第七讲：重复博弈（一）

重复博弈

- 当行动者之间的互动不是一次性的，而是会多次重复发生
- “未来”的存在是否能改变行动者此刻的策略？

例：囚徒困境(PD)

Prisoner’s Dilemma

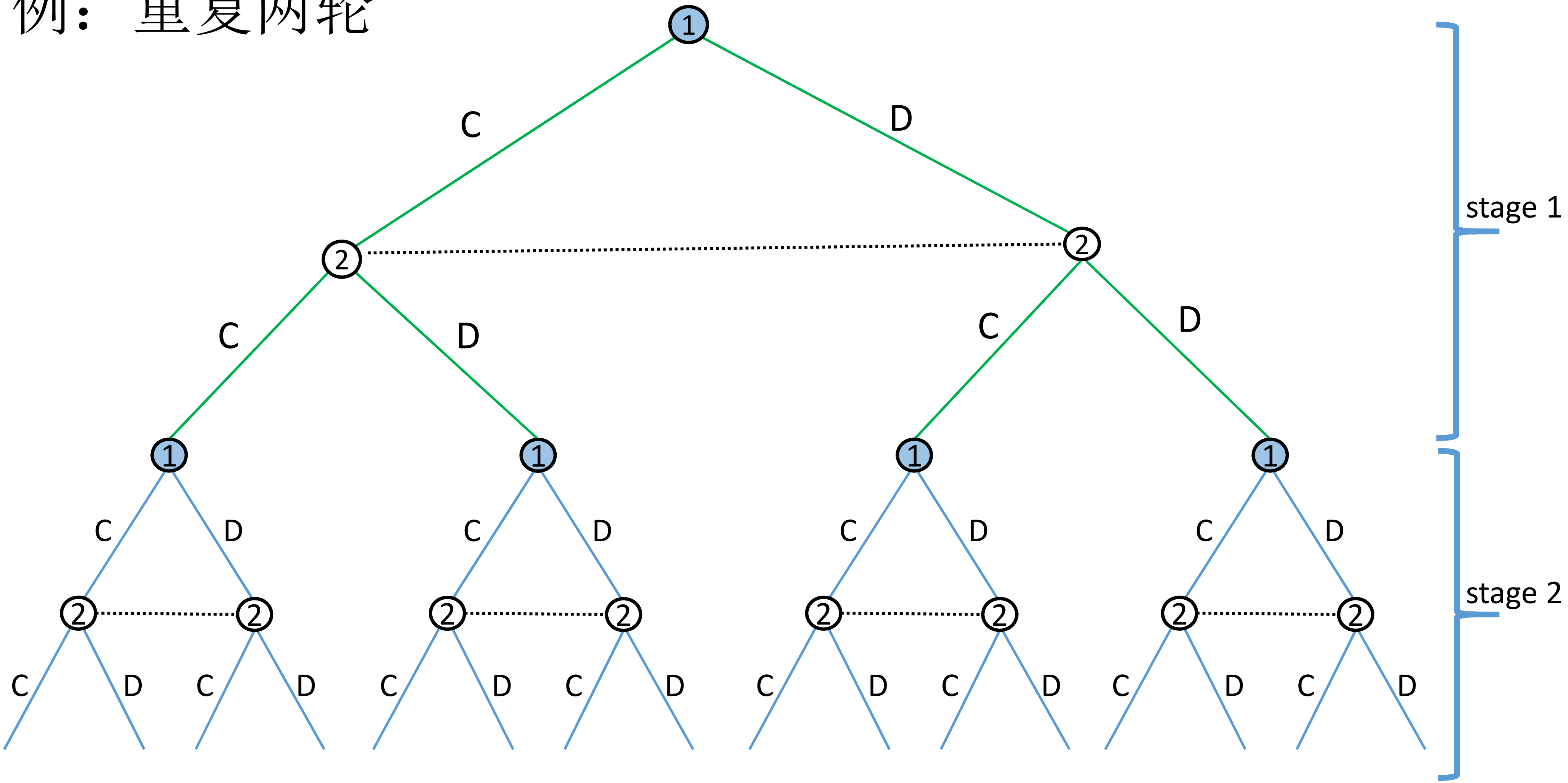
	C	D
C	2, 2	0, 3
D	3, 0	1, 1

- 将囚徒困境重复T次，甚至无限次

$$\boxed{PD} \longrightarrow \boxed{PD} \longrightarrow \boxed{PD} \longrightarrow \cdots$$

- 被重复的这个一次性博弈称为**阶段博弈(stage-game)**
  - 每一次博弈称为一个阶段（stage）、时期（period）或者一轮（round）

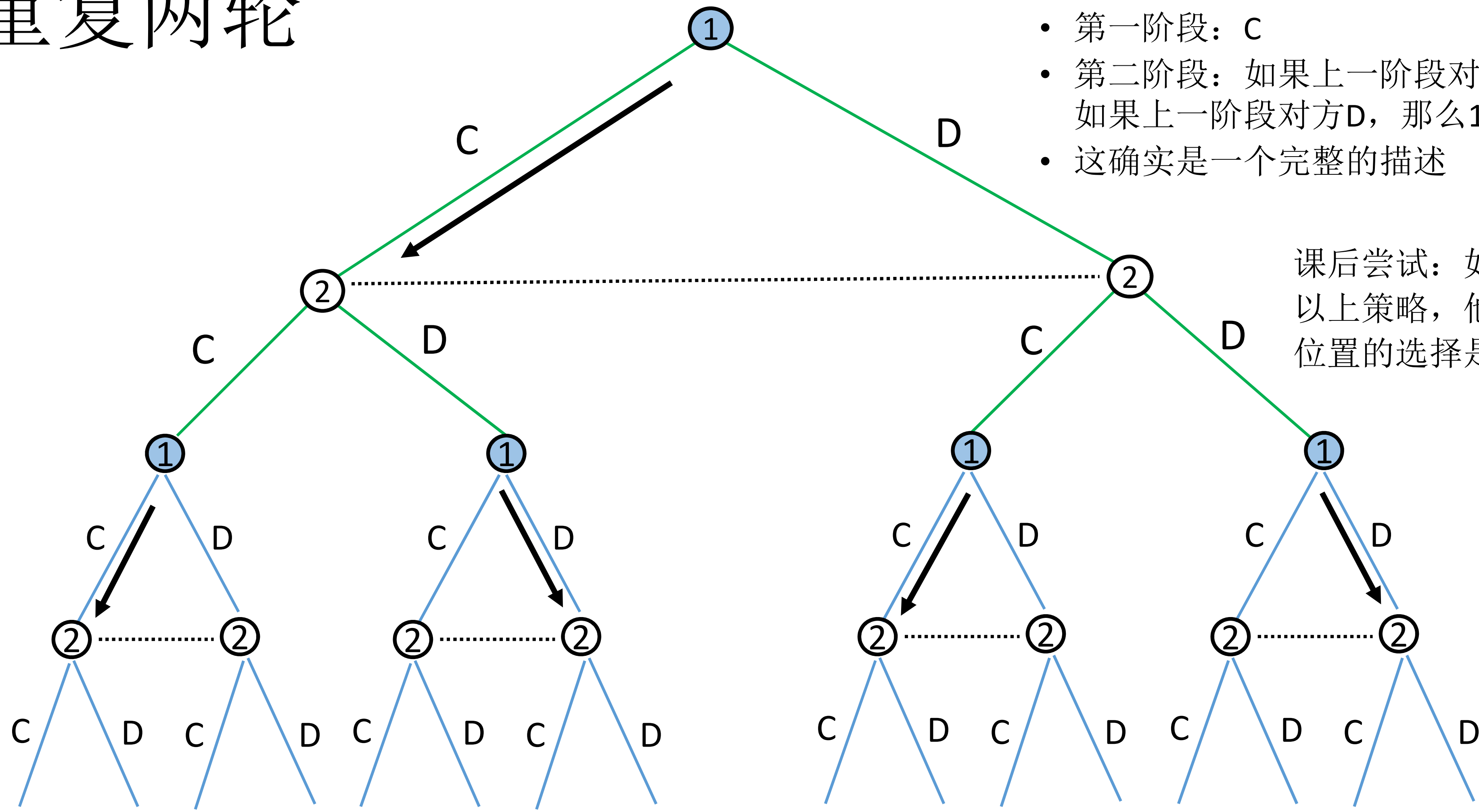
例：重复两轮


1
C
D
2
2
C
D
C
D
1
1
1
1
C
D
C
D
C
D
C
D
2
2
2
2
2
2
2
2
C
D
C
D
C
D
C
D
C
D
C
D
C
D
C
D
stage 1
stage 2

重复博弈中的策略

- 注意博弈者i的策略需指明在属于她的每一个信息集上如何行动
- 由于博弈树非常庞大，我们通常不是依赖博弈树而是对策略进行简练的（但需要是完整的）描述
- 例如，刚才重复两次的例子中，1的一个可能策略是“一报还一报”（tit-for-tat）
  - 第一阶段：合作(C)
  - 第二阶段：如果上一阶段对方合作(C)，那么1继续合作(C)；如果上一阶段对方背叛(D)，那么1也背叛(D)
- 以上的策略描述是否完整的？
  - 可试着考虑所有可能的历史 CC, CD, DC, DD 之后，该策略是否都指明了1该如何行动？

重复两轮



1的策略：tit-for-tat

- 第一阶段：C
- 第二阶段：如果上一阶段对方C，1继续C；如果上一阶段对方D，那么1也使用D
- 这确实是一个完整的描述

课后尝试：如果2也执行以上策略，他在各个决策位置的选择是什么？

反复使用纳什均衡策略：一定构成均衡

- **定理**：在任意一个（有限次或无限次）重复博弈中，在每一阶段都重复使用阶段博弈的一个纳什均衡策略组合，这构成该重复博弈的一个子博弈完善均衡
  - 例如：在重复的囚徒困境博弈中，双方在每一阶段都使用DD，构成一个子博弈完善均衡
- 证明思路
  - 有限重复：逆推
  - 无限重复：应用稍后介绍的“一次性偏离”原则（课后尝试自己证明）
- 除此之外，有可能出现其他新的均衡吗？
  - 区分有限次重复和无限次重复两种情况

有限重复(一)：当阶段博弈只有唯一的纳什均衡

- 例如：在重复有限次（T次）的囚徒困境博弈中，可能出现其他的均衡吗？
- 逆推归纳法
  - 在博弈的最后一个阶段，双方都会使用D
  - 给定这一点，在倒数第二阶段，双方的最优策略是D
  - …
- 可见：这里**唯一的**子博弈完善均衡是双方在每一阶段都使用D
- 推广：任何**有限次重复**的博弈，如果其阶段博弈只有**唯一**的纳什均衡，那么整个博弈只有唯一的子博奕完善均衡，即反复使用阶段博弈的纳什均衡

有限重复(二)：当阶段博弈有多个纳什均衡

- 虽然是有限重复，但如果阶段博弈具备**不止一个**纳什均衡，那么整个博弈可以出现新的子博弈完善均衡

- 基本思路：利用“好的”纳什均衡作为“奖励”，“差的”纳什均衡作为“惩罚”，引导博弈者在前面的阶段里选择所希望的博弈路径

例：阶段博弈有两个纳什均衡

	A	B	C
A	4, 4	0, 5	0, 0
B	5, 0	1, 1	0, 0
C	0, 0	0, 0	3, 3

- 右上的博弈有两个纳什均衡：(B, B), (C, C)
- (A, A)是一个对双方都更好的结果，但不是纳什均衡
- 如果该博弈**重复两次**，(A, A)有可能出现在均衡（的某个阶段）中吗？
- 基本思路：以较好的均衡(C, C)为奖励；以较差的均衡(B, B)为惩罚
- 考察如下策略组合，其中每一方的策略是
  - 第一阶段：使用A
  - 第二阶段：如果第一阶段的结果是AA，使用C；否则使用B
- 如果双方执行该策略组合，结果是第一阶段(A, A)，第二阶段(C, C)
- 该策略组合是否构成子博弈完善均衡？

策略组合，其中每一方的策略是

- 第一阶段：使用A
- 第二阶段：如果第一阶段的结果是AA，使用C；否则使用B

	A	B	C
A	4，4	0，5	0，0
B	5，0	1，1	0，0
C	0，0	0，0	3，3

- 可用逆推，判断是否构成子博弈完善均衡
- 第二阶段（给定对方策略，1的策略是否最优？）
  - 前一阶段是AA的话：对方使用C，1使用C显然最优
  - 前一阶段不是AA的话（是另外八个结果之一）：对方使用B，1使用B显然最优
- 第一阶段
  - 给定对方使用A，如果1**不偏离**指定策略：本阶段AA，收益4，下阶段CC，收益3，总收益4+3=7
  - 如果1**偏离**指定策略，在本阶段最大收益是5，但在下一阶段（对方将使用B），1最多只能获得1，总收益最大为6
  - 所以1没有动力偏离现策略

利用多重纳什均衡建立“奖励”和“惩罚”

- 在有限次重复博弈中，**如果阶段博弈具备不止一个纳什均衡**，那么可以利用“好的”纳什均衡作为后面的奖励，“差的”纳什均衡作为惩罚，引导博弈者在前面选择某种指定的博弈路径
- 需要“好的”和“差的”纳什均衡的收益差别足够大，以克服博弈者在之前阶段里所面临的偏离指定路径的“诱惑”

- 但是，如果阶段博弈只有唯一一个纳什均衡（例如囚徒困境），有限次重复博弈将无法产生新的均衡
- 考虑无限重复：博弈的重复次数是无限的，或者没有一个明确的终止时间
  - 例如，每次博弈之后，继续进行下一轮的可能性是δ
- 准备工作
  - 在无限重复博弈中，一般需要对未来的收益进行打折(discounting)

对未来收益打折 (Discounting)

- 同样的结果（例如吃一个蛋糕、看一部电影、或者得到100元），在**未来**实现该结果比起**现在**立刻实现该结果，对决策者（现在）来说，未来实现该结果所带来的效用**较低**
- 原因
  - 未来的不确定性：例如博弈是否会进行到下一个时间段
  - 人们的心理倾向：更偏好即时满足，而不是延迟满足
  - 利息的存在：今天的100元，存入银行，明年可以获得110元
- 对未来的结果，计算其“现值”时要打折——使用折扣系数或称贴现系数 (discount factor): δ（$0 < \delta < 1$）
  - 如果现阶段（阶段1）获得收益 u，其现值为 u
  - 如果下一阶段（阶段2）获得收益 u，折现值: δu
  - 如果在阶段t获得收益 u，折现值：$\delta^{t-1} u$

重复博弈中的平均收益和总收益

- 当博弈有T个阶段，如果博弈者在每个阶段的收益都是**相同的$\boldsymbol{u}$**，总收益的现值

$$V = u + \delta^1 u + \delta^2 u \ldots + \delta^{T-1} u = \frac{1-\delta^T}{1-\delta} \cdot u$$

- 当博弈**无限重复**（$\mathrm{T} = \infty$），若博弈者每阶段的收益都是u，总收益的现值为

$$\boxed{V = \frac{u}{1-\delta}}$$

- 反过来，现阶段一次性获得收益V，则相当于（从现在开始）在无限个阶段中每个阶段获得$\boxed{u = (1-\delta)V}$，这里u 称为**平均收益**

无限重复博弈

- 无限重复时，即使阶段博弈只有唯一的纳什均衡，整个博弈也可以出现新的均衡
- 思路：和之前类似，可以通过未来的奖励和惩罚，使得博弈者选择某种策略（例如合作）
- 下面以两人的囚徒困境为例

冷酷触发(Grim Trigger)

	C	D
C	2， 2	0， 3
D	3， 0	1， 1

- 无限重复的囚徒困境，阶段博弈如右上
- 考虑一种“冷酷触发”(grim trigger) 的策略，每一位博弈者
  - 在第一阶段使用C
  - 从第二阶段开始，如果双方在过去一直使用的是C，继续使用C
  - 否则（**任何人**曾经使用过D），从本阶段起**永远**使用D

$$\boxed{P_0\text{：}\ C} \xrightarrow[\text{任何人使用过D}]{} \boxed{P_1\text{：}\ D}$$

- 如果双方使用该策略，结果为在每一个阶段中都是(C, C)
- 该策略是否构成子博弈完善均衡？

验证子博弈完善（1）：要考虑每一类信息集

- 子博弈完善要求在**每一个**信息集上（历史之后），包括那些不在博弈路径上的信息集（历史），博弈者都没有动力偏离既定的策略组合
  - 而纳什均衡只需要考虑在均衡路径上的信息集，博弈者有没有动力偏离
- 因为信息集（历史）非常多，需要根据所讨论的策略组合将信息集（历史）划分为几类，分类讨论
  - 分类时注意：划分为同一类的信息集上，既定策略组合所指定的后续策略组合需要是相同的

验证子博弈完善（2）：只需考虑“一次性偏离”

- 在每一个（类）信息集上，是不是需要考虑其后所有可能的偏离方式？
  - 从一个给定的信息集出发，由于后续博弈有很多阶段，博弈者可选的偏离方式有很多种（例如在本阶段偏离和以后的某些阶段偏离）
  - 是否需要考虑所有这些可能的偏离方式？
- 不需要，只要检查“一次性偏离”
  - 称为 “one-period deviation”, or “one-shot deviation”

“一次性偏离”原则

- **一次性偏离**(one-period deviation)：给定一个策略组合，博弈者i在某一个信息集上改变了自己（那一步）的行动，但是在随后的后续博弈中保持自己的原策略不变（注意是**策略**不变）
- **定理**：在重复博弈中（包括有限重复和折扣系数小于1的无限重复），一个策略组合是子博弈完善均衡的**充分必要条件**是：给定其他人的策略，任意一个博弈者i都无法通过某个一次性偏离来提高自己的收益
  - 证明略
- 所以，在验证一个策略组合是否为子博弈完善均衡时，只需考察对于任意的信息集，博弈者i是否有动力在**该处**一次性改变其行动，给定其他人的策略和他自己在后续步骤中的策略不变

冷酷触发 (Grim Trigger)

	C	D
C	2，2	0，3
D	3，0	1，1

- 任何一方（i）的策略
  - 第一阶段使用C
  - 之后的阶段，如果从未有人使用过D，继续使用C
  - 如果**任何人**曾使用过D，永远使用D
- 信息集（历史）只有两类：之前没有人使用过D、之前有人使用过D

冷酷触发 (Grim Trigger)

	C	D
C	2, 2	0, 3
D	3, 0	1, 1

- 之前没有人使用过D， 策略组合要求双方继续合作 CC, CC, CC…
  - 给定2的策略，如果1**不偏离**：收益流为 2, 2, 2, 2…， 总收益 $2/(1-\delta)$
  - 给定2的策略，如果1**一次性偏离**：本阶段收益3，此后每阶段收益1，收益流为 3, 1, 1, 1…， 总收益$3 + \delta/(1-\delta)$
  - 当 $\delta \geq 1/2$ 时，1没有动力偏离

冷酷触发 (Grim Trigger)

	C	D
C	2，2	0，3
D	3，0	1，1

- 如果之前有人使用过D，策略组合要求双方转为永远背叛 DD, DD, DD…
  - 给定2的策略，如果1**不偏离**：收益流为1, 1, 1, 1…，总收益1/(1- δ)
  - 给定2的策略，如果1**一次性偏离**：本阶段（用C）收益0，以后每阶段（用D）收益1，收益流为 0, 1, 1, 1…，总收益 δ /(1- δ)
  - 1没有动力偏离
- 综上所述：当 δ ≥ 1/2 时，该策略组合是子博弈完善的纳什均衡

冷酷触发：小结

- 在合作阶段，只要博弈者足够重视未来，未来的无穷多轮的惩罚造成的损失足以抹去当下背叛所带来的一次性收益，博弈者没有动力背叛
- 在惩罚阶段，双方反复使用的是阶段博弈的纳什均衡(D, D)，所以是自我稳定的(self-enforcing)

有缺陷版本的冷酷触发

$$\boxed{P_0\text{：C}} \xrightarrow[\textbf{对方}\text{使用D}]{} \boxed{P_1\text{：D}}$$

- i的策略
  - 第一阶段使用C
  - 之后的阶段，**对方**从未使用过D，继续使用C
  - 如果**对方**使用过D，永远使用D

- 考察其是否构成子博弈完善均衡

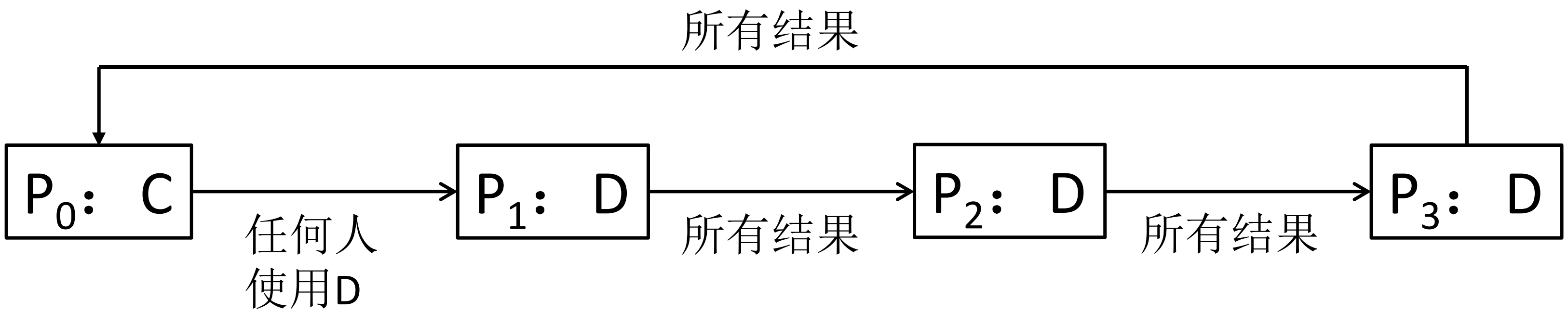
有缺陷版本的冷酷触发

P_0： C $\xrightarrow{\text{对方使用D}}$ P_1： D

- 从1的角度看，信息集分为几类?
  - 四类：a) 双方都未曾背叛；b) 1没有背叛但是2背叛过；c) 1背叛过但是2没有背叛过；d) 双方都背叛过
  - 为什么不是两类（2没有背叛过 vs. 2背叛过）？
    - 同类信息集上，策略组合指定的后续行动必须是同样的
- 容易验证，在a类信息集上，当$\delta \geq 1/2$，1没有动力偏离
  - 请课后自己具体验证
- 也容易验证，在d类信息集上，1没有动力偏离（课后验证）
- 在b类信息集上（1没有背叛而2背叛过），1有没有动力偏离？
  - 课后验证：1有动力偏离（假装这次没看见）
- 在c类信息集上（1背叛过但2没有背叛），1有没有动力偏离？
  - 课后验证：也有动力偏离（宁可提前一个阶段开始D）

有限惩罚 (Limited Punishment)

- 冷酷触发使用了最严厉的处罚：一旦有人使用D，转入**永远**使用D
- 是否可以使用比较温和的惩罚？
- 有限惩罚策略（只惩罚有限个阶段），例如
  - 合作期：一直使用C
  - 如果任何人使用D，转入惩罚期
  - 惩罚期：持续三个阶段，每个阶段使用D，然后转回合作期



能否构成子博弈完善均衡

	C	D
C	2，2	0, 3
D	3，0	1，1

- 合作期 （策略要求CC，并持续合作）
  - 1不偏离：获得 2, 2, 2, 2,…
  - 1一次性偏离：本阶段获得3，接下来三个阶段获得1，其后每个阶段2
  - 保证1不偏离需要 2+2δ+2δ^2+2δ^3 ≥ 3+ δ + δ^2+δ^3 ，即 $\delta^3+\delta^2+\delta-1\leq 0$
  - 可解出（用计算器）约为：$\delta > 0.55$
- 惩罚期（策略要求连续三个阶段双方为DD）
  - 无论是在惩罚期的哪个阶段（第1、2、或3阶段），给定对方执行既定策略
  - 1 如果在本阶段不偏离（用D）的收益为1，偏离（用C）的收益为0
  - 1 没有动力偏离
- 这里，惩罚期因为用的是阶段博弈的纳什均衡，所以是“自我稳定”的(self-enforcing)

- 可以看到，“有限惩罚”可以成为均衡，但是需要的δ值比前面的“冷酷触发”要大（这里三期惩罚需要$\delta > 0.55$）
- 是否还能构建其他的均衡？
- 能。事实上可以构建的均衡**非常多**。下一节课会介绍“无名氏定理”（Folk Theorem），告诉我们有哪些均衡（收益结果）是可以实现的