Game Theoretic Decidability and Undecidability

(Based on Hu-Kaneko (2014))

by Mamoru Kaneko,2015/08/19

Part I: Informal Part

Ex ante decision making in a game

- A situation is **interdependent**
- decision making is **Individualistic and Independent**.

Q1: What are possible decisions?

Q2: Does he, in the first place, reach a decision?

Nash (1951) solution theory NE=

#1 PD is solvable

	s_{21}	s_{22}
s_{11}	5 ,5	1, 6
s_{12}	6, 1	3, 3

#2 BS is unsolvable

	s_{21}	s_{22}
s_{11}	2, 1	0, 0
s_{12}	0, 0	1, 2

- How do we evaluate the above Q1 and Q2?

Structure of game theoretical decision making

A: Basic beliefs → **C: decision and prediction**

B: logical inference

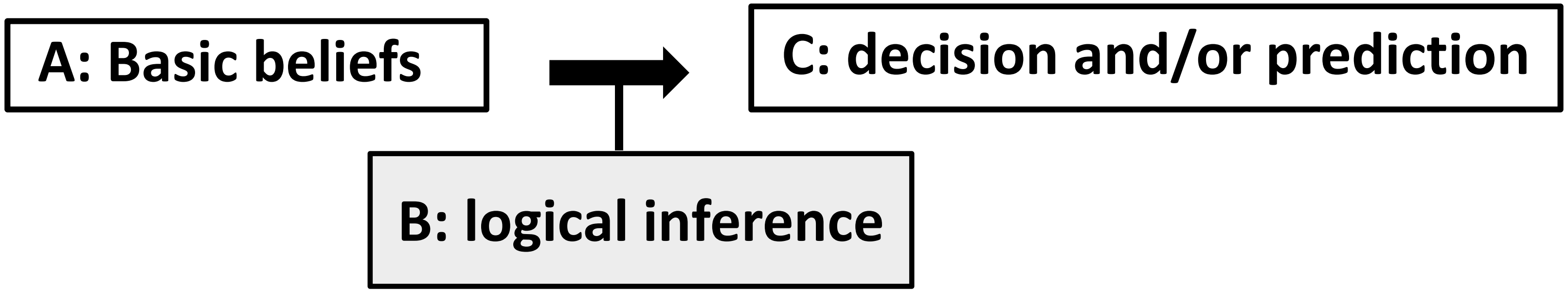
A: 1 Understanding of the situation – beliefs about the game

: 2 Decision/prediction criterion – how he and the other make decision.

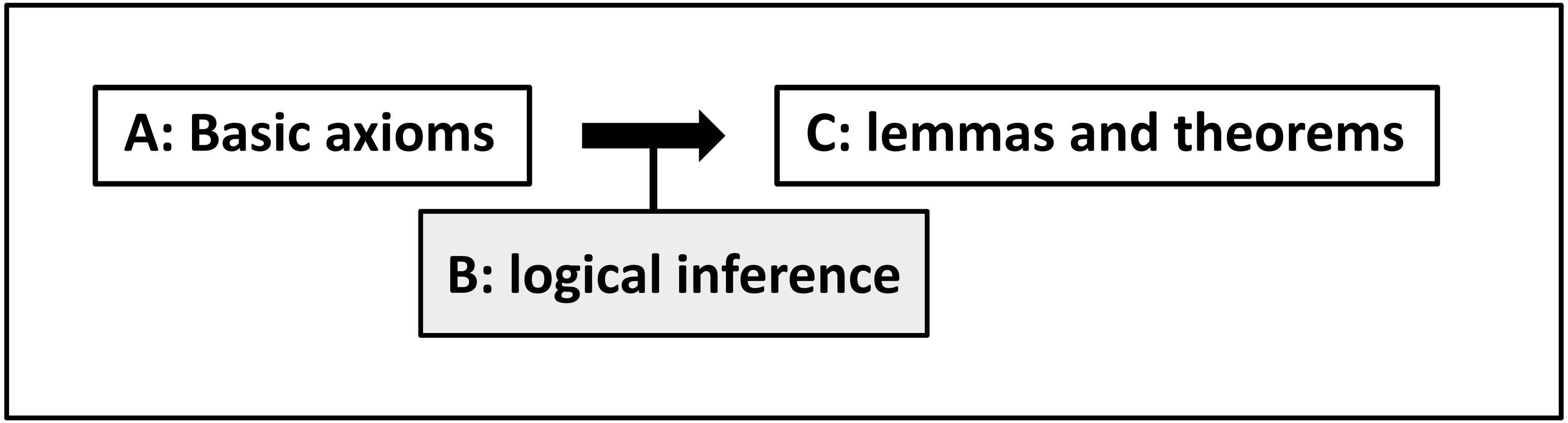
B: Logical abilities of the players

C: possible decisions and predictions

Q2(Does he, in the first place, reach a decision?) is relevant under the consistency of beliefs in A.



- **Parallelism to the Axiomatic method in Mathematics.**



We may recall Gödel's incompleteness theorem.

Recall Q2: Does he, in the first place, reach a decision?

Gödel's incompleteness theorem: Limitation on logical thinking

- **Gödel's incompleteness theorem**: Assuming the **consistency of PA** (Peano arithmetic), there is some (closed) formula A in PA such that

  (*) neither $\mathrm{PA} \vdash A$ nor $\mathrm{PA} \vdash \neg A$,

  - formula A is neither provable nor its negation is provable in PA.
- PA = the natural number theory in classical predicate logic
- An example of A is the formula expressing "consistency" of PA .

- (*) can be viewed as a statement on the ideal mathematician.
- If a game player is an ideal mathematician to make a decision, does he possibly have a similar difficulty?
- Yes, he does, yet **for a different but natural reason**!

Individual Player's Inference

$\mathrm{B_i}(\Gamma_i) \vdash \mathrm{B_i}(A)$;

1: Γ_i - - player i's beliefs - a finite set of (symbolic) **sentences**

2: A - - his deduced consequence from Γ_i;

3: $\vdash$ - - provability: existence of a proof in the logic EIR^2 .

1: Γ_i - player i's beliefs

- understanding of the game situation; preferences, etc.
- Prediction/decision criterion;

  e.g., dominant strategy criterion, but here, Nash theory!

2: A - his decision deduced from Γ_i;

3: Γ_i may include false beliefs, relative to the objective situation.

We take a specific set $\Delta_\mathrm{i}(g)$ of assumptions for as $\mathrm{B_i}(\Gamma_i)$.

Choice of $\Delta_{\mathrm{i}}(g)$: Nash (1951) theory

Decision (prediction) criterion:

Na_1: PL1 should choose a best response against all of his predictions about PL2's decisions based on Na_2:

Na_2: PL2 should choose a best response against all of his predictions about PL1's decisions base on Na_1.

Infinite Regress: PL1's inference for decision making:

$$\mathrm{B}_1(Na_1) \rightarrow \mathrm{B}_1\mathrm{B}_2(Na_2) \rightarrow \mathrm{B}_1\mathrm{B}_2\mathrm{B}_1(Na_1) \rightarrow \cdots$$

- Nash (1951) gave interchangeablity on the set of NE's $E(G)$:

  $E(G) = E(G)_1 \times E(G)_2$ (Product form)
- If G satisfies this condition, the game is *solvable*; and otherwise, it is *unsolvable*.
- "Nash equilibrium" should be distinguished from Nash theory; it is a component of Nash theory.

$\Delta_i(g)$ is the set of basic beliefs

1. Infinite regress of Na_i and Na_j;

$$\mathrm{B}_i(Na_i),\ \mathrm{B}_i\mathrm{B}_j(Na_j),\ \mathrm{B}_i\mathrm{B}_j\mathrm{B}_i(Na_i),\ \cdots$$

2. Infinite regress of g_i and g_j;

$$\mathrm{B}_i(g_i),\ \mathrm{B}_i\mathrm{B}_j(g_j),\ \mathrm{B}_i\mathrm{B}_j\mathrm{B}_i(g_i),\ \cdots$$

3. The axiom choosing the logically weakest formulae

Two cases on $g = (g_1, g_2)$ are crucial: solvable and unsolvable.

- PD

	s_{21}	s_{22}
s_{11}	5 ,5	1, 6
s_{12}	6, 1	3, 3

BS

	s_{21}	s_{22}
s_{11}	2, 1	0, 0
s_{12}	0, 0	1, 2

Decidability Theorem

- **Lemma: $\boldsymbol{\Delta_i(g)}$ is consistent.**

Decidability: Let g be a solvable game. Then, for any strategy $s_i \in S_i$,
either $\Delta_\mathrm{i}(g) \vdash \mathrm{B}_\mathrm{i}(\mathrm{I}_i(s_i))$ or $\Delta_\mathrm{i}(g) \vdash \mathrm{B}_\mathrm{i}(\neg \mathrm{I}_i(s_i))$.

- $\mathrm{I}_i(s_i)$ intends to mean "s_i is a possible decision for i.
- $\Delta_\mathrm{i}(g) \vdash \mathrm{B}_\mathrm{i}(\mathrm{I}_i(s_i))$ - - "i deduces s_i to be a possible decision".
- $\Delta_\mathrm{i}(g) \vdash \mathrm{B}_\mathrm{i}(\neg \mathrm{I}_i(s_i))$ - - "i deduces s_i not to be a possible decision".

	s_{21}	s_{22}
s_{11}	5 ,5	1, 6
s_{12}	6, 1	3, 3

Matching Pennies: No NE's

	s_{21}	s_{22}
s_{11}	1 , -1	-1, 1
s_{12}	-1, 1	1, -1

Undecidability Theorem

Undecidablity: Let g be an unsolvable game. Then, for some strategy $s_i \in S_i$,

$$\text{neither } \Delta_{\mathrm{i}}(g) \vdash \mathrm{B_i}(\mathrm{I}_i(s_i)) \quad \text{nor} \quad \Delta_{\mathrm{i}}(g) \vdash \mathrm{B_i}(\neg \mathrm{I}_i(s_i)).$$

– In the BS game, undecidability holds for either strategy.

	s_{21}	s_{22}
s_{11}	2, 1	0, 0
s_{12}	0, 0	1, 2

Undecidablity on predictions: Let g be an unsolvable game. Then, for some strategy $s_j \in S_j$,

$$\text{neither } \Delta_i(g) \vdash \mathrm{B}_i\mathrm{B}_j(\mathrm{I}_j(s_j)) \quad \text{nor} \quad \Delta_i(g) \vdash \mathrm{B}_i\mathrm{B}_j(\neg \mathrm{I}_j(s_j)).$$

Bridge between the Formalized and non-formalized theories

- $E(G)$ is a the set of NE's.
- A subset F of $E(G)$ is called a ***subsolution*** iff F is a maximal subset of $E(G)$ satisfying interchangeability.

> Let $F^1, \dots, F^k$ be the list of subsolutions of game G.
> Then, $\Delta_{\mathrm{i}}(g) \vdash \mathrm{B_i}(\mathrm{I}_i(s_i)) \Leftrightarrow\ s_i \in F^t$ for all $t = 1, \dots, k$.

BS

	s_{21}	s_{22}
s_{11}	2, 1	0, 0
s_{12}	0, 0	1, 2

	s_{21}	s_{22}
s_{11}	1, 1	0, 1
s_{12}	1, 0	0 , 0

- BS has two subsolutions $\{(s_{11}, s_{21})\}$ and $\{(s_{12}, s_{22})\}$
- The 2nd game also has two: $\{(s_{11}, s_{21}), (s_{12}, s_{21})\}$ and $\{(s_{11}, s_{21}), (s_{11}, s_{22})\}$; they have an intersection $\{(s_{11}, s_{21})\}$.

Gödel's incompleteness theorem

$$\mathrm{PA} \nvdash A \quad \text{nor} \quad \mathrm{PA} \nvdash \neg A$$

- PA: Peano Arithmetic;
- $\vdash$: the provability relation of classical predicate logic;
- PA is assumed to be consistent;
- A is, e.g., "consistency" of PA.

Ours

if g is solvable , either $\Delta_{\mathrm{i}}(g) \vdash \mathrm{B}_{\mathrm{i}}(\mathrm{I}_i(s_i))$ or $\Delta_{\mathrm{i}}(g) \vdash \mathrm{B}_{\mathrm{i}}\big(\neg \mathrm{I}_i(s_i)\big)$.

if g is unsolvable, $\Delta_{\mathrm{i}}(g) \nvdash \mathrm{B}_{\mathrm{i}}(\mathrm{I}_i(s_i))$ and $\Delta_{\mathrm{i}}(g) \nvdash \mathrm{B}_{\mathrm{i}}\big(\neg \mathrm{I}_i(s_i)\big)$.

- $\Delta_i(g)$: beliefs described previously;
- $\vdash$: provability relation of (propositional) **Epistemic Infinite-Regress Logic $\mathbf{EIR^2}$**;
- $\Delta_{\mathrm{i}}(g)$ is proved to be consistent;
- $\mathrm{I}_i(s_i)$ is expressed as
  - "Nash strategy" if g is solvable;
  - no game formula if g is unsolvable.

Axiom T and Common knowledge of a Nash strategy

- Let G be a solvable game.
- If we add Axiom T (truthfulness): $\mathrm{B}_i(A) \supset A$ to our logic,

then we have

2: $\Delta_i(g) \vdash I_i(s_i) \equiv \vee_{t_j} \mathrm{C}[\mathrm{Nash}(s_i; t_j)]$.

- - it is a Nash strategy with common-knowledge.

3: Our undecidability result is obtained in this case.

4: if g is solvable, the theory $(\boldsymbol{EIR^2}(\mathrm{T}), \Delta_i(g))$ is complete;
if g is unsolvable, the theory is incomplete.

◆ **However, we should not include Axiom T for various reasons.**

Two projects: Inductive Game Theory and Epistemic Logic

$\mathrm{B_i}(\Gamma_i) \vdash \mathrm{B_i}(A)$;

0: Γ_i - - player i's beliefs

1: What is the source for $\mathrm{B_i}(\Gamma_i)$?

2: Inductive game theory: we look for the source in experiences.

Characteristics of those projects:

- Symbolic
- Learning - - accumulation of information, for example, “information” is also symbolic
- **the status of semantics?**

Part II: Formal Part

Epistemic Logic KD^2 and its Extension EIR^2

We consider only the 2-person case.

Logical Construction:

1. **Primitive symbols**: preferences expressions and logical connectives
2. **Inductive Definition of formulae**
3. **Logical axioms and inference rules**
4. **Definitions of a proof, and provability**
5. **Non-logical axioms:** individual beliefs

- **Subtle relations between the outside analyst's viewpoint and a player's viewpoint.**

KD^2: Restrictions and Extensions

Epistemic Logics of Shallow depths GL_k $(0 \leq k < \omega)$
-- K-Suzuki ('03)

KD^2

Infinitary extensions
- - K-Nagashim ('96,'97)
- - Hu-K-Suzuki ('15)

Fixed-point extensions
- - Hu-K ('14)

The Epistemic Logic KD^2

Logical axioms for classical logic such as

L1: $A \supset (B \supset A)$;

L2: $(A \supset (B \supset C)) \supset ((A \supset B) \supset (A \supset C))$;

L3: $(\neg A \supset \neg B) \supset ((\neg A \supset B) \supset A))$;

L4: $\bigwedge\Phi \supset A$, where $A \in \Phi$;

L5: $A \supset \bigvee \Phi$, where $A \in \Phi$;

The three inference rules

$\dfrac{A \supset B \quad A}{B}$ MP $\quad \dfrac{\{A \supset B: \ B \in \Phi\}}{A \supset \bigwedge\Phi}$ ($\bigwedge$-rule) $\quad \dfrac{\{A \supset B: \ A \in \Phi\}}{\bigvee\Phi \supset B}$ ($\bigwedge$-rule)

Classical logic

K: $\mathbf{B}_i(\boldsymbol{A} \supset \boldsymbol{C}) \supset (\mathbf{B}_i(\boldsymbol{A}) \supset \mathbf{B}_i(\boldsymbol{C}))$;

D: $\neg\mathbf{B}_i(\neg A \wedge A)$;

$\dfrac{A}{\mathrm{B_i}(A)}$ Necessitation

Language: primitive symbols + formulae

- Preference symbols: $P_i(a_1, a_2; b_1, b_2)$, $a_i, b_i \in S_i$, $i = 1,2$;
- Decision/Prediction symbols: $\mathrm{I}_i(a_i)$, $a_i \in S_i$, $i = 1,2$;
- Logical connectives: $\wedge$(and), $\vee$(or), $\neg$(not), $\supset$ (implies);
- Belief operators: $\mathbf{B}_i(\cdot)$, $i = 1,2$;
- Infinite regress operators: $\mathbf{Ir}_i[\cdot\,;\cdot\,]$, $i = 1,2$.

Intended Interpretations:

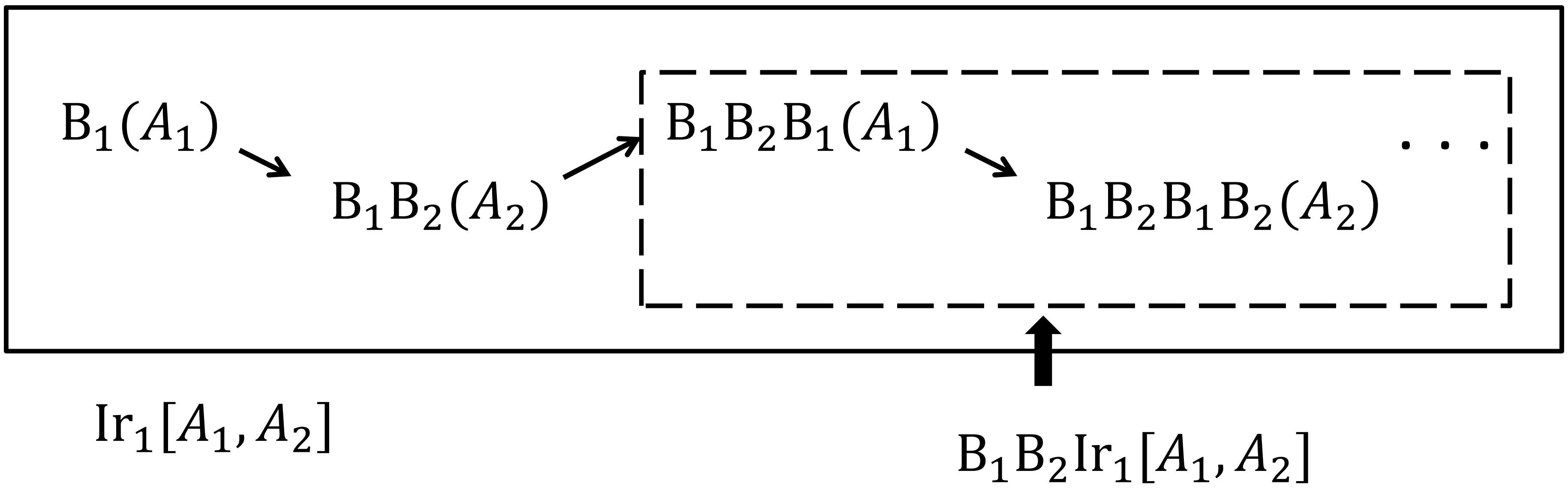
- $P_1(a_1, a_2; b_1, b_2)$: PL1 weakly prefers (a_1, a_2) to (b_1, b_2);
- $\mathrm{B}_1(A)$: PL1 believes A;
- $\mathrm{B}_1(\mathrm{I}_1(a_1))$: PL1 believes that a_1 is a possible decision for him;
- $\mathrm{B}_2\mathrm{B}_1(\mathrm{I}_1(a_1))$: PL2 predicts that a_1 is a possible decision for PL1.

Infinite Regress Formulae

Infinite regress $\mathbf{Ir_1}[A_1, A_2]$,

- To make $\mathbf{B}_1(A_1)$ meaningful, PL1 needs $\mathbf{B_1 B_2}(A_2)$;
- To have the latter, PL1 needs $\mathrm{B}_1\mathrm{B}_2\mathrm{B}_1(A_1)$; so on.

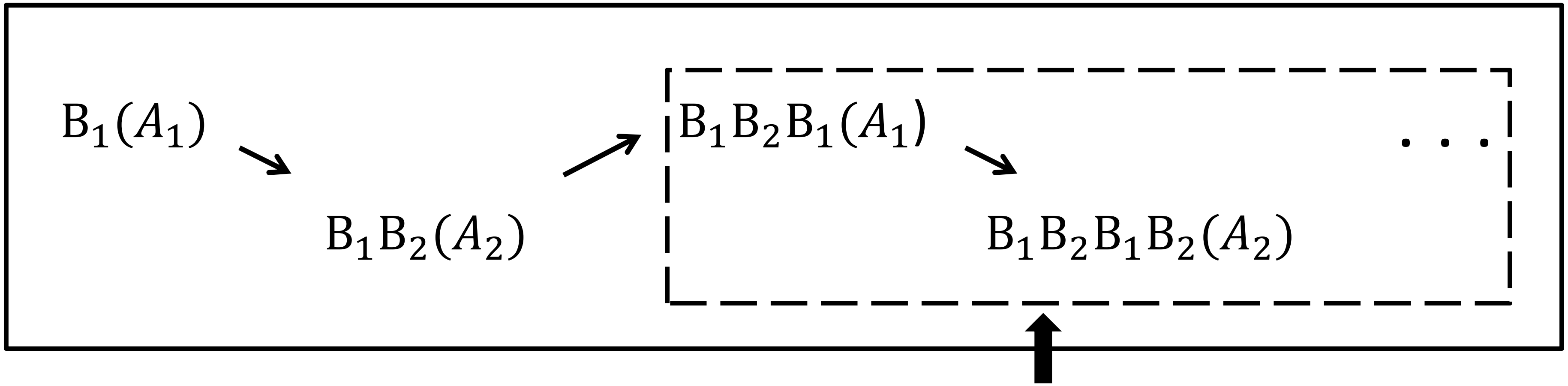
Individual Perspective for PL1:



The Fixed-point logic $\mathrm{IR}^2 = \mathrm{KD}^2$ + the following two.

$\mathrm{IRA_i}$: $\mathrm{Ir_i}[A_1, A_2] \supset \mathrm{B}_i(A_i) \wedge \mathrm{B}_i\mathrm{B}_j(A_j) \wedge \mathrm{B}_i\mathrm{B}_j(\mathrm{Ir}_i[A_1, A_2])$;

$$\frac{D_i \supset \mathrm{B}_i(A_i) \wedge \mathrm{B}_i\mathrm{B}_j(A_j) \wedge \mathrm{B}_i\mathrm{B}_j(D_j)}{D_i \supset \mathrm{Ir}_i[A_1, A_2]}$$ **IRI$_\mathbf{i}$** (choice of the logically weakest)



$\mathrm{Ir}_1[A_1, A_2]$

$\mathrm{B}_1\mathrm{B}_2\mathrm{Ir}_1[A_1, A_2]$

Proof - inference

A **proof** is a triple $(X, <; \psi)$ so that

- $(X, <)$ is a finite tree;
- ψ assigns a formula to each node of X;
- a formula attached to each leaf of $(X, <)$ by ψ is an instance of the logical axioms;
- for each non-leaf $x \in X$,

$$\frac{\{\psi(y): y \text{ is an immediate predecessor of } x\}}{\psi(x)}$$

forms an instance of inference rules.

- A formula A is *provable*, denoted by $\vdash A$, iff there is a proof $(X, <; \psi)$ with $\psi(x_0) = A$, where x_0 is the root of $(X, <; \psi)$.

- $\boldsymbol{\Gamma} \vdash \boldsymbol{A}$ **iff** $\vdash \boldsymbol{A}$ **or there is some finite nonempty subset** $\boldsymbol{\Phi}$ **of** $\boldsymbol{\Gamma}$ **such that** $\vdash \bigwedge\boldsymbol{\Phi} \supset \boldsymbol{A}$**.**

Decision (prediction) criterion:

Na_1: PL1 should choose a best response again all of his predictions about PL2's decisions based on Na_2:

Na_2: PL2 should choose a best response again all of his predictions about PL1's decisions base on Na_1.

These are described, taking beliefs into account, as follows:

- $N0_i$: $\wedge_{s_i \in S_i} [I_i(s_i) \supset \wedge_{s_j \in S_j} (\mathrm{B}_j(I_j(s_j)) \supset \mathrm{Bst}_i(s_i; s_j))]$

Additionally, we need to assume:

- $N1_i$: $\wedge_{s_i \in S_i} [I_i(s_i) \supset \vee_{s_j \in S_j} \mathrm{B}_j(I_j(s_j))]$
- $N2_i$: $\wedge_{s_i \in S_i} [I_i(s_i) \supset \mathrm{B}_j \mathrm{B}_i(I_i(s_i)]$

We denote $N0_i \wedge N1_i \wedge N2_i$ by $\mathrm{N012}_i$. We assume

- $\mathrm{I}r_i(\mathrm{N012}_i; \mathrm{N012}_j)$.

- ☐ $\mathbf{Ir}_i(g_i; g_j)$: infinite regress of the game:
- ☐ $\mathbf{Ir_i}(\mathrm{WF}_i; \mathrm{WF}_j)$: the choice of the deductively weakest $I_i(s_i)$.
- ◆ $\Delta_i(g) = \{\mathbf{Ir}_i(g_i; g_j), \mathbf{Ir}_i(\mathrm{N012}_i; \mathrm{N012}_j)\} \cup \mathbf{Ir_i}(\mathrm{WF}_i; \mathrm{WF}_j)$.

Lemma $\Delta_i(g)$ is consistent in the logic EIR^2.

Theorem 1. Let g be a solvable game. Then, $\Delta_\mathrm{i}(g) \vdash \mathrm{B_i}(\mathrm{I}_i(s_i)) \equiv \mathrm{B_i}(A_i(s_i))$ for some game formula $A_i(s_i)$.

Theorem 2. Let g be a solvable game. Then, for any strategy $s_i \in S_i$, either $\Delta_\mathrm{i}(g) \vdash \mathrm{B_i}(\mathrm{I}_i(s_i))$ or $\Delta_\mathrm{i}(g) \vdash \mathrm{B_i}(\neg \mathrm{I}_i(s_i))$.

Theorem 3. Assume Axiom T. Let g be a solvable game. Then, the theory (EIR^2(T), $\Delta_\mathrm{i}(g)$) is complete. i.e., for any A, $\Delta_\mathrm{i}(g) \vdash A$ or $\Delta_\mathrm{i}(g) \vdash \neg A$.

Theorem 4: Let g be an unsolvable game. Then, for some strategy $s_i \in S_i$, neither $\Delta_\mathrm{i}(g) \vdash \mathrm{B_i}(\mathrm{I}_i(s_i))$ nor $\Delta_\mathrm{i}(g) \vdash \mathrm{B_i}(\neg \mathrm{I}_i(s_i))$.

How should we interpret the decidability or undecidability result?

From the viewpoint of purely *ex ante* decision making

even in an interdependent situation;

- individualistic and independent decision making is
  - possible if the game is solvable;
  - Impossible if it is unsolvable.

- In a wider situation, one can bring his observation on the other's previous action ➔ Inductive game theory

References

- Hu, T., and M. Kaneko (2012), Critical Comparisons between the Nash Noncooperative Theory and Rationalizability, Logic and Interactive Rationality Yearbook 2012, Vol.II, eds. Z. Christo, et al. 203-226, http://www.illc.uva.nl/dg/?page_id=78
- Hu, T., and M. Kaneko (2014), Epistemic Infinite Regress Logic
- Hu, T., and M. Kaneko, & N.-Y. Suzuki (2014), Small Infinitary Epistemic Logics and Some Fixed-Point Logics, to be completed.
- Kaneko, M., (2002), Epistemic logics and their game theoretical applications: Introduction. *Economic Theory* 19, 7-62.
- Kaneko, M., (2004), *Game Theory and Mutual Misunderstanding*, Springer.
- Kaneko, M., J. J. Kline (2008), Inductive Game Theory: A Basic Scenario, *Journal of Mathematical Economics* 44, (2008), 1332--1363.
- Kaneko, M., and T. Nagashima, (1996,'97), Game logic and its applications Part I, *Studia Logica* 57, 325—354; and Part II 58 (1997), 273—303.
- Kaneko, M., and N.-Y. Suzuki, (2003), Epistemic models of shallow depths and decision making in games: Horticulture, (with N-Y. Suzuki), Journal of Symbolic Logic 68, 163-186.
- Nash, J. F., (1951), Non-cooperative Games, *Annals of Mathematics* 54, 286-295.
- Johansen, L., (1982), On the Status of the Nash Type of Noncooperative Equilibrium in Economic Theory, *Scand. J. of Economics* 84, 421-441.

- General Induction to the projects: 『ゲーム理論と蒟蒻問答』日本評論社、（2003）