Prediction/Decision Making in Epistemic Logic

Tai-Wei Hu

(based on papers with Mamoru Kaneko)

Northwestern University

SAET conference 2014, Tokyo, August 19, 2014

Outline

- Prediction and undecidability
- Nash theory: epistemic analysis
- Infinite regress logic
- Undecidability in Nash theory

Prediction and undecidability

Prediction/decision making in game theory

Payoff interdependence

- one player's optimal choice depends on other players' actions
- prediction about others' actions crucial to one's decision

Prediction/decision making in game theory

Payoff interdependence

- one player's optimal choice depends on other players' actions
- prediction about others' actions crucial to one's decision

Battle of Sexes

	Board Game	*Hiking*
Board Game	(3, 2)	(0, 0)
Hiking	(0, 0)	(2, 3)

How to make predictions?

How to make predictions?

Give up making predictions

- dominant strategy criterion, default choice

How to make predictions?

Give up making predictions

- dominant strategy criterion, default choice

Prediction by induction from past experiences

- treating players as nature and use probability distributions
- evolutionary game theory/learning theory

How to make predictions?

Give up making predictions

- dominant strategy criterion, default choice

Prediction by induction from past experiences

- treating players as nature and use probability distributions
- evolutionary game theory/learning theory

Prediction by inferences

- infer others' actions from their preferences and decision methods
- *ex ante* prediction-making is a process of logical inferences

Formal theory of inferences: proof theory

Proof theory treats "proofs" as mathematical objects

- a proof is a sequence of symbols, each element is either an *axiom*, or is derived from preceding elements following a *rule*
- a sentence A is provable, denoted by $\vdash A$, if a proof for A exists

Formal theory of inferences: proof theory

Proof theory treats "proofs" as mathematical objects

- a proof is a sequence of symbols, each element is either an *axiom*, or is derived from preceding elements following a *rule*
- a sentence A is provable, denoted by $\vdash A$, if a proof for A exists

Proof theory connected to model theory by completeness theorem

- completeness: for all sentences A,

$$\vdash A \text{ if and only if } A \text{ is "true" in every model}$$

Formal theory of inferences: proof theory

Proof theory treats "proofs" as mathematical objects

- a proof is a sequence of symbols, each element is either an *axiom*, or is derived from preceding elements following a *rule*
- a sentence A is provable, denoted by $\vdash A$, if a proof for A exists

Proof theory connected to model theory by completeness theorem

- completeness: for all sentences A,

$$\vdash A \text{ if and only if } A \text{ is "true" in every model}$$

Our proof theory approach highlights an undecidability result for prediction/decision making in games, using model theory as a tool

Undecidability (incompleteness)

Gödel's undecidability (incompleteness) theorem (1931): in a formal theory of arithmetic, Γ, there is a sentence A such that

$$\Gamma \nvdash A \text{ and } \Gamma \nvdash \neg A$$

- Γ, a set of consistent (nonlogical) axioms about arithmetic
- Φ is *decidable* (*complete*), if for all A, $\Phi \vdash A$ or $\Phi \vdash \neg A$
- Gödel proves that Γ is undecidable (incomplete)

Undecidability (incompleteness)

Gödel's undecidability (incompleteness) theorem (1931): in a formal theory of arithmetic, Γ, there is a sentence A such that

$$\Gamma \nvdash A \text{ and } \Gamma \nvdash \neg A$$

- Γ, a set of consistent (nonlogical) axioms about arithmetic
- Φ is *decidable* (*complete*), if for all A, $\Phi \vdash A$ or $\Phi \vdash \neg A$
- Gödel proves that Γ is undecidable (incomplete)

When undecidability arises, a player may get stuck in the reasoning process without reaching a satisfactory decision

Logical inferences and interpersonal beliefs

Logical inferences in game situations

Logical inferences and interpersonal beliefs

Logical inferences in game situations

- *ex ante* considerations require subjective inference for each player

Logical inferences and interpersonal beliefs

Logical inferences in game situations

- *ex ante* considerations require subjective inference for each player
- one player's inference may require simulated inferences for others

Logical inferences and interpersonal beliefs

Logical inferences in game situations

- *ex ante* considerations require subjective inference for each player
- one player's inference may require simulated inferences for others

Epistemic logic: proof-theoretical approach to prediction-making in games

- *belief operators* to model a player's subjective scope
- *epistemic axioms* to model simulated inferences

Logical inferences and interpersonal beliefs

Logical inferences in game situations

- *ex ante* considerations require subjective inference for each player
- one player's inference may require simulated inferences for others

Epistemic logic: proof-theoretical approach to prediction-making in games

- *belief operators* to model a player's subjective scope
- *epistemic axioms* to model simulated inferences

Players make decisions and predictions based on beliefs about preferences and decision criterion

Prediction/decision criterion

Decision criterion based on payoff maximization w.r.t. predictions

- “good” decision if best response against predicted actions from others
- independent decision-making: take *all* predictions into account

Prediction/decision criterion

Decision criterion based on payoff maximization w.r.t. predictions

- "good" decision if best response against predicted actions from others
- independent decision-making: take *all* predictions into account

Nash theory

- symmetric prediction/decision criterion
- prediction based on inference from other's decision criterion
- requires an infinite regress of beliefs

Prediction/decision criterion

Decision criterion based on payoff maximization w.r.t. predictions

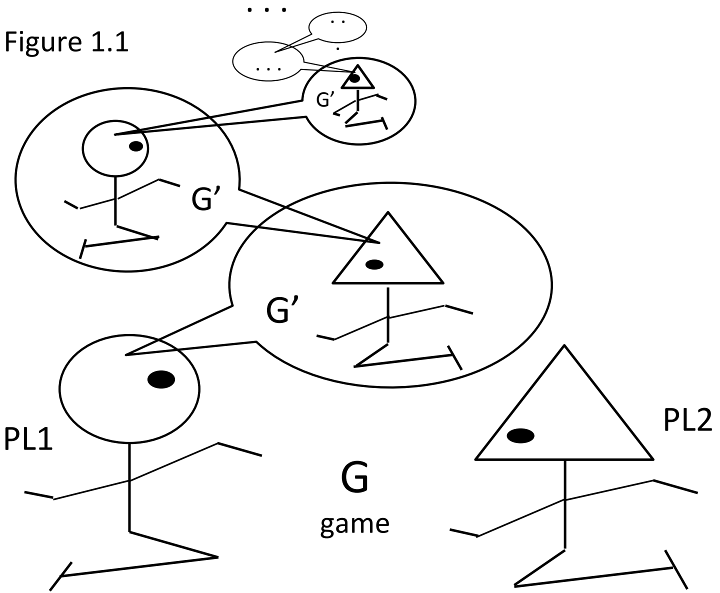
- "good" decision if best response against predicted actions from others
- independent decision-making: take *all* predictions into account

Nash theory

- symmetric prediction/decision criterion
- prediction based on inference from other's decision criterion
- requires an infinite regress of beliefs

Can a player reach a final decision from this infinite regress?

Figure 1.1



Undecidability in prediction/decision making

Let Γ_i represent player i's beliefs (or infinite regress) of preferences and decision criteria and let $\mathrm{I}_1(s_1)$ mean "s_1 is a good decision"

Undecidability in prediction/decision making

Let Γ_i represent player i's beliefs (or infinite regress) of preferences and decision criteria and let $\mathrm{I}_1(s_1)$ mean "s_1 is a good decision"

- Γ_i leads to decidability if for each s_i,
  - $\mathbf{B}_i(\Gamma_i) \vdash \mathbf{B}_i(\mathrm{I}_i(s_i))$ (positive decision), or
  - $\mathbf{B}_i(\Gamma_i) \vdash \mathbf{B}_i(\neg \mathrm{I}_i(s_i))$ (negative decision)

Undecidability in prediction/decision making

Let Γ_i represent player i's beliefs (or infinite regress) of preferences and decision criteria and let $\mathsf{I}_1(s_1)$ mean "s_1 is a good decision"

- Γ_i leads to decidability if for each s_i,
  - $\mathbf{B}_i(\Gamma_i) \vdash \mathbf{B}_i(\mathsf{I}_i(s_i))$ (positive decision), or
  - $\mathbf{B}_i(\Gamma_i) \vdash \mathbf{B}_i(\neg \mathsf{I}_i(s_i))$ (negative decision)
- Γ_i leads to undecidability if for some s_i,
  - $\mathbf{B}_i(\Gamma_i) \nvdash \mathbf{B}_i(\mathsf{I}_i(s_i))$ and $\mathbf{B}_i(\Gamma_i) \nvdash \mathbf{B}_i(\neg \mathsf{I}_i(s_i))$

Undecidability in prediction/decision making

Let Γ_i represent player i's beliefs (or infinite regress) of preferences and decision criteria and let $\mathrm{I}_1(s_1)$ mean "s_1 is a good decision"

- Γ_i leads to decidability if for each s_i,
  - $\mathbf{B}_i(\Gamma_i) \vdash \mathbf{B}_i(\mathrm{I}_i(s_i))$ (positive decision), or
  - $\mathbf{B}_i(\Gamma_i) \vdash \mathbf{B}_i(\neg \mathrm{I}_i(s_i))$ (negative decision)
- Γ_i leads to undecidability if for some s_i,
  - $\mathbf{B}_i(\Gamma_i) \nvdash \mathbf{B}_i(\mathrm{I}_i(s_i))$ and $\mathbf{B}_i(\Gamma_i) \nvdash \mathbf{B}_i(\neg \mathrm{I}_i(s_i))$

We characterize

- the class of games for which Nash theory leads to decidability
- the class of games for which Nash theory leads to undecidability

Example: decidable case

	L	R_1	R_2
U	$(5, 5)$	$(1, 0)$	$(1, 0)$
D_1	$(0, 1)$	$(2, -2)$	$(-2, 2)$
D_2	$(0, 1)$	$(-2, 2)$	$(2, -2)$

Example: decidable case

	L	R_1	R_2
U	$(5, 5)$	$(1, 0)$	$(1, 0)$
D_1	$(0, 1)$	$(2, -2)$	$(-2, 2)$
D_2	$(0, 1)$	$(-2, 2)$	$(2, -2)$

Under Nash theory,

- $\mathbf{B}_1(\Gamma_1) \vdash \mathbf{B}_1(\mathrm{I}_1(U))$
- $\mathbf{B}_1(\Gamma_1) \vdash \mathbf{B}_1(\neg \mathrm{I}_1(D_1)) \wedge \mathbf{B}_1(\neg \mathrm{I}_1(D_2))$

Example: undecidable case

	L	R
U	$(3, 2)$	$(0, 0)$
D	$(0, 0)$	$(2, 3)$

Example: undecidable case

	L	R
U	(3, 2)	(0, 0)
D	(0, 0)	(2, 3)

Under Nash theory,

- $\mathbf{B}_1(\Gamma_1) \nvdash \mathbf{B}_1(\mathsf{I}_1(U))$, $\mathbf{B}_1(\Gamma_1) \nvdash \mathbf{B}_1(\neg\mathsf{I}_1(U))$
- $\mathbf{B}_1(\Gamma_1) \nvdash \mathbf{B}_1(\mathsf{I}_1(D))$, $\mathbf{B}_1(\Gamma_1) \nvdash \mathbf{B}_1(\neg\mathsf{I}_1(D))$

Nash Theory

Nash solution of noncooperative games

$G = \langle \{1,2\}, \{S_1, S_2\}, \{h_1, h_2\} \rangle$, a two-person finite game

Nash solution of noncooperative games

$G = \langle \{1,2\}, \{S_1, S_2\}, \{h_1, h_2\} \rangle$, a two-person finite game

- $E \subseteq S_1 \times S_2$ is interchangeable iff $E = E_1 \times E_2 \neq \emptyset$
- interchangeability captures independence of players' decision-making

Nash solution of noncooperative games

$G = \langle\{1,2\},\{S_1,S_2\},\{h_1,h_2\}\rangle$, a two-person finite game

- $E \subseteq S_1 \times S_2$ is interchangeable iff $E = E_1 \times E_2 \neq \emptyset$
- interchangeability captures independence of players' decision-making
- E_i describes player i's decisions and E_j describes his predictions

Nash solution of noncooperative games

$G = \langle \{1,2\}, \{S_1, S_2\}, \{h_1, h_2\} \rangle$, a two-person finite game

- $E \subseteq S_1 \times S_2$ is interchangeable iff $E = E_1 \times E_2 \neq \emptyset$
- interchangeability captures independence of players' decision-making
- E_i describes player i's decisions and E_j describes his predictions

Solvable and unsolvable games (Nash, 1951)

- G is solvable if $E(G)$ (the set of Nash equilibria) is interchangeable and $E(G)$ is the solution
- otherwise, G is unsolvable
  - maximal $E \subseteq E(G)$ satisfying interchangeability is a *subsolution*

Decision criterion for Nash solutions

A candidate solution $E = E_1 \times E_2 \subset S$ satisfies

$\mathbf{N}_1$ If $s_1 \in E_1$, then s_1 is a best response against all $s_2 \in E_2$;

$\mathbf{N}_2$ If $s_2 \in E_2$, then s_2 is a best response against all $s_1 \in E_1$.

- for player 1, E_1 describes his "good" decisions and E_2 his predictions
- N_2 and N_2 can be viewed as a system of simultaneous equations

Prediction and interpersonal beliefs

In N_1-N_2 there is no distinction between decisions and predictions

- E_1 occurs in the scope of $\mathbf{B}_1(\cdot)$
- E_2 occurs in the scope of $\mathbf{B}_1\mathbf{B}_2(\cdot)$

Prediction and interpersonal beliefs

In N_1-N_2 there is no distinction between decisions and predictions

- E_1 occurs in the scope of $\mathbf{B}_1(\cdot)$
- E_2 occurs in the scope of $\mathbf{B}_1\mathbf{B}_2(\cdot)$

Derivation using N_1-N_2 requires to the following infinite regress (from player 1's perspective):

$\mathbf{B}_1(N_1)$		$\mathbf{B}_1\mathbf{B}_2\mathbf{B}_1(N_1)$		$\cdots\cdots$
$\downarrow$	$\nearrow$	$\downarrow$	$\nearrow$	$\downarrow$
$\mathbf{B}_1\mathbf{B}_2(N_2)$		$\mathbf{B}_1\mathbf{B}_2\mathbf{B}_1\mathbf{B}_2(N_2)$		$\cdots\cdots$

Derivation of final decisions

Positive decision: $\mathbf{B}_1(\Gamma_1) \vdash \mathbf{B}_1(\mathsf{I}_1(s_1))$

Negative decisions: $\mathbf{B}_1(\Gamma_1) \vdash \mathbf{B}_1(\neg \mathsf{I}_1(s_1))$

- $\mathsf{I}_1(s_1)$ means "s_1 is a good decision"
- Γ_1 includes
  - 1's belief about his decision criterion (N_1) and his preferences (g_1)
  - his belief about 2's belief about N_2 and g_2
  - his belief about 2's belief about his belief about N_1 and g_1, etc.

Derivation of final decisions

Positive decision: $\mathbf{B}_1(\Gamma_1) \vdash \mathbf{B}_1(\mathrm{I}_1(s_1))$

Negative decisions: $\mathbf{B}_1(\Gamma_1) \vdash \mathbf{B}_1(\neg\mathrm{I}_1(s_1))$

- $\mathrm{I}_1(s_1)$ means "s_1 is a good decision"
- Γ_1 includes
  - 1's belief about his decision criterion (N_1) and his preferences (g_1)
  - his belief about 2's belief about N_2 and g_2
  - his belief about 2's belief about his belief about N_1 and g_1, etc.

Undecidability: neither positive nor negative decision can be reached

$$\mathbf{B}_1(\Gamma_1) \nvdash \mathbf{B}_1(\mathrm{I}_1(s_1)) \text{ and } \mathbf{B}_1(\Gamma_1) \nvdash \mathbf{B}_1(\neg\mathrm{I}_1(s_1))$$

Infinite regress logic

Infinite regress logic IR^2

Language

- propositional variables: $\mathbf{p}_0$, $\mathbf{p}_1$, ...
- logical connectives: $\neg$, $\supset$, $\wedge$, $\vee$
- unary belief operators: $\mathbf{B}_1(\cdot)$, $\mathbf{B}_2(\cdot)$
- infinite regress operators: $\mathbf{Ir}_1(\cdot,\cdot)$, $\mathbf{Ir}_2(\cdot,\cdot)$

Infinite regress logic IR^2

Language

- propositional variables: $\mathbf{p}_0$, $\mathbf{p}_1$, ...
- logical connectives: $\neg$, $\supset$, $\wedge$, $\vee$
- unary belief operators: $\mathbf{B}_1(\cdot)$, $\mathbf{B}_2(\cdot)$
- infinite regress operators: $\mathbf{Ir}_1(\cdot,\cdot)$, $\mathbf{Ir}_2(\cdot,\cdot)$

Subjective perspectives

- $\mathbf{B}_i(A)$ means "i believes in A"
- $\mathbf{Ir}_i(A_i; A_j)$ means "i believes in A_i, i believes that j believes in A_j, i believes j believes i believes...."

Infinite regress and common knowledge

$\mathbf{Ir}_i(A_i; A_j)$ intends to capture

$$\mathbf{B}_i(A_i),\ \mathbf{B}_i\mathbf{B}_j(A_j),\ \mathbf{B}_i\mathbf{B}_j\mathbf{B}_i(A_i),\ \ldots$$

Infinite regress and common knowledge

$\mathbf{Ir}_i(A_i; A_j)$ intends to capture

$$\mathbf{B}_i(A_i),\ \mathbf{B}_i\mathbf{B}_j(A_j),\ \mathbf{B}_i\mathbf{B}_j\mathbf{B}_i(A_i),\ \ldots$$

$\mathbf{C}(A)$ (common knowledge of A) captures

$$A,\ \mathbf{B}_1(A),\ \mathbf{B}_2(A),\ \mathbf{B}_1\mathbf{B}_2(A),\ \mathbf{B}_2\mathbf{B}_1(A),\ \ldots$$

Infinite regress and common knowledge

$\mathbf{Ir}_i(A_i; A_j)$ intends to capture

$$\mathbf{B}_i(A_i),\ \mathbf{B}_i\mathbf{B}_j(A_j),\ \mathbf{B}_i\mathbf{B}_j\mathbf{B}_i(A_i),\ \ldots$$

$\mathbf{C}(A)$ (common knowledge of A) captures

$$A,\ \mathbf{B}_1(A),\ \mathbf{B}_2(A),\ \mathbf{B}_1\mathbf{B}_2(A),\ \mathbf{B}_2\mathbf{B}_1(A),\ \ldots$$

- $\mathbf{C}(A)$ is an objective notion, formulated from the analyst's perspective
- $\mathbf{Ir}_i(A_i; A_j)$ is a subjective concept, formulated from i's perspective

Epistemic axioms

Axioms and rules from epistemic logic

- K: $\mathbf{B}_i(A \supset B) \supset (\mathbf{B}_i(A) \supset \mathbf{B}_i(B))$
- D: $\neg\mathbf{B}_i(A \wedge \neg A)$
- NEC: from A infers $\mathbf{B}_i(A)$

Axiom and rule for $\mathbf{Ir}_i(\mathbf{A})$

- IRA_i : $\mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{B}_i(A_i) \wedge \mathbf{B}_i\mathbf{B}_j(A_j) \wedge \mathbf{B}_i\mathbf{B}_j\mathbf{Ir}_j(\mathbf{A})$
- IRI_i : from $D_i \supset \mathbf{B}_i(A_i) \wedge \mathbf{B}_i\mathbf{B}_j(A_j) \wedge \mathbf{B}_i\mathbf{B}_j(D_i)$ infer $D_i \supset \mathbf{Ir}_i(\mathbf{A})$

Epistemic axioms

Axioms and rules from epistemic logic

- K: $\mathbf{B}_i(A \supset B) \supset (\mathbf{B}_i(A) \supset \mathbf{B}_i(B))$
- D: $\neg\mathbf{B}_i(A \wedge \neg A)$
- NEC: from A infers $\mathbf{B}_i(A)$

Axiom and rule for $\mathbf{Ir}_i(\mathbf{A})$

- IRA_i : $\mathbf{Ir}_i(\mathbf{A}) \supset \mathbf{B}_i(A_i) \wedge \mathbf{B}_i\mathbf{B}_j(A_j) \wedge \mathbf{B}_i\mathbf{B}_j\mathbf{Ir}_j(\mathbf{A})$
- IRI_i : from $D_i \supset \mathbf{B}_i(A_i) \wedge \mathbf{B}_i\mathbf{B}_j(A_j) \wedge \mathbf{B}_i\mathbf{B}_j(D_i)$ infer $D_i \supset \mathbf{Ir}_i(\mathbf{A})$

A is provable, denoted $\vdash A$, if there is a sequence of formulae such that either each item is an axiom (or tautology) or is derived from previous items using inference rules

Undecidability in Nash Theory

Nash theory in IR^2

Given a finite 2-person game, $G = (\{S_1, S_2\}, \{h_1, h_2\})$, we use the following symbols to describe payoffs and decision/prediciton:

atomic preference formulae: $\mathrm{Pr}_i(s; t)$ for $i = 1, 2$, and $s, t \in S$

atomic decision/prediction formulae: $\mathrm{I}_i(s_i)$ for $s_i \in S_i$, $i = 1, 2$

- $\mathrm{Pr}_i(s; t)$ means that s is weakly preferred to t by player i
- $\mathrm{I}_i(s_i)$ means that s_i is a "good" decision for i
- $\mathbf{B}_j(\mathrm{I}_j(s_j))$ captures i's prediction that s_j is a "good" decision for j

Best responses and Nash equilibrium can be expressed by the Pr_i's

Prediction/decision criterion

Formalize N1-N2 in IR^2:

N0$_i$: $\wedge_{s \in S}[\mathsf{I}_i(s_i) \supset \langle \mathbf{B}_j(\mathsf{I}_j(s_j)) \supset \mathrm{best}_i(s_i; s_j)\rangle]$;

N1$_i$: $\wedge_{s_i \in S_i}[\mathsf{I}_i(s_i) \supset \mathbf{B}_j\mathbf{B}_i(\mathsf{I}_i(s_i))]$;

N2$_i$: $\wedge_{s_i \in S_i}[\mathsf{I}_i(s_i) \supset \vee_{s_j \in S_j}\mathbf{B}_j(\mathsf{I}_j(s_j))]$.

- $\mathrm{N0}_i$ corresponds directly to N_i, but distinguishes decisions from predictions
- $\mathrm{N1}_i$ assume correct predictability
- $\mathrm{N2}_i$ corresponds to non-emptiness of E_1 and E_2

Prediction/decision criterion

Formalize N1-N2 in IR^2:

N0$_i$: $\wedge_{s \in S}[\mathrm{I}_i(s_i) \supset \langle \mathbf{B}_j(\mathrm{I}_j(s_j)) \supset \mathrm{best}_i(s_i; s_j)\rangle]$;

N1$_i$: $\wedge_{s_i \in S_i}[\mathrm{I}_i(s_i) \supset \mathbf{B}_j \mathbf{B}_i(\mathrm{I}_i(s_i))]$;

N2$_i$: $\wedge_{s_i \in S_i}[\mathrm{I}_i(s_i) \supset \vee_{s_j \in S_j} \mathbf{B}_j(\mathrm{I}_j(s_j))]$.

- $\mathrm{N0}_i$ corresponds directly to N_i, but distinguishes decisions from predictions
- $\mathrm{N1}_i$ assume correct predictability
- $\mathrm{N2}_i$ corresponds to non-emptiness of E_1 and E_2

Auxiliary axiom WF^i: if a game formula $A_i(s_i)$ (consisting of preference formulae and belief operators) satisfies N0-N2, then it implies $\mathrm{I}_i(s_i)$

Decidability for solvable games

Let $\Delta_i = \{\mathbf{Ir}_i(g_i; g_j),\ \mathbf{Ir}_i(\mathrm{N}_i; \mathrm{N}_j),\ \mathbf{Ir}_i(WF^i; WF^j)\}$

- game formula (g_1, g_2) consists of the preferences in G
- $\mathrm{N}_i = \mathrm{N0}_i \wedge \mathrm{N1}_i \wedge \mathrm{N2}_i$

Decidability for solvable games

Let $\Delta_i = \{\mathbf{Ir}_i(g_i; g_j),\ \mathbf{Ir}_i(\mathrm{N}_i; \mathrm{N}_j),\ \mathbf{Ir}_i(WF^i; WF^j)\}$

- game formula (g_1, g_2) consists of the preferences in G
- $\mathrm{N}_i{=}\mathrm{N0}_i \wedge \mathrm{N1}_i \wedge \mathrm{N2}_i$

Theorem (Decidability for solvable games)

Let G be a solvable game. If s_i is a Nash strategy, then $\Delta_i \vdash \mathbf{B}_i(I_i(s_i))$; otherwise, $\Delta_i \vdash \mathbf{B}_i(\neg I_i(s_i))$.

Decidability for solvable games

Let $\Delta_i = \{\mathbf{Ir}_i(g_i; g_j),\ \mathbf{Ir}_i(\mathrm{N}_i; \mathrm{N}_j),\ \mathbf{Ir}_i(WF^i; WF^j)\}$

- game formula (g_1, g_2) consists of the preferences in G
- N_i=$\mathrm{N0}_i \wedge \mathrm{N1}_i \wedge \mathrm{N2}_i$

Theorem (Decidability for solvable games)

Let G be a solvable game. If s_i is a Nash strategy, then $\Delta_i \vdash \mathbf{B}_i(I_i(s_i))$; otherwise, $\Delta_i \vdash \mathbf{B}_i(\neg I_i(s_i))$.

- for solvable games, players can reach final decisions
- similar decidability result holds for any finite depth prediction criterion (such as dominant strategy criterion)

Undecidability for unsolvable games

Theorem (Undecidability for unsolvable games)

Let G be an unsolvable game. If s_i is not a Nash strategy, then $\Delta_i \vdash \mathbf{B}_i(\neg I_i(s_i))$. However, there exists a Nash strategy s_i such that

$$\Delta_i \nvdash \mathbf{B}_i(I_i(s_i)) \text{ and } \Delta_i \nvdash \mathbf{B}_i(\neg I_i(s_i)).$$

Undecidability for unsolvable games

Theorem (Undecidability for unsolvable games)

Let G be an unsolvable game. If s_i is not a Nash strategy, then $\Delta_i \vdash \mathbf{B}_i(\neg I_i(s_i))$. However, there exists a Nash strategy s_i such that

$$\Delta_i \nvdash \mathbf{B}_i(I_i(s_i)) \text{ and } \Delta_i \nvdash \mathbf{B}_i(\neg I_i(s_i)).$$

- for unsolvable games, players may get stuck in prediction/decision making process
- similar to Gödel's incompleteness theorem, but due to a different source—strategic unpredictability

Literature

Mathematical logic and epistemic logic

- *Introduction to Mathematical Logic* by Mendelson
- *Reasoning About Knowledge* by Fagin et al.
- "Epistemic logics and their game theoretical applications: Introduction," *Economic Theory* (2002) by Kaneko