A Course on Inductive Game Theory 3:

Transpersonal Understanding through Social Roles, and Emergence of Cooperation

by M. Kaneko, 2009 March 24

Aim: **Experiential origin/emergence** of **belief/knowledge** of the other's understanding about the game structure.

Key notions:

0: distinction between **persons** and **social roles**

1: **role switching**

2: experiences of both roles

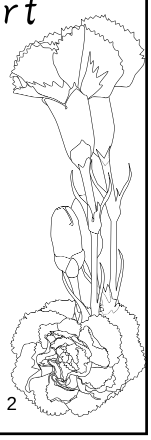
3: **transpersonal** understanding - - projection of one's experiences to the other.

1

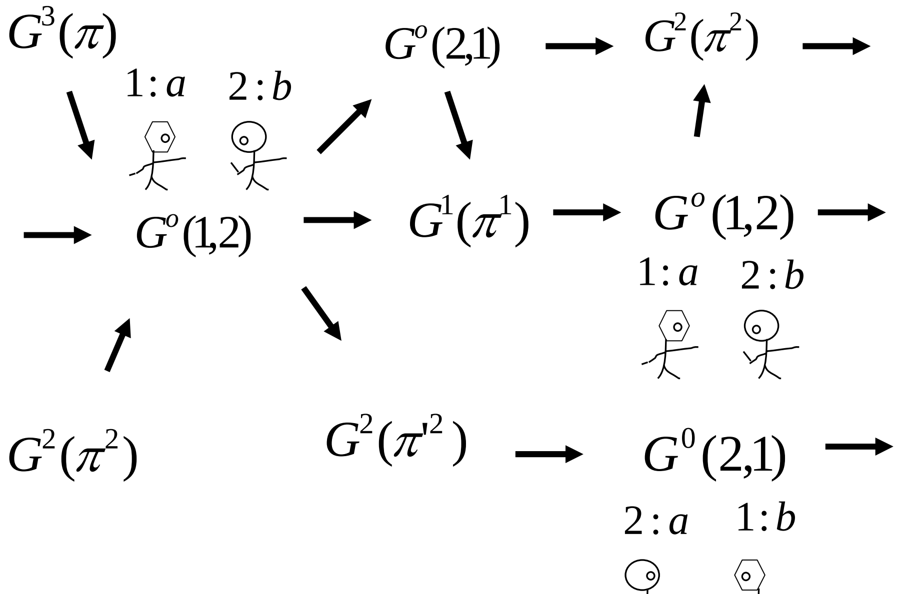
Broken Heart and her Agony

- The agony of a broken heart can only be understood by a person whose heart was once broken;

- yet he doubts her agony because he cannot explain her broken heart.



2



Social Web

3

2-role strategic game

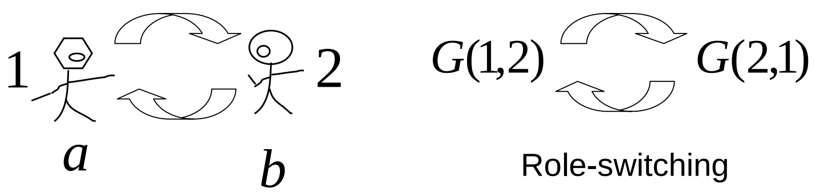
$G = (a, b, S_a, S_b, h_a, h_b)$, where

- a, b are social roles
- S_r is the (finite) set of available actions at roles $r = a, b$
- $h_r : S_a \times S_b \to \mathrm{R}$ is a payoff function at role $r = a, b$.

A role assignment

- A role assignment is a bijection $\pi : \{a, b\} \to \{1,2\}$.
- The 2-person game $G(\pi)$ is determined as

$$G(\pi) = (\pi(a), \pi(b), S_a, S_b, h_a, h_b).$$



Role-switching

4

A memory kit κ_i of person i is given as

$$\langle (s_a^o, s_b^o), (D_{ia}, D_{ib}), (h_{ia}, h_{ib}), (\rho_{ia}, \rho_{ib}) \rangle :$$

- (s_a^o, s_b^o) is the pair of regular actions;
- D_{ir} $(r = a, b)$ is the accumulated domains of experiences with $(s_a^o, s_b^o) \in D_{ia} \cup D_{ib} \subseteq S_a \times S_b$;
- $h_{ir} : D_{ir} \to R$ is the observed payoff function given as $h_{ir}(s_a, s_b) = h_r(s_a, s_b)$ for all $(s_a, s_b) \in D_{ir}$;
- (ρ_{ia}, ρ_{ib}) is a vector of frequency weights with $\rho_{ia}, \rho_{ib} \geq 0$ and $\rho_{ia} + \rho_{ib} = 1$.

5

Examples for the Accumulated Domains

1: Non-reciprocal Active Domains:

$$D_{1a}^N = \{(s_a, s_b^o) : s_a \in S_a\} \text{ and } D_{1b}^N = \phi;$$
$$D_{2a}^N = \phi \text{ and } D_{1b}^N = \{(s_a^o, s_b) : s_b \in S_b\}.$$

2: Reciprocal Active Domains:

$$D_{1a}^A = \{(s_a, s_b^o) : s_a \in S_a\} \text{ and } D_{1b}^A = \{(s_a^o, s_b) : s_b \in S_b\}.$$

3: Reciprocal Active-Passive Domains:

$$D_{1a}^{AP} = D_{1b}^{AP} = \{(s_a, s_b^o) : s_a \in S_a\} \cup \{(s_a^o, s_b) : s_b \in S_b\}.$$

We say that (D_{ia}, D_{ib}) is internally reciprocal iff

$$\text{Proj}(D_{ia}) = \text{Proj}(D_{ib}) \qquad (2.7)$$

where $\text{Proj}(T) := \{(s_a, s_b) \in T : s_a = s_a^o \text{ or } s_b = s_b^o\}$.

6

An Example

	s_{b1}	s_{b2}	s_{b3}
s_{a1}	3,3	10,2	3,1
s_{a2}	2,10	4,4	5,5
s_{a3}	1,3	5,5	4,4

Non-reciprocal Active Domains with $(s_a^o, s_b^o) = (s_{a1}, s_{b1})$:

$D_{1a}^N = \{(s_{a1}, s_{b1}), (s_{a2}, s_{b1}), (s_{a3}, s_{b1})\}$ and $D_{1b}^N = \phi.$

$\rho_{1a} = \rho_{2b} = 1.$

Reciprocal Active Domains:

$D_{1a}^A = \{(s_{a1}, s_{b1}), (s_{a2}, s_{b1}), (s_{a3}, s_{b1})\}$ and $D_{1b}^A = \{(s_{a1}, s_{b1}), (s_{a1}, s_{b2}), (s_{a1}, s_{b3})\}.$

$\rho_{1a} = \rho_{2b} = 1/2.$

Reciprocal Active-Passive Domains:

$D_{1a}^{AP} = D_{1b}^{AP} = \{(s_{a1}, s_{b1}), (s_{a2}, s_{b1}), (s_{a3}, s_{b1})\} \cup \{(s_{a1}, s_{b1}), (s_{a1}, s_{b2}), (s_{a1}, s_{b3})\}.$

$\rho_{1a} = \rho_{2b} = 1/2.$

7

An Informal Theory of Behavior and Accumulation of Memories behind a Memory Kit

Postulates for Behavior and Trials

BH0 (Switching the Roles): The role assignment changes from time to time, which is exogenously given.

BH1 (Regular Actions): Each person behaves regularly following his action s_r^o when he is assigned to role *r*.

BH2 (Occasional Deviations): Once in a while (infrequently), each person, taking role *r*, unilaterally and independently makes a trial deviation s_r from his regular action s_r^o, and then returns to his regular action.

8

Cognitive Postulates

After one play of the game, person *i* receives his short-term memory expressed as

$$\langle r, (s_a, s_b), h_{ir}(s_a, s_b) = h_r(s_a, s_b) \rangle$$

EP1 (Forgetfulness): If experiences are not frequent enough, then they would not be transformed into a long-term memory and disappear from a person's mind.

EP2(Habituation): A local (short-term) memory becomes lasting as a long-term memory in the mind of a person by habituation,

EP3 (Conscious Memorization Effort): A person makes a conscious effort to memorize the result of his own trials. These efforts are successful if they occur frequently enough relative to his trials.

EP4 (Sensitive with Active relative to Passive): A person is more (or not less) sensitive to his own active deviation than he is to his passive experiences.

9

Direct and Transpersonal Understandings

Let a memory kit of person *i* be given as

$$\kappa_i = \langle (s_a^o, s_b^o), (D_{ia}, D_{ib}), (h_{ia}, h_{ib}), (\rho_{ia}, \rho_{ib}) \rangle.$$

- The direct understanding is given as $g^{ii}(\kappa_i) = (a, b, S_a^i, S_b^i, h_a^{ii}, h_b^{ii})$:

ID1^i : $S_r^i = \{ s_r : (s_r ; s_{-r}) \in D_{ia} \cup D_{ib}$ for some $s_{-r} \}$ for $r = a, b$;

ID2^{ii} : for $r = a, b$, h_r^{ii} is defined over $S_a^i \times S_b^i$ as follows :

$$h_r^{ii}(s_a, s_b) = \begin{cases} h_r^{ii}(s_a, s_b) & \text{if } (s_a, s_b) \in D_{ir} \\ \theta_r & \text{otherwise} \end{cases} \quad (3.1)$$

- The transpersonal understanding is given as $g^{ij}(\kappa_i) = (a, b, S_a^i, S_b^i, h_a^{ij}, h_b^{ij})$:

ID2^{ij} : for $r = a, b$, h_r^{ij} is defined over : $S_a^i \times S_b^i$ as follows :

$$h_r^{ij}(s_a, s_b) = \begin{cases} h_{ir}(s_a, s_b) & \text{if } (s_a, s_b) \in D_{ir} \text{ and } (s_a, s_b) \in D_{i(-r)} \\ \theta_r & \text{otherwise} \end{cases} \quad (3.2)$$

10

Exercises:

1: Calculate the following domains in the example of page 7:

a: non-reciprocal domains;

b: reciprocal active domains;

c: reciprocal active-passive domains.

2: Calculate the *d*-understanding and *tp*-understanding for the above three domains.

Non-reciprocal Domains with $(s_a^o, s_b^o) = (\mathrm{s_{a1}}, \mathrm{s_{b1}})$:

$D_{1a}^N = \{(\mathrm{s_{a1}}, \mathrm{s_{b1}}), (\mathrm{s_{a2}}, \mathrm{s_{b1}}), (\mathrm{s_{a3}}, \mathrm{s_{b1}})\}$ and $D_{1b}^N = \phi$.

$\rho_{1\mathrm{a}} = \rho_{2b} = 1.$

	Sb1
Sa1	3, θ_b
Sa2	2, θ_b
Sa3	1, θ_b

g^{11}

	Sb1
Sa1	θ_a, θ_b
Sa2	θ_a, θ_b
Sa3	θ_a, θ_b

g^{12}

11

Transpersonal Postulates for the Other's Thoughts

DP1 (Direct Understanding of the Objective Situation): A person combines his accumulated experiences to construct his view on the situation in question.

TP2 (Projection of Self to the Other): A person projects his own experienced payoff onto the other person if he believes that the other knows his payoff at that experience.

EP3 (Experiential Reason to Believe): A person believes that the other knows a payoff only when the person has a sufficient experiential reason for the other to have the payoff.

12

Inductively Derived View

Let a memory kit of person i be given as

$$\kappa_i = \left\langle (s_a^o, s_b^o), (D_{ia}, D_{ib}), (h_{ia}, h_{ib}), (\rho_{ia}, \rho_{ib}) \right\rangle.$$

The i.d.view $\Gamma^i = \left\langle (s_a^o, s_b^o), (S_a^i, S_b^i), (\rho_{ia}, \rho_{ib}), H^{ii}, H^{ij} \right\rangle$ is given as follows : H^{ii}, H^{ij} are defined by : for all $([s_a, s_b], [t_a, t_b]) \in (S_a^i \times S_b^i)^2$,

- $H^{ii}([s_a, s_b]_a, [t_a, t_b]_b) = \rho_{ia} h_a^{ii}(s_a, s_b) + \rho_{ib} h_b^{ii}(t_a, t_b);$ (4.1)
- $H^{ij}([s_a, s_b]_a, [t_a, t_b]_b) = \rho_{ia} h_b^{ij}(s_a, s_b) + \rho_{ib} h_a^{ij}(t_a, t_b).$ (4.2)

Imagine
1): what are coming next?
2): what is the status of H^{ij}?

Continue…

13

Double Use of H^{ii} and H^{ij}

Suppose :

$$H^{ii}([s_a^o, s_b^o]_a, [s_a^o, s_b^o]_b) < H^{ii}([s_a, s_b^o]_a, [s_a, s_b^o]_b) \quad (4.3)$$

– person i thinks that it would be better to deviate from s_a^o to s_a, provided that person j deviates in the same manner.

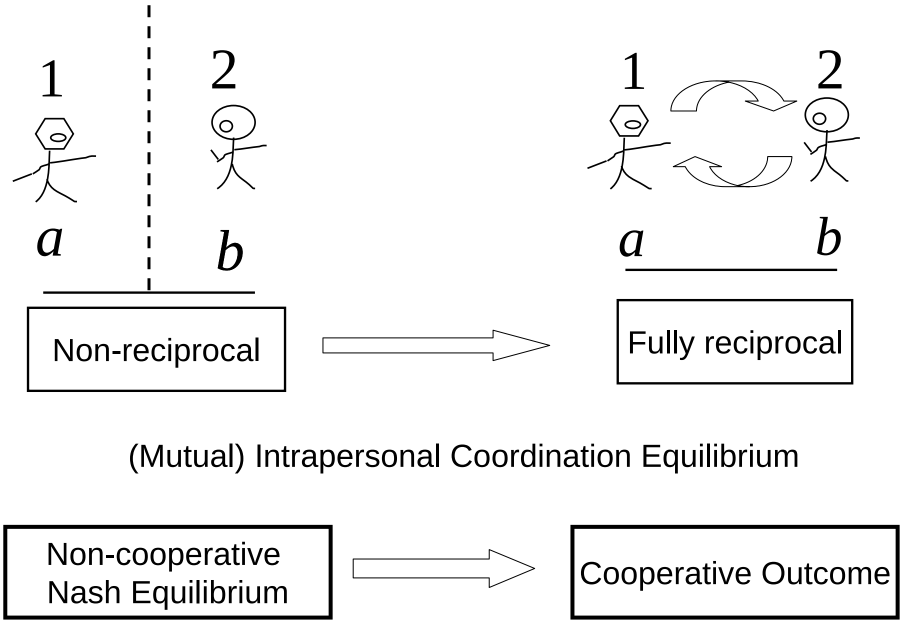
In the mind of person i, (4.3) is justified iff

$$H^{ij}([s_a^o, s_b^o]_a, [s_a^o, s_b^o]_b) < H^{ij}([s_a, s_b^o]_a, [s_a, s_b^o]_b) \quad (4.4)$$

– (in the mind of i) person j thinks that it would be better to deviate from s_a^o to s_a (provided that person i deviates in the same manner).

$$\rightarrow \begin{pmatrix} 1 & 2 \\ s_a^o & s_b^o \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 \\ s_a & s_b^o \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 1 \\ s_a & s_b^o \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 1 \\ s_a & s_b^o \end{pmatrix} \rightarrow$$

14



15

Definition 5.1: I.C.Equilibrium

Let $\Gamma^{i} = \left\langle (s_a^o, s_b^o), (S_a^i, S_b^i), (\rho_{ia}, \rho_{ib}), (H^{ii}, H^{ij}) \right\rangle$ be the i.d.view.

We say that (s_a^o, s_b^o) is an <u>i.c.equilibrium</u> iff for all $s_a \in S_a^i$

$$H^{ii}([s_a^o, s_b^o]_a, [s_a^o, s_b^o]_b) \geq H^{ii}([s_a, s_b^o]_a, [s_a, s_b^o]_b) \qquad (5.1)$$

$$H^{ij}([s_a^o, s_b^o]_a, [s_a^o, s_b^o]_b) \geq H^{ij}([s_a, s_b^o]_a, [s_a, s_b^o]_b)$$

and for all $s_b \in S_b^i$,

$$H^{ii}([s_a^o, s_b^o]_a, [s_a^o, s_b^o]_b) \geq H^{ii}([s_a^o, s_b]_a, [s_a^o, s_b]_b) \qquad (5.2)$$

$$H^{ij}([s_a^o, s_b^o]_a, [s_a^o, s_b^o]_b) \geq H^{ij}([s_a^o, s_b]_a, [s_a^o, s_{bb}^{o}]_b).$$

We say that (s_a^o, s_b^o) is a <u>mutual i.c.equilibrium</u> iff it is an i.c.equilibrium for $i = 1,2$.

16

Theorem 5.1: (Non-reciprocal Active Domain): Consider the non-reciprocal active domain $(D_{ia}^{N}, D_{ib}^{iN})$ where person i takes role a. Then the pair (s_a^o, s_b^o) is an i.c.equilibrium if and only if it is a Nash equilibrium in peron i's d-understanding g^{ii}.

Corollary 5.2: Consider the non-reciprocal active domain $(D_{ia}^{N}, D_{ib}^{iN})$ for $i = 1,2$. Then the pair (s_a^o, s_b^o) of regular actions is a mutual i.c.equilibrium if and only if it is a Nash equilibrium in G.

Reciprocal Domains

Theorem 6.1: (Utilitarian Criterion): Let (s_a^o, s_b^o) be an i.c.equilibrium for Γ^i with $(s_a^o, s_b^o) \in D_{ia} \cap D_{ib}$. Then, it holds that

(1): if $(s_a, s_b^o) \in D_{ia} \cap D_{ib}$, then $h_a(s_a^o, s_b^o) + h_b(s_a^o, s_b^o) \geq h_a(s_a, s_b^o) + h_b(s_a, s_b^o)$;

(2): if $(s_a^o, s_b) \in D_{ia} \cap D_{ib}$, then $h_a(s_a^o, s_b^o) + h_b(s_a^o, s_b^o) \geq h_a(s_a^o, s_b) + h_b(s_a^o, s_b)$.

We say that (D_{ia}, D_{ib}) is internally reciprocal iff $\mathrm{Proj}(D_{ia}) = \mathrm{Proj}(D_{ib})$, where $\mathrm{Proj}(T) := \{(s_a, s_b) \in T : s_a = s_a^o \text{ or } s_b = s_b^o\}$.

Theorem 6.2: (Existence): Let $\rho_{ia} = 1/2$. Then there is a pair (s_a^o, s_b^o) such that for any internally reciprocal domains (D_{ia}, D_{ib}) with $(s_a^o, s_b^o) \in D_{ia}$, the pair (s_a^o, s_b^o) is an i.c.equilibrium in Γ^i.

Prisoner's Dilemma

	s_{b1}	s_{b2}
s_{a1}	5,5	2,6
s_{a2}	6,2	3,3

(h_a, h_b)

IC.Eq. Full Rec.

IC.Eq. N-Rec.

	s_{b1}	s_{b2}
s_{a1}	5,30	2,36
s_{a2}	6,12	3,18

$(h_a, 6 \times h_b)$

Basic Assumption: Each person has no *a priori* knowledge of those payoff functions.
What are implications?

Exercise

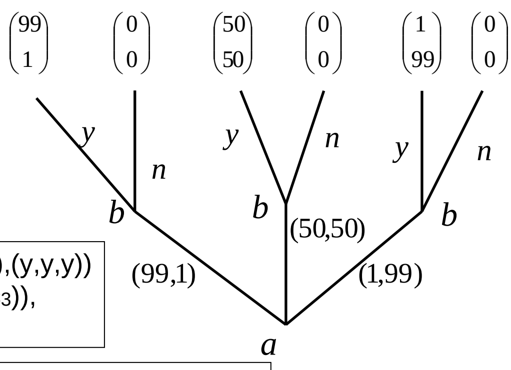
	s_{b1}	s_{b2}
s_{a1}	5,5	2,10
s_{a2}	10,2	3,3

IC.Eq. Full Rec.?

IC.Eq. N-Rec.?

19

Ultimatum Game



Each person has a strictly concave utility function *u* over [0,100].

Backward Induction Solution: ((99,1),(y,y,y))
NE: ((99,1),(y,c_2,c_3)) , ((50,50),(n,y,c_3)), ((1,99),(n,n,y))

I.C.Eq. in Full Rec.: <u>not</u> ((99,1),(y,y,y))
Positive: ((50,50),(y,y,y)) , ((99,1),(y,n,c_3)), ((1,99),(n,n,y))

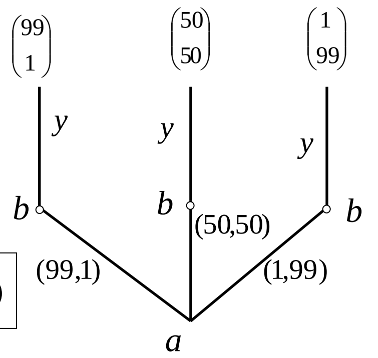
I.C.Eq. in Non-Rec.: NE's

c_1,c_2,c_3 are y or n.

20

Dictator Game

Each person has a strictly concave utility function *u* over [0,100].



Utility Maximization Solution: ((99,1),(y,y,y))

I.C.Eq. in Full Rec.: ((50,50),y)),(y,y,y))
in N-Rec.: ((99,1),(y,y,y))

21

Relationships to other disciplines

- Symbolic Interactionism from Mead (1934)

- the IGT approach is experiential;
- the sources for individual beliefs/knowledge are experiential;
- the sources for the other's thinking are also experiential
  - Lewis's (1969) "reason to believe";

- we follow the tradition of the symbolic interactionism;
- beliefs/knowledge are coming from social interactions - - in the beginning, these experiences are simply sequences of information pieces, but later form some structure with meanings.

22

Relationships to other disciplines

• (Symbolic) Epistemic Logic (Kaneko-Suzuki (2002)) :

$$B_i[\Gamma \to \Theta]$$

• person i has his basic beliefs Γ and infers conclusions Θ .
• the epistemic approach can talk about this derivation but it is incapable to talk about the origin of Γ.
• the IGT approach can talk about the origin/emergence of such starting beliefs Γ in social interactions and experiences.

• New problems in Epistemic Logic - - rejection of logical omniscience - - the sizes of logical inferences.

23

Relationships to other disciplines

• Experimental/Behavioral Economics

• the IGT approach provides a lot of hypothetical propositions to be tested and experimental designs for them.
Furthermore, indirect implications and testable propositions:
• patterned behavior, non-instantaneous maximizations, non-instantaneous logical inferences, etc.
Generally, negation of omniscience.
Dependence of Individual behavior upon the background social context.

Morality in the form of Utilitarianism - - again, experiential and emerging from social interactions (anthropological) – different from
• the rationalistic school of morality (Harsanyi (1953))
• Adam Smith’s moral sentiments - - a human born with such morality.

24