Logique et Représentation des Connaissances II - Logique épistémique

N. Maudet

2022–2023

LRC Master Informatique

Introduction

On interprète désormais la modalité $\Box$ comme une modalité de connaissance (knowledge)

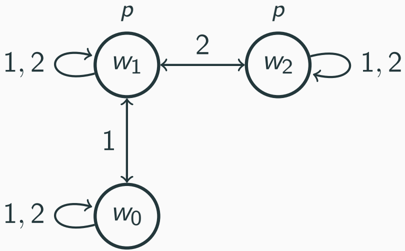
Ainsi $K_i p$ doit se lire « l'agent i sait que p »

Notez que l'on utilise désormais une modalité (indicée) par agent. On pourra donc désormais écrire des formules comm

$$K_1 K_2 p$$

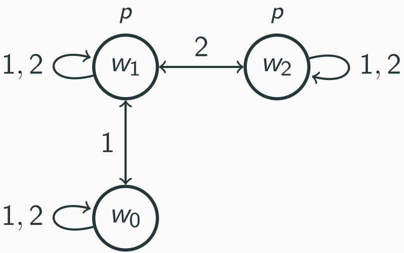
« *l'agent 1 sait que l'agent 2 sait que* p »

Exemple



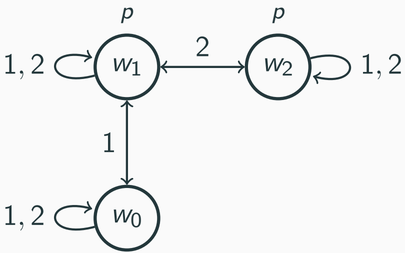
- $M, w_1 \models K_2 p$

Exemple



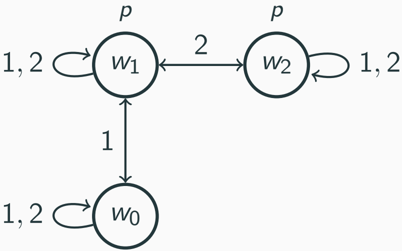
- $M, w_1 \models K_2 p$ ✔
- $M, w_1 \models K_1 p$

Exemple



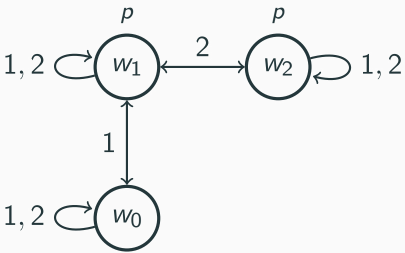
- $M, w_1 \models K_2 p$ ✔
- $M, w_1 \models K_1 p$ ✘
- $M, w_1 \models K_2 K_1 p$

Exemple



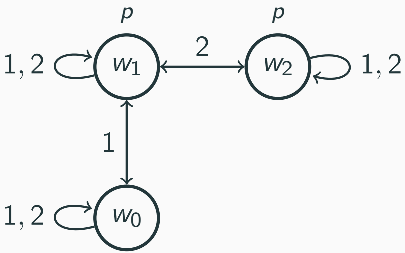
- $M, w_1 \models K_2 p$ ✔
- $M, w_1 \models K_1 p$ ✘
- $M, w_1 \models K_2 K_1 p$ ✘
- $M, w_1 \models K_2 K_1 \neg p$

Exemple



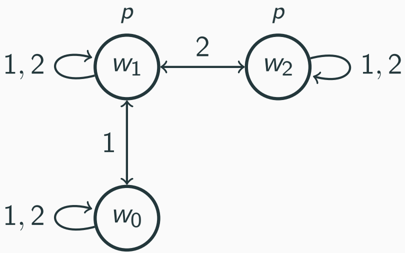
- $M, w_1 \models K_2 p$ ✔
- $M, w_1 \models K_1 p$ ✘
- $M, w_1 \models K_2 K_1 p$ ✘
- $M, w_1 \models K_2 K_1 \neg p$ ✘
- $M, w_1 \models K_2 \neg K_1 p$

Exemple



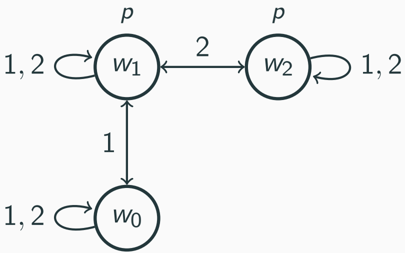
- $M, w_1 \models K_2 p$ ✔
- $M, w_1 \models K_1 p$ ✘
- $M, w_1 \models K_2 K_1 p$ ✘
- $M, w_1 \models K_2 K_1 \neg p$ ✘
- $M, w_1 \models K_2 \neg K_1 p$ ✘
- $M, w_1 \models K_1 \neg K_2 p$

Exemple



- $M, w_1 \models K_2 p$ ✔
- $M, w_1 \models K_1 p$ ✘
- $M, w_1 \models K_2 K_1 p$ ✘
- $M, w_1 \models K_2 K_1 \neg p$ ✘
- $M, w_1 \models K_2 \neg K_1 p$ ✘
- $M, w_1 \models K_1 \neg K_2 p$ ✘
- $M \models K_2 p$

Exemple



- $M, w_1 \models K_2 p$ ✔
- $M, w_1 \models K_1 p$ ✘
- $M, w_1 \models K_2 K_1 p$ ✘
- $M, w_1 \models K_2 K_1 \neg p$ ✘
- $M, w_1 \models K_2 \neg K_1 p$ ✘
- $M, w_1 \models K_1 \neg K_2 p$ ✘
- $M \models K_2 p$ ✘

Théorie de l'esprit

La capacité de réprésenter et de raisonner sur les connaissances des autres agents est une caractéristique de l'espèce humaine.
Il s'agit de la théorie de l'esprit.

Grâce à des travaux en psychologie du développement, on considère que cette capacité s'acquiert ente 3 et 5 ans.

Pour le tester, on procède par exemple à des false-belief tests (tâches de fausses croyances), comme les tests de Sally-Ann.

☞ https://www.youtube.com/watch?v=YGSj2zY2OEM

Wimmer and Perner. *Beliefs about beliefs : Representation and constraining function of wrong beliefs in young children's understanding of deception*. Cognition, 1983.

Exemple d'application : Interaction Humain-Robot

De l'utilité de la capacité de raisonner sur nos connaissances :

≪ Consider a service robot handling small logistic tasks in an office space. If the robot sees an employee searching for something the robot knows has been moved, it would be rude for the robot not to notify her. Conversely, it will be a great annoyance if the robot repeatedly informs employees about the location of things they already correctly know where are—or are not even looking for. ≫

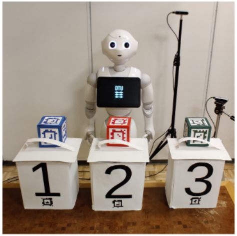


Figure 1: The Pepper robot and the two external cameras

Dissing and Bolander. *mplementing Theory of Mind on a Robot Using Dynamic Epistemic Logic*. IJCAI-2020.

Exemple d'application : Jeux coopératif

Plusieurs jeux reposent sur des principes épistémiques et requièrent ce genre de modélisation.

Le jeu Hanabi : un grand défi pour l'I.A !

Artificial Intelligence 280 (2020) 103216

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

ELSEVIER

The Hanabi challenge: A new frontier for AI research

Nolan Bard[a,*,1], Jakob N. Foerster[b,1,2], Sarath Chandar[c], Neil Burch[a], Marc Lanctot[a], H. Francis Song[d], Emilio Parisotto[e,3], Vincent Dumoulin[c], Subhodeep Moitra[c], Edward Hughes[d], Iain Dunning[d], Shibl Mourad[f], Hugo Larochelle[c], Marc G. Bellemare[c], Michael Bowling[a]

Exemple d'application : Protocoles à garantie épistémique

On dispose de 7 cartes : $\{0, 1, 2, 3, 4, 5, 6\}$.
Anne reçoit 3 cartes, Bob reçoit 3 cartes, Cathy reçoit 1 carte.

Problème : est-ce qu'il est possible pour Anne de communiquer publiquement de manière à ce que Bob connaisse le jeu de Anne, mais que Cathy ne sache rien (elle ne doit pas savoir qui détient une seule carte).

C'est le problème des Cartes Russes.

Comment le spécifier ?
Par ex. l'ignorance de Cathy, avec q_p : la carte q est détenue par p

$$\wedge_{q \in Cartes} \wedge_{p \in a,b} (\neg q_c \to \neg K_c q_p)$$

Quelle axiomatique pour la logique épistémique ?

(T) $K_i\phi \to \phi$: Axiome de la connaissance
« Si je sais qqchose, alors cette chose est vraie »

(4) $K_i\phi \to K_iK_i\phi$: Axiome d'introspection positive
« Si je sais qqchose, alors je sais que je le sais »

(5) $K\neg\phi \to K_i\neg K_i\phi$: Axiome d'introspection négative
« Si je ne sais pas qqchose, alors je sais que je ne le sais pas »

Il s'agit du système connu sous le nom de S5.

Si l'on se réfère aux propriétés sur les structures de Kripke, cela correspond à des graphes respectant la réflexivité, la symétrie, et la transitivité. (Autrement dit, des relations d'équivalence)

Comment interpréter $\Diamond$?

Si $\Box$ est la modalité de connaissance, on peut se demander comment doit s'interpréter la modalité $\Diamond$

Rappelons-nous qu'il s'agit de l'opérateur dual : $\Diamond\phi \equiv \neg\Box\neg\phi$

On peut donc le lire comment ≪ croire possible que ≫ et l'on notera $B_i p$ ≪ l'agent i croit possible que p ≫

Comment interpréter $\Diamond$?

Si $\Box$ est la modalité de connaissance, on peut se demander comment doit s'interpréter la modalité $\Diamond$

Rappelons-nous qu'il s'agit de l'opérateur dual : $\Diamond\phi \equiv \neg\Box\neg\phi$

On peut donc le lire comment « croire possible que » et l'on notera $B_i p$ « l'agent i croit possible que p »

De là on retrouve l'expression équivalente donnée au premier cours pour l'axiome (5) :

$$B\phi \to KB\phi \equiv \neg K\neg\phi \to K\neg K\neg\phi \equiv \neg K\psi \to K\neg K\psi$$

Example

Dans la logique S4 (axiomes T et 4, réflexivité + transitivité)

Montrons que :

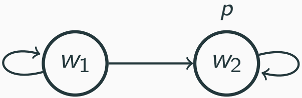
- $BK\phi \rightarrow \phi$ n'est pas valide.

Example

Dans la logique S4 (axiomes T et 4, réflexivité + transitivité)

Montrons que :

- $BK\phi \rightarrow \phi$ n'est pas valide.



Contre-exemple.

Example

Dans la logique S4 (axiomes T et 4, réflexivité + transitivité)

Montrons que :

- $BB\phi \rightarrow B\phi$ est valide.

Example

Dans la logique S4 (axiomes T et 4, réflexivité + transitivité)

Montrons que :

- $BB\phi \to B\phi$ est valide.

Par raisonnement sémantique. Supposons que la formule soit fausse. Il existe donc un modèle (réflexif et transitif) M et un monde w, tels que (i) $M, w \models BB\phi$ est vraie, tandis que (ii) $M, w \models B\phi$ est fausse. (i) signifie que il existe un monde accessible w' depuis lequel il existe un monde accessible w'' pour lequel on a $M, w'' \models \phi$. Par transitivité, w'' est accessible depuis w, contradiction avec (ii).

Example

Même question dans S5 ? $BK\phi \to \phi$ est-elle valide ?

Example

Même question dans S5 ? $BK\phi \to \phi$ est-elle valide ?
On cherche à montrer $\neg\phi \to \neg BK\phi$

1. $\neg\phi \to \neg K\phi$ (T)

Example

Même question dans S5 ? $BK\phi \to \phi$ est-elle valide ?
On cherche à montrer $\neg\phi \to \neg BK\phi$

1. $\neg\phi \to \neg K\phi$ (T)
2. $\neg K\phi \to K\neg K\phi$ (5)

Example

Même question dans S5 ? $BK\phi \to \phi$ est-elle valide ?
On cherche à montrer $\neg\phi \to \neg BK\phi$

1. $\neg\phi \to \neg K\phi$ (T)
2. $\neg K\phi \to K\neg K\phi$ (5)
3. $\neg\phi \to K\neg K\phi$ (MP, Prop.)

Example

Même question dans S5 ? $BK\phi \to \phi$ est-elle valide ?
On cherche à montrer $\neg\phi \to \neg BK\phi$

1. $\neg\phi \to \neg K\phi$ (T)
2. $\neg K\phi \to K\neg K\phi$ (5)
3. $\neg\phi \to K\neg K\phi$ (MP, Prop.)
4. $\neg\phi \to \neg\neg K\neg K\phi$ ($\equiv$)

Example

Même question dans S5 ? $BK\phi \to \phi$ est-elle valide ?
On cherche à montrer $\neg\phi \to \neg BK\phi$

1. $\neg\phi \to \neg K\phi$ (T)
2. $\neg K\phi \to K\neg K\phi$ (5)
3. $\neg\phi \to K\neg K\phi$ (MP, Prop.)
4. $\neg\phi \to \neg\neg K\neg K\phi$ ($\equiv$)
5. $\neg\phi \to \neg BK\phi$ ($\equiv$)

On peut aussi le montrer par raisonnement sémantique.

Savoir si vs. savoir que

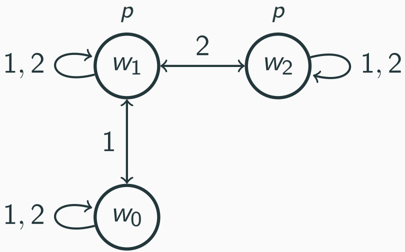
Notons la différence entre « savoir que ϕ" et savoir si ϕ »

Savoir si ϕ, c'est savoir que ϕ ou savoir que $\neg\phi$.

$$K_i^{si} p \equiv K_i p \vee K_i \neg p$$

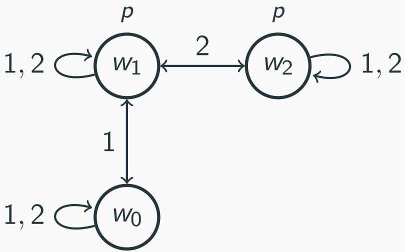
Attention : c'est bien différent de savoir que ϕ ou $\neg\phi$... qui est une tautologie (formule toujours vraie).

Exemple



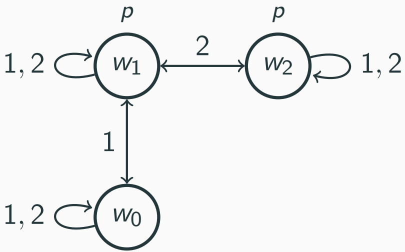
- $M, w_1 \models K_1 K_2^{si} p$

Exemple



- $M, w_1 \models K_1 K_2^{si} p$ ✔
- $M, w_1 \models K_2 K_1^{si} p$

Exemple



- $M, w_1 \models K_1 K_2^{si} p$ ✔
- $M, w_1 \models K_2 K_1^{si} p$ ✘

Exemple : Jeu de carte [Halpern et al.]

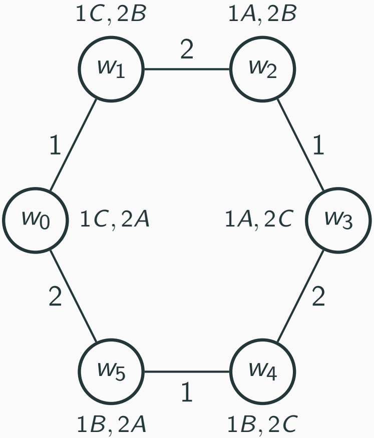
2 agents : $\{1, 2\}$, 3 cartes $\{A, B, C\}$

On distribue une carte à chaque agent, et la troisième carte reste cachée.

- un monde sera noté (x, y), où x est la carte donnée à 1, et y celle donnée y
- on utlisera les propositions $1A$, etc. pour indiquer que l'agent 1 reçoit la carte A.

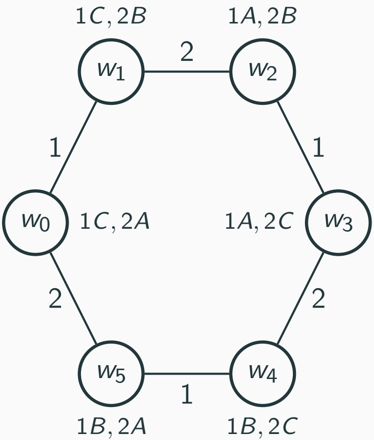
(On suppose que la structure est implicitement réflexive et symétrique pour simplifier la représentation graphique.)

Exemple : Jeu de carte [Halpern et al.]



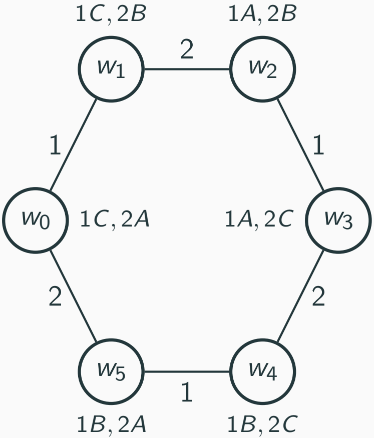
- $M, w_2 \models 1A \wedge 2B$

Exemple : Jeu de carte [Halpern et al.]



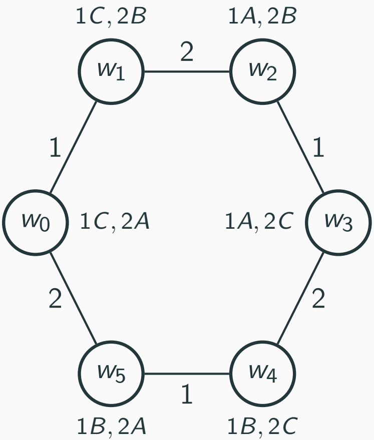
- $M, w_2 \models 1A \wedge 2B$ ✔
- $M, w_2 \models K_1(2B \vee 2C)$

Exemple : Jeu de carte [Halpern et al.]



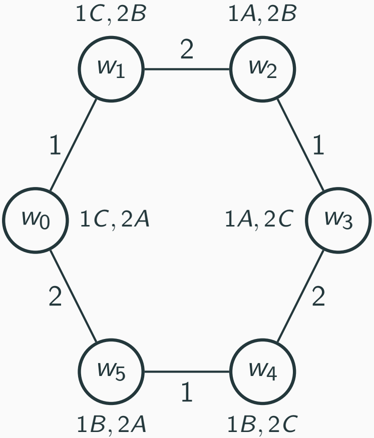
- $M, w_2 \models 1A \wedge 2B$ ✔
- $M, w_2 \models K_1(2B \vee 2C)$ ✔
- $M, w_3 \models K_1(2B \vee 2C)$

Exemple : Jeu de carte [Halpern et al.]



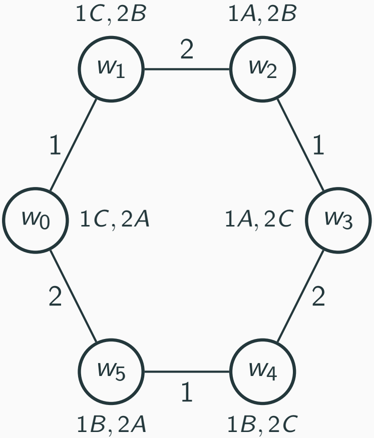
- $M, w_2 \models 1A \wedge 2B$ ✔
- $M, w_2 \models K_1(2B \vee 2C)$ ✔
- $M, w_3 \models K_1(2B \vee 2C)$ ✔
- $M \models \neg K_2 1A$

Exemple : Jeu de carte [Halpern et al.]



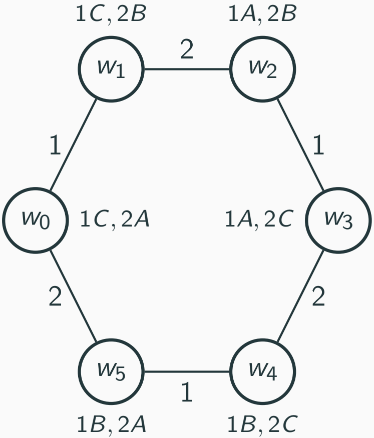
- $M, w_2 \models 1A \wedge 2B$ ✔
- $M, w_2 \models K_1(2B \vee 2C)$ ✔
- $M, w_3 \models K_1(2B \vee 2C)$ ✔
- $M \models \neg K_2 1A$ ✔
- $M \models B_2 1A$

Exemple : Jeu de carte [Halpern et al.]



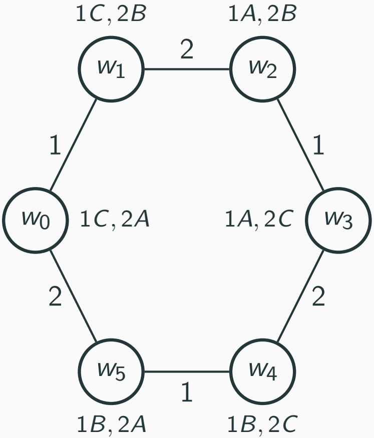
- $M, w_2 \models 1A \land 2B$ ✔
- $M, w_2 \models K_1(2B \lor 2C)$ ✔
- $M, w_3 \models K_1(2B \lor 2C)$ ✔
- $M \models \neg K_2 1A$ ✔
- $M \models B_2 1A$ ✘
- $M \models K_2 2A \lor B_2 1A$

Exemple : Jeu de carte [Halpern et al.]



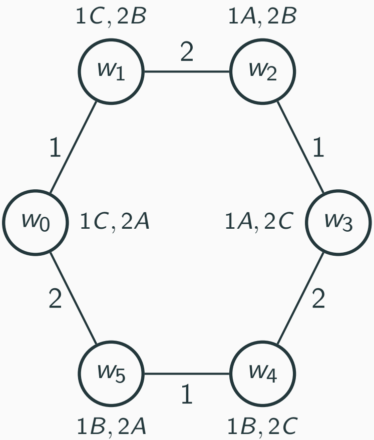
- $M, w_2 \models 1A \wedge 2B$ ✔
- $M, w_2 \models K_1(2B \vee 2C)$ ✔
- $M, w_3 \models K_1(2B \vee 2C)$ ✔
- $M \models \neg K_2 1A$ ✔
- $M \models B_2 1A$ ✘
- $M \models K_2 2A \vee B_2 1A$ ✔
- $M \models 1A \rightarrow K_1(2B \vee 2C)$

Exemple : Jeu de carte [Halpern et al.]



- $M, w_2 \models 1A \wedge 2B$ ✔
- $M, w_2 \models K_1(2B \vee 2C)$ ✔
- $M, w_3 \models K_1(2B \vee 2C)$ ✔
- $M \models \neg K_2 1A$ ✔
- $M \models B_2 1A$ ✘
- $M \models K_2 2A \vee B_2 1A$ ✔
- $M \models 1A \rightarrow K_1(2B \vee 2C)$ ✔
- $M \models 1A \rightarrow K_1 \neg K_2 1A$

Exemple : Jeu de carte [Halpern et al.]



- $M, w_2 \models 1A \wedge 2B$ ✔
- $M, w_2 \models K_1(2B \vee 2C)$ ✔
- $M, w_3 \models K_1(2B \vee 2C)$ ✔
- $M \models \neg K_2 1A$ ✔
- $M \models B_2 1A$ ✘
- $M \models K_2 2A \vee B_2 1A$ ✔
- $M \models 1A \rightarrow K_1(2B \vee 2C)$ ✔
- $M \models 1A \rightarrow K_1 \neg K_2 1A$ ✔

Le problème de l'omniscience logique

Remarquons que les systèmes axiomatiques tels que nous les avons définis (en fait tout système normal), modélisent des agents dôtés de capacité de raisonnement très forte, qui connaissent toutes les conséquences logiques de leurs connaissances.

En effet, avec la nécessitation N et l'axiome K de distribution, on observe que : à partir de $K_i\phi$ et de $\phi \to \psi$ on peut déduire que $K_i\psi$.

En particulier, un tel agent connait toutes les tautologies. Il s'agit donc ici d'une version idéalisée d'agent.

Complexité du raisonnement (1)

Quelle est la complexité algorithmique des tâches de raisonnement : satisfiabilité et model checking ?

- SAT : Etant donnée une formule ϕ, existe-t-il un modèle M et un monde w tel que $M, w \models \phi$?
- MCHECK : Etant donnée une formule ϕ et un modèle pointé (M, w), est-il vrai que $M, w \models \phi$?

Complexité du raisonnement (2)

La difficulté des problèmes va dépendre de la logique considérée et du nombre d'agents dans le modèle.

- SAT pour S5 avec 1 agent est NP-complet.
- SAT pour S4 avec 1 agent est PSPACE-complet
- SAT pour S5 avec ≥ 2 agents est PSPACE-complet
- MCHECK pour S4 et S5 est dans P

Rappel : P $\subseteq$ NP $\subseteq$ PSPACE

Attention : La complexité algorithmique est évaluée en fonction de la taille de l'entrée du problème... et celle-ci peut être grande !
On considère que la taille d'un modèle M est la somme du nombre de mondes et du nombre de relations d'accessibilité entre les mondes pour chaque agent.

Equivalence entre structures de Kripke

Cela amène à une question naturelle : est-il possible qu'il existe plusieurs modèles de Kripke différents, mais "équivalents" pour les formules de logique modale considérées ?

Cette idée est formalisée par la notion de bi-simulation.

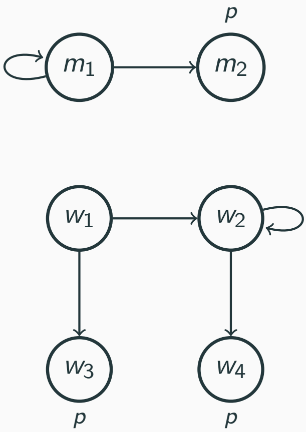
Bi-simulation

Soient deux modèles de Kripke M et M'. Une bi-simulation entre M et M' est une relation Z entre les mondes de M et M' telle que pour tout $(w, w') \in Z$:

- $I^{-1}(w) = I^{-1}(w')$ (les mondes ont la même valuation)
- s'il existe un monde t accessible pour s dans M, alors il existe un monde t' accessible pour s' dans M' tel que tZt'
- même condition dans l'autre sens

On note $(M, s) \leftrightarroweq (M', s')$ la relation de bi-similarité

Bi-simulation



- $M, m_1 \leftrightarroweq M', w_1$
- $M, m_1 \leftrightarroweq M', w_2$
- $M, m_2 \leftrightarroweq M', w_3$
- $M, m_2 \leftrightarroweq M', w_4$

Bi-simulation

Les modèles bi-similaires satisfont les mêmes formules de logique modale (pour les logiques étudiées pour le moment)

Si $(M, w) \leftrightarroweq (M', s')$ alors $(M, w) \models \phi \Leftrightarrow (M', w') \models \phi$