Plato's Second Puzzle:
Meaning and the Evolution of Signaling

Conor Mayo-Wilson

Munich Center for Mathematical Philosophy

Women's Summer School in Mathematical Philosophy
July 31st, 2014

Review

Review

- Two Platonic Puzzles: Justice and Meaning
- ABMs vs. population models
- First Puzzle: ABMs of evolution of cooperation

Today: Plato's second puzzle concerning meaning

We'll follow the same method we did yesterday.

ABMs of Cultural Evolution

Step 1: Build an ABM in which agents repeatedly play a game.

ABMS CULTURAL EVOLUTION

Step 2: Identify meaningful communication with a strategic profile in a game.

ABMS CULTURAL EVOLUTION

Step 3: We'll argue that **If**

- Players repeatedly play certain games in certain environments
- Learn to interact with one another over time in certain ways,

Then the strategic profile corresponding to meaningful communication will become prevalent.

WHAT STEPS MUST BE ALTERED?

To put it another way, consider what steps of yesterday's method must be altered in order to build an ABM of the evolution of language.

- Agents with properties = Preferences in some game
- Environment = Social Network
- Initial Conditions = Randomized
- Interaction rules = Imitation and Reinforcement Learning

WHAT STEPS MUST BE ALTERED?

To put it another way, consider what steps of yesterday's method must be altered in order to build an ABM of the evolution of language.

- Agents with properties = Preferences in some game
- ~~Environment = Social Network~~
- ~~Initial Conditions = Randomized~~
- ~~Interaction rules = Imitation and Reinforcement Learning~~

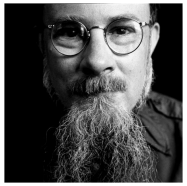
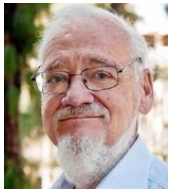
Idea: We can alter the game in the first step and leave the remaining parts of the ABM construction procedure untouched.

Meaning and Signaling Games

Central thesis: Many have argued that a particular strategic profile in a Lewis signaling game – called a "signaling system" – represents meaningful exchange of information

Just as cooperation, trust, etc. are represented by particular strategic profiles in prisoners' dilemmas, stag hunts, etc.

Meaning and Signaling Games

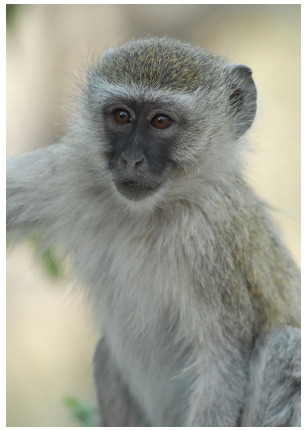




- Lewis [2008]'s definitions of **convention** and **signaling system** are an attempt to characterize, in game theoretic terms, when individuals' signals have acquired meaning.
- In Chapter 3, Skyrms [2010] gives an information-theoretic account of when signals convey information and how much.
- Millikan [2005] argues that these formal accounts are not sufficient, but need to be supplemented by a historical/evolutionary story about how the signals came to have causal properties.

Outline

Signaling Games

Example: Vervet monkey signals



Signaling Games

Example: Vervet signal

- Sender: A vervet monkey who sees a predator. He or she can
  - "Cough"
  - "Chutter", or
  - "Bark"
- Receiver: Another vervet monkey, who has not seen the predator yet.
- They both want each other to evade the predator, but ...
- Clearly, vervets did not schedule a meeting in which they decided that "cough" means that an eagle is approaching.

Signaling Games

Signaling games also have two players: sender and receiver.

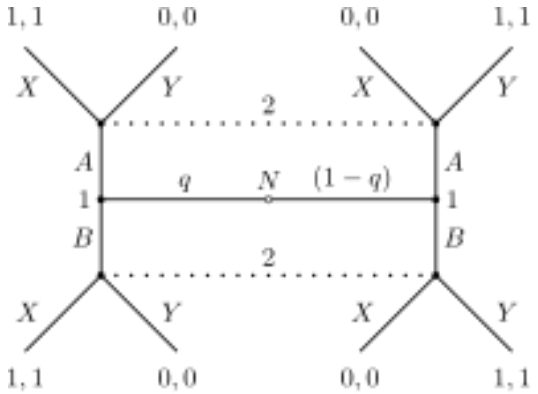
- Sender observes some state of the world (e.g., snake, eagle, or leopard).
- She then sends a signal to receiver (e.g., cough, chutter, or bark).
- The receiver then chooses an action (e.g. stand tall and back away, take cover in underbrush, scale a tree)
- The payoff that both receive depends upon the world and the receiver's action, e.g.,
  - Snakes are evaded by standing tall and backing away
  - Leopards are evaded by climbing trees, and
  - Eagles are evaded by hiding in the underbrush

Signaling Games

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.
- Nature's "Actions": Probability distributions over worlds W
- 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 Games



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?

Discussion: What aspects of human language are not captured by this game-theoretic analysis of meaning?

Limitations of simple signaling games

Discussion: What aspects of human language are not captured in this simple model?

Here are three that many notice:

- **Invention:** The set of signals and actions are fixed. But humans invent new words all the time.
- **Compositionality:** The signals are never combined to form larger meaningful utterances.
- **Conflict of Interest:** The interests of the sender and receiver are aligned. But meaning is conveyed even when there is conflict of interest.

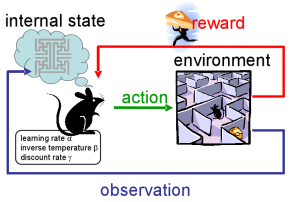
Limitations of simple signaling games

Luckily, modelers are already addressing these worries:

- **Invention:** [Alexander et al., 2012]
- **Compositionality:** [Barrett, 2013, 2014]
- **Conflict of Interest:** [Bergstrom and Lachmann, 1998] [Lachmann et al., 2001]

Outline

Two Types of Learning



Learning Rules:

- We can also use imitation rules and reinforcement learning in repeated signaling games.
- However, an important distinction arises because in signaling games, players do not choose simultaneously.

Two Types of Learning

Recall, a (receiver's) strategy in a signaling game is plan (i.e. function) consisting of conditionals of the form "If I see signal s, I will choose act a" for each possible state s.

Two Types of Learning

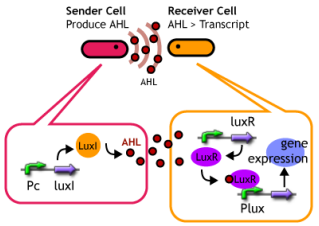
If I am unsuccessful when I chose action a after seeing signal s, I could change either

- **Learning Actions:** Only the part of my plan about how I should respond to signal s.
- **Learning Plans:** Several parts of my plan, including how I might respond to signals other than s.

Two Types of Learning

The two ways of learning are plausible (or implausible) in different contexts.

Learning Actions



Example: Bacteria likely only change how they respond to chemicals one at a time. They learn actions.

Learning Actions

A second example in which learning strategies is implausible: Imitation rules.

- When agents played simultaneous games like a Prisoners' dilemma, imitating one's neighbor meant imitating "Cooperate" or "Defect." One could easily imitate an entire strategy.
- Consider now a signaling game on a network.
- Suppose I see you (my neighbor) play a signaling game successfully in which you respond to signal s with action a.
- It seems that I should only be able to imitate how you respond to signal s; I cannot imitate your entire strategy because I may not have seen how you behave in other circumstances!

Learning Strategies

Nonetheless, there are circumstances in which humans clearly "learn plans."

- Suppose you learn that the correct answer to (the signal) "Is $5 > 3$?" is (the act of asserting) "Yes."
- Then you'll likely update your disposition to answer (the signal) "Is $5 < 3$?" with the answer "No."
- In this case, you've updated your response to one signal given your response to another different signal.
  - This seems related to compositionality, but I'm not sure why.

Programming Concepts

Programming Concepts:

- Procedures
- Agent Commands

References I

Alexander, J. M., Skyrms, B., and Zabell, S. L. (2012). Inventing new signals. *Dynamic games and applications*, 2(1):129–145.

Barrett, J. A. (2013). On the coevolution of basic arithmetic language and knowledge. *Erkenntnis*, 78(5):1025–1036.

Barrett, J. A. (2014). The evolution, appropriation, and composition of rules. *Synthese*, pages 1—14.

Bergstrom, C. T. and Lachmann, M. (1998). Signaling among relatives. III. talk is cheap. *Proceedings of the National Academy of Sciences*, 95(9):5100—5105.

Lachmann, M., Szmad, S., and Bergstrom, C. T. (2001). Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Sciences*, 98(23):13189—13194.

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

Millikan, R. G. (2005). *Language: A biological model*. Oxford University Press on Demand.

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