Classification of Dilemmas

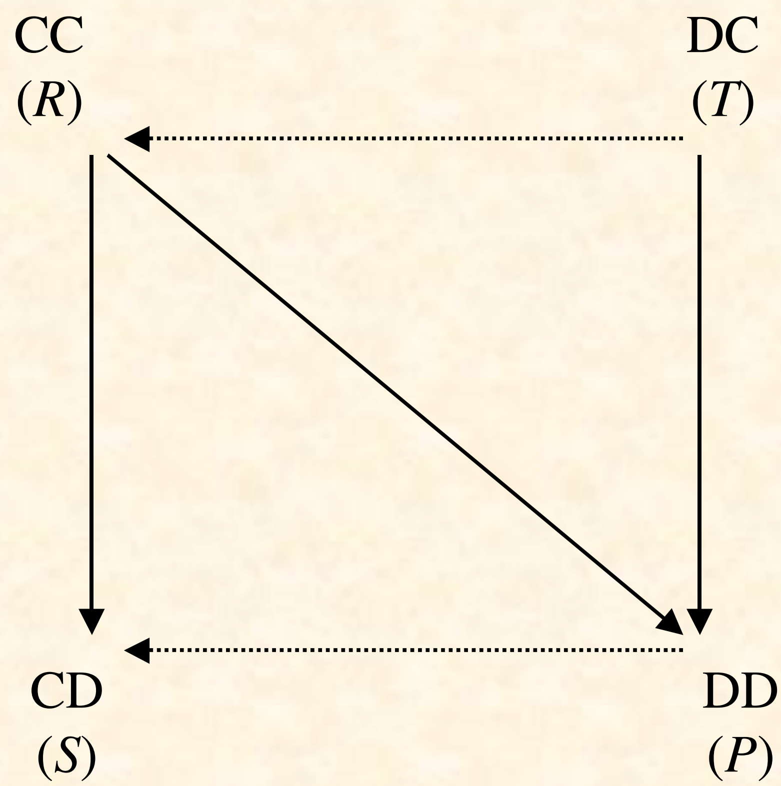
General Payoff Matrix

		Bob	
		cooperate	defect
Ann	cooperate	CC (*R*)	CD (*S*)
	defect	DC (*T*)	DD (*P*)

General Conditions for a Dilemma

- You always benefit if the other cooperates:
  - CC > CD *and* DC > DD
- You sometimes benefit from defecting:
  - DC > CC *or* DD > CD
- Mutual coop. is preferable to mut. def.
  - CC > DD
- Consider relative size of CC, CD, DC, DD
  - think of as permutations of *R*, *S*, *T*, *P*
  - only three result in dilemmas

Three Possible Orders



The three dilemmas: *TRSP*, *RTPS*, *TRPS*

The Three Dilemmas

- Chicken (*TRSP*)
  - DC > CC > CD > DD
  - characterized by mutual defection being worst
- Stag Hunt (*RTPS*)
  - CC > DC > DD > CD
  - better to cooperate with cooperator
- Prisoners' Dilemma (*TRPS*)
  - DC > CC > DD > CD
  - better to defect on cooperator

The Iterated Prisoners' Dilemma

and Robert Axelrod's Experiments

Assumptions

- No mechanism for enforceable threats or commitments
- No way to foresee a player's move
- No way to eliminate other player or avoid interaction
- No way to change other player's payoffs
- Communication only through direct interaction

Axelrod's Experiments

- 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

N.B. Unless DC + CD < 2 CC (i.e. $T + S < 2R$), can win by alternating defection/cooperation

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	2.0	0.5	2.166…
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 know **TFT** was previous winner
  - **TFT** wins again

Characteristics of Successful Strategies

- *Don't be envious*
  - at best **TFT** ties other strategies
- *Be nice*
  - i.e. don't be first to defect
- *Reciprocate*
  - reward cooperation, punish defection
- *Don't be too clever*
  - sophisticated strategies may be unpredictable & look random