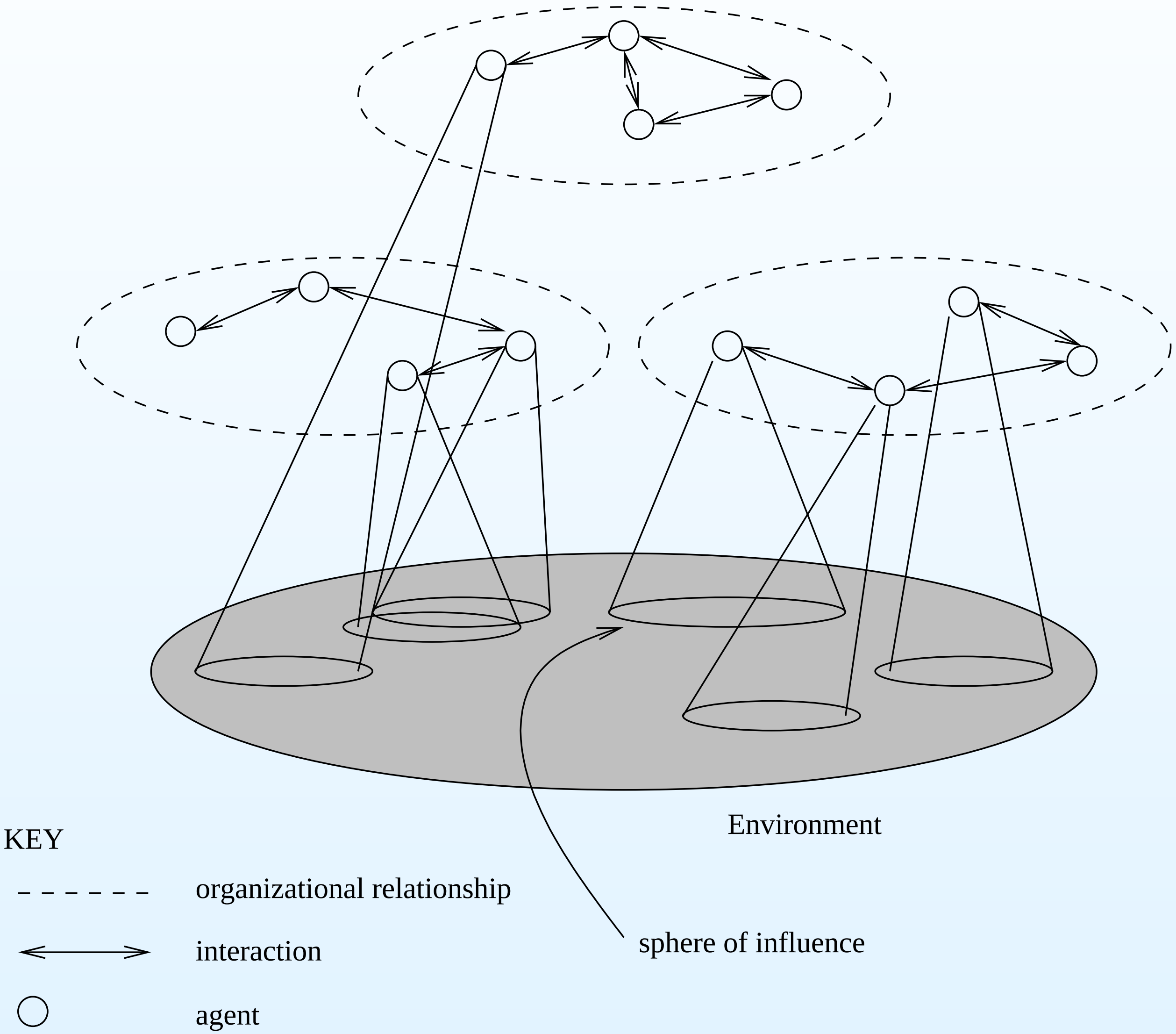
講義 *6,7:* マルチエージェント相互作用

Multiagent Interactions

内容

1. マルチエージェント システムとは?(1/2)



1. マルチエージェント システムとは?(2/2)

マルチエージェント システムは以下の様な多数のエージェント を含んでいる.

- 通信により相互作用する
- 環境において行動できる
- 異なる影響力の範囲をもつ
- 組織的関係により関連つけられる

2. 効用と選好 (1/2)

- 2 つのエージェントがいることを仮定する: $Ag = \{i, j\}$
- エージェントは利己的であると仮定する．つまり，環境がどのようになっているかに関する選好を持つ
- $\Omega = \{\omega_1, \omega_2, \ldots\}$ はエージェントが選好をもっている結果の集合であるとする
- 選好を効用関数によってとらえる

$$u_i : \Omega \to \mathbb{R}$$
$$u_j : \Omega \to \mathbb{R}$$

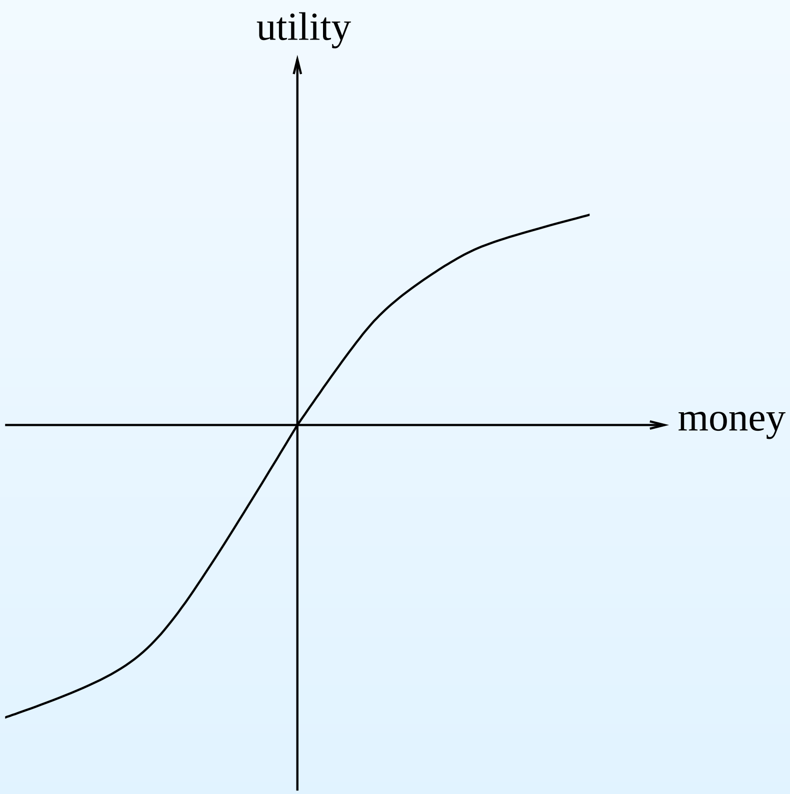
- 効用関数により結果の選好順位が得られる

$$\omega \succeq_i \omega' \text{ means } u_i(\omega) \geq u_i(\omega')$$
$$\omega \succeq_j \omega' \text{ means } u_j(\omega) \geq u_j(\omega')$$

2. 効用と選好 (2/2)

効用とは何か?

- 効用は金銭ではない (しかし，有益なアナロジになっている)
- 効用と金銭の典型的な関係



3. マルチエージェント遭遇 (1/2)

- これらのエージェントが行為を行う環境のモデルの必要性
  - エージェントは実行する行為を斉時的に選択．そして選択した行為の結果として，ある結果 Ω が生じる
  - 実際の結果は，行為の組合せに依存する
  - すべてのエージェントは $''C''$ (“cooperate”) か $''D''$(“defect”) の二つの行為だけが可能と仮定する
- 環境の挙動は以下の状態変換関数により与えられる

$$\tau : \underbrace{Ac}_{\textbf{agent i's action}} \times \underbrace{Ac}_{\textbf{agent j's action}} \to \Omega.$$

3. マルチエージェント 遭遇 (2/2)

- 以下は状態変換関数の一例
$\tau(D,D)=\omega_1, \tau(D,C)=\omega_2, \tau(C,D)=\omega_3, \tau(C,C)=\omega_4$
(この環境は両エージェント の行為に敏感)

- 他の状態変換関数の例は以下
$\tau(D,D)=\omega_1, \tau(D,C)=\omega_1, \tau(C,D)=\omega_1, \tau(C,C)=\omega_1$
(いずれのエージェント も環境に影響力がない)

- 他の状態変換関数の例は以下
$\tau(D,D)=\omega_1, \tau(D,C)=\omega_2, \tau(C,D)=\omega_1, \tau(C,C)=\omega_2$
(この環境はエージェント j により制御されている)

3.1 合理的行為

- 両エージェントが環境に影響力をもつケースを考える．以下が効用関数
  $u_i(\omega_1) = 1,\ u_i(\omega_2) = 1,\ u_i(\omega_3) = 4,\ u_i(\omega_4) = 4,$
  $u_j(\omega_1) = 1,\ u_j(\omega_2) = 4,\ u_j(\omega_3) = 1,\ u_j(\omega_4) = 4$

- 以下の記法を用いる
  $u_i(D,D) = 1,\ u_i(D,C) = 1,\ u_i(C,D) = 4,\ u_i(C,C) = 4,$
  $u_j(D,D) = 1,\ u_j(D,C) = 4,\ u_j(C,D) = 1,\ u_j(C,C) = 4$

- エージェント i の選好は以下のようになる
  $C,C \succeq_i C,D \succ_i D,C \succeq_i D,D$

- エージェント i にとって $''C''$ は合理的選択となる
  (なぜなら，i にとって C によるすべての結果は，D によるすべての結果より好ましいから)

3.1 利得行列

- 前のシナリオは利得行列で特徴つけられる

	i 裏切り	i 協調
j 裏切り	1 1	4 1
j 協調	1 4	4 4

- エージェント i は列プレーヤ
- エージェント j は行プレーヤ

3.2 支配戦略

- エージェント i の任意の戦略 s(C か D) が与えられたとき，可能な結果はたくさんある
- i が s_i を選んだことによるすべての結果が s_2 を選んだ場合の結果より好ましいとき，s_1 は s_2 を支配するという
- 合理的エージェントは決して支配された戦略を選ばない
- よって，何をすべきか決めるとき，支配された戦略を除去してよい
- 支配されていない戦略が常に唯一あるとは限らない

3.3 ナッシュ均衡

- 一般に，以下が成り立つとき，2 二つの戦略 s_1 と s_2 はナッシュ均衡にあるという
  1. エージェント i が戦略 s_1 をとるときに，エージェント j にとっては s_2 より良い戦略がない
  2. エージェント j が戦略 s_2 をとるときに，エージェント i にとっては s_1 より良い戦略がない
- いずれのエージェントもナッシュ均衡から離れる動機 (incentive) を持ち得ない
- ただし，残念なことに
  1. すべての相互作用シナリオにナッシュ均衡があるわけではない
  2. ある相互作用シナリオは 1 つより多くのナッシュ均衡を持つ

3.4 非協力零和ゲーム

- 両プレーヤの選好が対立する場合，そのシナリオは狭義に非協力的であるという
- 零和遭遇では，効用の和が零になる
  $u_i(\omega) + u_j(\omega) = 0 \; for \; all \; \omega \in \Omega$
- 零和は狭義に非協力的であることを意味する
- 日常生活においては零和遭遇はまれである．しかし，人々は多くのシナリオで，あたかもそれが零和シナリオであるかのように行動する．

4. 囚人のジレンマ (1/4)

二人の男が共同である罪の嫌疑をかけられており，別々の独房に入れられている．互いに通信するいかなる手段も，また，合意を得るいかなる方法もない．その二人の男は次のことを言われる．

1. もしも二人のうち一人が罪を自白し，そして他方が黙秘していたら，自白した者は自由の身として，他方は四年の刑となるだろう．
2. もしも両者とも自白したら，それぞれ二年の刑となるだろう．

両方の囚人は，もしもどちらも自白しなかったら，それぞれ一年の刑になるだろうということを知っている．

4. 囚人のジレンマ (2/4)

- 囚人のジレンマの利得行列は以下

	i 裏切り	i 協調
j 裏切り	2 2	0 4
j 協調	4 0	3 3

- 左上: 両方とも裏切れば，相互の裏切りにより両方とも罰せられる
- 右上: i が協調して j が裏切れは，i はお人良しの利得 0 を得て，j は i を出し抜いたことで 4 を得る
- 左して，j が協調して i が裏切れは，j はお人良しの利得 0 を得て，i は j を出し抜いたことで 4 を得る
- 右下: 相互協調に対する報酬

4. 囚人のジレンマ (3/4)

- 個合理的な行為は裏切りである．これにより 2 以上の利得が保証される．一方，協調した場合に保証される利得は 0 である
- よって，すべての可能な戦略の中で，裏切りが最良の戦略である．よって，両方のエージェントは裏切り，利得 2 を得る
- しかし，直観的にはこれは最良の結果ではない．
  両方とも協調したら，利得 3 が得られる．

4. 囚人のジレンマ (4/4)

- このパラドックスは，マルチエージェント 相互作用の基本的な問題である．
  これは，利己的なエージェント の社会では協調は起きないをことを意味している
- 実世界での例
  - 核兵器削減 (「自分だけとっておこう...」)
  - ただ乗り (公共交通機関)
- 囚人のジレンマは普遍的
- 協調を回復することは可能か?

4.1 繰り返し囚人のジレンマ

- 協調を回復するための方法: そのゲームを 1 回より多く行う もしも敵と再会することがわかっていれば，裏切ることへの動機付けは消えそうになる

- 無限回繰り返す囚人のジレンマゲームでは，協調は合理的選択となる

4.2 後向き帰納推論

- 両エージェントがゲームを厳密に n 回行うことを知っていると仮定する．
  n ラウンドでは，どちらもよりよい利得を得るために裏切ることに動機を付けを持つ．
  しかし，これにより $n-1$ 回目が本当の最後のラウンドとなり，ここでもまた裏切ることに動機を付けを持つ．
  これが後向き帰納推論の問題である．

- 囚人のジレンマを，事前に決められた固定有限回数繰り返す場合で，それが双方に共通にわかっている場合，裏切りが最良の戦略となる．

4.3 Axelrod のトーナメント

- 繰り返し囚人のジレンマをいろいろな敵と行うことを考える．自分の全体の利得を最大化するためには，どの戦略を選ぶべきか?

- Axelrod は，この問題を調べるためにプログラムで対戦するコンピュータトーナメントを開催した [Axelrod84]

4.4 Axelrodトーナメントでの戦略

ALLD ：常に裏切る「鷹派戦略」

TIT-FOR-TAT(しっぺ返し) ：

1. ラウンド $u = 0$ では協調
2. ラウンド $u > 0$ では $u-1$ ラウンドの敵の戦略を真似る

TESTER(試し屋) ：最初のラウンドは裏切る．敵が報復して来たらTIT-FOR-TATを演じる．さもなければ，協調と裏切りを適当にばらばらに行う

JOSS ：周期的に裏切ることを除けばTIT-FOR-TATと同じ

4.5 Axelrodトーナメントでの成功の秘訣

Axelrodはトーナメントで成功する戦略として以下の規則を示唆した．

- ねたむな!
  零和ゲームのようにプレイするな
- 上品であれ!
  協調で開始し，協調を繰り返せ
- 適切に報復せよ!
  裏切りは即座に罰せよ．しかしやりすぎるな．
- 心を広くもて!
  相手が協調に転じたら，すぐに協調を繰り返せ

5. チキンゲーム

- 他のタイプの遭遇を考える (臆病者ゲーム)

	i 裏切り	i 協調
j 裏切り	1 (i), 1 (j)	2 (i), 4 (j)
j 協調	4 (i), 2 (j)	3 (i), 3 (j)

(James Dean の映画「理由無き反抗」より)

- 囚人のジレンマとの違い: 相互裏切りは最も恐れるべき結果.
- 戦略 (C, D) と (D, C) がナッシュ均衡

6. 他の対称 2×2 ゲーム

Scenario	Preferences over outcomes	Comment
1.	$C, C \succ_i C, D \succ_i D, C \succ_i D, D$	cooperation dominates
2.	$C, C \succ_i C, D \succ_i D, D \succ_i D, C$	cooperation dominates
3.	$C, C \succ_i D, C \succ_i D, D \succ_i C, D$	stag hunt
4.	$D, C \succ_i C, C \succ_i C, D \succ_i D, D$	game of chicken
5.	$D, C \succ_i C, C \succ_i D, D \succ_i C, D$	prisoner's dilemma
6.	$D, D \succ_i D, C \succ_i C, C \succ_i C, D$	defection dominates
7.	$D, D \succ_i D, C \succ_i C, D \succ_i C, C$	defection dominates

参考文献

[Axelrod84 Axelrod, R. (1984) *The Evolution of Cooperation*. Basic Books, New York.