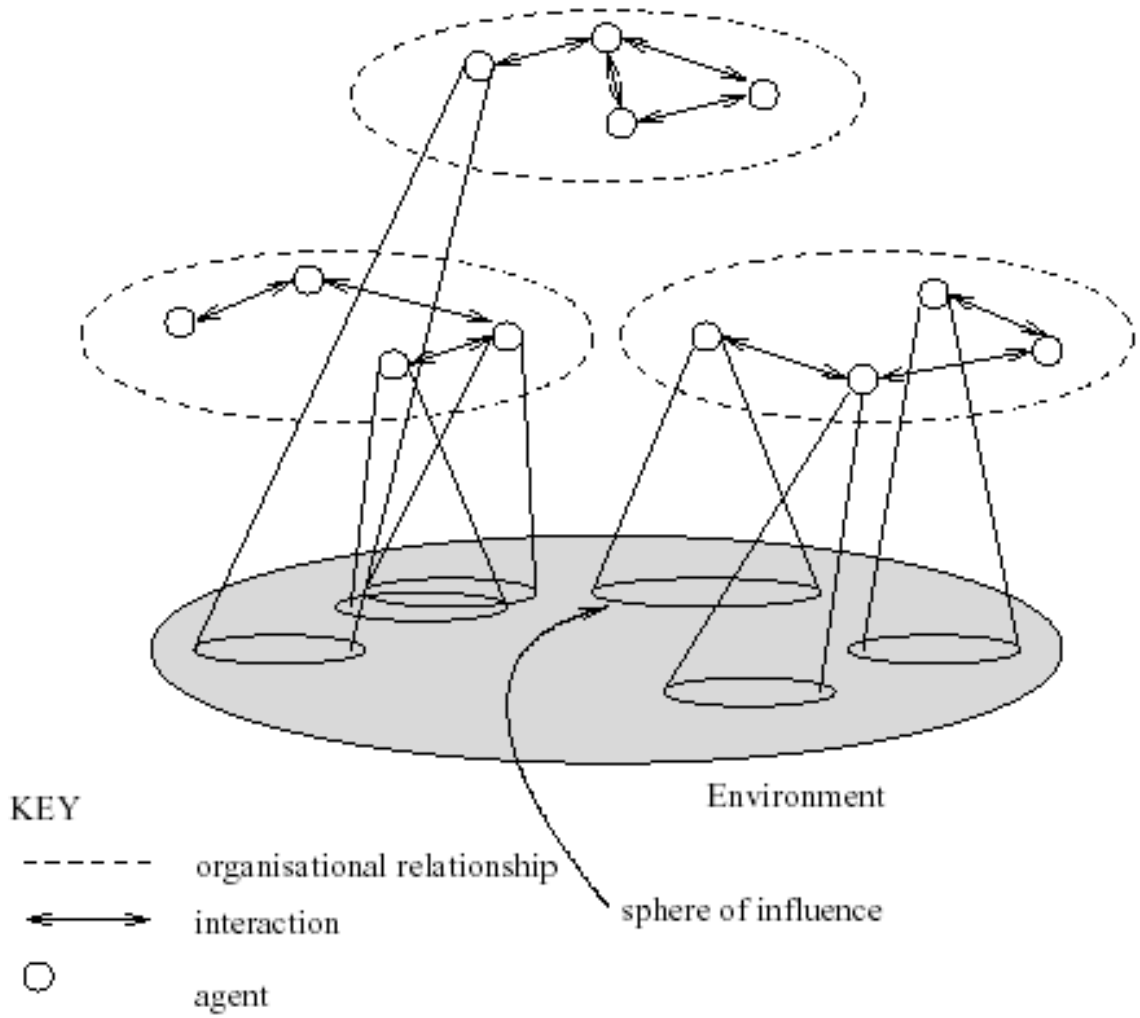
Lecture 7: Game theory

Simon Powers

SET10111 – Multi-Agent Systems

Agent societies



Utilities and Preferences

- Assume two agents
- Set of Environment states: $\Omega = \{\omega_1, \omega_2, \ldots\}$
- Utility functions of Agent i and Agent j:

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

- Preference ordering: state ω is weakly preferred by Ag. i over ω‘ iff

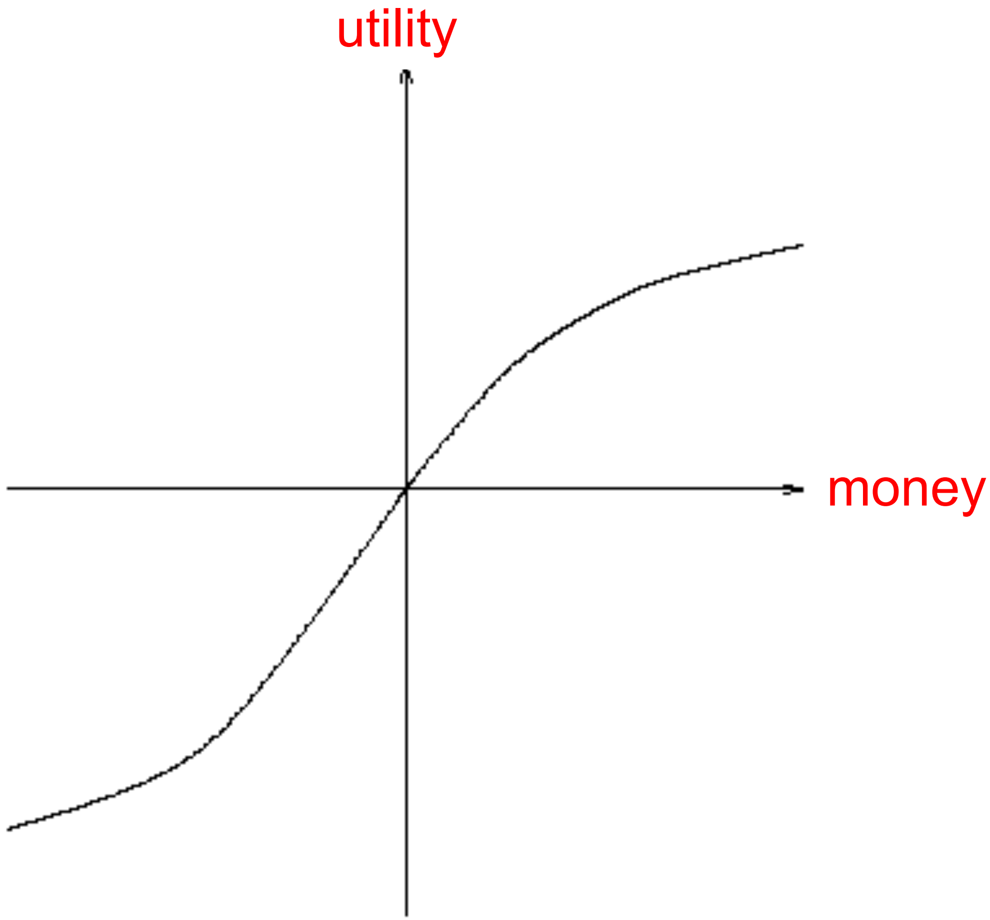
$u_i(\omega) \geq u_i(\omega')$ Abbreviated notation: $\omega \succeq \omega'$

- Preference ordering: state ω is strongly preferred by Ag. i over ω‘ iff

$u_i(\omega) > u_i(\omega')$ Abbreviated notation: $\omega \succ \omega'$

- Relation $\succeq$ is an ordering (reflexive, transitive, comparable)

What is utility?



Multi-agent encounters

- Agents simultaneously choose action to perform → result of the actions they select → outcome in Ω
- *Actual* outcome depends on the *combination* of actions
- Assume: each agent has just two possible actions *C* (“cooperate”) and *D* (“defect”)
- Environment behavior given by *state transformer function*:

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

Examples of state transformer functions

- $\tau(D, D) = \omega_1 \quad \tau(D, C) = \omega_2 \quad \tau(C, D) = \omega_3 \quad \tau(C, C) = \omega_4$

- $\tau(D, D) = \omega_1 \quad \tau(D, C) = \omega_1 \quad \tau(C, D) = \omega_1 \quad \tau(C, C) = \omega_1$

- $\tau(D, D) = \omega_1 \quad \tau(D, C) = \omega_2 \quad \tau(C, D) = \omega_1 \quad \tau(C, C) = \omega_2$

Examples of state transformer functions

- $\tau(D, D) = \omega_1 \quad \tau(D, C) = \omega_2 \quad \tau(C, D) = \omega_3 \quad \tau(C, C) = \omega_4$

  (environment is sensitive to actions of both agents.)

- $\tau(D, D) = \omega_1 \quad \tau(D, C) = \omega_1 \quad \tau(C, D) = \omega_1 \quad \tau(C, C) = \omega_1$

  (Neither agent has any influence in this environment.)

- $\tau(D, D) = \omega_1 \quad \tau(D, C) = \omega_2 \quad \tau(C, D) = \omega_1 \quad \tau(C, C) = \omega_2$

  (environment is controlled by *j*.)

Rational behaviour

- Assumption: Environment is sensitive to actions of both agents:

$$\tau(D,D)=\omega_1 \quad \tau(D,C)=\omega_2 \quad \tau(C,D)=\omega_3 \quad \tau(C,C)=\omega_4$$

- Utility functions:

$$u_i(\omega_1)=1 \quad u_i(\omega_2)=1 \quad u_i(\omega_3)=4 \quad u_i(\omega_4)=4$$
$$u_j(\omega_1)=1 \quad u_j(\omega_2)=4 \quad u_j(\omega_3)=1 \quad u_j(\omega_4)=4$$

- Short notation:

$$u_i(D,D)=1 \quad u_i(D,C)=1 \quad u_i(C,D)=4 \quad u_i(C,C)=4$$
$$u_j(D,D)=1 \quad u_j(D,C)=4 \quad u_j(C,D)=1 \quad u_j(C,C)=4$$

- → agent i's preferences: (also in short notation):

$$C,C \succeq_i C,D \quad \succ_i \quad D,C \succeq_i D,D$$

Rational behaviour

$$u_i(D,D)=1 \quad u_i(D,C)=1 \quad u_i(C,D)=4 \quad u_i(C,C)=4$$
$$u_j(D,D)=1 \quad u_j(D,C)=4 \quad u_j(C,D)=1 \quad u_j(C,C)=4$$

$$C,C \succeq_i C,D \quad \succ_i \quad D,C \succeq_i D,D$$

$$C,C \succeq_j D,C \quad \succ_j \quad C,D \succeq_j D,D$$

- "*C*" is the *rational choice* for i.

(Because i (strongly) prefers all outcomes that arise through *C* over all outcomes that arise through *D*.)

- "*C*" is the *rational choice* for j.

(Because j (strongly) prefers all outcomes that arise through *C* over all outcomes that arise through *D*.)

Payoff matrices

- Game theory: characterize the previous scenario in a *payoff matrix*:

	Agent *j* cooperates	Agent *j* defects
Agent *i* cooperates	4 4	1 4
Agent *i* defects	4 1	1 1

same as:

$$\begin{array}{llll} u_i(D,D)=1 & u_i(D,C)=1 & u_i(C,D)=4 & u_i(C,C)=4 \\ u_j(D,D)=1 & u_j(D,C)=4 & u_j(C,D)=1 & u_j(C,C)=4 \end{array}$$

- Agent i is "*row player*" – utility given in lower-left of cell
- Agent j is "*column player*" – utility given in upper-right of cell

Solution concepts

- How will a rational agent behave in any given scenario?
- Answered in *solution concepts*:
  - Dominant strategy
  - Nash equilibrium strategy
  - Pareto optimal strategy
  - Strategies that maximise social welfare

Dominant strategies

- A strategy s_i is dominant for player i if no matter what strategy s_j agent j chooses, i will do at least as well playing s_i as it would playing anything else.

- Unfortunately, there isn't always a dominant strategy.

(Pure Strategy) Nash Equilibrium

- In general, we will say that two strategies s_1 and s_2 are in Nash equilibrium if:

1. under the assumption that agent *i* plays s_1, agent *j* can do no better than play s_2; and
2. under the assumption that agent *j* plays s_2, agent *i* can do no better than play s_1.

- Neither agent has any incentive to deviate from a Nash equilibrium.
- Unfortunately:
  1. Not every interaction has a (pure strategy) Nash equilibrium.
  2. Some interactions have more than one Nash equilibrium.

The game of Matching Pennies

Players *i* and *j* simultaneously choose the face of a coin, either “heads” or “tails”.
If they show the same face, then *j* wins, while if they show different faces, then *i* wins.

Matching Pennies payoff matrix

	j **heads**	*j* **tails**
i **heads**	1 -1	-1 1
i **tails**	-1 1	1 -1

Mixed strategies for matching pennies

- NO pair of strategies forms a pure strategy Nash Equilibrium: whatever pair of strategies is chosen, somebody will wish they had done something else.

- The solution is to allow mixed strategies:
  - Play "heads" with probability 0.5
  - Play "tails" with probability 0.5

- This is a Nash Equilibrium strategy.
- Every finite game has a Nash Equilibrium in mixed strategies.
- But there might be more than one…

Pareto optimality

- An outcome is said to be *Pareto optimal* (or *Pareto efficient*) if there is no other outcome that makes one agent *better off* without making another agent *worse off*.

- If an outcome is Pareto optimal, then at least one agent will be reluctant to move away from it (because this agent will be worse off).

- If an outcome ω is *not* Pareto optimal, then there is another outcome ω' that makes *everyone* as happy, if not happier, than ω.

Social welfare

- The social welfare of an outcome ω is the sum of the utilities that each agent gets from ω:

$$\sum u_i(\omega)$$

- Think of it as the “total amount of money in the system”.
- As a solution concept, may be appropriate when the whole system (all agents) has a single owner (then overall benefit of the system is important, not individuals).

The Prisoner’s Dilemma

Two men are collectively charged with a crime and held in separate cells, with no way of meeting or communicating.

They are told that:

- if one confesses and the other does not, the confessor will be freed, and the other will be jailed for five years;
- if both confess, then each will be jailed for two years.

Both prisoners know that if neither confesses, then they will each be jailed for one year.

Prisoner's Dilemma payoff matrix

	Agent *j* cooperates	Agent *j* defects
Agent *i* cooperates	3 3	5 0
Agent *i* defects	0 5	2 2

Prisoner's Dilemma payoff matrix

	Agent *j* cooperates	**Agent *j* defects**
Agent *i* cooperates	3 3	5 0
Agent *i* defects	0 5	2 2

D is a dominant strategy.

(*D*, *D*) is the only Nash equilibrium.

All outcomes *except* (*D*, *D*) are Pareto optimal.

(*C*, *C*) maximises social welfare.

Rational behaviour

- suppose I cooperate: If j also cooperates → we both get payoff 3. If j defects → I get payoff 0. ➔ Best guaranteed payoff when I cooperate is 0
- suppose I defect: If j cooperates → I get payoff 5. If j also defects → both get payoff 2. ➔ Best guaranteed payoff when I defect is 2
- ➔ If I defect I'll get a minimum guaranteed payoff of 2. If I cooperate I'll get a minimum guaranteed payoff of 0.
- ➔ If prefer guaranteed payoff of 2 to guaranteed payoff of 0.
  ➔ I should defect

Rational behaviour

- Only one Nash equilibrium: (D,D). (“under the assumption that the other does D, one can do no better than do D”)
- Intuition says: (C,C) is better than (D,D) so why not (C,C)? → but if agent assumes that other agent does C it is BEST to do D! → seemingly “waste of utility“
- “Shocking” truth: defect is rational, cooperate is irrational

Examples

- Nuclear arms reduction (D: do not reduce, C: reduce)
- Paying for public goods like a TV license
- More generally, any form of exchange. Exchange is always sequential so why send the goods or be honest about their quality?
  - Buying a used car
  - Buying goods from an eBay seller

What does all this really mean?

- “Defect more rational than cooperate“ → Humans: Machiavellism (opposed to real altruism)
- Philosophical question: isn‘t even altruism ultimately some kind of optimisation towards OWN goals?!
- Further aspect: Strict rationalism (in case of prisoner‘s dilemma: defect) is usually only applied when sucker‘s payoff really hurts.
- What we have not yet regarded: Multiple sequential games between same players → “The shadow of the future” → What does it mean to rationalism and strategy?

The shadow of the future: Iterated Prisoner’s Dilemma

- Game is played multiple times. Agents can see all past actions of other agent.
- Course of reasoning:
  - If I defect, the other agent can punish me by defecting in the next run. (not a point in the one shot Prisoner‘s Dilemma game)
  - Testing cooperation (and possibly getting the sucker‘s payoff) is not tragic, because “on the long run” one (or several) sucker‘s payoff(s) is (are) “statistically” not important (can e.g. be equaled by gains through mutual cooperation)

- → in an iterated PD-game: cooperation is rational

The Folk Theorem of game theory

- In a game repeated for an indefinite number of times, with information about the past actions of other agents, cooperation will be a Nash equilibrium through conditional strategies.
  - “Cooperate if my partner cooperated last time, otherwise defect”
- Technically, any strategy that gives more than the minimax payoff will be an equilibrium.
- Doesn’t work if:
  - End time known
  - Insufficient information about the past actions of other agents.

So is the Folk Theorem useless?

- How many times will our agents meet???

- Institutions can provide reputational information about the past actions of other agents
  - e.g. Ebay's reputation system

Axelrod's Prisoner's Dilemma tournament

Some strategies competing:

- ALL-D: Always defect
- All-C: Always cooperate
- RANDOM: Choose D or C randomly
- TIT-FOR-TAT: On the first round cooperate; on round t do what opponent did on round t-1
- TESTER: Intention: Exploit "nice" programs that not punish D: On first round, test opponent with D. If opponent retaliates with D → play TIT-FOR-TAT. If opponent does C → play CCD.
- JOSS: like TIT-FOR-TAT, except periodically defect.

Axelrod’s Prisoner’s Dilemma tournament

- Winner was TIT-FOR-TAT (5 lines of Fortran code).
- Why? Overall Score of a strategy computed as average of performance against **all** other strategies. TIT-FOR-TAT was defeated by ALL-D.
- TIT-FOR-TAT won against “cooperative strategies“ → Again (as in analysis of single PD): being not too cooperative pays out.
- → Axelrod distilled some rules from outcome of tournament:

Axelrod’s Prisoner’s Dilemma tournament

1. Do not be envious.
2. Do not be first to defect.
3. Reciprocate C and D.
4. Don’t be too clever.

Snowdrift game (game of Chicken)

	Agent *j* cooperates	**Agent *j* defects**
Agent *i* cooperates	2 2	3 1
Agent *i* defects	1 3	0 0

Snowdrift game (game of Chicken)

	Agent *j* cooperates	Agent *j* defects
Agent *i* cooperates	2 2	3 1
Agent *i* defects	1 3	0 0

- Two (pure) Nash Equilibria: (D,C), (C,D)

(Assuming the other does D you can do no better than do C
Assuming the other does C you can do no better than do D)

- There is no dominant strategy.
- All outcomes except (D, D) are Pareto optimal.
- All outcomes except (D, D) maximise social welfare.

Pure coordination game (T-shirt game)

	Agent *j* wears red	Agent *j* wears blue
Agent *i* wears red	3 3	0 0
Agent *i* wears blue	0 0	3 3

Pure coordination game (T-shirt game)

	Agent *j* wears red	**Agent *j* wears blue**
Agent *i* wears red	3 3	0 0
Agent *i* wears blue	0 0	3 3

- Two Nash equilibria: (red, red), (blue ,blue)

- There is no dominant strategy.
- (red, red) and (blue, blue) are Pareto optimal.
- (red, red) and (blue, blue) maximise social welfare.

Stag-hunt game (risky coordination game)

	Agent *j* wears red	Agent *j* wears blue
Agent *i* wears red	3 3	1 0
Agent *i* wears blue	0 1	1 1

Stag-hunt game (risky coordination game)

	Agent *j* wears red	Agent *j* wears blue
Agent *i* wears red	3 3	1 0
Agent *i* wears blue	0 1	1 1

- Two Nash equilibria: (red, red), (blue ,blue)

- (blue, blue) is risk-dominant -- it has the largest basin of attraction (is less risky).
- (red, red) is payoff-dominant – it offers to each player at least as much payoff as the other Nash equilibria.
- (red, red) is Pareto optimal.
- (red, red) maximises social welfare.

Edinburgh Napier UNIVERSITY

Acknowledgements

- Much of this lecture is based on Chapter 11 of Wooldridge "An Introduction to MultiAgent Systems", 2nd edition
  - http://www.cs.ox.ac.uk/people/michael.wooldridge/pubs/imas/IMAS2e.html
  - And draws upon the accompanying slides by Dr. Georg Groh, TU-München, Germany http://www.cs.ox.ac.uk/people/michael.wooldridge/pubs/imas/alt-slides.zip