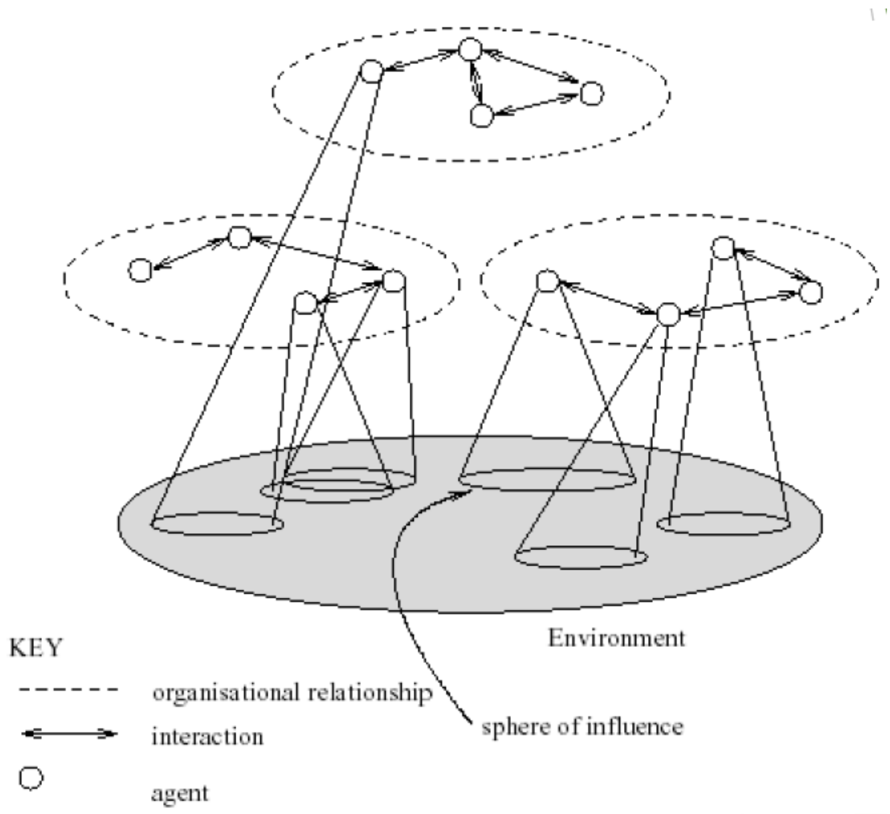
CPT302 W6 Multiagent Interactions



一个多代理系统包含许多代理:

- 它们通过交流 interact
- 可以在环境中执行活动
- 有不同的“势力范围 (spheres of influence)”（这些区域可能重合 (coincide)）

  注：spheres of influence 即这些代理的感知范围，上图环境中实线的圈
- 通过其他 (organizational) relationships 联系起来

  注：上图虚线的圈构成一个组织，代理在其中交互

Utilities and preferences

Self-interested agents: 每个 agent 对世界状态都有自己的preferences and desires (偏好和心愿)（non-cooperative game theory, 非合作博弈论）

Modelling preferences

Outcomes (states of the world):

$$\Omega = \{W_1, W_2, \ldots\}$$

Utility function:

$$u_i : \Omega \rightarrow R \text{ (real numbers)}$$

Preference Ordering

使用 utility function 导致 outcomes 存在优先顺序：

- Preference over w

$$u_i(w) \geq u_i\left(w^{\prime}\right) \Leftrightarrow w \succeq w^{\prime}$$

- Strict preference over w

$$u_i(w) > u_i\left(w^{\prime}\right) \Leftrightarrow w \succ w^{\prime}$$

utility function 将世界状态映射为数字，数字越大，说明 agent 越偏向该世界状态。

注：上面的 $\succ$ 与之前 behavior 之前的抑制完全不同；之前 b1 $\prec$ b2 意为 b1 抑制 b2；现在 $w \succ w^{\prime}$ 说明 w 更受偏向。

Multiagent Encounter

现在我把 interaction 看作一场游戏：世界的状态可以被看作是一场游戏的结果

- 假设我们只有两个 agent（玩家）Ag={i, j}
- 最终结果 Ω 取决于每个代理选择的行动的组合
- State transformer function：

$$\tau : \underbrace{Ac}_{\text{agent i's action}} \times \underbrace{Ac}_{\text{agent } j^{\prime} \text{ s action}} \rightarrow \Omega$$

Normal-form game (or strategic-form game)

游戏中的交互通常表示为元组（N, A, u），其中：

- N 是（有限的）玩家的集合
- $A = A_1 \times A_2 \times \ldots \times A_n$ ，其中 A_i 代表玩家 i 所能使用的一系列行动
- $U = (u_1, u_2, \ldots, u_n)$ 是 每个玩家的 utility function

Payoff matrix

	i: C	*i*: D
j: C	1, 1	1, 4
j: D	4,4	4, 1

上表代表有两个玩家 i 和 j，它们分别有两个可采取的行动 (或策略) C 和 D。每个策略被 utility function 转换为数字显示在表格中。

现在我们面临一些问题：作为一个 rational agent (理性代理)，我们希望最大化我们的 expected payoff (single-agent point of view)。然而，这在大多数情况下是不可行的，因为个人最佳策略取决于他人的选择 (multi-agent point of view)

Solution Concepts

Best response: 给定玩家 j 的策略 s_j，玩家 i 对 s_j 的最佳对策是使玩家 i 收益最高的策略 s_i 。

	i: C	i: D
j: C	1, 1	1, 4
j: D	4,4	4, 1

如上图所示，如果 j 采取 C，那么 i 的 best response 是 D (因为 i 的收益为 4)；如果 j 采取 D，那么 i 的 best response 是 C。

Dominant Strategy

主导策略就是：对于玩家 i 来说，不管另一个玩家 j 选择什么策略，i 的策略都可以至少和策略 s_i^* 一样好。那么 s_i^* 就是 dominant strategy。

如果 s_i^* 是对玩家 j 所有策略的最佳对策，则 s_i^* 是主导的。

	i: C	i: D
~~j: C~~	~~1, 4~~	~~1, 1~~
j: D	4,1	4, 4

如上图，玩家 j 有一个 dominant strategy D，j 只要选 D 就可以收益最大化；另一个策略 C (dominated strategy) 可以从表格中移除。玩家 i 没有主导策略。

Pareto Optimality

帕累托最优性（或帕累托效率）：如果没有其他结果使一个代理变得更好而不使另一个代理变得更糟，则这个结果被称为 Pareto optimal（或 Pareto efficient）。

Agent 1 \ Agent 2	C	D
C	3, 3	0, 5
D	5, 0	1, 1

上图中 (C,C), (D,C) 和 (C,D) 达到了 pareto optimal。

例如 (C,C)，它的右边和下面，都会导致一个值变大和另一个值变小；而右下角则是两个值都变小，这是 pareto optimal。又比如 (D,C)，它的上面，右面，和右上方，都会导致一个变小，另一个变大。

而 (D,D)，它的左上有另一个可以使得值变大，但不使另一个值变小的结果，因此它不是 pareto optimal。

Nash Equilibrium

Nash equilibrium for pure strategies

两个策略 s1 和 s2 处于纳什均衡，如果：

1. 代理 i 采取 s1，而代理 j 不能得到比采取 s2 更好的结果
2. 并且，代理 j 采取 s2，而代理 i 不能得到比采取 s1 更好的结果

因此这两个主体都没有任何偏离纳什均衡的动机，纳什均衡代表了 self-interested agents 所玩游戏的“理性”结果。

不幸的是，并非每个交互场景都有纳什均衡，而一些交互情景具有多个纳什均衡。

夫妻俩想一起看电影，妻子更喜欢 FilmA，丈夫更喜欢 FilmB。

	husband: FilmA	*husband:* FilmB
wife: FilmA	2, 1	0, 0
wife: FilmB	0, 0	1, 2

上图中 (A, A) 和 (B, B) 处于纳什均衡。

在 payoff matrix 中找到纯策略纳什均衡的简单方法：取第一个数字是这一列的最大值的单元格，然后检查第二个数字是否是这一行的最大值。

Social Welfare

Outcome ω 的 social welfare 是每个代理从 ω 获得的 utilities 之和：

$$\sum_{i \in Ag} u_i(\omega)$$

可以把它想象成"系统中的货币总量"。

Social welfare 将所有 agent 看作整体来考虑得失。

Example

The Prisoner’s Dilemma

囚徒困境：两个人被共同指控犯罪并被关押在单独的牢房中，无法见面或交流。他们被告知：

- 如果一个人招供而另一个人不招供 (confess)，供认者将被释放，另一个将被监禁十年；
- 如果两人都招供，那么每个人都将被监禁五年；
- 两个囚犯都知道，如果他们都不招供，那么他们每个人都将被监禁一年。

	Player 2 confesses	Player 2 does not confess
Player 1 confesses	(5,5)	(0,10)
Player 1 does not confess	(10,0)	(1,1)

注：上面的数字是刑期，所以是越小收益越大。

	Player 2 confesses	Player 2 does not confess
Player 1 confesses	(5,5) ***NASH***	(0,10) ***PARETO***
Player 1 does not confess	(10,0) ***PARETO***	(1,1) ***PARETO***

注：这里存在 dominant strategy。招供的获刑区间在 0-5，而不招供的区间在 1-10，显然招供更好。因此 dominant strategy 是招供。