COORDINATION GAMES

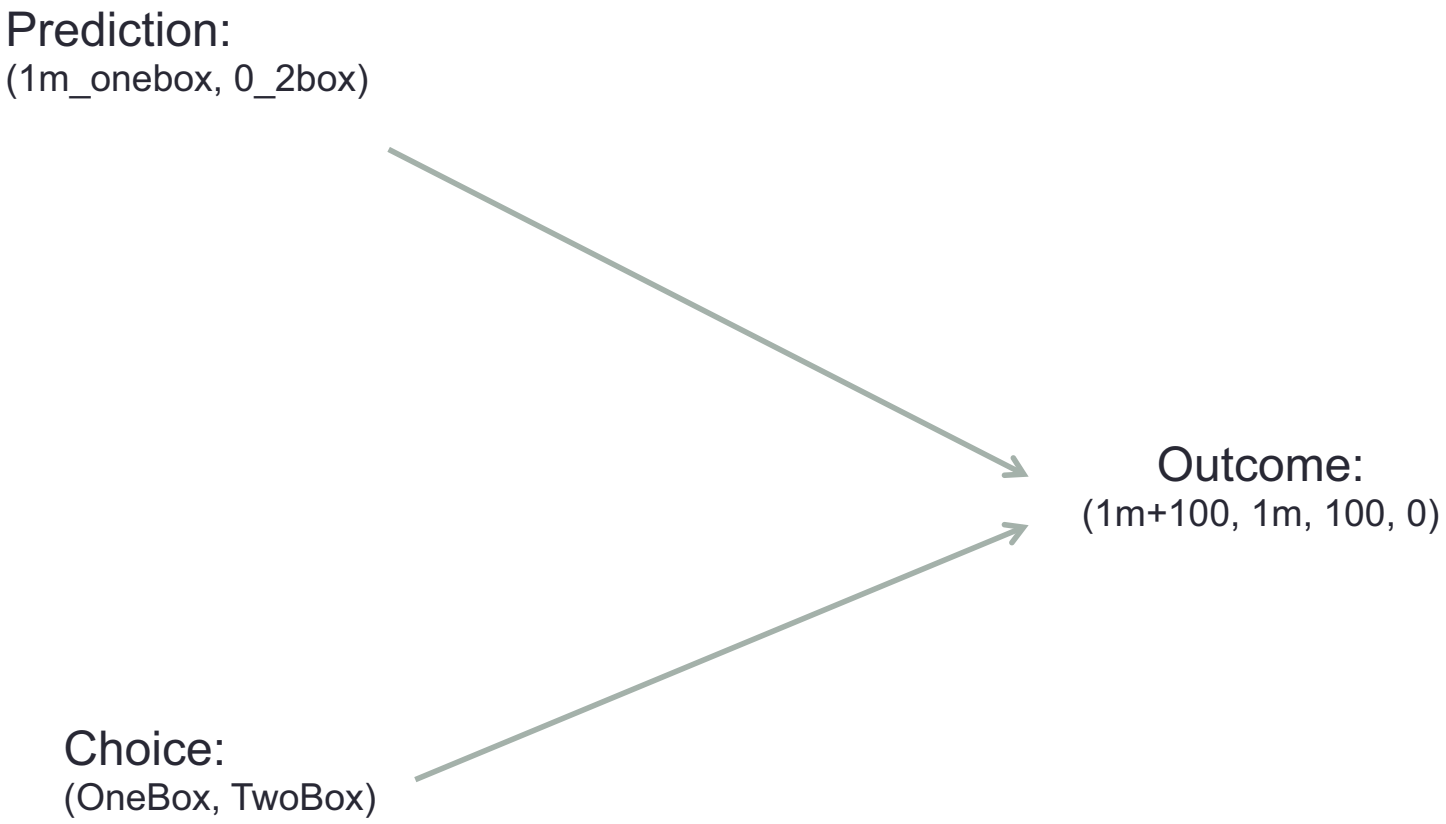
Nash Equilibria, Schelling Points and the Prisoner's Dilemma

Owain Evans, MIT Paradox, Monday 25 February 2013.

Newcomb's Paradox

Image by MIT OpenCourseWare.

Newcomb's Paradox: Causal Graph



Newcomb's Paradox: EDT problems

- EDT fails:
  - If Predictor perfect, then P($0 / OneBox) is undefined.
  - Assume I know I'm rational: conditional on any action, I must obtain the highest possible utility in this situation. So choose randomly.
  - Double transparent box Newcomb and other exotic problems.

Newcomb's Paradox: CDT problems

- Does Newcomb reward irrationality? Example of God who kills people for using "the best possible decision theory".
- Key observation: Reward for OneBoxing, but no dependence on how the decision is made. Best possible theory could succeed.
- Reflective-consistency:
  - CDT agent, given the chance to change DT before Newcomb
  - Or: CDT agent, enters an AI in a competition where one of challenges is NP
  - CDT will modify itself into a non-CDT theory.
  - Conclusion: CDT is not stable under intelligence and so no smart CDT agents will survive.

PD: Structure

	Cooperate	Defect
Cooperate	(4,4)	(1,5)
Defect	(5,1)	(2,2)

PD: Scenarios in Economics

Gains from Trade:

- Player 1 needs wool as much as Player 2 needs wheat.
- Trade is conducted by exchanging sealed boxes. By the time boxes can be opened, the other Player has left.

Cartel:

- Two sellers of a good. Prices are announced simultaneously and can't be changed.

Pivot: Coordination Games

- GOAL: Find a decision theory that “wins” on PD and NP and does well on standard problems.

- **Strategy**:
  - PD and NP are coordination problems. Study other coordination problems and find decision theories that solve them.

Why is game theory hard?

- One-player games are easy:
  - Work out the consequences of each action and take action with highest EV.
  - Example: Playing the lottery or 1-player casino games; betting on the weather.

Why is game theory hard?

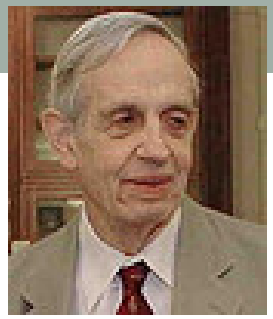
- Simple coordination game 1: (normal pset)

	Group Study	Cafe
Group Study	(1,1)	(0.5,0)
Cafe	(0,0.5)	(0,0)

- If players are symmetric, then simulation leads to infinite regress. (Can't happen with natural systems).

- With asymmetry, simulation is possible. Example: PD vs. religious law-follower.

Nash Equilibrium



- In physics and biology, you can sometimes make predictions even if simulation is intractable:
  - Complex slope, but easy to find the stable equilibria
  - Sex-ratio in biology
- Informal definition: A set of strategies/actions is a *Nash Equilibrium* if no player can do better by changing his strategy while everyone else's are held fixed.
- Idea: A non-Nash pair of actions is unstable because one of the players can do better by doing something else.

Nash Equilibrium

- Simple coordination game: we can predict outcome using NE, and if players each play NE, then they'll do well.
- PD: If each player plays NE, they both defect.
- Also for purely competitive games: e.g. Rock-Paper-Scissors. (No NE in pure strategies).

Image courtesy of TEDx Athens on Flickr. Available CC BY-NC.

- Extensively studied and applied in economics. Also studies of computational complexity of finding the NE by MIT's Daskalakis.

Schelling game: Rules

- You get a point if you and your partner provide the same answer.

- You should face away from each other and are not allowed to communicate in any way before writing your answer down.

Image courtesy of New America Foundation on Flickr. Available CC BY-NC-SA.

Multiple Equilibria Coordination Games

- Simple coordination game 2: (collaborative project)

	Group Study	Cafe
Group Study	(1,1)	(0,0)
Cafe	(0,0)	(0.5,0.5)

- Schelling game: (café is quiet)

	Group Study	Cafe
Group Study	(1,1)	(0,0)
Cafe	(0,0)	(1,1)

Schelling Games

- Many examples: Driving Problem, Rowing, Deciding on Linguistic Conventions

- How to resolve:
  - One player goes first
  - One player can simulate the other (asymmetric)

24.118 Paradox & Infinity
Spring 2013