Logique épistémique

Pierre Lescanne

29 novembre 2004 – 15 h 52

Il *croit* qu'il est Napoléon,
mais *tout le monde sait*
que c'est moi.

Plan

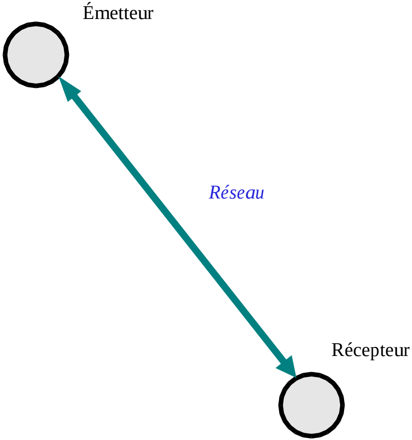
Plan

Un protocole émetteur-récepteur

Les noeuds ◯ transmettent les messages entre l'émetteur et le récepteur :

- ils *peuvent dupliquer* des messages,
- ils *peuvent perdre* des messages,
- cependant, ils ne peuvent pas *perdre indéfiniment* un même message.

C'est le principe d'Internet : «*faire de son mieux*» (en anglais «*the best effort*»). Le protocole s'appelle TCP (pour *Transmission Control Protocol*).


Émetteur
Réseau
Récepteur

Un protocole émetteur-récepteur (fin)

Tant que l'émetteur *ne sait pas* si le récepteur a reçu un message donné m_i, il le ré-émet.
Le récepteur accuse réception d'un message en émettant un message d'**accusé réception** ack_i tant qu'il *ne sait pas* si l'émetteur a reçu cet accusé réception.

Un protocole émetteur-récepteur (fin)

Tant que l'émetteur **ne sait pas** si le récepteur a reçu un message donné m_i, il le ré-émet.
Le récepteur accuse réception d'un message en émettant un message d'**accusé réception** ack_i tant qu'il **ne sait pas** si l'émetteur a reçu cet accusé réception.

Plan

L'attaque coordonnée

- Deux généraux et leurs armées sur deux collines,
- Ils doivent attaquer **ensemble** et chaque général doit être sûr que l'autre général attaquera en même temps.
- Ils communiquent par des messagers
  - qui mettent une heure pour aller d'un camp à l'autre,
  - qui peuvent se perdre dans le noir ou être capturés.

L'attaque coordonnée

- Deux généraux et leurs armées sur deux collines,
- Ils doivent attaquer **ensemble** et chaque général doit être sûr que l'autre général attaquera en même temps.
- Ils communiquent par des messagers
  - qui mettent une heure pour aller d'un camp à l'autre,
  - qui peuvent se perdre dans le noir ou être capturés.

Comment coordonner une attaque ?

Le général 1 envoie des messagers au général 2

Mais le messager peut être capturé ou être tué !

Mais le messager peut se perdre !

L'attaque coordonnée

Le général 1 choisit une heure pour l'attaque, disons H
et envoie une messager.

À l'arrivée du messager, le général 2 accepte l'heure H
et envoie une messager avec son accord.

L'attaque coordonnée

Le général 1 choisit une heure pour l'attaque, disons H
et envoie une messager.

À l'arrivée du messager, le général 2 accepte l'heure H
et envoie une messager avec son accord.

Le général 1 attaquera à l'heure H si il sait que le général 2 connaît l'heure qu'il a proposée et l'accepte.

L'attaque coordonnée

Le général 1 choisit une heure pour l'attaque, disons H
et envoie une messager.

À l'arrivée du messager, le général 2 accepte l'heure H
et envoie une messager avec son accord.

Le général 1 attaquera à l'heure H si il sait que le général 2 connaît l'heure qu'il a proposée et l'accepte.

Le général 2 attaquera à l'heure H si il (général 2) sait que le général 1 sait qu'il (général 2) connaît l'heure proposée et l'accepte.

Le général 1 doit envoyer un second messager avec un accord.

Le général 1 attaquera à l’heure H si il (général 1) sait que le général 2 sait qu’il (général 1) sait que le général 2 connaît l’heure proposée et l’accepte.

Le général 2 doit envoyer un second messager avec un accord.

Le général 1 attaquera à l'heure H si il (général 1) sait que le général 2 sait qu'il (général 1) sait que le général 2 connaît l'heure proposée et l'accepte.

Le général 2 doit envoyer un second messager avec un accord.

Le général 2 attaquera à l'heure H si il (général 2) sait que le général 1 sait qu'il (général 2) sait que le général 1 sait qu'il (général 2) connaît l'heure proposée et l'accepte.

Le général 1 doit envoyer un troisième messager avec un accord.

$\vdots$

L'attaque coordonnée

Le processus ne va jamais s'arrêter.

L’attaque coordonnée

Le processus ne va jamais s’arrêter.

*On peut démontrer qu’***avec des communications asynchrones***, une attaque coordonnée* **n’est pas possible***.*

L'attaque coordonnée

Le processus ne va jamais s'arrêter.

*On peut démontrer qu'***avec des communications asynchrones***, une attaque coordonnée* **n'est pas possible***.*

L'acquisition d'une **connaissance commune** *n'est pas possible de façon asynchrone.*

Plan

Le secrétaire américain à la Défense Donald Rumsfeld, lors d'un point de presse en février 2002 :

> *«Les informations annonçant que quelque chose n'a pas eu lieu m'intéressent toujours pour la bonne raison que, comme vous le savez, ce sont des nouvelles connues ; il y a des choses que nous savons que nous savons»*

> *«Nous savons aussi qu'il y a des choses inconnues ; ce qui revient à dire que nous savons qu'il y a certaines choses dont nous ne savons rien. Mais il existe aussi des nouvelles inexistantes que nous ne connaissons pas – ce sont celles dont nous ignorons si nous les connaissons.»*

Plan

Les as et les huit 1 / 3

Il y a huit cartes : quatre as et quatre 8.

Les as et les huit 1 / 3

Il y a huit cartes : quatre as et quatre 8.
Chaque joueur reçoit deux cartes qu'il ne regarde pas,
mais qu'il montre à tout le monde.

Les as et les huit 1 / 3

Il y a huit cartes : quatre as et quatre 8.
Chaque joueur reçoit deux cartes qu'il ne regarde pas,
mais qu'il montre à tout le monde.
Chaque joueur parle à son tour :

- Soit il dit *Je ne sais pas*,
- Soit il dit
  - *J'ai une paire*,
  - *J'ai un as et un huit*.

Les as et les huit 2 / 3

On fait autant de tours qu'il faut.
Il y a *toujours* un joueur qui peut deviner les cartes qu'il a.

Les as et les huit 2 / 3

On fait autant de tours qu'il faut.

Il y a *toujours* un joueur qui peut deviner les cartes qu'il a.

Comment cela se peut-il ?

Les as et les huit 3 / 3

1^{re} donne 1 : $A + A$ 2 : $8 + 8$ 3 : $8 + 8$

Les as et les huit 3 / 3

1re donne	1 : $A + A$	2 : $8 + 8$	3 : $8 + 8$
2e donne	1 : $A + A$	2 : $8 + 8$	3 : $A + A$

Les as et les huit $3\,/\,3$

1re donne	1 : $A + A$	2 : $8 + 8$	3 : $8 + 8$
2e donne	1 : $A + A$	2 : $8 + 8$	3 : $A + A$
3e donne	1 : $A + A$	2 : $8 + 8$	3 : $A + 8$

Les as et les huit 3 / 3

1re donne	1 : $A + A$	2 : $8 + 8$	3 : $8 + 8$
2e donne	1 : $A + A$	2 : $8 + 8$	3 : $A + A$
3e donne	1 : $A + A$	2 : $8 + 8$	3 : $A + 8$
4e donne	1^2 : $A + 8$	2 : $8 + 8$	3 : $A + 8$

Les as et les huit 3 / 3

1re donne	1 $: A + A$	2 $: 8 + 8$	3 $: 8 + 8$
2e donne	1 $: A + A$	2 $: 8 + 8$	3 $: A + A$
3e donne	1 $: A + A$	2 $: 8 + 8$	3 $: A + 8$
4e donne	$1^2 : A + 8$	2 $: 8 + 8$	3 $: A + 8$
5e donne	1 $: A + 8$	$2^2 : A + 8$	3 $: A + 8$

Les as et les huit 3 / 3

1re donne	$1 : A+A$	$2 : 8+8$	$3 : 8+8$
2e donne	$1 : A+A$	$2 : 8+8$	$3 : A+A$
3e donne	$1 : A+A$	$2 : 8+8$	$3 : A+8$
4e donne	$1^2 : A+8$	$2 : 8+8$	$3 : A+8$
5e donne	$1 : A+8$	$2^2 : A+8$	$3 : A+8$
6e donne	$1 : A+8$	$2 : A+8$	$3^2 : A+A$

Les as et les huit 3 / 3

1^{re} donne	1 : $A+A$	2 : $8+8$	3 : $8+8$
2^{e} donne	1 : $A+A$	2 : $8+8$	3 : $A+A$
3^{e} donne	1 : $A+A$	2 : $8+8$	3 : $A+8$
4^{e} donne	1^2 : $A+8$	2 : $8+8$	3 : $A+8$
5^{e} donne	1 : $A+8$	2^2 : $A+8$	3 : $A+8$
6^{e} donne	1 : $A+8$	2 : $A+8$	3^2 : $A+A$
7^{e} donne	1 : $8+8$	2 : $8+8$	3 : $A+A$

Les as et les huit 3 / 3

1^{re} donne	$1 : A+A$	$2 : 8+8$	$3 : 8+8$
2^{e} donne	$1 : A+A$	$2 : 8+8$	$3 : A+A$
3^{e} donne	$1 : A+A$	$2 : 8+8$	$3 : A+8$
4^{e} donne	$1^2 : A+8$	$2 : 8+8$	$3 : A+8$
5^{e} donne	$1 : A+8$	$2^2 : A+8$	$3 : A+8$
6^{e} donne	$1 : A+8$	$2 : A+8$	$3^2 : A+A$
7^{e} donne	$1 : 8+8$	$2 : 8+8$	$3 : A+A$
8^{e} donne	$1 : 8+8$	$2^2 : A+8$	$3 : A+A$

Les as et les huit 3 / 3

1^{re} donne	1 : $A+A$	2 : $8+8$	3 : $8+8$
2^{e} donne	1 : $A+A$	2 : $8+8$	3 : $A+A$
3^{e} donne	1 : $A+A$	2 : $8+8$	3 : $A+8$
4^{e} donne	1^2 : $A+8$	2 : $8+8$	3 : $A+8$
5^{e} donne	1 : $A+8$	2^2 : $A+8$	3 : $A+8$
6^{e} donne	1 : $A+8$	2 : $A+8$	3^2 : $A+A$
7^{e} donne	1 : $8+8$	2 : $8+8$	3 : $A+A$
8^{e} donne	1 : $8+8$	2^2 : $A+8$	3 : $A+A$
9^{e} donne	1 : $8+8$	2 : $A+8$	3^2 : $A+8$

Les as et les huit 3 / 3

1^{re} donne	1 : $A+A$	2 : $8+8$	3 : $8+8$
2^{e} donne	1 : $A+A$	2 : $8+8$	3 : $A+A$
3^{e} donne	1 : $A+A$	2 : $8+8$	3 : $A+8$
4^{e} donne	1^2 : $A+8$	2 : $8+8$	3 : $A+8$
5^{e} donne	1 : $A+8$	2^2 : $A+8$	3 : $A+8$
6^{e} donne	1 : $A+8$	2 : $A+8$	3^2 : $A+A$
7^{e} donne	1 : $8+8$	2 : $8+8$	3 : $A+A$
8^{e} donne	1 : $8+8$	2^2 : $A+8$	3 : $A+A$
9^{e} donne	1 : $8+8$	2 : $A+8$	3^2 : $A+8$
10^{e} donne	1^2 : $A+8$	2 : $8+8$	3 : $A+A$

Plan

Plan

Les modalités

Une modalité est un opérateur qui **transforme** une sentence en une autre sentence.
On crée un modalité K_A pour chaque agent A.
Une logique avec des modalités s'appelle une **logique modale**.

Qu'est-ce que la logique de la connaissance ?

- La **logique de la connaissance** ou **logique épistémique** est la logique qui formalise
  - «L'agent i sait que p», noté $K_i(p)$,
  - «p est une connaissance commune», noté $C(p)$.

La connaissance commune

$C(p)$ formalise des phrases comme

- «C'est un fait bien connu que p, sauf des fous.»
- «L'agent i *sait que* l'agent j *sait que* l'agent i *sait que* , etc.».

La connaissance commune

$C(p)$ formalise des phrases comme

- «C'est un fait bien connu que p, sauf des fous.»
- «L'agent i *sait que* l'agent j *sait que* l'agent i *sait que* , etc.».

On a besoin d'une modalité E, dite de «connaissance partagée», «Tout le monde *sait que* p»,

$$E_G(p) = \bigwedge_{i \in G} K_i(p).$$

La connaissance commune

$C(p)$ formalise des phrases comme

- «C'est un fait bien connu que p, sauf des fous.»
- «L'agent i *sait que* l'agent j *sait que* l'agent i *sait que* , etc.».

On a besoin d'une modalité E, dite de «connaissance partagée», «Tout le monde *sait que* p»,

$$E_G(p) = \bigwedge_{i \in G} K_i(p).$$

La **connaissance commune** *n'est pas* la **connaissance partagée**.

Plan

Les règles

C'est une logique qui se présente à la Hilbert (avec un symbole qui signifie **«est un théorème»**):

$$\frac{\vdash \varphi \qquad \vdash \varphi \Rightarrow \psi}{\vdash \psi}\,(MP)$$

La règle de généralisation de la connaissance

$$\frac{\vdash \varphi}{\vdash K_i \varphi}\,(GK)$$

Les axiomes 1 / 3

Il y a *tous les théorèmes de la logique propositionnelle classique*.

$$\frac{}{\vdash \varphi}\ (\textbf{Cl}) \qquad \text{si } \varphi \text{ est un théorème de la logique classique.}$$

Les axiomes 2 / 3

Il y a quatre axiomes.

Axiome (de distribution)

$$\dfrac{}{\vdash K_i\varphi \Rightarrow K_i(\varphi \Rightarrow \psi) \Rightarrow K_i\psi}\ (\mathbf{K})$$

Axiome (de la connaissance)

$$\dfrac{}{\vdash K_i\varphi \Rightarrow \varphi}\ (\mathbf{T})$$

Les axiomes 3 / 3

Axiome (d'introspection positive)

$$\frac{}{\vdash K_i\varphi \Rightarrow K_iK_i\varphi}\,(\mathbf{4})$$

Axiome (d'introspection négative)

$$\frac{}{\vdash \neg K_i\varphi \Rightarrow K_i\neg K_i\varphi}\,(\mathbf{5})$$

Attention

En logique modale **on n'a pas** la **règle de déduction**
«De $\Gamma, \varphi \vdash \psi$ je déduis $\Gamma \vdash \varphi \Rightarrow \psi$»

Les axiomes de la connaissance commune

Défintion de E_G

$$\frac{}{\vdash E_G(\varphi) \Leftrightarrow \bigwedge_{i \in G} K_i(\varphi)}\ (C1)$$

$C_G\varphi$ satisfait l'inégalité $\psi \Rightarrow \varphi \wedge E_G(\psi)$.

$$\frac{}{\vdash C_G\varphi \Rightarrow \varphi \wedge E_G(C_G\varphi)}\ (C2)$$

Les règles de la connaissance commune

$C_G\varphi$ est le plus petit dans un certain sens, c'est-à-dire si un ψ satisfait $\psi \Rightarrow \varphi \wedge E_G(\psi)$ alors $\psi \Rightarrow C_G\varphi$.

$$\frac{\vdash \psi \Rightarrow \varphi \wedge E_G(\psi)}{\vdash \psi \Rightarrow C_G\varphi}\,(RC1)$$

Une preuve

$$\vdash K_i(\varphi \Rightarrow \psi) \Rightarrow K_i\varphi \Rightarrow K_i\psi$$

Une preuve

$$\vdash K_i(\varphi \Rightarrow \psi) \Rightarrow K_i\varphi \Rightarrow K_i\psi$$

$$\dfrac{\dfrac{}{\vdash K_i\varphi \Rightarrow K_i(\varphi \Rightarrow \psi) \Rightarrow K_i\psi}(K) \qquad \dfrac{}{\vdash (K_i\varphi \Rightarrow K_i(\varphi \Rightarrow \psi) \Rightarrow K_i\psi) \Rightarrow (K_i(\varphi \Rightarrow \psi) \Rightarrow K_i\varphi \Rightarrow K_i\psi)}(CI)}{\vdash K_i(\varphi \Rightarrow \psi) \Rightarrow K_i\varphi \Rightarrow K_i\psi}(MP$$

Plan

Les modèles de Kripke 1 / 3

Un **modèle de Kripke** est un triplet $\mathcal{M} = (\mathcal{U}_{\mathcal{M}}, I_{\mathcal{M}}, \mathcal{R}_{\mathcal{M}})$ où

- $\mathcal{U}_{\mathcal{M}}$ est un ensemble dont les éléments sont appelés, suivant les auteurs,
  - des **mondes**,
  - des **mondes possibles**,
  - des **étapes (de raisonnement)**,
  - des **états**.

Les modèles de Kripke 2 / 3

- $I_{\mathcal{M}}$: *Variables* $\rightarrow \mathcal{P}(\mathcal{U}_{\mathcal{M}})$. Intuitivement $I_{\mathcal{M}}(p)$ est l'ensemble des mondes où la variable p est satisfaite.

Les mondes sont notés u, v, w.

Les modèles de Kripke 3 / 3

- $\mathcal{R}_{\mathcal{M}} = (R_1, ..R_n)$ est un ensemble de relations dites **relations d'accessibilité**.
  Si $u\, R_i\, v$ alors le monde v est accessible à partir de u pour i [1].

Les propriétés (transitivité, réflexivité, symétrie ou antisymétrie) des relations R_i jouent un rôle.

$I_{\mathcal{M}}$ doit satisfaire des propriétés de compatibilité avec les R_i.

[1] On verra plus tard ce que ça signifie.

Un jeu très simple 1 / 3

2 agents, 3 cartes $\{A, B, C\}$.

L'agent 1 reçoit une carte

L'agent 2 reçoit un carte

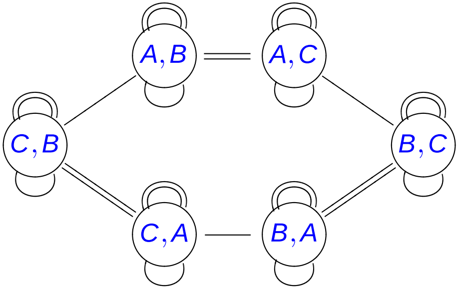
La troisième carte est retournée face contre la table

Il y a six mondes possibles :

$(A, B), (A, C), (B, A)(B, C), (C, A), (C, B)$.

Un jeu très simple 2 / 3

Dans le monde (A, B) l'agent 1 (sa relation d'accessibilité est notée par $=$) envisage deux mondes possibles à savoir (A, B) et (A, C).



Le modèle de Kripke $\mathcal{M}$.

Un jeu très simple 3 / 3

Les propositions primitives sont

- $1A$ le joueur (l'agent) 1 détient la carte A,
- $2A$ le joueur (l'agent) 2 détient la carte A,
- $1B$ le joueur (l'agent) 1 détient la carte B,
- $2B$ le joueur (l'agent) 2 détient la carte B,
- $1C$ le joueur (l'agent) 1 détient la carte C,
- $2C$ le joueur (l'agent) 2 détient la carte C.

Des assertions de forçage

$(A, B) \Vdash 1A \wedge 2B$,
$(A, B) \Vdash K_1(2B \vee 2C)$,
$(A, B) \Vdash K_1 \neg K_2(1A)$.
Pour tout monde u l'assertion $u \Vdash K_1(2A \vee 2B \vee 2C)$ est vraie
donc $\mathcal{M} \models K_1(2A \vee 2B \vee 2C)$.

Accessibilité et forçage

1. Si φ est une **variable** p:

$$\mathcal{M}, u \Vdash \varphi \qquad \text{ssi} \qquad u \in I_{\mathcal{M}}(p)$$

2. Si φ est une **conjonction** $\psi \wedge \theta$

$$\mathcal{M}, u \Vdash \varphi \qquad \text{ssi} \qquad \mathcal{M}, u \Vdash \psi \quad \text{et} \quad \mathcal{M}, u \Vdash \theta$$

3. Si φ est une **disjonction** $\psi \vee \theta$

$$\mathcal{M}, u \Vdash \varphi \qquad \text{ssi} \qquad \mathcal{M}, u \Vdash \psi \quad \text{or} \quad \mathcal{M}, u \Vdash \theta$$

4. Si φ est une **implication** $\psi \Rightarrow \theta$

$$\mathcal{M}, u \Vdash \varphi \qquad \text{ssi} \qquad \mathcal{M}, u \Vdash \psi \quad \text{implique} \quad \mathcal{M}, u \Vdash \theta$$

5. Si $\perp$ est **absurde**, alors $\mathcal{M}, u \nVdash \perp$.

Accessibilité et forçage

6. Si φ est une *modalité* $K_i(\psi)$ alors

$$u \Vdash K_i(\psi) \quad \text{ssi} \quad (\forall v \in \mathcal{U}_{\mathcal{M}})\ u\ R_i\ v \quad \text{implies} \quad v \Vdash \psi.$$

Cela signifie aussi que

l'agent i sait ψ dans le monde u

si et seulement si

dans chaque monde qu'il tient comme possible ψ est satisfaite.

Accessibilité et forçage

7. Si φ est une *modalité* $C_G(\psi)$ alors

$$u \Vdash C_G(\psi) \qquad \text{ssi} \qquad (\forall v \in \mathcal{U}_{\mathcal{M}})\ u\ \Big(\bigcup_{i \in G} R_i\Big)^*\ v \quad \text{implies} \quad v \Vdash \psi.$$

Cela signifie aussi que

$C_G(\psi)$ est satisfaite dans le monde u
si et seulement si
dans chaque monde accessible
par un chemin d'accessibilité, ψ est satisfaite.

Accessibilité et forçage

On doit avoir

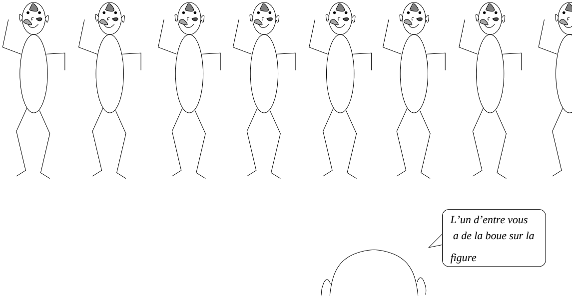
$$u \Vdash K_i\varphi \qquad \Longleftrightarrow \qquad (\forall v \in \mathcal{U}_{\mathcal{M}})\ v\ R_i\ u \Rightarrow v \Vdash \varphi.$$

Autrement dit, $u \Vdash K_i\varphi$ si et seulement si, dans tous les mondes accessibles à partir de u, on a φ.
Ou encore, l'agent i sait φ si dans tous les mondes qu'il peut envisager, φ est satisfaite.

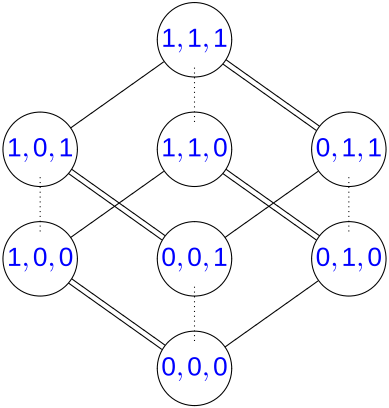
Plan

Les enfants sales

- Il y a n enfants dont certains ont la saleté sur le front.
- Le père déclare «L'un d'entre vous a de la saleté sur le front».
- Puis le père pose plusieurs fois (combien ?) la question «Avez-vous de la saleté sur le front ?».
- Comme les n enfants ont tous de la saleté sur le front.
- Après n questions du père, ils répondent tous ensemble «oui».

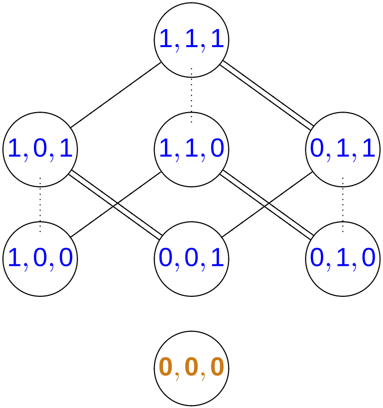

L'un d'entre vous
a de la boue sur la
figure

Le modèle de Kripke pour trois enfants sales

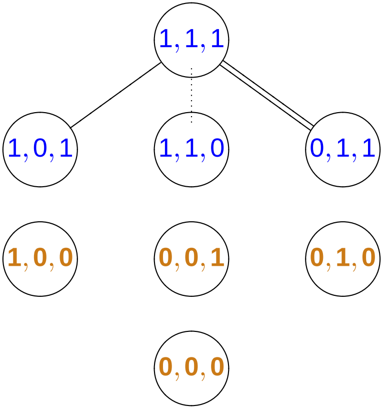


On abandonne les boucles de réflexivité.

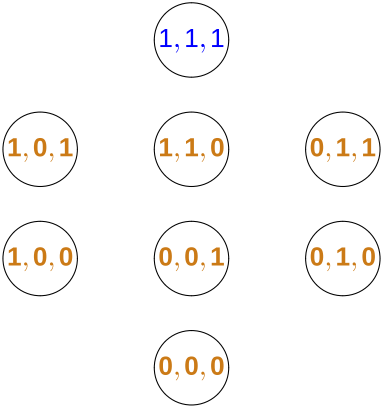
Après que le père a parlé



Après que le père a posé sa première question



Après que le père a posé sa deuxième question



Plan

Correction

Théorème

Si $\vdash \varphi$ *alors* $\models \varphi$.

Pourquoi pas la règle de déduction ?

Si on avait la règle de déduction «*De* $\Gamma, \varphi \vdash \psi$ *je déduis* $\Gamma \vdash \varphi \Rightarrow \psi$»
alors du jugement $\varphi \vdash K_i\varphi$ on aurait $\varphi \vDash K_i\varphi$,

c'est-à-dire «*Si dans tous les mondes de l'univers en question,* φ *est vrai, alors chaque agent* i *sait* φ»

on pourrait déduire $\vDash \varphi \Rightarrow K_i\varphi$

c'est-à-dire «Si φ est vrai alors chaque agent i sait φ».

Une preuve

On peut prouver $\vdash \varphi \Rightarrow K_i \neg K_i \neg \varphi$.

Une preuve

On peut prouver $\vdash \varphi \Rightarrow K_i \neg K_i \neg \varphi$.

$$\dfrac{\dfrac{}{\vdash \neg K_i\varphi \Rightarrow K_i\neg K_i\neg\varphi}\,(\mathbf{5}) \qquad \dfrac{\dfrac{}{\vdash \psi}\,(Cl) \qquad \dfrac{}{\vdash K_i\neg\varphi \Rightarrow \neg\varphi}\,(\mathbf{T})}{\vdash (\neg K_i\varphi \Rightarrow K_i\neg K_i\neg\varphi) \Rightarrow \varphi \Rightarrow K_i\neg K_i\neg\varphi}\,(MP)}{\vdash \varphi \Rightarrow K_i\neg K_i\neg\varphi}\,(MP)$$

où $\psi \equiv (K_i\neg\varphi \Rightarrow \neg\varphi) \Rightarrow (\neg K_i\varphi \Rightarrow K_i\neg K_i\neg\varphi) \Rightarrow \varphi \Rightarrow K_i\neg K_i\neg\varphi$
qui est un théorème classique.
Car c'est un instance de $(B \Rightarrow \neg A) \Rightarrow (\neg B \Rightarrow C) \Rightarrow (A \Rightarrow C)$.