Evolution of Cooperation and Trust

Models and Simulations in Philosophy
November 18th, 2013

Review and Today

- Last Week: Evolution of "Cooperation"
- This Week: Evolution of "Trust"

Review and Today

We used three types of models for analyzing whether cooperative behavior could emerge among rational agents.

Each model had a one-shot and repeated version.

Review

Review: Cooperation in Prisoners' Dilemmas

Classical Economics

- One-shot: Defecting is dominant.
- Finitely Repeated: Always defecting is the only strategy that survives iterated elimination of dominated strategies.

Review

Review: Cooperation in Prisoners' Dilemmas

Replicator Dynamics

- One-shot: Because defecting is dominant, it maximizes fitness. So defectors take over.
- Repeated: We'll start here today ...

Review

Review: Cooperation in Prisoners' Dilemmas

Network Models

- One-shot: Whether or not cooperation evolve
- Repeated: Same here, but Alexander doesn't pursue this question. Perhaps you could!

Outline

Repeated PDs and Replicator Dynamics

Does cooperative survive in the replicator dynamics if each agents plays a **repeated** prisoners' dilemma with a random member of the population?

Repeated PDs and Replicator Dynamics

Inspired by Axelrod [2006]'s PD tournament, [Alexander, 2007]

- Randomly assigns each agent in a large population a strategy for a repeated PD.
- Simulated the population's evolution according to the replicator dynamics.

Repeated PDs and Replicator Dynamics

Result:

- Strategies that sometimes cooperate and sometimes defect were the ones left after many stages of evolution.
- Commonly-discussed strategies, like Tit-for-Tat and Win-Stay-Lose-Shift, did not survive.

Moral: Cooperation can survive the replicator dynamics of repeated PDs.

Repeated PDs and Replicator Dynamics

That all sounds very straightforward, but we should think about two questions.

- What is a strategy in a repeated PD?
- What happens to a population when random mutation is possible?

Strategies in Repeated PDs

In games in which players act at different times, a strategy specifies what to do in light of past plays.

E.g., In chess, players obviously respond to each others' moves!

Strategies in Repeated PDs

Strategies in twice-repeated PDs:

- Opening move: Cooperate or Defect.
- A plan about how to response to the first play. There are four ways that first stage could have gone:
  - You and your opponent both cooperate
  - You and your opponent both defect
  - You defect and your opponent cooperates
  - You cooperate and your opponent defects

Strategies in Repeated PDs

Strategies in twice-repeated PDs:

- So you need a binary string of length $1 + 4 = 5$ to encode your strategy
  - First Digit: Opening Move
  - Four Digits: One for each possibility of the first stage
- So there are 32 possible strategies.

Strategies in Repeated PDs

Strategies in three-times-repeated PDs:

- Opening move: Cooperate or Defect.
- A plan about how to response to the first play. There are four ways that first stage could have gone.
- A plan about how to respond to the first **and** second plays. There are 16 ways the first two stages could have gone:
  - $\langle C, C\rangle$ followed by $\langle C, C\rangle$
  - $\langle C, C\rangle$ followed by $\langle C, D\rangle$
  - $\langle C, C\rangle$ followed by $\langle D, C\rangle$
  - And so on.

Strategies in Repeated PDs

Strategies in three-times-repeated PDs:

- You need a binary sequence of length $16 + 4 + 1 = 21$ digits to encode a strategy here.
- So there are $2^{21} = 2,097,152$ possible strategies

Number of Strategies in Repeated PDs

How many strategies are there in a four-times repeated PD? In a 5-times repeated?

Answer: Naively, 2^{53} and 2^{117} respectively.

By comparison, physicists estimate there have been 2^{68} seconds since the Big Bang.

Replicator Dynamics and Repeated PDs

Suppose Alexander assigned 500 agents a strategy for a four-times-repeated PD "at random" (i.e. each of the 2^{53} strategies is equally possible).

What is the probability no agent was assigned the strategy "Always Defect"?

Replicator Dynamics and Repeated PDs

Answer: 1.

Moral: No simulations were necessary to show that the population would consist of strategies that sometimes defect and sometimes cooperate. No agents were assigned said strategy initially.

Replicator Dynamics and Repeated PDs

But we could initialize the population to contain a minimum number of defectors, and rerun Alexander's simulations.

Question: In such a population, would Tit-For-Tat and Win-Stay-Lose-Shift still perish, as Alexander found?

Someone should run the simulation to find out ... (Hint, Hint).

Replicator Dynamics and Mutation

Another possibility is to consider mutation.

Suppose "mutants" with random strategies are inserted into the population at different stages.

Question: Does cooperative behavior survive?

Replicator Dynamics and Mutation

- Suppose every individual is playing the same strategy s, which cooperates at some point vs. itself.
- **Question:** Is s a best-response to itself?
- Answer: No. The argument is analogous to last week's "backward induction" proof: consider a strategy s^* just like s, except that it defects the last time that s cooperates against itself.

Replicator Dynamics and Mutation

- Suppose a population of individuals employ strategy s, which is not a best response to itself.
- **Question:** According to the replicator dynamics, what happens when a mutant playing s^* is introduced into the population?
- Answer: It invades, taking over the population. Why?
  - If s^* is a better response to s than is s itself, then s^* will have higher fitness.
  - By the replicator dynamics, the proportion of individuals playing s^* will increase.

Replicator Dynamics and Mutation

Moral 1: It's not clear that cooperative behavior survives in the replicator dynamics of a repeated PD **with mutation**.

This is an active area of research: do a Google search.

Replicator Dynamics and Mutation

Moral 2: If a population's composition is resistant to mutation, then agents are either playing strategies that constitute a Nash equilibrium of the game.

- There's a sticky point here about mixed strategies, which I'll skip. Think a bit about how the above "moral" should be rephrased when mixed strategies are employed in Nash Equilibria.

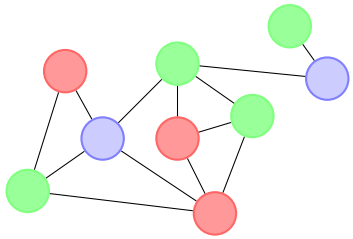
Repeated PDs on Networks

Question: According to Alexander, what happens if, on each stage of evolution, agents play repeated prisoners' dilemmas on the various types of networks?

It's a trick question. [Alexander, 2007] develops no models of this sort and runs no simulations.

Networks

In the past, I've shown you networks like this:



Nodes = Agents
Edges = Indicate which agents "interact"
Colors = "Type" of Agent

Dynamic Networks

But real networks change . . .

- Individuals find new friends and ditch old ones on Facebook.
- Computers in computer networks break and are sometimes replaced.
- Airports in airport networks are abandoned or shut down particular flights.
- Authors on the www add new pages, destroy old hyperlinks, etc.
- And so on.

Dynamic Networks

Two ways to change a network:

1. Add and delete agents.
2. Add and delete edges.

Alexander [2007] considers only modifications of the second type.

Dynamic Networks

There are several different ways of changing edges:

1. In the game-theoretic setting: form links with those with whom you earned higher payoffs in the past.
   - This is the model Alexander describes.
   - Perhaps unsurprisingly, cooperators stop interacting with defectors in PDs.
   - So cooperation can be sustained in a population, which self segregates according to strategy.
2. But there are lots of other methods for changing networks. See Bilgin and Yener [2006] for a survey.

Preferential Attachment

The most common model for dynamic networks is called preferential attachment: agents form new link to agents that have many existing neighbors.

1. The idea is that edges represent status, and agents try to gain status by forming links with those who have it.
2. Think of co-authorship among scientists: writing a paper with a famous scientist makes you look good.
3. Preferential attachment models evolve to produce power law degree distributions, which lends them some measure of empirical support for certain social networks.

Empirical Support for Dynamic Network Models

In what ways are the networks produced by Alexander's learning dynamics similar to and different from those of "real" social networks?

I don't know, but it would be interesting if someone investigated this (Hint, Hint).

Outline

Modeling Trust

Alexander models "trusting" behavior by the stag equilibrium of the stag hunt.

To see whether trust evolves, we can analyze the same types of models as before:

- Classical
- Replicator Dynamics
- Network Models
  - Lattice
  - Bounded degree
  - Small worlds
  - Dynamic

And in each model, agents might play a one-shot or a repeated stag hunt!

Today

Today: Brief explanation why classical and replicator dynamics cannot do the job in explaining the emergence of trust ...

Multiple Equilibria

	Stag	Hare	
Stag	$\langle 2, 2 \rangle$	$\langle 0, 1 \rangle$	
Hare	$\langle 1, 0 \rangle$	$\langle 1, 1 \rangle$	

The Stag Hunt has multiple equilibria. So we cannot simply predict that rational agents will end up in an equilibrium. This is called the problem of equilibrium selection.

Multiple Equilibria

	Stag	Hare	
Stag	$\langle 2, 2 \rangle$	$\langle 0, 1 \rangle$	
Hare	$\langle 1, 0 \rangle$	$\langle 1, 1 \rangle$	

Which if any strategic profiles would agents play if the employed the following decision rules?

- Dominance
- SEU Maximization
- Minimax

Multiple Equilibria

	Stag	Hare	
Stag	$\langle 2, 2 \rangle$	$\langle 0, 1 \rangle$	
Hare	$\langle 1, 0 \rangle$	$\langle 1, 1 \rangle$	

Which if any strategic profiles would agents play if the employed the following decision rules?

- Dominance - No action is dominant. No prediction.
- SEU Maximization - Depends upon probabilities. No prediction.
- Minimax - Hare equilibrium.

Multiple Equilibria

- The decision rules discussed thus far do not favor the stag equilibrium.
- Alexander discusses a second "classical" attempt to address equilibrium selection, which involves "risk dominance." I won't discuss this here, except to say
  - The strategy that is risk dominant depends upon the exact payoffs in the Stag Hunt.
  - Consequently, it does not provide an explanation of trust, if trust is identified with hunting stag in a stag hunt **regardless of exact payoffs**.

Multiple Equilibria

Moral: Existing "classical" explanation do not explain the evolution of trust, as conceived by Alexander.

Stag Hunt and The Replicator Dynamics

What about the replicator dynamics?

Stag Hunt and The Replicator Dynamics

It turns out absolutely everything depends upon

- The payoff matrix
- The initial proportion of stag hunters in the population.

Again, this does not make a good solution ...

Stag Hunt and The Replicator Dynamics

Network models do allow trust to evolve, but the payoffs, learning algorithms, and network structure can have differing degrees of strength in this explanation?

Purpose of Models

Question: What does all this tell us about cooperation and trust, especially if the models give different results?

In the last few classes, we'll talk about the purposes of modeling, the pitfalls, the advantages, and the disadvantages.

Purpose of Models

Answers for Today:

- How possible stories vs. How so
  - Given the problems with classical economic explanations, we are often just interested in explaining how it is **possible** that cooperation evolved.
- Provides motivation and framework for particular empirical investigations:
  - Many social scientists have characterized properties of real social networks.
  - Biologists can sometimes quantify the energy spent by organisms in acting; that is, they can measure the payoff structure.
  - Both biologists and social scientists study learning rules employed by organisms.

How Possible

Question: If we were just interested in "how possible" stories for the evolution of cooperation, then why consider so many models? Isn't one sufficient?

Robustness

Potential Answer: Robustness.

- "How possible" stories are not convincing if they are fragile, i.e., if slight changes to the model cause drastic changes in behavior.
- If many different models behave similarly, however, then "how possible" explanations become more convincing. Such behavior is said to be robust.
- Different models are more-or-less realistic in different ways and so may provide different reasons to believe a "how-possible" story.

Robustness

For discussions of robustness, see [Muldoon, 2007] and [Parker, 2011]; the former defends the value of robust models and the latter questions it.

Topics

Topics we'll discuss today: More of the following

- World Commands
- Agents: Turtles, Patches, and Links
- Agent Sets

References I

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

Axelrod, R. (2006). The evolution of cooperation: revised edition.

Bilgin, C. C. and Yener, B. (2006). Dynamic network evolution: Models, clustering, anomaly detection. *IEEE Networks*.

Muldoon, R. (2007). Robust simulations. *Philosophy of Science*, 74(5):873–883.

Parker, W. S. (2011). When climate models agree: The significance of robust model predictions. *Philosophy of Science*, 78(4):579–600.