

# Common Knowledge AND Global Games

This talk combines common knowledge with “global games” another advanced branch of game theory

See Stephen Morris’s work

Today we'll go back to a puzzle that arose during the Hawk-Dove lecture:

Why do we typically see the Bourgeois equilibrium and (almost) never see the anti-Bourgeois equilibrium?

**EVEN WHEN** there is no (or little) physical advantage to being the incumbent!

At the time, we didn't yet have the tools to answer this. Now that you've learned about common knowledge and the importance of higher order beliefs, we do

We'll formalize this with a toy model again

EVEN WHEN there is no (or little) physical advantage to being the incumbent!

But SOMETIMES there is. Once incorporated into the model, this SOMETIMES is enough to make Bourgeois the only REAL equilibrium

Here's the model...

We assume the incumbent has an  $X\%$  chance of winning a fight, where  $X \sim U[\frac{1}{2}, 1]$

Here's the payoff matrix...

(E)ntrant

H

D

H

$X(v) + (1-X)(-c), (1-X)(v) + X(-c)$

$v, 0$

(I)ncumbent

D

$0, v$

$\frac{1}{2}v, \frac{1}{2}v$

(E)ntrant

H

D

H

$x(2)+(1-x)(-4), (1-x)(2) + x(-4)$

2, 0

(I)ncumbent

D

0, 2

1, 1

We make the (realistic) assumption that each player does not EXACTLY know  $X$

Instead each gets a noisy signal of  $X$ . The amount of noise could be arbitrarily small.

(So you can be REALLY sure what  $X$  is and REALLY sure what other thinks  $X$  is, but  $X$  isn't common knowledge)

We model this as: player  $i$  gets signal  $X_i$  where  $X_i = X + \epsilon_i$  and  $\epsilon_i$  i.i.d.  $\sim U[-\epsilon, \epsilon]$  for some small  $\epsilon > 0$

In our example, we'll let  $\epsilon = .002$

(Ignore “edge” cases)

We now describe the normal form game...

The players: (I)ncumbent, (E)ntrant

The strategies:  $S_i: [\frac{1}{2}, 1] \rightarrow \{H, D\}$

E.g.,

$S_I(X_I)=H$  iff  $X_I > .7348$

$S_E(X_E)=H$  iff  $X_E$  is a rational number

What are the NE of this game?

Interestingly, there is EXACTLY one:

$$S_I(X_I) = H \text{ for all } X_I$$

$$S_E(X_E) = D \text{ for all } X_E$$

This equilibrium has a very cool implication:

Play Hawk if arrive first EVEN WHEN there is no (or little) advantage

(Except when  $\epsilon=0$ , in which case it can be common knowledge that there is no advantage!)

Here's the proof...

First let's show that:

$$S_I(X_I)=H \text{ for all } X_I$$

$$S_E(X_E)=D \text{ for all } X_E$$

is indeed a Nash equilibrium

Proof:

Is there any signal at which I can deviate and play D and do better? No. On those occasions, I will get 1. This is less than 2, which is what I is currently getting

Is there any signal at which E can deviate and play H and do better? No. On those occasions, E expects to get  $X_E(-4) + (1 - X_E)(2)$ . Since  $X_E \geq 1/2$ , this is  $\leq 0$  which is what E is currently getting

Next, we'll show that there is no other equilibrium

It will prove useful to define  $X^*$  as the value of  $X$  for which the incumbent is indifferent between playing H or D given entrant plays H

$$\rightarrow X^* 2 + (1 - X^*) (-4) = 0$$

$$\rightarrow X^* = 4/(2+4) = 2/3 = .667$$

Next, suppose the strategy pair  $S_I, S_E$  is a NE

$\rightarrow S_I(X_I) = H$  at least when for  $X_I \geq .667$  (b/c at .667, I is indifferent EVEN IF E were to play H everywhere, so he must CERTAINLY play H above)

$\rightarrow S_E(X_E) = D$  at least when  $X_E \geq .665$  (b/c above this D is a best response EVEN IF I plays D everywhere not yet specified)

$\rightarrow S_I(X_I) = H$  for  $X_I \geq 1/2$  (b/c at  $1/2$ , I is indifferent between D and H EVEN IF E plays D everywhere not yet specified)

$\rightarrow S_E(X_E) = D$  for  $X_E \geq 1/2$

☺ Notice this same logic would work for different  $c, v$  or any  $\epsilon > 0$ , albeit it might have taken more steps. (Prove this? What about other distributions of  $\epsilon$  and  $X$ ?)

The proof also shows us the  
learning/evolutionary process

Suppose that we start at a different strategy profile.

Claim: the population will evolve/learn to play Bourgeoisie equilibrium.

Here's the logic....

Regardless of where population starts all the incumbents will learn/evolve to play  $S_I(X_I) = H$  at least when for  $X_I \geq .667$

because any incumbent who doesn't play this will get lower payoff

Before long, all the entrants will learn/evolve to play  $S_E(X_E) = D$  at least when  $X_E \geq .665$

because any entrant who doesn't play this will get lower payoff

(In general, evolution/learning “iteratively eliminates strictly dominated strategies”)

Let's see if we can use this framework to explain additional puzzles...

Why isn't it the case that size acts as the uncorrelated asymmetry? I.e. why doesn't simply the bigger animal play hawk and the smaller play dove?

Turns out, this isn't a Nash equilibrium, once we recognize that there is some (albeit perhaps miniscule) uncertainty over size

Proof:

Suppose they play strategy: play hawk whenever estimate bigger

Suppose one animal thinks he is only slightly bigger. Then he estimates there is nearly a 50% chance the other also thinks he is slightly bigger. So he thinks other will play H with probability nearly 50%

If plays Hawk gets  $\frac{1}{2} \frac{1}{2} (v-c) + \frac{1}{2} v = \frac{3}{4}v - \frac{1}{4}c$

If plays Dove gets  $\frac{1}{2} (0) + \frac{1}{2} \frac{1}{2}v = \frac{1}{4}v$

If  $\frac{3}{4}v - \frac{1}{4}c > \frac{1}{4}v$ , i.e.  $v/c > \frac{1}{4}$ , is better off playing Dove

If  $v/c < \frac{1}{4}$ , then when animal is slightly smaller will strictly prefer to play hawk. Either way, our purported equilibrium won't hold unless  $v/c$  is exactly  $\frac{1}{4}$

Notice the “problem” arose because size differences can get arbitrarily small, i.e. size is continuous

Third and final animal question:

Can Bourgeois be an equilibrium, even if there is slight uncertainty over who arrived first?

Answer: yes.

Suppose is 5% error rate.

Then regardless of signal receive, cannot benefit from deviating...[insert proof]

Notice: this differs from size b/c arriving first is discrete/categorical. Noise works differently...arriving first is still evident p for sufficiently high p (here p is...). Whenever believe likely arrived first, believe other believed likely arrived first (.90) and whenever don't believe likely (.90) arrived first, believe other doesn't believe likely arrived first...this wasn't true for size...

Next application... categorical vs continuous norms

Suppose

- It is ONLY worth attacking countries if we are PRETTY CONFIDENT others will attack  
(e.g. we need to be confident 60%)
- We cannot detect EXACTLY how many civilians are killed  
(e.g. we get a signal which is uniformly distributed between the true value $\pm 10$ )

We will show: It is not possible to attack a country depending on our estimate of the number of civilians killed, regardless of how large the number is (e.g. attack if signal  $> 100,000$ )

Sketch of proof: Imagine that to attack, we need to be 60% confident that others attack, and that we set a rule that we attack any country whenever we estimate that the despot kills 100,000 civilians.

Suppose we estimate that 100,001 civilians were killed. Should we attack? No! There's a 45% chance France thinks there were  $< 100,000$  civilian casualties and won't attack, so we are better off not attacking.

OK so it's not an equilibrium...but only rarely can you benefit from deviating, so maybe if norm is 100,000 and get signal of 200,000 will attack?

Two reasons why not:

- 1) evolution/learning
- 2) if states are rational and believe other states are rational and...

## Evolution / Learning

When we get signal 101,000, we quickly learn not to attack. Same for France.

Shortly Thereafter, when we get 102,000, we will ALSO learn to not attack, because France won't attack at 101,000.

EVENTUALLY, we won't attack at 200,000 either (but this might take a REALLY long time...)

(Although, this gives us some sense of how “continuous norms” will “unravel.”

What about if the payoffs were such that want to attack so long as 45% sure? Only difference is unraveling will go in OTHER direction.

(So now we gain prediction of DIRECTION of unraveling. Cool!)

If states are rational, and expect each other to be rational, and expect the other to expect them to be rational:

The U.S. knows France would never attack if she received 101,000. So the U.S. won't attack at 102,000. The U.S. knows that France would anticipate this, so the U.S. knows France wouldn't attack at 103,000...

What about a norm against chemical weapons?

(A categorical norm)

Suppose each country gets a WRONG signal of whether a chemical weapon was used with probability .05

Recall our result from last class.  $p=.95$ , so there IS an equilibrium where punish when get signal other used chemical weapon (if we assume payoffs assumed last class)

The proof works the same way as our proofs did last class...

Let's assume the following payoffs

$$U(\text{both attack}) = -1$$

$$U(\text{attack alone}) = -2$$

$$U(\text{only other attacks}) = -1$$

$$U(\text{neither attacks}) = -2$$

Current payoffs to each country:

$$.95 * U(\text{both attack}) + .05 * U(\text{attack alone}) = -.95 - .1 = -1.05$$

Should either country deviate and not punish when it gets the signal? Then payoffs are:

$$.95 * U(\text{only others attack}) + .05 * U(\text{no one attacks}) = -1.9 - .1 = -2$$

Should a country deviate and punish when it doesn't get the signal? Then payoffs are:

$$U(\text{attack alone}) = -2$$

Neither deviation is worthwhile

For which payoffs does this work? What distribution on signals needs to be assumed?

Turns out our main CK theorem will play a role!

## Insights:

- 1) Which norms are “categorical”?
- 2) When do we expect to see categorical norms?
- 3) When ought we to see categorical norms?
- 4) What will happen to continuous norms over time?
- 5) Which emotions should vary continuously?
- 6) Why does the U.S. fight over a tree in N. Korea?
- 7) Why did the Nazis take away rights of Jews in slow steps?

MIT OpenCourseWare  
<http://ocw.mit.edu>

14.11 Insights from Game Theory into Social Behavior  
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.