Chapter 4: Correlated Convention

Outline

A. The puzzle: the curse of *symmetry*

How can we motivate a particular choice when there is no "sufficient reason" for preferring one of two or more indistinguishable outcomes?

B. Classical game theory and symmetry

Game of Chicken: two symmetric (pure) equilibria and a mixed equilibrium

$\Rightarrow$ Best classical solution (the mixed strategy equilibrium) is sub-optimal

Better solution: find an optimal joint strategy

C. How nature breaks symmetries

- Step 1: stable equilibria become unstable
- Step 2: slight perturbations push the system out of equilibrium

D. Correlated strategies

Joint strategies: specify a plan for each agent (so that optimal outcomes are obtained)

Randomized joint strategies: randomize between pure joint strategies

Correlated equilibrium: joint strategy such that no player has any incentive to depart.

Mechanisms for achieving correlated equilibrium:

1) external random process

2) learning/teaching behaviour

3) local pairing off

E. Examples

- signals
- property

A. Puzzle: the curse of symmetry

Two situations are symmetrical if there is no basis whatsoever for distinguishing them in any respect relevant to your choice (i.e., to probability or to utility).

Background: **Principle of Sufficient Reason**.

> **Leibniz:** God's problem of creating the best possible world (only solvable if there is a unique best one!)

In decision theory: How does a rational agent decide between two equally good options (indistinguishable in any important respect)?

> Ghazali's position: there must be some mechanism for making the decision available to any rational agent. But what? (Which randomizing mechanism to use?)

In game theory: What should one do in a mixed equilibrium? (Fixed payoff, assuming partner plays her equilibrium strategy)

B. Classical game theory and symmetry

Chicken

	Don't swerve	Swerve
Don't swerve	-100,-100	50,-10
Swerve	-10,50	0, 0

Two pure equilibria: (D,S); (S,D).
One symmetrical mixed equilibrium. Both play Don't swerve 5/14, Swerve 9/14.
Expected payoff (for both) of the mixed equilibrium: -50/14, or -3 4/7.
Sub-optimal: Both would do better if both played Swerve.

Hawks and Doves

	Hawk	Dove
Hawk	-25,-25	50,0
Dove	0,50	15,15

Basis: Winning = 50, Wasting time = -10, losing =0, injured/killed = -100

Same structure as Chicken.
Two pure equilibria: (H, D); (D, H).
One symmetrical mixed equilibrium. Both play Hawk 7/12, Dove 5/12.
Expected payoff of the mixed equilibrium: 25/4, or 6¼.
Sub-optimal: both would do better if both played Dove.

Classical Solution: Since there is no way to choose between the pure equilibria, select the mixed equilibrium.
Evolutionary Dynamics: Also selects the mixed equilibrium. The pure equilibria are meaningless, and in any case 100% dove and 100% hawk are both unstable.

The fact that the equilibrium is non-optimal is not a problem for evolutionary approach (which does not pretend to achieve optimality), but is a problem for the rationality-based approach.

C. How nature breaks symmetries

Ex: Compressed column.

With no load, a stable equilibrium: movement to left or right is corrected.

Under a load, this becomes an unstable equilibrium: movement to left or right becomes exaggerated and the column breaks.

In general:

1. A stable equilibrium becomes unstable under pressure.
2. Slight perturbations push the system out of the equilibrium to a solution.

D. Correlated strategies

In the game of Chicken, players can increase their utility by tossing a coin ahead of time. The loser agrees to swerve, the winner will not swerve. This increases the expected payoff for each player to

$$\tfrac{1}{2}(50) + \tfrac{1}{2}(-10) = 20$$

or, in the Hawk-Dove game, to

$$\tfrac{1}{2}(50) + \tfrac{1}{2}(0) = 25.$$

Joint strategy: a strategy that specifies choices for each player
Coordination game: one that permits joint strategies
Randomized joint strategy: randomizes between pure joint strategies
Correlated equilibrium: a joint strategy that attains an equilibrium. Neither player has any incentive to depart.

Examples of conventions:

- Drive on the right side of the road
- Yield to the car on the right
- In situations like Hawk-Dove, abide by the result of a coin toss

Link to convention:

A convention can be viewed as a *solution* to a coordination game.
More explicitly: a convention is a correlated equilibrium.

Upshot:
From the normative perspective, conventions don't "solve" anything, because there is still a problem of symmetry (which convention: left or right? Heads or tails?). Reason cannot solve this problem.

From the evolutionary perspective, conventions do solve a problem of non-optimality: where one or more conventions might evolve, any one is equally good, but only one will evolve **in fact**. "Nature" solves the problem.

<u>Mechanisms for achieving a correlated equilibrium (i.e., a convention)</u>

1) *Mutation*: emergence of new strategies plus an external random process

External random process: fair coin toss

New strategy: DH (Dove-Hawk)

Swerve just in case coin comes up heads.

DH does as well as anybody against the general population (6¼), and better against itself (25), so this group will take over if more than one of them emerge.

Significance: The emergence of the new strategy makes the 5/12 Dove, 7/12 Hawk equilibrium **unstable**. Just a small number of DH will take over the population.

(N.B. If HD mutants arise, who swerve just in case the coin comes up tails, then whichever of DH and HD is the larger group will take over. If the same size, then we still have a symmetry problem until some mutation or population fluctuation due to pairing off gives one group the upper hand.)

2) *Learning*: belief revision plus an external random process

External random process: fair coin toss

Belief revision: conditional probabilities about what the other players will do if the coin comes up heads/tails, revised according to some inductive rule (e.g., Bayes' Theorem)

Each player starts with some expectation (conditional probability) about what the others do, and acts to maximize expected payoffs given the result of the random process. This may be the default, i.e., the equilibrium probability.

Based on the other players' actions, the probabilities are updated for the next interaction.

Emergence of correlated convention:

Even if the players intially choose their actions by randomizing, conventions quickly become established.

This is easiest to see in the case of two players:

- If both do different things on turn 1 (A plays Hawk, B plays Dove) after a toss of Heads, then the corresponding slight revision of the conditional probabilities entails that each maximizes expected payoff in future by making this the rule for future Heads. [Ditto for Tails, and the choice of actions here is independent: players need not do the opposite of what they did on Heads.]
- If both do the same thing on turn 1 (A and B both play Hawk) after a toss of Heads, then both do the same *opposite* thing on the next toss of Heads (both play Dove) and this exactly cancels out the first revision. [Ditto for Tails, and the choice of actions is independent.] So there is once again a chance for different behaviour to emerge on the next round.

Very quickly, one of four conventions becomes established:

a) Heads: A plays Hawk and B plays Dove; Tails: reverse.
b) Heads: A plays Dove and B plays Hawk; Tails: reverse.
c) No matter what comes up, A plays Hawk and B plays Dove.
d) No matter what comes up, B plays Hawk and A plays Dove.

3) *Localization*: local or 'cultural' conventions

Problem: A pair of agents in repeated interactions will end up at one of the above four equilibria, but with equal probability for each.

So: in a very large population, we should expect all four strategies to emerge, yielding non-optimal outcomes. These are Symmetry problems at a higher level.

Focus just on the HD and DH equilibria. And now imagine that the coin has players' names on both sides, rather than Heads or Tails.

HD: player whose name comes up does H, other does D
DH: player whose name comes up does D, other does H

Two solutions:

i) *Small populations*. Slight asymmetries in the balance can tip things in favour of one of the two correlated equilibria.

ii) *Large populations*. If pairings are restricted to local match-ups, sub-populations will develop different conventions. The overall proportion may be 50/50.

E. Examples

1) *Signals and Rules*

Scenario 1: new traffic lights

- each light has two colours: orange and purple
- town officials leave before indicating which means “stop” and which means “go”

For each driver, the colour is essentially a random process.

Even without external directions, the townspeople can move to one of the two correlated equilibria: “go” on orange, or “go” on purple. We need only assume:

a) learning from observation of what most do;

b) a slight asymmetry emerges at some point in favour of one of these rules.

Scenario 2: intersections with no lights or stop signs

“Rule of the right”: driver on the right goes first.

“Rule of the left”: driver on the left goes first.

What actually emerged is the rule of the right.

2) *Property*

Empirical studies:

- people value something more if it is their property
- animal species use “ownership” to avoid costly fights:

  Hawk if owner; Dove if intruder

- in effect, ownership is the signal that breaks the symmetry.

The other correlated equilibrium is Dove if owner, Hawk if intruder.

- Stable, but rarely found.
- Explanation: even a slight asymmetry in fighting ability in favour of the owner, or in the value of the territory to the owner, will push things to the usual equilibrium.