

Part I : Standard beliefs in static games

Lecture 1:

Common belief in rationality

2 Common Belief in Rationality

Intuitive idea:

When making a choice, you should not only choose rationally yourself, but you should also **believe that your opponents choose rationally.**

Not only this, you should also **believe that your opponents believe that their opponents choose rationally.**

Etcetera.

2.1 Example: Where to locate my pub?

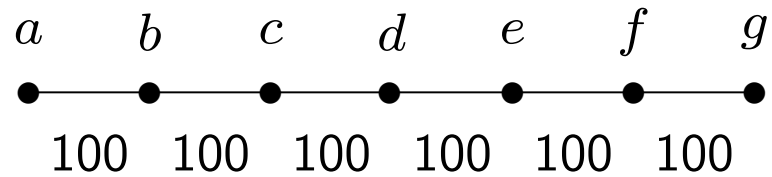
Story: You and your competitor should locate a pub somewhere on a long street.

There are seven possible locations: $\{a, b, c, d, e, f, g\}$.

Between every two locations there are 100 thirsty men.

Every man goes to the pub that is nearest to his house.

Which location(s) should you choose?



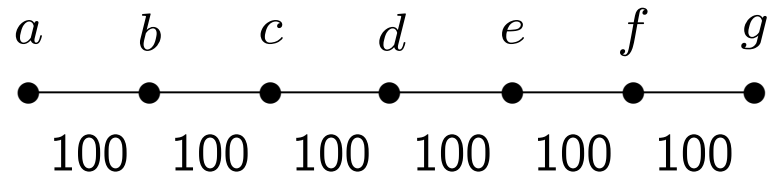
Location b is optimal if you believe that your opponent chooses a .

Location c is optimal if you believe that your opponent chooses b .

Location d is optimal if you believe that your opponent chooses c , d or e .

Location e is optimal if you believe that your opponent chooses f .

Location f is optimal if you believe that your opponent chooses g .



We say: b, c, d, e, f are **rational choices**, since they can be supported by **some** belief about the opponent's choice.

Choosing b is always better than choosing a .

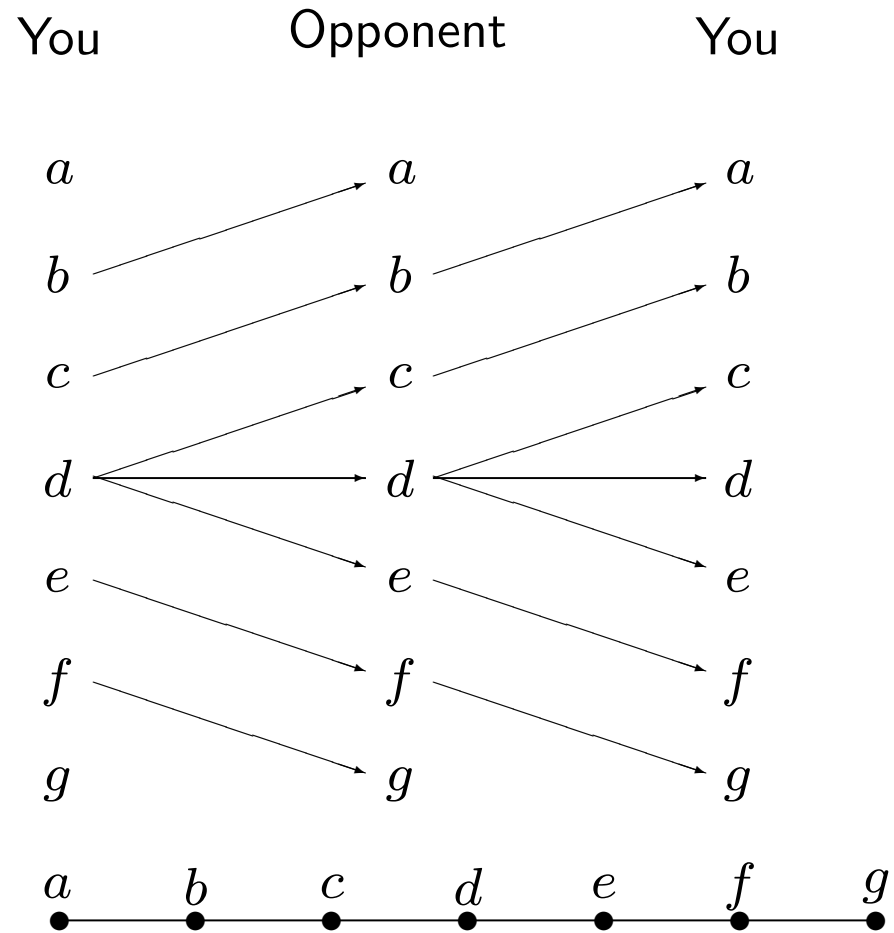
We say: a is **strictly dominated** by b .

Similarly, g is strictly dominated by f .

So, there are no beliefs for which a and g are optimal.

We say: a and g are **irrational choices**.

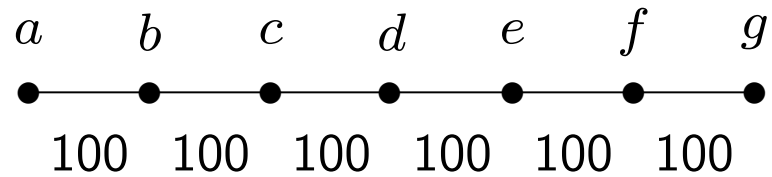
These findings can be summarized by a **beliefs diagram**:



Are each of your rational choices b, c, d, e and f **reasonable**? **No!**

You should not only choose rationally yourself, but you should also **believe that your opponent chooses rationally.**

If a rational choice can only be supported by believing that your opponent chooses irrationally, then this choice is **unreasonable.**

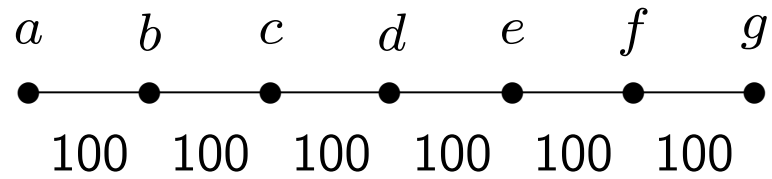


Choice b is only optimal if you believe, with positive probability, that your opponent **irrationally** chooses a .

Therefore, b is **unreasonable**.

Similarly, f is only optimal if you believe, with positive probability, that your opponent **irrationally** chooses g .

Hence, also f is **unreasonable**.



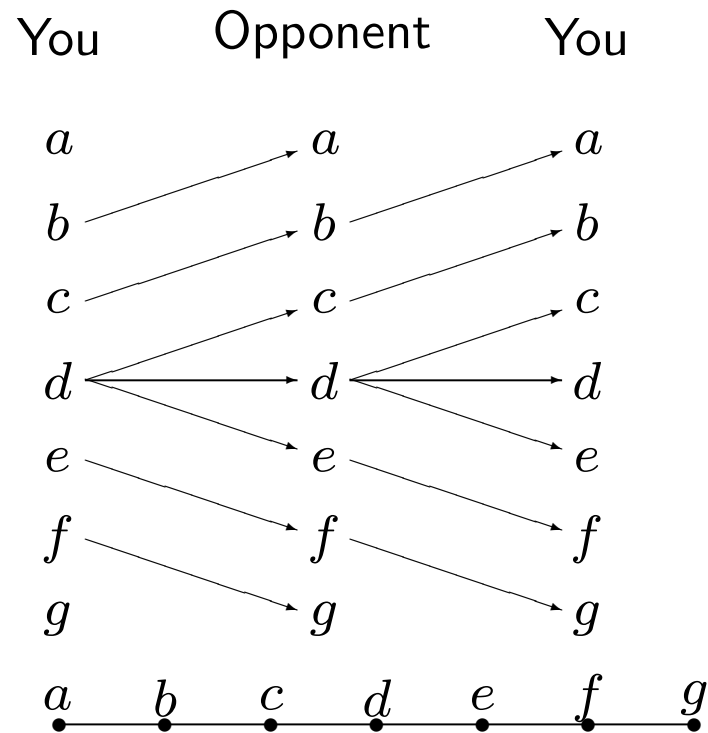
So, if you

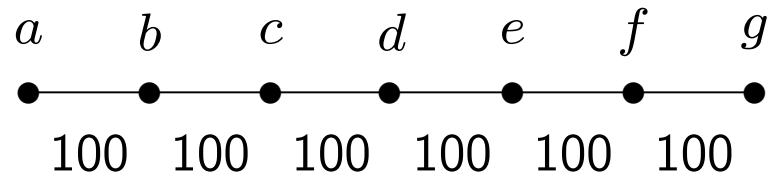
- **choose rationally,** and
- **believe that your opponent chooses rationally,**

you must choose from $\{c, d, e\}$.

Only c, d and e can be chosen rationally if you believe that your opponent chooses rationally.

Beliefs diagram

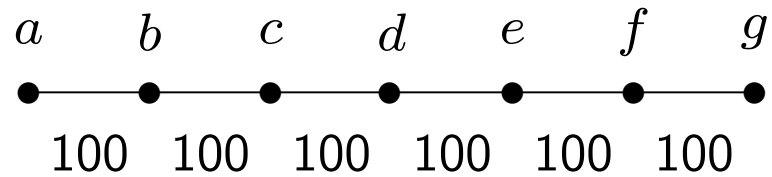




Are each of your choices c , d and e **reasonable**? **No!**

You should not only believe that your opponent chooses rationally, but you should also **believe that your opponent believes that you choose rationally.**

That is, you should believe that your opponent chooses from $\{c, d, e\}$.



However, if you believe that your opponent chooses from $\{c, d, e\}$, then only d is optimal for you.

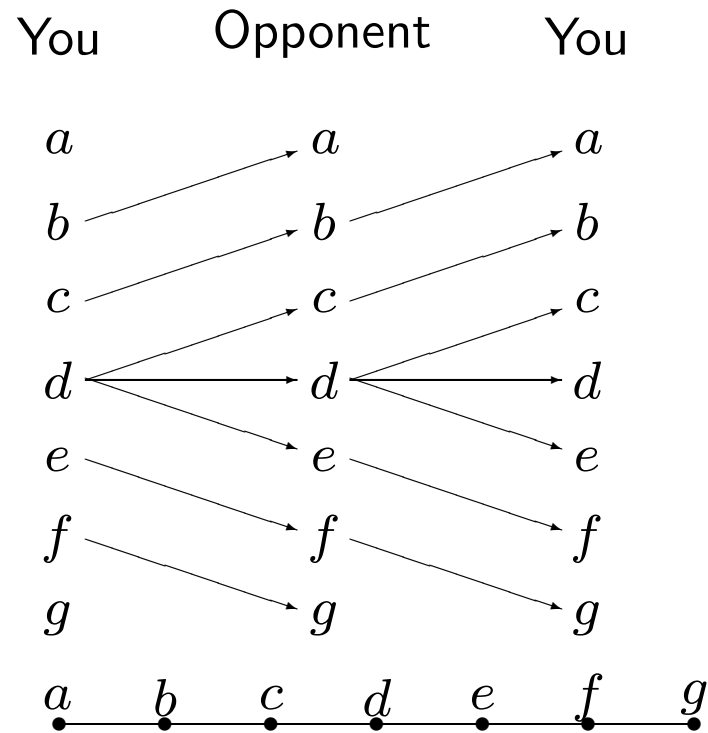
So, if you

- choose rationally,
- believe that your opponent chooses rationally, and
- believe that your opponent believes that you choose rationally,

then you must choose d .

So, **common belief in rationality** leads to a unique choice for you, namely d .

Beliefs diagram



2.2 Example: Going to a party.

Story: You and your friend are going to a party.

You must both decide which color to wear.

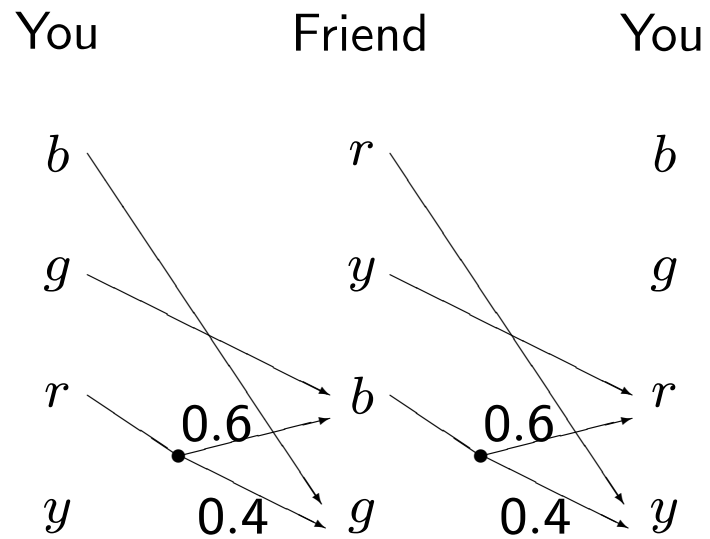
The preferences are as follows:

	blue	green	red	yellow
you	4	3	2	1
friend	2	1	4	3

If your friend chooses the same color as you, the utility would be 0 for both.

Which color(s) can be chosen under **common belief in rationality**?

Beliefs diagram



	blue	green	red	yellow
you	4	3	2	1
friend	2	1	4	3

	blue	green	red	yellow
you	4	3	2	1
friend	2	1	4	3

Yellow is an **irrational choice** for you:

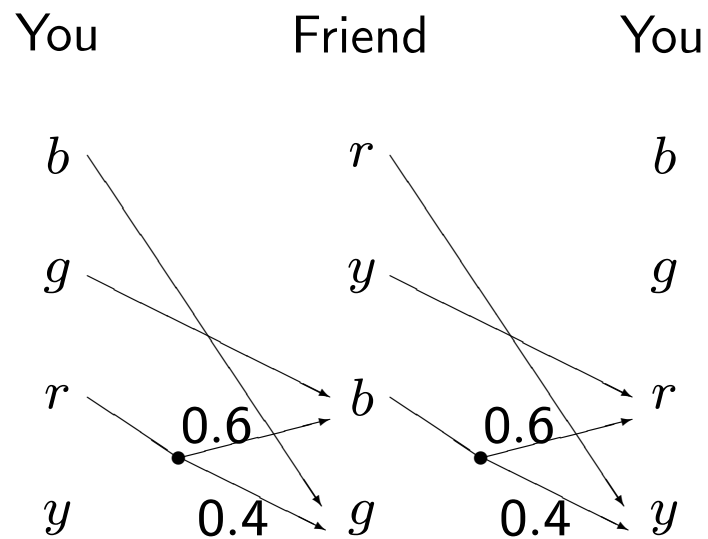
Suppose, you believe that your opponent chooses blue with prob. α , green with prob. β and red with prob. γ . Then,

$$\begin{aligned} u(b) &= 4(\beta + \gamma) \\ u(g) &= 3(\alpha + \gamma) \\ u(r) &= 2(\alpha + \beta), \end{aligned}$$

and hence

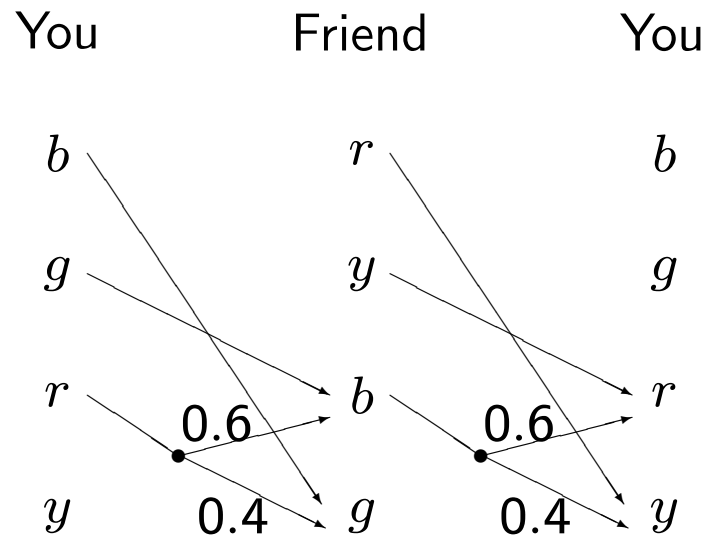
$$u(b) + u(g) + u(r) = 5\alpha + 6\beta + 7\gamma \geq 5,$$

so either $u(b) > 1$ or $u(g) > 1$ or $u(r) > 1$.



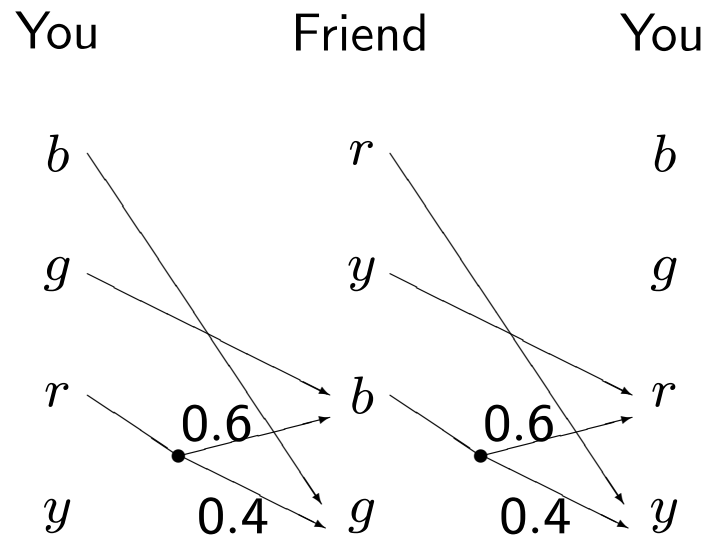
If you **believe that your friend chooses rationally**, you should **not** choose red.

	blue	green	red	yellow
you	4	3	2	1
friend	2	1	4	3



If you **believe that your friend chooses rationally**, and **believe that your friend believes that you choose rationally**, you should choose **blue**.

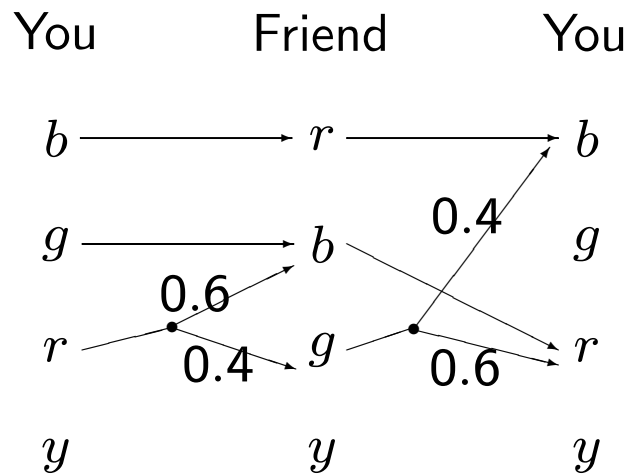
	blue	green	red	yellow
you	4	3	2	1
friend	2	1	4	3



So, **common belief in rationality** leads to a unique choice, namely wearing blue (your most preferred color).

	blue	green	red	yellow
you	4	3	2	1
friend	2	1	4	3

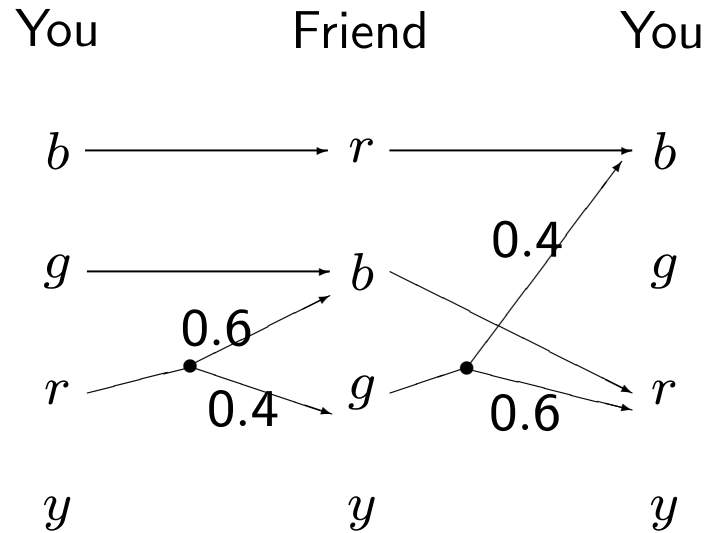
Things change if we change your friend's preferences.



Now, **common belief in rationality** allows for choosing blue, green and red.

	blue	green	red	yellow
you	4	3	2	1
friend	3	2	4	1

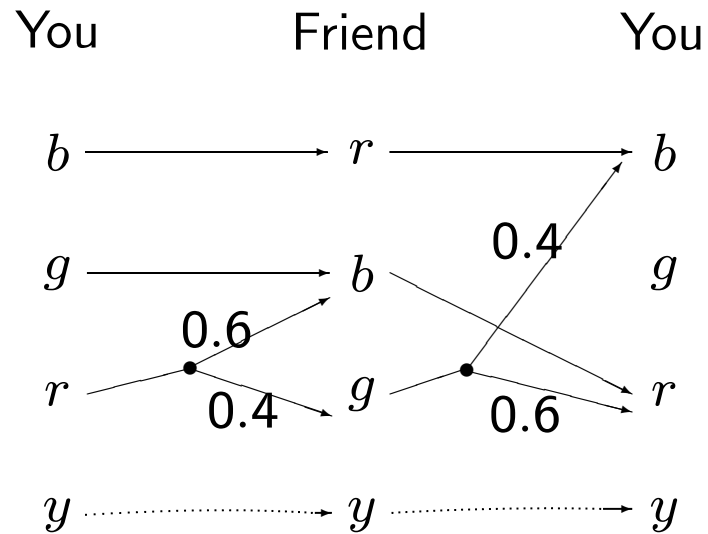
2.3 Beliefs Diagram



A **beliefs diagram** assigns to every **rational** choice a **probabilistic belief** about the opponents' choices for which this choice is optimal.

	blue	green	red	yellow
you	4	3	2	1
friend	3	2	4	1

Extended beliefs diagram



A beliefs diagram can be **extended** to an **extended beliefs diagram** by assigning beliefs to all **irrational** choices as well.

2.4 Epistemic Model: The Main Idea

Main idea:

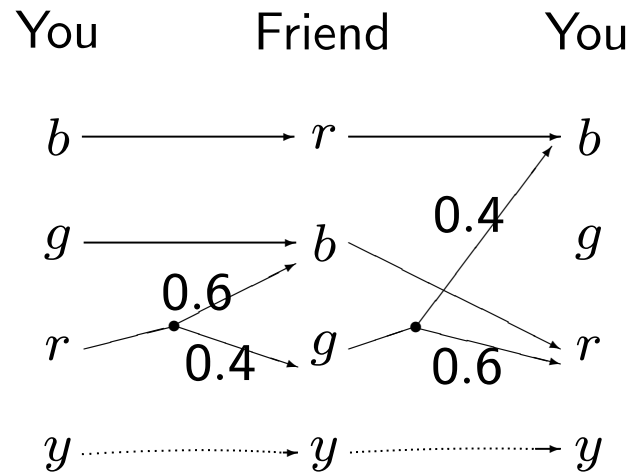
Construct an extended beliefs diagram.

Every **chain of arrows** in the extended beliefs diagram represents a **belief hierarchy**.

Every belief hierarchy is called a **type**.

All types together form an **epistemic model**.

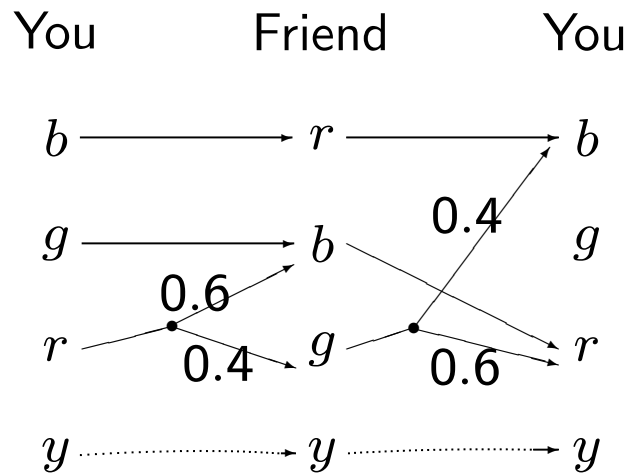
Extended beliefs diagram



Starting at your choice b , we obtain the following **belief hierarchy**:

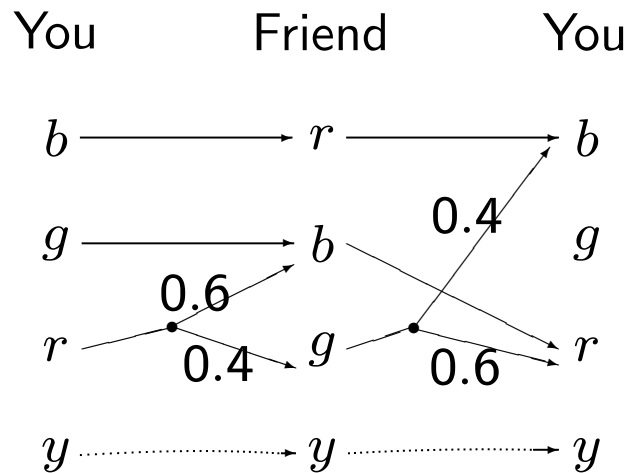
- you believe that your friend chooses r ,
- you believe that your friend believes that you choose b ,
- you believe that your friend believes that you believe that your friend chooses r ,

Extended beliefs diagram



Starting at your choice g , we obtain the following **belief hierarchy**:

- you believe that your friend chooses b ,
- you believe that your friend believes that you choose r ,
- you believe that your friend believes that you believe that, with prob. 0.6, your friend chooses b , and with prob. 0.4 your friend chooses g , ...



Starting at your choice r , we obtain the following **belief hierarchy**:

You believe that, with prob. 0.6,

- your friend chooses b , believes that you choose r , believes that you believe that ...

and you believe that, with prob. 0.4,

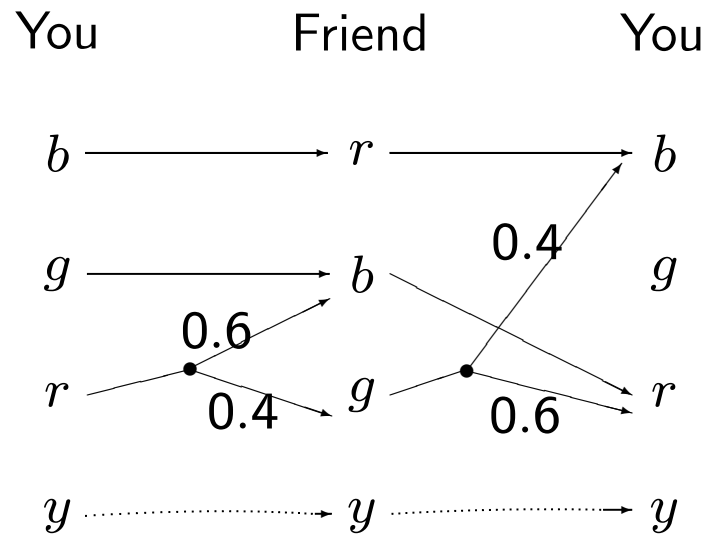
- your friend chooses g , believes that, with prob. 0.4, you choose b and believe that ..., and with 0.6 you choose r and believe that ...

There is a more compact way to represent these belief hierarchies:

Identify every belief hierarchy with a **type**.

Every type holds a belief about the opponent's choice **and the opponent's type**.

The belief hierarchy of a type can then be **deduced** from this model.



Types with their associated beliefs:

Types: $\{t_1^b, t_1^g, t_1^r, t_1^y\}$	Types: $\{t_2^b, t_2^g, t_2^r, t_2^y\}$
$b_1(t_1^b) = (r, t_2^r)$	$b_2(t_2^r) = (b, t_1^b)$
$b_1(t_1^g) = (b, t_2^b)$	$b_2(t_2^b) = (r, t_1^r)$
$b_1(t_1^r) = 0.6(b, t_2^b) + 0.4(g, t_2^g)$	$b_2(t_2^g) = 0.4(b, t_1^b) + 0.6(r, t_1^r)$
$b_1(t_1^y) = (y, t_2^y)$	$b_2(t_2^y) = (y, t_1^y)$

2.5 Epistemic Model: Formal Definition

Let $\Gamma = (C_i, u_i)_{i \in I}$ be a **finite, static game** where

- $I = \{1, \dots, n\}$ is the finite set of **players**,
- C_i is the finite set of **choices** for player i , and
- u_i is player i 's **utility function**.

u_i assigns to every combination of choices (c_1, \dots, c_n) a utility $u_i(c_1, \dots, c_n) \in \mathbf{R}$.

A **finite epistemic model** for Γ is a tuple $\mathbf{M} = (T_i, b_i)_{i \in I}$ where

- T_i is the finite set of **types** for player i , and
- b_i is a function that assigns to every type $t_i \in T_i$ a **probabilistic belief** $b_i(t_i) \in \Delta(C_{-i} \times T_{-i})$.

Here, $C_{-i} := \times_{j \neq i} C_j$, and $T_{-i} := \times_{j \neq i} T_j$.

$\Delta(C_{-i} \times T_{-i})$ denotes the set of probability distributions on $C_{-i} \times T_{-i}$.

Consider a type $t_i \in T_i$ and a choice $c_i \in C_i$. Then,

$$u_i(c_i, t_i) := \sum_{(c_{-i}, t_{-i}) \in C_{-i} \times T_{-i}} b_i(t_i)(c_{-i}, t_{-i}) u_i(c_i, c_{-i})$$

is t_i 's **expected utility** from choosing c_i .

Choice c_i is **rational** for type t_i if

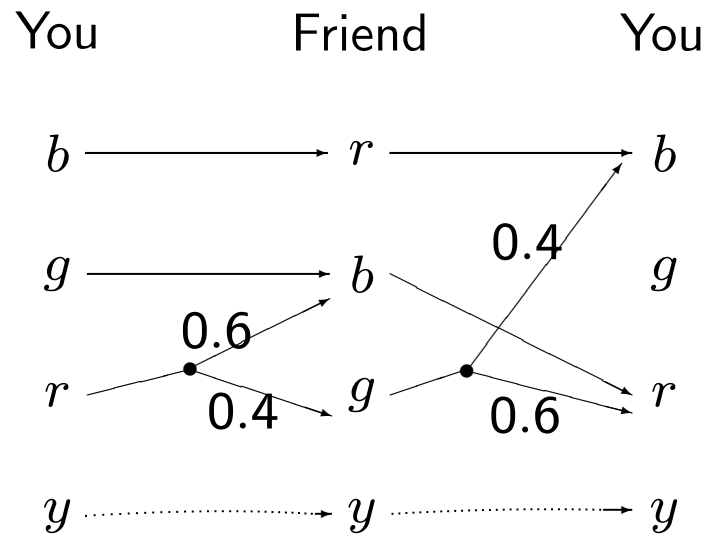
$$u_i(c_i, t_i) \geq u_i(c'_i, t_i)$$

for all other choices c'_i .

2.6 Common Belief in Rationality

With the definition of an epistemic model at hand, we can formally define **common belief in rationality**.

Say that type t_i **believes in the opponents' rationality** if $b_i(t_i)$ only assigns positive probability to choice-type pairs (c_j, t_j) where c_j is rational for t_j .



Types t_1^b, t_1^g and t_1^r believe in the opponent's rationality, but type t_1^y does not.

Types t_2^b, t_2^g and t_2^r believe in the opponent's rationality, but type t_2^y does not.

For every player j , let $E_j \subseteq T_j$ be a set of player j types.

Type t_i **believes in** E_j if $b_i(t_i)$ only assigns positive probability to player j types in E .

For example: Take $E_j :=$ set of player j types that believe in the opponents' rationality.

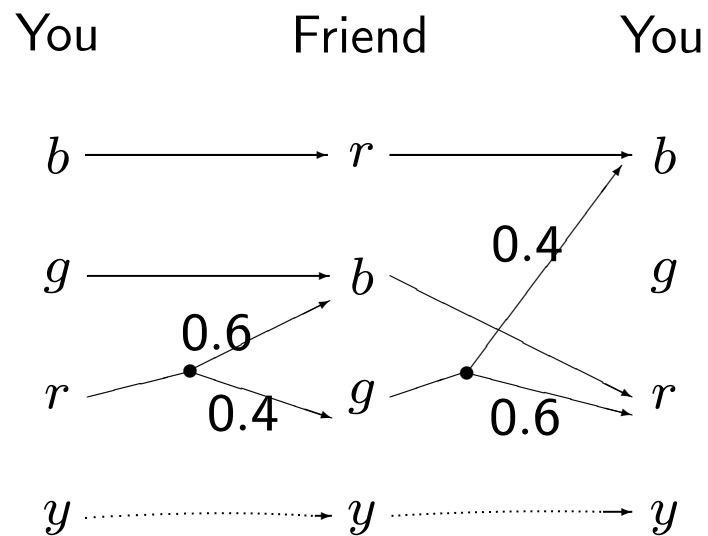
If t_i believes in E_j , then t_i believes that player j believes in his opponents' rationality.

Common belief in rationality (Tan and Werlang, 1988)

$$\begin{aligned} T_i^1 & : = \{t_i \in T_i \mid t_i \text{ believes in the opponents' rationality}\} \\ T_i^2 & : = \{t_i \in T_i^1 \mid t_i \text{ believes in } T_j^1 \text{ for all } j \neq i\} \\ & \vdots \\ T_i^k & : = \{t_i \in T_i^{k-1} \mid t_i \text{ believes in } T_j^{k-1} \text{ for all } j \neq i\} \\ & \vdots \end{aligned}$$

Type t_i expresses **common belief in rationality** if

$$t_i \in T_i^k \text{ for every } k.$$



Types t_1^b, t_1^g and t_1^r express **common belief in rationality**.

So, you can rationally choose b, g and r under common belief in rationality.

In general:

Player i can **rationally choose** $c_i \in C_i$ **under common belief in rationality** if:

there is an epistemic model $\mathbf{M} = (T_j, b_j)_{j \in I}$, and

a type $t_i \in T_i$ such that

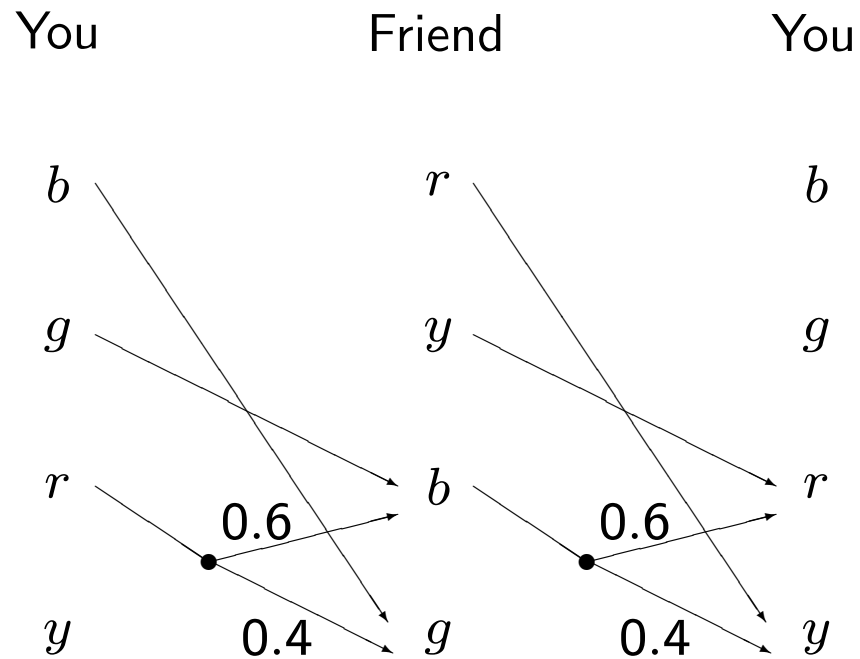
t_i expresses **common belief in rationality**, and

c_i is **rational** for t_i .

2.7 Algorithm

Question: Is there an **algorithm** that computes **all** choices for player i that can rationally be chosen under common belief in rationality?

First step: Try to characterize rational and irrational choices in an easy way.



Yellow is irrational. It is strictly dominated by the randomized choice in which you choose blue and green with probability 0.5.

	blue	green	red	yellow
you	4	3	2	1
friend	2	1	4	3

A choice c_i is **rational** if there is some probabilistic belief $b_i \in \Delta(C_{-i})$ such that

$$u_i(c_i, b_i) \geq u_i(c'_i, b_i)$$

for all other choices c'_i .

A **randomized choice** is a probability distribution $\mu_i \in \Delta(C_i)$.

Choice c_i is **strictly dominated** by the randomized choice μ_i if

$$u_i(c_i, c_{-i}) < u_i(\mu_i, c_{-i})$$

for every $c_{-i} \in C_{-i}$.

Lemma 2.1.

A choice c_i is **rational** if and only if it is **not strictly dominated** by a randomized choice.

For every player j , let $D_j \subseteq C_j$. Let $D_{-i} := \times_{j \neq i} D_j$.

Choice c_i is **strictly dominated on** D_{-i} if there is some randomized choice $\mu_i \in \Delta(C_i)$ such that

$$u_i(c_i, c_{-i}) < u_i(\mu_i, c_{-i})$$

for every $c_{-i} \in D_{-i}$.

Lemma 2.2.

Let $D_j \subseteq C_j$, and $D_{-i} := \times_{j \neq i} D_j$.

A choice c_i is **optimal for some belief in $\Delta(D_{-i})$** if and only if it is **not strictly dominated on D_{-i}** .

So, choice c_i is **rational** for player i if and only if c_i is not strictly dominated on C_{-i} .

Algorithm: Iterated strict dominance.

$$\begin{aligned} C_i^1 & : = \{c_i \in C_i \mid c_i \text{ not strictly dominated on } C_{-i}\} \\ C_i^2 & : = \{c_i \in C_i^1 \mid c_i \text{ not strictly dominated on } C_{-i}^1\} \\ & \vdots \\ C_i^k & : = \{c_i \in C_i^{k-1} \mid c_i \text{ not strictly dominated on } C_{-i}^{k-1}\} \\ & \vdots \end{aligned}$$

Choice c_i survives **iterated strict dominance** if $c_i \in C_i^k$ for every k .

Theorem 2.3: (Tan and Werlang, 1988)

A choice c_i can rationally be chosen under **common belief in rationality**

if and only if

c_i survives **iterated strict dominance**.

Proof:

(a) Assume that t_i is a type that expresses **common belief in rationality**.

To show: Every rational choice for t_i survives **iterated strict dominance**.

Suppose that \hat{c}_i is rational for t_i . We show that $\hat{c}_i \in C_i^k$ for every k .

Step 1:

$$C_i^1 := \{c_i \in C_i \mid c_i \text{ not strictly dominated on } C_{-i}\}.$$

Since \hat{c}_i is optimal for $b_i(t_i) \in \Delta(C_{-i})$, we know by Lemma 2.1 that \hat{c}_i is not strictly dominated on C_{-i} , so $\hat{c}_i \in C_i^1$.

Step 2:

$$C_i^2 := \{c_i \in C_i^1 \mid c_i \text{ not strictly dominated on } C_{-i}^1 \}.$$

Since t_i **believes in the opponents' rationality**, $b_i(t_i)$ should only assign positive probability to opponents' choices that are optimal for some belief.

So, by Step 1, $b_i(t_i) \in \Delta(C_{-i}^1)$.

Hence, c_i is optimal for $b_i(t_i) \in \Delta(C_{-i}^1)$.

By Lemma 2.2, it follows that c_i is not strictly dominated on C_{-i}^1 , so $c_i \in C_i^2$.

Step 3:

$$C_i^3 := \{c_i \in C_i^2 \mid c_i \text{ not strictly dominated on } C_{-i}^2 \}.$$

Since t_i **believes in the opponents' rationality**, and **believes that his opponents believe in their opponents' rationality**, $b_i(t_i)$ should only assign positive probability to opponents' choices that can rationally be made if the opponents believe in their opponents' rationality.

But then, by Step 2, $b_i(t_i) \in \Delta(C_{-i}^2)$.

So, c_i is optimal for $b_i(t_i) \in \Delta(C_{-i}^2)$.

Hence, by Lemma 2.2, c_i is not strictly dominated on C_{-i}^2 , so $c_i \in C_i^3$.

And so on. Hence, $c_i \in C_i^k$ for all k .

(b) To show: All choices that survive **iterated strict dominance** can rationally be chosen under common belief in rationality.

Let $(D_i)_{i \in I}$ be the sets of choices that survive iterated strict dominance.

By construction, every $d_i \in D_i$ is not strictly dominated on D_{-i} .

Hence, by Lemma 2.2, for every $d_i \in D_i$ there is some belief $b_i^{d_i} \in \Delta(D_{-i})$ under which d_i is optimal.

Make an epistemic model $(T_i, b_i)_{i \in I}$ where

$$T_i := \{t_i^{d_i} \mid d_i \in D_i\}$$

for all players i , and where

$$b_i(t_i^{d_i})((c_j, t_j)_{j \neq i}) := \begin{cases} b_i^{d_i}((c_j)_{j \neq i}), & \text{if } c_j \in D_j \text{ and } t_j = t_j^{c_j} \text{ for all } j \neq i \\ 0, & \text{otherwise} \end{cases} .$$

Then, d_i is rational for $t_i^{d_i}$ for every player i and every $d_i \in D_i$.

As a consequence, every $t_i^{d_i} \in T_i$ expresses **common belief in rationality**.

So, every $d_i \in D_i$ can rationally be chosen under common belief in rationality.

This completes the proof.

2.8 Related Models

In **common belief in rationality**, player i 's belief about j 's choice may be **correlated with** his belief about k 's choice.

For instance,

$$b_1(t_1) = \frac{1}{2}((t_2, a), (t_3, b)) + \frac{1}{2}((\hat{t}_2, c), (\hat{t}_3, d)).$$

In the model of **rationalizability** (Bernheim (1984) and Pearce (1984)) it is assumed that player i 's belief about j 's choice should be **independent** from i 's belief about k 's choice.

A choice c_i is **rationalizable** if it can rationally be chosen under

- common belief in rationality, and
- common belief in event that players' beliefs about opponents are **independent**.

For two-player games, rationalizability is equivalent to common belief in rationality.

In the model by Brandenburger and Friedenberg (2007) it is assumed that player i 's belief about j 's choice is **conditionally independent** from i 's belief about k 's choice.

For, instance, the belief

$$b_1(t_1) = \frac{1}{2}((t_2, a), (t_3, b)) + \frac{1}{2}((\hat{t}_2, c), (\hat{t}_3, d))$$

is **conditionally independent**:

Conditional on the opponents' types being t_2 and t_3 , the beliefs about the opponents' choices are independent.

Conditional on the opponents' types being \hat{t}_2 and \hat{t}_3 , the beliefs about the opponents' choices are independent.

The belief

$$b_1(t_1) = \frac{1}{2}((t_2, a), (t_3, b)) + \frac{1}{2}((t_2, c), (t_3, d))$$

is **not conditionally independent**:

Conditional on the opponents' types being t_2 and t_3 , the beliefs about the opponents' choices are correlated.

Brandenburger and Friedenberg (2007) assume common belief in rationality and common belief in event that players have **conditionally independent beliefs**.

2.9 References

Tommy Chin-Chiu Tan and Sérgio Ribeiro da Costa Werlang (1988): “The Bayesian foundations of solution concepts of games”, *Journal of Economic Theory* 45, 370-391.

Kaushik Basu (1994): “The traveler’s dilemma: Paradoxes of rationality in game theory”, *American Economic Review* 36, 391-396.

Douglas Bernheim (1984): “Rationalizable strategic behavior”, *Econometrica* 52, 1007-1028.

David Pearce (1984): “Rationalizable strategic behavior and the problem of perfection”, *Econometrica* 52, 1029-1050.

Adam Brandenburger and Amanda Friedenberg (2007): “Intrinsic correlation in games”, Forthcoming in *Journal of Economic Theory*, Downloadable from <http://pages.stern.nyu.edu/~abranden/>