HAL open science

Modélisation logique d'agents rationnels pour l'intelligence ambiante

Carole Adam, Fabrice Evrard, Benoit Gaudou, Andreas Herzig, Dominique Longin

To cite this version:

Carole Adam, Fabrice Evrard, Benoit Gaudou, Andreas Herzig, Dominique Longin. Modélisation logique d'agents rationnels pour l'intelligence ambiante. 14èmes Journées Francophones sur les Systèmes Multi-Agents (JFSMA 2006), Oct 2006, Annecy, France. pp.81-94. hal-03537168

HAL Id: hal-03537168

https://hal.science/hal-03537168v1

Submitted on 20 Jan 2022

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.

Modélisation logique d'agents rationnels pour l'intelligence ambiante

Carole Adam — Fabrice Évrard — Benoit Gaudou
Andreas Herzig — Dominique Longin

Institut de Recherche en Informatique de Toulouse
IRIT (UMR 5505) - Équipe LILAC
118 route de Narbonne, 31062 Toulouse cedex 9
{adam,gaudou,herzig,longin}@irit.fr

RÉSUMÉ. *Au sein de la communauté agent, les aspects émotionnels reçoivent de plus en plus d'attention depuis qu'il a été prouvé qu'elles sont essentielles pour des agents intelligents. En effet, d'un point de vue théorique, des résultats de psychologie et neurosciences cognitives ont établi les liens étroits qui existent chez les humains entre émotions et raisonnement (voir les travaux de Damasio) ou prise de décision. Et d'un point de vue pratique, de nombreux travaux ont montré l'intérêt des émotions chez les agents devant communiquer avec des humains : agents d'interface, tuteurs virtuels ... Pourtant, parmi les cadres logiques utilisés pour formaliser des* agents rationnels, *très peu intègrent ces aspects émotionnels. Dans cet article, nous caractérisons certaines émotions, telles que définies en psychologie cognitive par Lazarus, dans une logique modale BDI (Belief, Desire, Intention). Nous validons ensuite notre cadre formel grâce à l'étude d'un exemple issu de la problématique de l'Intelligence Ambiante.*

ABSTRACT. *In the agent community, emotional aspects receive more and more attention since they were proven to be essential for intelligent agents. Indeed, from a theoretical point of view, results from cognitive psychology and neuroscience have established the close links that exist in humans between emotions and reasoning (see Damasio's works) or decision making. And from a practical point of view, numerous research findings show the interest of emotions in agents communicating with humans : interface agents, pedagogical agents, ... However, among the logical frameworks used to formalize these* rational agents, *very few integrate these emotional aspects. In this paper, we characterize some emotions, as defined in cognitive psychology by Lazarus, in a BDI (Belief, Desire, Intention) modal logic. We then validate our framework with a case study illustrating the problematic of Ambient Intelligence.*

MOTS-CLÉS : *logique modale, intelligence ambiante, émotion*

KEYWORDS: *modal logic, ambient intelligence, emotion*

1. Introduction

L'*intelligence ambiante* désigne un ensemble d'interfaces intelligentes supportées par des technologies de réseau et de traitement des données enfouies dans les objets du quotidien comme le mobilier, les vêtements, les véhicules, les routes ou même les matériaux intelligents. Son but est d'être attentive aux caractéristiques spécifiques de chacun, de s'adapter aux besoins des utilisateurs, d'être capable de répondre intelligemment à des indications parlées ou gestuelles, et même d'engager un dialogue. Elle doit être non intrusive et le plus souvent invisible pour l'utilisateur au quotidien.

Dans ce qui suit, un *système d'intelligence ambiante* (SIA) désigne un (ensemble d')agent(s) dédié(s) à ces tâches. Dans ce cadre, nous pensons que l'émotion est un élément contextuel essentiel que le SIA doit prendre en compte (*i.e.* percevoir, analyser, générer) afin de répondre au mieux aux attentes de ses utilisateurs. Le SIA a donc besoin d'un modèle des émotions dans les cas suivants : *(C1)* pour calculer l'émotion générée chez l'utilisateur par un événement extérieur ; *(C2)* pour anticiper l'effet d'une de ses actions sur l'utilisateur et choisir ainsi l'action la plus appropriée ; *(C3)* pour comprendre les causes d'une émotion, détectée chez l'utilisateur en observant son comportement, en inférant *(C3b)* ou non *(C3a)* des hypothèses sur les croyances de l'utilisateur. Connaître l'émotion ressentie par l'utilisateur et les causes de cette émotion est primordial pour agir de manière réellement adaptée.

Dans cet article, notre but est donc de proposer un cadre formel pour des agents rationnels capables de manipuler un certain nombre d'émotions (c.-à-d. des *agents émotionnels*) destinés à s'intégrer dans un SIA. Ce cadre est celui des logiques dites BDI (pour *belief*, *desire*, *intention*). Ces logiques, fondées sur la philosophie du langage, des états mentaux, et de l'action, proposent de modéliser des agents *via* un certain nombre de concepts clés tels l'action et les *attitudes mentales* (croyances, buts, intentions, obligations, choix ...).

L'intérêt de ce cadre formel est qu'il est largement répandu dans la communauté agent internationale, et que ses propriétés sont maintenant bien connues : fort pouvoir explicatif du comportement de l'agent, vérifiabilité formelle, cadre théorique rigoureux et bien établi (tant du point de vue philosophique que de celui de la logique formelle).

Notre démarche consiste donc à étendre de façon minimale une logique BDI afin de pouvoir gérer des émotions. Nous montrons dans ce qui suit (cf. Section 3) que cette extension passe par la définition de deux opérateurs capturant ce qu'un agent (n')aime (pas)[1], et qu'à partir de ces deux opérateurs, des attitudes mentales traditionnelles (croyance, choix), et de l'action et du temps, nous pouvons définir un certain nombre d'émotions (cf. Section 4) qui ont été identifiées en psychologie (cf. Section 2). Nous illustrons le fonctionnement de notre logique par un exemple de SIA où un agent *maison* va prendre en compte et gérer les émotions d'un des ses habitants, éventuellement avec l'aide d'autres agents du SIA (cf. Section 5).

1. Comme nous le montrons par la suite, nous utilisons deux opérateurs plutôt qu'un car ces notions sont bipolaires plutôt que complémentaires ou duales.

2. Analyses de l'émotion

2.1. *Représentation des émotions*

La psychologie propose trois types de modèles explicatifs des émotions : *les modèles dimensionnels* les représentent avec trois dimensions, qui sont généralement la valence (*valence*), l'excitation (*arousal*), et la puissance (*stance*) [RUS 97] ; *les modèles discrets* considèrent les émotions comme des mécanismes adaptatifs basiques et universels, conçus au cours de l'évolution pour favoriser la survie ([DAR 72], [EKM 92]) ; *les modèles cognitifs* soutiennent que les émotions sont déclenchées par un processus cognitif évaluant la relation entre l'individu et son environnement [LAZ 91] : l'individu évalue en permanence, consciemment ou non, l'impact des stimuli sur ses buts et désirs (internal needs ...). Alors que les deux premiers types de modèles sont principalement descriptifs, le dernier est normatif et est donc plus adapté à notre but d'expliquer comment émergent les émotions de l'utilisateur.

Ortony et al. [ORT 88] ont proposé une typologie des émotions (la typologie OCC, ci-dessous) qui est basée sur la théorie de l'évaluation cognitive de Lazarus [LAZ 91]. Ils considèrent trois types de stimuli physiques : les événements, les actions d'agents, et les objets, qui peuvent être évalués suivant différents critères comme l'agréabilité, l'attribution causale, ou la probabilité. Cela leur permet de distinguer vingt-deux émotions qu'ils regroupent dans plusieurs catégories. Cette typologie a été très souvent utilisée en informatique pour développer des agents virtuels émotionnels [GRA 04].

2.2. *Détermination des émotions d'un humain*

Les développeurs d'agents ont étudié différentes méthodes pour déterminer l'émotion ressentie par l'utilisateur [2]. Prendinger et Ishizuka [PRE 05] surveillent ses signaux physiologiques et la direction de son regard, et utilisent des travaux existants [PIC 97, LEV 03] pour associer un label d'émotion à ces informations. Cette méthode permet de détecter les moindres changements de l'émotion du sujet, mais est assez intrusive, et désobéit donc à un des principes de l'intelligence ambiante.

Une autre méthode est explorée par Jaques et al. [JAQ 04]. Leur agent pédagogique déduit l'émotion de son élève en évaluant les événements depuis son point de vue (grâce à un modèle de l'utilisateur), via une fonction d'évaluation basée sur la typologie OCC. Cette méthode fournit seulement une spéculation sur l'état émotionnel de l'utilisateur, mais elle est plutôt efficace associée à un bon modèle de ses attitudes mentales, et surtout, elle n'est pas du tout intrusive. De plus, elle permet aussi à l'agent d'anticiper la réaction émotionnelle de l'utilisateur à ses actions, et ainsi d'essayer d'induire une émotion particulière chez lui par le choix d'une action adaptée. C'est une caractéristique très importante, car il a été prouvé que l'émotion influence tous les

2. dans le cas où il ne l'exprime pas directement par des mimiques ou des gestes. Sinon des connaissances comportementales (cf. Section 3) permettent de la déduire.

aspects du raisonnement humain (Damasio [DAM 94]), par exemple l'apprentissage [PIA 89]. Finalement, cette seconde méthode est plus adaptée à la problématique de l'intelligence ambiante.

Dans la section suivante nous présentons notre cadre qui est une extension des logiques BDI, et qui doit permettre de représenter les attitudes mentales et les émotions de l'utilisateur.

3. Architecture logique d'agent émotionnel

La théorie logique initiale T_0 de l'agent est constituée : de connaissances initiales pouvant évoluer au cours du temps en fonction d'actions d'observation de l'agent ou d'actions d'information des autres agents du SIA ; et d'axiomes non-logiques globaux *i.e.* persistant dans tous les mondes. Les connaissances initiales comprennent des connaissances factuelles (CF), *i.e.* connaissances sur les faits du monde (par exemple : il fait beau) ; et des connaissances épistémiques (CE), *i.e.* des connaissances sur les états mentaux des autres agents, humains ou pas (par exemple : l'utilisateur ne sait pas s'il pleut dehors). Les axiomes globaux comprennent des connaissances des lois du monde (CLM) (par exemple si un verre tombe par terre il se casse), et des connaissances comportementales (CC) (par exemple si l'utilisateur claque la porte c'est qu'il est en colère).

Le modèle de l'utilisateur MU est un sous-ensemble de $CF \cup CE \cup CC$.

Dans cette partie nous présentons notre logique qui est basée sur celle développée par Herzig et Longin [HER 04], elle-même étant une refonte de celle de Cohen et Levesque [COH 90]. Nous y ajoutons notamment l'opérateur de probabilité défini par Herzig dans [HER 03].

Notons AGT l'ensemble des agents, ACT l'ensemble des actions, et $ATM = \{p, q...\}$ l'ensemble des formules atomiques. Les formules complexes sont notées A, B... Dans ce qui suit nous présentons différents axiomes de notre logique[3].

La croyance. La croyance caractérise ce que l'agent, de manière privée, pense être vrai (ce qui n'impose pas que cela soit effectivement vrai). On note $Bel_i\, A$ pour signifier « l'agent i croit que A est vrai ».

De manière standard, l'opérateur de croyance est utilisé dans une logique modale dite normale (KD45) [CHE 80]. Les logiques modales normales sont caractérisées par exemple par l'axiome et la règle d'inférence suivants, sous lesquels l'ensemble des croyances de l'agent est fermé.

$$\text{si } A \text{ alors } Bel_i\, A \qquad (\mathrm{RN}_{Bel_i})$$

$$Bel_i\, A \wedge Bel_i\, (A \rightarrow B) \rightarrow Bel_i\, B \qquad (\mathrm{K}_{Bel_i})$$

3. Pour des raisons de place, seule les axiomes seront présentés en détails. La sémantique ne sera qu'évoquée en fin de section ; pour plus de détails se référer à [COH 90, HER 03, HER 04].

Pour une logique de type KD45, nous avons aussi les axiomes suivants, qui caractérisent le fait que les croyances d'un agent sont consistantes (D_{Bel_i}), et que l'agent est conscient de ce qu'il croit (4_{Bel_i}) et de ce qu'il ignore (5_{Bel_i}).

$$Bel_i\, A \rightarrow \neg Bel_i\, \neg A \qquad (D_{Bel_i})$$

$$Bel_i\, A \rightarrow Bel_i\, Bel_i\, A \qquad (4_{Bel_i})$$

$$\neg Bel_i\, A \rightarrow Bel_i\, \neg Bel_i\, A \qquad (5_{Bel_i})$$

La Probabilité. Pour modéliser les émotions, nous aurons besoin d'une notion de croyance plus faible que la précédente. Nous allons utiliser l'opérateur P_i défini par [HER 03], basé sur la notion de mesure de probabilité subjective. $P_i\, A$ signifie que « pour l'agent i A est plus probable que $\neg A$ ». La logique de P_i est beaucoup plus faible que celle de la croyance, en particulier, elle n'est pas normale : la règle (RN_{Bel_i}) et l'axiome (K_{Bel_i}) n'ont pas de contrepartie en termes de P_i. L'auteur exhibe tout de même un certain nombre de propriétés intéressantes, en particulier la règle d'inférence et les axiomes suivants :

$$\text{si } A \rightarrow B \text{ alors } P_i\, A \rightarrow P_i\, B \qquad (RM_{P_i})$$

$$P_i\, \top \qquad (N_{P_i})$$

$$P_i\, A \rightarrow \neg P_i\, \neg A \qquad (D_{Prob})$$

La croyance et la probabilité sont profondément liées. Citons simplement ici les principaux axiomes de [HER 03] :

$$Bel_i\, A \rightarrow P_i\, A \qquad (1)$$

$$P_i\, A \rightarrow Bel_i\, P_i\, A \qquad (2)$$

$$\neg P_i\, A \rightarrow Bel_i\, \neg P_i\, A \qquad (3)$$

Le choix. Le choix correspond à ce que l'agent préfère : $Choice_i\, A$ signifie : « l'agent i préfère que A soit vrai ». Comme pour l'opérateur de croyance, le choix est défini dans une logique KD45, et obéit donc aux axiomes suivants, signifiant que les choix d'un agent sont consistants (D_{Choice_i}), et que l'agent est en accord avec ses choix (4_{Choice_i}) et non-choix (5_{Choice_i}).

$$Choice_i\, A \rightarrow \neg Choice_i\, \neg A \qquad (D_{Choice_i})$$

$$Choice_i\, A \rightarrow Choice_i\, Choice_i\, A \qquad (4_{Choice_i})$$

$$\neg Choice_i\, A \rightarrow Choice_i\, \neg Choice_i\, A \qquad (5_{Choice_i})$$

À l'instar de Cohen et Levesque [COH 90][4] et Rao et Georgeff [RAO 91], le choix est réaliste, *i.e.* un agent ne peut pas préférer quelque chose qu'il croit impossible, c.-à-d. :

$$Bel_i\ A \rightarrow Choice_i\ A \tag{4}$$

Les agents sont conscients de leurs choix (comme ils l'étaient de leurs croyances) :

$$Choice_i\ A \rightarrow Bel_i\ Choice_i\ A \tag{5}$$

$$\neg Choice_i\ A \rightarrow Bel_i\ \neg Choice_i\ A \tag{6}$$

Les goûts. $Like$ représente une préférence déconnectée de la réalité. $Like_i\ A$ (resp. $Dislike_i\ A$) signifie : « l'agent i aime (resp. déteste) que A ». Ces opérateurs se rapprochent beaucoup de ceux de choix et de désir. Ils diffèrent cependant du choix car ils sont déconnectés de la réalité : $Like_i\ A \wedge Bel_i\ \neg A$ est satisfiable contrairement à $Choice_i\ A \wedge beli \neg A$. Ils diffèrent également du désir car ils sont bipolaires : une formule A peut plaire à l'agent i ($Like_i\ A$), lui déplaire ($Dislike_i\ A$) ou lui être indifférente.

Pour simplifier, nous nous plaçons dans une logique modale standard de type KD pour les opérateurs $Like$ et $Dislike$ [5]. Remarquons qu'il n'existe pas de forte relation de dépendance entre $Like$ et $Dislike$; donc **ni** $Like_i\ A \rightarrow Dislike_i\ \neg A$, **ni** son inverse $Dislike_i\ A \rightarrow Like_i\ \neg A$ ne sont valides. Comme pour les autres attitudes mentales, nous faisons les hypothèses d'introspection positive et négative pour $Like$ ((IP$_{Like}$) et (IN$_{Like}$)) et pour $Dislike$:

$$Like_i\ A \rightarrow Bel_i\ Like_i\ A \tag{IP$_{Like}$}$$

$$\neg Like_i\ A \rightarrow Bel_i\ \neg Like_i\ A \tag{IN$_{Like}$}$$

L'action. Les opérateurs dynamiques $After_\alpha\ A$ et $Before_\alpha\ A$ signifient que « après (resp. avant) toute exécution de α A est vraie ». Ils sont définis dans la logique temporelle standard K^t, *i.e.* ils sont dans une logique modale normale avec les axiomes de conversion :

$$A \rightarrow After_\alpha\ \neg Before_\alpha\ \neg A \tag{7}$$

$$A \rightarrow Before_\alpha\ \neg After_\alpha\ \neg A \tag{8}$$

4. Comme Cohen et Levesque, dans [HER 04], Herzig et Longin utilise la croyance et le choix pour définir une notion d'intention non-primitive (contrairement à [HER 02a] où le choix n'est pas défini dans la logique et l'intention est primitive).

5. Cela implique notamment que $Like_i\ A$ pour toute tautologie A, ce qui semble contre-intuitif. Mais nous pouvons nous contenter de cette logique simple sans conséquences sur la définition des émotions ; en effet si on définit $Like'_i A \stackrel{def}{=} Like_i\ A \wedge \neg Dislike_i\ A$, alors $Like'_i \top$ n'est pas tautologique, et c'est ce $Like'_i$ que nous utiliserons dans les définitions des émotions.

Le temps. Nous n'utilisons pas ici la logique du temps définie dans [HER 04], mais conservons une logique du temps linéaire avec les opérateurs G (GA signifie que « dorénavant A sera vraie ») et H (HA signifie que « jusqu'à maintenant A était vraie »), définis dans une logique KT4 avec confluence et axiomes de conversion (l'axiome (T_G) spécifiant en particulier que l'instant présent est inclus dans le futur) :

$$GA \rightarrow A \tag{T_G}$$

$$GA \rightarrow GGA \tag{4_G}$$

$$\neg G \neg GA \rightarrow G \neg G \neg A \tag{G_G}$$

$$A \rightarrow G \neg H \neg A \tag{9}$$

$$A \rightarrow H \neg G \neg A \tag{10}$$

Nous y ajoutons l'opérateur « next » X (XA signifie qu'à l'instant suivant A sera vraie) et son inverse X^{-1} ($X^{-1}A$ signifie qu'à l'instant précédent A était vraie), définis dans une logique KD avec les axiomes de conversion correspondants.

Sémantique. Nous utilisons une sémantique des mondes possibles. La sémantique est développée dans [HER 04] pour la croyance, le choix et le temps, dans [HER 03] pour la probabilité. Les opérateurs $Like$ et $Dislike$ sont définis dans une logique modale normale, leur sémantique est donc standard.

Le cadre présenté ici n'est toujours pas assez expressif pour calculer toutes les variables d'évaluation utilisées dans la typologie OCC[6], donc nous considérerons uniquement l'évaluation d'événements relativement à leurs conséquences (en termes de nouvelles attitudes mentales) pour l'agent lui-même. Dans la typologie OCC cela concerne trois catégories d'émotions : le bien-être si l'événement est évalué *a posteriori* ; la perspective (basée sur des croyances à propos du futur) si c'est un événement potentiel évalué *a priori* ; la confirmation si cet événement confirme ou infirme une attente / perspective. La section suivante présente les huit émotions situées dans ces catégories.

4. Formalisation d'émotions

Nous définissons dans cette partie deux critères d'évaluation d'OCC puis les émotions qu'on peut caractériser à partir d'eux.

4.1. *Critères d'évaluation*

Le critère d'agréabilité. Il caractérise un stimulus procurant une sorte de bien-être à l'individu, pour différentes raisons, non exhaustivement : parce qu'il l'apprécie,

6. C'est notamment le cas à cause de l'absence des concepts de norme et d'obligation. Nous prévoyons de les intégrer par la suite.

ou parce qu'il participe à la satisfaction d'un de ses choix. Nous définissons donc $Pleasant_i\, A$ et $Unpleasant_i\, A$ de la manière suivante :

$$Pleasant_i\, A \stackrel{def}{=} (Like_i\, A \wedge \neg Dislike_i\, A) \vee (\neg Dislike_i\, A \wedge X^{-1}\, Choice_i\, X\, A \wedge X^{-1} \neg Bel_i\, X\, A)$$

$$Unpleasant_i\, A \stackrel{def}{=} (Dislike_i\, A \wedge \neg Like_i\, A) \vee (\neg Like_i\, A \wedge X^{-1}\, Choice_i\, X \neg A \wedge X^{-1} \neg Bel_i\, X \neg A)$$

A peut donc être agréable d'une part si l'agent aime que A (et ne déteste pas simultanément que A), et d'autre part s'il ne déteste pas que A et préférait à l'instant précédent que A arrive ($X^{-1}\, Choice_i\, X\, A$) sans en être sûr ($X^{-1} \neg Bel_X\, A$).

Étant données ces définitions, on déduit les propriétés suivantes :

$$\neg Pleasant_i\, \top \wedge \neg Unpleasant_i\, \top \wedge \neg Pleasant_i\, \bot \wedge \neg Unpleasant_i\, \bot \tag{11}$$

$$Pleasant_i\, A \rightarrow \neg Unpleasant_i\, A \tag{12}$$

(11) signifie que nous sommes indifférents aux tautologies et aux contradictions (elles ne sont ni plaisantes ni déplaisantes). (12) signifie que ce qui est plaisant ne peut pas être déplaisant, et inversement (ce qui est plutôt intuitif). Mais ces définitions ne permettent de déduire ni $(Pleasant_i\, A \wedge Pleasant_i\, A') \rightarrow Pleasant_i\, (A \wedge A')$ ni son inverse : A and A' peuvent être plaisantes toutes les deux mais pas quand elles sont associées ; l'exemple classique étant qu'il peut être plaisant d'être marié à Anne, et être plaisant d'être marié à Betty, sans qu'il soit plaisant d'être bigame.

Le critère de probabilité. Il caractérise un stimulus que l'individu considère comme probable. $Expect_i\, A$ signifie « l'agent i croit que A est plus probable que $\neg A$ » :

$$Expect_i\, A \stackrel{def}{=} P_i\, A \tag{Def$_{Expect_i}$}$$

et $Envisage_i\, A$ signifie « l'agent i n'est pas certain que A est fausse » :

$$Envisage_i\, A \stackrel{def}{=} \neg Bel_i\, \neg A \tag{Def$_{Envisage_i}$}$$

On remarque que $Expect_i\, A \rightarrow Envisage_i\, A$, ce qui est assez intuitif.

Dans le reste de cette section, nous spécifions les conditions d'existence des huit émotions que nous voulons caractériser.

4.2. *Conditions d'existence des émotions*

Dans la typologie OCC, les émotions résultent de l'occurrence de trois types de stimuli (événements, actions, et objets : ici nous considérons uniquement les événements), qui dans notre cadre changent les attitudes mentales de l'agent et peuvent ainsi valider des conditions particulières. Les émotions sont donc des abréviations du langage, équivalentes à leurs conditions d'existence.

Joie / Tristesse. La joie (resp. la tristesse) est l'émotion ressentie par un individu quand il se produit un événement plaisant (resp. déplaisant) pour lui.

– $Joy_i A \stackrel{def}{=} Bel_i\, A \wedge Pleasant_i\, A$

– $Sadness_i A \stackrel{def}{=} Bel_i\, A \wedge Unpleasant_i\, A$

Espoir / Crainte. Un individu ressent de l'espoir (resp. de la crainte) quand il s'attend à ce qu'un certain événement se produise, mais qu'il envisage que ce soit l'événement contraire qui arrive, et que ce second événement est plaisant (resp. déplaisant) pour lui.

– $Hope_i A \stackrel{def}{=} Expect_i\, \neg A \wedge Pleasant_i\, A \wedge Envisage_i\, A$

– $Fear_i A \stackrel{def}{=} Expect_i\, \neg A \wedge Unpleasant_i\, A \wedge Envisage_i\, A$

Satisfaction / Confirmation de crainte / Soulagement / Déception. Ces quatre émotions sont déclenchées quand un événement se produit qui confirme ou infirme un espoir ou une peur passé.

– $Satisfaction_i A \stackrel{def}{=} Bel_i\, A \wedge X^{-1} Hope_i A$

– $Disappointment_i \neg A \stackrel{def}{=} Bel_i\, \neg A \wedge X^{-1} Hope_i A$

– $Relief_i \neg A \stackrel{def}{=} Bel_i\, \neg A \wedge X^{-1} Fear_i A$

– $FearConfirmed_i A \stackrel{def}{=} Bel_i\, A \wedge X^{-1} Fear_i A$

On remarque que la satisfaction implique la joie, et la confirmation de crainte implique la tristesse. Ces implications semblent intuitivement correctes.

5. Étude de cas

Nous voulons à présent appliquer notre cadre logique à quatre scénarios différents correspondant aux différents cas (C1) à (C3b) évoqués en introduction où l'agent a besoin d'émotions. Dans tous les cas, nous considérons un SIA de gestion intelligente d'une maison administrée par l'agent m. Nous notons h un habitant de cette maison.

Cas (C1) : évaluation d'un événement extérieur du point de vue de l'utilisateur

Par définition, dès que l'agent m croit que l'état mental de h vérifie les conditions entrant dans la composition d'une certaine émotion, alors *de facto* m croit que h ressent l'émotion en question. Ainsi, si m croit que h croit qu'il fait beau (*i.e.* $Bel_m\, Bel_h\, beau$) et qu'il croit également que c'est plaisant pour h ($Bel_m\, Pleasant_h\, beau$) alors par définition il croit que h est joyeux qu'il fasse beau (*i.e.* $Bel_h\, Joy_m beau$)[7].

7. Ici, l'émotion considérée est positive. Le but de m peut alors soit être de la maintenir, soit de la prendre en compte dans certaines situations particulières (par exemple s'il a une mauvaise nouvelle à annoncer à h).

Cas (C2) : pré-évaluation de l'effet émotionnel d'une action de l'agent sur l'utilisateur

Dans certains cas, un agent ayant une action à accomplir peut tirer parti de son impact émotionnel potentiel. Cela peut être le cas, par exemple, quand la production ou la suppression d'une certaine émotion chez le destinataire de l'action (physique ou linguistique[8]) constitue l'effet visé (communément appelé *effet rationnel* dans la communauté agent [FIP 02]), ou quand différentes actions ayant le même effet (informatif ou physique) ont des effets émotionnels différents (ceux-ci constituent donc le critère de sélection de l'action parmi celles possibles).

Pour illustrer le premier cas, supposons que m sache que h ressent une émotion négative (par exemple de la tristesse) car il pleut (et qu'il ne peut pas aller se promener), soit $Bel_m\, Sadness_h\, ilPleut$. Certaines lois comportementales peuvent pousser m à aider h à mettre en place des stratégies de *coping*[9] soit en informant h qu'il ne pleut plus dès que m l'apprend, soit en focalisant l'attention de h sur autre chose[10]. Dans le premier cas, une telle loi serait de la forme : $Bel_m\, (Sadness_h \varphi \wedge \neg\varphi) \rightarrow Intend_m\, Bel_h\, \neg\varphi$ (en d'autres termes : si m croit que h est triste à propos de φ alors que lui-même sait que φ est faux, il va adopter l'intention d'en informer h).

Concernant le second cas, supposons que m sache que l'instant d'avant h espérait jouer aux échecs avec Jean (soit : $Bel_m\, X^{-1} Hope_h\, echecsJean$), mais que finalement Jean ne vient pas ($Bel_m\, \neg vientJean$). Supposons que m ainsi que tous les agents aient différentes lois du monde de la forme $echecsX \rightarrow vientX$ où $X \in \{Jean, Pierre, Paul\}$ signifiant que si X joue aux échecs avec h alors nécessairement c'est que X vient chez h. Dans ces conditions, si m informe h que Jean ne vient pas, m peut en déduire logiquement que h saura qu'il ne jouera pas aux échecs avec Jean (soit : $Bel_m\, Bel_h\, \neg echecsJean$) ce qui, par définition, va le décevoir (*i.e.* $Disappointment_h \neg echecsJean$). m apprend alors que Pierre et Paul se proposent de remplacer Jean pour jouer aux échecs avec h, elle doit choisir l'adversaire qui plaira le plus à h. m sait que h apprécie que Paul lui rende visite ($Bel_m\, Pleasant_h\, vientPaul$), mais est indifférent à la venue de Pierre ($Bel_m\, \neg Pleasant_h\, vientPaul$). On peut alors démontrer que si m informe h que Paul va jouer aux échecs avec lui alors m croira que h est heureux que Paul vienne chez lui ($Bel_m\, Joy_h\, vientPaul$) (et non pas heureux qu'il joue aux échecs avec lui, ce qui lui est indifférent ; ce qui lui était plaisant était de jouer aux échecs avec Jean). On peut aussi démontrer que si m annonce à h qu'il

8. Nous nous plaçons dans le cadre de la théorie des actes de langage où la production d'un énoncé est vu comme une action à part entière, avec des préconditions d'exécution et des effets (cf. [SEA 69, SEA 85, BRA 87, VAN 90] par exemple).

9. Le coping est le choix par l'agent d'une stratégie dans le but de supprimer ou atténuer une émotion négative qu'il ressent (par exemple minimiser ou inhiber totalement ses causes). On considère ici que le SIA peut aider l'utilisateur dans cette tâche, par exemple en l'informant s'il se trompe sur l'objet d'une émotion négative et qu'il n'a donc pas à la ressentir.

10. Ce dernier cas (que nous ne traitons pas ici) requiert qu'on soit capable de gérer le *degré d'activation* de certaines croyances, où ce degré rend compte de l'accessibilité de cette croyance à la conscience. Voire à ce sujet les travaux en psychologie cognitive de John Anderson [AND 90, AND 93, AND 98].

va jouer avec Pierre, h n'en éprouvera aucune émotion particulière. m a tout intérêt à demander à Paul de venir plutôt qu'à Pierre.

Cas (C3a) observation et explication du comportement

Un matin h est visiblement stressé sans que m connaisse l'origine de ce stress. Supposons que m sait que h a une réunion dans la matinée où il doit faire une présentation, et que, parallèlement, une connaissance sur le monde indique à m que si on a préparé sa présentation, on s'attend à ce que sa réunion se passe bien tout en envisageant le cas où elle se passe mal (*i.e.* $Bel_m\,(Bel_h\,presentationOk \rightarrow Expect_h\,reunionOk \wedge Envisage_h\,\neg reunionOk))$[11].

De deux choses l'une : soit m croit que le fait que la réunion se passe bien est plaisant pour h ($Bel_m\,Pleasant_h\,reunionOk$), et dans ce cas il en déduit que si h croit que sa présentation n'est pas au point alors h espère que sa réunion se passera bien ($Bel_m\,(Bel_h\,\neg presentationOk \rightarrow Hope_h reunionOk)$) ; soit m croit que le fait que la réunion de h se passe mal est non-plaisant pour h (*i.e.* $Bel_m\,Unpleasant_h\,\neg reunionOk$), et à ce moment-là il croit que si h croit que sa présentation est bonne alors h craint que sa réunion se passe quand même mal (soit $Bel_m\,(Bel_h\,presentationOk \rightarrow Fear_h \neg reunionOk)$).

Nous nous plaçons dans ce second cas (le plus réaliste dans le contexte où h est stressé). Si m croit que h croit que sa présentation est au point (*i.e.* $Bel_m\,(Bel_h\,presentationOk)$, il en déduira que h craint que sa réunion ne le soit pas, sinon il n'en déduira rien.

Cas (C3b) : observation du comportement et hypothèse d'explication

Supposons que l'habitant h rentre le soir à la maison et l'agent intelligent m remarque que h affiche de la tristesse (autrement dit, après son observation[12], m croit que h est triste à propos d'un certain φ, soit : $Bel_m\,Sadness_h \varphi$). Cependant m ne connaît pas l'origine de cette émotion, *i.e.* il n'existe pas dans ses connaissances une formule φ vérifiant la définition de la tristesse (*i.e.* telle que $Bel_m\,Bel_h\,\varphi \wedge Bel_m\,Unpleasant_h\,\varphi$). Si on suppose que, d'après ses connaissances factuelles (ensemble CF défini Section 3)[13] l'agent m sait que h avait une réunion, on voudrait que m puisse en déduire que h croit que sa réunion s'est mal passée (soit $Bel_m\,Bel_h\,\neg reunionOk$) et que c'est à cause de ça qu'il est triste. Ainsi il pourra

11. On peut également imaginer que m dispose d'une loi pour le cas où la réunion n'a pas été préparée. Par exemple : $Bel_m\,(Bel_h\,\neg presentationOk \rightarrow Expect_h\,\neg reunionOk \wedge Envisage_h\,reunionOk)$.
12. Dans des travaux antérieurs, nous avons étudié les actions de perception (appelées *sensing actions* ou *knowledge gathering actions* dans la littérature cf. [HER 00, HER 02b]).
13. Ces connaissances peuvent être des connaissances initiales, ou des connaissances acquises *via* l'exécution par m d'actions d'observation, ou *via* des actions d'information exécutées par d'autres agents de la maison (par exemple, dans le cadre du rendez-vous, nous pouvons supposer l'existence d'un agent gérant l'agenda de h).

essayer de lui remonter le moral à ce sujet, ou lui proposer des services afin qu'il se relaxe.

Pour l'instant, nous n'avons aucune raison de supposer que m sait si la réunion s'est bien passée (soit : $\neg BelIf_m\, reunionOk$), bien qu'il sache *a priori* que le fait que la réunion se soit mal passée est déplaisant pour h [14] (soit : $Bel_m\, Unpleasant_h\, \neg reunionOk$), et qu'il croie que h sait si la réunion s'est bien passée (soit : $Bel_m\, BelIf_h\, reunionOk$).

Par ailleurs, d'après les définitions des émotions dont il dispose, m sait que si h croit que sa réunion s'est bien passée alors il sera content (*i.e.* $Bel_m\, (Bel_h\, reunionOk \rightarrow Joy_h reunionOk)$ et que si elle ne s'est pas bien passée, il sera triste (*i.e.* $Bel_m\, (Bel_h\, \neg reunionOk \rightarrow Sadness_h reunionOk)$. Or, comme m croit que h est triste, il infère que h croit que sa réunion ne s'est pas bien passée ($Bel_m\, Bel_h\, \neg reunionOk$) qui est ce qu'on voulait démontrer[15].

Nous avons traités les 4 cas identifiés en introduction. Nous ne sommes pas rentrés dans les détails de la modélisation du domaine, et n'avons volontairement pas explicité les preuves du raisonnements pour ne pas complexifier inutilement les exemples.

6. Conclusion

D'autres chercheurs se sont récemment intéressés à la formalisation logique des émotions. Nous analysons ci-dessous l'originalité de notre travail par rapport à deux autres contributions à ce domaine en pleine expansion.

Les travaux de Meyer [MEY 04] sont principalement orientés vers le lien entre les émotions et la satisfaction d'un plan. Il a développé KARO, une logique spécifique de l'action, de la croyance, et du choix, et il l'utilise pour exprimer une règle de génération pour chacune de ses quatre émotions (joie, tristesse, colère et peur). Ces règles sont assez complexes, et n'associent pour l'instant pas de degré d'intensité aux émotions générées. De plus seules quatre émotions sont décrites. Par contre Meyer étudie l'influence que les émotions exercent ensuite sur l'action, un point très important que nous ne traitons pas du tout pour l'instant.

Ochs et al. [OCH 05] mettent l'accent sur les expressions faciales émotionnelles pour des agents incarnés. Ils fournissent une formalisation basée sur la théorie de l'interaction rationnelle de Sadek [SAD 97], une logique de la croyance, l'intention, et l'incertitude. Ils restent très proches de la typologie OCC, mais n'en considèrent que quatre émotions (joie, tristesse, espoir et peur), auxquelles ils associent un degré d'intensité dépendant de leurs degrés d'incertitude. De plus ils étudient le problème essentiel du mélange des émotions. Leurs résultats ne sont pas très formels, proposant une sorte de mélange spatial des expressions faciales des émotions. Cependant c'est

14. connaissance épistémique issue de CE (défini Section 3)
15. Notons que nous avons ici un exemple où interviennent des inférences non triviales en logique épistémique.

déjà une première prise en compte de ce problème que nous avons pour notre part complètement ignoré jusqu'ici.

Pour conclure, nous argumentons que notre modèle, basé sur des définitions des émotions issues de la psychologie, reste assez simple à manipuler. Bien sûr il nous faudra encore le complexifier pour exprimer plus d'émotions, mais nous pensons que sa construction modulaire permet de réaliser facilement cette extension. De plus il est déjà utilisable en l'état, comme le prouve l'étude de cas que nous avons réalisée. Dans des travaux ultérieurs, nous envisageons plusieurs orientations de recherche, comme l'intégration de degrés d'intensité des émotions, permettant une adaptation plus fine, ou l'étude de stratégies d'adaptation à partir des théories psychologiques du coping. Nous pensons pouvoir établir un résultat de complétude, puis implémenter ce cadre dans le démonstrateur automatique de théorèmes développé au sein de notre équipe : Lotrec [GAS 05]. Toutes ces améliorations vont dans le sens d'agents toujours plus utiles aux êtres humains.

7. Bibliographie

[AND 90] ANDERSON J., *The Adaptive Character of Thought*, LEA, Hillsdale, NJ, 1990.

[AND 93] ANDERSON J., *Rules of the Mind*, LEA, NJ, 1993.

[AND 98] ANDERSON J., LEBIERE C., *The Atomic Components of Thought*, LEA, Mahwah, NJ, 1998.

[BRA 87] BRATMAN M. E., *Intention, Plans, and Practical Reason*, Harvard University Press, Cambridge, MA, 1987.

[CHE 80] CHELLAS B. F., *Modal Logic : an introduction*, Cambridge University Press, 1980.

[COH 90] COHEN P. R., LEVESQUE H. J., « Intention is Choice with Commitment », *Artificial Intelligence Journal*, vol. 42, n° 2–3, 1990.

[DAM 94] DAMASIO A. R., *Descartes' Error : Emotion, Reason, and the Human Brain*, Putnam Pub Group, 1994.

[DAR 72] DARWIN C. R., *The expression of emotions in man and animals*, Murray, London, 1872.

[EKM 92] EKMAN P., « An Argument for Basic Emotions », *Cognition and Emotion*, vol. 6, 1992, p. 169-200.

[FIP 02] FIPA (FOUNDATION FOR INTELLIGENT PHYSICAL AGENTS), « FIPA Communicative Act Library Specification », 2002, URL : http ://www.fipa.org/repository/aclspecs.html.

[GAS 05] GASQUET O., HERZIG A., LONGIN D., SAADE M., « LoTREC : Logical Tableaux Research Engineering Companion », BECKERT B., Ed., *International Conference on Automated Reasoning with Analytic Tableaux and Related Methods (TABLEAUX 2005), Koblenz (14–17 septembre), Germany*, n° 3702 LNCS, Springer Verlag, 2005, p. 318–322.

[GRA 04] GRATCH J., MARSELLA S., « A Domain-independent framework for modeling emotion », *Journal of Cognitive Systems Research*, vol. 5, n° 4, 2004, p. p. 269-306.

[HER 00] HERZIG A., LANG J., LONGIN D., POLACSEK T., « A logic for planning under partial observability », *Proc. Seventeenth National Conf. on Artificial Intelligence (AAAI-2000)*, Austin, Texas, 2000, p. 768–773.

[HER 02a] HERZIG A., LONGIN D., « A logic of intention with cooperation principles and with assertive speech acts as communication primitives », CASTELFRANCHI C., JOHNSON W. L., Eds., *Proc. 1st Int. Joint Conf. on Autonomous Agent and Multi-Agent System (AAMAS 2002) , Bologna*, New York, 15-19 juillet 2002, ACM Press, p. 920-927.

[HER 02b] HERZIG A., LONGIN D., « Sensing and revision in a modal logic of belief and action », VAN HARMELEN F., Ed., *Proc. of 15th European Conf. on Artificial Intelligence (ECAI 2002)*, Amsterdam, 2002, IOS Press, p. 307–311.

[HER 03] HERZIG A., « Modal probability, belief, and actions », *Fundamenta Informaticae*, vol. 57, n° 2-4, 2003, p. 323–344.

[HER 04] HERZIG A., LONGIN D., « C&L intention revisited », DUBOIS D., WELTY C., WILLIAMS M.-A., Eds., *Proc. 9th Int. Conf. on Principles on Principles of Knowledge Representation and Reasoning (KR2004)*, AAAI Press, 2004, p. 527–535.

[JAQ 04] JAQUES P. A., VICARI R. M., PESTY S., BONNEVILLE J.-F., « Applying Affective Tactics for a Better Learning. », *In Proceedings of the 16th European Conference on Artificial Intelligence (ECAI 2004)*, 2004.

[LAZ 91] LAZARUS R. S., *Emotion and Adaptation*, Oxford University Press, 1991.

[LEV 03] LEVENSON R. W., « Autonomic specificity and emotion », DAVIDSON R., SCHERER K., GOLDSMITH H., Eds., *Handbook of affective sciences*, Oxford Univ. Press, 2003.

[MEY 04] MEYER J. J., « Reasoning about Emotional Agents », DE MÁNTARAS R. L., SAITTA L., Eds., *16th European Conf. on Artif. Intell. (ECAI)*, 2004, p. 129–133.

[OCH 05] OCHS M., NIEWIADOMSKI R., PELACHAUD C., SADEK D., « Intelligent Expressions of Emotions », *1st International Conference on Affective Computing and Intelligent Interaction ACII*, China, October 2005.

[ORT 88] ORTONY A., CLORE G., COLLINS A., *The cognitive structure of emotions*, Cambridge University Press, Cambridge, MA, 1988.

[PIA 89] PIAGET J., « Les Émotions », RIMÉ B., SCHERER K., Eds., *Les relations entre l'intelligence et l'affectivite dans le developpement de l'enfant*, p. 75–95, Delachaux et Niestlé, Neuchâtel-Paris, 1989.

[PIC 97] PICARD R. W., *Affective Computing*, The MIT Press, 1997.

[PRE 05] PRENDINGER H., ISHIZUKA M., « Human physiology as a basis for designing and evaluating affective communication with life-like characters », *IEICE Transactions on Information and Systems*, vol. E88-D, n° 11, 2005, p. 2453–2460.

[RAO 91] RAO A. S., GEORGEFF M. P., « Modeling Rational Agents within a BDI-Architecture », ALLEN J. A., FIKES R., SANDEWALL E., Eds., *Proc. 2nd Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'91)*, Morgan Kaufmann, 1991.

[RUS 97] RUSSELL J. A., « How shall an emotion be called ? », PLUTCHIK R., CONTE H., Eds., *Circumplex models of personality and emotions*, American Psychol. Assoc., 1997.

[SAD 97] SADEK D., BRETIER P., PANAGET F., « ARTIMIS : Natural Dialogue Meets Rational Agency », *Proc. of Int. Joint Conf. on Art. Intellig. (IJCAI'97)*, 1997, p. 1030–1035.

[SEA 69] SEARLE J. R., *Speech acts : An essay in the philosophy of language*, Cambridge University Press, New York, 1969.

[SEA 85] SEARLE J. R., VANDERVEKEN D., *Foundation of illocutionary logic*, Cambridge University Press, 1985.

[VAN 90] VANDERVEKEN D., *Meaning and Speech Acts*, Cambridge University Press, 1990.