# Motiviation: Auctions

FCC spectrum auction

- Bidder A winning license 37 for $1M.
- Bidders A and B competing for license 63.
- Simultaneously, Bidder B bids:
  - license 37: $1.1M ← threat!
  - license 63: $13,000,037

*First steps toward agents that can reason this way:*

*Negotiation without explicit communication!*

# Matrix Game Model

Simple, yet instructive model for 2-player interactions.

$$M_1 = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}, M_2 = \begin{bmatrix} u & v & w \\ x & y & z \end{bmatrix}$$

Player 1 chooses a row, player 2 chooses a column.

Player $i$ payoff determined by entry in $M_i$.

*Iterated matrix game*, repeat over unbounded *stages*.

# Policy Types

Generally: action choice conditioned on full history.

Usually: finite amount of history.

Deterministic: choose the same action in every stage

Memoryless (0): fixed probability distribution

Bigram (1): condition action on previous action choice

Repeated interaction: influence future behavior (threats).

Game theory literature: “folk theorems”.

# Leader Strategies

If your opponent learns, stubbornness and threats help.

Leader: Assume opponent is learning how to respond.

*We describe general strategies that can issue threats to lead learners to cooperate.*

- Bully
- Godfather

# Bully

Bully is a deterministic, memoryless policy:

$$M_1 = \begin{bmatrix} 1 & 2 & 6 \\ 5 & 2 & 9 \end{bmatrix}, M_2 = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 5 & 2 \end{bmatrix}$$

$M_1$: leader's payoff matrix, $M_2$ follower's payoff matrix.

Oligopoly lit.: "Stackelberg leader" (Fudenberg and Levine 98)

# Godfather

Finite-state: “makes its opponent an offer it can’t refuse.”

$$M_1 = \begin{bmatrix} 1 & 2 & 6 \\ 5 & 2 & 9 \end{bmatrix}, M_2 = \begin{bmatrix} 4 & 1 & 3 \\ 1 & 5 & 2 \end{bmatrix}$$

- Security level $(2, \sim 2.7)$. Dominating cell $(6, 3)$.
- Lead with cell action.
- Punish uncooperativeness with security level.

Threat: “Play your action from the cell, or I’ll force you to get no more than your security level no matter what.”

Generalization of tit-for-tat (Axelrod 84).

# Experiments

Bully, Godfather, $Q_0$ and $Q_1$ vs. $Q_0$ & $Q_1$ in several games

Parameters:

- $\varepsilon = 0.1$
- 30,000 stages of learning
- average payoff over the final 5,000 stages
- mean and standard deviation over 100 experiments

# Test Games

We used games with a common structure:

- $2 \times 2$ bimatrix games ("cooperate", "defect")
- symmetric payoffs

$$M_1 = \begin{bmatrix} 3 & y \\ x & 1 \end{bmatrix}, M_2 = \begin{bmatrix} 3 & x \\ y & 1 \end{bmatrix}$$

Games:

- deadlock
- assurance
- prisoner's dilemma
- chicken

# Deadlock: An Obvious Choice

Always better off cooperating:

$$M_1 = \begin{bmatrix} 3 & 2 \\ 0 & 1 \end{bmatrix}, M_2 = \begin{bmatrix} 3 & 0 \\ 2 & 1 \end{bmatrix}$$

Bully cooperates. Godfather cooperates, defect as threat.

| | $Q_0$ | $Q_1$ | Bully | GF |
|---|---|---|---|---|
| $Q_0$ | 2.8 | 2.8 | 3.0 | 2.8 |
| $Q_1$ | 2.8 | 2.8 | 3.0 | 2.8 |

# Assurance: Suboptimal Preference

More important to match the other than to cooperate:

$$M_1 = \begin{bmatrix} 3 & 0 \\ 2 & 1 \end{bmatrix}, M_2 = \begin{bmatrix} 3 & 2 \\ 0 & 1 \end{bmatrix}$$

Q-learners coordinate with no particular bias.

| | $Q_0$ | $Q_1$ | Bully | GF |
|---|---|---|---|---|
| $Q_0$ | 1.4* | 1.5* | 2.8 | 1.4* |
| $Q_1$ | 1.9* | 1.7* | 2.8 | 2.8 |

(Stars mark numbers with high variance, more than $0.15$).

# PD: Incentive to Defect

Better off defecting:

$$M_1 = \begin{bmatrix} 3 & 0 \\ 5 & 1 \end{bmatrix}, M_2 = \begin{bmatrix} 3 & 5 \\ 0 & 1 \end{bmatrix}$$

Bully defects, Godfather is tit-for-tat.

| | $Q_0$ | $Q_1$ | Bully | GF |
|---|---|---|---|---|
| $Q_0$ | 1.2* | 1.2* | 1.2 | 1.4* |
| $Q_1$ | 1.2 | 1.2 | 1.2 | 2.9 |

Godfather lures $Q_0$ to cooperate for short periods of time.

# Chicken: Incentive to Exploit

Each player is better off choosing the opposite:

$$M_1 = \begin{bmatrix} 3 & 1.5 \\ 3.5 & 1 \end{bmatrix}, M_2 = \begin{bmatrix} 3 & 3.5 \\ 1.5 & 1 \end{bmatrix}$$

Feign stupidity! Learning problem is meta-chicken.

Godfather+$Q_1$ reaches mutual cooperation

| | $Q_0$ | $Q_1$ | Bully | GF |
|---|---|---|---|---|
| $Q_0$ | 2.5* | 2.5* | 3.4 | 2.8 |
| $Q_1$ | 2.4* | 2.9 | 3.4 | 2.9 |

Bully overpowers others, but loses to self (unlike GF).