



**HAL**  
open science

## Intention et logique épistémique dynamique

Alix Goudyme, Nathalie Chetcuti-Sperandio, Sylvain Lagrue, Tiago de Lima

► **To cite this version:**

Alix Goudyme, Nathalie Chetcuti-Sperandio, Sylvain Lagrue, Tiago de Lima. Intention et logique épistémique dynamique. Rencontres des Jeunes Chercheurs en Intelligence Artificielle (RJCIA 2019), Jul 2019, Toulouse, France. pp.1-9. hal-02161427

**HAL Id: hal-02161427**

**<https://hal.science/hal-02161427v1>**

Submitted on 20 Jun 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Intention et logique épistémique dynamique

A. Goudyme<sup>1</sup>   N. Chetcuti-Sperandio<sup>1</sup>   S. Lagrue<sup>2</sup>   T. de Lima<sup>1</sup>

<sup>1</sup>CRIL UMR CNRS 8188/Université d'Artois

<sup>2</sup>HEUDIASYC UMR CNRS 7253/Université de Technologie de Compiègne  
{goudyme,chetcuti,delima}@cril.fr   sylvain.lagrue@hds.utc.fr

## Résumé

*Modéliser l'intention permet d'expliquer les décisions prises par des agents. Dans ce travail, nous proposons une modélisation de l'intention dans des jeux épistémiques représentés en logique épistémique dynamique. Étant données une propriété et une suite d'actions déjà réalisée par un joueur dans un tel jeu, nous proposons une méthode capable de déterminer si le joueur a eu l'intention d'obtenir cette propriété. Un exemple de l'utilisation de cette méthode est donné en utilisant une version simplifiée du jeu collaboratif Hanabi.*

## Mots Clef

Logique épistémique dynamique, intention, jeux épistémiques.

## Abstract

*Modeling intention is essential to explain decisions made by agents. In this work, we propose a model of intention in epistemic games, represented in dynamic epistemic logic. Given a property and a sequence of actions already performed by a player in such a game, we propose a method able to determine whether the player had the intention to obtain the property. An example of the method is given using a simplified version of the collaborative game Hanabi.*

## Keywords

Dynamic epistemic logic, intention, epistemic games.

## 1 Introduction

Pouvoir déterminer le but poursuivi par un agent ou par un groupe d'agents à partir de ses connaissances aurait un intérêt certain dans des domaines tels que l'économie, les jeux et, bien sûr, l'intelligence artificielle. Notre objectif dans cet article est d'être capables de découvrir l'intention des agents à partir, d'une part, des connaissances de ceux-ci sur les actions à leur disposition et, d'autre part, des actions qu'ils ont finalement réalisées. Nous avons choisi de nous restreindre aux jeux épistémiques. L'univers des jeux permet de travailler dans un cadre défini et d'avoir un contrôle total sur les connaissances des joueurs et des actions

qu'ils peuvent effectuer. Les jeux épistémiques désignent des jeux à information incomplète où la réussite au jeu dépend essentiellement des connaissances des joueurs sur l'état du jeu et sur les connaissances des autres joueurs. Des exemples de ce type de jeux sont le Cluedo ou encore Hanabi.

La logique épistémique dynamique est une logique qui traite de la connaissance d'agents et de son évolution à la suite d'événements [1, 2, 4]. Elle apparaît ainsi comme un formalisme adapté à la modélisation logique de l'intention dans les jeux épistémiques, en l'occurrence dans Hanabi, jeu sur lequel s'est porté notre choix pour évaluer nos travaux.

Dans la suite de cet article, nous présenterons tout d'abord la logique épistémique dynamique, puis le jeu Hanabi, avant de détailler nos travaux de modélisation de l'intention dans ce jeu.

## 2 Logique épistémique dynamique

### 2.1 Logique modale

La logique modale [3] s'appuie sur les connecteurs de la logique propositionnelle ( $\top, \perp, \neg, \wedge, \vee, \rightarrow, \leftrightarrow$ ) auxquels sont ajoutés deux nouveaux connecteurs représentant la *possibilité*  $\diamond$  et la *nécessité*  $\square$ . Par exemple, la formule  $\diamond\neg(\phi \leftrightarrow \psi)$  signifie « il est possible que  $\phi$  et  $\psi$  ne soient pas équivalents ». L'opérateur  $\diamond$  peut s'écrire  $\neg\square\neg$ .

Une des sémantiques les plus utilisées pour la logique modale est celle proposée par Kripke [8]. Un modèle de Kripke se base sur la notion de *mondes possibles*. On note chaque monde  $M_n$  (où  $n$  est un entier),  $M$  représente l'ensemble des mondes  $M_n$ . Un *cadre de Kripke* est défini comme l'ensemble formé de  $M$  et de  $R$ , où  $R$  est une relation binaire sur  $M$ . Cette relation est appelée *relation d'accessibilité* et définit les mondes accessibles depuis chacun des mondes. Par exemple, dans le système que nous allons utiliser, la relation lie les mondes envisageables pour un joueur. Un cadre de Kripke est souvent représenté sous forme de graphe orienté où les sommets sont les mondes et où les arêtes relient les mondes accessibles.

Enfin, soit  $\mathbb{P}$  l'ensemble de propositions, un *modèle de Kripke* est un triplet formé de  $M$ , de  $R$  et de  $h$ , où  $h$  est la fonction de valuation des propositions. La fonction de valuation associe à chaque proposition de  $\mathbb{P}$  un sous-ensemble de mondes. Un monde peut alors être vu comme un modèle au sens de la logique propositionnelle. Le triplet  $U = (M, R, h)$  est aussi appelé *Univers*.

La sémantique de la logique modale est régie par les dix règles suivantes :

1.  $\models_{M_j} \top$
2.  $\not\models_{M_j} \perp$
3.  $\models_{M_j} p$  ssi  $M_j \in h(p)$
4.  $\models_{M_j} \neg\psi$  ssi  $\not\models_{M_j} \psi$
5.  $\models_{M_j} \phi \wedge \psi$  ssi  $\models_{M_j} \phi$  et  $\models_{M_j} \psi$
6.  $\models_{M_j} \psi \vee \phi$  ssi  $\models_{M_j} \psi$  ou  $\models_{M_j} \phi$  ou les deux
7.  $\models_{M_j} \psi \rightarrow \phi$  ssi  $\models_{M_j} \psi$  implique  $\models_{M_j} \phi$
8.  $\models_{M_j} \psi \leftrightarrow \phi$  ssi,  $\models_{M_j} \psi$  ssi  $\models_{M_j} \phi$
9.  $\models_{M_j} \Box\psi$  ssi pour tous les mondes  $M_i$  de  $U$  tels que  $M_j R M_i$   $\models_{M_i} \psi$
10.  $\models_{M_j} \Diamond\psi$  ssi pour au moins un des mondes  $M_i$  de  $U$  tels que  $M_j R M_i$   $\models_{M_i} \psi$

La règle 3 stipule qu'une variable propositionnelle  $p$  est vraie dans un monde si et seulement si ce monde appartient à  $h(p)$ . Les règles 9 et 10 concernent les modalités :  $\Box\phi$  est vraie dans un monde si et seulement si tous les mondes accessibles vérifient  $\phi$  et  $\Diamond\phi$  est vraie si au moins un monde accessible vérifie  $\phi$ .

Une formule est *valide* dans un univers  $U$  si et seulement si elle est vraie dans tous les mondes de  $U$ . Une formule est valide si et seulement si elle est valide dans tous les univers.

Du fait de la sémantique qui vient d'être définie, cette logique modale est dite *normale* puisqu'elle admet les axiomes et règles d'inférence suivants :

1. les tautologies de la logique propositionnelle
2. Kripke (K) :  $\Box(\psi \rightarrow \phi) \rightarrow (\Box\psi \rightarrow \Box\phi)$
3. axiome Df :  $\neg\Diamond\neg\psi \leftrightarrow \Box\psi$
4. Modus Ponens (MP) :  $\frac{\psi \rightarrow \phi, \psi}{\phi}$
5. Nécessitation (RN) :  $\frac{\psi}{\Box\psi}$

La règle RN signifie que si  $\phi$  est valide dans un univers, alors  $\Box\phi$  est aussi valide dans l'univers (et par extension si  $\phi$  est valide alors  $\Box\phi$  est valide).

## 2.2 Logique épistémique

La logique épistémique est un type de logique modale qui traite de la notion de connaissance et de croyance d'agents [6]. L'agent croit certaines choses et n'a pas d'information sur d'autres. Il imagine alors plusieurs mondes possibles. Appelons  $N$  l'ensemble des agents.

Pour chaque agent  $\alpha$ , nous introduisons un opérateur (ou connecteur) de croyance/connaissance que l'on peut noter  $B_\alpha$ . Pour se situer par rapport à la sous-section précédente, l'opérateur de croyance correspond à l'opérateur  $\Box$ . En effet, ce qu'un agent croit doit être vrai dans tous les mondes qu'il imagine. L'opérateur  $\Diamond$  est souvent représenté  $\hat{B}_\alpha$ . Par ailleurs, chaque opérateur se voit attribuer une relation binaire  $R_\alpha$ . La relation  $R_\alpha$  lie alors des mondes *indistinguishables* pour l'agent  $\alpha$ , c'est-à-dire des mondes dont les propositions communes sont crues par l'agent  $\alpha$ . La sémantique de la logique modale ne change pas pour la logique épistémique sauf pour les règles 9 et 10 qui deviennent simplement pour chaque agent  $\alpha$  :

- $\models_{M_j} B_\alpha\psi$  ssi pour tous les mondes  $M_i$  de  $U$  tels que  $M_j R_\alpha M_i$   $\models_{M_i} \psi$
- $\models_{M_j} \hat{B}_\alpha\psi$  ssi pour au moins un monde  $M_i$  de  $U$  tel que  $M_j R_\alpha M_i$   $\models_{M_i} \psi$

Enfin, les agents peuvent raisonner sur les connaissances des autres agents. En d'autres termes, les agents ont des croyances sur les croyances des autres agents. Et pour aller plus loin, ils peuvent imaginer des mondes qu'ils savent faux mais potentiellement vrais pour d'autres agents. Par exemple, la formule  $B_1 B_2 p \wedge \neg B_2 B_1 B_2 p$  signifie que l'agent 1 croit que l'agent 2 croit  $p$  mais que l'agent 2 ne croit pas que l'agent 1 croit que l'agent 2 croit  $p$ .

## 2.3 Système S5

La logique épistémique présente deux familles d'axiomes fréquemment utilisées, le système KD45 et le système S5. Le système KD45 traite de croyances, c'est-à-dire, de connaissances qui pourraient être fausses. Le système S5 traite de connaissances vraies et c'est celui que nous allons utiliser.

En effet, dans les jeux, si l'on respecte les règles, les joueurs ne peuvent croire que des vérités, même pour les jeux où le mensonge est autorisé. Dans ce dernier cas, les joueurs, sachant que les autres peuvent mentir, ne prennent pas pour vérité les affirmations des autres joueurs. Le seul moyen pour qu'une connaissance soit fautive est que les règles n'aient pas été respectées.

Le système S5 est un système normal augmenté de deux axiomes :

- axiome 5 :  $\neg B_\alpha\psi \rightarrow B_\alpha(\neg B_\alpha\psi)$
- axiome T :  $B_\alpha\psi \rightarrow \psi$

L'axiome 5 stipule que l'agent sait ce qu'il ne sait pas. L'axiome T implique que si un agent sait quelque chose, cette chose est forcément vraie. L'ensemble des axiomes du système S5 implique alors que les mondes reliés entre eux par une même relation forment des cliques, que les mondes se relient eux-mêmes et que toutes les relations sont dans les deux sens.

## 2.4 Actions

On peut rendre la logique épistémique dynamique en y ajoutant des actions. Celles-ci permettent de décrire des changements d'état du monde ou bien l'apprentissage de nouvelles connaissances, pour plus de détails, se référer à [11, 7, 1, 2, 4]. Par exemple, les actions permettent de représenter la pioche d'une carte ou bien l'ouverture d'une porte. L'action transforme un univers en un autre univers. Elle se compose d'une pré-condition et d'une post-condition et fonctionne de la manière suivante : chaque monde du modèle vérifiant la pré-condition est modifié par la post-condition, les autres mondes sont supprimés et les liens entre les mondes restants sont gardés.

**Définition 1** Une action est un couple formé d'une pré-condition et d'une post-condition, noté  $a = (pre(a), post(a))$ . La pré-condition est définie comme une formule du langage. La post-condition est une fonction qui, à chaque proposition  $p \in \mathbb{P}$ , associe un élément de  $\{\top, \perp, p, \neg p\}$ .

En particulier, l'action "nop" est définie comme le couple  $nop = (pre(nop), post(nop))$ , où  $pre(nop) = \top$  et  $post(nop)(p) = p$  pour tout  $p \in \mathbb{P}$ . Cette action représente l'action de « ne rien faire », qui n'a pas de pré-condition (i.e., peut être exécutée à tout moment) et ne change pas l'état du monde (i.e., la valeur de chaque proposition  $p$  reste la même).

On note  $A$  l'ensemble des actions possibles et  $U = (M, R, h)$  le modèle épistémique courant. L'exécution de l'action  $a \in A$ , dans  $U$  engendre le modèle  $U|_a = (M|_a, R|_a, h|_a)$  où :

- $M|_a = \{M_i \in M \text{ tel que } \models_{M_i}^U pre(a)\}$  est la restriction à l'ensemble des mondes vérifiant la pré-condition de  $a$ .
- $R|_a = R \cap (M|_a \times M|_a)$  est la restriction des relations aux mondes de  $M|_a$ .
- $h|_a(p) = \{M_i \in M \text{ tel que } \models_{M_i}^U post(a)(p)\} \cap M|_a$  est la restriction de la valuation aux mondes de  $M|_a$  et réaffectation de la valeur des variables propositionnelles.

**Exemple.** Soit l'exemple présenté dans la figure 1, l'état épistémique initial  $U$  est présenté à gauche. Il y a trois mondes, et une variable propositionnelle  $p$ . Le monde  $M_0$  présente un double cerclage : il s'agit du monde réel. Le monde réel est celui qui représente la réalité (c'est-à-dire ce qui est vrai à l'instant actuel). Les mondes  $M_0$  et  $M_2$  appartiennent à  $h(p)$ , le monde  $M_1$  n'appartient pas à  $h(p)$ . Soit l'action  $a$  définie par la pré-condition  $pre(a) = p$  et la post-condition  $post(a)(p) = \perp$ , l'exécution de l'action  $a$  sur l'univers  $U$  est présentée à droite : les mondes  $M_0$  et  $M_2$  vérifient la pré-condition, ils sont donc conservés et la variable propositionnelle acquiert une nouvelle valeur, celle de la post-condition.

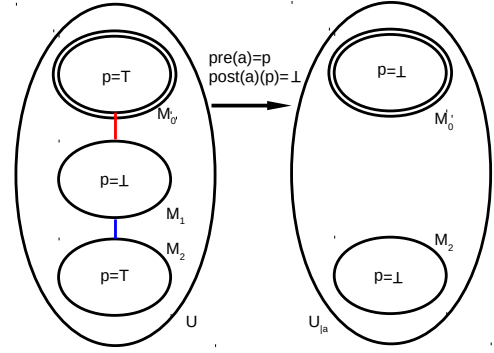


FIGURE 1 – Exemple d'action

## 3 Hanabi

Hanabi est un jeu de cartes coopératif au tour par tour de 2 à 5 joueurs dont le but est de gagner le plus de points (voir figure 2). Les cartes sont de 5 couleurs différentes (rouge, jaune, bleu, vert, blanc) et de 5 valeurs différentes (de 1 à 5). Pour chaque couleur, il y a trois cartes 1, deux 2, deux 3, deux 4 et une 5. Pour gagner des points, il faut reconstituer les 5 piles de cartes de couleur identique, dans l'ordre croissant des valeurs. On marque 1 point par carte posée, soit un maximum de 25 points. Les joueurs ont en main 4 ou 5 cartes (suivant le nombre de joueurs) ; la particularité de Hanabi est que les joueurs ne voient pas leurs propres cartes mais peuvent voir celles des autres. Pendant son tour, un joueur peut effectuer l'une des trois actions suivantes : donner un indice, défausser une carte, poser une carte. Il existe une pioche, une table, une défausse, ainsi que 8 jetons indices et 3 jetons de vie. Chaque joueur ne pouvant utiliser que ses propres cartes, la difficulté réside donc dans la capacité à donner de bons indices et à poser les bonnes cartes.



FIGURE 2 – Présentation jeu Hanabi

**Donner un indice.** Un joueur ne peut donner un indice qu'à un seul joueur. Cet indice porte sur l'en-

semble de ses cartes, doit être complet et ne peut concerner qu'une seule des 10 caractéristiques que peuvent avoir les cartes (couleur ou valeur). Par exemple, si le joueur 1 dit au joueur 3 "tes 1ère et 4ème cartes sont bleues" alors les autres cartes ne le sont pas. Enfin, lorsqu'un joueur donne un indice, un jeton indice est consommé. Cette action n'est donc possible que s'il reste des jetons indices.

**Défausser une carte.** Le joueur prend une carte de sa main et la pose sur la table, face visible. Cette action permet de piocher une nouvelle carte, s'il en reste, et de regagner un jeton indice.

**Poser une carte.** Lorsque le joueur choisit de poser l'une de ses cartes, il existe deux cas de figure, soit la carte permet d'alimenter l'une des piles, soit elle ne le permet pas et finit dans la défausse, face visible, entraînant la perte d'un jeton de vie. Une carte peut être posée sur la pile de sa couleur si la carte au sommet de cette pile est la carte précédente (de valeur immédiatement inférieure). Une carte de valeur 1 peut être posée si elle permet de débiter une pile d'une nouvelle couleur. Enfin, poser une carte de valeur 5 permet de compléter une pile et de gagner un jeton de vie.

**Fin de la partie.** La partie se termine s'il ne reste plus de jeton de vie, si le nombre maximal de points est atteint ou après un tour complet suivant la pioche de la dernière carte de la pioche.

## 4 Intention

### 4.1 Principes

L'objectif de notre travail est de déterminer *a posteriori* les intentions qu'un joueur a eues au cours d'une partie. En d'autres termes, nous souhaitons expliquer les actions effectuées par un joueur lors d'une partie. Notre idée se base sur le principe suivant : *par rapport à tous les résultats potentiels imaginés par le joueur, a-t-il effectué l'action qui le menait vers le meilleur résultat espéré ?* Dans la suite, nous allons expliquer cette idée plus en détail.

Soit un joueur et une propriété  $p$ , exprimée comme une formule en logique propositionnelle.

**Premier principe.** Nous associons à un univers  $U$  une valeur  $v(U, p)$  appelée fréquence de la propriété  $p$  qui, comme son nom l'indique, est égale à la fréquence des mondes vérifiant  $p$ , indistinguables du monde réel pour le joueur (voir figure 3). Cette idée a été initialement proposée par Markus Eger dans sa thèse [5].

**Deuxième principe.** Supposons que le joueur n'imagine qu'un monde possible, le monde réel, et qu'il a deux actions possibles  $a$  et  $b$  (voir figure 4). L'action  $a$  mène vers un univers de valeur 0 et l'action  $b$  vers un univers de valeur 1. La meilleure action pour la propriété  $p$  est donc l'action  $b$ . Si le joueur a effectué l'action  $b$ , alors il avait l'intention d'obtenir la propriété  $p$ ,

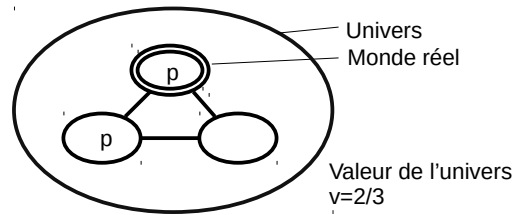


FIGURE 3 – Illustration du premier principe

dans le cas contraire, il n'en avait pas l'intention. En outre, on affecte à l'univers initial la valeur de l'univers obtenu par la meilleure action.

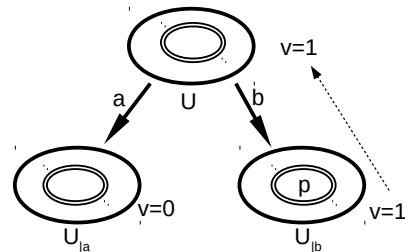


FIGURE 4 – Illustration du deuxième principe simple

Supposons maintenant que le joueur imagine deux mondes possibles, avec toujours deux actions  $a$  et  $b$  (voir figure 5). Dans le premier monde, l'action  $a$  mène vers un univers de valeur 0,5 et l'action  $b$  vers un univers de valeur 1. Dans le second monde, l'action  $a$  mène vers un univers de valeur 0,5 et l'action  $b$  vers un univers de valeur 0. Les mondes étant équivalents pour le joueur, il y a autant de chance pour lui de se trouver dans l'un que dans l'autre. On associe donc aux actions la moyenne des valeurs des univers vers lesquels elles mènent. La valeur de  $a$  est donc de 0,5 et la valeur de  $b$  est de 0,5 également. Comme auparavant, nous affectons à l'univers initial la valeur de la meilleure action (ou des meilleures actions comme ici). Dans le cas présent, peu importe quelle action a été exécutée, l'obtention de la propriété  $p$  ne peut pas être considérée comme intentionnelle puisque toutes les actions permettent de l'obtenir avec autant de chance.

En revanche, l'intention d'obtenir une propriété est claire dans le cas où il n'y a qu'une seule meilleure action pour cela. Pour d'autres pourcentages de meilleures actions parmi l'ensemble des actions, il est

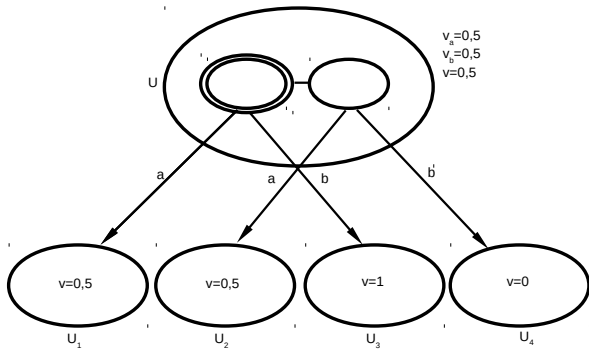


FIGURE 5 – Illustration du deuxième principe

plus difficile de mesurer l'intention. Ainsi s'il semble toujours pertinent de parler d'intention lorsqu'il y a 2 meilleures actions parmi 10 actions pour obtenir une propriété, pour 5 meilleures actions parmi 10 on pourrait parler d'aléatoire. Et à 9 actions parmi 10, il est difficile de parler d'intention d'obtenir la propriété. Cependant, dans ce dernier cas, nous avons alors une meilleure action parmi 10 pour obtenir la non-propriété. Établir des seuils d'intention fait partie de nos réflexions actuelles.

**Troisième principe.** Dans un troisième exemple, le joueur imagine de nouveau deux mondes, mais cette fois-ci, c'est à un autre joueur de jouer et celui-ci a trois actions possibles :  $a$ ,  $b$  et  $c$  (voir figure 6). Le second joueur ayant ses propres objectifs, les différentes actions sont plus ou moins intéressantes au regard de ses objectifs personnels et on appelle *utilité* la mesure de cet intérêt. Cependant ici, ce qui sera pris en compte n'est pas l'utilité définie par le second joueur mais ce que le premier joueur imagine être l'utilité pour le second joueur.

Supposons d'abord que le premier monde est le monde réel, le premier joueur imagine l'utilité des actions, par exemple  $u(a) = 0,6$ ,  $u(b) = 0,6$  et  $u(c) = 0,3$ . Le premier joueur suppose que le second joueur est rationnel et qu'il va donc effectuer une action d'utilité maximale. Les deux actions  $a$  et  $b$  ayant une utilité maximale, le premier joueur pense que le second joueur effectuera l'une de ces deux actions avec une équiprobabilité. Ces deux actions mènent vers deux univers différents dont les valeurs sont de 0,8 et 0,6. La moyenne de la valeur de ces deux actions, qui est 0,7, est affectée au premier monde. Nous effectuons à présent le même raisonnement avec le second monde. Les actions ont des utilités différentes, les univers ainsi générés ont des valeurs différentes et, finalement, nous trouvons, pour ce second monde, une valeur de 0,3. Pour terminer, nous faisons la moyenne des valeurs

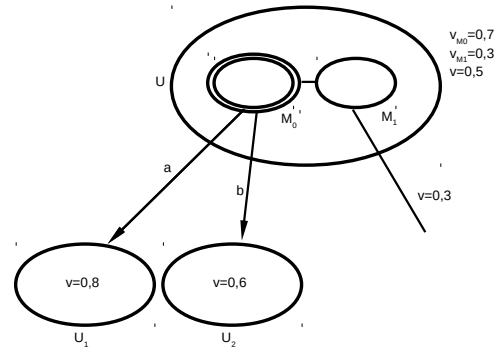


FIGURE 6 – Illustration du troisième principe

des deux mondes, ce qui donne 0,5 et nous affectons cette valeur à l'univers.

La méthode de détermination de l'intention d'un joueur que nous proposons s'appuiera sur ces 3 principes.

**Exemple.** Soient 3 joueurs et une propriété  $p$  arbitraire. On souhaite déterminer la meilleure action pour le joueur 1 pour obtenir la propriété  $p$  en 3 tours. On commence par développer l'arbre des univers comme décrit précédemment mais sur 3 tours ; cet arbre comporte 4 niveaux. Le 1<sup>er</sup> niveau est composé de l'univers initial. Le 2<sup>me</sup> niveau est composé de l'ensemble des univers imaginés par le joueur 1 après son tour. Le 3<sup>me</sup> niveau est composé de l'ensemble des univers possibles pour le joueur 1 après son tour et celui du joueur 2. Le 4<sup>me</sup> niveau est composé de l'ensemble des univers possibles pour le joueur 1 après le tour des joueurs 1, 2 et 3. On calcule la valeur des univers du 4<sup>me</sup> niveau à l'aide du premier principe. Ceci nous permet de calculer ensuite la valeur des univers du 3<sup>me</sup> et du 2<sup>me</sup> niveaux grâce au troisième principe. Enfin, la valeur de l'univers du 1<sup>er</sup> niveau est calculée au moyen du deuxième principe. Ceci permet de vérifier si l'action effectuée par le joueur 1 est celle qui, du point de vue du joueur en question, avait une meilleure chance d'obtenir la propriété  $p$ .

**Remarque.** La notion de valeur d'un univers est proche de celle d'utilité. En effet, pour les exemples les plus simples, la valeur de l'univers est simplement la fréquence d'apparition de la propriété recherchée au sein de l'univers, en d'autres termes, la probabilité de l'obtenir. Étant donné qu'un univers est lié à l'action qui permet de l'atteindre, on peut attribuer à chaque action une mesure de son utilité pour obtenir la propriété. Cela définit donc une utilité. Cependant, nous préférons ne pas utiliser ce terme car nous donnons des valeurs à des états et non pas à des actions.

## 4.2 L'intention dans les jeux épistémiques

La notion d'agent devient ici plus importante, puisqu'elle n'est plus seulement caractéristique des relations d'accessibilité entre les mondes mais également des actions possibles. En effet, dans la présentation de la logique épistémique dynamique, la notion d'agent apparaît uniquement pour donner une signification aux différentes relations d'accessibilité (par exemple, la relation d'accessibilité  $R_3$  représente les connaissances de l'agent 3). Il n'est pas utile de préciser la relation qui se rapporte à un agent en particulier, les raisonnements sur la relation suffisent. Ce que nous allons faire ici est d'associer à la fois un agent à sa relation et aux actions qu'il va entreprendre. Il est donc important de construire un système possédant un lien entre l'ensemble des agents, l'ensemble des relations et l'ensemble des actions. On s'attarde donc à séparer les entités, il faut donc ré-associer les agents et leurs relations d'accessibilité, c'est ce qui est décrit dans les deux définitions suivantes :

**Définition 2** Soit  $M$  un ensemble de mondes,  $N$  un ensemble d'agents et  $R^M$  un ensemble de relations d'accessibilité sur  $M$  vérifiant  $\text{Card}(R^M) = \text{Card}(N)$ . On appelle fonction d'association des relations entre  $N$  et  $R^M$  toute fonction bijective de  $N$  vers  $R^M$ .

La définition 2 permet d'associer une relation d'accessibilité à tout agent et inversement.

**Définition 3** Soit  $\mathbb{P}$  un ensemble de propositions,  $M$  un ensemble de mondes,  $N$  un ensemble d'agents,  $h$  une fonction de valuation,  $R^M$  un ensemble de relations d'accessibilité sur  $M$  tel que  $\text{Card}(R^M) = \text{Card}(N)$  et  $f_r$  une fonction d'association des relations entre  $N$  et  $R^M$ . On appelle univers le 6-uplet  $U = (M, N, R^M, f_r, \mathbb{P}, h)$ .

Pour des raisons de lisibilité, on notera indifféremment  $f_r(\alpha)$  et  $R_\alpha$  la relation d'accessibilité liée à l'agent  $\alpha$ . Pour rappel, dans ce système, la relation  $R_\alpha$  lie les mondes indistinguables pour le joueur  $\alpha$ , pour qui ils ont donc tous autant de chances d'être le monde réel. Attribuer le concept de tour de jeu à la logique épistémique peut se révéler compliqué et contraignant pour les définitions. Introduire les actions simultanées de tous les joueurs (*action jointe*) a été préféré. La notion de tour de jeu est alors modélisée en utilisant l'action *nop* : à chaque tour, un joueur joue pendant que les autres effectuent l'action *nop* (ne font rien). La définition suivante d'action jointe se base sur la logique ATDEL [10].

**Définition 4** Soit  $\mathbb{P}$  un ensemble de propositions et  $A$  un ensemble d'actions sur  $\mathbb{P}$ . On note  $A^n$  l'ensemble formé par  $A^n = \times_1^n A$ . On appelle  $a^n = (a_1, a_2, \dots, a_n)$

action jointe un élément de  $A^n$  défini par une pré-condition jointe et une post-condition jointe noté  $a^n = (\text{pre}(a^n), \text{post}(a^n))$ . La pré-condition jointe vérifie  $\text{pre}(a^n) = \bigwedge_{i=1}^n \text{pre}(a_i)$ . La post-condition jointe vérifie, pour chaque élément  $p$  de  $\mathbb{P}$  :

$$\text{post}(a^n)(p) = \begin{cases} \top & \text{si } \forall a_i, \text{post}(a_i)(p) = \top \\ \perp & \text{si } \forall a_i, \text{post}(a_i)(p) = \perp \\ \neg p & \text{si } \forall a_i, \text{post}(a_i)(p) = \neg p \\ \text{sinon } p & \end{cases}$$

Pour simplifier, nous emploierons  $a$  à la place de  $a^n$  lorsqu'il n'y aura pas d'ambiguïté.

**Définition 5** Soit un univers  $U = (M, N, R^M, f_r, \mathbb{P}, h)$ , un ensemble  $A$  d'actions et une action jointe  $a \in A^n$  où  $n = \text{card}(N)$ . L'univers  $U|_a = (M|_a, N, R^M|_a, f_r, \mathbb{P}, h|_a)$  engendré par l'exécution de l'action jointe  $a$  vérifie :

- $M|_a = \{M_i \in M \text{ tel que } \models_{M_i}^U \text{pre}(a)\}$  est la restriction à l'ensemble des mondes vérifiant la pré-condition de  $a$ .
- $\forall \alpha \in N$  on a  $R_{\alpha|_a} = R_\alpha \cap (M|_a \times M|_a)$  est la restriction des relations aux mondes de  $M|_a$
- $h|_a(p) = \{M_i \in M \text{ tel que } \models_{M_i}^U \text{post}(a)(p)\} \cap M|_a$  est la restriction de la valuation aux mondes de  $M|_a$  et réaffectation de la valeur des variables propositionnelles

Vient ensuite le concept de l'utilité des actions des autres joueurs vu par le joueur qui nous intéresse. On définit donc une fonction d'utilité qui va associer à chaque monde et à chaque joueur une valeur.

**Définition 6** Soit  $A$  un ensemble fini d'actions et  $U = (M, N, R^M, f_r, \mathbb{P}, h)$  un univers. On appelle fonction d'utilité  $u$  toute fonction qui associe à un triplet  $(a, \alpha, M_i) \in A \times N \times M$  un réel  $x \in \mathbb{R}$ . On appelle  $\Omega$  Univers Actionnable Mesuré (UAM) tout triplet  $(U, A, u)$ .

**Définition 7** Soit  $\Omega = (U, A, u)$ , un monde  $M_i \in M$  et un agent  $\alpha \in N$ . On note  $A_{M_i, \alpha}^\Omega$  l'ensemble des actions les plus utiles pour  $\alpha$  pour le monde  $M_i$  où  $A_{M_i, \alpha}^\Omega = \{a \in A \mid \forall b \in A, u(a, \alpha, M_i) \geq u(b, \alpha, M_i)\}$

**Proposition 1** Soit  $\Omega = (U, A, u)$  et  $\alpha \in N$ .  $\forall M_i \in M$ , l'ensemble  $A_{M_i, \alpha}^\Omega$  est non vide et fini.

*Preuve* : Nous avons  $A_{M_i, \alpha}^\Omega \subseteq A$ . Comme  $A$  est fini, alors  $A_{M_i, \alpha}^\Omega$  est fini. De plus, comme  $A$  est fini,  $\forall \alpha \in N, \forall M_i \in M, \{u(a, \alpha, M_i) \mid a \in A\}$  est fini et possède donc un plus grand élément. De ce fait,  $A_{M_i, \alpha}^\Omega$  est non vide. QED.

À présent, nous allons définir formellement le premier principe, c'est-à-dire le calcul de la fréquence d'une variable propositionnelle au sein des mondes indistinguables du monde réel pour le joueur.

**Définition 8** Soit  $U = (M, N, R^M, f_r, \mathbb{P}, h)$  un univers,  $M_0 \in M$  et  $\alpha \in N$ . On note  $U^{M_0}$  l'univers tel que  $M_0$  est le monde réel et on appelle  $I_\alpha^{U^{M_0}}$  ensemble des mondes envisageables de 1<sup>er</sup> degré pour  $\alpha$  l'ensemble de tous les mondes  $M_i \in M$  tel que  $M_0 R_\alpha M_i$ .

**Définition 9** Soit un univers  $U^{M_0}$ , une formule  $p$  et un agent  $\alpha$ . On appelle  $pr_\alpha^{U^{M_0}}(p)$  rapport de présence de  $p$  pour  $\alpha$  le rapport :  $pr_\alpha^{U^{M_0}}(p) = \frac{\text{Card}(I_\alpha^{U^{M_0}} \cap h(p))}{\text{Card}(I_\alpha^{U^{M_0}})}$  si  $I_\alpha^{U^{M_0}} \neq \emptyset$  sinon  $pr_\alpha^{U^{M_0}}(p) = 0$

Ce qui nous permet de calculer la valeur de l'univers suivant une propriété pour  $k$  tours de jeu à l'avance.

**Définition 10** Soit un univers actionnable mesuré  $\Omega = (U^{M_0}, A, u)$ , un agent  $\alpha \in N$ , une formule  $p$  et un entier positif  $k$ . On appelle  $v(U^{M_0}, A, u, \alpha, p, k)$  valeur de l'univers  $U^{M_0}$  pour l'agent  $\alpha$  à l'ordre  $k$  le réel suivant, avec  $A_{M_i}^\alpha = A_{M_i,1}^\alpha \times A_{M_i,\alpha-1}^\alpha \times A_{M_i,\alpha+1}^\alpha \times A_{M_i,n}^\alpha$  :  $v(U^{M_0}, A, u, \alpha, p, k \geq 1) = \max_{a \in A}(\text{moymonde})$  où  $\text{moymonde} = \sum_{M_i \in I_\alpha^{U^{M_0}}} \text{moyaction} \times \frac{1}{\|I_\alpha^{U^{M_0}}\|}$  où  $\text{moyaction} = \sum_{(a_1, \dots, a_{n-1}) \in A_{M_i}^\alpha} \text{valeur} \times \frac{1}{\|A_{M_i}^\alpha\|}$  où  $\text{valeur} = v(U_{|(a_1, \dots, a_n)}^{M_i}, A, u, \alpha, p, k - 1)$  et  $v(U^{M_0}, A, u, \alpha, p, 0) = pr_\alpha^{U^{M_0}}(p)$

Le maximum caractérise le deuxième principe, tandis que la double somme caractérise le troisième principe. Lorsque c'est le tour du joueur qui nous intéresse, la seule action exécutée par les autres joueurs est l'action *nop*. La deuxième somme ne possède alors qu'un seul élément, il n'y a donc pas de moyenne sur les actions des autres joueurs, ce qui correspond au deuxième principe. Lorsque c'est le tour de l'un des autres joueurs, la seule action maximum pour le joueur est l'action *nop*, ses autres actions donnant des univers vides et donc des valeurs nulles. Il y a donc une moyenne sur les actions et sur les mondes, ceci correspond au troisième principe.

**Définition 11** Soit un univers actionnable mesuré  $\Omega = (U^{M_0}, A, u)$ , un agent  $\alpha \in N$ , une formule  $p$ , une action  $a \in A$  et un entier strictement positif  $k$ . On appelle  $v_a(U^{M_0}, A, u, \alpha, p, k)$  valeur de l'univers  $U^{M_0}$  pour l'agent  $\alpha$  à l'ordre  $k$  suivant l'action  $a \in A$  le réel suivant, avec  $A_{M_i}^\alpha = A_{M_i,1}^\alpha \times A_{M_i,\alpha-1}^\alpha \times A_{M_i,\alpha+1}^\alpha \times A_{M_i,n}^\alpha$  :  $v_a(U^{M_0}, A, u, \alpha, p, k \geq 1) = \text{moymonde}$  où  $\text{moymonde} = \sum_{M_i \in I_\alpha^{U^{M_0}}} \text{moyaction} \times \frac{1}{\|I_\alpha^{U^{M_0}}\|}$  où  $\text{moyaction} = \sum_{(a_1, \dots, a_{n-1}) \in A_{M_i}^\alpha} \text{valeur} \times \frac{1}{\|A_{M_i}^\alpha\|}$  où  $\text{valeur} = v(U_{|(a_1, \dots, a_n)}^{M_i}, A, u, \alpha, p, k - 1)$

**Définition 12** Soit un univers actionnable mesuré  $\Omega = (U^{M_0}, A, u)$ , un agent  $\alpha \in N$ , une variable propositionnelle  $p \in \mathbb{P}$  et un entier strictement positif  $k$ , on appelle  $Best(\Omega, p, \alpha, k)$  ensemble des meilleures actions pour obtenir une propriété

propriété  $p$  au  $k$ -ième tour pour le joueur  $\alpha$  l'ensemble suivant :  $Best(\Omega, p, \alpha, k) = \{a \in A \mid \forall b \in A, v_a(U^{M_0}, A, u, \alpha, p, k) \geq v_b(U^{M_0}, A, u, \alpha, p, k)\}$

Ces deux dernières définitions permettent de caractériser les meilleures actions suivant le joueur pour obtenir une propriété  $k$  tours de jeu à l'avance. Ainsi, en se donnant une propriété  $p$ , nous pouvons vérifier si le joueur a effectué la meilleure suite d'actions pour l'obtenir. Si tel est le cas, alors l'obtention de  $p$  est intentionnelle, sinon nous considérons que  $p$  a été obtenue par accident. Par exemple, si l'agent  $\alpha$  a effectué l'action  $a$  et si  $a \in Best(\Omega, p, \alpha, k)$  alors cela était intentionnel. Cela devient faux si toutes les actions font partie des meilleures actions ( $Best(\Omega, p, \alpha, k) = A$ ), dans ce cas aucune action n'est intentionnelle.

### 4.3 Utilité

Le calcul de la meilleure action demande de définir une utilité pour les actions, ou plus exactement, une utilité du point de vue du joueur pour les actions des autres joueurs. Trouver la meilleure façon de modéliser l'utilité pour un jeu est un travail de recherche en lui-même et ce n'est pas celui que nous avons voulu effectuer ici. Nous proposons donc une définition d'utilité basée sur la récompense de fin de partie qui est calculable par n'importe quel joueur pour n'importe quel autre joueur dès lors que l'information sur la récompense de fin de partie est accessible à tous les joueurs.

Le joueur imagine plusieurs mondes possibles et pour chaque monde qu'il imagine, les autres joueurs peuvent imaginer d'autres mondes. Dans la figure 7, pour le joueur 1, le monde  $M_0$  et le monde  $M_1$  sont indistinguables du monde réel  $M_0$ . Pour le joueur 2, les mondes  $M_0$  et  $M_5$  sont indistinguables et les mondes  $M_1$  et  $M_2$  sont indistinguables.

Nous commençons par définir l'utilité des actions pour le joueur 2 d'après le joueur 1. Le joueur 1 sait que le joueur 2 hésite soit entre la clique de mondes  $M_0$  et  $M_5$ , soit entre la clique de mondes  $M_1$  et  $M_2$ . Il faut donc traiter ces deux cliques séparément. Nous nous focalisons d'abord sur la première clique. Pour calculer l'utilité d'une action du joueur 2 suivant le monde de la clique, on effectue le raisonnement suivant :

- on suppose que ce monde est le monde réel ;
- on regarde le score maximum que l'on pourrait atteindre à la fin du jeu après avoir effectué l'action jointe du joueur 2 et des autres joueurs (ici le joueur 1 seulement) ;
- on calcule la moyenne sur l'ensemble des actions des autres joueurs (ici le joueur 1 seulement) ;
- on définit cette moyenne comme l'utilité de l'action pour ce monde.

Comme les joueurs envisagent ces mondes de manière équiprobable, pour comparer les actions, il faut calculer la moyenne des utilités sur l'ensemble des mondes de la clique. On obtient l'utilité de l'action pour cette



clique. Enfin, on définit l'utilité d'une action du joueur 2 d'un monde possible pour le joueur 1 comme étant égale à l'utilité de la clique qu'elle génère pour le joueur 2. Pour les autres mondes, la valeur de l'utilité, n'ayant pas d'importance, est fixée arbitrairement.

Pour ce qui concerne les utilités du joueur 1 pour le joueur 1 elles n'ont pas d'importance, elles sont donc aussi fixées arbitrairement.

Pour les jeux au tour par tour, seule l'action *nop* ne donne pas un univers vide lorsque ce n'est pas au joueur de jouer. On ne peut donc pas à priori donner de valeurs à l'utilité pour les autres actions. Donner une valeur identique à chaque univers vide ne changerait pas l'ordre des valeurs des utilités des actions. On peut aussi sauter cette étape de calcul de moyenne sur les actions des autres joueurs. Dans les deux cas, l'action de plus grande utilité sera la même.

#### 4.4 Exemple

Nous allons utiliser le jeu Hanabi dans une version simplifiée pour illustrer l'utilisation des définitions précédemment données. Dans le jeu simplifié, il n'y a plus que les cartes 1, 2 et 3 rouges et deux joueurs, il y a toujours 8 jetons indices et 3 jetons de vie. Il n'y a plus que trois actions (D, P et I) : défausser sa carte, poser sa carte, donner un indice (qui correspond ici à donner la valeur de la carte de l'autre). Lors de la donne, le joueur 1 avait reçu la carte 1, le joueur 2 avait reçu la carte 2 et la pioche contenait donc le 3. La figure 7 présente l'univers initial. La partie s'est déroulée, le joueur 1 a commencé par donner un indice "tu as la carte 2" et après le tour du joueur 2, le joueur 1 connaît sa carte et la carte 2 n'est pas défaussée. Nous allons donc vérifier si la propriété « le joueur 2 n'a pas défaussé et le joueur 1 connaît sa carte », noté  $p = \neg D : 2 \wedge (B_1 J1 : 1 \vee B_1 J1 : 2 \vee B_1 J1 : 3)$  a été obtenue intentionnellement.

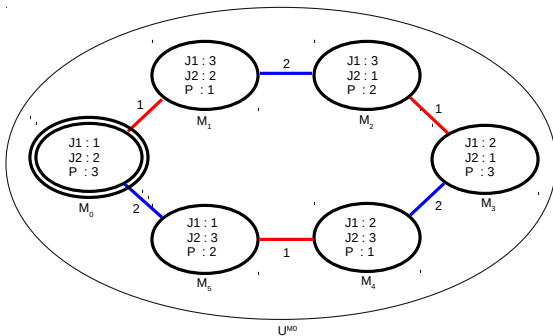


FIGURE 7 – Univers initial

À partir de l'univers initial, figure 7, le joueur 1 imagine deux mondes possibles :  $M_0$  et  $M_1$ . Nous devons donc considérer deux cas de figure : l'un où le

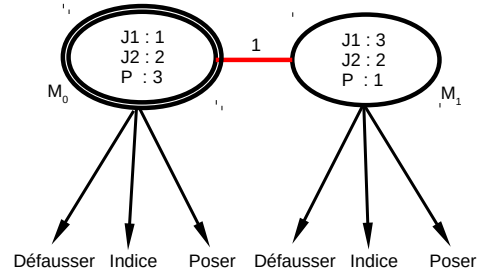


FIGURE 8 – Initialisation du calcul de la valeur des actions du joueur 1

monde réel est  $M_0$  et l'autre où le monde réel est  $M_1$  (voir figure 8). Nous calculons la valeur de l'action « donner un indice ». Le joueur 1 va déclarer « tu as la carte de valeur 2 » ceci a pour conséquence que les mondes ne vérifiant pas « le joueur 2 à la carte de valeur 2 » disparaissent. Dans cet exemple, effectuer cette action depuis le monde  $M_0$  (respectivement le monde  $M_1$ ) donne l'univers  $U_{|I}^{M_0}$  (respectivement  $U_{|I}^{M_1}$ , seul le monde dit "réel" diffère) présenté en figure 9. Dans cet univers  $U_{|I}^{M_0}$ , le joueur 2 a les trois mêmes actions possibles. Si le monde  $M_0$  est le monde réel, l'action D et P ont une utilité de 1 et l'action I a une utilité de 3. Le joueur 1 pense donc que le joueur 2 va effectuer cette action, qui permettra d'obtenir la propriété avec une fréquence de 1. Si le monde  $M_1$  est le monde réel, les trois actions D, P et I ont une utilité de 1. Pour le joueur 1, le joueur 2 va effectuer équiprobablement l'une de ces trois actions. Les actions D et P ne permettent pas d'obtenir la propriété, l'action I permet d'obtenir la propriété avec une fréquence de 1. La valeur de cet univers est donc :  $\frac{1}{2} + \frac{1}{3}$ .

On effectue le même raisonnement pour  $U_{|I}^{M_1}$ , qui, dans cet exemple, présente les mêmes calculs que  $U_{|I}^{M_0}$ . En calculant la moyenne de la valeur de ces deux univers on obtient la valeur de l'action « donner un indice » :

$$v_{Indice}(U^{M_0}, A, u, 1, p, 2) = \frac{\frac{1}{2} + \frac{1}{3}}{2} + \frac{\frac{1}{2} + \frac{1}{3}}{2} = \frac{2}{3}$$

De même, on trouve  $v_{Défausser}(U^{M_0}, A, u, 1, p, 2) = \frac{1}{3}$  et  $v_{Poser}(U^{M_0}, A, u, 1, p, 2) = \frac{2}{3}$ . Les meilleures actions sont donc l'action de donner un indice et l'action de poser. Le joueur avait donc bien l'intention d'obtenir la propriété « je connais ma carte, et le joueur 2 n'a pas défaussé » au bout de deux tours.

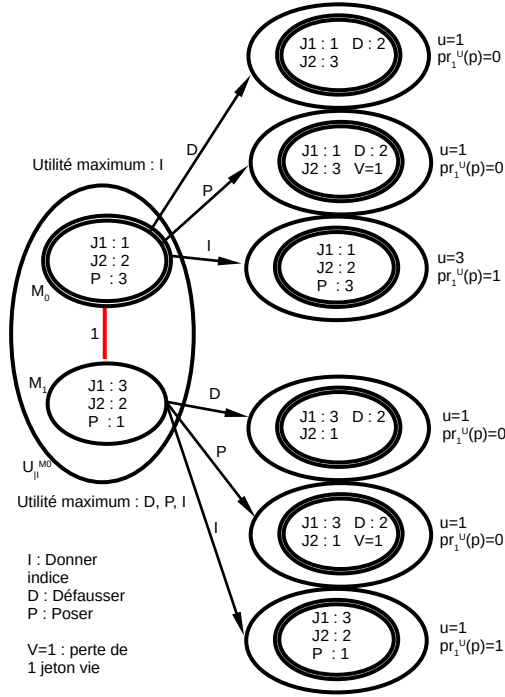


FIGURE 9 – Calcul de la valeur de l'action « donner un indice » pour le monde  $M_0$

## 5 Conclusion

Nous avons défini une première modélisation de l'intention puisqu'il nous est à présent possible de vérifier si une propriété donnée a été obtenue intentionnellement ou non.

Ceci n'est pas le premier travail qui intègre intention et logique épistémique. Le travail de Lorini et Herzig [12] modélise l'intention via des opérateurs de tentatives réussies ou échouées. Toutefois, leur logique modélise le passage du temps avec une ligne droite (c'est-à-dire qu'il n'y a qu'un seul futur possible), il n'y a pas d'embranchements. Il est donc beaucoup moins naturel de capturer une sémantique de jeu.

Dans notre approche, il est à noter que plus on prend une formule spécifique à l'univers dans lequel on est arrivé, plus il est certain que les actions seront considérées comme intentionnelles. Une formule plus générale permettrait donc a priori de mieux caractériser l'intention. La définition de la généralité d'une formule au sein d'un ensemble d'univers est une des voies pour continuer le travail présenté dans cet article.

Par ailleurs, les mondes sont ici considérés comme équiprobables, il pourrait être intéressant, dans des travaux futurs, d'intégrer des logiques pondérées comme celle présentée par Legastelois [13].

Notre modélisation prend en compte des joueurs capables d'imaginer tous les mondes possibles et tous les univers qui en découleraient suivant les actions qu'ils

pensent les plus pertinentes. Si une machine a déjà des ressources limitées, ce même travail pour un humain est encore plus difficile. L'une des pistes envisagées pour tenir compte de cette limitation serait de considérer une action comme meilleure à partir du moment où sa valeur dépasse un seuil donné (par exemple, 80% de la véritable valeur maximale).

Un logiciel est en cours de développement, il nous permettra de confronter les concepts que nous avons construits à la réalité du jeu Hanabi. Il permettra également de tester l'utilité que nous avons définie et de voir s'il est possible de retrouver des stratégies telles que celles définies dans [14].

## Références

- [1] A. Baltag and L.S. Moss. *Logic for epistemic programs* Synthese, 139 :165-224, 2004
- [2] A. Baltag and L.S. Moss. and S. Solecki. *The logic of common knowledge, public announcements and private suspicions* Technical Report SEN-R9922, CWI, Amsterdam, 1999
- [3] B.F. Chellas. *Modal Logic*, Cambridge University Press, pp. 3-160, 1980.
- [4] H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*, Springer, 2007.
- [5] M. Eger. *Intentional Agents for Doxastic Games*, North Carolina State University, 2018
- [6] R. Fagin and J.Y Halpern and Y. Moses and M.Y. Vardi. *Reasoning about knowledge*, MIT Press, pp. 1-232, 2003.
- [7] J.D. Gerbrandy and W. Groeneveld. *Reasoning about information change*, Journal of Logic, Language, and Information, 6 :147-169, 1997
- [8] S.A. Kripke. *Semantical Analysis of Modal Logic*, The Journal of Symbolic Logic, 1959.
- [9] D.K. Lewis. *Convention : A Philosophical Study*, 1969
- [10] T. de Lima. *Alternating-time temporal dynamic epistemic logic*, Journal of Logic and Computation, vol 24, pp. 1145-1178, 2014
- [11] J.M. Plaza. *Logics of public communications* Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems, pp. 201-2016, 1989
- [12] E. Lorini and A. Herzig *A logic of intention and attempt* Kluwer Academic Publishers, 2007
- [13] B. Legastelois *Extension pondérée des logiques modales dans le cadre des croyances graduelles* Université Pierre et Marie Curie - Paris VI, 2017
- [14] B. Bouzy *Playing Hanabi Near-Optimally* Université Descartes - Paris V, 2017