

Bonanno, Giacomo

**Working Paper**

## Two lectures on the epistemic foundations of game theory

Working Paper, No. 07-2

**Provided in Cooperation with:**

University of California Davis, Department of Economics

*Suggested Citation:* Bonanno, Giacomo (2007) : Two lectures on the epistemic foundations of game theory, Working Paper, No. 07-2, University of California, Department of Economics, Davis, CA

This Version is available at:

<https://hdl.handle.net/10419/31389>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

TWO LECTURES ON THE  
EPISTEMIC FOUNDATIONS  
OF GAME THEORY

Giacomo Bonanno

Department of Economics  
University of California,  
Davis, CA 95616 – 8578  
USA

e-mail: [gbonanno@ucdavis.edu](mailto:gbonanno@ucdavis.edu)  
<http://www.econ.ucdavis.edu/faculty/bonanno>

*Abstract*

*This working paper contains the slides of two invited lectures on the Epistemic Foundations of Game Theory, delivered at the Royal Netherlands Academy of Arts and Sciences (KNAW) on February 8, 2007.*

Royal Netherlands Academy of Arts and Sciences (KNAW)  
Master Class

Amsterdam, February 8th, 2007

# Epistemic Foundations of Game Theory

## Lecture 1

Giacomo Bonanno

(<http://www.econ.ucdavis.edu/faculty/bonanno/>)

# QUESTION:

What strategies can be chosen by *rational* players who *know* the structure of the game and the preferences of their opponents and who *recognize* each other's rationality and knowledge?

Keywords: knowledge, rationality, recognition of each other's knowledge and rationality

# Modular approach

Module 1: representation of belief and knowledge of an individual (Hintikka, 1962; Kripke, 1963).

Module 2: extension to many individuals.  
Common belief and common knowledge  
("recognition of each other's belief / knowledge")

Module 3: definition of rationality in games  
(relationship between choice and beliefs)

**QUESTION: what are the implications of rationality and common belief of rationality in games?**

# Module 1

## representation of beliefs and knowledge of an individual

Finite set of states  $\Omega$  and a binary relation  $\mathcal{B}$  on  $\Omega$ .

$\alpha \mathcal{B} \beta$  means “at state  $\alpha$  the individual considers state  $\beta$  possible”

Notation:  $\mathcal{B}(\omega) = \{\omega' \in \Omega : \omega \mathcal{B} \omega'\}$  set of states considered possible at  $\omega$

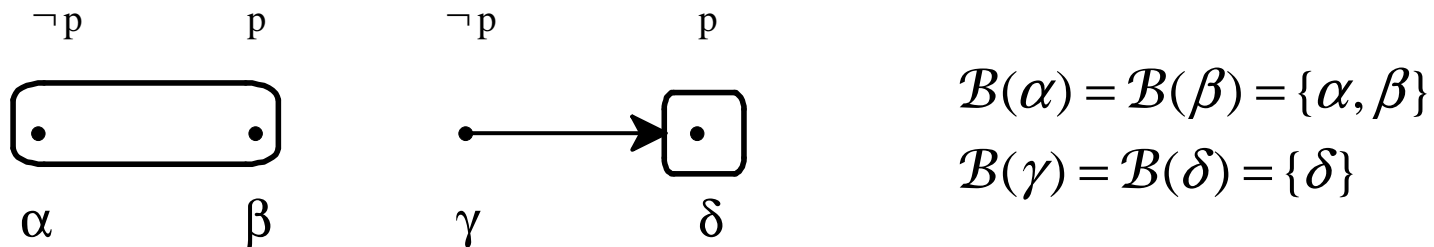
PROPERTIES  $\forall \omega, \omega' \in \Omega,$

1.  $\mathcal{B}(\omega) \neq \emptyset$  seriality
2. if  $\omega' \in \mathcal{B}(\omega)$  then  $\mathcal{B}(\omega') \subseteq \mathcal{B}(\omega)$  transitivity
3. if  $\omega' \in \mathcal{B}(\omega)$  then  $\mathcal{B}(\omega) \subseteq \mathcal{B}(\omega')$  euclidean

Belief operator on events:  $B : 2^\Omega \rightarrow 2^\Omega$

For  $E \subseteq \Omega$ ,  $\omega \in BE$  if and only if  $\mathcal{B}(\omega) \subseteq E$

EXAMPLE:



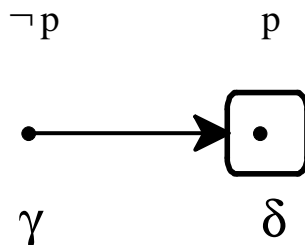
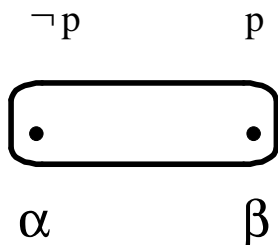
Let  $E = \{\beta, \delta\}$  : the event that represents the proposition  $p$

Then  $BE = \{\gamma, \delta\}$

## Properties of the belief operator: $\forall E \subseteq \Omega$

1.  $BE \subseteq \neg B\neg E$  (consistency:  
follows from seriality of  $\mathcal{B}$ )
2.  $BE \subseteq BBE$  (positive introspection:  
follows from transitivity of  $\mathcal{B}$ )
3.  $\neg BE \subseteq B\neg BE$  (negative introspection:  
follows from euclideaness of  $\mathcal{B}$ )

*Mistaken beliefs are possible:* at  $\gamma$   $p$  is false but the individual believes  $p$

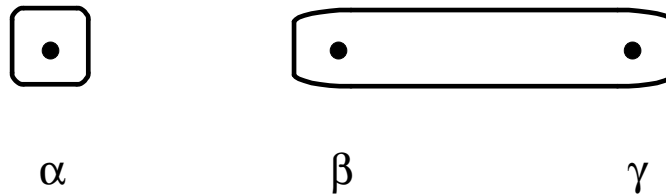


If  $E = \{\beta, \delta\}$ , then  
 $\gamma \notin E$  but  $\gamma \in BE = \{\gamma, \delta\}$



# KNOWLEDGE

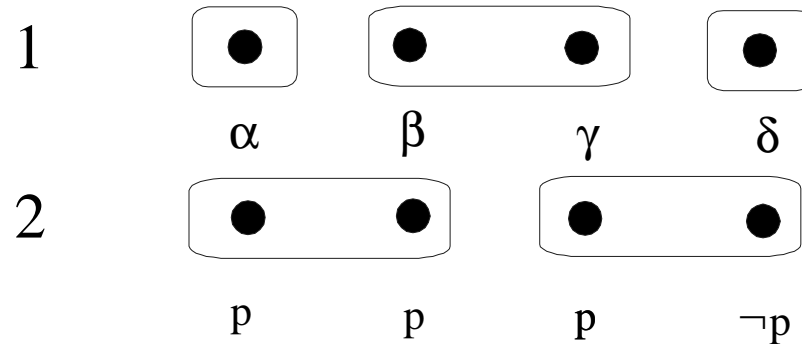
If - in addition to the previous properties - the "doxastic accessibility" relation  $\mathcal{B}$  is *reflexive* ( $\forall \omega \in \Omega, \omega \in \mathcal{B}(\omega)$ ) then it is an *equivalence relation* - giving rise to a *partition* of the set of states - and the associated belief operator satisfies the additional property that  $\forall E \subseteq \Omega, BE \subseteq E$  (beliefs are correct). In this case we speak of *knowledge* and the associated operator is denoted by  $K$  rather than  $B$



## Module 2

### interactive belief and common belief

Set of individuals  $N$  and a binary relation  $\mathcal{B}_i$  for every  $i \in N$



Let  $E = \{\alpha, \beta, \gamma\}$  : the event that represents the proposition  $p$

Then  $K_1 E = \{\alpha, \beta, \gamma\}$ ,  $K_2 E = \{\alpha, \beta\}$

$K_1 K_2 E = \{\alpha\}$ ,  $K_2 K_1 K_2 E = \emptyset$

An event  $E$  is *commonly believed* if (1) everybody believes it, (2) everybody believes that everybody believes it, (3) everybody believes that everybody believes that everybody believes it, etc.

Define the “everybody believes” operator  $B^e$  as follows:

$$B^e E = B_1 E \cap B_2 E \cap \dots \cap B_n E$$

The common belief operator  $B_*$  is defined as follows:

$$B_* E = B^e E \cap B^e B^e E \cap B^e B^e B^e E \cap \dots$$

Let  $\mathcal{B}_*$  be the *transitive closure* of  $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \dots \cup \mathcal{B}_n$

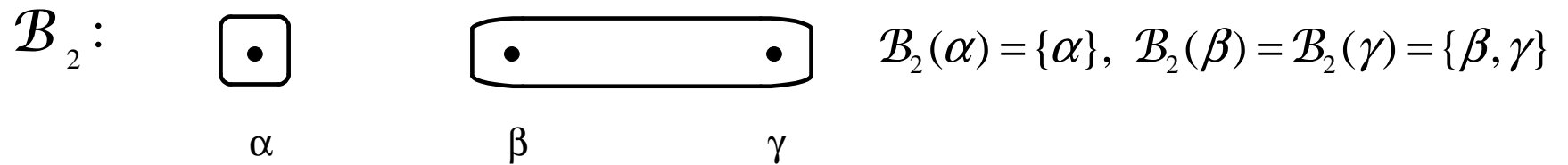
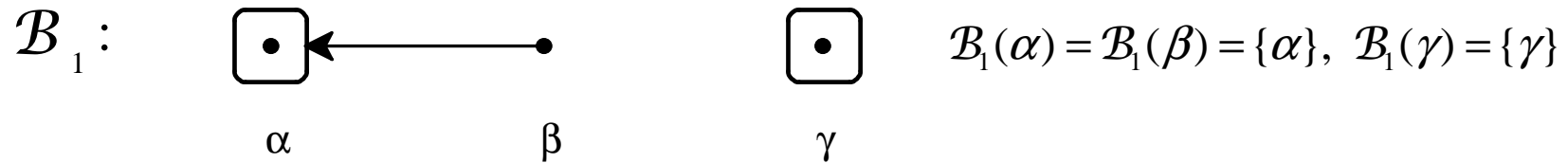
Thus  $\omega' \in \mathcal{B}_*(\omega)$  if and only if there exists a sequence

$\langle \omega_1, \dots, \omega_m \rangle$  in  $\Omega$  such that

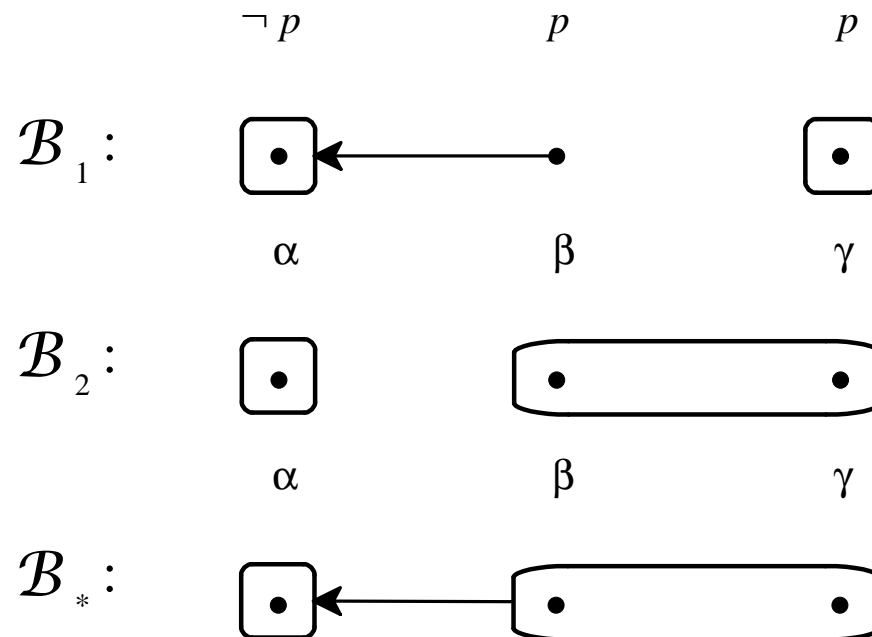
(1)  $\omega_1 = \omega$

(2)  $\omega_m = \omega'$

(3) for every  $j = 1, \dots, m$  there exists an individual  $i \in N$  such that  $\omega_{j+1} \in \mathcal{B}_i(\omega_j)$



**PROPOSITION.**  $\omega \in B_*E$  if and only if  $B_*(\omega) \subseteq E$ .



Let  $E = \{\beta, \gamma\}$  : the event that represents the proposition  $p$

Then  $B_1E = \{\gamma\}$ ,  $B_2E = \{\beta, \gamma\}$ ,  $B_*E = \emptyset$

In fact, while  $\gamma \in B_1B_2E = \{\gamma\}$ ,  $\gamma \notin B_2B_1E = \emptyset$

# Module 3

## Models of games and Rationality

**Definition.** A finite strategic-form game *with ordinal payoffs* is a quintuple

$$\langle N, \{S_i\}_{i \in N}, O, \{\succeq_i\}_{i \in N}, z \rangle$$

$N = \{1, \dots, n\}$  is a set of *players*

$S_i$  is a finite set of *strategies* or choices of player  $i \in N$

$O$  is a set of *outcomes*

$\succeq_i$  is player  $i$ 's ordering of  $O$  ( $o \succeq_i o'$  means that, for player  $i$ , outcome  $o$  is at least as good as outcome  $o'$ )

$z: S \rightarrow O$  (where  $S = S_1 \times \dots \times S_n$ ) associates an outcome with every strategy profile  $s \in S$

**Definition.** Given a strategic-form game with ordinal payoffs

$$\langle N, \{S_i\}_{i \in N}, O, \{\geq_i\}_{i \in N}, z \rangle$$

a reduced form of it is a triple

$$\langle N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N} \rangle$$

where  $u_i : S \rightarrow \mathbb{R}$  is such that  $u_i(s) \geq u_i(s')$  if and only if  $z(s) \geq_i z(s')$

↙ player  $i$ 's utility function

		Player 2		
		e	f	g
P l a y e r  1	A	3, 2	3, 1	0, 1
	B	2, 3	2, 2	3, 1
	C	1, 2	1, 2	4, 1
	D	0, 2	0, 3	1, 3

SAME AS

		Player 2		
		e	f	g
P l a y e r  1	A	9, 6	6, 4	0, 4
	B	4, 9	3, 3	2, 0
	C	2, 5	2, 5	8, 2
	D	1, 0	0, 8	1, 8

**Definition.** An *epistemic model* of a strategic-form game is an interactive belief structure together with  $n$  functions

$$\sigma_i : \Omega \rightarrow S_i \quad (i \in N)$$

Interpretation:  $\sigma_i(\omega)$  is player  $i$ 's chosen strategy at state  $\omega$

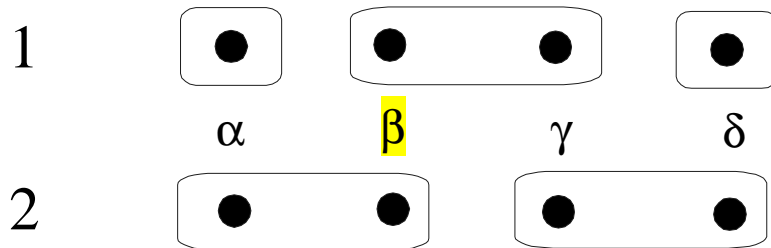
Restriction: if  $\omega' \in \mathcal{B}_i(\omega)$  then  $\sigma_i(\omega') = \sigma_i(\omega)$

(no player has mistaken beliefs about her own strategy)



# EXAMPLE

		Player 2		
		e	f	g
P l a y e r  1	A	3, 2	3, 1	0, 1
	B	2, 3	2, 2	3, 1
	C	1, 2	1, 2	4, 1
	D	0, 2	0, 3	1, 3



At every state each player knows his own strategy

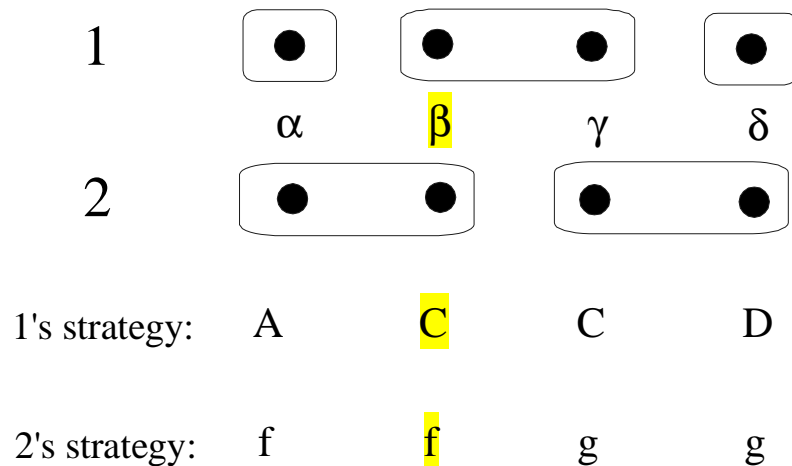
At state  $\beta$  player 1 plays C (he knows this) not knowing whether player 2 is playing f or g and player 2 plays f (she knows this) not knowing whether player 1 is playing A or C

# RATIONALITY

Non-probabilistic (no expected utility) and very weak notion of rationality

**Definition.** Player  $i$  is **IRRATIONAL** at state  $\omega$  if there is a strategy  $s_i$  (of player  $i$ ) which she believes to be better than  $\sigma_i(\omega)$  (that is, if she believes that she can do better with another strategy)

Player  $i$  is **RATIONAL** at state  $\omega$  if and only if she is not irrational



		Player 2		
		e	f	g
P l a y e r 1	A	3, 2	3, 1	0, 1
	B	2, 3	2, 2	3, 1
	C	1, 2	1, 2	4, 1
	D	0, 2	0, 3	1, 3

Player 1 is rational at state  $\beta$

Let  $s_i$  and  $t_i$  be two strategies of player  $i$ :  $s_i, t_i \in \mathcal{S}_i$

$s_i \succ_i t_i$  is interpreted as “strategy  $s_i$  is better for player  $i$  than strategy  $t_i$ ”

$s_i \succ_i t_i$  is true at state  $\omega$  if  $u_i(s_i, \sigma_{-i}(\omega)) > u_i(t_i, \sigma_{-i}(\omega))$

that is,  $s_i$  is better than  $t_i$  against  $\sigma_{-i}(\omega)$

profile of strategies chosen by the players other than  $i$

		Player 2			
		E	F	G	
	$\alpha$	$\beta$	$\gamma$		P l a y e r  1
	●	●	●		
1's strategy:	A	C	C		
2's strategy:	E	F	G		
	A	B	C	A	
	B	C	A	B	
	C	A	B	C	
	E	F	G	E	
	F	G	E	F	
	G	E	F	G	

A	3, 2	1, 1	0, 1
B	2, 3	2, 2	3, 1
C	1, 2	0, 2	4, 1

- $A \succ_1 B$     $B \succ_1 A$     $C \succ_1 B$
- $A \succ_1 C$     $B \succ_1 C$     $C \succ_1 A$
- $B \succ_1 C$     $A \succ_1 C$     $B \succ_1 A$
- $E \succ_2 F$     $F \succ_2 G$     $F \succ_2 G$    etc.

Let  $\|s_i \succ_i t_i\| = \{\omega \in \Omega : u_i(s_i, \sigma_{-i}(\omega)) > u_i(t_i, \sigma_{-i}(\omega))\}$  event that  $s_i$  is better than  $t_i$

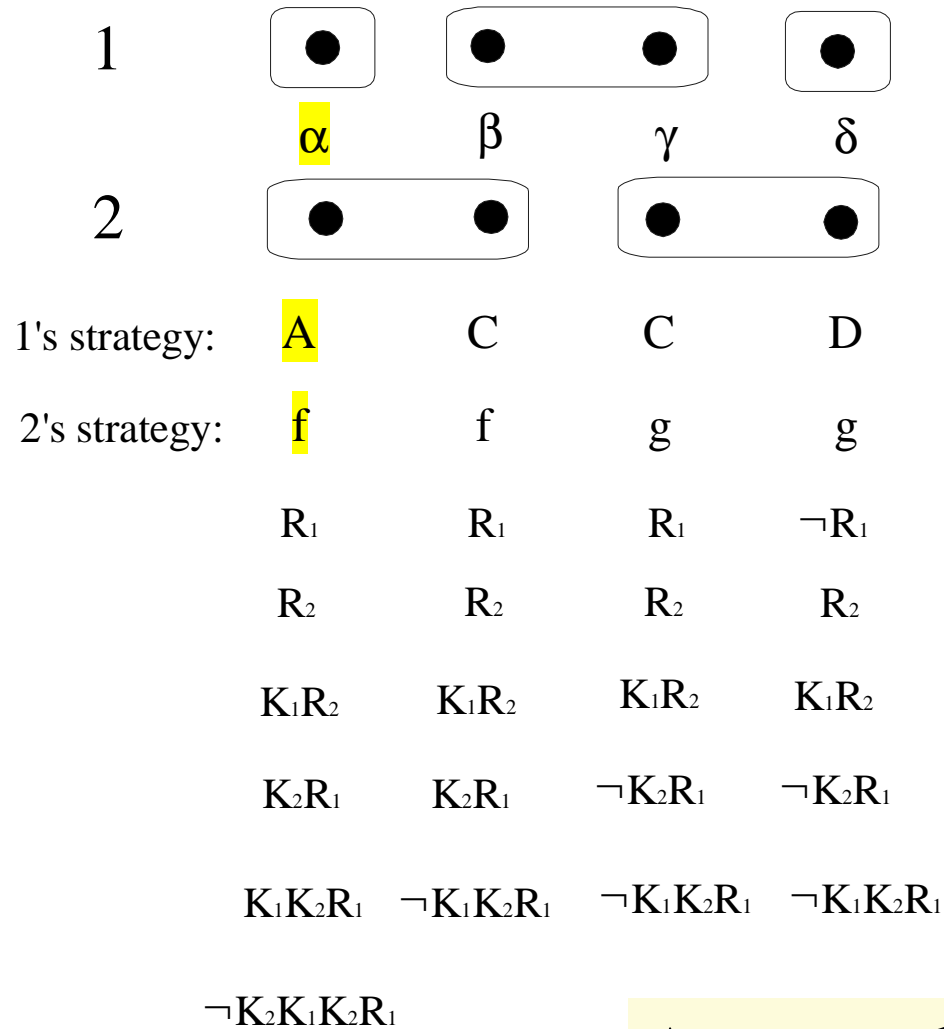
If  $s_i \in S_i$ , let  $\|s_i\| = \{\omega \in \Omega : \sigma_i(\omega) = s_i\}$  event that player  $i$  chooses  $s_i$

Let  $\mathbf{R}_i$  be the event representing the proposition “player  $i$  is rational”

$$\|s_i\| \cap B_i \|t_i \succ_i s_i\| \subseteq \neg \mathbf{R}_i$$

$$\neg \mathbf{R}_i = \bigcup_{s_i \in S_i} \bigcup_{t_i \in S_i} (\|s_i\| \cap B_i \|t_i \succ_i s_i\|)$$

$$\mathbf{R} = \mathbf{R}_1 \cap \dots \cap \mathbf{R}_n \quad \text{all players are rational}$$



Player 2

		e	f	g
Player 1	A	3, 2	3, 1	0, 1
	B	2, 3	2, 2	3, 1
	C	1, 2	1, 2	4, 1
	D	0, 2	0, 3	1, 3

$$R_1 = \{\alpha, \beta, \gamma\}, R_2 = \{\alpha, \beta, \gamma, \delta\}$$

$$K_1R_2 = \{\alpha, \beta, \gamma, \delta\}, K_2R_1 = \{\alpha, \beta\}$$

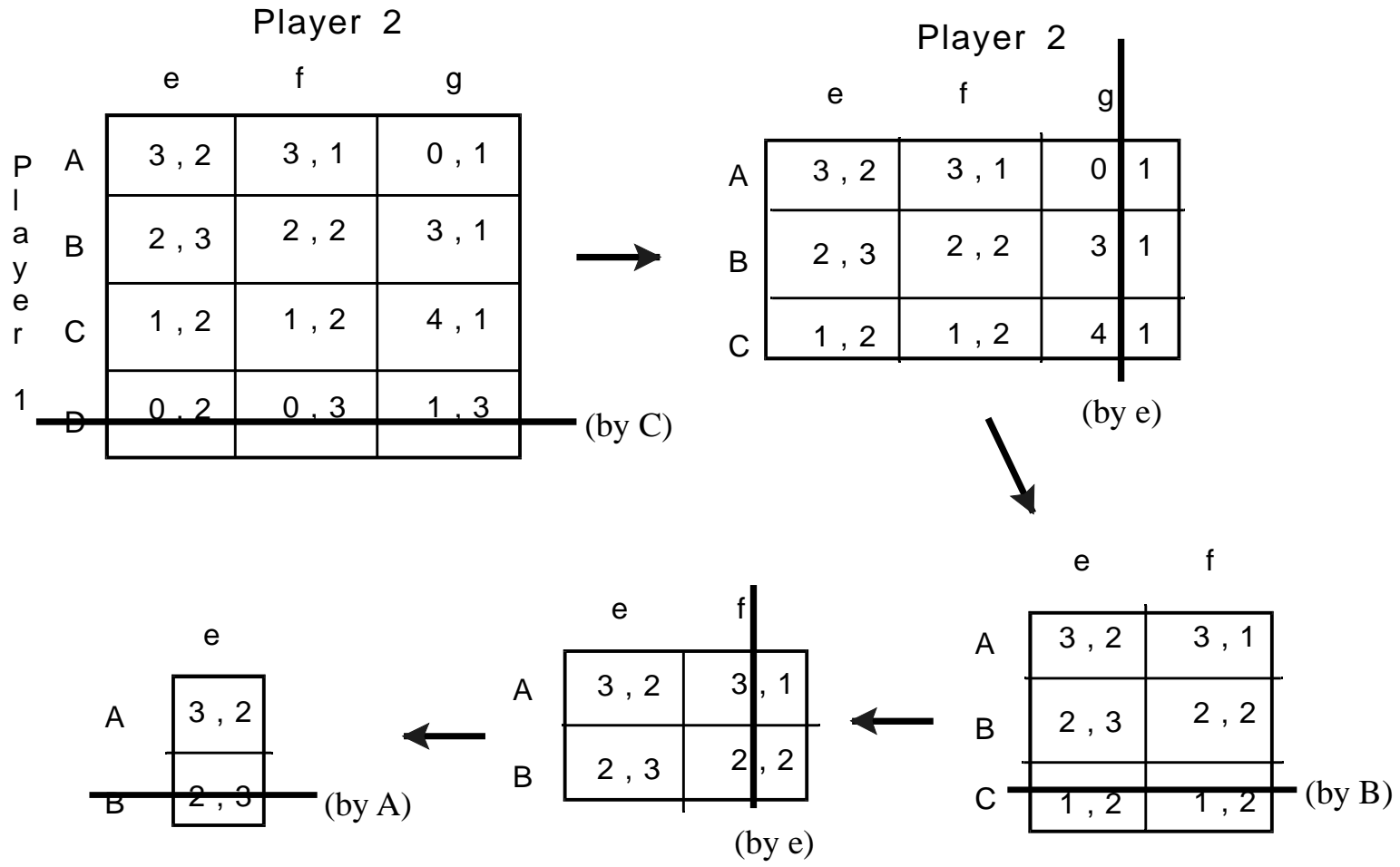
$$K_1K_2R_1 = \{\alpha\}, K_2K_1K_2R_1 = \emptyset$$

At state  $\alpha$  there is mutual knowledge of rationality but not common knowledge of rationality

Let  $S_{-i} = S_1 \times \dots \times S_{i-1} \times S_{i+1} \times \dots \times S_n$  set of strategy profiles of all players except  $i$

**Definition.** Let  $s_i, t_i \in S_i$ . We say that  $t_i$  is *strictly dominated* by  $s_i$  if  $u_i(t_i, s_{-i}) < u_i(s_i, s_{-i})$  for all  $s_{-i} \in S_{-i}$

### ITERATED DELETION OF STRICTLY DOMINATED STRATEGIES



Let  $G$  be a strategic-form game with ordinal payoffs and  $G^\infty$  be the game obtained after applying the procedure of Iterated Deletion of Strictly Dominated Strategies.

Let  $S^\infty$  denote the strategy profiles of game  $G^\infty$

Given a model of  $G$ , let  $S^\infty$  denote the event  $\{\omega \in \Omega : \sigma(\omega) \in S^\infty\}$

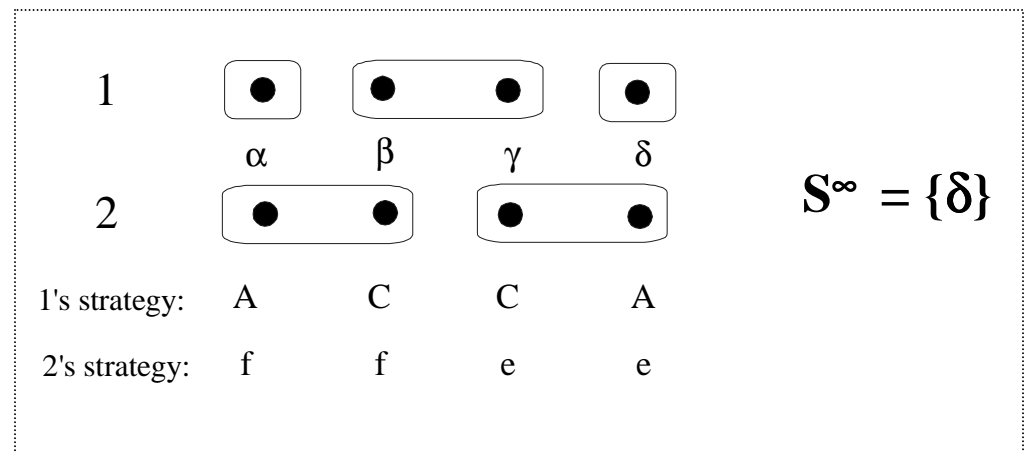
**G**

		Player 2		
		e	f	g
P l a y e r  1	A	3, 2	3, 1	0, 1
	B	2, 3	2, 2	3, 1
	C	1, 2	1, 2	4, 1
	D	0, 2	0, 3	1, 3

**G<sup>∞</sup>**

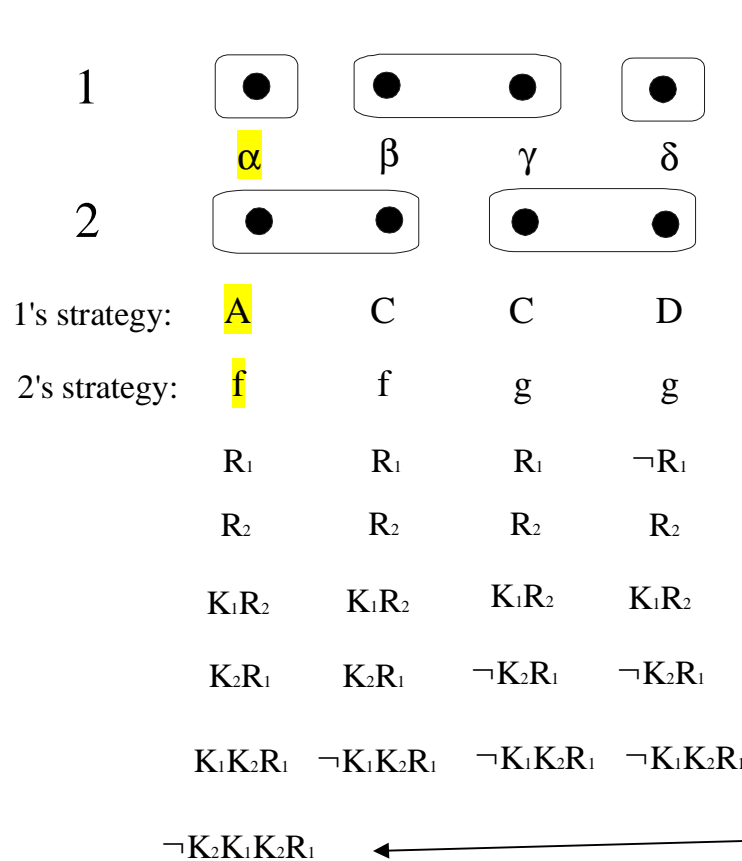
		e
A	3, 2	

$$S^\infty = \{(A, e)\}$$



# PROPOSITION 1. $B_*R \subseteq S^\infty$

If at a state it is commonly believed that all players are rational, then the strategy profile chosen at that state belongs to the game obtained after applying the iterated deletion of strictly dominated strategies.



		Player 2		
		e	f	g
Player 1	A	3, 2	3, 1	0, 1
	B	2, 3	2, 2	3, 1
	C	1, 2	1, 2	4, 1
	D	0, 2	0, 3	1, 3

At state  $\alpha$  there cannot be common knowledge of rationality since  $\sigma(\alpha) \neq (A, e)$



Every normal operator  $B$  satisfies the property that if  $E \subseteq F$  then  $BE \subseteq BF$ .

$B_*$  is a normal operator. Thus from  $B_*R \subseteq S^\infty$

it follows that  $B_*B_*R \subseteq B_*S^\infty$ .

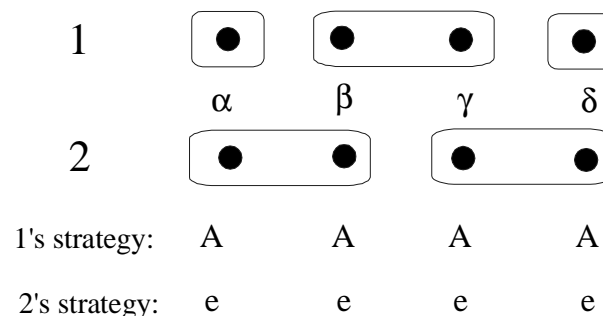
By transitivity of  $\mathcal{B}_*$  we have that

$B_*E \subseteq B_*B_*E$  for every event  $E$ .

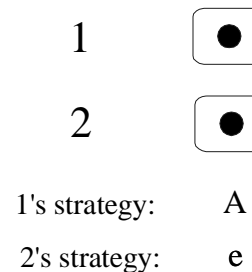
Thus  $B_*R \subseteq B_*B_*R$ .

It follows that  $B_*R \subseteq B_*S^\infty$

		Player 2		
		e	f	g
Player 1	A	3, 2	3, 1	0, 1
	B	2, 3	2, 2	3, 1
	C	1, 2	1, 2	4, 1
	D	0, 2	0, 3	1, 3

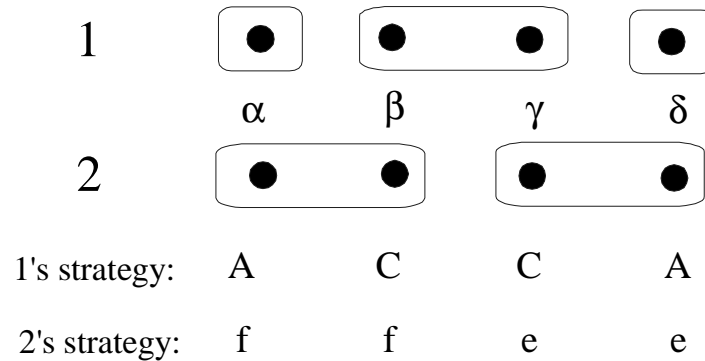


Same as:



REMARK. In general it is not true that  $S^\infty \subseteq B_*R$

		Player 2		
		e	f	g
Player 1	A	3, 2	3, 1	0, 1
	B	2, 3	2, 2	3, 1
	C	1, 2	1, 2	4, 1
	D	0, 2	0, 3	1, 3



$$S^\infty = \{\delta\}$$

$$K_*R = \emptyset$$

$$R_1 = \{\alpha, \delta\}, \quad R_2 = \{\alpha, \beta, \gamma, \delta\}$$

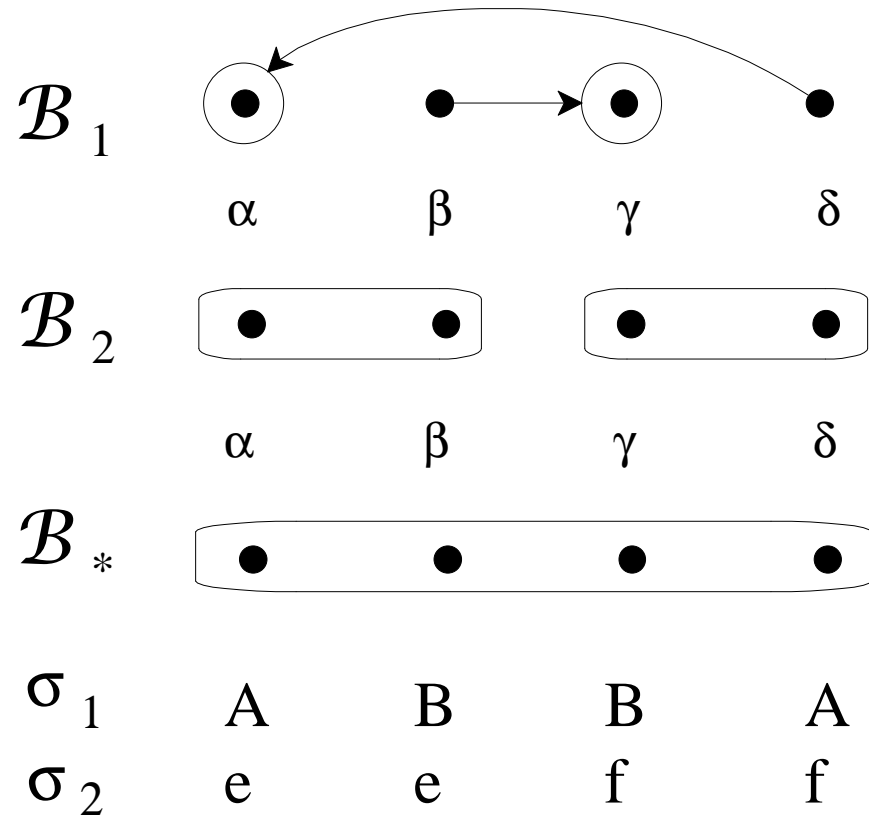
$$K_2R_1 = \emptyset$$

**PROPOSITION 2.** Fix a strategic-form game with ordinal payoffs  $G$  and let  $s \in S^\infty$ . Then there exists an epistemic model of  $G$  and a state  $\omega$  such that  $\sigma(\omega) = s$  and  $\omega \in B_*R$ .

EXAMPLE

		Player 2	
		e	f
P   1	A	3, 2	3, 3
	B	2, 3	4, 2

In this game every strategy profile survives iterative deletion



In this model  $R = B_*R = \Omega$  and every strategy profile occurs at some state

**REMARK.** Given the above notion of rationality, *there is no difference between common belief of rationality and common knowledge of rationality.* The previous two propositions can be restated in terms of knowledge and common knowledge.

**PROPOSITION 1'.**  $K_*R \subseteq S^\infty$

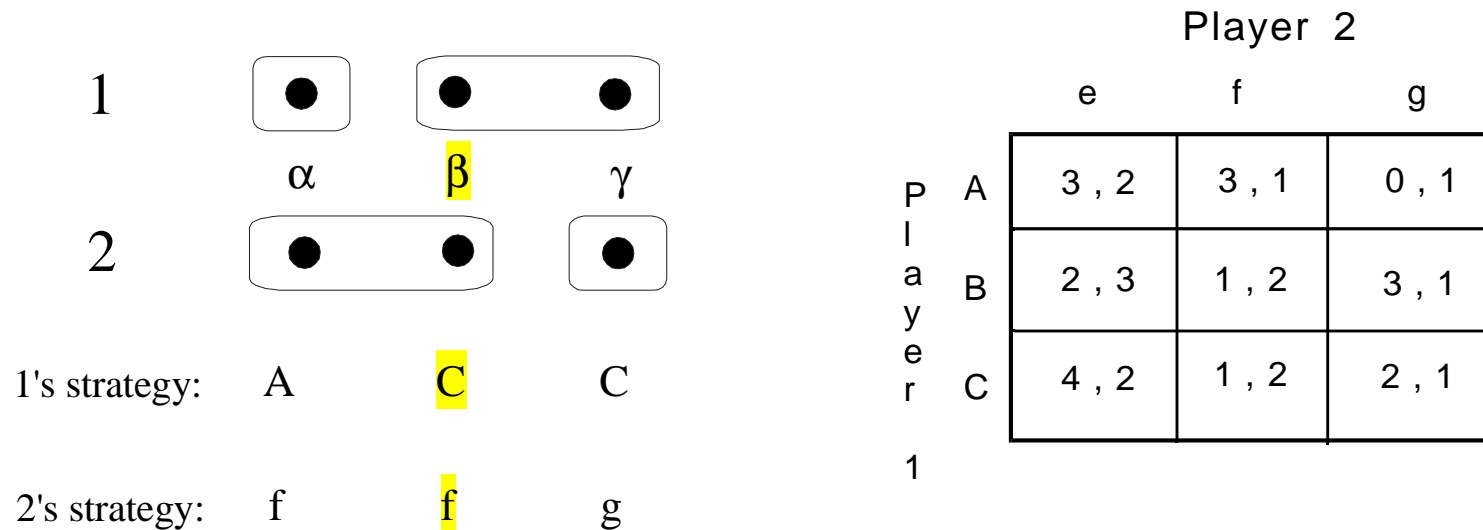
**PROPOSITION 2'.** Fix a strategic-form game with ordinal payoffs  $G$  and let  $s \in S^\infty$ . Then there exists an epistemic model of  $G$  and a state  $\omega$  such that  $\sigma(\omega) = s$  and  $\omega \in K_*R$ .

# STRONGER NOTION OF RATIONALITY

Still non-probabilistic (no expected utility)

**Definition.** Player  $i$  is **IRRATIONAL** at state  $\omega$  if there is a strategy  $s_i$  which she believes to be at least as good as  $\sigma_i(\omega)$  and she considers it possible that  $s_i$  is better than  $\sigma_i(\omega)$

Player  $i$  is **RATIONAL** at state  $\omega$  if and only if she is not irrational



Player 1 is irrational at state  $\beta$ : B is at least as good as C at both  $\beta$  and  $\gamma$  and it is better than C at  $\gamma$

$$R_1 = \{\alpha\}, R_2 = \emptyset$$

Player  $i$  is **IRRATIONAL** at state  $\omega$  if there is a strategy  $s_i$  which she believes to be at least as good as  $\sigma_i(\omega)$  and she considers it possible that  $s_i$  is better than  $\sigma_i(\omega)$

$$\|s_i\| \cap B_i \|t_i \succeq_i s_i\| \cap \neg B_i \neg \|t_i \succ_i s_i\| \subseteq \neg \mathbf{R}_i$$

$$\neg \mathbf{R}_i = \bigcup_{s_i \in S_i} \bigcup_{t_i \in S_i} (\|s_i\| \cap B_i \|t_i \succeq_i s_i\| \cap \neg B_i \neg \|t_i \succ_i s_i\|)$$

$$\mathbf{R} = \mathbf{R}_1 \cap \dots \cap \mathbf{R}_n \quad \text{all players are rational}$$

**Definition.**

Given a game  $G = \langle N, \{S_i\}_{i \in N}, O, \{\succeq\}_{i \in N}, z \rangle$ , a subset of strategy profiles  $X \subseteq S$  and a strategy profile  $x \in X$ , we say that  $x$  is **inferior relative to  $X$**  if there exist a player  $i$  and a strategy  $s_i \in S_i$  of player  $i$  (thus  $s_i$  need not belong to the projection of  $X$  onto  $S_i$ ) such that:

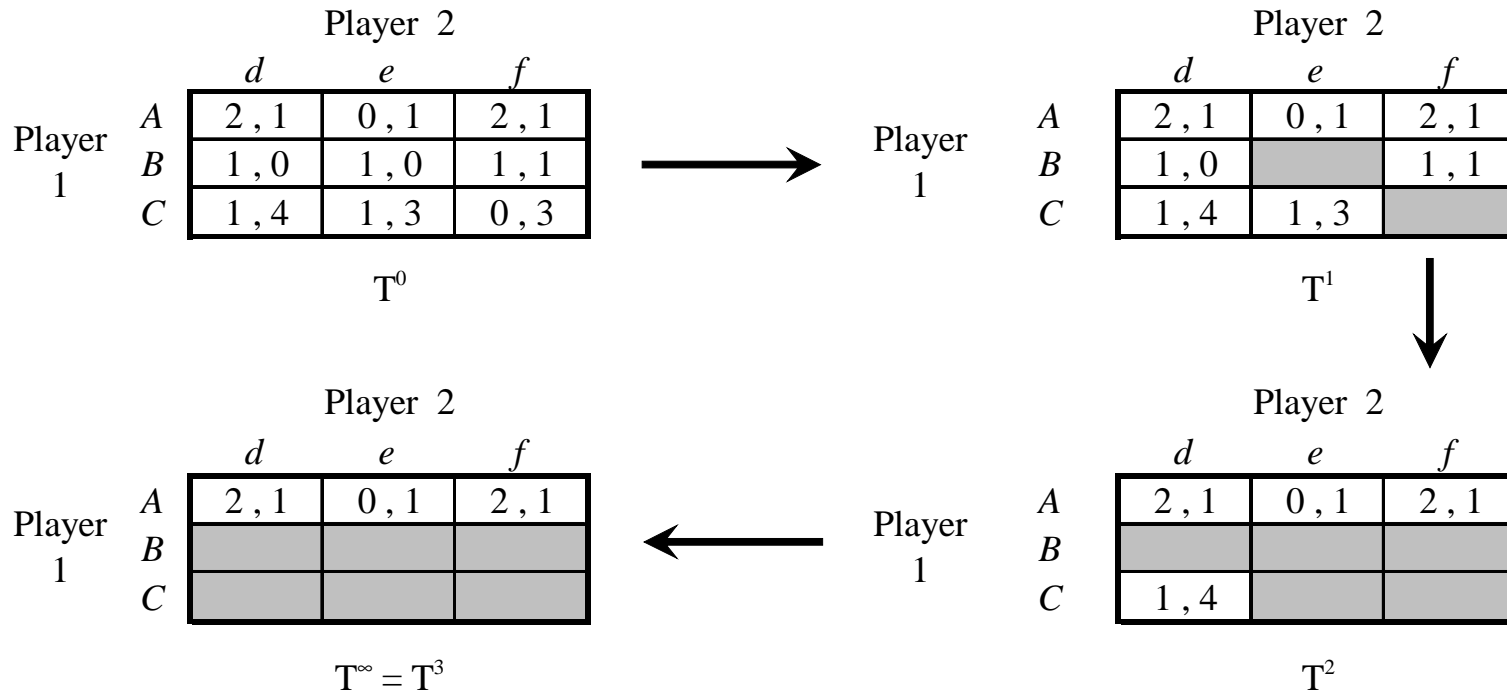
1.  $z(s_i, x_{-i}) \succ_i z(x_i, x_{-i})$  and
2. for all  $s_{-i} \in S_{-i}$ , if  $(x_i, s_{-i}) \in X$  then  $z(s_i, s_{-i}) \succeq_i z(x_i, s_{-i})$ .

**Iterated Deletion of Inferior Profiles :** for  $m \in \mathbb{N}$  define

$T^m \subseteq S$  recursively as follows:  $T^0 = S$  and, for  $m \geq 1$ ,

$T^m = T^{m-1} \setminus I^{m-1}$ , where  $I^{m-1} \subseteq T^{m-1}$  is the set of strategy profiles

that are inferior relative to  $T^{m-1}$ . Let  $T^\infty = \bigcap_{m \in \mathbb{N}} T^m$ .



$T^0 = S = \{(A, d), (A, e), (A, f), (B, d), (B, e), (B, f), (C, d), (C, e), (C, f)\}$ ,  $I^0 = \{(B, e), (C, f)\}$  (the elimination of  $(B, e)$  is done through player 2 and strategy  $f$ , while the elimination of  $(C, f)$  is done through player 1 and strategy  $B$ );

$T^1 = \{(A, d), (A, e), (A, f), (B, d), (B, f), (C, d), (C, e)\}$ ,  $I^1 = \{(B, d), (B, f), (C, e)\}$  (the elimination of  $(B, d)$  and  $(B, f)$  is done through player 1 and strategy  $A$ , while the elimination of  $(C, e)$  is done through player 2 and strategy  $d$ );

$T^2 = \{(A, d), (A, e), (A, f), (C, d)\}$ ,  $I^2 = \{(C, d)\}$  (the elimination of  $(C, d)$  is done through player 1 and strategy  $A$ );

$T^3 = \{(A, d), (A, e), (A, f)\}$ ,  $I^3 = \emptyset$ ; thus  $T^\infty = T^3$ .



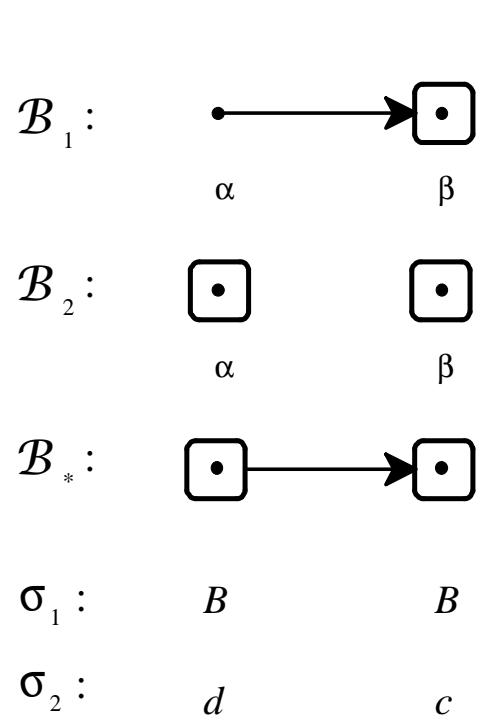
### **PROPOSITION 3.** $K_*R \subseteq T^\infty$

If at a state it is commonly **known** that all players are rational, then the strategy profile chosen at that state belongs to the game obtained after applying the iterated deletion of Inferior strategy profiles.

**PROPOSITION 4.** Fix a strategic-form game with ordinal payoffs  $G$  and let  $s \in T^\infty$ . Then there exists an epistemic model of  $G$  and a state  $\omega$  such that  $\sigma(\omega) = s$  and  $\omega \in K_*R$ .

NOT TRUE if we replace common knowledge with common belief

		Player 2	
		<i>c</i>	<i>d</i>
Player 1	<i>A</i>	1, 1	1, 0
	<i>B</i>	1, 1	0, 1



$$R_1 = \{\alpha, \beta\}, \quad R_2 = \{\alpha, \beta\}$$

There is common belief of rationality at every state and yet at state  $\alpha$  the strategy profile played is  $(B, d)$  which is inferior

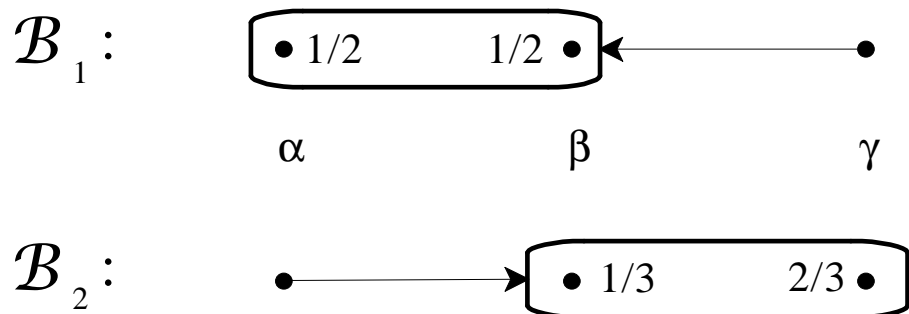
$$T^\infty = \{(A, c), (B, c)\}$$

$$S^\infty = \{(A, c), (A, d), (B, c), (B, d)\}$$

# PROBABILISTIC BELIEFS

**Definition.** A *Bayesian frame* is an interactive belief frame together with a collection  $\{p_{i,\omega}\}_{i \in N, \omega \in \Omega}$  of probability distributions on  $\Omega$  such that

- (1) if  $\omega' \in \mathcal{B}_i(\omega)$  then  $p_{i,\omega'} = p_{i,\omega}$
- (2)  $p_{i,\omega}(\omega') > 0$  if and only if  $\omega' \in \mathcal{B}_i(\omega)$   
 (the support of  $p_{i,\omega}$  coincides with  $\mathcal{B}_i(\omega)$ )



**Definition.** A strategic-form game *with von Neumann-Morgenstern payoffs* is a quintuple

$$\langle N, \{S_i\}_{i \in N}, O, \{U_i\}_{i \in N}, z \rangle$$

where

$N = \{1, \dots, n\}$  is a set of *players*

$S_i$  is the set of *strategies* of player  $i \in N$

$O$  is a set of *outcomes*

$U_i: O \rightarrow \mathbb{R}$  is player  $i$ 's von Neumann-Morgenstern utility function

$z: S \rightarrow O$  (where  $S = S_1 \times \dots \times S_n$ ) associates an outcome with every strategy profile  $s \in S$

Its reduced form is a triple  $\langle N, \{S_i\}_{i \in N}, \{\pi_i\}_{i \in N} \rangle$  where  $\pi_i(s) = U_i(z(s))$ .

An *epistemic model* of a strategic-form game is a Bayesian frame together with  $n$  functions

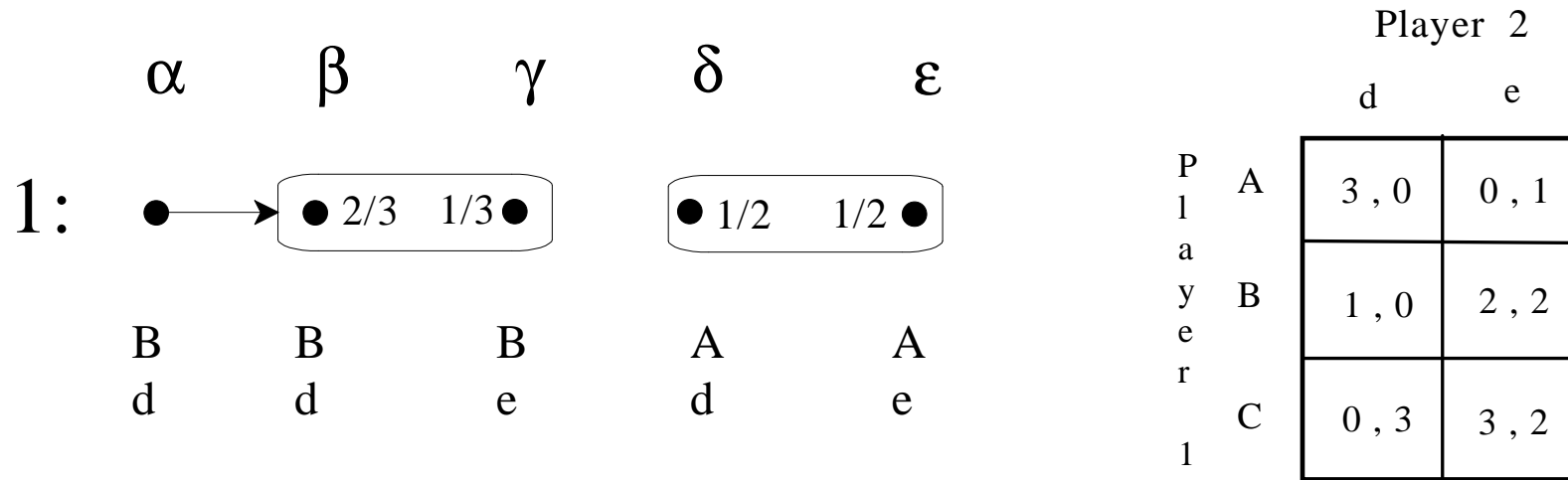
$$\sigma_i : \Omega \rightarrow S_i \quad (i \in N)$$

such that if  $\omega' \in \mathcal{B}_i(\omega)$  then  $\sigma_i(\omega') = \sigma_i(\omega)$

Stronger definition of Rationality than the previous ones

Player  $i$  is **RATIONAL** at state  $\alpha$  if her choice at  $\alpha$  maximizes her expected payoff, given her beliefs at  $\alpha$ : for all  $t_i \in S_i$

$$\sum_{\omega \in \mathcal{B}_i(\alpha)} \pi_i(\sigma_i(\alpha), \sigma_{-i}(\omega)) p_{i,\alpha}(\omega) \geq \sum_{\omega \in \mathcal{B}_i(\alpha)} \pi_i(t_i, \sigma_{-i}(\omega)) p_{i,\alpha}(\omega)$$



$$R_1 = \{\delta, \epsilon\}$$

Player 1 is not rational at  $\alpha$  because her expected payoff is  $\frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 2 = \frac{4}{3}$

while if she had chosen strategy A her payoff would have been  $\frac{2}{3} \cdot 3 + \frac{1}{3} \cdot 0 = 2$

On the other hand, Player 1 *is* rational at  $\delta$  because her expected payoff is  $\frac{1}{2} \cdot 3 + \frac{1}{2} \cdot 0 = \frac{3}{2}$

and if she had chosen strategy B her payoff would have been  $\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 2 = \frac{3}{2}$

and if she had chosen strategy C her payoff would have been  $\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 3 = \frac{3}{2}$

## What are the implications of Common Belief of this stronger notion of rationality?

**Definition.** A mixed strategy of player  $i$  is a probability distribution over  $S_i$ . The set of mixed strategies of player  $i$  is denoted by  $\Delta(S_i)$ .

Let  $t_i \in S_i$  and  $v_i \in \Delta(S_i)$ . We say that  $t_i$  is *strictly dominated* by  $v_i$  if,

for every  $s_{-i} \in S_{-i}$ ,  $\pi_i(t_i, s_{-i}) < \sum_{s_i \in S_i} v_i(s_i) \pi_i(s_i, s_{-i})$

		Player 2	
		d	e
P l a y e r  1	A	3 , 0	0 , 1
	B	0 , 0	2 , 2
	C	0 , 3	3 , 2

In this game strategy  $B$  of player 1 is

strictly dominated by the mixed strategy  $\left( \begin{matrix} A & C \\ \frac{1}{6} & \frac{5}{6} \end{matrix} \right)$

ITERATIVE  
 DELETION  
 OF PURE  
 STRATEGIES  
 THAT ARE  
 STRICTLY  
 DOMINATED  
 BY (POSSIBLY  
 MIXED)  
 STRATEGIES

		Player 2		
		e	f	g
Player 1	A	3, 0	1, 0	0, 1
	B	1, 1	0, 2	1, 1
	C	0, 0	4, 1	2, 2
	D	0, 3	1, 0	3, 2

(a) The game G

B is strictly dominated by  $(1/2 A, 1/2 D)$

		Player 2		
		e	f	g
Player 1	A	3, 0	1, 0	0, 1
	C	0, 0	4, 1	2, 2
	D	0, 3	1, 0	3, 2

(b) The game  $G^1$

Now f is strictly dominated by g

		Player 2	
		e	g
Player 1	A	3, 0	0, 1
	C	0, 0	2, 2
	D	0, 3	3, 2

(c) The game  $G^2$

Now C is strictly dominated by  $(1/6 A, 5/6 D)$

		Player 2	
		e	g
Player 1	A	3, 0	0, 1
	D	0, 3	3, 2

(d) The game  $G^3 = G^\infty$

No strategy is strictly dominated



Let  $G$  be a strategic-form game with von Neumann-Morgenstern payoffs and  $G^\infty$  be the game obtained after applying the procedure of Iterated Deletion of Pure Strategies that are Strictly Dominated by Possibly Mixed Strategies.

Let  $S_m^\infty$  denote the pure-strategy profiles of game  $G^\infty$

Given a model of  $G$ , let  $S_m^\infty$  be the event  $\{\omega \in \Omega : \sigma(\omega) \in S_m^\infty\}$

**PROPOSITION 5.**  $B_*R \subseteq S_m^\infty$

**PROPOSITION 6.** Fix a strategic-form game with von Neumann-Morgenstern payoffs  $G$  and let  $s \in S_m^\infty$ . Then there exists a Bayesian model of  $G$  and a state  $\omega$  such that  $\sigma(\omega) = s$  and  $\omega \in B_*R$ .

Given this stronger notion of rationality, *there is a difference between common belief of rationality and common knowledge of rationality*. The implications of common knowledge of rationality are stronger.

With knowledge, a player's beliefs are always correct and are believed to be correct by every other player. Thus there is *correctness and common belief of correctness* of everybody's beliefs.

**Definition.** Given a strategic-form game with von Neumann-Morgenstern payoffs  $G$ , a pure-strategy profile  $x \in X \subseteq S$  is *inferior relative to  $X$*  if there exists a player  $i$  and a (possibly mixed) strategy  $\nu_i$  of player  $i$  (whose support can be any subset of  $S_i$ , not necessarily the projection of  $X$  onto  $S_i$ ) such that:

- (1)  $\pi_i(x_i, x_{-i}) < \sum_{s_i \in S_i} \pi_i(s_i, x_{-i}) \nu_i(s_i)$  ( $\nu_i$  yields a higher expected payoff than  $x_i$  against  $x_{-i}$ )
- (2) for all  $s_{-i} \in S_{-i}$  such that  $(x_i, s_{-i}) \in X$ ,  $\pi_i(x_i, s_{-i}) \leq \sum_{s_i \in S_i} \pi_i(s_i, s_{-i}) \nu_i(s_i)$

		Player 2		
		D	E	F
Player 1	A	2 , 0	2 , 2	0 , 2
	B	2 , 2	1 , 2	5 , 1
	C	2 , 0	1 , 0	1 , 5

Here  $(C, F)$  is inferior relative to  $S$  (for player 1,  $B$  weakly dominates  $C$  and is strictly better than  $C$  against  $F$ )

and  $(A, D)$  is inferior relative to  $S$  (for player 2,  $E$  weakly dominates  $D$  and is strictly better than  $D$  against  $A$ )

ITERATED  
DELETION  
OF  
INFERIOR  
PURE  
STRATEGY  
PROFILES

		Player 2		
		D	E	F
Player 1	A	2 , 0	2 , 2	0 , 2
	B	2 , 2	1 , 2	5 , 1
	C	2 , 0	1 , 0	1 , 5

(a)

$$S_s^0 = S, D_s^0 = \{(A, D), (C, F)\}$$

		Player 2		
		D	E	F
Player 1	A		2 , 2	0 , 2
	B	2 , 2	1 , 2	5 , 1
	C	2 , 0	1 , 0	

(b)

$$S_s^1 = \{(A, E), (A, F), (B, D), (B, E), (B, F), (C, D), (C, E)\}$$

$$D_s^1 = \{(C, E), (B, F)\}$$

		Player 2		
		D	E	F
Player 1	A		2 , 2	0 , 2
	B	2 , 2	1 , 2	
	C	2 , 0		

(c)

$$S_s^2 = \{(A, E), (A, F), (B, D), (B, E), (C, D)\},$$

$$D_s^2 = \{(B, E)\}.$$

		Player 2		
		D	E	F
Player 1	A		2 , 2	0 , 2
	B	2 , 2		
	C	2 , 0		

(d)

$$S_s^3 = S_s^\infty = \{(A, E), (A, F), (B, D), (C, D)\},$$

$$D_s^3 = \emptyset.$$

Let  $G$  be a strategic-form game with von Neumann-Morgenstern payoffs and  $G^\infty$  be the game obtained after applying the procedure of Iterated Deletion of Inferior Pure-Strategy Profiles.

Let  $S_s^\infty$  denote the pure-strategy profiles of game  $G^\infty$

Given a model of  $G$ , let  $S_s^\infty$  be the event  $\{\omega \in \Omega : \sigma(\omega) \in S_s^\infty\}$

**PROPOSITION 7.**  $K_*\mathbf{R} \subseteq S_s^\infty$

**PROPOSITION 8.** Fix a strategic-form game with von Neumann-Morgenstern payoffs  $G$  and let  $s \in S_s^\infty$ . Then there exists a Bayesian model of  $G$  and a state  $\omega$  such that  $\sigma(\omega) = s$  and  $\omega \in K_*\mathbf{R}$ .

		Player 2		
		D	E	F
Player 1	A	2 , 0	2 , 2	0 , 2
	B	2 , 2	1 , 2	5 , 1
	C	2 , 0	1 , 0	1 , 5

In this game  $S^\infty = S_m^\infty = S$

while  $S_s^\infty = \{(A, E), (A, F), (B, D), (C, D)\}$

Thus every strategy profile is compatible with *common belief* of rationality while only  $(A, E)$ ,  $(A, F)$ ,  $(B, D)$  and  $(C, D)$  are compatible with *common knowledge* of rationality

## CREDITS

The link between the iterated deletion of strictly dominated strategies and the informal notion of common belief of rationality was first shown by Bernheim (1984) and Pearce (1984)

The first explicit epistemic characterization was provided by Tan and Werlang (1998) using a universal type space.

The state space formulation used in Propositions 5 and 6 is due to Stalnaker (1994), but it was implicit in Brandenburger and Dekel (1987).

Propositions 7 and 8 are due to Stalnaker (1994) (with a correction given in Bonanno and Nehring, 1996b).

To my knowledge, Propositions 1, 2, 3 and 4 have not been explicitly stated before.

References and further details can be found in

Battigalli, Pierpaolo and Bonanno Giacomo, “Recent results on belief, knowledge and the epistemic foundations of game theory”, *Research in Economics*, 53 (2), June 1999, pp. 149-225.

For a syntactic version of Propositions 1, 2, 3 and 4 see  
Giacomo Bonanno, A syntactic approach to rationality in games, Working Paper,  
University of California, Davis (<http://www.econ.ucdavis.edu/faculty/bonanno/PDF/CBR.pdf>)

Royal Netherlands Academy of Arts and Sciences (KNAW)  
Master Class

Amsterdam, February 8th, 2007

# Epistemic Foundations of Game Theory

## Lecture 2

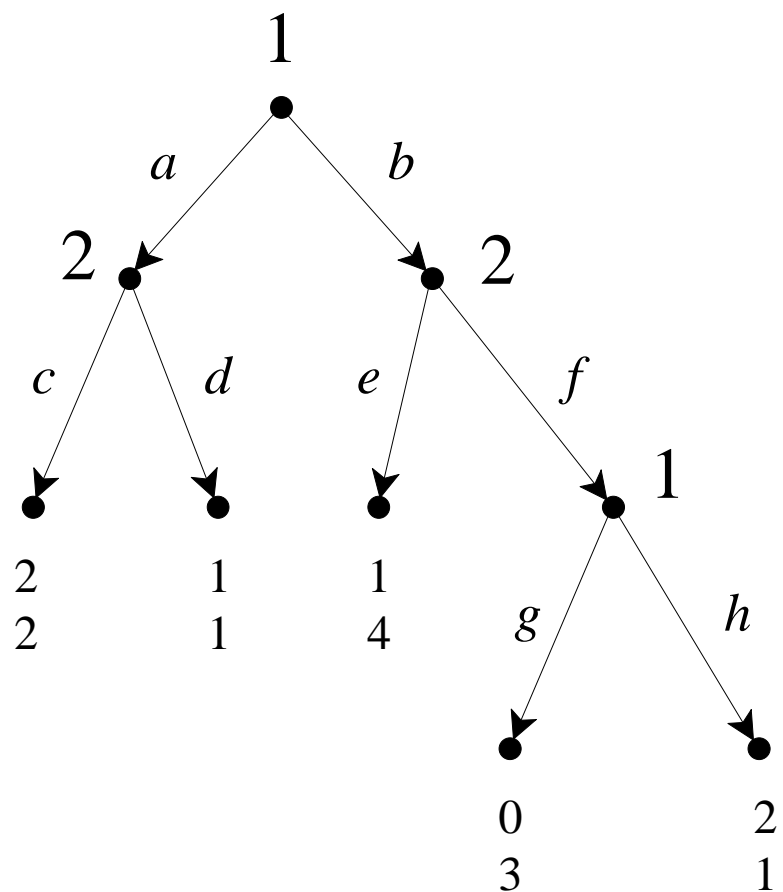
Giacomo Bonanno

(<http://www.econ.ucdavis.edu/faculty/bonanno/>)

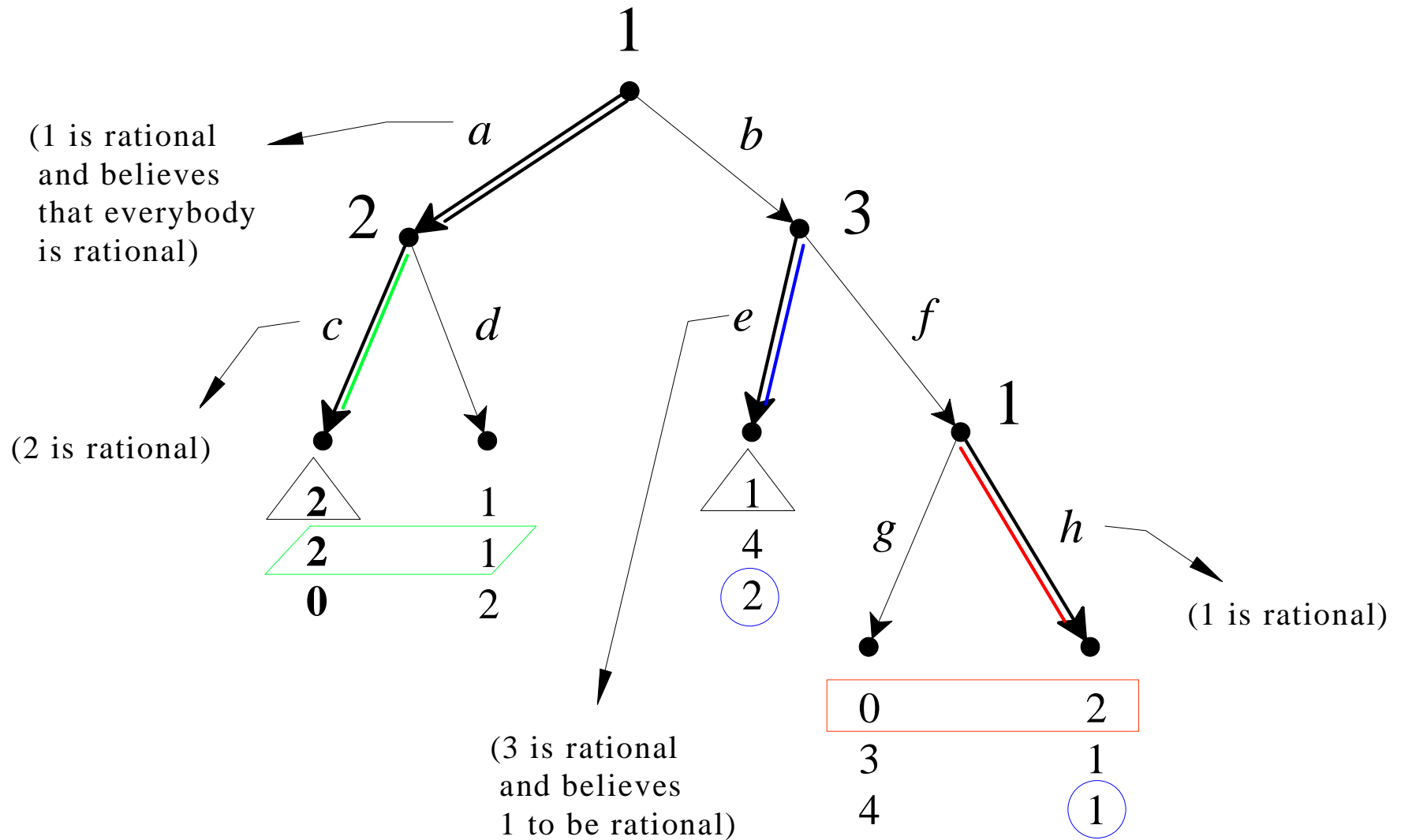


# EXTENSIVE GAMES WITH PERFECT INFORMATION

- tree
- $n$  players
- assignment of one player to every non-terminal node
- assignment of an *ordinal* payoff to every player at every terminal node



# BACKWARD-INDUCTION SOLUTION



# STRATEGIES IN PERFECT-INFORMATION GAMES

Non-terminal nodes are called *decision nodes*

$X$  : set of decision nodes

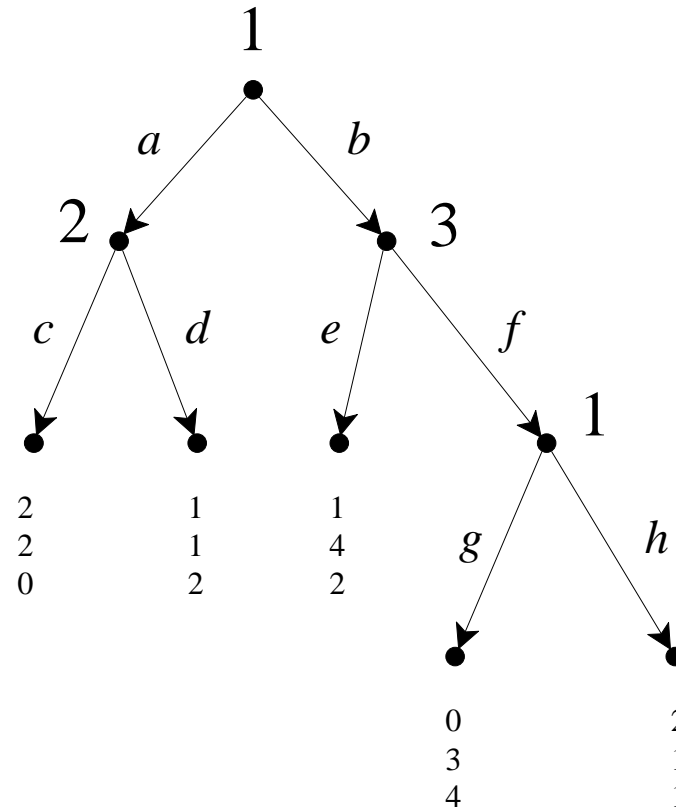
$X_i$  : set of decision nodes assigned to player  $i$

**Definition.** A strategy of player  $i$  is a function that assigns to every

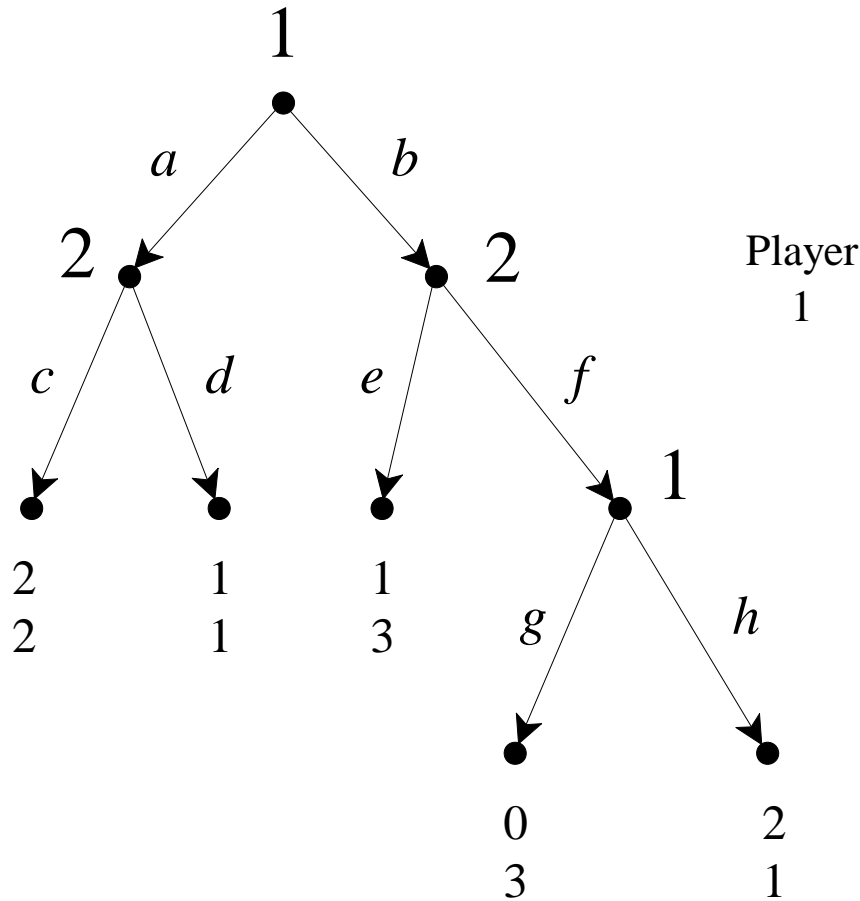
$x \in X_i$  a choice at  $x$

Player 1's strategies:

$(a,g)$ ,  $(a,h)$ ,  $(b,g)$  and  $(b,h)$



# THE STRATEGIC FORM OF A PERFECT-INFORMATION GAME



	Player 2			
	$ce$	$cf$	$de$	$df$
$ag$	2, 2	2, 2	1, 1	1, 1
$ah$	2, 2	2, 2	1, 1	1, 1
$bg$	1, 3	0, 3	1, 3	0, 3
$bh$	1, 3	2, 1	1, 3	2, 1

# EPISTEMIC MODEL OF A PERFECT-INFORMATION GAME

(Knowledge based)

- Set of states  $\Omega$
- Equivalence relation  $\mathcal{K}_i$  on  $\Omega$  for every player  $i$
- For every player  $i$  a function  $\sigma_i : \Omega \rightarrow S_i$  satisfying  
if  $\omega' \in \mathcal{K}_i(\omega)$  then  $\sigma_i(\omega') = \sigma_i(\omega)$

Thus a standard epistemic model for the associated strategic form

## Recall from Lecture 1:

Let  $s_i$  and  $t_i$  be two strategies of player  $i$ :  $s_i, t_i \in S_i$

$s_i \succ_i t_i$  is interpreted as “strategy  $s_i$  is better for player  $i$  than strategy  $t_i$ ”

$s_i \succ_i t_i$  is true at state  $\omega$  if  $u_i(s_i, \sigma_{-i}(\omega)) > u_i(t_i, \sigma_{-i}(\omega))$

that is,  $s_i$  is better than  $t_i$  against  $\sigma_{-i}(\omega)$

profile of strategies chosen  
by the players other than  $i$

Let  $\|s_i \succ_i t_i\| = \{\omega \in \Omega : u_i(s_i, \sigma_{-i}(\omega)) > u_i(t_i, \sigma_{-i}(\omega))\}$  event that  $s_i$  is better than  $t_i$

If  $s_i \in S_i$ , let  $\|s_i\| = \{\omega \in \Omega : \sigma_i(\omega) = s_i\}$  event that player  $i$  chooses  $s_i$

Let  $\mathbf{R}_i^{EA}$  be the event representing the proposition “player  $i$  is *ex ante* rational”

$$\|s_i\| \cap K_i \|t_i \succ_i s_i\| \subseteq \neg \mathbf{R}_i^{EA}$$

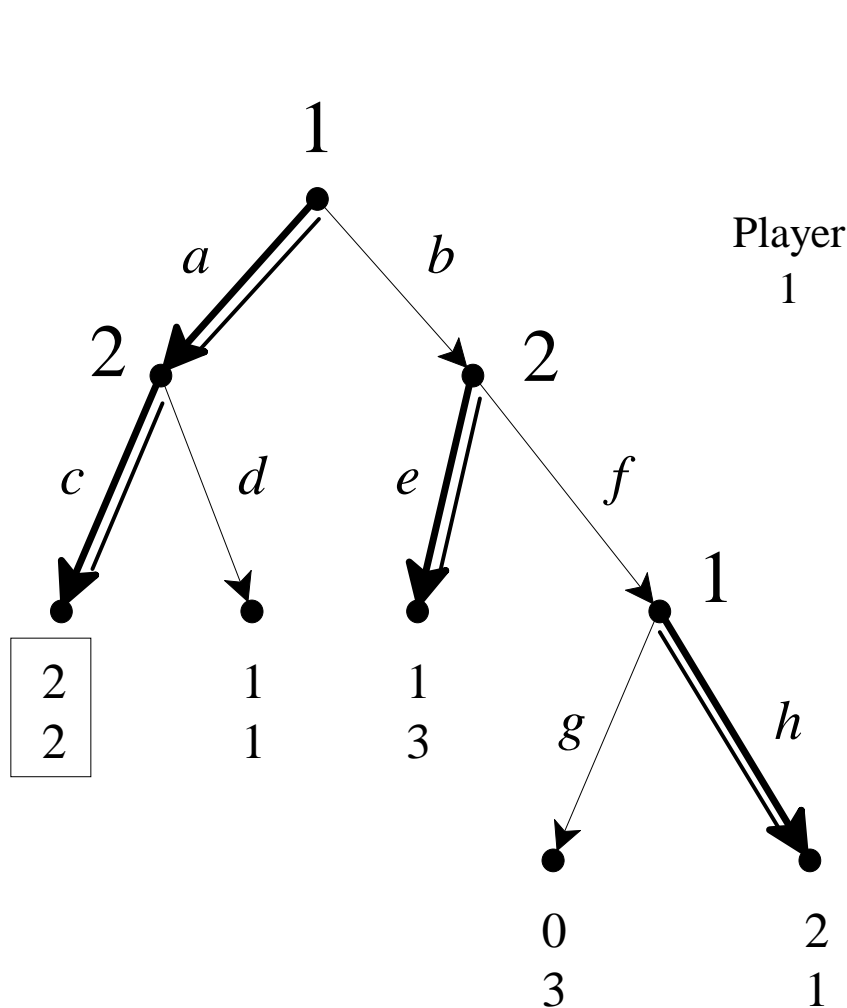
$$\neg \mathbf{R}_i^{EA} = \bigcup_{s_i \in S_i} \bigcup_{t_i \in S_i} (\|s_i\| \cap K_i \|t_i \succ_i s_i\|)$$

$$\mathbf{R}^{EA} = \mathbf{R}_1^{EA} \cap \dots \cap \mathbf{R}_n^{EA} \quad \text{all players are rational}$$

### Recall from Lecture 1:

**PROPOSITION:** if at a state there is common knowledge of *ex ante* rationality then the strategy profile chosen at that state belongs to the game obtained by applying the iterated deletion of strictly dominated strategies; conversely, for every such strategy profile there is a model and a state where (1) the strategy profile is chosen and (2) there is common knowledge of *ex ante* rationality.

This notion of rationality is not sufficient to yield backward induction



		Player 2			
		<i>ce</i>	<i>cf</i>	<i>de</i>	<i>df</i>
<i>ag</i>	2, 2	2, 2	1, 1	1, 1	
<i>ah</i>	2, 2	2, 2	1, 1	1, 1	
<i>bg</i>	1, 3	0, 3	1, 3	0, 3	
<i>bh</i>	1, 3	2, 1	1, 3	2, 1	

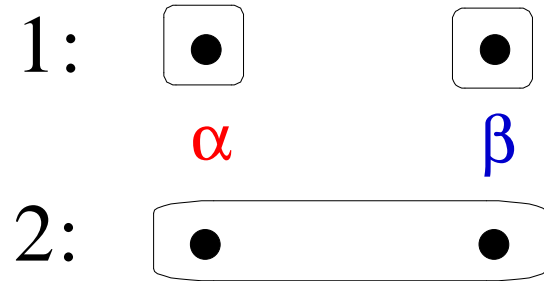
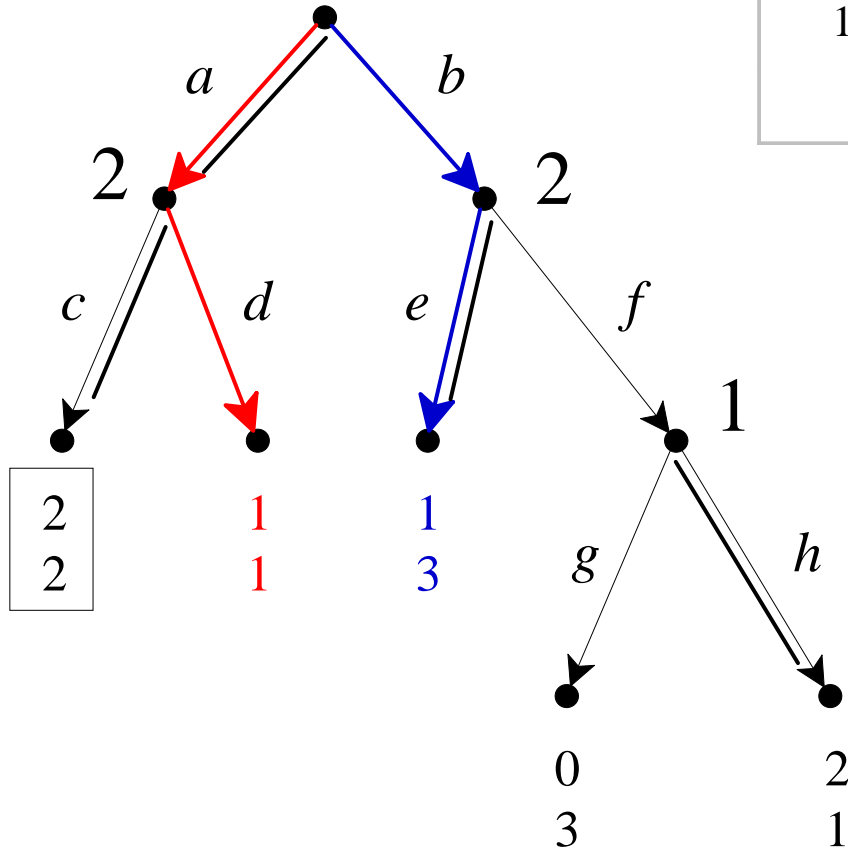
Here there are no strictly dominated strategies

Thus every strategy profile is consistent with common belief/knowledge of *ex ante* rationality



For example:

		Player 2			
		<i>ce</i>	<i>cf</i>	<i>de</i>	<i>df</i>
Player 1	<i>ag</i>	2, 2	2, 2	1, 1	1, 1
	<i>ah</i>	2, 2	2, 2	1, 1	1, 1
	<i>bg</i>	1, 3	0, 3	1, 3	0, 3
	<i>bh</i>	1, 3	2, 1	1, 3	2, 1



1's strategy: *ah* *bh*

2's strategy: *de* *de*

(For 2 *ce* better than *de* at  $\alpha$  but not at  $\beta$ , thus at  $\alpha$  she does not know that *ce* is better.)

Here: *ex ante* rationality and common knowledge of *ex ante* rationality at both states.

Let  $\mathbf{R}_i^{EA/S}$  be the event representing the proposition “player  $i$  is *ex ante* rational in a strong sense”

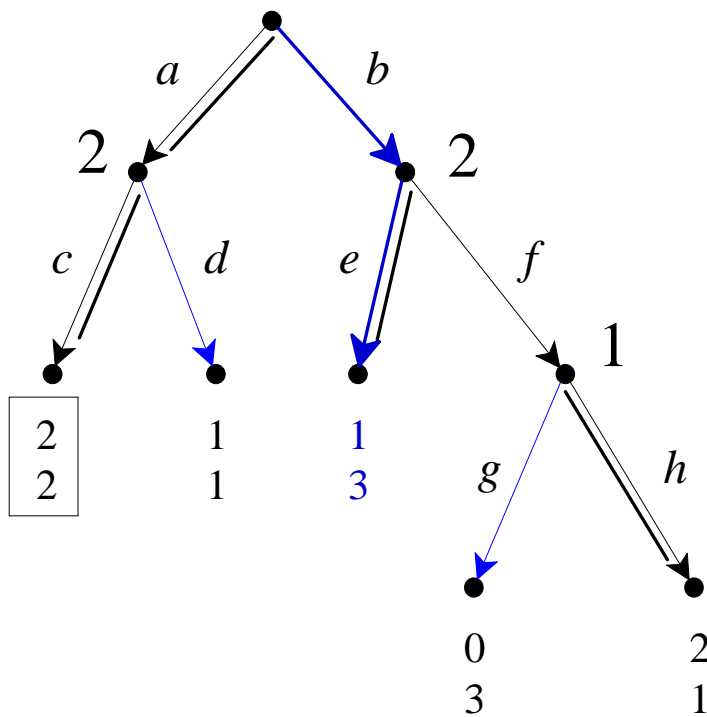
$$\|s_i\| \cap K_i \|t_i \succeq_i s_i\| \cap \neg K_i \neg \|t_i \succ_i s_i\| \subseteq \neg \mathbf{R}_i^{EA/S}$$

$$\neg \mathbf{R}_i^{EA/S} = \bigcup_{s_i \in S_i} \bigcup_{t_i \in S_i} (\|s_i\| \cap K_i \|t_i \succeq_i s_i\| \cap \neg K_i \neg \|t_i \succ_i s_i\|)$$

$$\mathbf{R}^{EA/S} = \mathbf{R}_1^{EA/S} \cap \dots \cap \mathbf{R}_n^{EA/S} \quad \text{all players are rational in a strong sense}$$

### Recall from Lecture 1:

**PROPOSITION:** if at a state there is common knowledge of *ex ante* rationality in a strong sense then the strategy profile chosen at that state belongs to the set  $T^\infty$  of strategy profiles that survive the iterated deletion of inferior profiles; conversely, for every such strategy profile there is a model and a state where (1) the strategy profile is chosen and (2) there is common knowledge of *ex ante* rationality in a strong sense.



- 1:  ●  
 $\alpha$
- 2:  ●

1's strategy:  $bg$   
 2's strategy:  $de$

	$ce$	$cf$	$de$	$df$
$ag$	2, 2	2, 2	1, 1	1, 1
$ah$	2, 2	2, 2	1, 1	1, 1
$bg$	1, 3	0, 3	1, 3	0, 3
$bh$	1, 3	2, 1	1, 3	2, 1

player 1 using  $ah$   
 player 2 using  $ce$

	$ce$	$cf$	$de$
$ag$	2, 2	2, 2	1, 1
$ah$	2, 2	2, 2	1, 1
$bg$	<input type="checkbox"/>	<input type="checkbox"/>	1, 3
$bh$	1, 3	<input type="checkbox"/>	1, 3

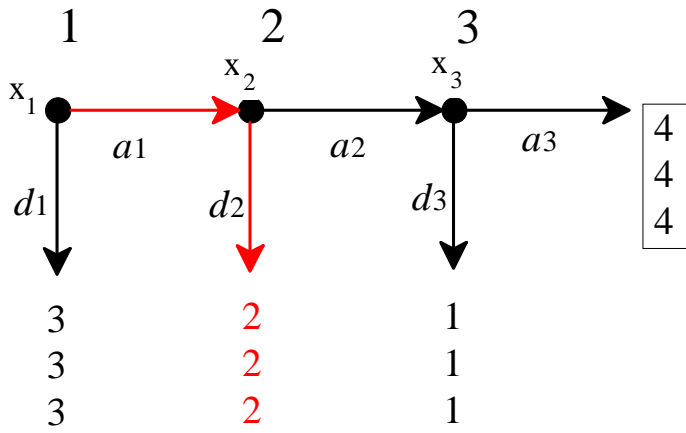
player 2 using  $cf$   
 player 1 using  $ah$

	$ce$	$cf$	$de$
$ag$	2, 2	2, 2	<input type="checkbox"/>
$ah$	2, 2	2, 2	<input type="checkbox"/>
$bg$	<input type="checkbox"/>	<input type="checkbox"/>	1, 3
$bh$	<input type="checkbox"/>	<input type="checkbox"/>	1, 3

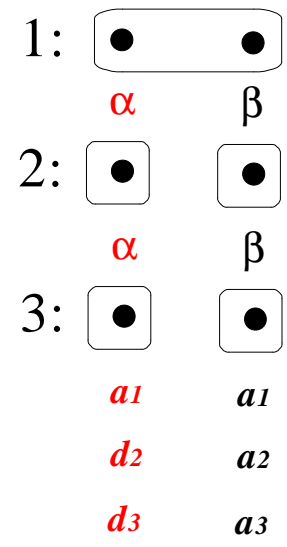
Thus even common knowledge of *ex ante* rationality in a strong sense is not sufficient to yield backward induction

In this example all the strategy profiles in  $T^\infty$  are Nash equilibria. Is it the case that common knowledge of *ex ante* rationality in the strong sense gives Nash equilibrium **play** in perfect information games?

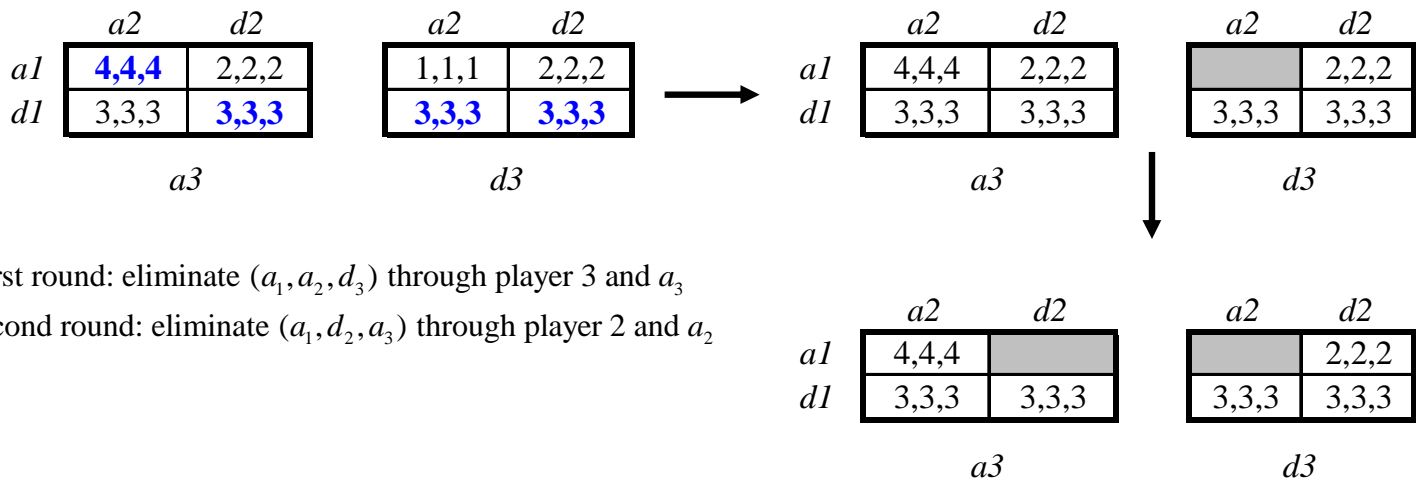
The answer is NO!



4  
4  
4  
Backward induction solution



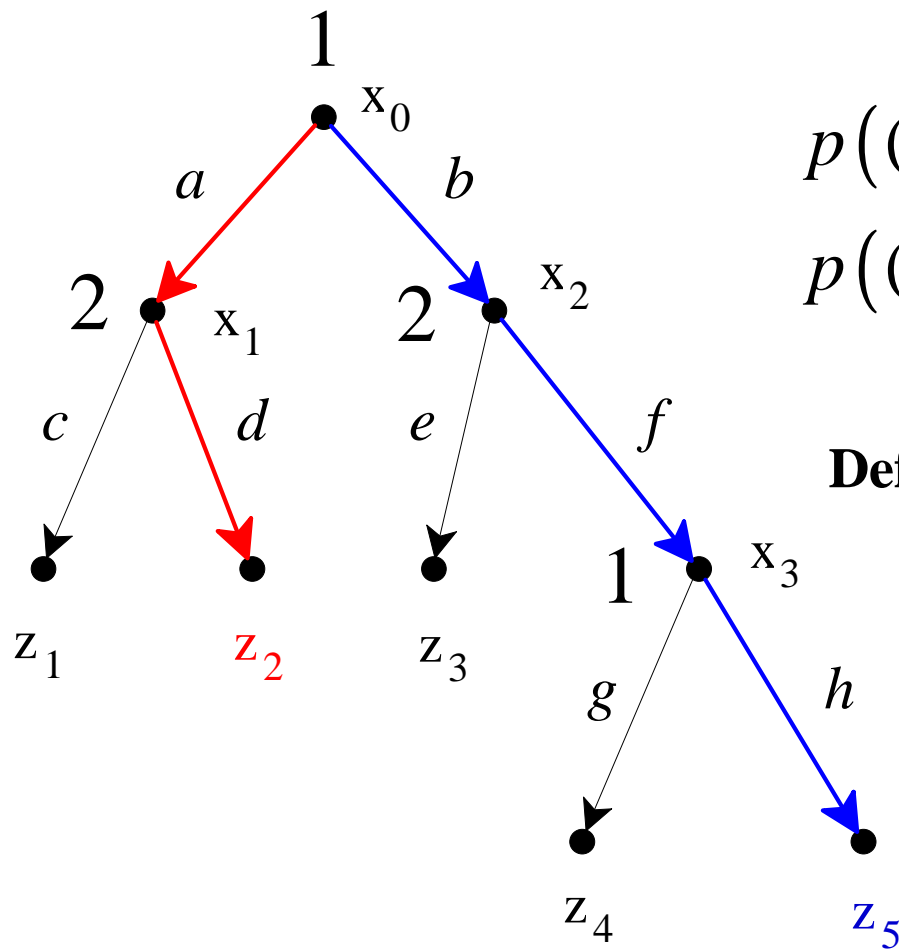
There is no Nash equilibrium that yields the play  $a_1 d_2$  (the Nash equilibria are marked in blue)



First round: eliminate  $(a_1, a_2, d_3)$  through player 3 and  $a_3$   
 second round: eliminate  $(a_1, d_2, a_3)$  through player 2 and  $a_2$

# Going beyond *ex ante* rationality

Given a strategy profile  $s$ , let  $p(s)$  be the associated play



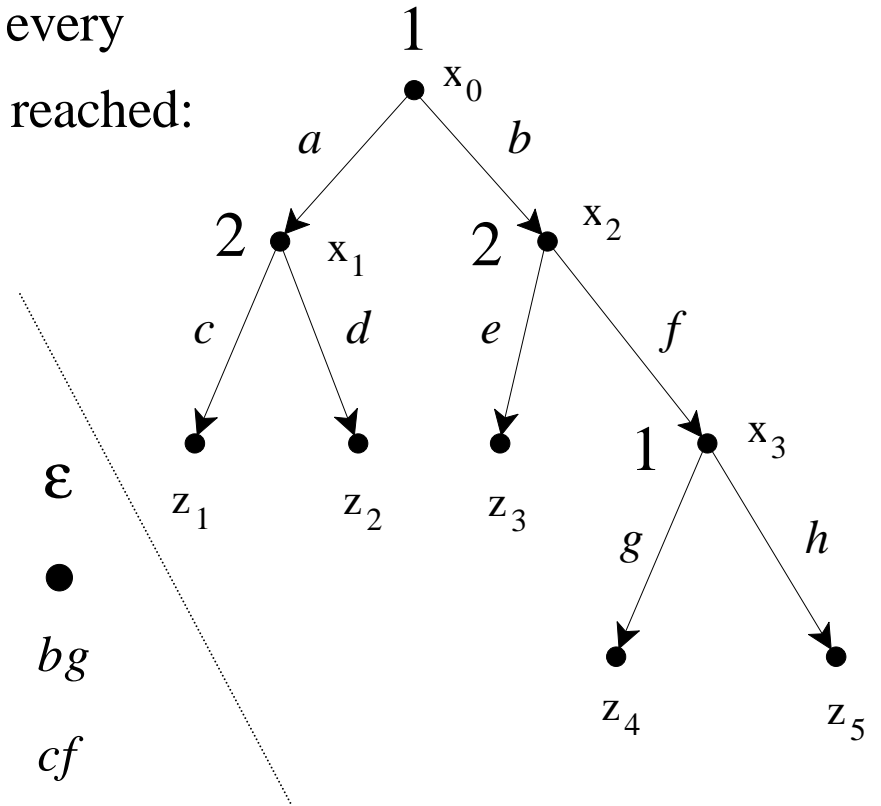
$$p((ag, df)) = x_0 x_1 z_2$$

$$p((bh, df)) = x_0 x_2 x_3 z_5$$

**Definition.** At state  $\omega$  node  $x$  is *reached* if and only if  $x \in p(\sigma(\omega))$ .

**Definition.** Given an epistemic model, for every node  $x$ , let  $\|x\|$  be the event that node  $x$  is reached:

$$\|x\| = \{\omega \in \Omega : x \in p(\sigma(\omega))\}$$



	$\alpha$	$\beta$	$\gamma$	$\delta$	$\varepsilon$
	●	●	●	●	●
1's strategy:	$ag$	$bh$	$bg$	$bh$	$bg$
2's strategy:	$df$	$df$	$ce$	$de$	$cf$

$$\|x_1\| = \{\alpha\}, \quad \|x_2\| = \{\beta, \gamma, \delta, \varepsilon\}$$

$$\|x_3\| = \{\beta, \varepsilon\}, \quad \|z_1\| = \emptyset, \quad \|z_2\| = \{\alpha\}, \quad \text{etc.}$$

Let  $E, F \subseteq \Omega$  be two events.

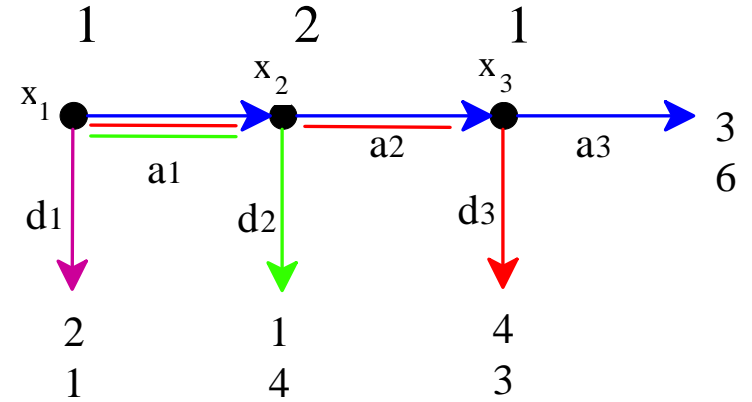
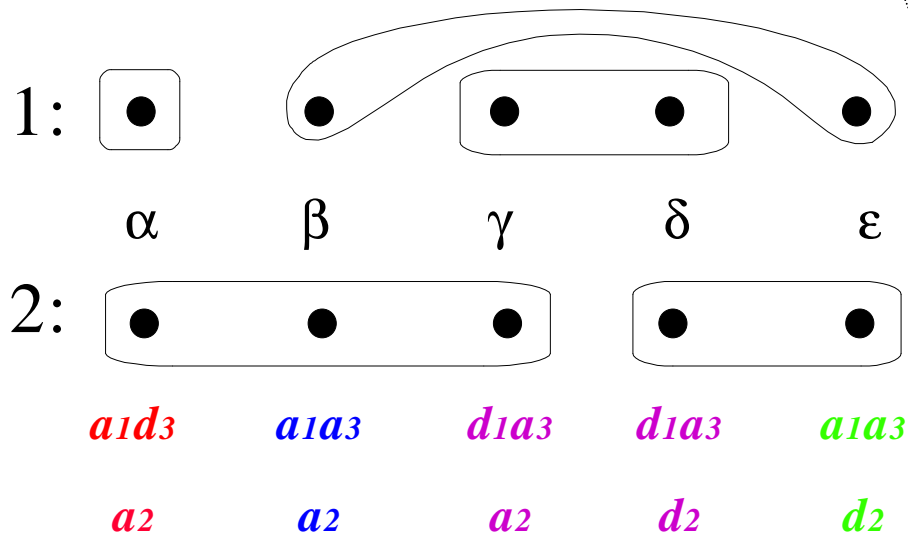
Denote by  $E \rightarrow F$  the event  $\neg E \cup F$  (if  $E$  then  $F$ )

Let  $\mathbf{R}_i^{RN}$  be the event representing the proposition “player  $i$  is rational *at reached nodes*”

$$\text{if } x \in X_i \quad \|x\| \cap \|s_i\| \cap K_i (\|x\| \rightarrow \|t_i \succ_i s_i\|) \subseteq \neg \mathbf{R}_i^{RN}$$

$$\neg \mathbf{R}_i^{RN} = \bigcup_{x \in X_i} \bigcup_{s_i \in S_i} \bigcup_{t_i \in S_i} (\|s_i\| \cap K_i (\|x\| \rightarrow \|t_i \succ_i s_i\|) \cap \|x\|)$$

$$\mathbf{R}^{RN} = \mathbf{R}_1^{RN} \cap \dots \cap \mathbf{R}_n^{RN} \quad \text{all players are rational at reached nodes}$$



$$\|d_2 \succ_2 a_2\| = \{\alpha\} \quad \|x_2\| = \{\alpha, \beta, \epsilon\} \quad \neg \|x_2\| \cup \|d_2 \succ_2 a_2\| = \{\alpha, \gamma, \delta\}$$

$K_2(\|x_2\| \rightarrow \|d_2 \succ_2 a_2\|) = \emptyset$  Thus player 2 is rational at nodes  $\alpha$  and  $\beta$  and trivially at  $\gamma$ .

$$\|a_2 \succ_2 d_2\| = \{\beta, \epsilon\} \quad \|x_2\| = \{\alpha, \beta, \epsilon\} \quad \neg \|x_2\| \cup \|a_2 \succ_2 d_2\| = \{\beta, \gamma, \delta, \epsilon\}$$

$$K_2(\|x_2\| \rightarrow \|a_2 \succ_2 d_2\|) = \{\delta, \epsilon\} \quad \|x_2\| \cap \|d_2\| \cap K_2(\|x_2\| \rightarrow \|a_2 \succ_2 d_2\|) = \{\epsilon\}$$

Thus **player 2 is** trivially rational at state  $\delta$ , and **irrational at  $\epsilon$** .

$$K_* \mathbf{R} = \emptyset$$



# Backward Induction terminating games

**Definition.** A *BI terminating game* is a perfect information game where

- (1) at each decision node there is a choice the terminates the game (it leads to a terminal node) and
- (2) the backward-induction solution prescribes a terminating choice at every decision node.

The best-known example is the **centipede game** ( $n$  is the number of decision nodes)

$$\begin{aligned}
 u_1(z_1) &= 2 && \text{for } 1 < k \leq n \\
 u_2(z_1) &= 1 && u_1(z_k) = u_2(z_{k-1}) \\
 &&& u_2(z_k) = u_1(z_{k-1}) + 2
 \end{aligned}$$

If  $n$  is even

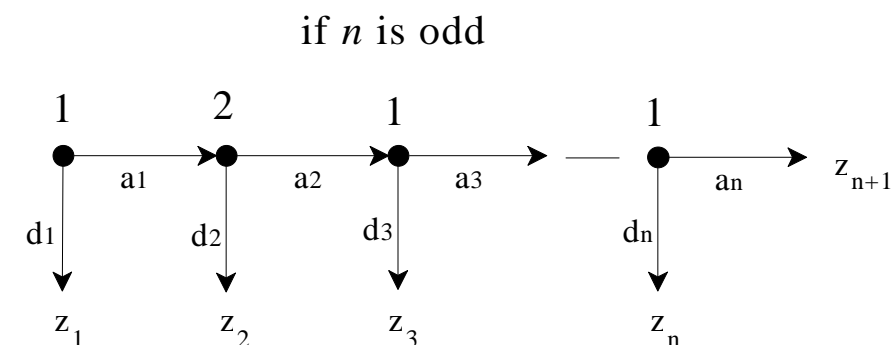
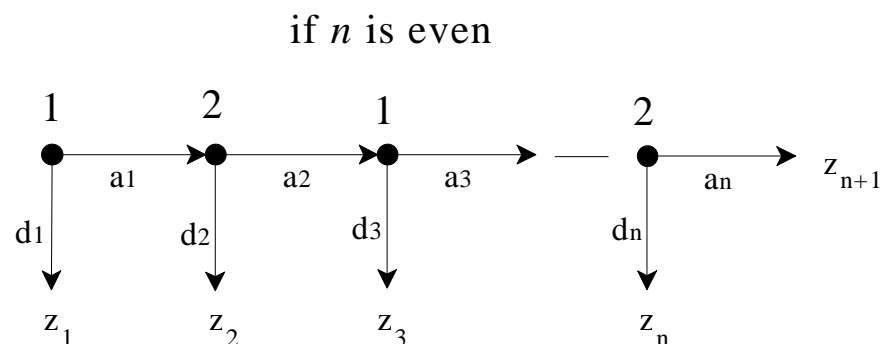
$$u_1(z_{n+1}) = u_1(z_n) + 1$$

$$u_2(z_{n+1}) = u_2(z_n) - 1$$

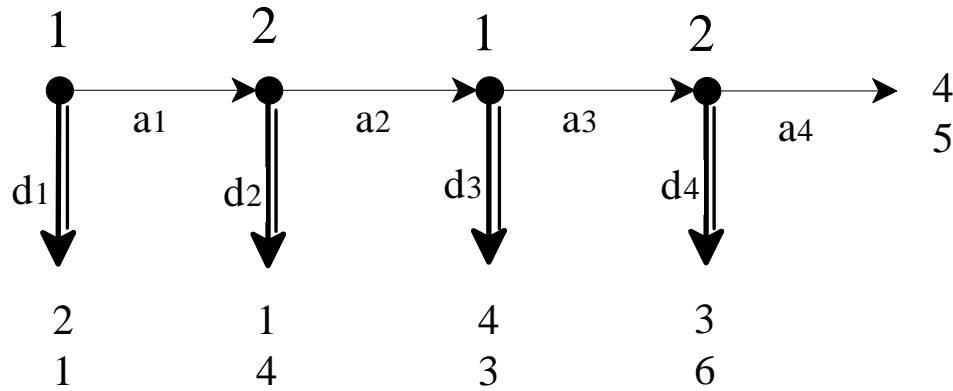
If  $n$  is odd

$$u_1(z_{n+1}) = u_1(z_n) - 1$$

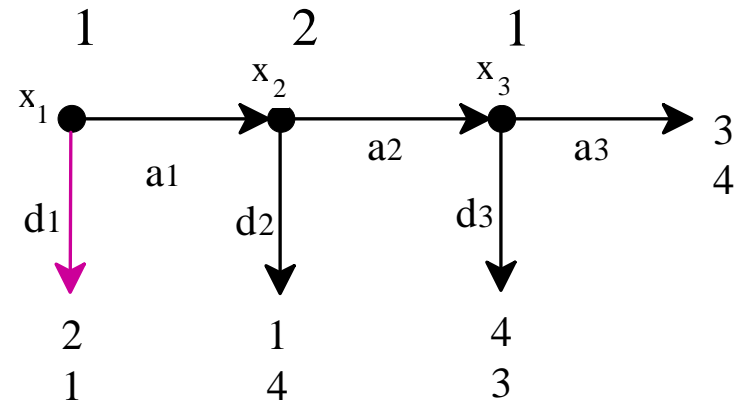
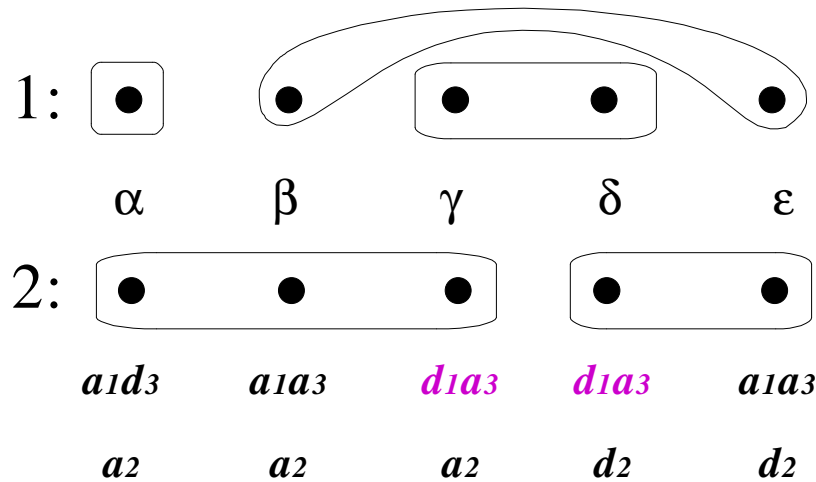
$$u_2(z_{n+1}) = u_2(z_n) + 1$$



$n = 4$



**Definition.** Given an epistemic model of a **BI** terminating game, let **BI** be the event that the backward-induction **play** obtains, that is,  $\mathbf{BI} = \{\omega \in \Omega : p(\sigma(\omega)) = x_1 z_1\}$



$\mathbf{BI} = \{\gamma, \delta\}$

**PROPOSITION 1.** In every BI terminating game,  $K_*R^{RN} \subseteq BI$

**PROPOSITION 2.** For every BI terminating game, there is a model of it where  $K_*R^{RN} \neq \emptyset$

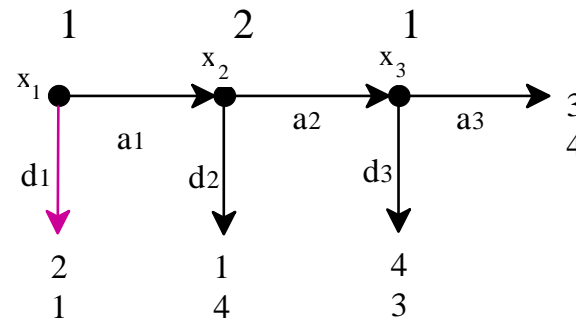
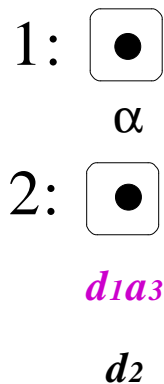
Aumann, R., A note on the centipede game, *Games and Economic Behavior*, 1998, 23: 97-105.

Broome, J. and W. Rabinowicz, Backwards induction in the centipede game, *Analysis*, 1999, 59:237-242.

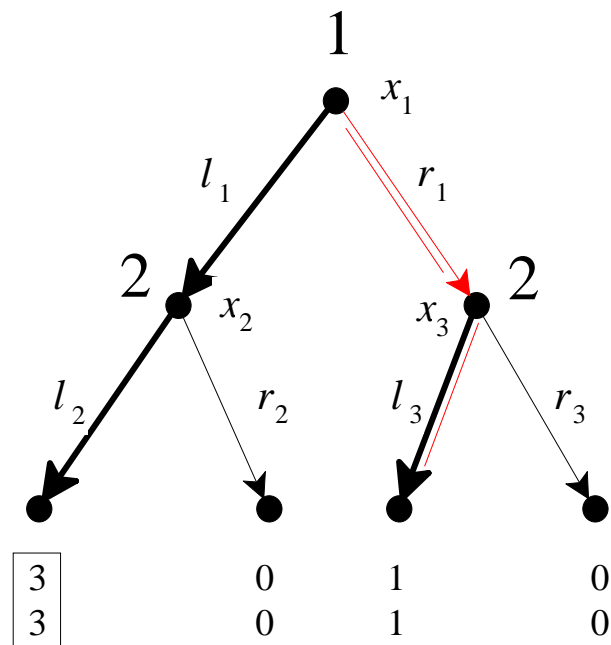
Rabinowicz, W., Grappling with the centipede, *Economics and Philosophy*, 1998, 14: 95-126.

Sugden, R., Rational choice: a survey of contributions from economics and philosophy, *Economic Journal*, 1991, 101:751-785.

Note: it is not necessarily the case that if  $\omega \in \Omega$  is such that at  $\omega$  there is common knowledge of rationality then  $\sigma(\omega)$  coincides with the backward-induction **strategy profile**. What is true is that player 1's strategy assigns the terminating choice to the root.



In general perfect-information games common knowledge of Rationality at Reached Nodes does **not** yield the backward-induction play.

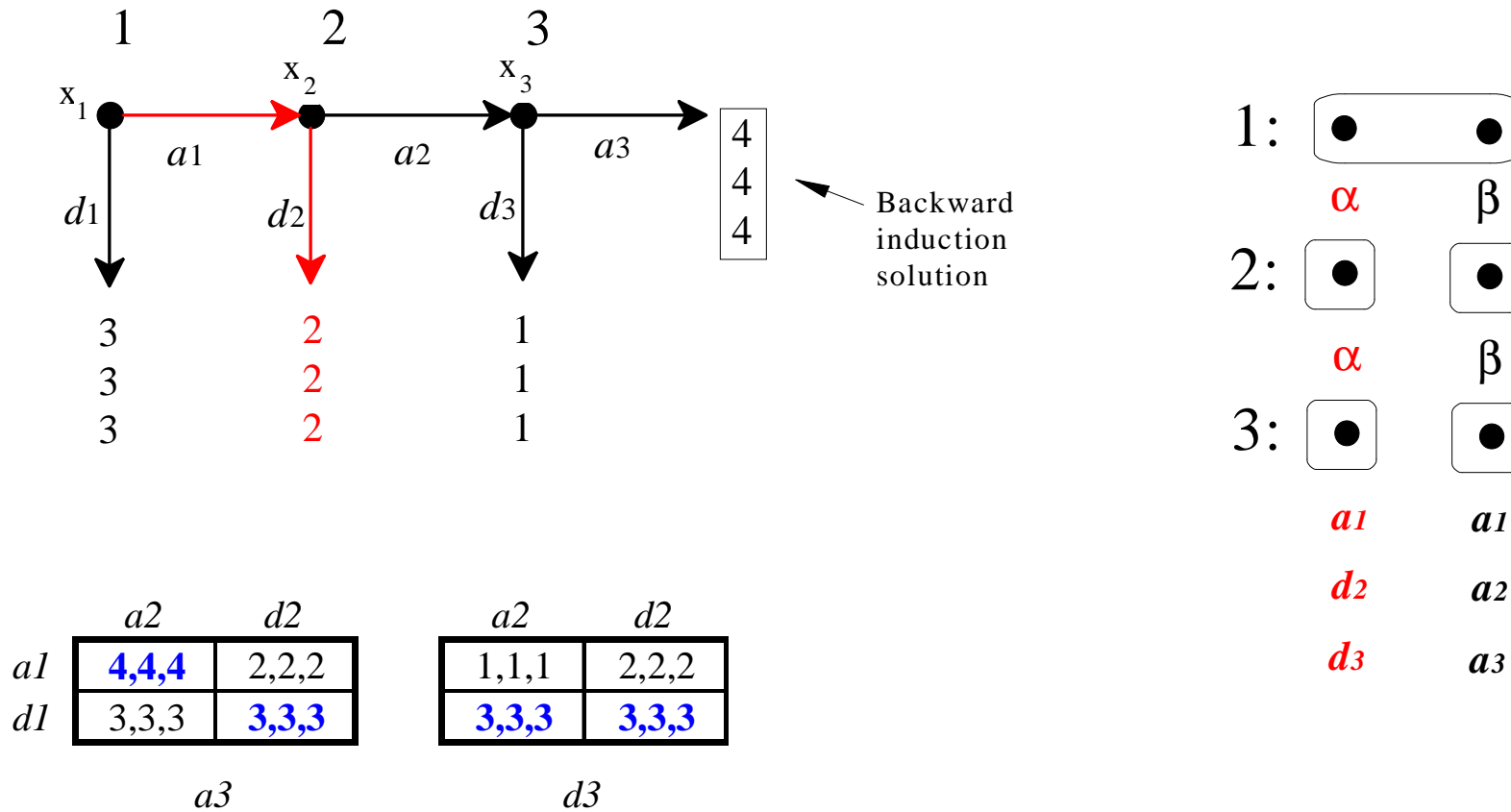


- 1: ●
- $\alpha$
- 2: ●
- r1*
- r2l3*

The backward induction play is  $l_1l_2$  while in this model we get  $r_1l_3$

$(r_1, r_2l_3)$  is a Nash equilibrium. Does common knowledge of Rationality at Reached Nodes at least yield a play that can be sustained by a Nash equilibrium?

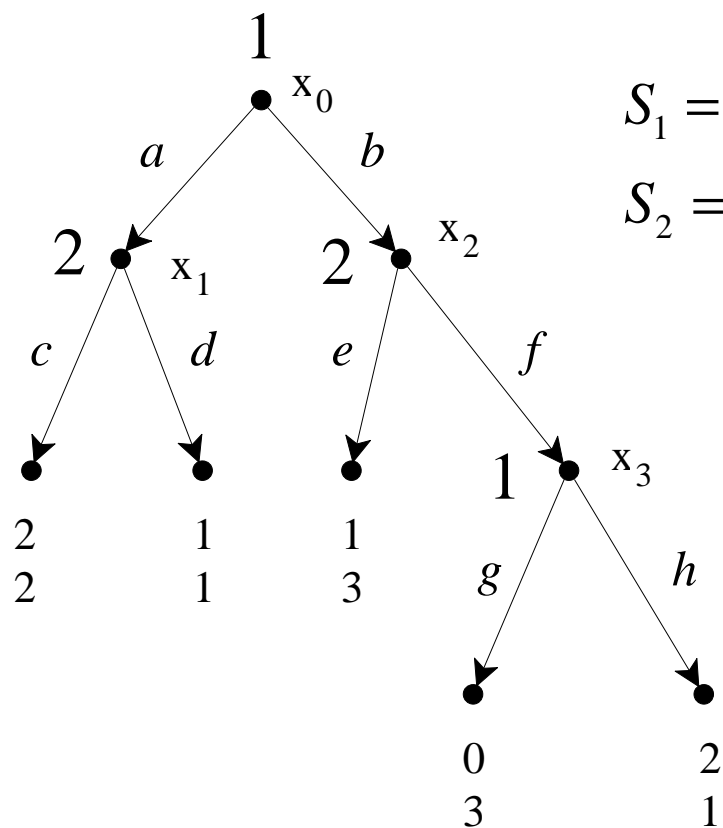
**NO!** In general, common knowledge of Rationality at Reached Nodes does not yield Nash equilibrium play



The Nash equilibria are marked in blue

## Dealing with general perfect-information games

Let  $x \in X_i$  be a decision node of player  $i$ . Denote by  $S_i^x$  the set of player  $i$ 's strategies in the subgame that starts at node  $x$ .



$$S_1 = \{ag, ah, bg, bh\}, \quad S_1^{x_3} = \{g, h\}$$

$$S_2 = \{cd, cf, de, df\}, \quad S_2^{x_1} = \{c, d\}, \quad S_2^{x_2} = \{e, f\}$$

Let  $x$  be a decision node of player  $i$  and let  $s_i^x, t_i^x \in S_i^x$   
be two strategies of player  $i$  in the subgame that starts at node  $x$

$s_i^x \succ_i t_i^x$  is interpreted as "for player  $i$ , strategy  $s_i^x$  is better than strategy  $t_i^x$   
in the subgame that starts at node  $x$ "

$s_i^x \succ_i t_i^x$  is true at state  $\omega$  if, **starting from node  $x$ ,**  
 $s_i^x$  gives a higher payoff to player  $i$  than  $t_i^x$  against  $\sigma_{-i}(\omega)$

Let  $\|s_i^x \succ_i t_i^x\|$  be the event that  $s_i^x \succ_i t_i^x$  is true.

If  $x$  is a node of player  $i$ , let  $\sigma_i(\omega)|_x$  denote the restriction of  
 $\sigma_i(\omega)$  to the subgame that starts at  $x$

If  $s_i^x \in S_i^x$ , let  $\|s_i^x\| = \{\omega \in \Omega : s_i^x = \sigma_i(\omega)|_x\}$

# SUBSTANTIVE RATIONALITY (Aumann, GEB 1995)

Recall that if  $E, F \subseteq \Omega$ ,  $E \rightarrow F$  is the event  $\neg E \cup F$  (if  $E$  then  $F$ )

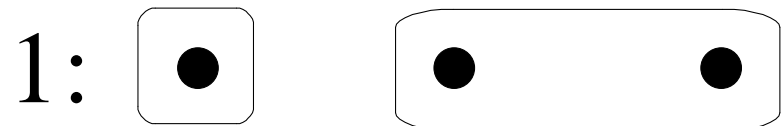
Let  $\mathbf{R}_i^{SR}$  be the event representing the proposition “player  $i$  is substantively rational”

$$\text{if } x \in X_i \quad \left\| s_i^x \right\| \cap K_i \left( \left\| t_i^x \succ_i s_i^x \right\| \right) \subseteq \neg \mathbf{R}_i^{SR}$$

$$\neg \mathbf{R}_i^{SR} = \bigcup_{x \in X_i} \bigcup_{s_i \in S_i^x} \bigcup_{t_i \in S_i^x} \left( \left\| s_i^x \right\| \cap K_i \left( \left\| t_i^x \succ_i s_i^x \right\| \right) \right)$$

$$\mathbf{R}^{SR} = \mathbf{R}_1^{SR} \cap \dots \cap \mathbf{R}_n^{SR} \quad \text{all players are substantively rational}$$





$\alpha$

$\beta$

$\gamma$



$a_1 d_3$

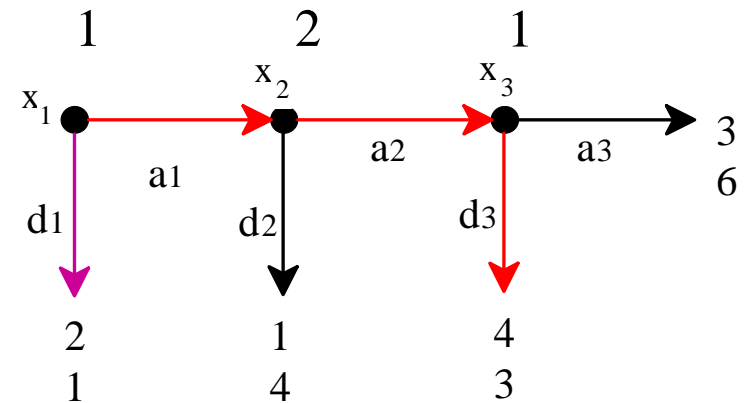
$d_1 d_3$

$d_1 d_3$

$a_2$

$a_2$

$d_2$



$$\mathbf{R}_2^{EA} = \{\alpha, \beta, \gamma\} \text{ (ex ante rationality)}$$


$$\mathbf{R}_2^{RN} = \{\beta, \gamma\} \text{ (rationality at reached nodes)}$$

$$\mathbf{R}_2^{SR} = \{\gamma\} \text{ (substantive rationality)}$$

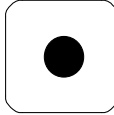
**PROPOSITION 3.** In every perfect information game,  $K_*R^{SR} \subseteq BI$

**PROPOSITION 4.** For every perfect information game, there is a model of it where  $K_*R^{SR} \neq \emptyset$

Aumann, R., Backward induction and common knowledge of rationality, *Games and Economic Behavior*, 1995, 8: 6-19.

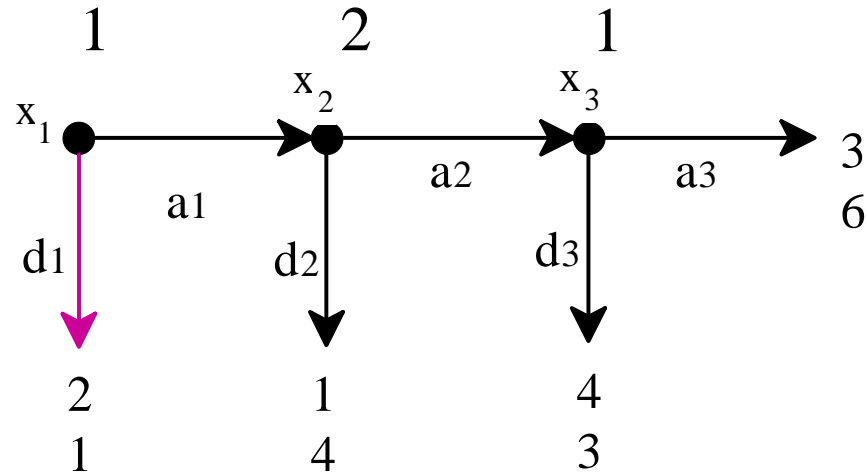
1: 

$\alpha$

2: 

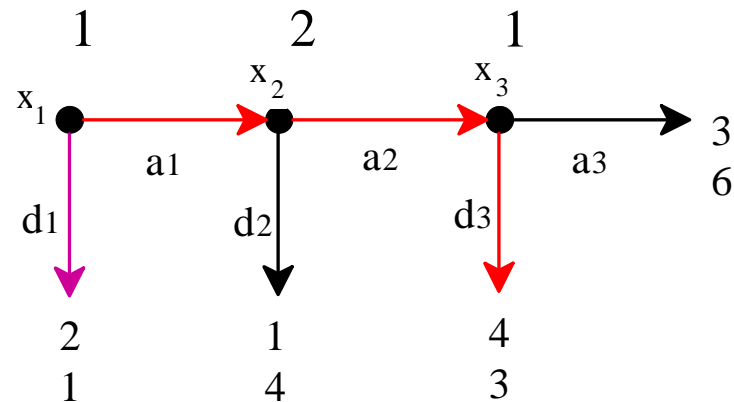
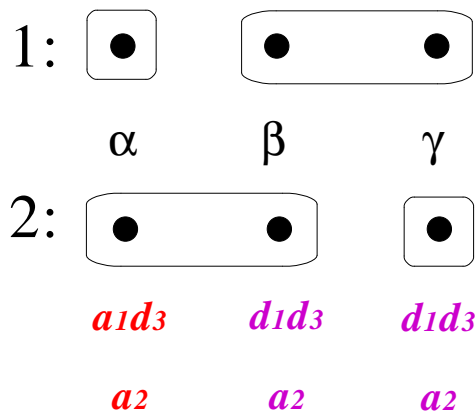
$d_1 a_3$

$d_2$



Why is player 2 substantively irrational at state  $\alpha$ ? What is true at state  $\alpha$  that makes player 2 substantively irrational?

At state  $\alpha$  player 2 is not taking any actions, because her node  $x_2$  is not reached. In fact, at state  $\alpha$  player 2 *knows* that her node is not reached. So what makes her irrational (according to the notion of substantive rationality) must be her *plan* to choose  $d_2$  *if her decision node were to be reached*. This is a *counterfactual* statement.



The association of a strategy profile with every state gives rise to *two types of counterfactuals*:

- (1) An objective statement about what the relevant player would do at a node that is not reached.
- (2) (With the help of the partitions) a subjective statement about what a player believes would happen if he were to take a different action from the one he is actually taking.

- (1) Thus at state  $\gamma$  it is true that **player 2** would take action  $a_2$  if her node  $x_2$  were to be reached (although it is not in fact reached and she knows that it is not reached)
- (2) At states  $\beta$  and  $\gamma$  **player 1** knows that if he were to take action  $a_1$  instead of  $d_1$  at the root (he knows that he is taking  $d_1$ ) then his payoff would be 4 (the payoff associated with  $a_1a_2d_3$ )

Modeling counterfactuals indirectly through strategies is not satisfactory. We have abandoned the modular approach suggested in Lecture 1, since there exists a module that deals with counterfactuals.

## Modeling Counterfactuals

For every  $\omega \in \Omega$ , let  $\mathcal{P}_\omega$  be a relation on  $\Omega$  satisfying,  $\forall \alpha, \beta \in \Omega$ ,

(1) either  $\alpha \in \mathcal{P}_\omega(\beta)$  or  $\beta \in \mathcal{P}_\omega(\alpha)$  (completeness)

(2) if  $\beta \in \mathcal{P}_\omega(\alpha)$  then  $\mathcal{P}_\omega(\beta) \subseteq \mathcal{P}_\omega(\alpha)$  (transitivity)

(3) if  $\alpha \in \mathcal{P}_\omega(\beta)$  and  $\beta \in \mathcal{P}_\omega(\alpha)$  then  $\alpha = \beta$  (antisymmetry)

(4)  $\omega' \in \mathcal{P}_\omega(\omega)$ , for all  $\omega' \in \Omega$  (centeredness)

The interpretation of  $\beta \in \mathcal{P}_\omega(\alpha)$  or  $\alpha \mathcal{P}_\omega \beta$  is that state  $\alpha$  is at least as close to

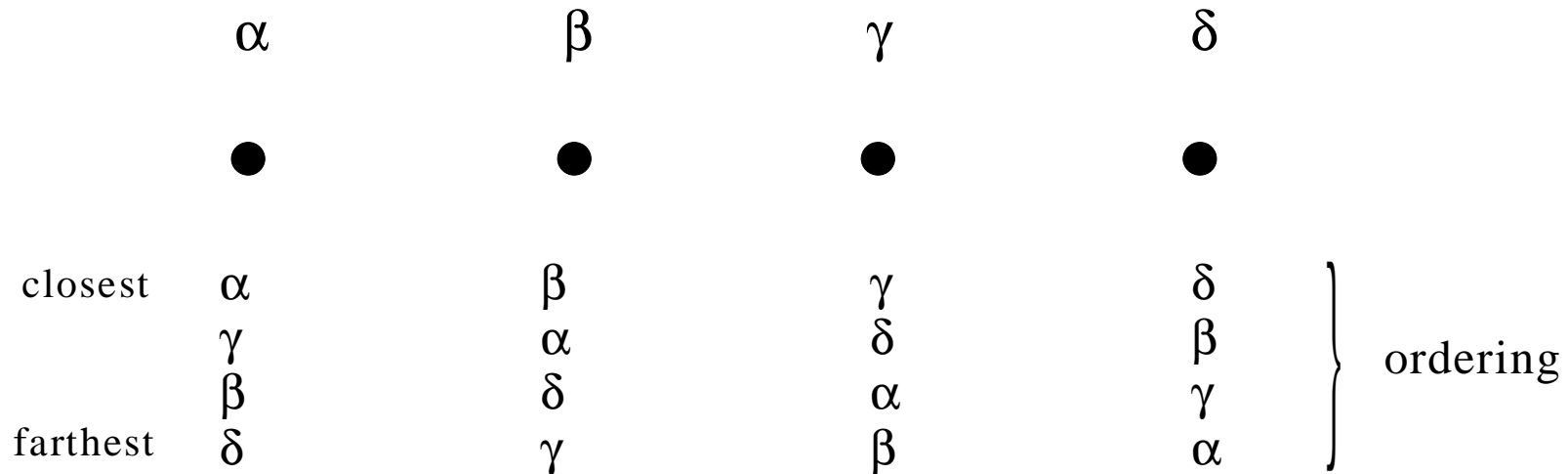
to state  $\omega$  as state  $\beta$  is. Thus, for every state  $\omega$ , the closeness relation  $\mathcal{P}_\omega$  determines

a strict ordering of the set of states based on closeness to  $\omega$ , with  $\omega$  itself being the closest state.

$\mathcal{P}_\omega(\alpha)$  = set of states that are not closer to  $\omega$  than  $\alpha$  is.

# REPRESENTATION

$$\Omega = \{\alpha, \beta, \gamma, \delta\}$$



Given a state  $\omega$  and an event  $E$ , denote by  $\min(\omega, E)$  the closest state to  $\omega$  that belongs to event  $E$ . Thus if  $\omega \in E$ , then  $\min(\omega, E) = \omega$ .

In the above example, if  $E = \{\beta, \delta\}$  then  $\min(\alpha, E) = \beta$

Recall that, if  $E, F \subseteq \Omega$  are two events,  $E \rightarrow F$  denotes the event  $\neg E \cup F$  (if  $E$  then  $F$ ). Thus  $\omega \in E \rightarrow F$  if either  $\omega \notin E$  or  $\omega \in E \cap F$ .

$\rightarrow$  represents the material conditional, which is true whenever the antecedent is false

We use the symbol  $\rightsquigarrow$  to denote the counterfactual conditional.

Thus  $E \rightsquigarrow F$  is interpreted as “if  $E$  were the case then  $F$  would be the case”

**Definition.**  $E \rightsquigarrow F = \{ \omega \in \Omega : \min(\omega, E) \in F \}$

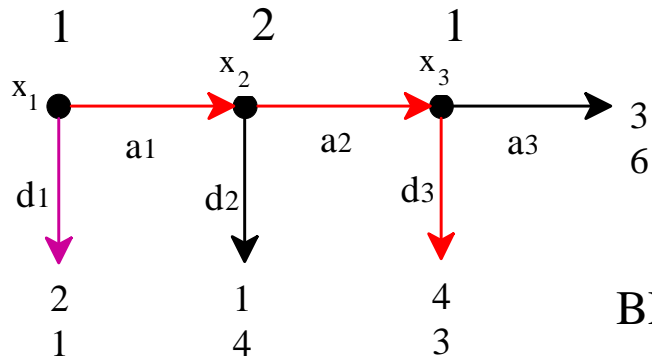
	$\alpha$	$\beta$	$\gamma$	$\delta$	
	●	●	●	●	
closest	$\alpha$	$\beta$	$\gamma$	$\delta$	If $E = \{\beta, \delta\}$ and $F = \{\alpha, \gamma, \delta\}$ then $E \rightsquigarrow F = \{\gamma, \delta\}$ while $E \rightarrow F = \{\alpha, \gamma, \delta\}$
	$\gamma$	$\alpha$	$\delta$	$\beta$	
	$\beta$	$\delta$	$\alpha$	$\gamma$	
farthest	$\delta$	$\gamma$	$\beta$	$\alpha$	

Note that, for all  $E, F \subseteq \Omega$ ,  $E \rightsquigarrow F \subseteq E \rightarrow F$

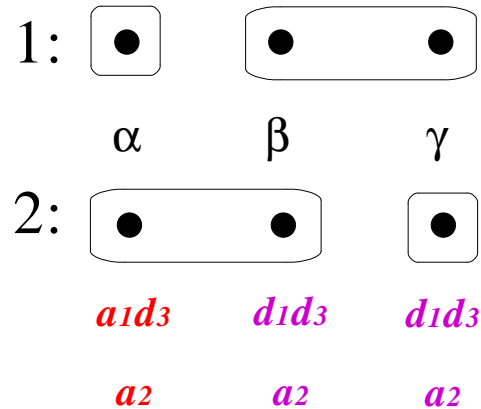
# MODELING STRATEGIES WITH COUNTERFACTUALS

Given a perfect information game define an epistemic model of it as before, but with the following changes:

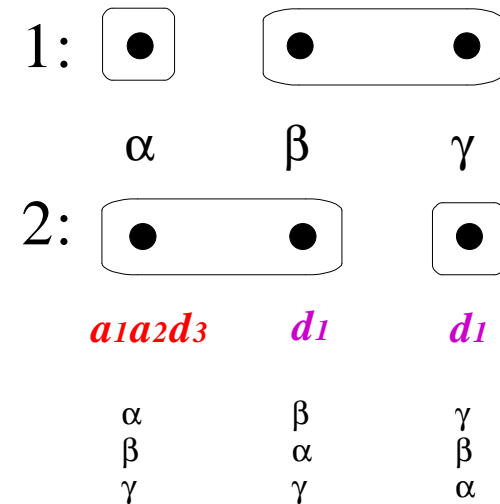
- (1) replace the  $n$  functions  $\sigma_i : \Omega \rightarrow S$  with a single function  $d : \Omega \rightarrow P$  where  $P$  is the set of plays of the game written in terms of actions taken,
- (2) add a set of closeness relations  $\{\mathcal{P}_\omega\}_{\omega \in \Omega}$



BEFORE



NOW

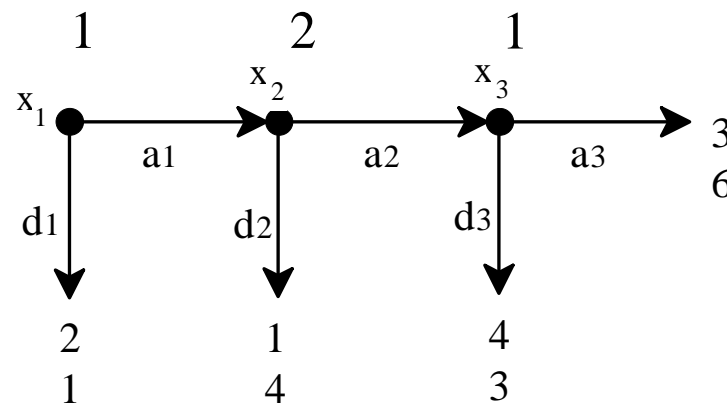
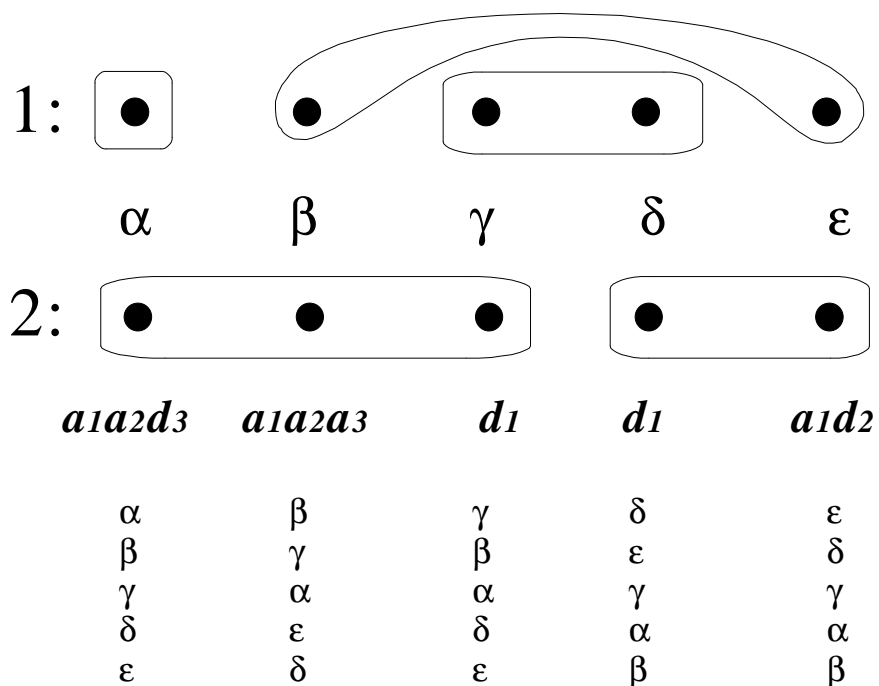




We add two more requirements:

(3) for every play there is at least one state where that play is realized

(4) if, at a state, node  $x$  of player  $i$  is reached and he takes action  $a$  there, then he knows that if  $x$  is reached he takes action  $a$ :  $\|a\| \subseteq K_i(\|x\| \rightarrow \|a\|)$



$$\|x_2\| = \{\alpha, \beta, \epsilon\}, \quad \|a_2\| = \{a, \beta\}$$

$$\|x_2\| \rightarrow \|a_2\| = \neg\|x_2\| \cup \|a_2\| = \{\alpha, \beta, \gamma, \delta\}$$

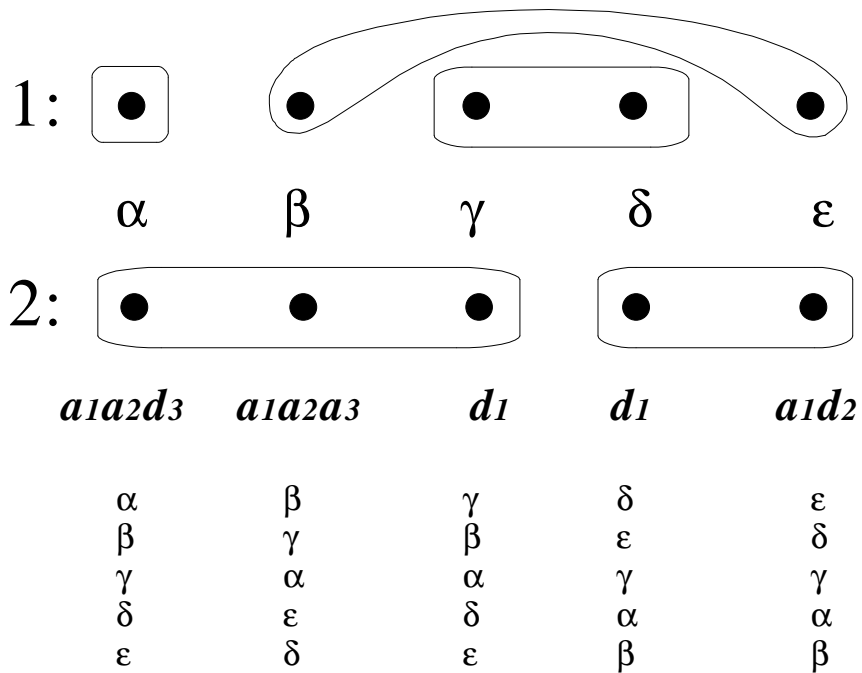
$$K_2(\|x_2\| \rightarrow \|a_2\|) = \{\alpha, \beta, \gamma\}$$

# EXTRACTING STRATEGIES FROM A MODEL

Given a model we can extract a strategy profile at every state as follows.

If  $s_i$  is a strategy of player  $i$  and  $x_i$  is a decision node of player  $i$ , denote by  $s_i(x_i)$  the choice prescribed by  $s_i$  at  $x_i$ .

Define  $\sigma_i(\omega)$  as follows:  $\sigma_i(\omega)(x_i) = c_i$  if and only if  $\omega \in \parallel x_i \parallel \rightsquigarrow \parallel c_i \parallel$



$$\sigma_1(\alpha) = a_1d_3, \quad \sigma_1(\beta) = a_1a_3$$

$$\sigma_1(\gamma) = d_1a_3 \quad (\text{for node } x_3 \text{ we use state } \beta)$$

$$\sigma_1(\delta) = d_1d_3 \quad (\text{for node } x_3 \text{ we use state } \alpha)$$

$$\sigma_1(\epsilon) = a_1d_3 \quad (\text{for node } x_3 \text{ we use state } \alpha)$$

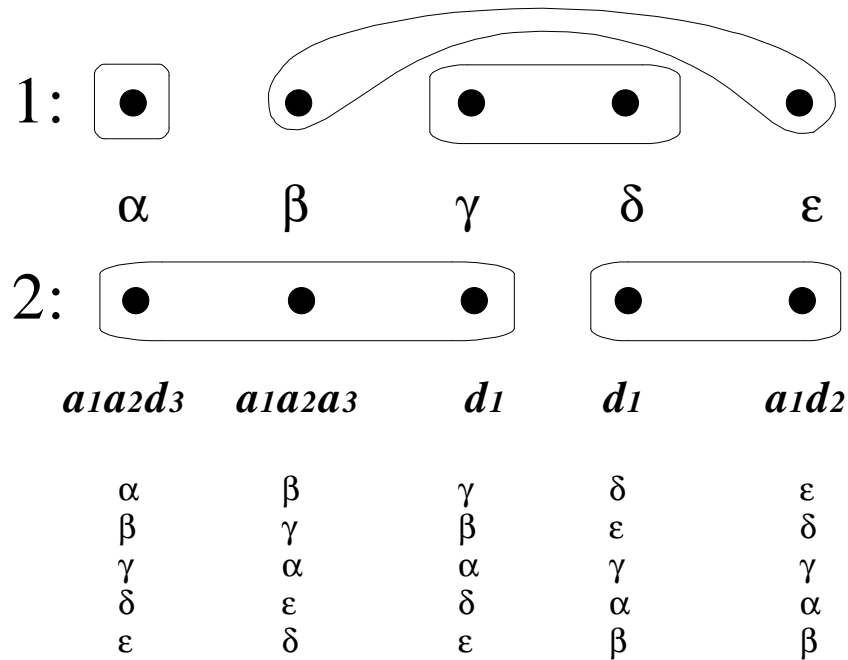
$$\sigma_2(\alpha) = a_2, \quad \sigma_2(\beta) = a_2$$

$$\sigma_2(\gamma) = a_2 \quad (\text{for node } x_2 \text{ we use state } \beta)$$

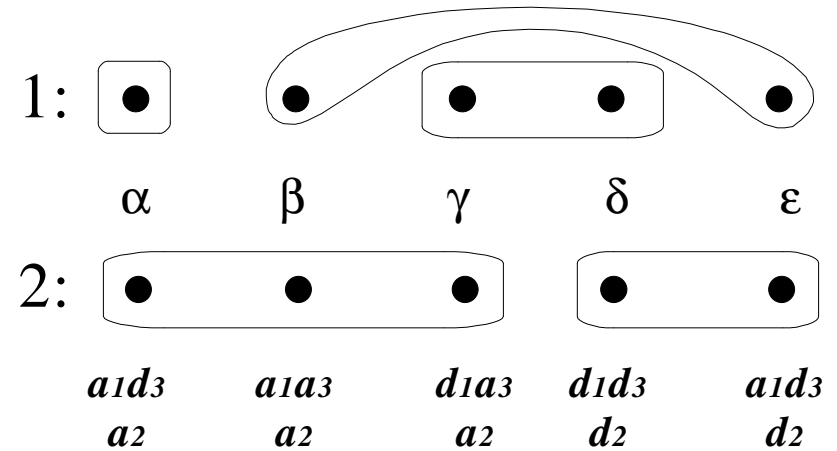
$$\sigma_2(\delta) = d_2 \quad (\text{for node } x_2 \text{ we use state } \epsilon)$$

$$\sigma_2(\epsilon) = d_2$$

From



We get



In this model it is not true that players know their own strategies. E.g. player 1 at state  $\gamma$

In order for a counterfactual model to give rise to a standard model based on strategies, we need to impose a further condition:

$$(5) \quad (\|x_i\| \rightsquigarrow \|c_i\|) \rightarrow K_i (\|x_i\| \rightsquigarrow \|c_i\|)$$

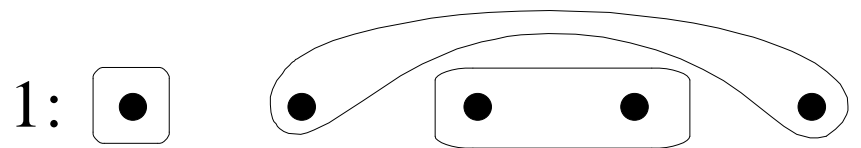
## RE-DEFINING RATIONALITY AT REACHED NODES

Let  $x_i$  be a decision node of player  $i$  and  $c_i$  and  $c_i'$  be two choices of player  $i$  at  $x_i$ .

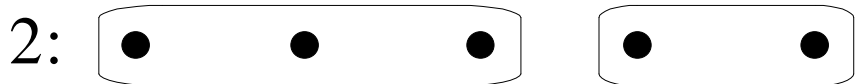
If  $m$  is a number, let  $\|\pi_i = m\|$  be the event that player  $i$ 's payoff is  $m$ .

If  $k$  and  $\ell$  are numbers, let  $\|k > \ell\| = \Omega$  if  $k > \ell$  and  $\|k > \ell\| = \emptyset$  otherwise.

$$\|c_i\| \cap \|\pi_i = k\| \cap K_i \left( \|x_i\| \rightarrow \left( \|c_i'\| \rightsquigarrow \|\pi_i = \ell\| \right) \right) \cap \|\ell > k\| \subseteq \neg R_i^{RN}$$



1:  $\alpha$     $\beta$     $\gamma$     $\delta$     $\epsilon$



2:  $a_1 a_2 d_3$     $a_1 a_2 a_3$     $d_1$     $d_1$     $a_1 d_2$

$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$
$\beta$	$\gamma$	$\beta$	$\epsilon$	$\delta$
$\gamma$	$\alpha$	$\alpha$	$\gamma$	$\gamma$
$\delta$	$\epsilon$	$\delta$	$\beta$	$\beta$
$\epsilon$	$\delta$	$\epsilon$	$\alpha$	$\alpha$

$\pi_1 = 4$   
 $d_1 \rightsquigarrow \pi_1 = 2$   
 $\pi_2 = 3$   
 $d_2 \rightsquigarrow \pi_2 = 4$   
 $R_1$   
 $R_2$

$\pi_1 = 3$   
 $d_1 \rightsquigarrow \pi_1 = 2$   
 $\pi_2 = 6$   
 $d_2 \rightsquigarrow \pi_2 = 4$   
 $R_1$   
 $R_2$

$\pi_1 = 2$   
 $a_1 \rightsquigarrow \pi_1 = 3$   
 $\pi_2 = 1$   
 no choices  
 by 2  
 $R_1$   
 $R_2$

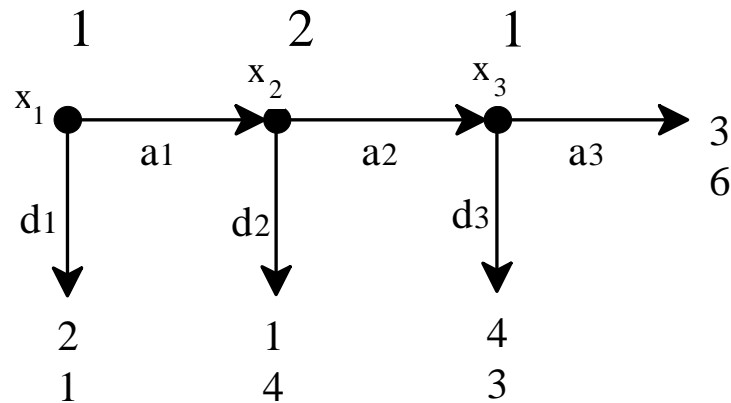
use  $\beta$

$\pi_1 = 2$   
 $a_1 \rightsquigarrow \pi_1 = 1$   
 $\pi_2 = 1$   
 no choices  
 by 2  
 $R_1$   
 $R_2$

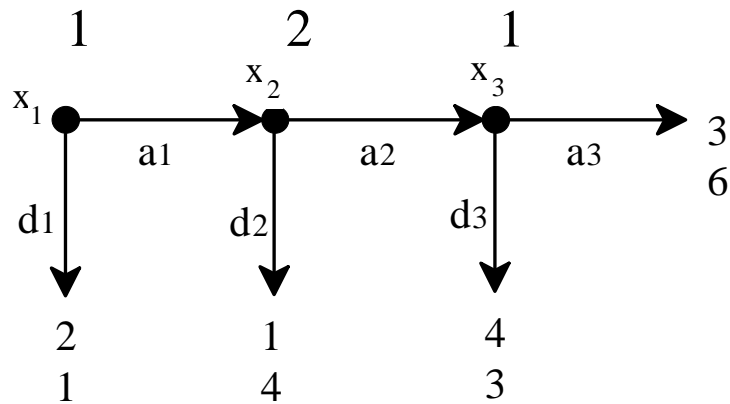
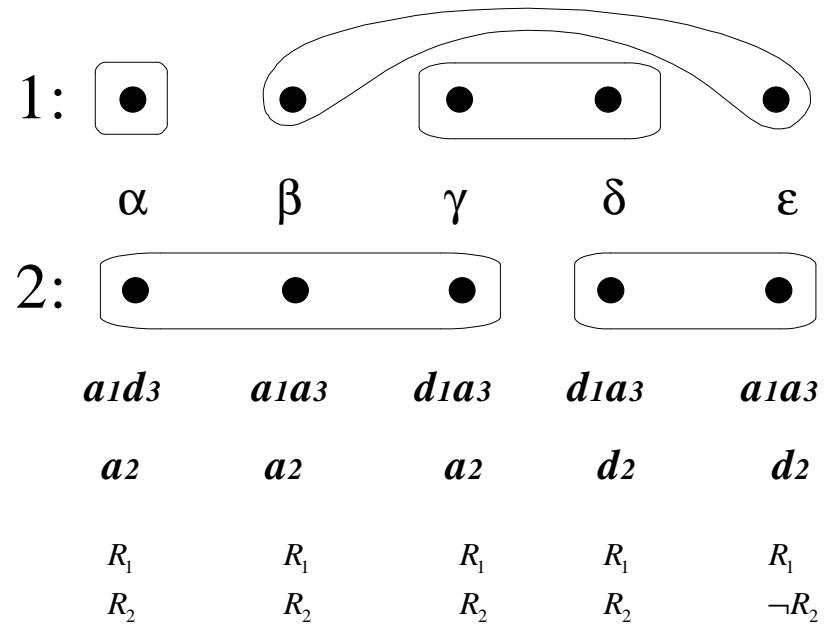
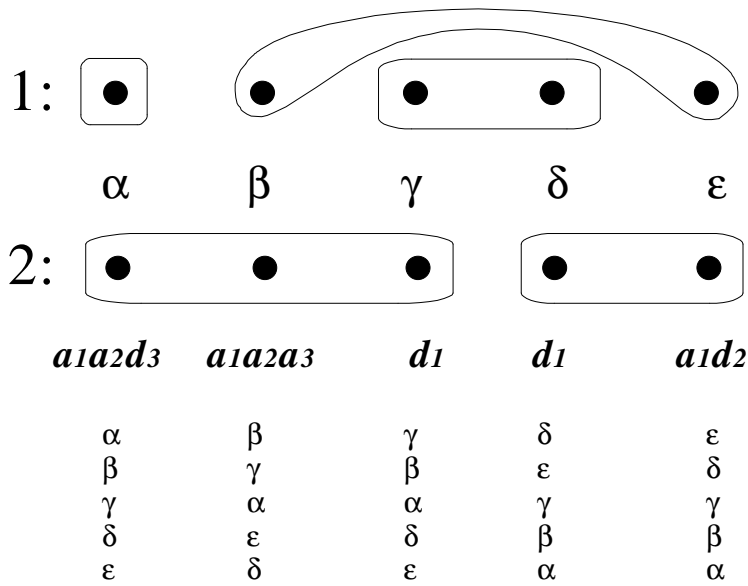
use  $\epsilon$

$\pi_1 = 1$   
 $d_1 \rightsquigarrow \pi_1 = 2$   
 $\pi_2 = 4$   
 $a_2 \rightsquigarrow \pi_2 = 6$   
 $R_1$   
 $\neg R_2$

use  $\beta$



Thus no common knowledge of rationality at any state.



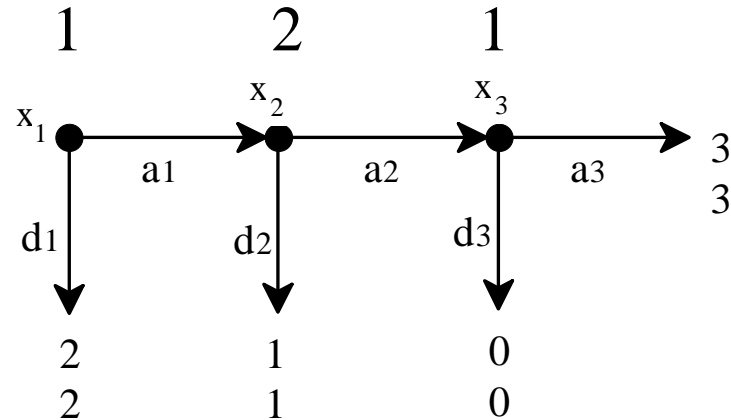
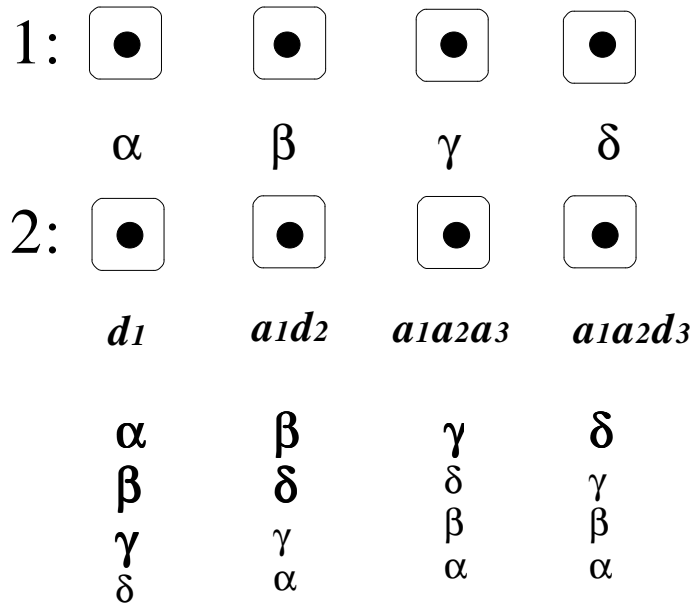
The corresponding strategy-based model

Redefining substantive rationality (Stalnaker's notion)

$$\mathbf{R}_i^{SR} = \bigcap_{x_i \in X_i} \left( \|x_i\| \rightsquigarrow \mathbf{R}_i^{RN} \right)$$

rationality at all nodes: reached and un-reached

Does common knowledge of substantial rationality so defined imply the backward-induction play?



$$R_1^{RN} = \{\alpha, \gamma\}$$

$$R_2^{RN} = \{\alpha, \beta, \gamma\}$$

At state  $\alpha$  there is common knowledge of substantive rationality. The following is true at  $\alpha$ :

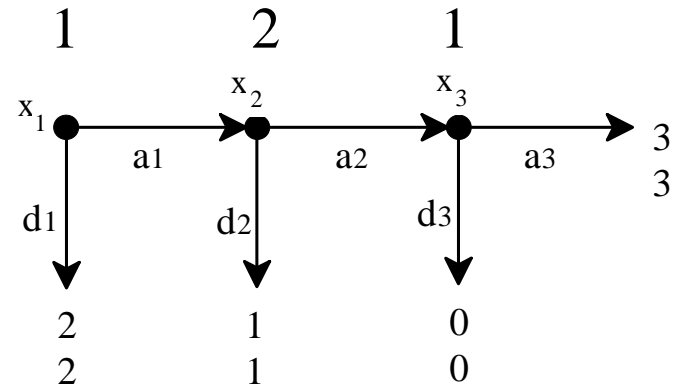
- (1) 1 is materially rational at  $x_1$  : 1 knows that if he played  $a_1$  then 2 would play  $d_2$ . [**state  $\beta$** ]
- (2) 2 is materially rational (does not do anything) but also substantively rational: if  $x_2$  were reached [**state  $\beta$** ] then player 2 would be materially rational (she would play  $d_2$  knowing that if she played  $a_2$  then 1 would play  $d_3$ ) [**state  $\delta$** ].
- (3) 1 is substantively rational at  $x_3$  : if  $x_3$  were reached he would play  $a_3$  [**state  $\gamma$** ].

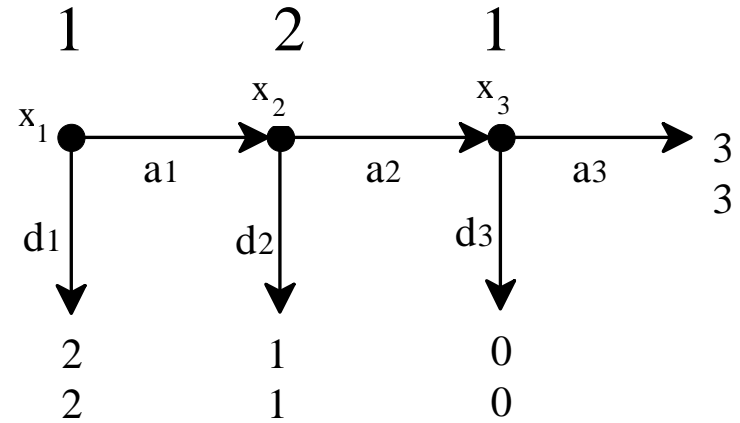
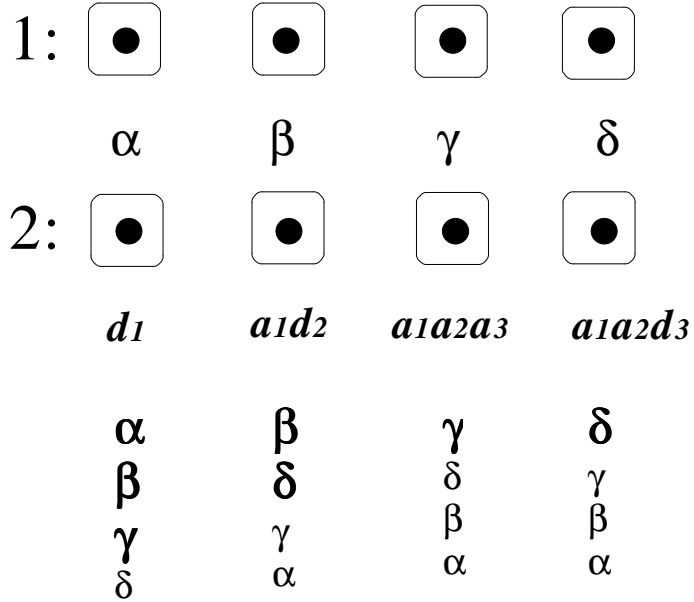


Stalnaker (1998 p. 48)

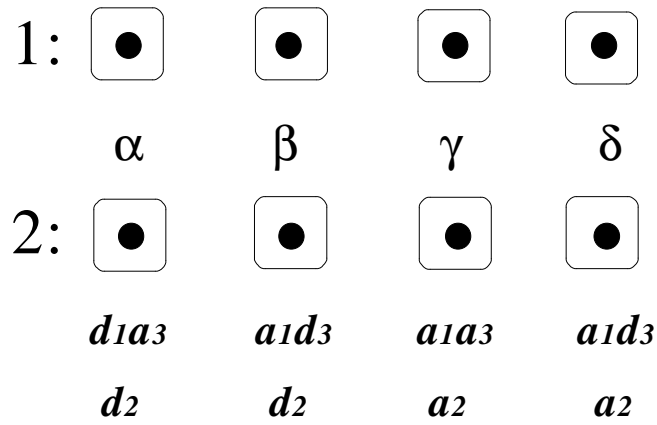
Player 2 has the following initial belief: player 1 would choose  $a_3$  on her second move *if* she had a second move. This is a causal ‘if’ – an ‘if’ used to express 2’s opinion about 1’s *disposition to act* in a situation that they both know will not arise. Player 2 knows that since player 1 is rational, if she somehow found herself at her second node, she would choose  $a_3$ . But to ask what player 2 would believe about player 1 *if* he learned that he was wrong about 1’s first choice is to ask a completely different question – this ‘if’ is epistemic; it concerns player 2’s belief revision policies, and not player 1’s disposition to be rational. No assumption about player 1’s substantive rationality, or about player 2’s knowledge of her substantive rationality, can imply that player 2 should be disposed to maintain his belief that she will act rationally on her second move even were he to learn that she acted irrationally on her first.

1:				
	$\alpha$	$\beta$	$\gamma$	$\delta$
2:				
	$d_1$	$a_1d_2$	$a_1a_2a_3$	$a_1a_2d_3$
	$\alpha$	$\beta$	$\gamma$	$\delta$
	$\beta$	$\delta$	$\delta$	$\delta$
	$\gamma$	$\gamma$	$\beta$	$\gamma$
	$\delta$	$\alpha$	$\alpha$	$\alpha$

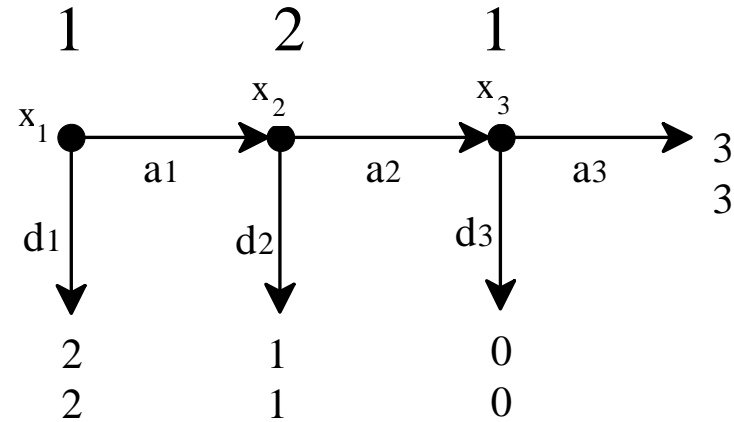
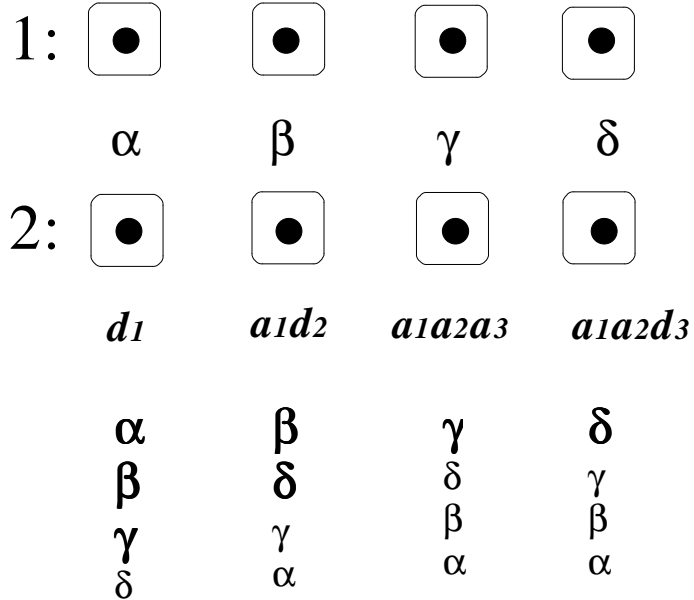




The corresponding strategy-based model is:



According to Aumann, player 2 is not substantively rational at  $\alpha$ : player 2 is planning to play  $d_2$  knowing that player 1 would play  $a_3$ .



$$\alpha \in K_2 \left( \|x_3\| \rightsquigarrow a_3 \right) \text{ and also } \alpha \in \|x_2\| \rightsquigarrow K_2 \left( \|x_3\| \rightsquigarrow d_3 \right)$$

Thus what player 2 believes about player 1's behavior in the hypothetical world where node  $x_3$  is reached changes going from node  $x_1$  (where the game ends without node  $x_2$  being reached) to the hypothetical world where  $x_2$  is reached. *If one imposes the constraint that such changes cannot happen, then common knowledge of substantive rationality implies the backward-induction play.*

## ADDITIONAL REFERENCES

- Aumann, R., Backward induction and common knowledge of rationality, *Games and Economic Behavior*, 1995, 8: 6-19.
- Halpern, J., Substantive rationality and backward induction, *Games and Economic Behavior*, 2001, 37: 425-435.
- Samet, D., Hypothetical knowledge and games with imperfect information, *Games and Economic Behavior*, 1996, 17: 230-251.
- Stalnaker, R., Belief revision in games: forward and backward induction, *Mathematical Social Sciences*, 1998, 36: 31–56