Introduction to Signaling:
Lewis on Convention and Signaling Games

Models and Simulations in Philosophy
November 24th, 2013

First Class: Two Platonic Puzzles: Morality and Meaning

Each puzzle really consisted of three different questions:

First Class: Two Platonic Puzzles: Morality and Meaning

Each puzzle really consisted of three different questions:

- Definition: What is morality? What makes an assertion meaningful?

First Class: Two Platonic Puzzles: Morality and Meaning

Each puzzle really consisted of three different questions:

- Definition: What is morality? What makes an assertion meaningful?
- Evolution: How did morality and/or meaningful language evolve?

First Class: Two Platonic Puzzles: Morality and Meaning

Each puzzle really consisted of three different questions:

- Definition: What is morality? What makes an assertion meaningful?
- Evolution: How did morality and/or meaningful language evolve?
- Stability: Why does moral behavior persist? Why is language stable?

Review

Class Thus Far: The first puzzle about morality

We've seen tentative answers to three questions for morality . . .

Definition: [Gauthier, 1967] considered a definition of morality that characterized moral actions as those belonging to particular strategic profiles in games.

Evolution: [Alexander, 2007]'s models show boundedly -rational agents might learn to cooperate, trust, and play fair in certain games.

Stability: The learning algorithms concerned by Alexander also promoted stability agents imitated one another: once a population is full of cooperators, there are no defectors to imitate.

Preview

This Month and Next: The second puzzle - Meaning

Definition: Lewis [2008]'s definitions of **convention** and **signaling system** are an attempt to characterize, in game theoretic terms, when individuals' signals have acquired meaning.

Evolution: [Skyrms, 2004] and [Skyrms, 2010] will provide models that show how agents might learn to play a signaling system.

Stability: Both Lewis' definition and Skyrms' models provide explanations as to why conventional meanings might persist.

Today's Class

Today: Lewis' analysis of coordination, convention, and signaling system.

Today's Class

Lewis argues that language is a type of convention, and

Convention is a type of coordination equilibria.

So we'll start with coordination equilibria and work backwards.

Outline

Outline

Outline

Outline

Outline

Outline

Common vs. Conflict of Interest

What is a coordination game?

Intuitively, we've seen games in which players have "conflicting" interests and ones in which they have "common" or "similar" interests.

Common vs. Conflict of Interest

- Conflict of Interest: Prisoners' Dilemma
- Common/Similar Interests: Stag Hunt
- Both conflict and common interest: Hawk-Dove

What is a coordination game?

There have been several different attempts to characterize "coordination games", in which players have similar interests. Here are two from economics:

A coordination game is a game with multiple Nash equilibria in which

- Definition [Cooper et al., 1990]: Players have identical preferences over all outcomes (i.e. strategic profiles) of the game.
- Definition [Colman, 1997]: Players have identical preferences over all Nash equilibia.

What is a coordination game?

Both definitions fail to count Bach-or-Stravinsky as a coordination game:

	Bach	Stravinsky
Bach	1,2	0,0
Stravinsky	0,0	2,1

What is a coordination game?

Both definitions fail to count Bach-or-Stravinsky as a coordination game:

	Bach	Stravinsky
Bach	1,2	0,0
Stravinsky	0,0	2,1

Intuitively, this seems like precisely the type of game in which agents are "coordinating" their behaviors.

Lewis' Definition

What is Lewis' definition of a coordination equilibrium?

Nash Equilibria vs. Coordination Equilibria

Contrast: A strategic profile s is a Nash equilibrium if no player strictly prefers to change **her own** action if other players actions are held fixed.

Nash Equilibria vs. Coordination Equilibria

Contrast: A strategic profile s is a Nash equilibrium if no player strictly prefers to change **her own** action if other players actions are held fixed.

Lewis: A strategic profile s is a coordination equilibrium if if no player strictly prefers that **some** player (potentially himself) changes actions, if all others' actions are held fixed.

Nash Equilibria vs. Coordination Equilibria

So a coordination equlibrium is a special type of Nash equilibrium.

Let's consider some examples.

Prisoners' Dilemma: Coordination Equilibria

	Cooperate	Defect
Cooperate	2,2	0,3
Defect	3,0	1,1

Question: Which strategic profiles (if any) are coordination equilibria in a one-shot prisoner's dilemma?

Prisoners' Dilemma: Coordination Equilibria

	Cooperate	Defect
Cooperate	2,2	0,3
Defect	3,0	1,1

Question: Which strategic profiles (if any) are coordination equilibria in a one-shot prisoner's dilemma?

Answer: None. Coordination equilibria are Nash, and so only $\langle D, D \rangle$ is a candidate. But each player would prefer that the other cooperates if she defects.

Hawk-Dove: Coordination Equilibria

	Dove	Hawk
Dove	2,2	1,3
Hawk	3,1	0,0

Question: Which strategic profiles (if any) are coordination equilibria in Hawk-Dove?

Hawk-Dove: Coordination Equilibria

	Dove	Hawk
Dove	2,2	1,3
Hawk	3,1	0,0

Question: Which strategic profiles (if any) are coordination equilibria in Hawk-Dove?

Answer: None. Coordination equilibria are Nash, and so only $\langle H, D \rangle$ and $\langle D, H \rangle$ are candidates. But the Dove player would prefer the Hawk player switches to Dove.

Bach-or-Stravinsky: Coordination Equilibria

	Bach	Stravinsky
Dove	1,2	0,0
Stravinsky	0,0	2,1

Question: Which strategic profiles (if any) are coordination equilibria in a Stag-Hunt?

Bach-or-Stravinsky: Coordination Equilibria

	Bach	Stravinsky
Dove	1,2	0,0
Stravinsky	0,0	2,1

Question: Which strategic profiles (if any) are coordination equilibria in a Stag-Hunt?

Answer: Both Nash. No players would prefer to change either his action or his opponent's action.

Stag Hunt: Coordination Equilibria

	Stag	Hare
Stag	2,2	0,1
Hare	1,0	1,1

Question: Which strategic profiles (if any) are coordination equilibria in a Stag-Hunt?

Stag Hunt: Coordination Equilibria

	Stag	Hare
Stag	2,2	0,1
Hare	1,0	1,1

Question: Which strategic profiles (if any) are coordination equilibria in a Stag-Hunt?

Answer: Both Nash. If we change "strict" to "weak" in Lewis' definition, only the stag equilibria would be a coordination one.

Coordination Equilibria

Moral: There are coordination equilibria (in Lewis' sense) in precisely those games that game-theorists typically called "coordination games."

So one could tentatively define a coordination game as one containing a coordination equilibrium.

Class Activity

Class Activity:

- Lewis provides several examples of conventions at the outset of the book.
- In small groups, pick three examples from list.
- Write a game matrix for a two-person game that you think represents the payoffs of interactions among individuals in the example.
- Are there any coordination equilibria in the example? If so, are they the conventions?
- If the example involves interactions with more than two individuals, try to generalize your payoff matrix to three or more individuals. What are the coordination equilibria? Are they the conventions?

Lewis' Rough Definition: A convention is a regularity R in the behavior of members of population P when they are in recurrent situation S if, in any instance S among members of P,

- Everyone conforms to R
- Everyone expects everyone else to conform to R
- Everyone prefers that everyone conform to R on condition that others do in S, since S is a coordination problem and uniform conformity to R is a proper coordination equilibrium in S.

Lewis' Rough Definition: A convention is a regularity R in the behavior of members of population P when they are in recurrent situation S if, in any instance S among members of P,

- Everyone conforms to R
- Everyone expects everyone else to conform to R
- Everyone prefers that everyone conform to R on condition that others do in S, since S is a coordination problem and uniform conformity to R is a proper coordination equilibrium in S.

We'll come back to modifications to this definition in a moment ...

This is a neat definition, but we haven't yet discussed why:

- Convention should be analyzed in terms of coordination equilibria.
- Language should be analyzed in terms of convention.

This is a neat definition, but we haven't yet discussed why:

- Convention should be analyzed in terms of coordination equilibria.
- Language should be analyzed in terms of convention.

Let's take these questions one at a time.

Coordination and Convention

Why are coordination equilibria good candidates for representing conventions?

Primary Answer: The definition seems to fit a lot of the examples.

Primary Answer: The definition seems to fit a lot of the examples.

But there are also general reasons to think it will fit more examples.

Coordination and Convention

Conventions are coordinated behaviors:

- A convention is typically a behavior that is common and widespread: drink from the water glass on the right, speak German, drive on the right hand side of the road.

Coordination and Convention

Conventions are coordinated behaviors:

- A convention is typically a behavior that is common and widespread: drink from the water glass on the right, speak German, drive on the right hand side of the road.
- So some behavior is conventional only if every individual employs the same action (in similar circumstances).

Coordination and Convention

Conventions are coordinated behaviors:

- A convention is typically a behavior that is common and widespread: drink from the water glass on the right, speak German, drive on the right hand side of the road.
- So some behavior is conventional only if every individual employs the same action (in similar circumstances).
- Hence, when individuals follow a convention, their actions will be "coordinated."

Coordination and Convention

Conventions are equilibria:

- A convention would be extremely unstable if a substantial number of individuals preferred to deviate from it.
  - E.g. Suppose a sizable population preferred to wear jeans at black-tie events (regardless of what others think). Then there would be no convention against wearing jeans.
- So a convention has to be a Nash equilibrium, or at least a strategy profile in which an overwhelming number of individuals would not prefer to deviate (even with a small number of deviators).

Coordination and Convention

Conventions are coordination equilibria:

- Many conventions are such that individuals do not strictly prefer others to deviate.
- E.g., I strictly prefer that *you* continue to drive on the right side of the road.
- E.g., I couldn't care less if one of you decided to break the convention of wearing matching clothes.

Coordination and Convention

There needs to be more than one possible convention:

Coordination and Convention

There needs to be more than one possible convention:

- Something can count as "conventional" only if another convention could have been adopted.

Coordination and Convention

There needs to be more than one possible convention:

- Something can count as "conventional" only if another convention could have been adopted.
- Lewis wants to define a convention in terms of some type of strategic profile that may not be unique.

Coordination and Convention

There needs to be more than one possible convention:

- Something can count as "conventional" only if another convention could have been adopted.
- Lewis wants to define a convention in terms of some type of strategic profile that may not be unique.
- The Stag Hunt that there may be multiple coordination equilibria in a game: this is what Lewis wants!

But the non-uniqueness of coordination equilibria raises the other two other questions about convention at the outset:

- Evolution:: How does a convention come about? That is, why does one convention rather than another come to be?
- Stability: How is it maintained? That is, why does one convention persist if another could be adopted?

Stability of a Convention

Let's start with the stability question.

Nash Equilibria and Stability

Nash equilibria, in general, are not necessarily stable.

- Suppose we will play a stag hunt.
- In the past, you and I have both played stag (creating a Nash equilibrium), but,
- I have good reason to suspect you will play hare next.

What should I do?

Moral: If I don't expect you to play part of a Nash equilibrium, it might be rational for me to do something else.

Suppose, now, I expect you to play Stag.

What should I do?

In general, suppose we are playing some game and I expect you to play your half of a Nash equilibrium.

What should I do?

In general, suppose we are playing some game and I expect you to play your half of a Nash equilibrium.

What should I do?

Answer: Play the other half. By definition, a Nash equilibrium is one in which each player performs a best response to all others.

Definition of Convention

Lewis' Rough Definition: A convention is a regularity R in the behavior of members of population P when they are in recurrent situation S if, in any instance S among members of P,

- Everyone conforms to R
- Everyone expects everyone else to conform to R
- Everyone prefers that everyone conform to R on condition that others do in S, since S is a coordination problem and uniform conformity to R is a proper coordination equilibrium in S.

Evolution of Convention

But how do conventions come about?

In economists' terms, how is the equilibrium **selected**?

Evolution of Convention

> *Agreement*, *salience*, or *precedent*, we have seen, can solve a coordination problem by producing a system of concordant *first- and higher-order mutual expectations*.

[Lewis, 2008], pp. 52.

Higher-Order Expectations

Watch the "poison scene" from the Princess Bride.

Iterated Elimination of Dominated Strategies

Remember this game from a few weeks back?

	Left	Center	Right
Top	0,2	3,1	2,3
Middle	1,4	2,2	4,1
Bottom	2,1	4,4	3,2

Iterated Elimination of Dominated Strategies

Remember this game from a few weeks back?

	Left	Center	Right
Top	0,2	3,1	2,3
Middle	1,4	2,2	4,1
Bottom	2,1	4,4	3,2

I asked you questions like

- Suppose column is rational, and
- Suppose Row knows that Column is rational
- **And** Row knows that Column knows that Row is rational.
- Then what outcomes will Row consider?

Moral: Higher-order knowledge helped agents to select an equilibrium.

In previous classes, our theorems only assumed common-knowledge of the rationality of players,

Higher Order Knowledge

Moral: Higher-order knowledge helped agents to select an equilibrium.

In previous classes, our theorems only assumed common-knowledge of the rationality of players,

But there might also be common knowledge of facts about which actions (several of which might be rational) that agents may choose.

Definition of Convention

Lewis' Definition: A convention is a regularity R in the behavior of members of population P when they are in recurrent situation S if, in any instance S among members of P, it is common knowledge that

- Everyone conforms to R
- Everyone expects everyone else to conform to R
- Everyone prefers that everyone conform to R on condition that others do in S, since S is a coordination problem and uniform conformity to R is a proper coordination equilibrium in S.

Higher Order Knowledge

Where does common knowledge come from?

Lewis gives three sources:

- Agreement
- Salience
- Precedent

Higher Order Knowledge

Skyrms' criticism:

- Agreement:

Higher Order Knowledge

Skyrms' criticism:

- Agreement: Lewis admits this cannot be used to explain the origin of language.

Higher Order Knowledge

Skyrms' criticism:

- Agreement: Lewis admits this cannot be used to explain the origin of language.
- Precedent: Puts the cart before the horse

Higher Order Knowledge

Skyrms' criticism:

- Agreement: Lewis admits this cannot be used to explain the origin of language.
- Precedent: Puts the cart before the horse
- Salience: Are the coordination equilibria representing meaningful communication salient? ...

Outline

Outline

Outline

Outline

Outline

Outline

Example: Paul Revere's Ride

- Sender: Lighthouse attendant who observes if the British are coming by sea or land. He can show one lantern or two.

Example: Paul Revere's Ride

- Sender: Lighthouse attendant who observes if the British are coming by sea or land. He can show one lantern or two.
- Receiver: Paul Revere, who must ride to warn the colonists whether the British are coming by land or sea.

Example: Paul Revere's Ride

- Sender: Lighthouse attendant who observes if the British are coming by sea or land. He can show one lantern or two.
- Receiver: Paul Revere, who must ride to warn the colonists whether the British are coming by land or sea.
- The attendant and Revere both want to make the right warning, but

Example: Paul Revere's Ride

- Sender: Lighthouse attendant who observes if the British are coming by sea or land. He can show one lantern or two.
- Receiver: Paul Revere, who must ride to warn the colonists whether the British are coming by land or sea.
- The attendant and Revere both want to make the right warning, but
- They haven't agreed upon how many lanterns to use for each situation!

Signaling Games

Signaling games also have two players: sender and receiver.

- Sender observes some state of the world (e.g., by land or by sea).
- She then sends a signal to receiver (e.g., one lantern or two).
- The receiver then chooses an action (e.g. shout "by land" or shout "by sea" as you ride)
- The payoff that both receive depends upon the world and the receiver's action.

Formally, in cooperative signaling games:

- There are finite sets of states of the world W, a finite number of signals S, and finitely many actions A.
- Sender's actions: A function from worlds W to signals S.
- Receiver's actions: Functions from signals to acts.
- The payoffs to sender and receiver are the same, and they are determined by the state of the world and the action taken by the receiver.

Signaling System

- Assumption: For each state of the world w, there is at least one action a_w that is optimal.
- A signaling system is a pair of strategies $\langle f, g \rangle$ such that $g(f(w)) = a_w$ for all worlds w.
- **Question:** Is a signaling system a Nash equilibrium? A coordination equilibrium?

Programming Concept Today: Recursion

References I

Alexander, J. M. (2007). *The structural evolution of morality*. Cambridge University Press Cambridge.

Colman, A. M. (1997). Salience and focusing in pure coordination games. *Journal of Economic Methodology*, 4(1):61–81.

Cooper, R. W., DeJong, D. V., Forsythe, R., and Ross, T. W. (1990). Selection criteria in coordination games: Some experimental results. *The American Economic Review*, 80(1):218–233.

Gauthier, D. P. (1967). Morality and advantage. *The Philosophical Review*, 76(4):460–475.

Lewis, D. (2008). *Convention: A philosophical study*. Wiley-Blackwell.

Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge University Press.

Skyrms, B. (2010). *Signals: Evolution, learning, & information*. Oxford University Press.