第八讲：重复博弈（二）

无名氏定理（Folk Theorem）

社群性奖惩（Community Enforcement）

重复博弈中的多种均衡

- 上一讲里，我们看到无限重复博弈里往往存在多个均衡
  - 例如，无限重复的囚徒困境中，“冷酷触发”、“有限惩罚”、以及作业里的“一报还一报”等都可以构成子博弈完善均衡
  - **基本结构**：指定的（合作）策略 + 惩罚策略
- 均衡策略很多，可以导致的收益结果也很多
- 下面，我们不一一讨论均衡的策略，而是聚焦于均衡的**收益结果**：均衡中可能出现哪些收益结果？
  - 本讲中的收益结果都是指博弈者在无穷个阶段里每个阶段的**平均收益**
  - 准备工作：一个重复博弈中，行动者所有可能的策略组合（不考虑是否构成均衡）可产生的收益结果有哪些？

可行收益 (Feasible Payoffs)

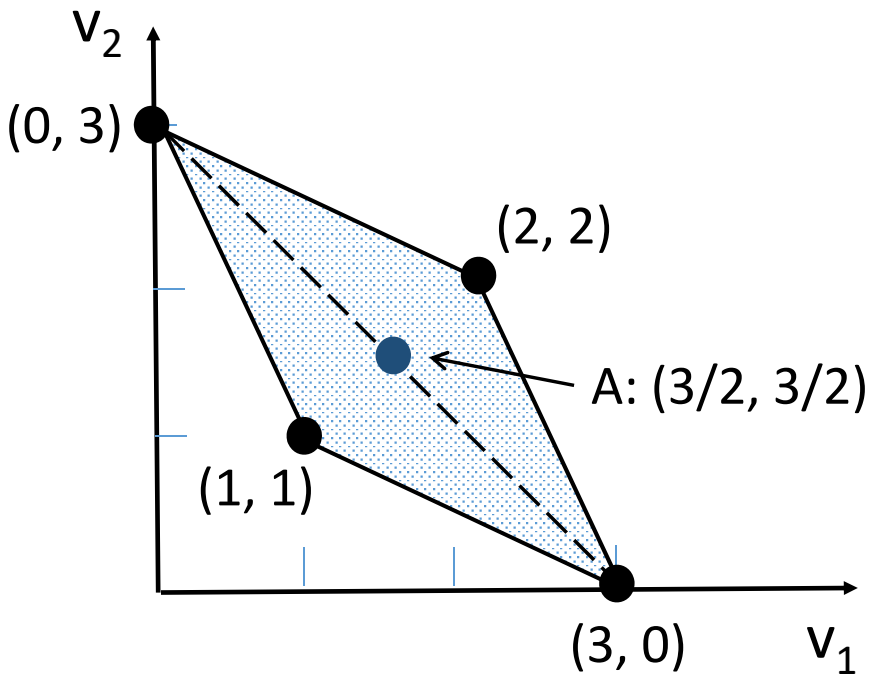
	C	D
C	2, 2	0, 3
D	3, 0	1, 1

- 在一个重复博弈中，任意一个可能的策略组合（**无论**它是否构成均衡）产生的平均收益，称为该博弈的一个“可行收益”
- 可行收益集：包含了该博弈所有可能的收益结果；如何确定这个集合？
- 以无限重复的囚徒困境为例（阶段博弈的收益矩阵如上）
  - 在任何一个阶段，博弈的结果或者是(C, C), (C, D), (D, C), (D, D)中的一种，或者这几种结果按照某一个概率分布随机出现
  - 那么在任何一个阶段，收益一定是(2, 2), (0, 3), (3, 0), (1, 1) 中的某一个，或者这几个结果按照一定概率加权平均
  - 整个重复博弈的（平均）收益，一定是这几个收益结果（或其加权平均）按其出现的比例加权平均，即这几个收益结果的某种加权平均

重复博弈的可行收益集

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

- 一个重复博弈的可行收益集
  - 就等于它的阶段博弈的单纯策略组合的收益结果的所有加权平均值之集
- 即，以单纯策略组合的收益为“端点”构建的“凸包” (Convex Hull)
  - 凸包（定义）：能够把给定的点都包含在内的最小的凸集
  - 凸包中的每一个点，都是其端点值的某种加权平均
  - 例如：A点，(3/2, 3/2) = 1/2*(3,0) + 1/2*(0,3)

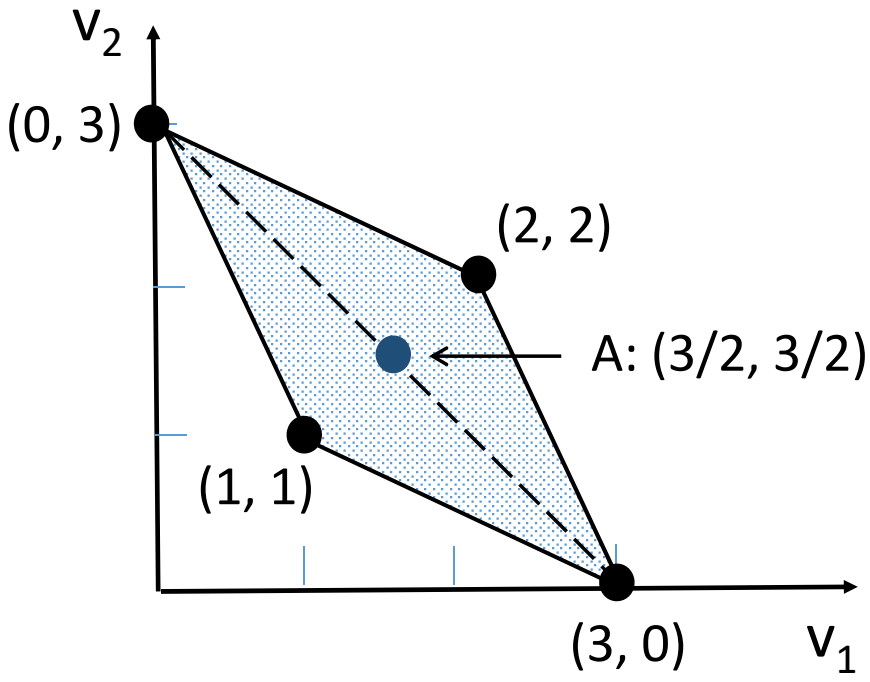


- 这个集合中的任何一个点都是一个可行收益，即博弈中存在至少一个策略组合可以产生该收益结果

产生可行收益的策略组合

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

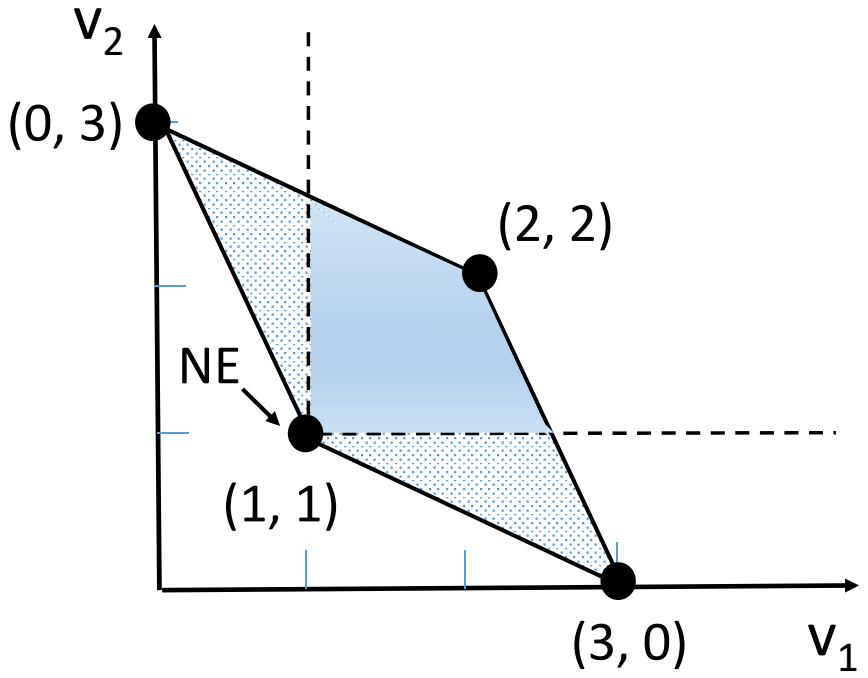
- 例如：什么策略组合能产生A点的收益结果？
  - 需要有一半的结果是(3,0)，一半是(0,3)
- 一种办法：使用“公共随机装置”
  - 想象博弈者共同抛一个硬币，正面朝上使用(D,C)，反面朝上使用(C,D)
  - 每阶段的期望收益：(3/2, 3/2)
- 另一种方法：轮换交替
  - 按照1/2和1/2的比例，博弈者交替使用(D,C)和(C,D)，收益流为(3,0), (0,3), (3,0), (0,3)…
  - 当δ --> 1，平均收益趋近(3/2, 3/2)



Friedman 无名氏定理（Folk Theorem）

- 定理：在一个无限重复博弈中，当δ足够接近1，任何优于纳什收益的可行收益都可以成为该博弈的均衡收益
  - 所谓“纳什收益”是指阶段博弈的纳什均衡中的收益
  - 这里“优于”定义为在一个结果中每个人的收益都**大于**在另一结果中的收益
- 这个定理对于任何无限重复的博弈都成立
- 如何证明？
  - 针对区域内的任何点，构建均衡策略，实现该收益

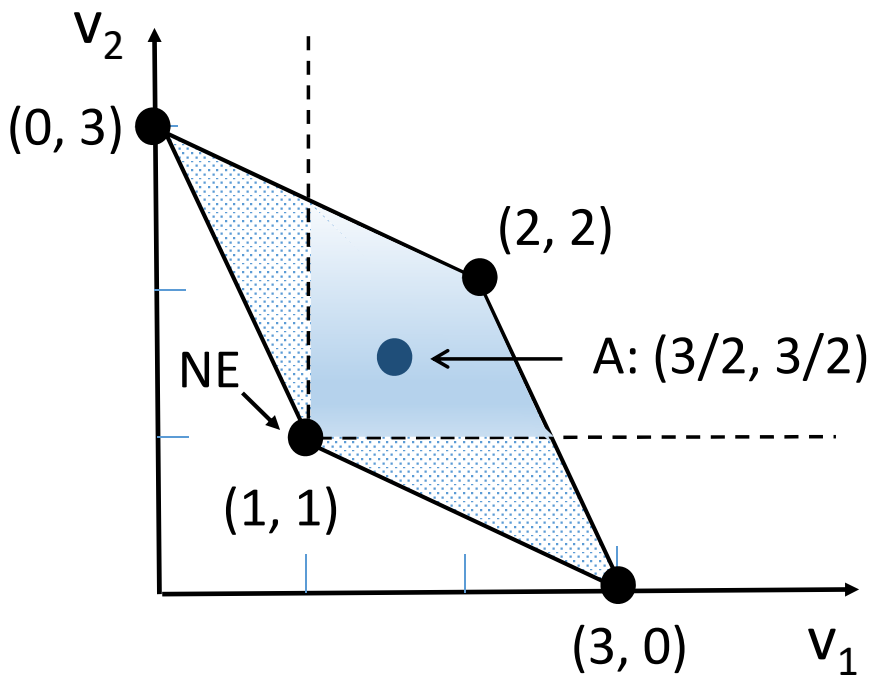
	C	D
C	2, 2	0, 3
D	3, 0	1, 1



Friedman无名氏定理

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

- 以A点为例
- 均衡策略：“合作”策略 + 惩罚策略
  - 在“合作状态”，指定各方按照能够产生目标收益的某种策略行动
    - 例如要产生A点的收益，指定双方在每一轮用1/2, 1/2的概率随机使用DC和CD（或者交替使用这两者）
  - 惩罚策略：如果任何人偏离指定路径，那么转入永远使用阶段博弈的纳什均衡策略，如这里的DD
  - 显然，这里使用的是一种“冷酷触发”型的策略



检验子博弈完善性

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

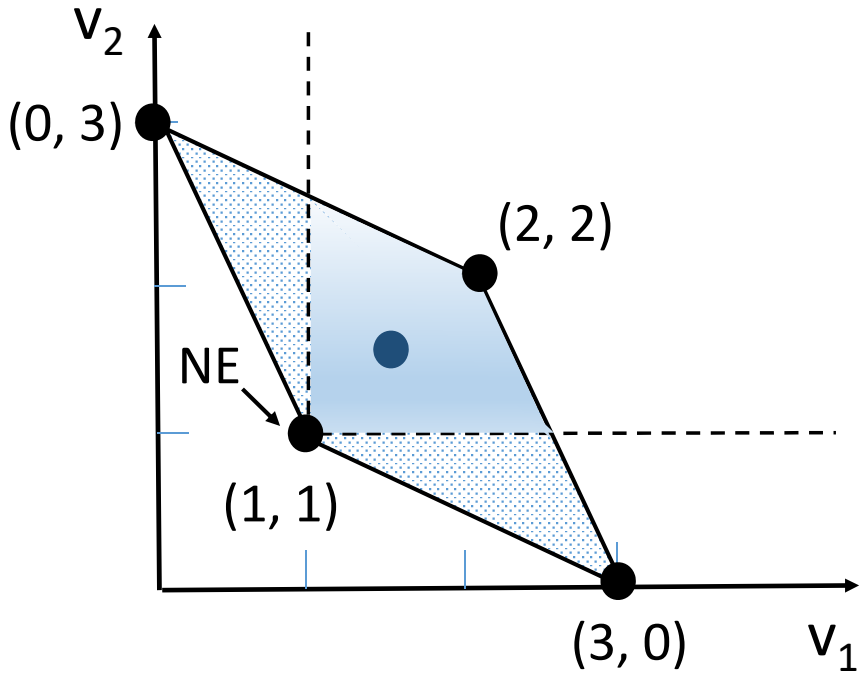
- 以A点(3/2, 3/2)为例
- 在合作期，即尚未发生偏离的历史之后
  - 指定策略是双方按公共随机装置*，以1/2, 1/2的概率随机使用DC和CD
    - 如果1不偏离：收益流是3/2, 3/2, 3/2…，平均收益3/2
    - 如果1一次性偏离（用自己本阶段优势策略D），现阶段期望收益(3+1)/2=2，之后每阶段1，收益流2,1,1…，平均收益 $1 + (1-\delta)1 = 2 - \delta$
    - 不偏离需要3/2 ≥ 2-δ，即 $\delta \geq 1/2$
- 在惩罚期，已经发生过偏离的历史之后
  - 指定策略是双方永远使用DD
  - 给定2的策略是永远使用D，1的最佳应对也是永远使用D
- 所以，以上策略组合是子博弈完善的，只要$\delta \geq 1/2$
- 该逻辑显然可推广到可行收益集中的任意优于纳什收益的点，只要δ足够接近1，任何一次性偏离的有限收获总能能被无限多个阶段的惩罚所抵消

*如果不用公共随机装置，而是要求博弈者交替使用某个单纯策略，证明思路类似

小结

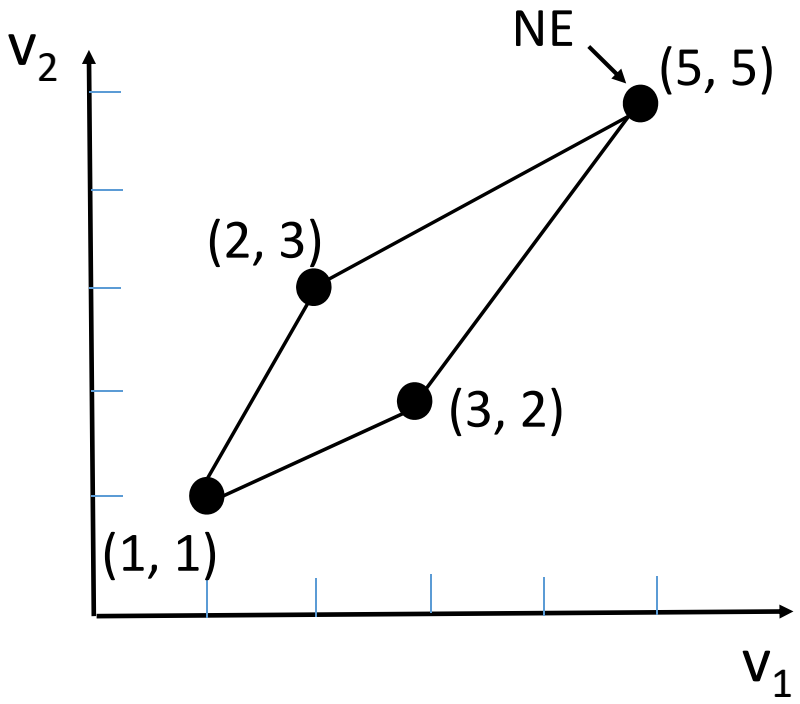
- Friedman无名氏定理以永远重复阶段博弈的纳什均衡作为惩罚
  - 也可以称为“纳什威慑”或冷酷触发型的无名氏定理
- 惩罚策略本身，因为是使用阶段博弈的纳什均衡策略，所以是自我稳定的
- 在合作期，任何优于惩罚期收益的结果都可以稳定（当δ足够大）
  - 因为未来无限个阶段的惩罚造成的损失，足以抵消现阶段一次性越轨行为带来的任何有限收益（当δ足够大）

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



- 不过，纳什收益不一定是一个博弈中可以执行的最严厉的惩罚
- 右边的博弈：纳什均衡的收益为(5, 5)
- 如果靠“纳什威慑”，无限重复中只有(5, 5)可以成为均衡收益
- 是否存在其他的均衡收益？
  - 那么要思考最严厉的惩罚能严厉到什么程度

	C	D
C	5, 5	2, 3
D	3, 2	1, 1



最小最大值 (Minmax Value)

	C	D
C	5, 5	2, 3
D	3, 2	1, 1

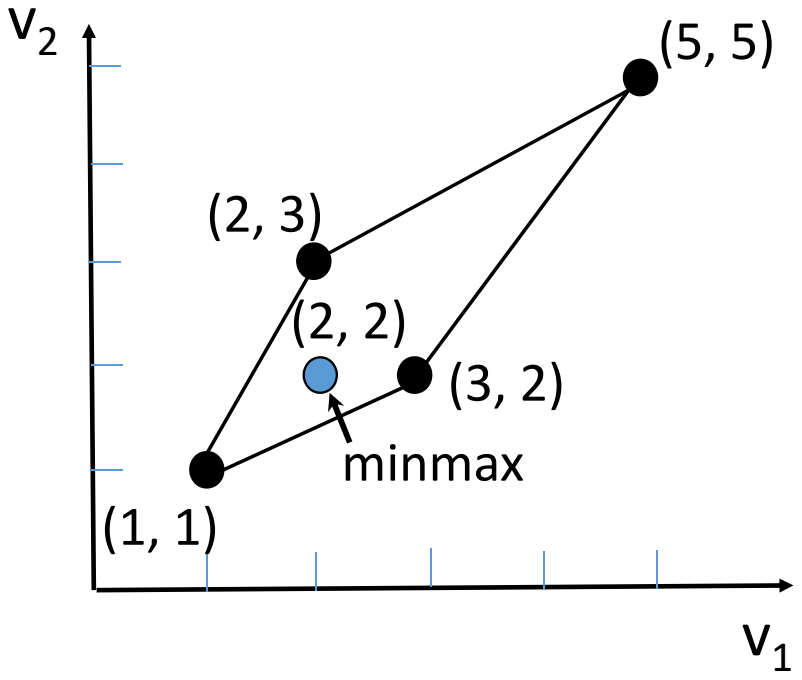
- 最小最大值的定义*
  - 在阶段博弈里，当博弈者i总是选择自己的最优反应的前提下，对方（其他人）能将他的收益最小化到什么程度
  - 也可以说，是当对方使用对i最不利的策略时，i能获得的最大收益
- 在右上的博弈中，对博弈者1而言
  - 如果对方选C，他的最大收益是5，如果对方选D，他的最大收益是2
  - 他的minmax收益：2
  - 我们称博弈者2针对1的minmax策略是D；博弈者1的minmax应对是C
- 对博弈者2而言，minmax收益也是2
- Minmax 收益组合：(2, 2)

*为简单起见，本课中我们只考虑单纯策略下的minmax

最小最大值 (Minmax Value)

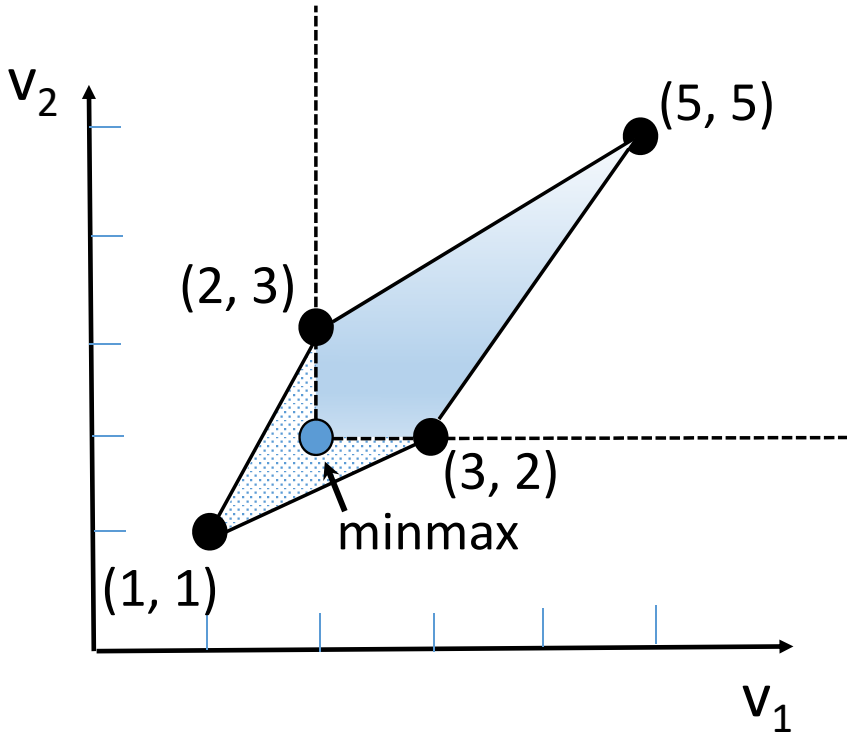
	C	D
C	5, 5	2, 3
D	3, 2	1, 1

- Minmax 收益组合：(2, 2)
- Minmax 收益是重复博弈的任何均衡中一个博弈者平均收益的**下限**
  - 因为博弈者只要每个阶段使用自己的minmax应对，就可以保证获得至少minmax的平均收益



Fudenberg & Maskin 无名氏定理

- 在一个无限重复的两人博弈中*，当δ足够接近1，任何优于minmax 收益的可行收益都能成为均衡收益
  - “优于”仍然是指在一个结果中每个人的收益都大于在另一结果中的收益



*当博弈者人数N>2，该命题仍然成立，但要求可行收益集合满足一个条件——“完全维度”(full-dimensionality)，即其必须有N维，这里不详细介绍

- 均衡策略（两人博弈*）：合作策略 + 惩罚策略
  - 合作状态：按照需要的比例随机（或交替）使用某几个端点策略
  - 如果任何博弈者偏离以上策略，进入惩罚状态（惩罚期）
  - 惩罚期（若期内无人偏离）持续T阶段，结束后回到合作状态
  - 惩罚期内，每一位博弈者采用针对对方的minmax策略
    - 使得对方可能获得的最大收益不超过其minmax收益，此时每人收益小于等于自己的minmax收益
  - 惩罚期时间（T）需要足够长，以抵消在合作期偏离可能带来的最大的一次性额外收益
  - 惩罚期内如果任何人偏离指定策略，惩罚期**重新开始**
    - 惩罚期内的一次性偏离的最高收益是自己的minmax收益，而延长一个阶段的惩罚，损失的是未来一个阶段的合作收益(大于minmax)，当δ足够大，不合算

* 三人以上的博弈，均衡策略要复杂一些（惩罚期后回到一个带奖励性的 “新合作期”，之前执行惩罚的博弈者获得较高的收益），这里不详细介绍，本课中不要求掌握

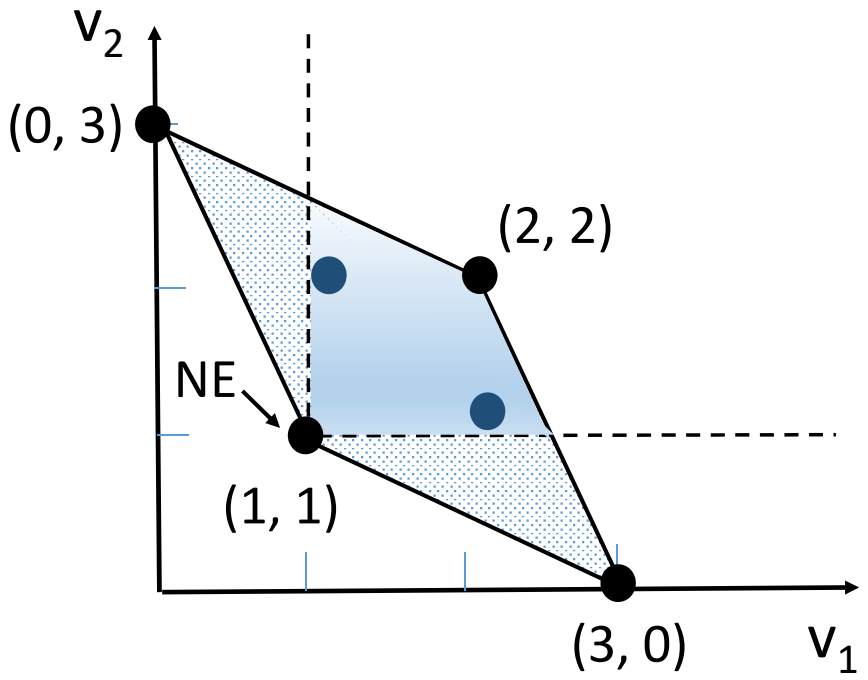
小结：重复博弈和无名氏定理

- 无名氏定理显示，在无限重复博弈中，存在非常多的均衡结果
  - 只要δ足够大，相当大范围内的可行收益都可以在均衡中实现
- “Almost anything goes”

Almost Anything Goes

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

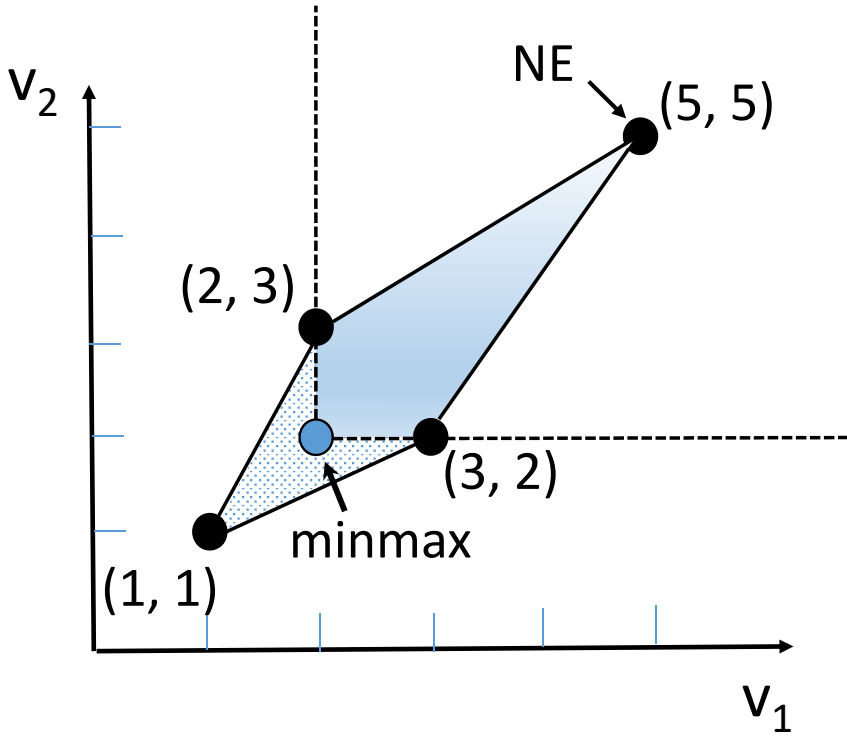
- 不仅合作型结果是可能的均衡
  - 例如（2，2）
- 不合作的结果也是可能的均衡
  - 例如（1，1）
- 不平等的、“剥削型”的结果也是可能的均衡
  - 例如（1.1，2）
  - 反过来（2, 1.1）也是可能的



Almost Anything Goes

- 比阶段博弈中的纳什均衡更差的结果也能成为均衡结果

	C	D
C	5, 5	2, 3
D	3, 2	1, 1



小结：重复博弈和无名氏定理

- 无名氏定理显示，在无限重复博弈中，存在非常多的均衡结果
  - 只要δ足够大，相当大范围内的可行收益都可以在均衡中实现
- “Almost anything goes”
- 这有助于解释现实中的长期关系和群体生活的多样性
  - 并非所有的长期关系和群体生活都是合作、平等、和谐的；不合作、不平等和压迫也可能长期存在

如果博弈重复，但是对象不断改变

- 在较大的人群或社会中，有时候两个个体在互动之后再次相遇的概率很小，但是会和社群中的其他个体继续展开类似博弈
  - 例如：滴滴打车、网购平台（eBay, 淘宝）上的卖家和买家、等等
  - 博弈在重复，但是互动对象不断改变
- 如果不通过其他强制机关（例如法院），个体间合作还能形成吗？
- 能；可借助某种社群性奖惩（community enforcement）和声誉机制（reputational mechanism）

随机相遇模型 (Radom Match)

	C	D
C	1, 1	-1, 2
D	2, -1	0, 0

- 在一个规模较大（N）的人群中，每一个时期人群中的个体都随机地配对并进行两人博弈，例如右上方的囚徒困境
  - 每个阶段博弈结束后，下一阶段重新进行随机配对
- 这里设定：每个人的行动都是公共信息（perfect public monitoring）
- 如何构建支持合作的均衡策略？
- 考虑一种“极冷酷触发”策略 (really grim trigger)
  - 每个人第一轮都使用C
  - 如果之前没有任何人使用过D，每个人继续使用C
  - 如果之前有任何人使用过D，每个人从本轮起都永远使用D
  - 也即是说，只要有一个人背叛，所有人都转入永远背叛
- 这能否构成子博弈完善均衡？

	C	D
C	1, 1	-1, 2
D	2, -1	0, 0

- 只有两种可能的历史：没有人使用过D，曾有人用过D
- 从博弈者i的角度看
- 如果从没有人使用过D，既定策略要求此次i与j都用C，以后的轮次中 i 与其他对手相遇也都用C
  - i 如果不偏离，本轮收益为1，之后每轮收益也是1，收益流是1, 1, 1, …，折现值为 $1/(1-\delta)$
  - i 如果一次性偏离，本轮收益为2，但之后每一轮收益只有0，收益流为2, 0, 0…，折现值为 2
  - 当$\delta \geq 1/2$时，i 没有动力偏离
- 如果曾有人使用过D，策略要求本轮i与j都用D，以后的轮次中i与任何其他对手相遇也都用D
  - 给定他人用D，i的最优反应是D，i显然没有动力偏离
- 可见，这种“极冷酷触发”是子博弈完善的

- 但是以上的策略意味着，一旦发生了背叛，不仅背叛者受到惩罚，所有人都会受到惩罚（都转入永远用D）
- 能不能只惩罚“有罪者”呢？
- 设想这样规定：在合作期，如果某个成员j使用了D，那么之后所有j参加的互动中，双方都用D（永远惩罚）；但其他历史清白的成员互动时，双方继续用C
  - 就好像每一个人都有一个声誉的标签或记录，一开始是“好”，如果没有犯错，继续保持“好”，一旦犯错，则**永远**改为“差”
  - 每一次相遇中，如果双方的声誉都是好的，那么合作；只要有有一方的声誉是差的，则双方不合作
- 只要δ足够大，这显然可以构成纳什均衡（请课下自己验证）
- 但能否构成子博弈完善均衡?

- 子博弈完善要求考虑每一种历史后博弈者是否有动力偏离
- 考虑当历史是绝大多数成员都曾经背叛过，i（自己没有背叛过）碰上另一个还没有背叛过的（假定是i以外唯一的一个）成员k时
  - i如果用C，可确保之后再遇到k时还可继续合作，但每一次配对再碰到k的概率很小，仅1/(N-1)，而之后碰上任何其他人，按既定策略都是双方用D收益为0，所以i的收益流是1, $\frac{1}{N-1}$, $\frac{1}{N-1}$, … , 折为现值是1+ $\frac{\delta}{1-\delta} \cdot \frac{1}{N-1}$
  - i如果改为用D，本次可以获得收益2，之后每一轮收益为0，收益现值为2
  - 我们注意到他偏离的损失仅仅是以后再碰到k时无法合作，而以后碰上任何其他人，由于本来就无法和他们合作，并无额外的损失
  - 如果N足够大，i偏离的一次性额外收益将会大于他之后无法与k合作的损失，即 $1 > \frac{\delta}{(1-\delta)(N-1)}$
- 可见，对于给定的折扣系数δ，只要N足够大，以上策略**无法**构成子博弈完善均衡

- 考虑把惩罚变得温和一些：允许在有限个惩罚阶段后，原谅犯过错误的博弈者，恢复他的声誉
- 考察以下策略
  - 合作期：一开始每一个配对中双方用C，若无人偏离，下一轮大家继续用C
  - 在合作期如果有博弈者j偏离，转入对于j的惩罚期
  - 惩罚阶段，在j参与的配对中，双方采取针对对方的minmax策略(D)
  - 惩罚阶段（期内若无人偏离）持续T个阶段，之后回到合作期（受罚者声誉恢复）
  - 惩罚阶段，如果受罚者j偏离，惩罚阶段重新开始；如果施罚的一方k偏离，则下一轮开始针对k的T阶段惩罚，针对j的惩罚自动结束
  - 在惩罚时期，如果在其他的（本该合作的）配对中发生了某一方的偏离行为，那么转而开始针对新的偏离者的T阶段惩罚，原来的惩罚自动结束（受罚者声誉恢复）
  - 任何时期，如果同时发生多起偏离行为，自动忽略
- 这样，任何一个时期，在整个社群中最多只有一个人需要被惩罚，或者说最多只有一个人的声誉是“差”的

- 可以构成子博弈完善均衡
- 合作期（没有任何人需要被惩罚）
  - 谁偏离，就会引发针对他的惩罚；惩罚期内每个阶段他的损失是合作收益(1)，只要惩罚期T足够长，就足以抵消任何一次性偏离的额外收益
- 惩罚期（有一个j需要被惩罚）
  - 受罚者j：偏离会导致惩罚期重新开始计时；由于偏离的最高收益是自己的minmax，而推迟一阶段回到合作的损失是合作收益(大于minmax)，当δ足够大，不合算
  - 施罚者k：不偏离只需本轮执行一次惩罚，之后T-1轮每次再碰到j（需要执行惩罚）可能性只是1/(N-1)；偏离则导致自己被连续惩罚T次，显然不合算
  - 其他人：与合作期的基本逻辑类似，不偏离可以持续获得合作收益，偏离可获得一次性额外收益但会受到T个阶段惩罚，损失T个阶段的合作收益
    - 这里与合作期的一个小差别是，由于存在一个声誉差的博弈者j，我在未来的最多T-1次互动中每次有1/(N-1)的概率遇到j需要用D，所以我不偏离的收益略有下降，在T-1个阶段中每阶段收益是 1·(N-2)/(N-1)（而不是 1），但由于N较大，这个差别很小，不会造成实质改变

小结：社群奖惩和声誉机制

- 以上策略里，使用了社群奖惩（community enforcement）机制
  - 一个博弈者j背叛了i，会在未来受到i以外的其他的社群成员的惩罚
- 其中的信息和协调机制，可以说是一种声誉机制（reputational mechanism)
  - 任何偏离（越轨）行为都会导致越轨者的声誉变坏，受到社群的惩罚
- 以上策略可用于证明随机相遇模型版本的“无名氏定理”
  - 即只要δ足够大，任何优于阶段博弈的minmax收益的收益结果，都可以成为整个博弈的均衡中的平均收益结果

- 但是以上策略中，任何时期只有一个声誉差和接受惩罚的人，似乎和经验现实有较大差距
- 也可以修改上述模型，允许同一时期有多个声誉差和接受惩罚的个体
- 这就需要确保实施惩罚的个体有足够的动力保持和其他声誉好的个体合作，尤其是在大多数人都是声誉差的时候
  - 基本思路是，当好人惩罚坏人时，坏人采用某种“悔过策略”，使得好人的收益较高，坏人的收益较低
- 以下仅供了解

考虑如下策略（仅供了解，不需要掌握）

- 每个人一开始声誉都是“好”的，一旦违背如下规定，则声誉变“坏”，需接受T阶段惩罚，之后声誉恢复为好（受罚期间如背离，则重新开始T阶段的计时）；惩罚期内规定
  - 好人遇到好人：双方合作
  - 坏人遇到坏人：双方使用针对对方的minmax 策略
  - 好人遇到坏人：好人使用针对坏人的minmax 策略，而坏人必须使用某种“悔过”策略 r，此时好人的收益比自己的minmax 收益高，而坏人的收益比自己的minmax 收益低
    - 例如在PD game 中，好人用D，坏人用C，即可
    - 这就要求阶段博弈中存在这样的策略 r；并非所有的博弈都满足该要求
- 当δ足够大，可以选择合适的T，保证即使大多数人都是坏人的历史后，好人仍然会执行既定策略
  - 因为继续做好人（执行惩罚）比起做坏人（被惩罚）来说，收益更高

作业四、期中考试

- 下周四（4月16日）期中考试，请提前10分钟左右到课堂准备
- 作业四已经在教学网发布，下周四带到课堂提交
- 稍后会公布作业四的参考答案，供大家考前参考