Prisoners’ Dilemma: The Story

- Invented by Merrill Flood & Melvin Dresher in 1950s
- Studied in game theory, economics, political science
- Two criminals have been caught
- They cannot communicate with each other
- If both confess, they will each get 10 years
- If one confesses and accuses other:
  - confessor goes free
  - accused gets 20 years
- If neither confesses, they will both get 1 year on a lesser charge

Prisoners’ Dilemma
Payoff Matrix

		Bob	
		cooperate	defect
Ann	cooperate	–1, –1	–20, 0
	defect	0, –20	–10, –10

- defect = confess, cooperate = don’t
- payoffs < 0 because punishments (losses)

Ann's "Rational" Analysis (Dominant Strategy)

		Bob	
		cooperate	defect
Ann	cooperate	–1, –1	–20, 0
	defect	0, –20	–10, –10

- if cooperates, may get 20 years
- if defects, may get 10 years
- ∴, best to defect

Bob's "Rational" Analysis (Dominant Strategy)

		Bob	
		cooperate	defect
Ann	cooperate	–1, –1	–20, 0
	defect	0, –20	–10, –10

- if he cooperates, may get 20 years
- if he defects, may get 10 years
- ∴, best to defect

Suboptimal Result of "Rational" Analysis

		Bob	
		cooperate	defect
Ann	cooperate		–20, 0
	defect	0, –20	–10, –10

- each acts individually rationally ⇒ get 10 years (dominant strategy equilibrium)
- "irrationally" decide to cooperate ⇒ only 1 year

Summary

- Individually rational actions lead to a result that all agree is less desirable
- In such a situation you cannot act unilaterally in your own best interest
- Just one example of a (game-theoretic) *dilemma*
- Can there be a situation in which it would make sense to cooperate unilaterally?
  – **Yes**, if the players can expect to interact again in the future

The Iterated Prisoners' Dilemma

and Robert Axelrod's Experiments

Iterated Game

- In simulation, the endpoint of the game is unknown to the players, making it essentially an infinitely iterated game
- Each player has a memory of the previous three or more) rounds on which to base his strategy
- Strategies are deterministic - for a given history *h* players will always make the same move
- With 4 possible configurations in each round and a history of 3, each strategy is comprised of 4^3 = 64 moves

Axelrod's Experiments (1984)

- Intuitively, expectation of future encounters may affect rationality of defection
- Various programs compete for 200 rounds
  - encounters each other and self
- Each program can remember:
  - its own past actions
  - its competitors' past actions
- 14 programs submitted for first experiment

IPD Payoff Matrix

		B	
		cooperate	defect
A	cooperate	3, 3	0, 5
	defect	5, 0	1, 1

Indefinite Number of Future Encounters

- Cooperation depends on expectation of indefinite number of future encounters
- Suppose a known finite number of encounters:
  - No reason to C on last encounter
  - Since expect D on last, no reason to C on next to last
  - And so forth: there is no reason to C at all

Analysis of Some Simple Strategies

- Three simple strategies:
  - **ALL-D**: always defect
  - **ALL-C**: always cooperate
  - **RAND**: randomly cooperate/defect
- Effectiveness depends on environment
  - **ALL-D** optimizes local (individual) fitness
  - **ALL-C** optimizes global (population) fitness
  - **RAND** compromises

Expected Scores

⇓ playing ⇒	ALL-C	RAND	ALL-D	Average
ALL-C	3.0	1.5	0.0	1.5
RAND	4.0	2.25	0.5	2.25
ALL-D	5.0	3.0	1.0	3.0

Result of Axelrod's Experiments

- Winner is Rapoport's **TFT** (Tit-for-Tat)
  - cooperate on first encounter
  - reply in kind on succeeding encounters
- Second experiment:
  - 62 programs
  - all known **TFT** was previous winner
  - **TFT** wins again

Expected Scores

⇓ playing ⇒	ALL-C	RAND	ALL-D	TFT	Avg
ALL-C	3.0	1.5	0.0	3.0	1.875
RAND	4.0	2.25	0.5	2.25	2.25
ALL-D	5.0	3.0	1.0	$1+4/N$	2.5+
TFT	3.0	2.25	$1-1/N$	3.0	2.3125–

N = #encounters

The Genetic Algorithm

The Model

- Darwinian Survival of the Fittest
- Genetic representation of entities
- Fitness function
- Select most fit individuals to reproduce
- Mutate
- Traits of most fit will be passed on
- Over time, the population will evolve to be more fit, optimal

GA’s and the Prisoner’s Dilemma

- Population: 100 individuals
- Chromosome: 64-bit string representing a strategy
- A bit: a choice between the Cooperate or Defect move played for a specific configuration of past moves

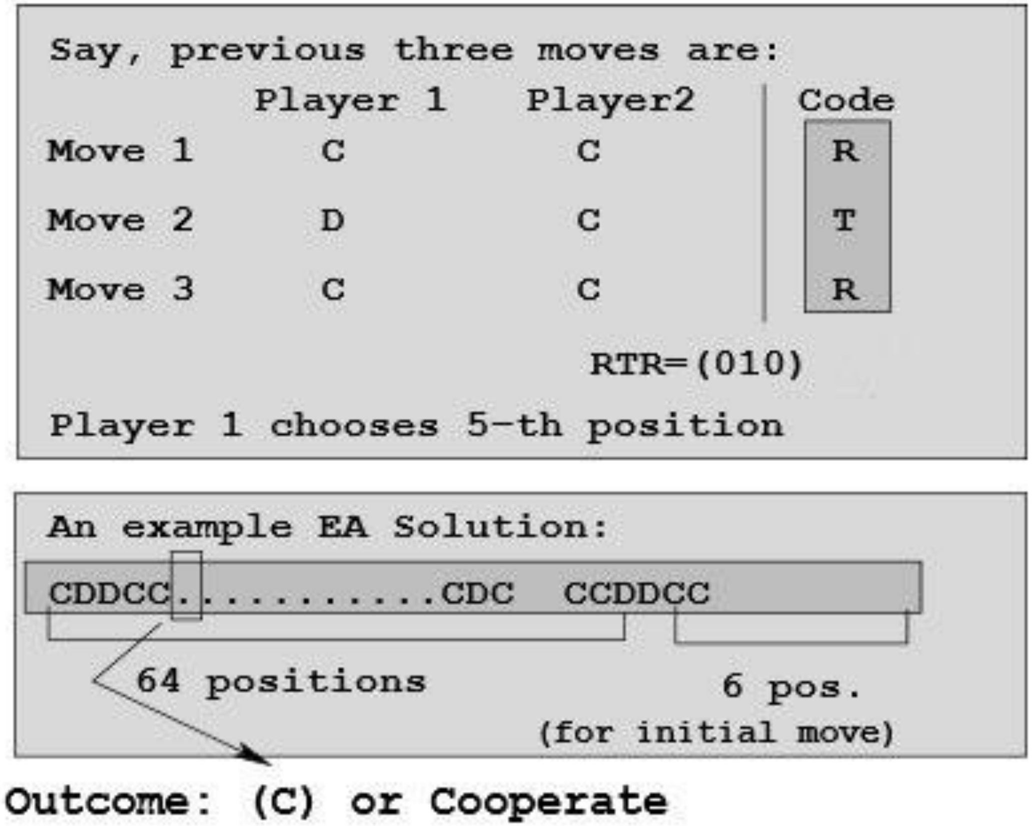
- Individuals = strategies
- How to encode a strategy by a string?

- Let memory depth of previous moves=1

Fix a canonical order of cases:

	A	B
– Case 0:	C	C
– Case 1:	D	C
– Case 2:	C	D
– Case 3:	D	D

e.g. strategy encoding (for A): ‘CDCD’



Encoding a strategy for IPD.

String Position	Represented History	Move
0	CCCCCC	C
1	CCCCCD	D
2	CCCCDC	D
3	CCCCDD	D
4	CCCDCC	C
5	CCCDCD	C
6	CCCDDC	C
7	CCCDDD	D
8	CCDCCC	C
9	CCDCCD	D
10	CCDCDC	D
11	CCDCDD	D
12	CCDDCC	D
13	CCDDCD	D
14	CCDDDC	C
15	CCDDDD	C
16	CDCCCC	C
17	CDCCCD	C
18	CDCCDC	D
19	CDCCDD	C
20	CDCDCC	C
21	CDCDCD	D
22	CDCDDC	D
23	CDCDDD	C
24	CDDCCC	C
25	CDDCCD	C
26	CDDCDC	D
27	CDDCDD	C
28	CDDDCC	D
29	CDDDCD	C
30	CDDDDC	C
31	CDDDDD	D

String Position	Represented History	Move
32	DCCCCC	D
33	DCCCCD	C
34	DCCCDC	D
35	DCCCDD	D
36	DCCDCC	C
37	DCCDCD	C
38	DCCDDC	D
39	DCCDDD	D
40	DCDCCC	D
41	DCDCCD	C
42	DCDCDC	C
43	DCDCDD	C
44	DCDDCC	C
45	DCDDCD	D
46	DCDDDC	D
47	DCDDDD	C
48	DDCCCC	C
49	DDCCCD	D
50	DDCCDC	C
51	DDCCDD	C
52	DDCDCC	C
53	DDCDCD	C
54	DDCDDC	D
55	DDCDDD	D
56	DDDCCC	C
57	DDDCCD	C
58	DDDCDC	D
59	DDDCDD	D
60	DDDDCC	C
61	DDDDCD	C
62	DDDDDC	D
63	DDDDDD	C

GA’s and PD

- Fitness: Each player competes against every other for 64 consecutive rounds, and a cumulative score is maintained
- Selection: Roulette Wheel selection
- Reproduction: Random point crossover with replacement
- Mutation rate 0.001

- **Generations: 1,000 generations**

Tit for Tat (TFT)

- The action chosen is based on the opponent's last move.
  - On the first turn, the previous move cannot be known, so always cooperate on the first move.
  - Thereafter, always choose the opponent's last move as your next move.

- Key Points of Tit for Tat

  - Nice; it cooperates on the first move.
  - Regulatory; it punishes defection with defection.
  - Forgiving; it continues cooperation after cooperation by the opponent.
  - Clear; it is easy for opponent to guess the next move, so mutual benefit is easier to attain.

Tit for Two Tat (TF2T)

- Same as Tit for Tat, but requires two consecutive defections for a defection to be returned.
  - Cooperate on the first two moves.
  - If the opponent defects twice in a row, choose defection as the next move.

- Key Points of Tit for Two Tat
  - When defection is the opponent’s first move, this strategy outperforms Tit for Tat
  - Cooperating after the first defection causes the opponent to cooperate also. Thus, in the long run, both players benefit more points.

Suspicious Tit for Tat (STFT)

- Always defect on the first move.
- Thereafter, replicate opponent’s last move.

- Key Points of Suspicious Tit for Tat
  - If the opponent’s first move is defection, this strategy outperforms Tit for Tat
  - However, it is generally worse than Tit for Tat.
    - The first move is inconsequential compared to getting stuck in an infinite defection loop.

Free Rider (ALLD)

- Always choose to defect no matter what the opponent’s last turn was.

- This is a dominant strategy against an opponent that has a tendency to cooperate.

Always Cooperate (ALLC)

- Always choose to cooperate no matter what the opponent’s last turn was.

- This strategy can be terribly abused by the Free Rider Strategy.
  - Or even a strategy that tends towards defection.

- Use and compare different optimization methods or combination of methods (GA, hill climbing, tabu search, exhaustive search…) to find the best possible strategies.

- Try different parameters of the optimization methods (like population size or mutation probability for GA)

- Try different memory depth: 3, 4, 5…

- Try to compete against human designed strategies.

- Compared the obtained strategies with TFT, TF2T or STFT