无限次重复社会困境博弈中的新策略和新方法

赵衍才
无锡城市职业技术学院，江苏 无锡 214153
DOI:10.61369/SE.2025090010

摘　　要： 以囚徒困境为代表的社会困境问题一直是博弈论关注的焦点之一，其中的冷酷策略不允许纠错和试探的机会，从而导致一旦一次失误就会永远不合作的结论。但另一方面，现实世界中的重复博弈存在广泛的合作现象，这就导致理论与现实的不一致。本文首先将传统的冷酷策略加以改进，提出了一种新的策略，3- 触发策略；然后证明了双方都采用 3- 触发策略是无限次社会困境博弈的纳什均衡。本文尝试用新的策略和新的方法探讨无限次重复博弈现象中的合作演化机制，以更好地解释现实中的重复博弈现象，探索社会困境问题的解决方法。
关 键 词： 囚徒困境；重复博弈；冷酷策略；3- 触发策略

New Strategies and Methods for Infinite Repetition of Social Dilemma Games

Zhao Yancai
WuXi City College of Vocational Technology, Wuxi, Jiangsu 214153

Abstract： Social dilemmas, epitomized by the Prisoner's Dilemma, have long been a focal point in game theory. The Cold Strategy, which prohibits error correction and trial-and-error opportunities, leads to a conclusion where a single misstep results in permanent non-cooperation. However, repeated games in the real world demonstrate widespread cooperative behavior, creating a discrepancy between theoretical predictions and practical outcomes. This paper first improves the traditional Cold Strategy by proposing a novel 3-trigger strategy. It then proves that both parties adopting this 3-trigger strategy constitutes a Nash equilibrium in infinite-repetition social dilemma games. By employing this innovative strategy and methodology, the study explores cooperative evolution mechanisms in infinite-repetition game scenarios, aiming to better explain real-world repeated game phenomena and develop solutions for social dilemma challenges.
Keywords： **prisoner's dilemma; repeated games; cold strategy; 3-trigger strategy**

引言

博弈论的研究在经济学，社会学，计算机科学，生物学和物理学的领域里的探究起着非常重要的作用 [1]。本文的概念和术语主要来自于文献 [2-6]。

以囚徒困境为代表的社会困境问题在人类经济和社会活动中普遍存在，这种困境反映了个体的最优行为与集体最优行为的不一致，如何化解社会困境已成为整个社会科学关注的焦点。用博弈论的语言可将社会困境问题描述为：两人（或多人）形成的纳什均衡策略不是集体最优的，至少存在一个帕累托改进，即通过合作，在原有的纳什均衡基础上增加所有人的收益 [10]。因此，社会困境问题向我们提出了这样一个重要的科学问题：何种机制可以实现社会或经济系统中的合作红利？

在经济社会生活中大量存在重复博弈。重复博弈中有一个经典的演化策略是：开始试图合作，一旦一方选择了不合作，另一方随即也以不合作进行报复并一直进行下去，在博弈论里面称之为冷酷策略。文章 [14] 研究了重复博弈中的另一种演化模式，轮换策略，文章 [15-16] 对不同领域中的重复博弈进行了深入研究。

传统的冷酷策略不允许试探和纠错，因此不利于合作的达成，这与现实中广泛存在的合作现象不符。本文中，我们对传统的冷酷策略进行改进，提出 3- 触发策略，这种策略允许试探和纠错，从而有利于合作的达成。我们证明了双方都采用 3- 触发策略是无限次囚徒困境博弈的纳什均衡。本文尝试用新的策略和新的方法探讨重复博弈现象中的合作演化机制，以更好地解释现实中的重复博弈现象，探索社会困境问题的解决方法。

作者简介：赵衍才（1969—），江苏无锡人，汉族，教授，博士，研究方向：运筹优化，博弈论。

一、重复社会困境博弈

囚徒困境 (prisoners' dilemma) 是博弈论中最著名的博弈论模型之一，它是实际生活中许多现象的一个抽象概括，也因此被许多文章所研究[8-13]

囚徒困境讲的是两个嫌疑犯作案后被警察抓住，分别被关在不同的屋子里接受审讯。每个人都有两种选择：揭发对方和保持沉默。如果两人都揭发对方，则因都有罪各判刑8年，如果两人都沉默，则因证据不足各判刑1年，如果其中一人揭发对方但对方沉默，则因揭发有功抗拒从严，揭发的放出去，沉默的判刑10年。

囚徒困境的纳什均衡是(揭发,揭发)：即不论乙揭发还是沉默，甲的最优策略都是揭发；同样，不论甲揭发还是沉默，乙的最优策略也是揭发。囚徒困境反映了一个很深刻的问题，这就是个人理性与集体理性的矛盾。如果两个人都沉默,各判刑1年，显然比都揭发各判刑8年好。但这个帕累托最优却不能达到,因为任何一方若单独改变策略将遭受更大的损失。

为了便于对以囚徒困境为代表的社会困境问题进行一般化的分析，我们将具体的囚徒困境模型一般化为下面的社会困境模型，其得益矩阵如表1。而且我们按照通常的做法，将其中的得益值均设置为正数。

表1：社会困境模型

乙甲	合作	不合作
合作	k_1a,k_1a	$0,k_2a$
不合作	$k_2a,0$	a,a

在上述社会困境模型中，始终假设$a>0$且$k_2>k_1>1$。

读者不难发现，同囚徒困境一样，上述社会困境类型的单次博弈的纳什均衡是(不合作,不合作)，即不论乙合作还是不合作，甲的最优策略都是不合作；同样，不论甲合作还是不合作，乙的最优策略也是不合作。

在考虑重复博弈各阶段得益时，我们考虑一个更为一般的情况：重复博弈每个阶段的得益有时间上的先后之分，这在只有少数几次重复而且每次重复间隔时间并不很长的情况下可能并不重要，但如果重复次数很多，而且每次重复间隔时间又较长的有限次重复博弈，或者是无限次重复博弈，得益的时间先后就不能不考虑[19]。解决这一点的流行方法是将后一阶段的得益乘以某个贴现系数 $\delta(0<\delta<1)$，折算成当前阶段的得益(现值)[5,17]。有了贴现系数δ，若一个T次重复博弈的某博弈方在某博弈路径各阶段得益分别为$\pi_1,\pi_2,\cdots,\pi_T$，则该博弈方的总得益的现值为：

$$\pi=\pi_1+\delta\pi_2+\delta^2\pi_3+\cdots+\delta^{T-1}\pi_T=\sum_{t=1}^{T}\delta^{t-1}\pi_t$$

在一个无限次重复博弈中，设某博弈方各阶段得益为$\pi_1,\pi_2,\cdots$,则该博弈方总得益的现值就是：

$$\pi=\pi_1+\delta\pi_2+\delta^2\pi_3+\cdots=\sum_{t=1}^{\infty}\delta^{t-1}\pi_t$$

重复博弈理论的最大贡献是对人们之间的合作行为提供了理性解释：在囚徒困境中，一次博弈的唯一纳什均衡是不合作即（不合作，不合作），但如果博弈多次重复，合作就可能出现[20]。

二、冷酷策略下的重复社会困境博弈

对于在无限次重复博弈中博弈方所采用的策略，目前最流行的是冷酷策略：即首先试探合作，一旦发觉对方不合作则以一直不合作相报复的策略。冷酷策略是重复博弈中实现合作和提高均衡效率的关键机制，是重复博弈分析的重要构件之一[17]。对于有限次重复囚徒困境博弈，目前所用的是逆推归纳法。关于囚徒困境和重复博弈的更多研究见文献[7-11]。下面的几个定理来自于文献[4-5]。

定理A 单次囚徒困境博弈中，（揭发，揭发）是其唯一的纳什均衡。

定理B 囚徒困境博弈无限次，双方都采取冷酷策略是一个纳什均衡，且一直（揭发，揭发）是其纳什均衡结果。

对于无限次囚徒困境博弈，据作者所知，现有的文献或者是没有提及贴现因子δ（实际上是$\delta=1$），或者虽有提及但是只研究了当$\delta>1/8$的情况。下面我们对$\delta>1/8$时的纳什均衡增加一种策略组合，并对$\delta<1/8$的情况也进行研究，下面的定理可看做是对现有定理B的一个补充完善，从而形成对无限次囚徒困境博弈纳什均衡的一个更为完整的表述。

定理3.1（1）当$\delta>1/8$时，双方都采取冷酷策略是无限次囚徒困境博弈的一个纳什均衡，且一直（不合作，不合作）是一个纳什均衡结果；（2）当$\delta<1/8$时，双方一直互相揭发是无限次囚徒困境博弈的一个纳什均衡，双方都采取冷酷策略不是无限次囚徒困境博弈的一个纳什均衡。

证明。（1）是已有定理B的结论，故下面证明（2）。

（2）中所述当$\delta<1/8$时双方都采取冷酷策略不是无限次囚徒困境博弈的一个纳什均衡，我们对其进行证明的方法是，可以先假设甲方已经采用了冷酷策略，然后证明当$\delta>1/8$时，同样采用冷酷策略也是乙方的最优策略，但当$\delta<1/8$时，采用冷酷策略不是乙方的最优策略。

设甲方选择冷酷策略，即一直采用沉默，直到乙方采用了揭发，则甲以一直揭发进行报复。我们来分析乙方的最优选择。

设乙在第 $k(k\geq1)$个阶段首次选择了揭发，则乙在第 k 个阶段得到的收益是0而不是-1，这样对他似乎有利，但他的这个机会主义行为将触发甲的此后永远揭发的惩罚，从而自己也只能在此后的阶段一直采用揭发[18]。因此乙从第 $k+1$个阶段以后的收益都是-8，其总收益为

$$\pi(k)=-1+\delta-1+\cdots+\delta^{k-2}(-1)+\delta^{k-1}0+\delta^{k}-8+\cdots$$

$$=\sum_{0\leq i\leq k-2}\delta^{i}(-1)+\sum_{i\geq k}\delta^{i}(-8)$$

$$=(-1)\frac{1-\delta^{k-1}}{1-\delta}+(-8)\frac{\delta^{k}}{1-\delta}$$

$$=-\frac{1-\delta^{k-1}+8\delta^{k}}{1-\delta}$$

将上式对 k 求导数，有

$$\pi'(k) = -\frac{-\delta^{k-1}ln\delta + 8\delta^{k}ln\delta}{1-\delta}$$

$$= \frac{\delta^{k-1}(1-8\delta)ln\delta}{1-\delta}$$

若 $\delta > 1/8$，则 $\pi'(k) > 0$，函数 $\pi(k)$ 为增函数，则当 $k \to \infty$，即乙永远选择沉默时，$\pi(k)$ 最大。这说明，当 $\delta > 1/8$ 时，只要甲采用触发策略，乙的最优策略也是触发策略。同理，只要乙采用触发策略，则甲的最优策略也是触发策略。双方都采用触发策略构成一个纳什均衡。

若 $\delta < 1/8$，则 $\pi'(k) < 0$，函数 $\pi(k)$ 为减函数，则当 $k = 1$，即乙方永远选择揭发时，$\pi(k)$ 最大。这说明，当 $\delta < 1/8$ 时，虽然甲方采用冷酷策略，但乙方的最优策略不是冷酷策略，而是一直揭发。同理，假定乙方采用冷酷策略，但甲方的最优策略不是冷酷策略，而是一直揭发。这说明当 $\delta < 1/8$ 时双方都采用冷酷策略不构成纳什均衡，一直互相揭发是其纳什均衡。

三、3- 触发策略下的社会困境博弈

首先，我们将传统的冷酷策略看作是1- 触发策略，进而提出一种新的策略，3- 触发策略；然后证明双方都采用3- 触发策略是无限次社会困境博弈的纳什均衡；最后对于有限次社会困境博弈，我们引入冷酷策略，从总收益的角度用逆推的方法证明了有限次社会困境博弈的一个均衡状态，这个均衡状态中出现了更多的相互合作。本文尝试用新的策略和新的方法探讨重复博弈现象中的合作演化机制，以更好地解释现实中的重复博弈现象，探索社会困境问题的解决方法。

定义 3- 触发策略。一开始双方都采取合作的态度，即（合作，合作），直到一方率先选择了不合作，即出现了（不合作，合作）或（合作，不合作）或（不合作，不合作），不妨以（不合作，合作）的情形为例。则在第二阶段，选择不合作的甲方将被乙方所原谅，乙方将甲方的这次不合作暂时理解为一次失误，此时的局面是（X，合作），其中 X 可能是合作，也可能是不合作。第三阶段及后续，如果 X 是合作，则说明甲纠正了失误，则乙在第三阶段继续选择合作，但在后续阶段一旦甲再次出现不合作，则触发乙以一直不合作作为惩罚；如果 X 是不合作，则这次是甲的第二次不合作，乙认为甲是故意不合作，触发乙在第三阶段及后续以一直不合作作为惩罚。同理，在第四阶段及后续，甲方也类似地原谅乙方一次。对于其它两种情形，可类似给出3- 触发策略的定义。

显然，3- 触发策略有利于达成合作，这也是“最大诚意”原则的一个落实和体现。

下面论述3- 触发策略在无限次社会困境博弈中的应用。我们首先给出3- 触发策略下无限次社会困境博弈的纳什均衡如下。

定理4.1对于无限次社会困境博弈，在双方都恰好存在一次随机的失误（即不合作）的情况下，双方都采取3- 触发策略是无限次囚徒困境博弈的一个纳什均衡，且纳什均衡结果具有下面的三种形式之一。

序列1:（合作，合作）---（合作，不合作）（合作，合作）---（不合作，合作）（合作，合作）---；

序列2:（合作，合作）---（不合作，合作）（合作，合作）---（合作，不合作）（合作，合作）---；

序列3:（合作，合作）---（合作，合作）（不合作，不合作）（合作，合作）---。

证明。我们以序列1为例进行证明。证明双方都采取3- 触发策略是纳什均衡的方法是，先假设甲方已经采用了3- 触发策略，然后证明，采用3- 触发策略是乙方的最优反应。

首先可以断定，乙方不会故意选择不合作，因为否则，这次故意不合作加上因随机失误一次导致的不合作，乙方就会有两次不合作，而这会导致甲方的一直报复。在甲方一直报复的情况下，乙方也只能一直选择不合作。而这样的情况下乙方的总收益显然小于序列1下乙方的总收益，故乙方不会故意选择不合作，从而在未出现失误的情况下，双方会一直合作。

当序列1中乙方因失误出现不合作时，即（合作，不合作）时，由于甲方已采取了3- 触发策略，故乙清楚，在下个阶段甲的决策是选择合作原谅乙方，则乙的最优反应必定是合作，因为否则，会触发甲的一直报复，从而导致乙的总收益减少。乙的这个最优反应与3- 触发策略一致。

当序列1中后续阶段甲第一次出现不合作，即（不合作，合作）时，按照假设条件，这是甲的唯一一次失误。根据甲的3- 触发策略，甲接下来肯定要选择合作以纠正失误，乙也清楚这一点，那么乙的最优反应也必定是合作，因为否则，乙前面已经有一次不合作，这次的不合作会触发甲的一直报复，从而导致乙的总收益减少。乙的这个最优反应与3- 触发策略也是一致的。

总之，这意味着乙的最优策略也是3- 触发策略，从而双方都采用3- 触发策略是整个博弈的纳什均衡。同理可以证明在序列2和序列3的情形下，双方都采用3- 触发策略也是整个博弈的纳什均衡。

事实上，在双方都不会出现失误的情况下，即理想状态下，双方都采取3- 触发策略也是无限次囚徒困境博弈的一个纳什均衡。

定理4.2对于无限次社会困境博弈，在双方都不存在失误的情况下，双方都采取3- 触发策略是整个博弈的纳什均衡，且均衡结果为双方一直合作。

证明。我们首先断言，在双方都不会失误的情况下，按照3- 触发策略，尽管每个人都有一次被原谅的机会，但都不会选择故意的不合作，因为否则，假如甲首先选择了一次故意的不合作，乙会原谅甲的这次不合作，但在后面的阶段，乙肯定也要选择一次故意的不合作，当然甲也会原谅乙的这次不合作，但是，这样对应的序列中，双方的总得益显然小于双方一直合作的总得益，故双方都不会选择故意的不合作。

其它的证明类似于定理4.1。

对比传统的冷酷策略，3- 触发策略有以下改进：

（1）3- 触发策略比冷酷策略更易达成合作。虽然定理 B 中的无限次囚徒困境博弈的纳什均衡状态也是合作，但那是理想状态下，因为它没有纠错和试探机制。事实上，现实中的很多不合作或互相伤害现象就是因为互相不给对方纠错和试探的机会而造成的。

（2）3- 触发策略是一个较为合理的“亮剑”模式。亮剑是必须的，但仓促的亮剑效果也不好。3- 触发策略允许“试探”的过程，但又不能让试探一直进行下去，若对方多次不合作后仍不采取惩罚措施，就没有威慑，那对手就可能会选择更多的不合作，从而导致自身利益不断受损。只有让对手意识到触发机制的存在，才能促成合作局面。3- 触发策略也正应了 “事不过三”。

参考文献

[1] 袁硕，郭雷 . 随机自适应动态博弈 [J]. 中国科学，数学， 2016，3：1367-1382.

[2]Joel Wawon. Strategy：an introduction to game theory，Third edition [M]. W. W. Norton company，2013.

[3]Robert Gibbons. A Primer in Game Theory，Third edition[M]，Pearson Academic，1992.

[4] 常金华，陈梅 . 博弈论通识十八讲 [M]. 北京：北京大学出版社， 2017年 .

[5] 张维迎 . 博弈论与信息经济学 [M]. 上海：上海三联书店，上海人民出版社，1996年 .

[6] 谢识予 . 经济博弈论 [M]. 上海：复旦大学出版社， 2003年 .

[7] 全吉，周亚文，王先甲，社会困境博弈中群体合作行为演化研究综述 . 复 杂 系 统 与 复 杂 性 科 学，2020，17(1)：1-14.

[8] 刘华，李莹，赵建立，葛美侠 . 沉默策略对囚徒困境博弈合作水平的影响 . 数学的实践与认识，2016，46 (20)：240-246.

[9]Zeng Weijun，Ai Hongfeng，Zhao Man. A symmetrical expectations of future interaction and cooperation in the iterated prisoner’s dilemma game [J]. Applied mathematics and computation，2019，359：148-164 .

[10]Wang Xianjia，Lv Shaojie . The roles of particle swarm intelligence in the prisoner’s dilemma based on continuous and mixed strategy systems on scale-free networks[J]. Applied mathematics and computation，2019，355：213-220 .

[11]Lu Fuxiao，Tang Wansheng， Liu Guowei，et al. Cooperative advertising：A way escaping from the prisoner’s dilemma in a supply chain with sticky price [J]. Omega-international journal of management science，2019，86：87-106.

[12] 黄少安，张苏. 人类的合作及其演进研究. 中国社会科学，2013，7：77-89.

[13] 张衔，魏中许. 如何破解人类合作之谜——与黄少安教授商榷. 中国社会科学，2016,8：90-94.

[14] 汪燕，重复博弈中触发策略与轮换策略的比较 . 市场论坛，2016，2：48-50.

[15] 单而芳，曾晗，蔡 蕾 . 互联网租车市场的演化博弈分析 . 运 筹 与 管 理，2020，29 (8)： 12-19.

[16] 牛文举，夏晶 . 面向策略型消费者的产品创新与定价策略 . 运筹与管理，2021，30(5)：154-160.

[17] 博弈论（11）下载 . 互联网文档资源（ https://ishare.iask.）.

[18] 孙昌庆，廖瑞华 . 企业管理软件开发项目收付款问题的博弈分析 . 中国管理信息化，2022，25(3)：91-94.

[19] 叶枫，吕世平 . 重复博弈：“合作”实现中美纺织品贸易博弈的双赢 . 云南财贸学院 .

