Learning to Signal
with 2 kinds of Trial and Error

Brian Skyrms

Suppesfest March 10, 2012

1. Low Rationality Game Theory

High rationality game theory has strong cognitive assumptions.

Can the same results be gotten with low rationality learning dynamics? Or not?

What predicts actual behavior?

A Pioneering Work (1960)

Suppes and Atkinson

Markov Learning Models for Multiperson Interactions

2. Two Kinds of
Trial and Error Learning

Herrnstein-Roth-Erev
Reinforcement

Probability of a specific choice is proportional to accumulated rewards from that choice in the past.

Herrnstein, R. J. “On the Law of Effect.” *Journal of the Experimental Analysis of Behavior* 13: 243-266, 1970.

Roth and Erev *GEB* 1995. Erev and Roth *AER* 1998.

Probe and Adjust

Mostly an agent just keeps doing the same thing, occasionally she tries something at random (*probe).*

If the probe gives a better payoff than she got last time she switches to it, if worse she goes back, if a tie she flips a coin (*adjust).*

Skyrms 2010. Huttegger and Skyrms forthcoming.

3. Signaling Games

D. K. Lewis *Convention* 1969

Signaling Games (for our purposes today)

Nature chooses a situation with uniform probability from $s_1 \ldots . s_n$.

Sender observes the situation and chooses a signal from $t_1 \ldots . t_m$

Receiver observes the signal and guesses the situation (from $s_1 \ldots . s_n$.)

Both players are paid 1 if the act matches the situation, 0 otherwise.

We apply trial and error to acts, not strategies.

Agents do not need to *have* strategies, or know the they are playing a game.

4. Reinforcement

Simplest Game

m=n=2

Playing the simplest game with Roth-Erev reinforcement learning

Sender has one urn for each state.

Initially, each urn has one A ball and one B ball.

Receiver has one urn for each signal.

Initially, each urn has one 1 ball and one 2 ball.

Repeated trials with reinforcement of the balls drawn on a trial.

Convergence to a signaling system with probability 1.

Argiento, R., R. Pemantle, B. Skyrms and S. Volkov. "Learning to Signal: Analysis of a Micro-Level Reinforcement Model." *Stochastic Processes and their Applications* 2009.

5. Probe and Adjust Simplest Game

Sketch of playing the simplest game with Probe and Adjust

Sender remembers what happened last time in each situation.

Receiver remembers last time for each signal.

State of the system: map from situations to signals + map from signals to guesses.

(Assumption: either sender or receiver probes and adjusts at one time.)

Agents learn to signal (in the appropriate sense) with probability 1

Signaling systems are the unique absorbing states.

There is a positive probability path from any state to a signaling system.

How do these results generalize?

6. Reinforcement Learning General Case

N situations, M signals, N acts

(Hu et al - Yilei Hu Thesis 2010)

Bipartite Graph

- Consider the bipartite graph that maps a situation to a signal if there is positive probability.

Key Property *P*

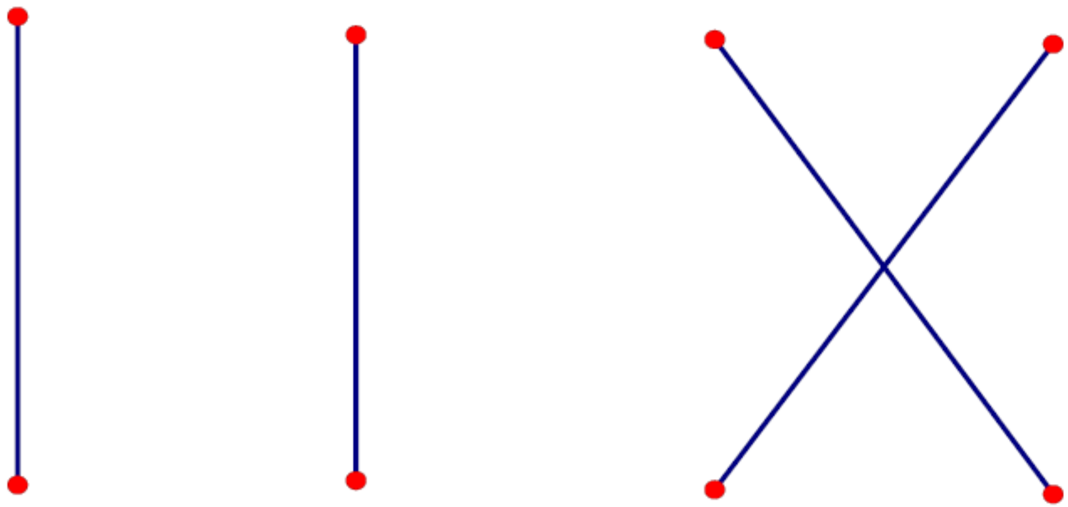
- Each connected component of the graph has just one state or just what signal.

- Each vertex has an edge.

    Each state or signal is connected to something.

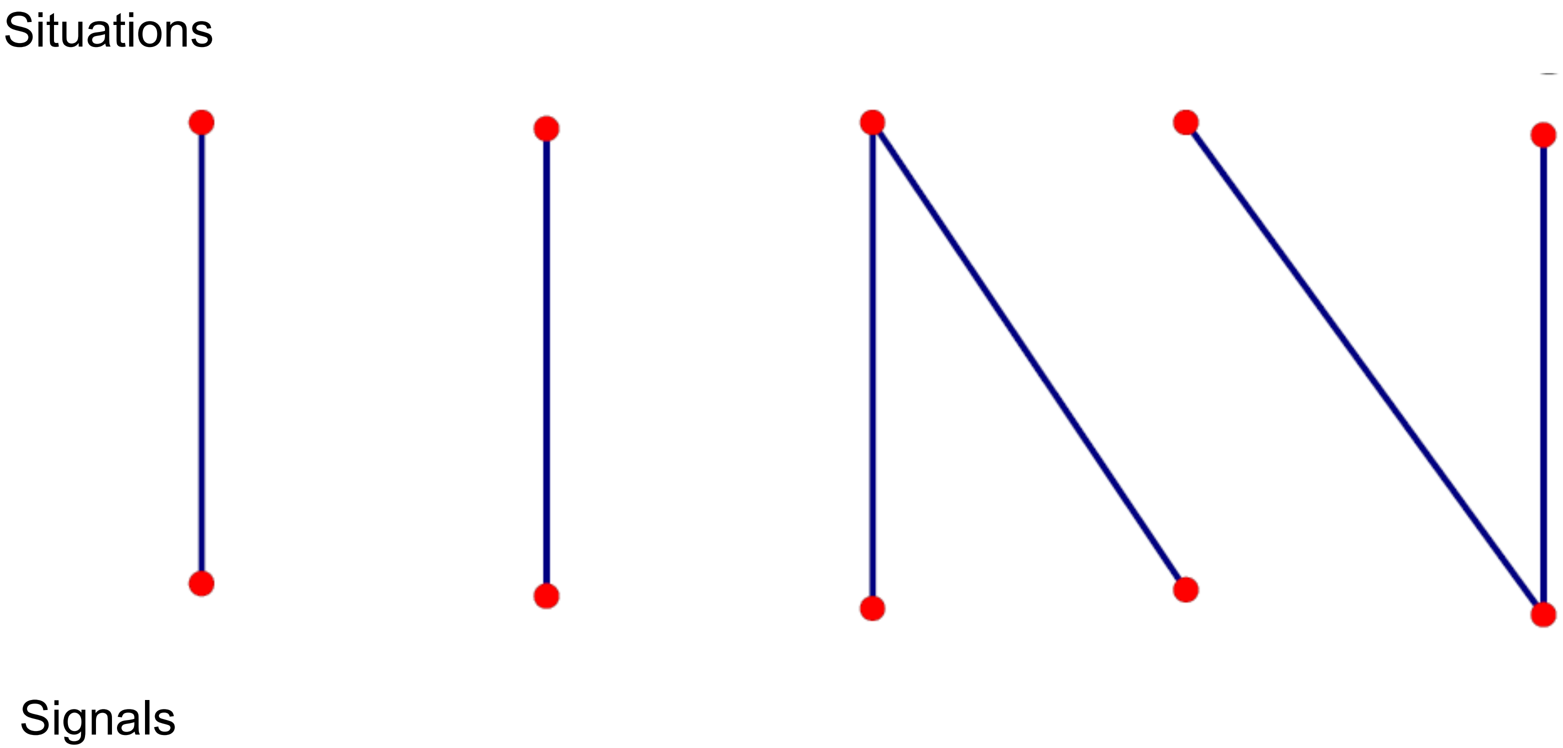
Example: Signaling System

Situations



Signals

Example: Synonyms & Bottlenecks



Th. If a bipartite graph has property *P*, then reinforcement learning converges to it with positive probability.

Signaling systems, synonyms and bottlenecks all have positive probability.

7. Probe and Adjust

$M=N$

Just like the simplest case

8. Probe and Adjust

M>N

Too many signals

Example: N=2,M=3.

No absorbing states,

but multiple absorbing sets.

Learn to signal with probability 1.

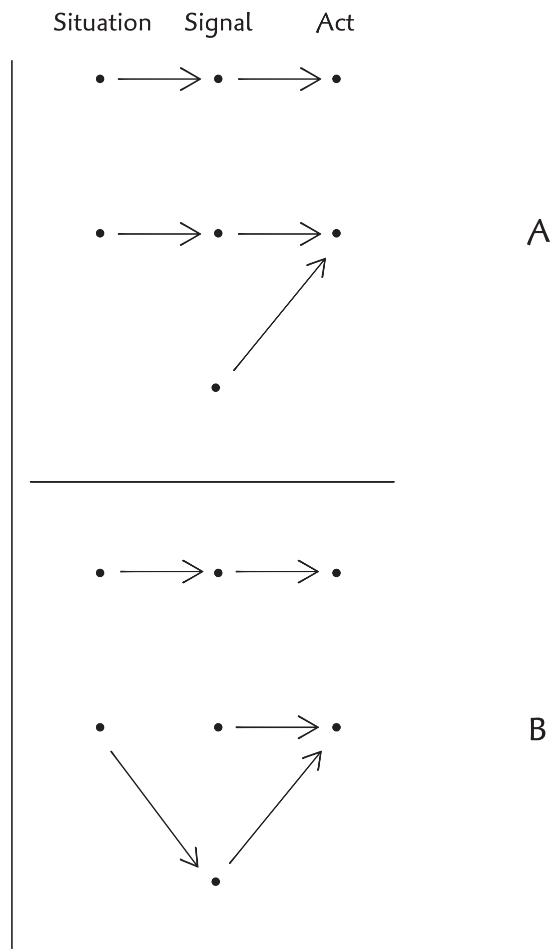


figure 3

9. Probe and Adjust

$M<N$

Too few signals

Now we must have bottlenecks.

For example

Sender	**Receiver**
s1 => t1	t1 => a1
s2 => t2	t2 => a2
s3 => t2	

12 efficient equilibria

One big absorbing set of efficient equilibria.

Positive probability path from any state to this set.

9. Unequal situation probabilities

Reinforcement.

Probe and adjust.

10. Conclusions

In the simplest game both forms of trial and error *always* learn to signal.

In *all* signaling games, both forms of trial and error *sometimes* learn to signal,

…..and *probe and adjust*

always learns to signal.

Thank you.

Moving towards the center

Reinforcement learning can learn faster by decreasing the initial weights.

Probe and adjust can be modified by making probability of a probe payoff-dependant.

Extreme forms of these modifications give almost the same rule:

- Win stay
- Lose, probe with some probability.

This always learns to signal.

--- and there is no problem with simultaneous probes.

Reinforcement in Bandit Problem

Chosen?		*Reinforced?*		*New Value of p(R)*	
1. [p(R)	*	q	*	(N p(R) +1)/(N+1)]	+
2. [p(R)	*	(1-q)	*	p(R)]	+
3. [(1-p(R))	*	p	*	(N p(R))/(N+1)] +	
4. [(1-p(R)	*	(1-p)	*	p(R)]	

- *(1/N+1) p(R) (1- p(R) (q-p)* *(expected increment)*

d p(R)/dt = pr(R) (1-pr(R) (q-p) *(mean field dynamics)*

(1) Starting from a signaling system every *probe* changes a payoff from 1 to 0. Then *adjust* returns to the signaling system.

(2) If a state is not a signaling system, some probe either gives the same payoff or a greater one. Thus some probe leads away with positive probability.

(3) *Start with S1*. The composition of sender and receiver functions g(f(S1) map it to an act. If it is A1, move on. If it is not A1, nature chooses the situation and the receiver to probe. Receiver probes A1, and adjusts to choose A1 for that signal since the probe moved payoff from 0 to 1.

Continue as follows:

- Consider Sn. If sender maps it to a signal that does not yet appear on the path, proceed at above. The composition of sender and receiver functions g(f(Sn) maps it to an act. If the act is An, move on. If it is not An, nature chooses the situation and the receiver to probe. Receiver probes An, and adjusts to choose An for that signal since the probe moved payoff from 0 to 1.
- If sender maps it to a signal already visited on this path [f (S1) --- f(Sn-1)] then nature chooses the situation, sender probes an unused signal [not now in the range of f]. There must be one since in this case more than one signal is mapped to the same situation. Previous payoff must have been a 0, since the old signal led to Aj (j<n). so adjust sticks with the probe with positive probability.
- If g(f(Sn) = An move on. Otherwise Nature chooses receiver to probe, receiver probes An, and adjusts by keeping g(f(Sn)=An, since the probe changes zero to 1
- Next Si.

More Problems

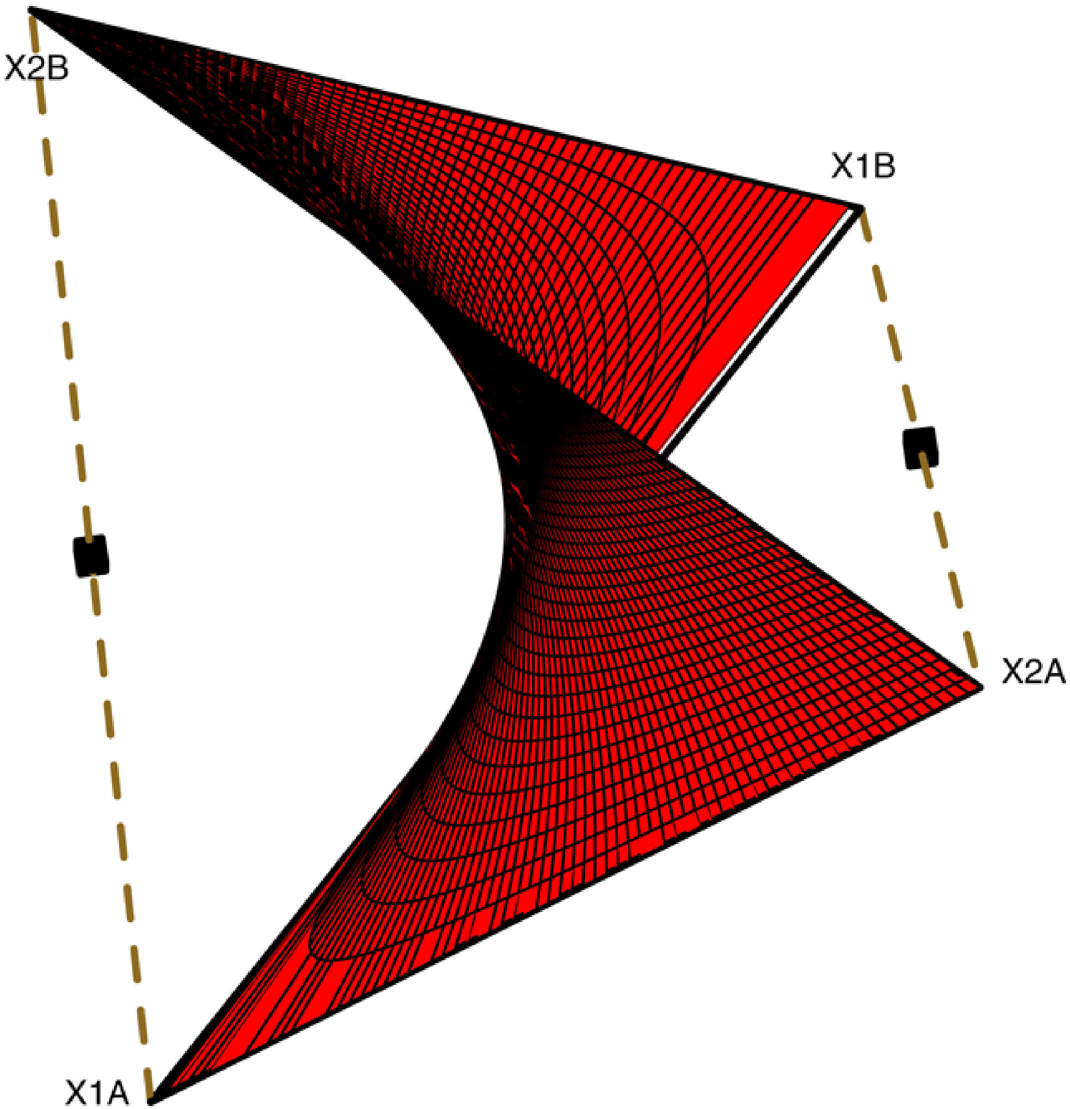
Chains o => o => … o => o

(work in progress Jonathan Kariv)

Inventing New Signals

(in progress Alexander, Skyrms, Zabell)

Unequal Payoffs, Probabilities, Signal Costs



Signaling at <1/2,0,0,1/2>, <0,1/2,1/2,0>