化方程式：

$$\dot{\alpha}_t(1)=w_K(1, \alpha_t)-\alpha_t(1). \tag{7}$$

进而讨论 $S(K)$-抽样均衡的稳定性问题。由于稳定 $S(K)$-抽样均衡的正式定义与定义2完全类似，我们在此略过。

定理6 对于任意的合作成本分布函数，至少存在一个稳定的 $S(K)$-抽样均衡。

证明：假设 $F(b)=0$，且 $w_K(1,\alpha)\leqslant\alpha(1)\ \ \forall\alpha(1)\in[0,1]$。那么，易知 $\alpha^*=(0,1)$（即全部博弈者均选择不合作）是一个 $S(K)$-抽样均衡。此外，由于$\dot{\alpha}_t(1)\leqslant 0$ 对于任意的起始点 $\alpha(1)$和时间 t 均成立，$\alpha^*=(0,1)$必然满足 Hirsch and Smale（1974）的稳定性条件。

接下来，假设 $F(b)>0$。因为式（6）是关于 $\alpha(1)$的连续函数，且定义域和值域均为$[0,1]$，又因为差值函数 $\delta_K(\alpha(1))\equiv w_K(1,\alpha)-\alpha(1)$ 在 $\alpha(1)=0$ 处严格大于零且在 $\alpha(1)=1$ 处小于零，利用介值定理（intermediate value theorem）可以判断，必然存在一个同时满足以下两个条件的概率分布 $\alpha^*\in\Delta A$：(i) $w_K(1,\alpha^*)=\alpha^*(1)$；(ii) 存在 $\varepsilon>0$，使得如果$|\alpha(1)-\alpha^*(1)|<\varepsilon$，则有$(w_K(1,\alpha)-\alpha(1))(\alpha(1)-\alpha^*(1))\leqslant 0$。条件（i）表明 α^*是一个 $S(K)$-抽样均衡。条件（ii）表明，如果演化方程式（7）的初始概率 $\alpha_0(1)$稍微高于（低于）均衡概率 $\alpha^*(0)$，我们总是有$\dot{\alpha}_t(1)\leqslant 0$（$\dot{\alpha}_t(1)\geqslant 0$）。根据 Hirsch and Smale（1974）和 Sethi（2000），以上条件足以保证 $\alpha^*(1)$是演化方程式（7）的一个稳定的静止点，因此它对应的概率分布 α^*是一个稳定的 $S(K)$-抽样均衡。

定理6的一个直接推论是，当存在唯一的 $S(K)$-抽样均衡时，它必然是稳定的。然而，与定理1不同，定理6并不排除多个 $S(K)$-抽样均衡同时存在的可能。以 $K=2$ 为例。此时，我们有$c^0=b$，$c^1=(B+b)/2$ 和$c^2=B$，进而有

$$w_2(1,\alpha)=F(b)+(F((B+b)/2)-F(b))\cdot 2\cdot\alpha(1)(1-\alpha(1))$$
$$+(F(B)-F((B+b)/2))\cdot\alpha(1)^2.$$

若 $F(b)=0$，则此时 $\alpha(1)=0$ 必为方程 $w_2(1,\alpha)=\alpha(1)$的一个解，即存在一个无人合作的 $S(2)$-抽样均衡。这一均衡也是在 $K=1$ 的情形下的唯一均衡。然而，如果我们进一步有 $F((B+b)/2)>0.5$，则易证差值函数 $\delta_2(\alpha(1))\equiv w_2(1,\alpha)-\alpha(1)$关于 $\alpha(1)$在 $\alpha(1)=0$ 附近是严格递增的。故由介值定理可知，均衡条件 $w_2(1,\alpha)=\alpha(1)$必然还存在一个非零解。换而言之，与单次抽样的情形不同，即使在 $b\leqslant\underline{c}$的情况下，只要合作成本较低的人群比例足够高（$F((B+b)/2)$足够大），我们依然可能在人群中实现一定水平的合作。值得注意的是，在这个例子中，合作水平为零的抽样均衡是不稳定的：如果人群的初始合作水平稍稍大于零，那么根据演化方程式（7），他们将在后续的互动中进一步提高其合作的概率。与之相反，合作水平非零的抽样均衡是稳定的。以下定理将上述理论直觉拓展到更为一般化的情况。

定理7 对于任意 $K\geqslant 2$，只要 $F(c^1)>\dfrac{1}{K}$，那么总存在一个有非零比例的博弈者合作的稳定 $S(K)$-抽样均衡。

证明：当 $F(b)>0$ 时，$\alpha(1)=0$ 不可能为均衡条件 $w_K(1,\alpha)=\alpha(1)$的解，故此时对于任意取值的 $F(c^1)$，都存在一个合作水平非零的稳定 $S(K)$-抽样均衡（参见定理6）。

下面我们考虑 $F(b)=0$ 的情形。对式（6）求导，并对其在 $\alpha(1)=0$ 的点取值，我们有

$$\frac{\partial w_K(1,\ \alpha)}{\partial \alpha(1)}\Big|_{\alpha(1)=0}=F(c^1)\cdot K>1.$$

这意味着，函数 $\delta_K(\alpha(1))\equiv w_K(1,\ \alpha)-\alpha(1)$ 在零点附近的取值必然是递增的。因此，由连续性和介值定理可知，此时均衡条件 $w_K(1,\ \alpha)=\alpha(1)$ 必然存在一个符合定理 6 证明中的条件（i）和（ii）的非零解。根据 Hirsch and Smale（1974）和 Sethi（2000），该非零解必然对应一个稳定的 $S(K)$-抽样均衡。

总结而言，引入博弈者多次抽样的可能性并没有彻底改变模型的均衡性质，说明了我们在第三部分中关于 $S(1)$-抽样均衡的分析结论具有一定的稳健性。更进一步地，我们发现与单次抽样的情形相比，多次抽样均衡中合作出现的可能性更高。这一结果与新近的文献中提出的多次抽样均衡比单次抽样均衡更容易出现合作（参见如 Sandholm et al.，2020；Arigapudi et al.，2021）的发现是一致的。需要指出的是，多次抽样均衡下更容易合作的机制与重复博弈完全不同：前者是因为合作成本较低的博弈者在人群中的占比较高，多次抽样经历能放大合作相对于非合作的支付优势；后者则是靠与博弈历史相关的奖惩机制促进合作。

五、模型的应用

接下来，我们通过简单的福利分析，讨论几个较为重要的现实问题。为了避免冗长的技术性讨论，我们将把注意力集中到 $S(1)$-抽样均衡。

（一）国民财富的积累：信任与合作的价值

近年来中国经济从过去 30 多年的高速增长逐渐步入“新常态”，以前粗放式的、主要依靠高投资以及廉价劳动力的增长模式不可持续。这意味着中国今后的增长要更多地依靠技术进步以及其他一些“无形”的因素。事实上，社会信任这种“无形”因素对国民财富的积累越来越受到关注：国际上大量研究表明，以社会信任为核心的社会资本可以显著促进一国的经济增长（如 Knack and Keefer，1997；Zak and Knack，2001；Durlauf and Fafchamps，2005）。尤为值得我们注意的是，最近涌现的很多实证文献论证了社会信任与经济增长之间的正向关系在中国也存在，如刘凤委等（2009）、崔巍和陈琨（2016）、吕朝凤等（2019）等。

尽管大量的实证文献讨论了社会信任对经济有正面促进作用，从直观上也不难理解，比较遗憾的是目前关于社会信任与财富积累（经济增长）之间的理论探讨大多局限于完全理性模型（参见如崔巍和陈琨，2016；Francois and Zabojnik，2005；Zak and Knack，2001），因此可能会遇到我们在第三部分第（三）小节提到的问题。本文模型为此提供了一种仅要求经济体参与者具有有限理性的理论支撑，其结论可能更为稳健。为简便起见，假设 $B,\ b\in(\underline{c},\ \overline{c})$。根据第三部分第（三）小节的分析，我们的模型预测均衡时人群中采取合作行动（或选择信任他人）的概率为 $\alpha^*(1)=\dfrac{F(b)}{1-(F(B)-F(b))}$（同时也是合作行动为类型 $c\in(\underline{b},\ B)$ 带来更高抽样支付的概率）。据此，我们得到以下的社会总福利：

$$W^F(B, b) = \alpha^*(1)\int_{\underline{c}}^{B}(B-c)\,\mathrm{d}F(c) + (1-\alpha^*(1))\int_{\underline{c}}^{b}(b-c)\,\mathrm{d}F(c). \tag{8}$$

不难发现，社会总福利（或国民财富）将随着合作或信任收益 B 或 b 的提升而提升：$\partial W^F/\partial B > 0$，$\partial W^F/\partial b > 0$。这其中当然包含了收益 B 或 b 的直接效应：给定整个社会的信任程度（即人群中选择信任他人的比例），B 或 b 的提升将直接提高信任和合作带来的经济收益。然而，我们的分析表明，即使增加 B 或 b 并不具有直接收益（比如当它们伴随较高的征税成本时），它们也可以通过提高社会信任程度 $\alpha^*(1)$ 的方式（参见定理 3），间接提高国民的总财富。

现实中有一些较为直观的政策可以改变 B 以及 b，比如 Olson（1965）在《集体行动的逻辑》中提议第三方对于人们的合作行为予以额外奖励。另外还可以考虑一些与制度建设方面有关的政策，比如公正的司法体系、清晰的产权保护以及契约的监督与维护等方面的加强。本文的一个重要启示在于，当政策制定者以增加社会财富为目标对上述公共资源的投入和维护进行权衡分析时，必须全盘考虑它们对促进群体信任的内生影响。

（二）文化、信任与集体合作

普遍认为，文化作为一种群体内世代传承的习俗和信念，决定了人们的价值观念和认知，潜移默化地影响个体的行为模式（张航等，2021）。本小节的内容主要是围绕本文模型的主要结果，为新近出现的一系列从不同角度分析文化与经济之间关系的经验研究提供一定的理论支持。

特别地，我们注意到近年来国内学者非常关注宗族、方言等文化因素如何通过影响群体内部的信任与合作的机制，进而影响人们的各种经济行为。例如，潘越等（2019）讨论了宗族因素可以缓解民营企业的融资约束，其中一个重要的影响机制就是同宗同族的群体拥有更高的信任水平。针对更为具体的市场交易情景，张航等（2021）发现当二手房买卖双方的方言越接近时，最终成交的价格越低。根据作者的论证，这是因为方言作为文化的一种主要载体激发了卖方对买方的提携与帮助意愿，而这种意愿的产生可能是因为使用同一种方言促进了人们之间相互的认同感。马双和赵文博（2019）则发现地域中的方言越多样，流动人口的收入增长越受到抑制。文章给出的解释是不同方言的使用导致信任以及其他人际关系的阻力增加，进而影响了收入增长。

本文提出的简单理论框架可以帮助我们更好地理解上述实证结果背后的机制。一方面，我们认为宗族以及方言具有非常独特而又清晰的文化因素，非常容易在特定的同类人群中形成认同感、凝聚力，比如共同的习俗、饮食习惯、行为模式等能明显降低人际中的各种交易成本以及阻力。这一论断也有实证研究的支持。例如，卢盛峰和陈悦（2019）的研究便明确论证了方言可以显著提升人际融合状态。如果把这些阻力理解为合作或信任活动的成本，那么我们第三部分第（二）小节的比较静态分析表明，当这些成本整体下降时，群体的信任程度会上升。举例而言，假设群体 F 的成员文化背景差异比较大，而群体 G 的成员则具有方言、宗族等方方面面共同的文化背景。在这种情况下，群体 F 面对的合作成本整体上很可能会比群体 G 的更高。特别地，如果这两者对应的合作成本分布满足一阶随机占优条件的话，由定理 4 可知群体 F 的均衡合作、信任水

平 $\alpha^F(1)$ 将低于群体 G 的水平 $\alpha^G(1)$。若进一步假设 F 局部随机占优 G（local stochastic dominance，即 F 和 G 满足 $F'(c)/F(c)\leqslant G'(c)/G(c)\ \ \forall c\in[\underline{c}, \overline{c}]$），则能得到群体 F 的社会总福利比群体 G 的（分别记为 W^F 和 W^G）更低：

$$\begin{aligned}W^F &= \alpha^F(1)\int_{\underline{c}}^{B}(B-c)\,\mathrm{d}F(c)+(1-\alpha^F(1))\int_{\underline{c}}^{b}(b-c)\,\mathrm{d}F(c)\\ &\leqslant \alpha^G(1)\int_{\underline{c}}^{B}(B-c)\,\mathrm{d}G(c)+(1-\alpha^G(1))\int_{\underline{c}}^{b}(b-c)\,\mathrm{d}G(c)=W^G,\end{aligned}$$

其中不等式成立是因为 F 与 G 之间的局部随机占优关系意味着 $\mathbb{E}_F[c\mid c\leqslant\hat{c}]\geqslant \mathbb{E}_G[c\mid c\leqslant\hat{c}]\ \ \forall\hat{c}\in[\underline{c}, \overline{c}]$（参见 Riley，2012，p. 242）。

另一方面，我们的理论模型说明了文化的“同质性”（homogeneity）本身并不足以解释前述的经验发现：一个在合作成本方面高度同质的群体（如共同信仰某种宗教）并不一定比另一个同质性较低的群体有更高的合作程度。为了理解这一点，我们可以考虑一个较为极端的例子：假设群体 F 中一半人的合作成本为 $b+\frac{\varepsilon}{2}$，一半人的成本为 $b+\varepsilon$，其中 $b>\underline{c}$；而群体 G 中有一半人的合作成本为 $b-\varepsilon$，另一半人的合作成本为 $b+\varepsilon$（此处 $\varepsilon>0$）。在这个例子中，我们可以认为群体 F 比群体 G 更为同质化，然而根据定理 1 我们依然有 $\alpha^F(1)=0<\alpha^G(1)$。换言之，我们之所以观察到前述文献中的实证结果，很可能并不是因为文化背景的共同性本身有利于经济绩效的增长。例如，一个来自同一地方的群体可能会由于共同的成长环境，呈现较高的同质化。然而，如果该地方的文化缺乏如潮汕文化中对社会救助以及慈善体系的重视（陈友义，2014），那么该群体也可能难以实现较高的合作水平和经济绩效。①

六、结　　语

为了理解人类的集体合作行动，经济学家进行了大量的研究，也取得了丰硕的成果。然而，经典的理性选择模型的均衡假设往往过于苛刻，同时也在预测和解释实验观察方面遇到了一定的困难。有鉴于此，Ostrom（1998）公开呼吁研究集体合作的学者需要重视来自行为经济学的理论创新和实验发现。作为一个回应，本文的主要贡献在于以下两个方面。第一，虽然文献中已有一些从行为经济学的角度理解合作行为的理论工作，但它们主要关注的是博弈者在现实中的行为动机（或目标函数）与传统理性模型的差异，如利他主义、互惠激励等（参见如 Sethi and Somanathan，2006；韦倩等，2019）。本文关注的则是集体合作博弈的参与者在决策过程中的有限理性问题。第二，我们将经典的抽样均衡解概念拓展到博弈参与者具有异质性的情形，并证明了它在集体合作博弈模型中普遍具有唯一性，因此能够比基于完全理性模型的解概念带来更清晰的比较静态分析结果，便于建立可供检验的实证预测。未来的研究可以通过考虑博弈者自愿选择加入不同的群体（如通过移民）的激励，进而分析不同群体的合作水平与不同类型的博弈者的参与及合作行动的选择之间的内生关系。

① 从二阶随机占优也可得类似的结论：单纯的异质性本身并非重点，绝对成本的尾部分布（例如前文提及的潮汕宗族文化中的社会救助以及慈善体系）等细节因素却十分关键。比如期望相同的两个分布，设 F 二阶随机占优 G，说明 G 比 F 异质程度更高，容易构造出例子，F 的合作程度既可能高于 G，也可能低于 G。

参 考 文 献

[1] Alesina, A., Y. Algan, P. Cahuc, and P. Giuliano, "Family Values and the Regulation of Labor", *Journal of the European Economic Association*, 2015, 13, 599-630.

[2] Anauati, M. V., B. Feld, S. Galiani, and G. Torrens, "Collective Action: Experimental Evidence", *Games and Economic Behavior*, 2016, 99, 36-55.

[3] Arigapudi, S., Y. Heller, and I. Milchtaich, "Instability of Defection in the Prisoner's Dilemma Under Best Experienced Payoff Dynamics", *Journal of Economic Theory*, 2021, 197, 105174.

[4] Aumann, R., and A. Brandenburger, "Epistemic Conditions for Nash Equilibrium", *Econometrica*, 1995, 63, 1161-1180.

[5] Boyd, R., H. Gintis, S. Bowles, and P. J. Richerson, "The Evolution of Altruistic Punishment", *Proceedings of the National Academy of Sciences*, 2003, 100, 3531-3535.

[6] Burlando, R. M., and F. Guala, "Heterogeneous Agents in Public Goods Experiments", *Experimental Economics*, 2005, 8, 35-54.

[7] Carlsson, H., and E. Van Damme, "Global Games and Equilibrium Selection", *Econometrica*, 1993, 61, 989-1018.

[8] 陈友义，"潮汕慈善文化形成的主要因素探析"，《汕头大学学报：人文社会科学版》，2014年第30期，第18—23页。

[9] Costa-Gomes, M. A., and G. Weizsäcker, "Stated Beliefs and Play in Normal-Form Games", *The Review of Economic Studies*, 2008, 75, 729-762.

[10] 崔巍、陈琨，"社会信任对经济增长的影响——基于经济收敛模型的视角"，《经济与管理研究》，2016年第37期，第14—22页。

[11] Durlauf, S., and M. Fafchamps, "Social Trust", In: Aghion, P. and S. Durlauf (eds.), *Handbook of Economic Growth*. Amsterdam: Elsevier, 2005, 1639-1699.

[12] Fischbacher, U., S. Gächter, and E. Fehr, "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment", *Economics Letters*, 2001, 71, 397-404.

[13] Fischbacher, U., and S. Gächter, "Heterogeneous Social Preferences and the Dynamics of Free Riding In Public Goods", *IZA Discussion Paper*, No. 20112006.

[14] Francois, P., and J. Zabojnik, "Trust, Social Capital, and Economic Development", *Journal of the European Economic Association*, 2005, 3, 51-94.

[15] Friedman, E., and J. Ward, "Stochastic Choice and Noisy Beliefs in Games: An Experiment", *Mimeo*, 2019.

[16] Goerg, S. J., and R. Selten, "Experimental Investigation of Stationary Concepts in Cyclic Duopoly Games", *Experimental Economics*, 2009, 12, 253-271.

[17] Hardin, G., "The Tragedy of the Commons", *Science*, 1968, 162, 1243-1248.

[18] Hirsch, M. W., and S. Smale, *Differential Equations, Dynamical Systems and Linear Algebra*. San Diego: Academic Press, 1974.

[19] 黄少安、张苏，"人类的合作及其演进：研究综述和评论"，《中国社会科学》，2013年第7期，第79—91页。

[20] Kennedy, D., and C. Norman, "What Don't We Know", *Science*, 2005 309, 75.

[21] Knack, S., and P. Keefer, "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation", *The Quarterly Journal of Economics*, 1997, 112, 1251-1288.

[22] 刘凤委、李琳、薛云奎，"信任，交易成本与商业信用模式"，《经济研究》，2009年第8期，第60—72页。

[23] 卢盛峰、陈悦，"语言的力量：讲本地话增进了流动人口的社会融合吗?"，《经济科学》，2019年第4期，第118—128页。

[24] 吕朝凤、陈汉鹏、S. L'opez-Leyva，"社会信任，不完全契约与长期经济增长"，《经济研究》，2019年第3期，第4—20页。

[25] 马双、赵文博，“方言多样性与流动人口收入——基于 CHFS 的实证研究”，《经济学》（季刊），2019 年第 1 期，第 393—414 页。

[26] Medina, L. F., *A Unified Theory of Collective Action and Social Change*. Ann Arbor: University of Michigan Press, 2007.

[27] Morris, S., and H. Shin, "Global Games: Theory and Applications", In: Dewatripon, M., L. P. Hansen, and S. Turnovsky (eds.), *Advances in Economics and Econometrics*. Cambridge: Cambridge University Press, 2003, 56-114.

[28] Olson, M. J., *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge: Harvard University Press, 1965.

[29] Osborne, M. J., and A. Rubinstein, "Games with Procedurally Rational Players", *American Economic Review*, 1998, 88, 834-847.

[30] Osborne, M. J., and A. Rubinstein, "Sampling Equilibrium, with an Application to Strategic Voting", *Games and Economic Behavior*, 2003, 45, 434-441.

[31] Ostrom, E., "A Behavioral Approach to the Rational Choice Theory of Collective Action: Presidential Address, American Political Science Association, 1997", *American Political Science Review*, 1998, 92, 1-22.

[32] Ostrom, E., "Tragedy of the Commons", *The New Palgrave Dictionary of Economics*, 2008, 2.

[33] 潘越、宁博、纪翔阁、戴亦一，“民营资本的宗族烙印：来自融资约束视角的证据”，《经济研究》，2019 年第 7 期，第 94—110 页。

[34] Riley, J. G., *Essential Microeconomics*. Cambridge: Cambridge University Press, 2012.

[35] Rowthorn, R., and R. Sethi, "Procedural Rationality and Equilibrium Trust", *The Economic Journal*, 2008, 118, 889-905.

[36] Salant, Y., and J. Cherry, "Statistical Inference in Games", *Econometrica*, 2020, 88, 1725-1752.

[37] Sandholm, W. H., S. S. Izquierdo, and L. R. Izquierdo, "Stability for Best Experienced Payoff Dynamics", *Journal of Economic Theory*, 2020, 185, 104957.

[38] Schelling, T., *The Strategy of Conflict*. Cambridge: Harvard University Press: 1960.

[39] Selten, R., and T. Chmura, "Stationary Concepts for Experimental 2x2-Games", *American Economic Review*, 2008, 98, 938-966.

[40] Sethi, R., "Stability of Equilibria in Games with Procedurally Rational Players", *Games and Economic Behavior*, 2000, 32, 85-104.

[41] Sethi, R., and E. Somanathan, "A Simple Model of Collective Action", *Economic Development and Cultural Change*, 2006, 54, 725-747.

[42] 韦倩、孙瑞琪、姜树广、叶航，“协调性惩罚与人类合作的演化”，《经济研究》，2019 年第 7 期，第 140—151 页。

[43] Zak, P. J., and S. Knack, "Trust and Growth", *The Economic Journal*, 2001, 111, 295-321.

[44] 张航、范子英、杨现领，“方言、文化认同与住房市场交易”，《经济学》（季刊），2021 年第 2 期，第 693—712 页。

[45] 周业安、连洪泉、陈叶烽、左聪颖、叶航，“社会角色，个体异质性和公共品自愿供给”，《经济研究》，2013 年第 1 期，第 123—136 页。

Cooperation Between Boundedly Rational Agents: Theory and Applications

HU Tao LIU Shuo*
(Peking University)

Abstract: We study a model of cooperation between boundedly rational players, where they choose actions according to the outcome of a sampling procedure. We solve the model by generalizing the concept of sampling equilibrium (Osborne and Rubinstein, 1998) to settings with agent heterogeneity. We analyze the existence, uniqueness and stability of the equilibrium, and show that its comparative statics crucially depend on the distribution of players' costs of cooperative efforts. Our results can shed light on topics such as the relationship between trust and growth, and the effect of culture on economic behavior.

Keywords: cooperation; bounded rationality; sampling equilibrium
JEL Classification: D82, C70, H41

* Corresponding Author: Liu Shuo, New Guanghua Building, Peking University, No. 5 Yiheyuan Road, Beijing 100871, China; Tel: 86-10-62747635; E-mail: shuo. liu@gsm. pku. edu. cn.