Game Theory

Electronic Mail Game

Tetsuya Hoshino

April 3, 2022

In this note, we study Rubinstein's (1989) insightful example that illustrates the sensitivity of equilibrium prediction to the assumption of common knowledge.

1 Electronic Mail Game

There are two players 1 and 2, each of whom chooses to either "attack" (action A) or "not attack" (action N) a target enemy. There are two states $\theta \in \Theta \equiv \{G, B\}$, where states G and B are realized with common priors $\mu \in (0,1)$ and $1-\mu$. State G is a "good" state to attack the target, so that a coordinated attack can defeat it down. State B is a "bad" state to attack the target, so that a coordinated attack cannot defeat it down. Assume that the payoff structure is defined by the following tables:

	A	N
A	$1, 1$	$-2, 0$
N	$0, -2$	$0, 0$

Table 1: State G

	A	N
A	$-2, -2$	$-2, 0$
N	$0, -2$	$0, 0$

Table 2: State B

We can read this payoff structure as follows: Action N costs payoff -2 (regardless of the state or the opponent's action), but a successful coordinated attack (A, A) at state G yields payoff 3 for each, so that each receives payoff 1 from the successful coordination.

Electronic Mail Communication Consider the following communication protocol:

Round 0: Player 1 observes a realized state θ (but player 2 observes nothing).

Round 1: Player 1 sends an email to player 2 if $\theta = G$, but does not if $\theta = B$.[1]

Round 2: Player 2 sends an email to player 1 after receiving the round-1 email.

Round 3: Player 1 sends an email to player 2 after receiving the round-2 email.

$\vdots$

This protocol is repeated ad infinitum, but each email does not arrive at the destination with small probability $\epsilon > 0$. Hence, the protocol is repeated until an email fails to arrive. After the communication ends, each player simultaneously chooses action A or N.

[1] Each email is "blank." The fact that player 1 has sent the round-1 email means that state G is realized.

	0	1	2	3	4	$\cdots$
0	$1-\mu$	0	0	0	0	$\cdots$
1	$\mu\epsilon$	$\mu(1-\epsilon)^1\epsilon$	0	0	0	$\cdots$
2	0	$\mu(1-\epsilon)^2\epsilon$	$\mu(1-\epsilon)^3\epsilon$	0	0	$\cdots$
3	0	0	$\mu(1-\epsilon)^4\epsilon$	$\mu(1-\epsilon)^5\epsilon$	0	$\cdots$
4	0	0	0	$\mu(1-\epsilon)^6\epsilon$	$\mu(1-\epsilon)^7\epsilon$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Table 3: the type distribution of the electronic mail game

The communication is "automated." Thus, we need not consider players' incentives about whether they want to receive or send emails. The above protocol may seem to be dynamic, but they make a decision only when they choose action A or N.

Types How do we represent this information structure? We define player i's type t_i as the number of emails she has sent. For example, at state B, we have a type profile $t=(0,0)$ since player 1 sends no email, while at state G, if player 1's email fails to arrive at player 2, then we have a type profile $t=(1,0)$. The type distribution is given by Table 3. Each cell denotes the (ex-ante) probability of the correspond type profile. This distribution is well defined because all cells have non-negative probabilities and the sum of them is equal to 1.

1.1 Higher-Order Knowledge*

Player i's type t_i represent her **higher-order knowledge**, which is her knowledge about player $-i$'s knowledge. We see players' knowledge at a given $t=(t_1,t_2)$. For brevity, let K_iE denote that player i knows an event E.[2] For example, K_1G denotes that player 1 knows state G, and K_2K_1G denotes that player 2 knows player 1 knows state G. Also, let $\neg K_iE$ denote that player i does not know an event E. For example, $\neg K_2G$ denotes that player 2 does not know state G, and $\neg K_1K_2G$ denotes that player 1 does not know player 2 knows state G.

$t=(0,0)$: K_1B but $\neg K_2B$.

$t=(1,0)$: K_1G but $\neg K_2G$.

- Player 1 knows a realized state θ.
- Player 2 has not received any email.

$t=(1,1)$: K_1G, K_2K_1G, but $\neg K_1K_2G$.

$t=(2,1)$: K_1G, K_2K_1G, $K_1K_2K_1G$, but $\neg K_2K_1K_2G$.

- Player 1 knows player 2 knows state G if and only if she gets player 2's confirmation.
- Player 2 has the same knowledge (K_2K_1G) as long as he has the same type $t_2=1$.

[2] K_i is called a knowledge operator, which formalizes the meaning of "know." We omit the formal definition.

Note (Two Armies Problem): A similar problem is known as the Coordinated Attack Problem or the Two Armies Problem in computer science. Here is a story modified to my liking.

Armies 1 and 2 attempt to attack the target in the valley. Army 1 learns that the enemy is weak enough that if both armies attack, they can win, but army 2 cannot observe (because of the thick forest).

Army 1 sends a carrier pigeon to let army 2 know the plan of a coordinated attack. However, a pigeon sometimes may fail to arrive at army 2. It might fly somewhere, or it might be shot. Then, even after receiving the pigeon, army 2 worries, "does army 1 know the pigeon arrived?" Army 1 may not attack because it thinks, "if the pigeon doesn't arrive, army 2 doesn't know the attack plan and doesn't attack." Then, army 2 decides to send back the pigeon (to let army 1 know army 2 understand the plan). Even if the pigeon comes back to army 1, army 1 still worries, "does army 2 know the pigeon arrived? If army 2 doesn't know that, it may not attack." To confirm, army 1 sends the pigeon again... *How many times does the pigeon have to go back and forth (for both armies to launch a coordinated attack)?*

Note that if player 2 knows player 1 knows state G, then he knows state G. That is, $K_2K_1G \Rightarrow K_2G$. Hence, at any type profile $t \geq 1$, both players know state G (but it does not mean that they know they know state G, etc.).

Type t_i represents the "height" of player i's knowledge. For each $t_i \geq 1$, player i's higher-order knowledge is as follows:

$$\underbrace{K_1K_2K_1K_2\cdots K_2K_1}_{t_1 \geq 1 \text{ number of } K_1\text{'s}} G \qquad \text{or} \qquad \underbrace{K_2K_1K_2K_1\cdots K_2K_1}_{t_2 \geq 1 \text{ number of } K_2\text{'s}} G.$$

As her type t_i gets larger, her knowledge gets "closer" to the common-knowledge situation.

Remark 1. It is impossible that players 1 and 2 obtain the common knowledge of state G. The common knowledge requires that players 1 and 2 exchange emails infinitely many times, but that event occurs with probability zero. □

Coffee Break ♨. It is so confusing!! Pause and ponder. □

1.2 Equilibrium

Complete-Information Game Suppose that a state θ is common knowledge. Then, the game being played is simple. There are two cases to consider:

1. The complete-information game of state G has two Nash equilibria: (A,A) and (N,N).[3]

2. The complete-information game of state B has a unique Nash equilibrium: (N,N).

"Nearby" Complete-Information Game If the error probability ϵ is small, then the email game will be "close" to the complete-information game of state G (with high probability). Since the complete-information game of state G has two pure-strategy Nash equilibrium, we may expect that the email game has those equilibria, but it is not true.

Theorem 1. *For any sufficiently small $\epsilon > 0$, the electronic mail game has a unique Bayesian Nash equilibrium, which plays action profile (N,N) for every type profile t.*[4]

Proof. The proof resembles the mathematical induction. We start from the top-left cell $(0,0)$ (with probability $1-\mu$) and then descend to the bottom-right cells in Table 3.

$t=(0,0)$**:** Player 1 of type $t_1 = 0$ knows state B, so that she plays N. Player 2 has a posterior

$$\mathbb{P}(\theta = B \mid t_2 = 0) = \frac{1-\mu}{1-\mu+\mu\epsilon} \approx 1,$$

so that player 2 of type $t_2 = 0$ also plays N.

$t=(1,0)$**:** Player 1 of type $t_1 = 1$ has a posterior

$$\mathbb{P}(t_2 = 0 \mid t_1 = 1) = \frac{\mu\epsilon}{\mu\epsilon + \mu(1-\epsilon)\epsilon} = \frac{1}{2-\epsilon} > \frac{1}{2}.$$

Hence, player 1 assigns probability more than $\frac{1}{2}$ to player 2 playing N (by case $t=(0,0)$). If player 1 plays A then her payoff is below $-\frac{1}{2}$, but if she plays N then her payoff is zero. Hence, player 1 of type $t_1 = 0$ plays N.

$t=(1,1)$**:** Player 2 of type $t_2 = 1$ has a posterior

$$\mathbb{P}(t_1 = 1 \mid t_2 = 1) = \frac{\mu(1-\epsilon)\epsilon}{\mu(1-\epsilon)\epsilon + \mu(1-\epsilon)^2\epsilon} = \frac{1}{2-\epsilon} > \frac{1}{2}.$$

Hence, player 2 assigns probability more than $\frac{1}{2}$ to player 1 playing N (by case $t=(1,0)$). If player 2 plays A then his payoff is below $-\frac{1}{2}$, but if he plays N then his payoff is zero. Hence, player 2 of type $t_2 = 1$ plays N.

$t=(2,1)$**:** Player 1 of type $t_1 = 2$ has a posterior

$$\mathbb{P}(t_2 = 1 \mid t_1 = 2) = \frac{\mu(1-\epsilon)^2\epsilon}{\mu(1-\epsilon)^2\epsilon + \mu(1-\epsilon)^3\epsilon} = \frac{1}{2-\epsilon} > \frac{1}{2}.$$

[3] We restrict our attention to pure strategies.

[4] As the proof shows, the equilibrium is indeed a unique rationalizable strategy profile.

Hence, player 1 assigns probability more than $\frac{1}{2}$ to player 2 playing N (by case $t = (1, 1)$). If player 1 plays A then her payoff is below $-\frac{1}{2}$, but if she plays N then her payoff is zero. Hence, player 1 of type $t_1 = 2$ plays N.

This argument continues ad infinitum.[5] By induction, we complete the proof. ■

Remark 2. Theorem 1 holds for every $\mu \neq 1$. This means that even if state G is realized with ex-ante probability 99%, if there is a 1% risk of state B, the Nash equilibrium (A, A) in the complete-information game of state G is fragile. □

Remark 3. The induction argument in the proof is called the **infection argument** or the **contagion argument**. Playing N is strictly dominant for type $t_i = 0$. Playing N is infected to the next type $t_1 = 1$. Playing N is again infected to the next type $t_2 = 1$. Playing N is yet again infected to the next type $t_1 = 2$, and so forth. This infection continues ad infinitum. As a result, all the types are infected to play N. □

Implications The electronic mail game has far-reaching implications.

1. If we approximate an "almost" complete-information game by an (exact) complete-information game, then our equilibrium prediction could be misleading, because the two games can have different equilibrium predictions.
    - This leads to the question about when we can approximate an "almost" complete-information game by a complete-information game. That is, how should we formalize the concept of "almost" common knowledge (e.g., Monderer & Samet, 1989)?
2. If we add noise to a complete-information game, then we may select a "robust" Nash equilibrium. Complete-information games often have multiple Nash equilibria, but by adding some noise, we may screen out fragile ones.
    - This leads to the question about which equilibrium is fragile or robust to what kind of noise, and which equilibrium we should select under noise (e.g., Carlsson & van Damme, 1993; Kajii & Morris, 1997; Weinstein & Yildiz, 2007).

1.3 Is More Information Better?

Suppose that neither player obtains any information about a realized state θ. That is, they have to play the game according to their prior. Then, each player i receives payoff $v_i(a) \equiv \mu u_i(a, G) + (1 - \mu) u_i(a, B)$ from action profile a, where $u_i : A \times \Theta \to \mathbb{R}$ is player i's payoff function defined by Tables 1 and 2. This payoff v_i is represented by Table 4.

[5] It is straightforward to formalize the argument by induction. We omit it.

	A	N
A	$3\mu - 2, 3\mu - 2$	$-2, 0$
N	$0, -2$	$0, 0$

Table 4: the game under the prior

If prior μ of state G is high enough that $\mu > \frac{2}{3}$, then there are two Nash equilibria: (A, A) and (N, N). In particular, each player i receives payoff $3\mu - 2 > 0$ from the former equilibrium.

More information does not necessarily make players better-off. If they receive no information then they gain positive payoffs in *some* equilibrium, but if they receive some information (through the email communication) then they gain zero payoff in *any* equilibrium.

References

Carlsson, H., & van Damme, E. (1993). Global games and equilibrium selection. *Econometrica*, *61*(5), 989–1018.

Kajii, A., & Morris, S. (1997). The robustness of equilibria to incomplete information. *Econometrica*, *65*(6), 1283–1309.

Monderer, D., & Samet, D. (1989). Approximating common knowledge with common beliefs. *Games and Economic Behavior*, *1*(2), 170–190.

Rubinstein, A. (1989). The electronic mail game: Strategic behavior under "almost common knowledge". *American Economic Review*, *79*(3), 385–391.

Weinstein, J., & Yildiz, M. (2007). A structure theorem for rationalizability with application to robust predictions of refinements. *Econometrica*, *75*(2), 365–400.