

Common Knowledge: Formalizing the Social Applications

Today and Thursday we'll take a step in the direction of formalizing the social puzzles, such as omission commission.

First, a reminder of the set-up and the theorem...

The set-up...

We have an information structure:

$$\langle \Omega, \pi = (\pi_1, \pi_2), \mu \rangle$$

And a coordination game (or any game with multiple equilibria)

And we showed...

Define a non-constant equilibrium as an equilibrium in which $\exists i$ s.t. S_i is not a constant function.

\exists a non constant equilibrium

Iff

$\exists E, F \subseteq \Omega$ s.t. $E \cap F = \emptyset$, E is p^* -evident and F is $(1-p^*)$ -evident, where $p^* = (d-b)/(d-b+a-c)$

Let's create a stylized example

Suppose Moshe and Erez always fight over who has to grade the homeworks. Erez currently grades the hw. Suppose Moshe apologizes for this imbalance.

Claim: It is an equilibrium for Moshe to grade the homework after he apologizes, even if Moshe's apology isn't always heard, but not if the apology was indirect.

Let's formalize this one step at a time

First the game

We can model who grades the HW as a hawk dove game (hawk means doesn't grade).

This will work just like a coordination game.

Now the information structure, which is a little boring in this case

States: $\Omega = \{\text{apologizes, didn't}\}$

Partitions: $\pi_M = \pi_E = \{ \{\text{apologizes}\}, \{\text{didn't}\} \}$

Prior: μ

Recall that strategies are functions from signals to actions.

Define the following strategy pair:

$$s^*_M(\text{apologizes}) = D$$

$$s^*_M(\text{didn't}) = H$$

$$s^*_E(\text{apologizes}) = H$$

$$s^*_E(\text{didn't shake hands}) = D$$

We can directly show this is a NE bc neither player can benefit by changing his strategy for any signal.

Define the payoff-irrelevant events E and F as follows

E = apologizes

F = didn't

Let's check if they are p -evident for all values of p

Recall E is p -evident if $\forall \omega \in E, \forall i \mu(E | \pi_i(\omega)) \geq p$

Only $\omega \in E$ is apologizes

$$\mu(\text{apologizes} | \pi_E(\text{apologizes}) = \text{apologizes}) = 1$$

$$\mu(\text{apologizes} | \pi_M(\text{apologizes}) = \text{apologizes}) = 1$$

Similarly for F

Thus, according to the theorem, there must be an equilibrium where each player plays Hawk in at least one state and plays dove in at least one state.

So they can condition their behavior on whether Moshe apologies...

What happens if there is some chance Erez doesn't hear Moshe's apology?

$\Omega = \{M \text{ apologizes and E hears, } M \text{ apologizes and E doesn't hear, } M \text{ doesn't apologize}\}$

$\pi_M = \{ \{M \text{ apologizes and E hears, } M \text{ apologizes and E doesn't hear}\}, \{M \text{ doesn't apologize}\} \}$

$\pi_E = \{ \{M \text{ apologizes and E hears}\}, \{M \text{ apologizes and E doesn't hear, } M \text{ doesn't apologize}\} \}$

Prior: Let's assume that whenever Moshe apologizes there's a high probability that Erez hears (.95)

Let's assume that Moshe apologizes with relatively low probability (.3)

$$\mu(M \text{ apologizes and E hears}) = .3 * .95 = .285$$

$$\mu(M \text{ apologizes and E doesn't hear}) = .3 * .05 = .015$$

$$\mu(M \text{ doesn't apologize}) = .7$$

Let's fill in values for the H-D game...

	H	D
H	$\frac{1}{2}(2-4)$	2
D	0	$\frac{1}{2}2$

Consider the strategy pair

$$s^*_M(\{M \text{ apologizes and } E \text{ hears, } M \text{ apologizes and } E \text{ doesn't hear}\}) = s^*_M(\text{"apologizes"}) = D$$
$$s^*_M(\{\text{doesn't apologize}\}) = H$$
$$s^*_E(\{M \text{ apologizes and } E \text{ hears}\}) = H$$
$$s^*_E(\{M \text{ apologizes and } E \text{ doesn't hear, } \{M \text{ doesn't apologize}\}) = s^*_E(\text{"doesn't hear"}) = D$$

Notice this strategy pair is an equilibrium

Suppose Moshe apologizes. Can he do better by playing hawk?

$$U_M(H | \text{apologizes}) = \Pr\{E \text{ plays D} | \text{apologizes}\} U_M(H, D) + \Pr\{E \text{ plays H} | \text{apologizes}\} U_M(H, H) = \dots$$
$$<$$
$$U_M(D | \text{apologizes}) = \Pr\{E \text{ plays D} | \text{apologizes}\} U_M(D, D) + \Pr\{E \text{ plays H} | \text{apologizes}\} U_M(D, H) = \dots$$

Do the same thing or the other three deviations

Suppose Moshe doesn't apologize. Can he do better by playing dove?

Suppose Erez hears an apology. Can he do better by playing dove?

Suppose Erez doesn't hear an apology. Can he do better by playing hawk?

What if Moshe apologizes indirectly?

The KEY property of indirect speech, we argue, is that **SOMETIMES** even when Erez understand it, Moshe can't tell that Erez understood it.

States:

Moshe either doesn't apologize or apologizes indirectly

When Moshe apologizes indirectly, Erez either gets it or doesn't get it

Moshe sometimes can tell whether Erez gets it, but not always

$\Omega = \{(\text{apologizes, gets it, can tell}), (\text{apologizes, gets it, can't tell}), (\text{apologizes, doesn't get it}), (\text{doesn't apologize})\}$

Partitions:

$$\pi_M = \{ \{(\text{apologizes, gets it, can tell})\}, \\ \{(\text{apologizes, doesn't get it}), (\text{apologizes, gets it, can't tell})\}, \\ \{(\text{doesn't apologize})\} \}$$
$$\pi_E = \{ \{(\text{apologizes, gets it, can tell}), (\text{apologizes, gets it, can't tell})\}, \\ \{(\text{apologizes, doesn't get it}), (\text{doesn't apologize})\} \}$$

Priors:

$$\Pr\{\text{Moshe apologizes}\} = .3$$

$$\Pr\{\text{Erez gets it}\} = .95$$

$$\Pr\{\text{Moshe can tell}\} = .25$$

Let's skip straight to the posteriors:

$$\Pr\{\text{Erez gets it and Moshe can tell} \mid \text{Moshe apologizes}\} = .95 * .25 =$$

$$\Pr\{\text{Erez gets it and Moshe can't tell} \mid \text{Moshe apologizes}\} = .95 * .75 =$$

$$\Pr\{\text{Erez doesn't get it} \mid \text{Moshe apologizes}\} = .05$$

$$\Pr\{\text{Moshe doesn't apologize}\} = .3$$

Consider the strategy pair:

$$s_M(\text{apologize and can tell}) = D$$

$$s_M(\text{apologize and can't tell}) = D$$

$$s_M(\text{don't apologize}) = H$$

$$s_E(\text{gets it}) = H$$

$$s_E(\text{doesn't get it}) = D$$

Not Nash. Moshe can deviate by playing Hawk when apologizes and can't tell:

$$U_M(H \mid \text{apologize and can't tell}) =$$

$$U_M(D \mid \text{apologize and can't tell}) =$$

Now consider the strategy pair:

$$s_M(\text{apologize and can tell}) = D$$

$$s_M(\text{apologize and can't tell}) = H$$

$$s_M(\text{don't apologize}) = H$$

$$s_E(\text{gets it}) = H$$

$$s_E(\text{doesn't get it}) = D$$

Still not Nash. Erez can do better by deviating when gets it:

$$U_E(H \mid \text{gets it}) =$$

$$U_E(D \mid \text{gets it}) =$$

Notice that we didn't give Moshe the option of choosing whether to apologize

One can model this, and if one does, gain another insight: only see apologies if w/o apology at pareto dominated equilibrium (cool!)

Now we've seen that innuendos can't be used to switch equilibria

Why would you ever want to use one, then?

Sometimes I may just want you to have the information but avoid the risk of us switching equilibria

It's not obvious that you would—for this need to formalize

Key insight: in some games, first order knowledge matters, not common knowledge.

E.g., costly signaling. Player 2 is the only one who moves. His move depends on his guess of player 1's type, but he doesn't care about coordinating with player 1

Will formalize, if time permits

Next application... omission/commission

First, let's formalize the notion of coordinated punishment

In a coordinated punishment game, player 1 takes an action that can be "good" or "bad"

Players 2 and 3 get a signal of player 1's action, then decide whether to pay a cost to punish player 1

In all subsequent periods, players 2 and 3 observe each other's punishment choices and decide whether to punish each other

Player 1's actions are G, B

Player 1's payoff from G is -1

Player 1's payoff from B is 0

When Player 1 takes the good action, Players 2 and 3 get signal "good".
When Player 1 takes the bad actions, Players 2 and 3 get signal "bad" with probability p and "good" with probability $1-p$

Players 2 and 3 can pay a cost in any period of 1 to punish other players by 2

To simplify the math, we'll assume the discount rate is approximately 1

Claim 1: if $p = .75$, there exists a Nash equilibrium where player 1 plays G, and players 2 and 3 punish 1 iff they see B, and punish the other player anytime the other was expected to punish the previous period and did not

Claim 2: if $p = .25$, there is no Nash equilibrium where players punish only when they see the bad signal

Proof of claim 2:

Player 2 benefits from deviating to not punishing when gets the bad signal

Gains 1 for not punishing

With probability .25, player 3 gets the bad signal, too, and punishes player 2

With probability .75, player 3 doesn't get the bad signal and doesn't punish

$$\text{Net gain} = 1 - .25*2 + .75*0 = .5$$

We assume key distinction between omission and commission is p is high for commission (it is likely that any observer can tell a bad deed was committed)

Thus

If p is low, even when 2 got the “bad” signal, he can’t punish 1.

(can’t punish omission even when bad intentions are clear to you)

In contrast, if p is high, there is an equilibrium where 2 punishes 1 when he gets the bad signal.

(can punish commission)

Let's conclude by connecting these applications to the theorem

The theorem taught us that we can only condition behavior in coordination games on events that create common knowledge

This is what we saw in each case

apologies create common knowledge, even if sometimes go unheard

Innuendos and omission don't so can't affect hawk dove games or coordinated punishment.

Appendix of unused slides

Suppose Moshe and Erez are negotiating over a bowl of ramen after class

We can think of the ramen as a contested resource, and of Moshe and Erez as players in a Hawk-Dove game

Erez currently “owns” the ramen

So he plays hawk over it, and Moshe plays dove

Erez agrees to “sell” the ramen to Moshe over a handshake

Erez plays the strategy: play dove if we shake hands, hawk otherwise

Moshe plays the strategy: play hawk if we shake hands, dove otherwise

Is this a NE?

Beliefs:

$$\mu(\text{shook hands} \mid \text{shook hands}) = 1$$

$$\mu(\text{didn't shake hands} \mid \text{didn't shake hands}) = 1$$

For which SETS B is there an equilibrium in which player A doesn't play any element of B.

(next class we will answer this question formally thereby addressing “categorical” norms)

OK, let's formalize a simple example...

Suppose I have a research assistant

The RA says he worked overtime, but I have a camera in the lab that the RA doesn't know about, and I know the RA was lying and didn't work over time.

I want the RA to know that I can monitor him so that he doesn't lie in the future, but I worry that if I tell him this explicitly, he'll put less attention into the hours he does work.

Let's formalize why

There are two aspects that I care about, which we will model as two separate games

Both I and the RA get the sum of the payoffs from the two games

The first aspect is the amount of effort we put in to our joint task. We model this as a “minimal effort game”:

	H	L
H	$\frac{1}{2}v - c$	$-c$
L	0	0

Assume $v > 2c > 0$

Notice this is a coordination game. Each player wants to put in high effort only if he expects the other to put in high effort

	H	L
H	$\frac{1}{2}$	-2
L	0	0

The second aspect...

The RA chooses what to put on his timecard: the true number of hours or more

We assume he prefers to put more iff he believes he can't get caught

	Camera	No Camera
T	0, 10	0, 10
C	-10, -20	-10, 15

Me → (points to the 'Camera' column)
 RA → (points to the 'Camera' column)

Assume $v > 2c > 0$

Notice this is a coordination game. Each player wants to put in high effort only if he expects the other to put in high effort

Now let's combine the games...

I have three options:

Explicitly say I have a video camera

Allude to the camera

Say nothing

Information structure...

As with the apology innuendo, the states are ...

I...

allude to the camera

say nothing

explicitly say I have a camera

When I allude to the camera, the RA...

gets it

doesn't get it

I can sometimes tell whether the RA gets it, but not always

Priors:

$$\Pr\{\text{RA gets it}\} = .95$$

$$\Pr\{\text{I can tell}\} = .25$$

Let's skip straight to the posteriors:

$$\Pr\{\text{RA gets it and I can tell}\} = .95 * .25 =$$

$$\Pr\{\text{RA gets it and I can't tell}\} = .95 * .75 =$$

$$\Pr\{\text{RA doesn't get it}\} = .05$$

Then so long as I allude to the camera such that most RAs wouldn't understand that I caught them cheating, then even if I know this RA would realize I caught her cheating, she still continues to put in high effort

Because she doesn't know that I know she realized it

Claim: There is a Nash equilibrium where, when I can tell...

I allude to the camera

And The RA continues to put in high effort and stops cheating

But had I used explicit speech, he would have stopped cheating and we both would stop putting in effort

Proof

Should I deviate and explicitly say there's a camera?

No, then the RA stops putting in effort, so I lose $\frac{1}{2}$

Should I deviate and say nothing?

Then RA lies and I lose 10

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>.