Last update: October 25, 2012

Game theory

CMSC 421, Section 17.5

Chapter 17, Section 5: Game theory

◇ In Chapter 5 we looked at 2-player perfect-information zero-sum games

◇ We'll now look at games that might have one or more of the following:

- more than 2 players
- imperfect information
- nonzero-sum outcomes

◇ Recall that an agent's *strategy* is a specification of what the agent will do in every game state where it's the agent's move.

The Prisoner's Dilemma

- Scenario: The police have arrested two suspects for a crime.
- They tell each prisoner they'll reduce his/her prison sentence if he/she betrays the other prisoner
- Each prisoner must choose between two actions:
  - cooperate with the other prisoner, i.e., don't betray him
  - defect (betray the other prisoner).
- Years in prison, represented as negative utility values:

		P_2: Cooperate	P_2: Defect
P_1	Cooperate	P_1: -2, P_2: -2	P_1: -5, P_2: 0
	Defect	P_1: 0, P_2: -5	P_1: -4, P_2: -4

- Non-zero-sum
- Imperfect information
  - neither player knows the other's move until after *both* players have moved

Notation

◇ Add 5 so the numbers are ≥ 0; abbreviate names or omit them

	C	D
C	$3, 3$	$0, 5$
D	$5, 0$	$1, 1$

- P_1 chooses a row, and gets the 1st payoff in each square
- P_2 chooses a column, and gets the 2nd payoff in each square

◇ In the book, the roles of P_1 and P_2 are interchanged

- In the Prisoner's Dilemma, it doesn't matter because the game is *symmetric* (same game if you interchange the two players)

◇ But not all games are symmetric:

	1	2
1	$2, -2$	$-3, 3$
2	$-3, 3$	$4, -4$

Strategies

◇ Players $P_1, \ldots, P_n$

- $S_i = \{\text{all possible strategies for } P_i\}$
- s_i will always refer to a strategy in S_i

◇ *Strategy profile*: an n-tuple $(s_1, s_2, \ldots, s_n)$, one strategy for each player

◇ *Utility*: $U_i(s_1, \ldots, s_n)$ = payoff for P_i with strategy profile $(s_1, \ldots, s_n)$

◇ s_i *strongly dominates* s'_i if P_i always does better with s_i than with s'_i:

- $\forall\ s_1, \ldots, s_{i-1},\ s_{i+1}, \ldots, s_n,$
  $U_i(s_1, \ldots, s_{i-1}, s_i, s_{i+1}, \ldots, s_n) > U_i(s_1, \ldots, s_{i-1}, s'_i, s_{i+1}, \ldots, s_n)$

◇ s_i *weakly dominates* s'_i if P_i never does worse with s_i than with s'_i, and there is at least one case where P_i does better with s_i than with s'_i

- $\forall\ s_1, \ldots, s_{i-1},\ s_{i+1}, \ldots, s_n,\ U_i(\ldots, s_i, \ldots) \geq U_i(\ldots, s'_i, \ldots)$
- $\exists\ s_1, \ldots, s_{i-1},\ s_{i+1}, \ldots, s_n\quad U_i(\ldots, s_i, \ldots) > U_i(\ldots, s'_i, \ldots)$

Dominant strategy equilibrium

- s_i is a *(strongly, weakly) dominant* strategy if it (strongly, weakly) dominates every $s'_i \in S_i$.
- *Dominant strategy equilibrium*: a strategies $(s_1, \ldots, s_n)$ such that each s_i is dominant for player P_i
  - Thus P_i will do best by using s_i rather than a different strategy,
    - regardless of what strategies the other players use

- The Prisoner's Dilemma has a dominant strategy equilibrium
  - What is it?

	C	D
C	$3, 3$	$0, 5$
D	$5, 0$	$1, 1$

- What can happen if you don't play your dominant strategy:
  - http://www.youtube.com/watch?v=ED9gaAb2BEw

Pareto optimality

◇ A strategy profile $(s_1, \ldots, s_n)$ is *Pareto optimal* if there's no strategy profile $(s'_1, \ldots, s'_n)$ that gives all players higher payoffs

	C	D
C	$3, 3$	$0, 5$
D	$5, 0$	$1, 1$

◇ (C,C) is Pareto optimal
- So are (C,D) and (D,C)

◇ (D,D) is the *only* strategy profile that isn't Pareto optimal

Coordination games

◇ Not every game has a dominant strategy equilibrium

◇ Example: which side of the road?

- 2 people driving toward each other in a country with no traffic rules
- Each needs to decide which side of the road to drive on

	L	R
L	1, 1	0, 0
R	0, 0	1, 1

◇ Why did I use 1 and 0 for the payoffs?

◇ How to decide which side of the road?

(1) guess

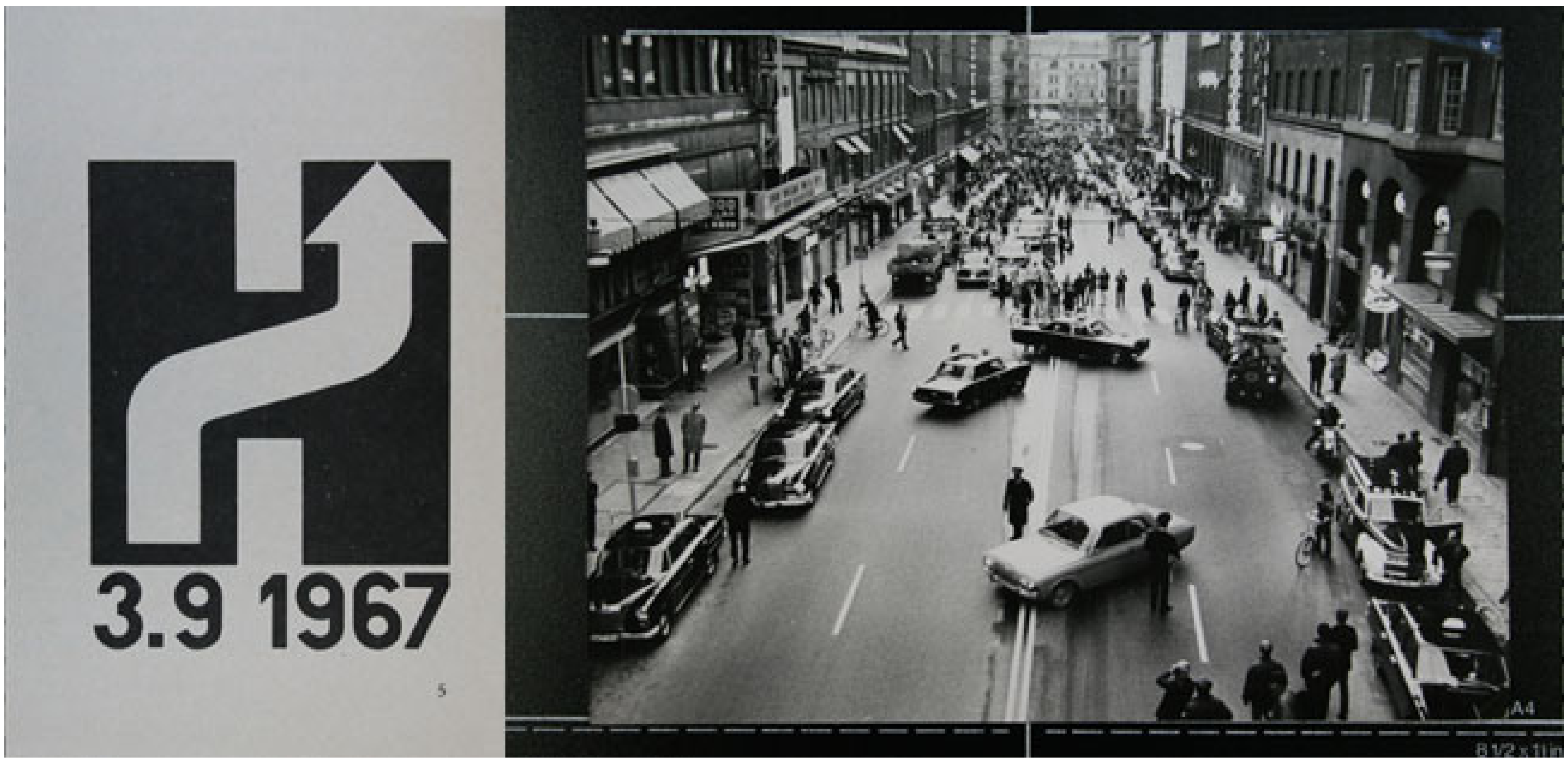
(2) change the rules of the game

Changing the rules of the game

◇ *Mechanism design*: set up the rules of the game, to give each agent an incentive to choose a desired outcome

- E.g., pass a law saying what side of the road to drive on

◇ Sweden on September 3, 1967:



Best response and Nash equilibrium

◇ Suppose players $P_1, \ldots, P_n$ have chosen the following strategy profile:

- $\sigma = (s_1, \ldots, s_{i-1}, s_i, s_{i+1}, \ldots, s_n)$

◇ P_i's strategy s_i is a *best response* to the other players' strategies in σ if for every strategy $s_i' \in S_i$,

- $U_i(s_1, \ldots, s_{i-1}, s_i, s_{i+1}, \ldots, s_n) \geq U_i(s_1, \ldots, s_{i-1}, s_i', s_{i+1}, \ldots, s_n)$
- i.e., if P_i switches to a different strategy and nobody else does, then P_i won't do any better, and might do worse

◇ Suppose that for **every** player i, P_i's strategy s_i is a best response to the other players' strategies in σ

- Then σ is a *Nash equilibrium* (named after John Nash)

◇ Basically a local optimum:

- No player can benefit from *unilaterally* switching to a different strategy

Example

	L	R
L	$1, 1$	$0, 0$
R	$0, 0$	$1, 1$

- ♢ Two Nash equilibria: (L, L) and (R, R)

- ♢ Every game has a Nash equilibrium
  - (subject to a condition I'll describe later)
- ♢ Every dominant strategy equilibrium is a Nash equilibrium
  - but not vice versa

Mixed strategies

- Two-finger Morra:
  - Two players: E (*Even*) and O (*Odd*). Each holds up 1 or 2 fingers:

O

E		one	two
	one	$2, -2$	$-3,\ 3$
	two	$-3,\ 3$	$4, -4$

- If this game has a Nash equilibrium, then what is it?
  - Not (one, one): O can do better by changing to "two"
  - Not (one,two): E can do better by changing to "two"
  - Likewise, not (two,one) or (two,two)
- There is no equilibrium in *pure* (deterministic) strategies
- Equilibrium: both players use a *mixed* (randomized) strategy:
  - [Pr(one)=7/12, Pr(two)=5/12]

Von Neumann's maximin technique

◇ Suppose O's strategy is [Pr(one)=q, Pr(two)=$1-q$]

- If E plays *one*, E's expected utility is $2q - 3(1-q) = 5q - 3$
- If E plays *two*, E's expected utility is $-3q + 4(1-q) = 4 - 7q$

◇ E's *best response* is to choose whichever move (*one* or *two*) produces a greater expected utility. If E does this, E's expected utility is

$$U_{E|q} = \max(5q - 3, 4 - 7q)$$

◇ O's expected utility is the negative of E's, so O's best strategy is to choose a value of q that minimizes $U_{E|q}$:

$$\arg\min_q U_{E|q} = \arg\min_q(\max(5q - 3, 4 - 7q))$$

◇ This occurs where the line $y = 5q - 3$ intersects the line $y = 4 - 7q$

$$5q - 3 = 4 - 7q \;\Rightarrow\; 12q = 7 \;\Rightarrow\; q = 7/12$$

◇ If O uses this, then E's expected utility is $-1/12$, and O's is $1/12$, regardless of what move E makes

- E can't benefit by unilaterally changing to a different strategy

Von Neumann's maximin technique

◇ Suppose E's strategy is [Pr(one)=p, Pr(two)=$1-p$]

- If O plays *one* then E's expected utility is $2p - 3(1-p) = 5p - 3$.
- If O plays *two* then E's expected utility is $-3p + 4(1-p) = 4 - 7p$.

◇ O's best response is to choose whichever move (*one* or *two*) produces a greater expected utility for O (i.e., a smaller expected utility for E)

- This gives E the following expected utility:

$$U_{E|p} = \min(5p - 3, 4 - 7p)$$

so E's best strategy is to choose p that maximizes $U_{E|p}$:

$$\arg\max_p U_{E|p} = \arg\max_p(\min(5p - 3, 4 - 7p))$$

◇ This occurs where the lines $y = 5p - 3$ and $y = 4 - 7p$ intersect

$$5p - 3 = 4 - 7p \;\Rightarrow\; 12p = 7 \;\Rightarrow\; p = 7/12$$

◇ If E uses this, then E's expected utility is $-1/12$, and O's is $1/12$, regardless of what move O makes

- O can't benefit by unilaterally changing to a different strategy

Von Neumann's maximin technique

◇ Suppose that

- E's strategy is [Pr(one)=7/12, Pr(two)=5/12]
- O's strategy also is [Pr(one)=7/12, Pr(two)=5/12]

◇ Then

- E's expected utility is $-1/12$ and O's is $1/12$

◇ If either player unilaterally changes to a different strategy, the expected utilities don't change

- So we have a Nash equilibrium

◇ When can we expect players to choose a Nash equilibrium?

- Requires *common knowledge of rationality*

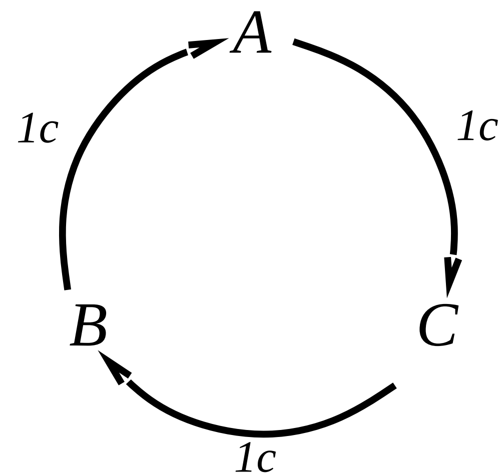
Rational preferences

- Let $O = \{o_1, \ldots, o_k\}$ be the set of all possible outcomes of some choice
- For every pair of outcomes $o, o' \in O$, which do you prefer?
  - Either you prefer o, or you prefer o', or both are equally preferable
- There are mathematical axioms defining when these preferences are decision-theoretically *rational*
  - e.g., rational preferences must be transitive:
    - prefer o to o' and prefer o' to o'' $\Rightarrow$ prefer o to o''
- This isn't psychological rationality, it's mathematical consistency
  - But if someone's preferences aren't decision-theoretically rational, they can be induced to do things that seem self-evidently irrational

Example: intransitive preferences

◇ Suppose that an agent

- prefers A to B
- prefers B to C
- prefers C to A



◇ Such an agent would

- trade C plus some money to get B
- trade B plus some money to get A
- trade A plus some money to get C

◇ Such an agent can be induced to give away all its money

Principle of maximum expected utility (MEU)

◇ **Theorem:**

- Every rational set of preferences correspond to a *utility function*
  - ◇ a function that assigns a numeric *utility value* to each outcome
- Choices that satisfy the preferences

  = choices that maximize expected utility

Common knowledge

- A fact is *common knowledge* if
  - Everyone knows it
  - Everyone knows that everyone knows it
  - Everyone knows that everyone knows that everyone knows it

    . . .
  - And so on, *ad infinitum*

- Knowing that everyone knows something can make a big difference
  - http://www.youtube.com/watch?v=3-son3EJTrU

Nash equilbrium

- ◇ Consider a game that has a unique Nash equilibrium
  - If all of the players are decision-theoretically rational and if there is common knowledge of rationality
  - then we can expect them to choose the Nash equilibrium

- ◇ If a game has more than one Nash equilibrium, then it's more complicated

	L	R
L	$1, 1$	$0, 0$
R	$0, 0$	$1, 1$

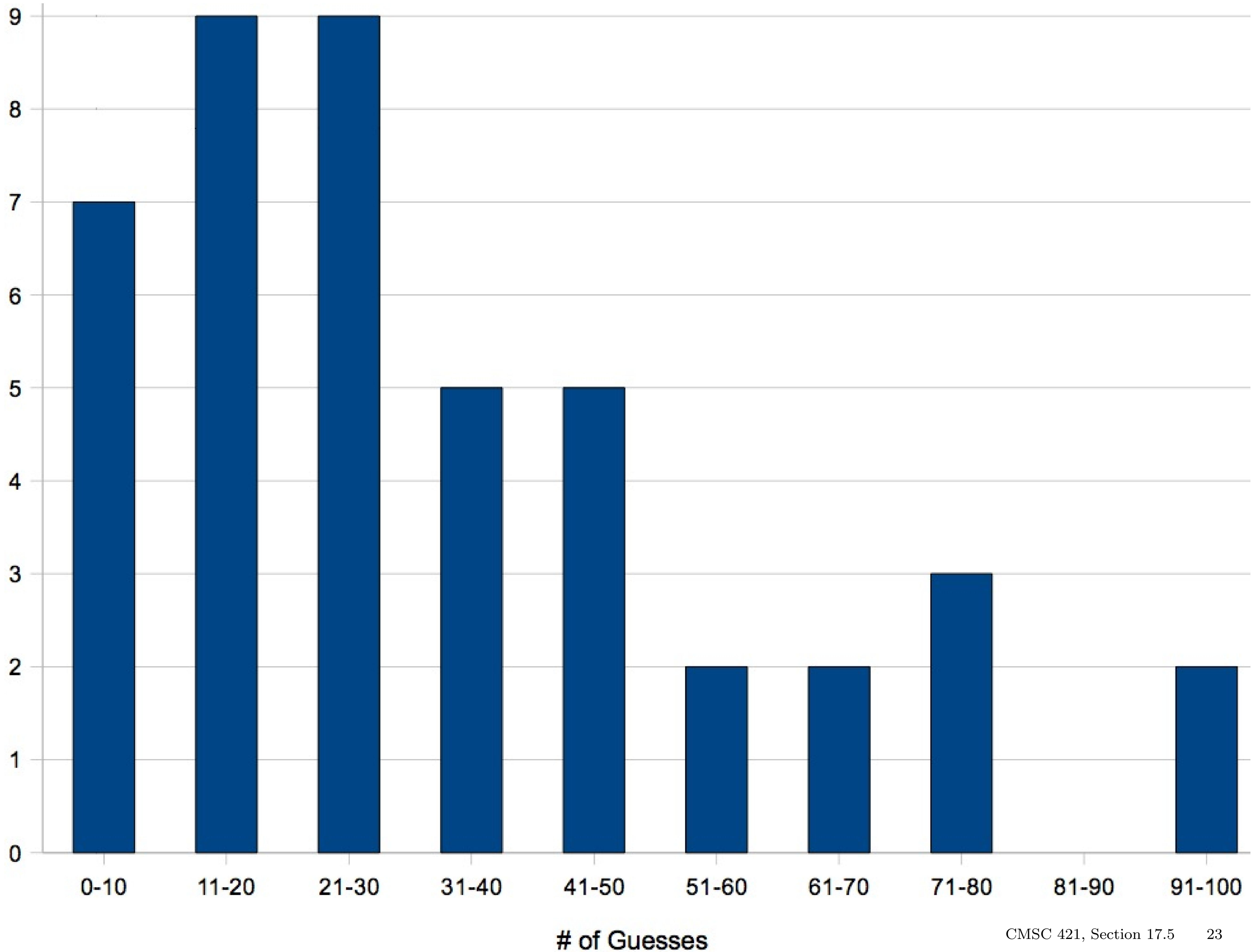
  - Need to either guess or communicate

The p-beauty contest

- ◇ Earlier this semester, we played the following game:
  - Everyone chooses a number in the range from 0 to 100.
  - The winner(s) are whoever chooes a number that's closest to 2/3 of the average of all of the numbers.
- ◇ This game is famous among economists and game theorists
  - It's called the *p-beauty contest*; I used $p = 2/3$
- ◇ What does game theory tell us about it?

Nash equilibrium for the p-beauty contest

- ◇ We can find a Nash equilibrium using *backward induction*
- ◇ All of the numbers are ≤ 100
  - average $\leq 100 \Rightarrow 2/3$ of the average < 67
- ◇ If rationality is common knowledge, they'll all choose numbers < 67
  - average $< 67 \Rightarrow 2/3$ of the average < 45
- ◇ If rationality is common knowledge, they'll all choose numbers < 45
  - average $< 45 \Rightarrow 2/3$ of the average < 30
  - ...
- ◇ Nash equilibrium: everybody chooses 0


9
8
7
6
5
4
3
2
1
0
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100
# of Guesses

Why choose a non-equilibrium strategy?

◇ Limitations in reasoning ability

- Maybe you didn't calculate the Nash equilibrium correctly, or you didn't know how to calculate it, or you didn't know the concept

◇ Incorrect utilities

- Maybe the payoff matrix doesn't represent your actual preferences

◇ Opponent modeling

- If an opponent uses a non-equilibrium strategy, your best response will be a non-equilibrium strategy
- If you can predict the other agents' likely actions, you can play an approximation of your best response
- You may be able to do much better that way

Rock-Paper-Scissors

	Rock	Paper	Scissors
Rock	0, 0	-1, 1	1, -1
Paper	1, -1	0, 0	-1, 1
Scissors	-1, 1	1, -1	0, 0

◇ Nash equilibrium:

- Both players choose randomly, probability 1/3 for each move
- Expected utility = 0

Rock-Paper-Scissors

- International rock-paper-scissors programming competition, 1999
  - www.cs.ualberta.ca/~darse/rsbpc1.html
- Round-robin tournament:
  - 55 programs, 1000 iterations for each pair of programs
  - Lowest possible score = –55000; highest possible score = 55000
- Average over 25 tournaments:
  - Lowest score (*Cheesebot*): –36006
  - Highest score (*Iocaine Powder*): 13038

Prisoner's Dilemma

	C	D
C	$3, 3$	$0, 5$
D	$5, 0$	$1, 1$

- ◇ (D,D) is a dominant strategy equilibrium, but it isn't Pareto optimal
- ◇ (C,C) is Pareto optimal, but it's not a Nash equilibrium

- ◇ How to get both players to choose C?
  - Each player must be willing to forego the personal gain that he/she would get from defecting
  - Each player has to trust the other to do the same

- ◇ How to make this happen?

Repeated games

- In a *repeated* or *iterated* game, some game G is played multiple times by the same set of players
  - G is called the *stage game*
  - Each occurrence of G is called an *iteration*, *round*, or *stage*
- Usually each player knows what all players did in the previous iterations, but not what they're doing in the current iteration
  - Thus, imperfect information
- Usually the final score is the sum of the payoffs in all the iterations

Iterated Prisoner's Dilemma

	C	D
C	$3, 3$	$0, 5$
D	$5, 0$	$1, 1$

◇ *Iterated Prisoner's Dilemma*: play the Prisoner's Dilemma repeatedly

- Score is the sum of the payoffs in all the iterations

◇ If you defect and they cooperate, you get a short-term gain

- But they might punish you next time by defecting

◇ You can both do well if you both cooperate with each other

◇ How to establish and maintain cooperation,

- without letting them take advantage of you?

Some well-known strategies

- Tit for Tat (TFT):
  - Move 1: cooperate
  - Move i: do what the other player did on move $i-1$

Iteration	TFT	other player
1	C	C
2	C	D
3	D	C
4	C	C

- Tit for Two Tats: cooperate unless the other player defected twice
- GRIM: if the other player ever defects, never cooperate again
- AllC: always cooperate
- AllD (the "hawk" strategy): always defect
- Tester: defect on round 1, cooperate on round 2
  - If the opponent defects on round 2
    - Cooperate on round 3 and play Tit-for-Tat from then on
  - Otherwise, randomly intersperse cooperation and defection

TFT with other players

◇ Axelrod's famous tournaments

- *The Evolution of Cooperation*, 1985

◇ In these tournaments, TFT usually did best

- It could establish and maintain cooperations with many other players
- It could prevent malicious players from taking advantage of it

TFT	AllC	TFT	AllD	TFT	Grim	TFT	TFT	TFT	Tester
C	C	C	D	C	C	C	C	C	D
C	C	D	D	C	C	C	C	D	C
C	C	D	D	C	C	C	C	C	C
C	C	D	D	C	C	C	C	C	C
C	C	D	D	C	C	C	C	C	C
C	C	D	D	C	C	C	C	C	C
C	C	D	D	C	C	C	C	C	C
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

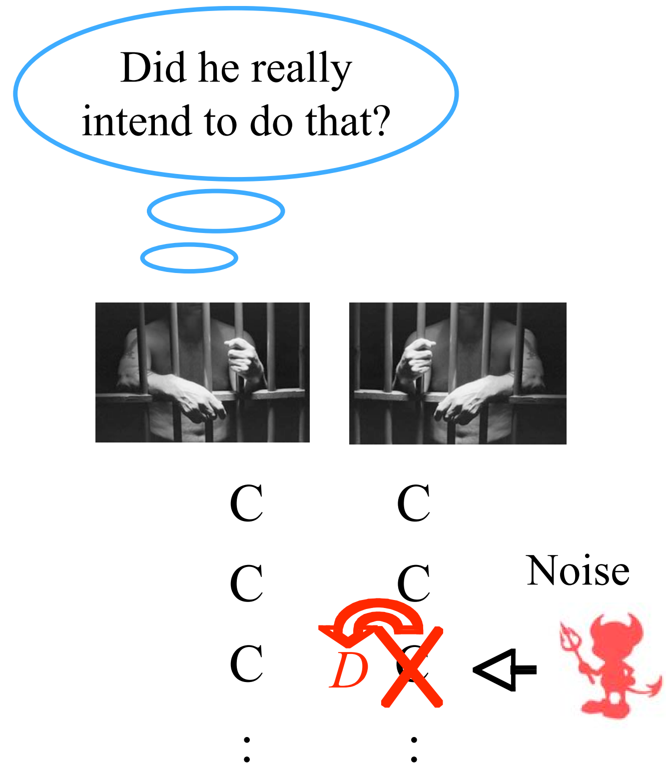
◇ Axelrod also looked at analogies with various human behaviors

Example: trench warfare in World War I

- Incentive to cooperate:
  - If I attack the other side, then they'll retaliate and I'll get hurt
  - If I don't attack, maybe they won't either
- Result: tacit cooperation
  - Even though the soldiers were supposed to be enemies, they tried to avoid attacking each other

Iterated Prisoner's Dilemma with Noise

- On each move, a nonzero probability that C will be recorded as D, and vice versa
- Can use this to model accidents or misinterpretations



Example of noise

◇ Story from a British army officer in World War I:

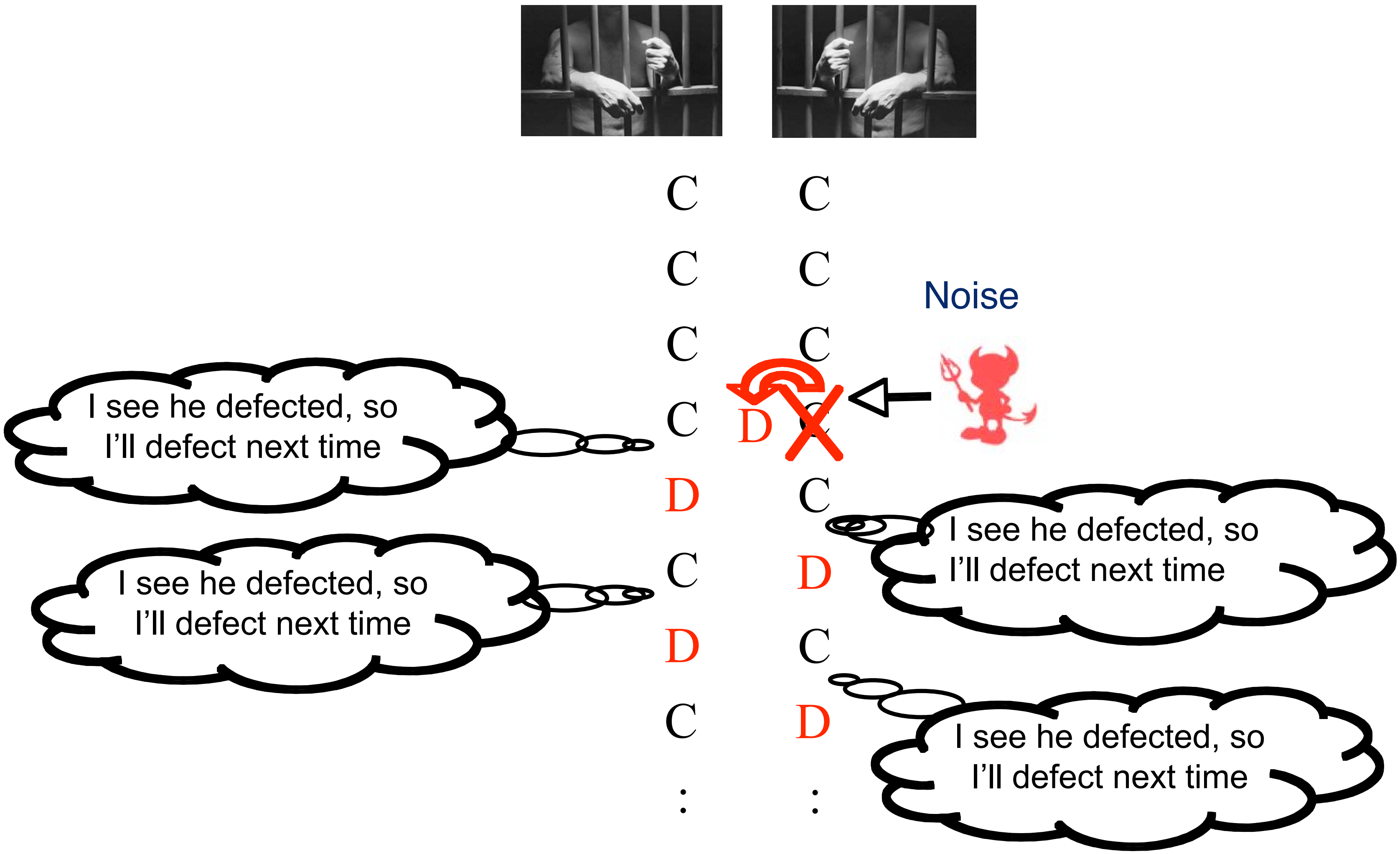
> I was having tea with A Company when we heard a lot of shouting and went out to investigate. We found our men and the Germans standing on their respective parapets. Suddenly a salvo arrived but did no damage. Naturally both sides got down and our men started swearing at the Germans, when all at once a brave German got onto his parapet and shouted out: "We are very sorry about that; we hope no one was hurt. It is not our fault. It is that damned Prussian artillery."

◇ The salvo wasn't the German infantry's intention

- They didn't expect it, and didn't want it

Effects of noise

◇ Noise causes big problems for Tit-for-Tat and similar strategies:



Some strategies for the noisy IPD

- One idea: be more forgiving in the face of apparent defections
  - Tit-For-Two-Tats (TFTT)
    - Retaliate only if the other player defects twice in a row
  - Generous Tit-For-Tat (GTFT)
    - Forgive randomly: small probability of cooperation if the other player defects
  - Pavlov (Win-Stay, Lose-Shift)
    - Repeat my previous move if I got 3 or 5 points last time
    - Reverse my previous move if I got 0 or 1 points last time
    - If the other player defects continuously, Pavlov will alternatively cooperate and defect
- Problem: more forgiving $\Rightarrow$ can be exploited by an unscrupulous opponent

Discussion

◇ The British army officer's story:

a brave German got onto his parapet and shouted out: "We are very sorry about that; we hope no one was hurt. It is not our fault. It is that damned Prussian artillery."

◇ The apology avoided a conflict

- It was convincing because it was consistent with the German infantry's past behavior
- The British had ample evidence that the German infantry wanted to keep the peace

◇ Principle: if you can tell which actions are affected by noise, you can avoid reacting to the noise

DBS

- ♢ Author: Tsz-Chiu Au (PhD graduate of mine, now a professor in Korea)
  - A program to play the noisy IPD
  - Tries to detect when noise occurs, and respond to the move that the other player *intended* rather than the one that was recorded

- ♢ Based on observations of the other player's recent behavior, DBS builds a simple, approximate model of their strategy
  - gives a probabilistic prediction of their next move

- ♢ DBS uses the model to filter out the noise
  - If the model says the player will cooperate (or defect) with probability 1, and you see them do the opposite, assume you saw noise
  - But if that happens too many times, assume their strategy has changed

    $\Rightarrow$ Build a new model of their strategy, based on their recent behavior

- ♢ DBS uses the model to decide what move to make next
  - game-tree search, using the model to predict the other player's moves

DBS's strategy model

- Here's what DBS's model of other player's strategy looks like
  - Four probabilities of the following form:

    Pr[they'll choose C | my last move, their last move]

- This can correctly represent simple strategies, but not complicated ones:
  - Can correctly represent TFT:
    - Pr[C | C, C] = 1
    - Pr[C | C, D] = 1
    - Pr[C | D, C] = 0
    - Pr[C | D, D] = 0
  - Can't correctly represent TFTT:
    - TFTT's next move depends on the last *two* iterations

- Why is this OK?

20th Anniversary IPD Competition

- Category 2:
  - Iterated Prisoner's Dilemma with Noise
  - 165 programs

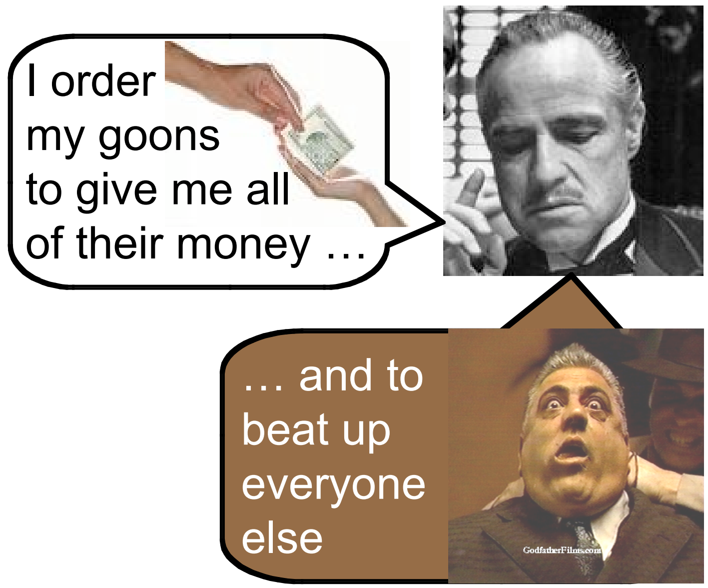
Master-and-slaves strategies. Each of them had 19 other programs feeding points to it.

Instances of DBS

Rank	Program	Avg. score
1	BWIN	433.8
2	IMM01	414.1
3	DBSz	408.0
4	DBSy	408.0
5	DBSpl	407.5
6	DBSx	406.6
7	DBSf	402.0
8	DBStft	401.8
9	DBSd	400.9
10	lowESTFT_classic	397.2
11	TFTIm	397.0
12	Mod	396.9
13	TFTIz	395.5
14	TFTIc	393.7
15	DBSe	393.7
16	TTFT	393.4
17	TFTIa	393.3
18	TFTIb	393.1
19	TFTIx	393.0
20	mediumESTFT_classic	392.9

How BWIN and IMM01 worked

- Each participant could enter up to 20 agents
  - Some of them wrote agents that could recognize each other by exchanging coded sequences of C and D moves
- Once they recognized each other, the 20 agents worked as a team:
  - 1 master and 19 slaves
- When a slave plays with its master, it cooperates and the master defects
  - ⇒ master gets 5 points, slave gets nothing



- When a slave plays with an agent not in its team, the slave defects
  - ⇒ the other agent gets ≤ 1 points
- BWIN and IMM01 were the masters of two master-and-slave teams

Analysis

- Average score of each master-slaves team was much lower than DBS's
  - If BWIN and IMM01 each had ≤ 10 slaves, DBS would have placed 1st
  - If BWIN and IMM01 had no slaves, they would have done badly

- Unlike BWIN and IMM01, DBS had no slaves
  - None of the DBS programs even knew that the others were there

- DBS established cooperations with many other agents
  - Could do this despite the noise, because it could filter out the noise

Homework (only 30 points this time)

1. Do Problem 17.16 in the book.

2. Consider the following game:

	H	D
H	–2, –2	6, 0
D	0, 6	3, 3

(a) Find all dominant strategy equilibria. If there are none, explain why.

(b) Find all Nash equilibria.

3. For each of the following strategies, can DBS's strategy model represent it correctly? If so, write the representation. If not, explain why not.

(a) AllC.

(b) Random (choose C or D at random, with 0.5 probability for each).

(c) GTFT (If the other player's last move was C, choose C. Otherwise, choose D with probability 0.9, and C with probability 0.1.)

(d) Pavlov.

(e) Grim. (Think carefully about this one, it's tricky.)