Information-Interaction-Intelligence, numéro hors série

Formalisation de l'évolution des croyances dans le Calcul des Situations

Robert Demolombe[⋆], Pilar Pozos Parra[†⋆]

[⋆] ONERA Toulouse,
France
Robert.Demolombe@cert.fr

[†] ONERA Toulouse,
France
Pilar.Pozos@cert.fr

Résumé

La formalisation des interactions entre agents pose le problème de la modélisation de l'évolution des croyances des agents. Nous nous appuyons sur une solution proposée par les chercheurs du Cognitive Robotic Group de Toronto pour un seul agent, qui est exprimée dans le Calcul des Situations. Nous montrons les problèmes que pose cette solution dans le cas de plusieurs agents, et nous proposons une solution qui permet de les résoudre.

Mots-clés : Multi Agents, Croyances, Calcul des Situations

1 INTRODUCTION

L'un des problèmes les plus importants et les plus difficiles concernant la formalisation des aspects dynamiques est le "frame problem". Le "frame problem" consiste à exprimer l'idée, qui paraît très simple, qu'après avoir réalisé une action les propositions qui ne sont pas liées logiquement aux effets de l'action ne changent pas de valeur de vérité. De nombreuses solutions ont été étudiées depuis les débuts de l'Intelligence Artificielle [1, 5, 4] qui

---

[⋆] ONERA/DTIM, 2 Avenue E. Belin B.P. 4025, 31055 Toulouse Cedex, France. Cette recherche a été partiellement financée par une bourse attribuée par CONACYT.

ont presque toutes en commun de sortir du cadre de la logique classique du premier ordre.

Une solution a été proposée par R. Reiter [8] et elle a été étendue par d'autres chercheurs du Cognitive Robotic Group de l'Université de Toronto [10, 6, 11]. Dans [3] nous avons proposé une solution alternative, qui présente de fortes restrictions mais qui a pu être partiellement implémentée. La solution de Reiter est exprimée dans le Calcul des Situations, qui est une logique classique du premier ordre avec égalité, dans laquelle on distingue pour chaque prédicat dont la valeur de vérité peut changer dynamiquement un argument (le dernier) qui a un type particulier, le type "situation".

Par exemple, le fait que le verre x est cassé dans la situation s est représenté par : $cass\acute{e}(x, s)$. Les situations sont dénotées soit par des constantes, par exemple la situation initiale S_0, soit par des variables, généralement dénotées par : s, s', s'',..., soit par des termes de la forme $do(a, s)$, où le premier argument est un terme de type action et le second un terme de type situation. Si $tomber(x)$ et $r\acute{e}parer(x)$ sont du type action, alors les termes suivants sont du type situation : $do(tomber(X1), S_0)$, $do(r\acute{e}parer(x), do(tomber(x), S_0))$, $do(a, s)$,...

La signification intuitive de la formule atomique : $cass\acute{e}(X1, do(tomber(X1, S_0)))$ est que dans la situation qui résulte de la réalisation de l'action $tomber(X1)$ à partir de la situation S_0, le verre $X1$ est cassé. Les variables de type action et de type situation peuvent être quantifiées comme les autres variables. Par exemple, les formules suivantes font partie du langage du Calcul des Situations : $\forall s\, cass\acute{e}(X1, do(tomber(X1), s))$, $\exists a \exists x\, \neg cass\acute{e}(x, do(a, do(tomber(x), S_0)))$. On verra plus loin que la possibilité de quantifier sur les actions et sur les situations joue un rôle clé dans la solution qui est proposée pour le frame problem.

On présente cette solution d'abord sur un exemple. Supposons que dans toute situation, si on réalise l'action $tomber(x)$, il en résulte que le verre x est cassé, et que si on réalise l'action $r\acute{e}parer(x)$, il en résulte que x n'est plus cassé. Ceci s'exprime par :

$(S1)$ $\forall s \forall a \forall x (a = tomber(x) \rightarrow cass\acute{e}(x, do(a, s)))$

$(S2)$ $\forall s \forall a \forall x (a = r\acute{e}parer(x) \rightarrow \neg cass\acute{e}(x, do(a, s)))$

La solution du frame problem découle de l'hypothèse qu'il n'y a pas d'autre action, ou condition, que celles exprimées dans l'antécédent de $(S1)$ (respectivement $(S2)$) qui ont pour effet que dans la situation $do(a, s)$ on a $cass\acute{e}(x)$ qui est vraie (respectivement fausse). Plus simplement, on suppose que ces conditions donnent une définition complète des situations qui ont pour effet que $cassé(x)$ est vraie (respectivement fausse). Ces hypothèses s'expriment formellement par:

$(S3)$ $\forall s \forall a \forall x(\neg cassé(x,s) \wedge cassé(x, do(a,s)) \rightarrow a = tomber(x))$

$(S4)$ $\forall s \forall a \forall x(cassé(x,s) \wedge \neg cassé(x, do(a,s)) \rightarrow a = réparer(x))$

On montre facilement que l'ensemble $(S1)$, $(S2)$, $(S3)$ et $(S4)$, est équivalent à (S).

(S) $\forall s \forall a \forall x(cassé(x, do(a,s)) \leftrightarrow a = tomber(x) \vee cassé(x,s) \wedge \neg(a = réparer(x)))$

(S) est appelé un "axiome de changement d'état" ("successor state axiom" dans [9]). Cet axiome suffit à déterminer la valeur de vérité de $cassé(x)$ en $do(a,s)$ quelle que soit l'action réalisée.

Par exemple, si on a $\neg cassé(X1, S_0)$, on en déduit $\neg cassé(X1, do(tomber(X2), S_0))$ et $cassé(X2, do(tomber(X2), S_0))$. Ce qui signifie qu'après avoir laissé tomber le verre $X2$ celui-ci est cassé, mais le verre $X1$ reste intact.

Dans le cas général, pour chaque prédicat p qui peut changer de valeur de vérité d'une situation à l'autre, on suppose que l'on a un axiome de changement d'état de la forme de (S_p). Pour simplifier les notations on a omis les variables autres que a et s.

(S_p) $\forall s \forall a(p(do(a,s)) \leftrightarrow \Gamma_p^+(a,s) \vee p(s) \wedge \neg\Gamma_p^-(a,s))$

Les conditions Γ_p^+ et Γ_p^- concernent la situation s et l'action a, mais pas la situation $do(a,s)$. De plus, on suppose qu'elles ne peuvent pas être toutes les deux vraies dans la même situation. Soit $\neg\exists s \exists a(\Gamma_p^+(a,s) \wedge \Gamma_p^-(a,s))$.

L'ensemble des axiomes tels que (S_p) déterminent la valeur de vérité de chaque formule atomique après avoir réalisé n'importe quelle action. Ce qui résout le frame problem.

2 EXTENSION À L'ÉVOLUTION DES CROYANCES D'UN AGENT

Le Calcul des Situations a été étendu aux croyances d'un agent par Scherl et Levesque dans [10]. D'un point de vue formel cette extension repose sur l'idée d'exprimer en logique classique du premier ordre la définition de la modalité de croyance. Pour cela le langage est étendu avec un prédicat binaire $K(s', s)$ [1] dont les arguments sont du type situation. Ce prédicat joue

1. L'usage [2] veut que dans une relation d'accessibilité $R(w, w')$ le monde w' représente une alternative au monde w par rapport à la croyance. Nous avons respecté ici la notation utilisée dans [9] où l'ordre des arguments est inversé.

le même rôle qu'une relation d'accessibilité dans la sémantique des mondes possibles. La différence est qu'il relie des situations au lieu de mondes possibles, et qu'il est introduit dans l'axiomatique et non dans la sémantique. Dans [6], Lakemeyer et Levesque ont utilisé ce formalisme pour exprimer qu'un agent "sait seulement si" une proposition est vraie.

Le fait que dans une situation s un agent croit que ϕ est vraie est dénoté par $Knows(\phi, s)$. Ici ϕ est une formule obtenue à partir d'une formule du Calcul des Situations dans laquelle tous les arguments des prédicats de type situation ont été supprimés. On suppose qu'ils concernent tous la situation s.

Par exemple, à partir de la formule $cass\acute{e}(X1, s) \wedge \neg cass\acute{e}(X2, s)$ on obtient $\phi_1 = cass\acute{e}(X1) \wedge \neg cass\acute{e}(X2)$, et $Knows(cass\acute{e}(X1) \wedge \neg cass\acute{e}(X2), s)$.

La signification attribuée à $Knows(cass\acute{e}(X1) \wedge \neg cass\acute{e}(X2), s)$ est que dans la situation s l'agent croit que dans toutes les situations s' compatibles avec ses croyances (c'est-à-dire telles que $K(s', s)$), on a $cass\acute{e}(X1, s') \wedge \neg cass\acute{e}(X2, s')$ qui est vraie. C'est pour cette raison qu'on impose que les croyances portent sur des formules dans lesquelles les arguments de type situation réfèrent tous la situation s qui apparaît dans $Knows(\phi, s)$.

Si on a $Knows(\phi, s)$ on note $\phi[s']$ la formule obtenue à partir de ϕ en rajoutant l'argument s' à chaque prédicat qui peut changer de valeur de vérité. On a alors la définition :

$Knows(\phi, s) \stackrel{\text{def}}{=} \forall s'(K(s', s) \rightarrow \phi[s'])$

d'où

$Knows(cass\acute{e}(X1) \wedge \neg cass\acute{e}(X2), s) \stackrel{\text{def}}{=} \forall s'(K(s', s) \rightarrow cass\acute{e}(X1, s') \wedge \neg cass\acute{e}(X2, s'))$.

Les formules telles que ϕ peuvent elles-mêmes contenir des modalités. On peut avoir, par exemple, $Knows(\neg Knows(cass\acute{e}(X2)), S_0)$ qui exprime qu'en S_0 l'agent croit qu'il ne croit pas que $X2$ est cassé. D'après la définition précédente on a :

$Knows(\neg Knows(cass\acute{e}(X2)), S_0) \stackrel{\text{def}}{=} \forall s'(K(s', s) \rightarrow \neg Knows(cass\acute{e}(X2), s'))$

et on a aussi :

$Knows(cass\acute{e}(X2), s') \stackrel{\text{def}}{=} \forall s''(K(s'', s') \rightarrow cass\acute{e}(X2, s''))$

d'où :

$Knows(\neg Knows(cass\acute{e}(X2)), S_0) \stackrel{\text{def}}{=}$
$\forall s'(K(s', s) \rightarrow \exists s''(K(s'', s') \wedge \neg cass\acute{e}(X2, s'')))$

Pour définir la façon dont les croyances évoluent quand des actions sont réalisées on distingue deux aspects.

Le premier concerne l'évolution de la relation d'accessibilité $K(s', s)$. Il faut définir les situations s'' qui sont accessibles après avoir réalisé l'action a à partir de s, c'est-à-dire telles que $K(s'', do(a, s))$. Pour cela on distingue deux types d'actions :

– celles qui permettent d'acquérir de l'information sur le monde. On les appelle "actions de perception" ("sensing action" en anglais). Chaque action de perception permet à l'agent de savoir si une formule ϕ associée à cette action est vraie, ou fausse, dans la situation où il se trouve. Ces actions ont pour effet d'éliminer des situations considérées comme possibles mais qui sont inconsistantes avec l'information acquise par l'observation. On suppose que ces actions peuvent modifier les croyances mais qu'elles ne modifient pas le monde.

– celles qui modifient le monde. Pour chaque situation possible elles projettent leurs effets sur la situation suivante, et éventuellement elles modifient les croyances en conséquence.

Le deuxième aspect concerne la détermination des valeurs de vérité des atomes dans les situations s'' accessibles depuis $do(a, s)$ par K. Pour cela on utilise les axiomes de changement d'état tels que (S_p).

Nous allons illustrer ces deux aspects avec l'exemple précédent, mais en considérant qu'il y a un seul verre pour simplifier les notations. On a alors :

(S') $\forall s \forall a(cass\acute{e}(do(a, s)) \leftrightarrow a = tomber \vee cass\acute{e}(s) \wedge \neg(a = r\acute{e}parer))$

On considère la nouvelle action $observer$ qui permet de savoir dans une situation donnée si le verre est, ou non, cassé. Supposons, par exemple, qu'en S_0 l'agent ignore si le verre est cassé, et que le verre ne soit pas cassé. On a alors : $\neg Knows(cass\acute{e}, S_0) \wedge \neg Knows(\neg cass\acute{e}, S_0) \wedge \neg cass\acute{e}(S_0)$.

Il y a donc une situation s'_1 et une situation s'_2 accessibles depuis S_0 où on a respectivement $\neg cass\acute{e}(s'_1)$ et $cass\acute{e}(s'_2)$. Le fait qu'après avoir réalisé $observer$ l'agent sait que le verre n'est pas cassé, c'est-à-dire qu'on a $Knows(\neg cass\acute{e}, do(observer, S_0)))$, est exprimé formellement par le fait que les seules situations s' accessibles depuis S_0 qui ont un successeur $s'' = do(observer, s')$ accessible depuis $do(observer, S_0)$ sont celles pour lesquelles la valeur de vérité de $cass\acute{e}$ est la même en s' qu'en S_0. Dans l'exemple de la figure 1 c'est le cas des situations telles que s'_1. Plus généralement on a :

$\forall s \forall s'' \forall a(a = observer \rightarrow$
$(K(s'', do(a, s)) \leftrightarrow \exists s'(K(s', s) \wedge s'' = do(a, s) \wedge (cass\acute{e}(s') \leftrightarrow cass\acute{e}(s))))).$

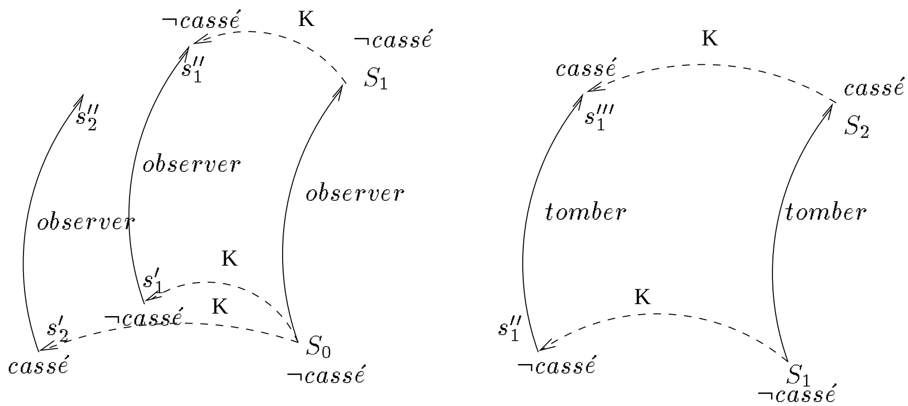

FIG. 1 – *Evolution des croyances d'un agent. Action de perception et action sur le monde.*

Comme on suppose que l'action $observer$ ne change pas la valeur de vérité des atomes, dans toutes les situations telles que $s_1'' = do(observer, s_1')$ on a $\neg cassé$ qui est vraie, et donc on a bien $Knows(\neg cassé, do(observer, S_0)))$.

Soit S_1 la situation $S_1 = do(observer, S_0)$. En S_1, après avoir réalisé une action qui n'est pas une action de perception, par exemple $tomber$, les nouvelles croyances de l'agent sont déterminées par les situations accessibles depuis $S_2 = do(tomber, S_1)$. Si, par exemple s_1'' est accessible depuis S_1, alors $s_1''' = do(tomber, s_1'')$ est accessible depuis S_2. On a donc $\forall s'''(K(s''', S_2) \leftrightarrow \exists s''(K(s'', S_1) \wedge s''' = do(tomber, s'')))$. Plus généralement on a :

$\forall s \forall s'' \forall a(\neg(a = observer) \rightarrow$
$(K(s'', do(a, s)) \leftrightarrow \exists s'(K(s', s) \wedge s'' = do(a, s'))))$.

D'autre part la valeur de vérité de $cassé(do(tomber, s_1''))$ est déterminée par l'axiome (S'). D'après (S') on a $\forall s(cassé(do(tomber, s))$. Comme toutes les situations s''' accessibles depuis S_2 sont telles que $s''' = do(tomber, s'')$, on a $cassé(s''')$ qui est vraie dans toutes ces situations. On a donc $Knows(cassé, S_2)$.

La forme générale de l'axiome définissant l'évolution de la relation d'accessibilité K est la suivante :

$$
\begin{array}{ll}
(S_K) & \forall s \forall s'' \forall a (K(s'', do(a,s)) \leftrightarrow \exists s'(K(s',s) \wedge s'' = do(a,s') \wedge (\\
& (\neg(a = \alpha_1) \wedge \ldots \wedge \neg(a = \alpha_n)) \\
& \vee\ a = \alpha_1 \wedge (\phi_1(s) \leftrightarrow \phi_1(s')) \\
& \ldots \\
& \vee\ a = \alpha_n \wedge (\phi_n(s) \leftrightarrow \phi_n(s'))))
\end{array}
$$

dans lequel chaque α_i dénote une action de perception qui permet de connaître la valeur de vérité de la formule ϕ_i. On peut remarquer que dans cette formalisation de l'évolution des croyances on fait implicitement deux hypothèses.

La première est que chaque fois qu'une action a été réalisée l'agent sait qu'elle a été réalisée. En effet, les croyances d'un agent évoluent chaque fois qu'une action a été réalisée. Dans le cas où on ne considère qu'un seul agent, et qu'il n'y a pas d'action "exogène" (réalisée par l'environnement ou par un autre agent), l'hypothèse est acceptable, mais dans le cas où on considère plusieurs agents, elle ne l'est plus.

La deuxième hypothèse est que l'agent sait quels sont les effets des actions. En effet, l'évolution de ses croyances est déterminée, entre autres, par les axiomes de changement d'état qui sont les mêmes en s et en s' telles que $K(s', s)$. Là aussi, dans le cas de plusieurs agents, il se peut que certains agents croient que les choses évoluent d'une certaine manière, et que d'autres croient qu'elles évoluent d'une manière différente. Donc, dans de nombreux cas cette hypothèse n'est pas acceptable.

C'est pour ne pas être contraint par ces deux hypothèses que nous proposons une variante de cette formalisation dans la section suivante.

3 EVOLUTION DES CROYANCES DE PLUSIEURS AGENTS

Pour exprimer formellement le fait que les croyances de chaque agent peuvent évoluer de façons différentes on suppose qu'il y a une relation d'accessibilité K_i pour chaque agent i. Ceci permet de définir une modalité $Knows_i(\phi, s)$ qui représente ce que croit l'agent i. On a la définition suivante.

$Knows_i(\phi, s) \stackrel{\text{def}}{=} \forall s'(K_i(s', s) \rightarrow \phi[s'])$

Pour ne pas être contraint par les hypothèses implicites que nous avons mentionnées dans la section précédente on considère que :

– l'ensemble des actions dont la réalisation est connue par un agent est donné explicitement, et peut être différent d'un agent à l'autre,

– les axiomes de changement d'état dépendent des types de situations auxquelles on les applique. Selon qu'une situation représente une situation réelle, ou une situation correspondant aux croyances de tel ou tel agent, on peux avoir des axiomes de changement d'état différents.

Nous allons voir, d'abord sur un exemple, comment ces deux principes peuvent être exprimés formellement.

On considère maintenant deux agents : une mère et un petit enfant. Les actions que peut réaliser la mère sont : $tomber_m$, $réparer_m$ et $observer_m$, et celles que peut réaliser l'enfant sont : $tomber_e$ et $observer_e$. On a les relations d'accessibilité K_m et K_e pour représenter respectivement les croyances de la mère et de l'enfant.

On suppose que les actions dont la réalisation est connue de la mère sont $tomber_m$, $réparer_m$, $observer_m$ et $tomber_e$, celles connues de l'enfant sont : $tomber_e$, $observer_e$ et $réparer_m$.

Dans [7], Lespérance, Levesque et Reiter proposent une solution dans laquelle un agent est toujours (sauf dans les situations qui suivent immédiatement une observation) dans l'ignorance au sujet des faits qui ont pu être modifiés par des actions dont il n'a pas connaissance.

Selon cette proposition, dans notre exemple, comme l'enfant ignore si la mère a réalisé l'action $tomber_m$, même s'il vient d'observer que le verre n'est pas cassé, il n'exclut pas, dans la situation suivante, que la mère l'ait laissé tombé et qu'il soit cassé. Et donc il sera dans l'ignorance par rapport au fait qu'il est cassé, ou non. Cette attitude nous paraît excessivement prudente, ou sceptique.

Dans la solution que nous présentons on considère, à l'inverse, qu'un agent raisonne comme si les actions dont il n'a pas connaissance n'avaient pas eu lieu. Ceci revient pour lui à ignorer ces actions tant qu'il n'a pas observé leurs effets. C'est une attitude plus confiante, ou crédule.

Dans notre proposition, le fait que la réalisation de l'action $tomber_m$ n'est pas connue de l'enfant est exprimé formellement par le fait que les situations accessibles depuis $do(tomber_m, s)$ par la relation K_e sont **les mêmes** que celles accessibles depuis s par K_e. Cela signifie intuitivement que les situations imaginées par l'enfant sont les mêmes en s et en $do(tomber_m, s)$, et donc que ses croyances sont inchangées après la réalisation de $tomber_m$.
Formellement on a :
$\forall s \forall s'' \forall a(a = tomber_m \rightarrow (K_e(s'', do(a, s)) \leftrightarrow K_e(s'', s)))$.

Pour les actions dont la réalisation est connue de l'enfant l'évolution des croyances est définie de la façon suivante :

$\forall s \forall s'' \forall a (\neg(a = tomber_m) \rightarrow$
$(K_e(s'', do(a, s)) \leftrightarrow \exists s'(K_e(s', s) \wedge s'' = do(a, s') \wedge ($
$\quad a = tomber_e \vee a = réparer_m \vee$
$\quad a = observer_e \wedge (cassé(s) \leftrightarrow cassé(s'))))))$

Pour représenter l'évolution des croyances quel que soit le type d'action on a alors :

$\forall s \forall s'' \forall a (K_e(s'', do(a, s)) \leftrightarrow$
$\quad (K_e(s'', s) \wedge \neg(a = tomber_e) \wedge \neg(a = observer_e) \wedge \neg(a = réparer_m)) \vee$
$\quad \exists s'(K_e(s', s) \wedge s'' = do(a, s') \wedge ($
$\quad a = tomber_e \vee a = réparer_m \vee$
$\quad a = observer_e \wedge (cassé(s) \leftrightarrow cassé(s'))))$

On a de la même manière l'axiome suivant pour représenter l'évolution des croyances de la mère :

$\forall s \forall s'' \forall a (K_m(s'', do(a, s)) \leftrightarrow$
$\quad (K_m(s'', s) \wedge \neg(a = tomber_m) \wedge \neg(a = réparer_m) \wedge \neg(a = observer_m) \wedge \neg(a = tomber_e)) \vee$
$\quad \exists s'(K_m(s', s) \wedge s'' = do(a, s') \wedge ($
$\quad a = tomber_m \vee a = réparer_m \vee a = tomber_e$
$\quad a = observer_m \wedge (cassé(s) \leftrightarrow cassé(s'))))$

La forme générale de l'axiome définissant l'évolution d'une relation K_i est :

$(S_{K_i})\ \ \forall s \forall s'' \forall a (K_i(s'', do(a, s)) \leftrightarrow$
$\quad (K_i(s'', s) \wedge \neg(a = \alpha_1) \wedge \ldots \wedge \neg(a = \alpha_n) \wedge \neg(a = \beta_1) \wedge \ldots \wedge \neg(a = \beta_m)) \vee$
$\quad \exists s'(K_i(s', s) \wedge s'' = do(a, s') \wedge$
$\quad (a = \beta_1 \vee \ldots \vee a = \beta_m$
$\quad \vee\ a = \alpha_1 \wedge (\phi_1(s) \leftrightarrow \phi_1(s'))$
$\quad \ldots$
$\quad \vee\ a = \alpha_n \wedge (\phi_n(s) \leftrightarrow \phi_n(s'))))$

où les alphas sont les actions de perception réalisées par l'agent i, et les bétas sont les autres actions dont la réalisation est connue de l'agent i. La forme de cet axiome pourrait être généralisée en définissant l'ensemble des actions connues par l'agent i dans la situation s à l'aide d'une formule $\psi_i(a, s)$, au lieu de donner explicitement ces actions. La formule $\psi_i(a, s)$ pourrait exprimer, par exemple, que a est connue de i en s si l'action a est réalisée par un agent qui est dans la même pièce que i ou qui est dans un voisinage proche.

Pour déterminer l'évolution des croyances il faut aussi donner les axiomes de changement d'état qui s'appliquent aux situations accessibles par K_m ou K_e.

Supposons, par exemple, qu'en S_0 le verre ne soit pas cassé, et que la mère et l'enfant sachent que le verre n'est pas cassé. Soit $\neg cassé(S_0) \wedge Knows_m(\neg cassé, S_0) \wedge Knows_e(\neg cassé, S_0)$. Dans les situations s'_1 et s'_2 respectivement accessibles par K_e et K_m on a donc $\neg cassé(s'_1)$ et $\neg cassé(s'_2)$ (voir figure 2).

Si l'enfant croit que le verre ne se casse pas si on le laisse tomber, et qu'en S_0 il fasse tomber le verre (action $tomber_e$), alors après avoir fait tomber le verre (i.e. dans la situation $do(tomber_e, S_0)$) il croit que le verre n'est pas cassé. Donc dans les situations telles que $do(tomber_e, s'_1)$, qui sont accessibles depuis $do(tomber_e, S_0)$ par K_e, on doit avoir $\neg cassé(do(tomber_e, s'_1))$. Par contre, si la mère croit que le verre est cassé après l'avoir laissé tomber, alors en $do(tomber_e, S_0)$ elle croit que le verre est cassé. Donc, dans les situations telles que $do(tomber_e, s'_2)$, accessibles depuis $do(tomber_e, S_0)$ par K_m, on doit avoir $cassé(do(tomber_e, s'_2))$.

Il faut donc distinguer l'axiome de changement d'état qui s'applique en s'_1 et celui qui s'applique en s'_2. Pour cela on introduit les prédicats $croit_m(s)$ et $croit_e(s)$ qui expriment intuitivement qu'en s les axiomes de changement d'état qui s'appliquent sont ceux qui correspondent respectivement aux croyances de la mère et de l'enfant. Ces axiomes peuvent être, par exemple :

$(S_{cassé,m})$ $\forall s(croit_m(s) \rightarrow (cassé(do(a,s)) \leftrightarrow a = tomber_e \vee a = tomber_m \vee cassé(s) \wedge \neg(a = réparer_m)))$

$(S_{cassé,e})$ $\forall s(croit_e(s) \rightarrow (cassé(do(a,s)) \leftrightarrow cassé(s) \wedge \neg(a = réparer_m)))$

Pour définir les situations où s'appliquent les axiomes de changement d'état correspondant à la réalité on introduit le prédicat $réel(s)$, qui signifie intuitivement que s représente une situation réelle. On a, par exemple :

$(S_{réel})$ $\forall s(réel(s) \rightarrow (cassé(do(a,s)) \leftrightarrow a = tomber_e \vee a = tomber_m \vee cassé(s) \wedge \neg(a = réparer_m)))$

Dans le cas général, pour chaque prédicat p, et pour chaque agent i, ou pour les situations réelles, on doit avoir un axiome de changement d'état de la forme:

$(S_{p,i})$ $\forall s \forall a(croit_i(s) \rightarrow ((p(do(a,s)) \leftrightarrow \Gamma^+_{p,i}(a,s) \vee p(s) \wedge \neg\Gamma^-_{p,i}(a,s))$

$(S_{réel})$ $\forall s \forall a(réel(s) \rightarrow ((p(do(a,s)) \leftrightarrow \Gamma^+_{p,0}(a,s) \vee p(s) \wedge \neg\Gamma^-_{p,0}(a,s))$

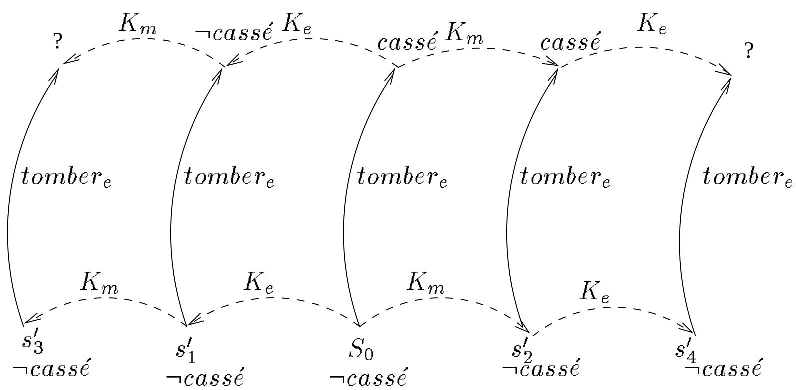

FIG. 2 – *Evolution des croyances respectives des agents.*

Supposons maintenant qu'en S_0 l'enfant croit que la mère croit que le verre n'est pas cassé, et que la mère croit que l'enfant croit lui aussi que le verre n'est pas cassé. Soit $Knows_e(Knows_m(\neg cassé), S_0)$ et $Knows_m(Knows_e$ $(\neg cassé), S_0)$. On a donc $Knows_m(\neg cassé, s'_1)$ et $Knows_e(\neg cassé, s'_2)$.

Soient s'_3 et s'_4 des situations respectivement accessibles depuis s'_1 par K_m et depuis s'_2 par K_e (voit figure 2), la question qui se pose est alors : faut-il considérer s'_3 comme une situation qui représente des croyances de la mère, ou comme une situation qui représente des croyances de l'enfant sur les croyances de la mère?

Si on considère que s'_3 représente une croyance de la mère, soit $croit_m(s'_3)$, alors l'axiome de changement d'état qui s'applique en s'_3 est $(S_{cassé,m})$. Mais ceci revient à supposer que l'enfant croit que sa mère croit que si on laisse tomber un verre il se casse, bien que lui croit le contraire. Cela pourrait être le cas, mais c'est peu vraisemblable. Il faut donc pouvoir éventuellement représenter que ce n'est pas le cas, c'est-à-dire qu'en s'_3 c'est $(S_{cassé,e})$ qui s'applique, et qu'on a $croit_e(s'_3)$.

Si on considère maintenant que s'_4 représente les croyances de la mère sur les croyances de l'enfant, on peut accepter que la mère sache que l'enfant croit que le verre ne se casse pas s'il tombe, et dans ce cas on acceptera d'appliquer en s'_4 l'axiome $(S_{cassé,e})$.

D'un point de vue pratique, comment peut-on exprimer la nature des situations? On ne peut pas, pour chaque situation s, donner l'information $croit_i(s)$, car il y a une infinité de situations. Il paraît naturel cependant de considérer que les agents croient que les effets des actions sont toujours les mêmes, c'est-à-dire que les situations qui succèdent à une situation s sont de même nature que celle-ci. Ceci s'exprime formellement par l'axiome :

(CR_1) $\forall s(croit_i(s) \leftrightarrow croit_i(do(a,s)))$

Mais même en se limitant aux situations accessibles depuis la situation initiale, il en reste une infinité. On propose alors d'attribuer aux agents deux attitudes possibles qui peuvent être décrites par un nombre très limité de faits de la forme $croit_i(s)$.

La première suppose que chaque agent i sait comment évoluent les croyances d'un autre agent j. Par exemple, ça peut être le cas de la mère vis-à-vis de l'enfant. Cette attitude s'exprime formellement par le fait que la nature d'une situation est déterminée uniquement par la relation d'accessibilité qui accède à cette situation. Par exemple, on a $croit_m(s'_2)$ parce que s'_2 est accessible par K_m, et $croit_e(s'_4)$ parce que s'_4 est accessible par K_e. Dans le cas général on a alors le schéma d'axiome :

(CR_2) $\forall s \forall s'(K_i(s',s) \rightarrow croit_i(s'))$

On notera que l'axiome $(CR2)$, même s'il ne résout pas de manière générale les problèmes des connaissances mutuelles mentionnés par Fagin et al. dans [4], il permet de déduire des formules de la forme :
$Knows_{i_1}(\ldots(Knows_{i_p}(croit_{i_p}))\ldots,s)$.
En effet, de $\forall s \forall s'_p(K_{i_p}(s'_p,s) \rightarrow croit_{i_p}(s'_p))$ on peut déduire :
$\forall s \forall s'_1 \ldots \forall s'_p(K_{i_1}(s'_1,s'_2) \wedge \ldots \wedge K_{i_{p-1}}(s'_{p-1},s'_p) \wedge K_{i_p}(s'_p,s) \rightarrow croit_{i_p}(s'_p))$.

La deuxième approche suppose que chaque agent i croit que les croyances des autres agents évoluent comme ses propres croyances. Par exemple, ça peut être, le cas de l'enfant qui croit que sa mère croit, comme lui, qu'un verre ne se casse pas si on le laisse tomber. Dans ce cas la nature d'une situation sera déterminée par l'"agent réel" qui imagine les croyances des autres agents. Par "agent réel" on veut dire celui qui est dans une situation réelle et non dans une situation imaginée par un autre agent. Dans ce cas on a, par exemple, $croit_e(s'_1)$ parce que s'_1 est accessible par K_e depuis S_0 et qu'on a $réel(S_0)$, et on a $croit_e(s'_3)$, bien que la relation qui accède à s'_3 soit K_m, parce que s'_3 est accessible depuis s'_1 et qu'on a $croit_e(s'_1)$. Dans le cas général on a alors les schémas d'axiomes :

(CR_3) $\forall s \forall s'(réel(s) \wedge K_i(s',s) \rightarrow croit_i(s'))$

(CR_4) $\forall s \forall s'(croit_i(s) \wedge K_j(s',s) \rightarrow croit_i(s'))$

Dans le cas où le concepteur qui doit formaliser une application connaît

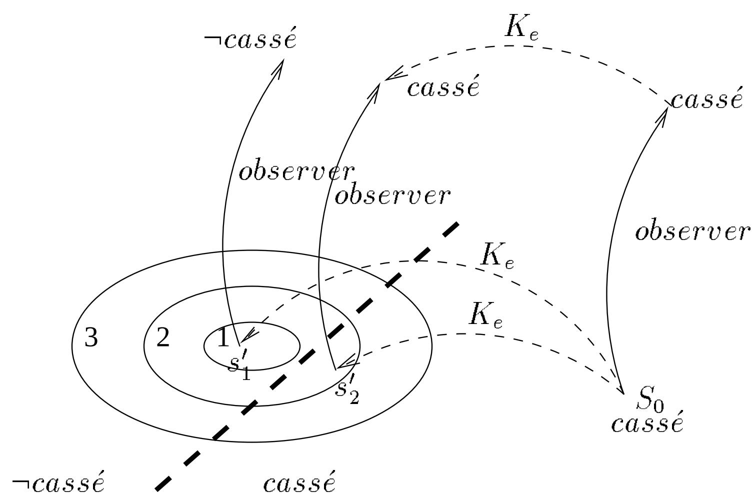

FIG. 3 – *Evolution de croyances ayant différents degrés de plausibilité.*

l'évolution des croyances des agents, mais ne connaît pas les croyances qu'ont les agents sur l'évolution des croyances des autres agents, il peut faire l'hypothèse, par défaut, que les agents ont la deuxième attitude ci-dessus. Il se peut aussi que les agents eux-mêmes adoptent cette attitude, par défaut, quand ils ne connaissent pas l'évolution des croyances des autres agents. Par exemple, si l'enfant ignore comment évoluent les croyances de sa mère.

Que l'on choisisse l'une ou l'autre de ces attitudes, il ne doit y avoir qu'un seul axiome de changement d'état par prédicat qui s'applique à une situation donnée. On doit donc avoir les schémas d'axiomes suivants pour i différent de j:

(CR_5) $\neg\exists s(croit_i(s) \wedge croit_j(s))$
(CR_6) $\neg\exists s(réel(s) \wedge croit_j(s))$

4 RÉVISION DES CROYANCES

La formalisation proposée dans [10] et [6] pose un problème pour la révision des croyances. Considérons l'exemple décrit dans la section 2. Si en

S_0 l'agent croit que le verre n'est pas cassé, soit $Knows(\neg cassé, S_0)$, on a $\forall s'(K(s', S_0) \rightarrow \neg cassé(s'))$, et comme on a $cassé(S_0)$, on a $\neg\exists s'(K(s', S_0) \wedge (cassé(S_0) \leftrightarrow cassé(s'))$, et d'après (S_K) il n'y a aucune situation accessible depuis $do(observer, S_0)$ par K. La conséquence est que pour n'importe quelle formule ϕ on a $Knows(\phi, do(observer, S_0))$ et $Knows(\neg\phi, do(observer, S_0))$. Ce qui évidemment n'est pas acceptable.

Dans [11] Shapiro, Pagnucco, Lespérance et Levesque ont proposé une nouvelle version de cette formalisation dans laquelle on n'a plus ce problème. L'idée est d'assigner à chaque situation s un degré de plausibilité défini par la fonction $pl(s)$. On définit d'abord une relation d'accessibilité K_{max} qui accède aux situations les plus plausibles (celles pour lesquelles $pl(s')$ est minimal) $K_{max}(s', s) \stackrel{\text{def}}{=} K(s', s) \wedge \forall s''(K(s'', s) \rightarrow pl(s') \leq pl(s''))$. On peut alors définir une nouvelle modalité de croyance $Bel(\phi, s)$ avec K_{max} :

$Bel(\phi, s) \stackrel{\text{def}}{=} \forall s'(K_{max}(s', s) \rightarrow \phi[s'])$

De plus on suppose que toutes les situations qui succèdent à une situation donnée ont la même plausibilité, soit : $\forall s \forall a(pl(do(a, s)) \leftrightarrow pl(s))$.

Si on considère l'exemple décrit sur la figure 3, on a $Bel(\neg cassé, S_0)$ car dans toutes les situations accessibles de S_0 où $pl(s')$ est minimal on a $\neg cassé(s')$. Après avoir réalisé l'action $observer$, comme on a $cassé(S_0)$, les seules situations s' qui ont un successeur sont celles pour lesquelles on a $cassé(s')$. Dans l'exemple de la figure 3, pour ces situations s' la valeur minimale de la plausibilité est 2, et pour les situations qui leur succèdent elle reste égale à 2. Donc, pour toutes les situations accessibles par K_{max} depuis $do(observer, S_0)$ on a $cassé(do(observer, s'))$, et on a $Bel(cassé, do(observer, S_0))$, comme souhaité.

Par contre, si dans toutes les situations s' telles que $K(s', S_0)$ on a $\neg cassé(s')$, alors il n'y a pas de situation accessible depuis $do(observer, S_0)$, et l'agent a des croyances inconsistantes, comme on l'a vu dans la section 2. Pour que ce cas ne se produise pas il faut rajouter des conditions supplémentaires (voir Théorème 3 dans [11]).

Il faut noter aussi que dans cette formalisation on peut résoudre le problème de la révision, mais le frame problem n'a pas de solution évidente. Supposons, par exemple, qu'en S_0 l'agent croit que le verre n'est pas cassé et qu'il est froid, soit $Bel(\neg cassé, S_0)$ et $Bel(froid, S_0)$, et que parmi les situations s' les plus plausibles ($pl(s')$ minimal) où on a $cassé(s')$ on ait $froid(s'_2)$ pour l'une d'entre elles, et $\neg froid(s'_3)$ pour une autre. Dans ce cas on a $froid(do(observer, s'_2))$ et $\neg froid(do(observer, s'_3))$, et les situations $do(observer, s'_2)$ et $do(observer, s'_3)$ font partie des situations les plus plausibles accessibles depuis $do(observer, S_0)$. Donc on a $\neg Bel(froid, do(observer, S_0))$ et $\neg Bel(\neg froid, do(observer, S_0))$. Ce qui montre que la croyance $Bel(froid, S_0)$ n'est pas persistante après avoir réalisé l'action $observer$, alors qu'elle devrait l'être, car cette action, intuitivement, ne modifie pas la croyance de l'agent dans le fait que le verre est froid.

L'extension du formalisme que nous venons de voir peut être appliquée directement à la formalisation présentée dans la section 3. Il suffit d'introduire autant de fonctions $pl_i(s)$ qu'il y a d'agents i, et de définir K_{max_i} par

$K_{max_i}(s', s) \stackrel{\text{def}}{=} K_i(s', s) \wedge \forall s''(K_i(s'', s) \rightarrow pl_i(s') \leq pl_i(s''))$

La modalité $Bel_i(\phi, s)$ est alors définie par :

$Bel_i(\phi, s) \stackrel{\text{def}}{=} \forall s'(K_{max_i}(s', s) \rightarrow \phi[s'])$

5 Conclusion

Nous avons vu que le Calcul des Situations permet de formaliser une solution très intéressante du frame problem qui s'applique à l'évolution du monde et des croyances.

Nous avons montré que cependant la définition de l'évolution des croyances présentée dans [10] et [6] suppose qu'un agent connaisse toutes les actions qui sont réalisées et sache comment les choses évoluent dans la réalité. Ces deux hypothèses sont trop fortes dans le cas où on considère plusieurs agents.

Dans la formalisation que nous avons présentée on n'est pas obligé de faire ces hypothèses. En effet, dans les axiomes tels que (S_{K_i}) on distingue les actions dont la réalisation est connue de l'agent i et les autres. D'autre part il se peut que les agents aient des croyances différentes sur l'évolution du monde. Ceci est formalisé à l'aide des prédicats $croit_i(s)$ et $réel(s)$ et en donnant des axiomes de changement d'état $(S_{p,i})$ correspondant aux croyances de chaque agent i. Nous avons vu que l'extension des prédicats $croit_i(s)$ peut être définie à l'aide des schémas d'axiomes (CR_1), (CR_2) (ou (CR_3) et (CR_4)), (CR_5) et (CR_6). Enfin, la formalisation de la révision des croyances proposée par Levesque dans [11] se transpose facilement dans la solution que nous avons proposée.

Remerciements. Nous remercions Yves Lespérance pour ses nombreux commentaires et suggestions qui ont contribué à améliorer la qualité de cet article.

RÉFÉRENCES

[1] F. M. Brown, éditeur. *The Frame Problem in Artificial Intelligence*. Morgan Kaufmann Publishers, 1987.

[2] B. F. Chellas. *Modal Logic: An introduction*. Cambridge University Press, 1988.

[3] R. Demolombe et M. P. Pozos Parra. A simple and tractable extension of situation calculus to epistemic logic. In Z. W. Ras et S. Ohsuga, éditeurs, *Proc. of 12th Internatioonal Symposium ISMIS 2000*. Springer. LNAI 1932, 2000.

[4] R. Fagin, J.Y. Halpern, Y. Moses et M.Y. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.

[5] N. Friedman et J.Y. Halpern. A knowledge based framework for belief change, part ii: Revision and update. In *Proc. of the 4th International Conference on Principle of Knowledge Representation and Reasoning (KR-94)*, pages 190–201, 1994.

[6] G. Lakemeyer et H. Levesque. AOL: a logic of acting, sensing, knowing and only knowing. In *Proc. of the 6th Int. Conf. on Principles of Knowledge Representation and Reasoning*, pages 316–327, 1998.

[7] Y. Lespérance, H. Levesque et R. Reiter. A situation calculus approach to modeling and programming agents. In A. Rao et M. Wooldbridge, éditeurs, *Foundations and Theories of Rational Agents*. Kluwer, 1999.

[8] R. Reiter. The frame problem in the situation calculus: a simple solution (sometimes) and a completeness result for goal regression. In V. Lifschitz, éditeur, *Artificial Intelligence and Mathematical Theory of Computation: Papers in Honor of John McCarthy*, pages 359–380. Academic Press, 1991.

[9] R. Reiter. Knowledge in Action: Logical Foundations for Describing and Implementing Dynamical Systems. Technical report, University of Toronto, 1999.

[10] R. Scherl et H. Levesque. The Frame Problem and Knowledge Producing Actions. In *Proc. of the National Conference of Artificial Intelligence.* AAAI Press, 1993.

[11] S. Shapiro, M. Pagnuco, Y. Lespérance et H. Levesque. Iterated belief change in the situation calculus. In *Proc. of the 7th Conference on Principles on Knowledge Representation and Reasoning (KR2000)*. Morgan Kaufman Publishers, 2000.