

# Proposition de sujet de recherche

Sujet : Vers un modèle pour décrire et implémenter la  
dynamique du raisonnement

Pour Mr. Belloula Anouar Sadat  
Par Brahim Chaib-draa

February 16, 1998

## 1 Le contexte

Le système *intentionnel* tel que présenté par Dennett [6], nous semble intéressant pour la modélisation informatique. Il y a plusieurs raisons à cela : (1) c'est un modèle sur lequel on peut se baser pour analyser et concevoir des systèmes informatiques intelligents (appelés *agents*); (2) c'est un modèle qui aide à comprendre et à expliquer les systèmes complexes; (3) c'est un modèle qui fait ressortir certaines régularités et certains motifs d'actions qui sont indépendants de l'implémentation physique des agents; (5) c'est un modèle qui facilite la conception d'agents informatiques capables d'interagir efficacement. Il est cependant, fort utile de se poser la question de la légitimité de l'attribution d'attitudes mentales à des agents artificiels. Dans la communauté "informatique", la majorité des chercheurs pensent qu'il est légitime et utile d'attribuer des états mentaux à des agents artificiels. McCarthy, l'un des pionniers en IA, soutient, par exemple, que :

*Attribuer des attitudes mentales à une machine est légitime* lorsqu'une telle attribution exprime la même information à propos de la machine, qu'elle exprime à propos d'une personne. Cette attribution est *utile* quand elle nous permet de comprendre la structure de la machine, son comportement ancien ou futur, ou comment la réparer ou l'améliorer. ... exprimer raisonnablement et brièvement ce qui est connu à propos de l'état d'une machine dans une situation particulière peut nécessiter l'attribution de qualités mentales. [[22], trad. libre]

Ceci dit, il est donc avantageux d'avoir un modèle représentant les attitudes mentales. Ce modèle nous donne la possibilité de décrire le comportement d'un système sans avoir recours à une description détaillée de son implémentation en machine ou de ses composants physiques. Nous pouvons, également, l'utiliser pour évaluer un système, en jugeant si ses attitudes mentales sont conformes à des spécifications préétablies. Les attitudes mentales qui ont été les plus étudiées en informatique sont les croyances, les connaissances, les intentions, les désirs, les préférences, les obligations et les engagements. Parmi les thèmes de recherche où ces attitudes ont un rôle central, on peut citer :

1. l'interaction entre les connaissances et l'action [23];
2. la compréhension du langage naturel et la théorie des actes du langage [2, 3, 5, 11, 20];
3. la formalisation d'agents rationnels basée sur une architecture intégrant croyances, désirs et intentions [25, 4];
4. la programmation orientée agent [28, 30];
5. les logiques pour la spécification et la vérification des systèmes multiagents [36];
6. les logiques pour la formalisation d'agents à rationalité limitée [15];
7. spécification et raisonnement sur les connaissances dans les systèmes distribués [8];
8. la robotique cognitive [18, 19].

Malheureusement, la plupart de ces travaux de recherche se sont concentrés sur les aspects statiques des attitudes mentales. C'est ainsi que jusqu'à récemment, un agent artificiel est modélisé comme une entité ayant des attitudes mentales qui ne changent pas au cours du temps. C'est grâce aux travaux

menés en philosophie qu'on est venu à s'intéresser aux aspects dynamiques. En fait, la philosophie des sciences reconnaît deux types de changements : les grands et les petits. Les grands changements sont appelés par Kuhn [16] *les révolutions scientifiques*; la relativité d'Einstein en est un exemple. Les petits changements, étudiés principalement par Popper [24] et Lakatos [17], représentent généralement ce qui change suite au progrès évolutif. Un aspect significatif de ces petits changements est l'aspect de *révision des croyances*, c'est à dire, le changement de croyances qu'un agent opère quand il est confronté à de nouvelles informations provenant de son monde extérieur. La révision est ici considérée dans son sens le plus général, à savoir, celui de changement. Dans ce cas, les opérations de changements de croyances par rapport aux propositions, dûs aux nouvelles informations peuvent être classés comme suit [1, 12]:

1. *Expansion*: une nouvelle croyance avec toutes ses conséquences possibles  $\phi$  doit être acceptée par un système de croyances  $K$ . Le système de croyances qui en résulte, en étendant  $K$  par  $\phi$  et ses conséquences logiques, est noté  $K + \phi$ .
2. *Révision*: Un système de croyances  $K$  doit être étendu à une nouvelle croyance  $\phi$ , typiquement inconsistante avec ce système. Pour garder le système de croyances résultant, consistant, il faudrait rétracté certaines anciennes croyance de  $K$ . L'opération de révision de  $K$  par  $\phi$  est noté  $K * \phi$ .
3. *Contraction*: une nouvelle croyance  $\phi$  doit être rétractée d'un système de croyances  $K$  sans l'ajout de nouvelles informations. Pour que le système de croyances résultant soit consistant, il faudrait aussi abandonner certaines autres croyances de  $K$ . L'opération de contraction de  $K$  avec  $\phi$  est notée  $K \circ \phi$ .

Ces trois aspects ont donné lieu, ces dernières années, à une intense activité de recherche. Cependant, bien des points restent en suspend, particulièrement : 1) la dynamique des autres attitudes mentales, autres que les croyances; 2) l'influence des actes du langage et/ou du discours sur cette dynamique, 3) l'influence de la crédibilité, de la sincérité et de la compétence de la source des nouvelles informations sur cette dynamique, 4) le choix d'un modèle qui décrit et implémente la dynamique du raisonnement. La présente proposition vise justement à *contribuer* à ces différents points.

## 2 Les objectifs

Les principaux objectifs de recherche consistent à :

1. élaborer une nouvelle sorte de position, en plus des trois proposées par Dennett, et qu'on appellerait *position sociale*, en vue d'incorporer les états mentaux partagées, les conventions et les différents modes de vie qui sont à la base du langage;
2. étudier la révision des croyances dans le cadre de la théorie des actes du langage en déterminant les états mentaux du locuteur et de l'interlocuteur dans le cadre d'un transfert d'information. Il convient d'étudier, entre autres, l'incidence de la crédibilité, la sincérité, la compétence et les préférences des intervenants à un dialogue, sur le processus de révision des croyances;
3. étendre la révision aux états mentaux, autres que les croyances, comme par exemple : les intentions, les engagements et les plans;

4. contribuer au raisonnement dynamique en proposant un modèle qui puisse aider à décrire et à implémenter ce type de raisonnement;
5. mettre à l'épreuve le modèle proposé;

### 3 La méthodologie adoptée

#### 3.1 Réalisation de l'objectif 1

L'utilisation d'un langage donné requiert de la part de ses utilisateurs, l'acceptation de certaines conventions au niveau de la grammaire et de la sémantique. Ces conventions jouent un rôle crucial quand on essaye d'expliquer ou de prédire le comportement linguistique d'un individu donné. Or, si on se fie uniquement aux trois sortes de positions de Dennett, on ne peut expliquer pleinement ces conventions. Dès lors il convient d'étudier la faisabilité d'une quatrième sorte de position qui rend compte de l'"aspect social" de l'individu, et qui permettrait d'expliquer i) les états mentaux partagés (connaissances communes, intentions communes, engagements communs, etc.), ii) les conventions, iii) l'interprétation dialogique des actes du discours [32]. La nécessité d'étudier cette quatrième sorte s'explique également par le fait que les états mentaux partagés (ce que Searle appelle l'intentionnalité partagée [26]) ne peuvent se réduire à la somme d'intentions individuelles. Une fois cette "position sociale" étudiée, le langage sera considéré du point de vue de cette nouvelle position, ce qui nous permettrait d'étudier la révision des états mentaux, engendrés par les actes du langage, en tenant compte des attitudes mentales sociales.

#### 3.2 Réalisation de l'objectif 2

Les ouvrages de Searle [27] et de Vanderveken [33, 34] enrichissent la théorie des actes en proposant un modèle formel, la logique illocutoire, pour traiter les actes et les relations entre eux. Il apparaît clairement au niveau de ce modèle, qu'un acte n'est pas primitif et que son accomplissement repose sur les cinq composantes suivantes : 1) l'intention ou le but de l'énoncé, 2) le mode de réalisation, 3) les conditions préparatoires, 4) les conditions de sincérité, 5) le degré de force. Ce sont ces cinq notions qui sont primitives, l'acte illocutoire n'est qu'un concept dérivé. Le problème ici est d'étudier les liens de ces notions avec la révision des croyances faisant ainsi ressortir : la fiabilité, la sincérité, la compétence et les préférences des interlocuteurs. C'est ainsi par exemple, qu'il convient d'évoquer, en utilisant le mode de réalisation d'une force illocutoire [33, 34], la position d'autorité du locuteur qui, sous certaines conditions qu'il convient de délimiter, donnerait lieu à une révision partielle ou totale des croyances de l'interlocuteur.

#### 3.3 Réalisation de l'objectif 3

On se posera ici la question de savoir comment les intentions, les engagements, les buts et les plans changent suite à un transfert d'information entre agents. Autrement dit, comment l'expansion, la révision et la mise à jour contribuent aux changements des autres états mentaux des agents? Par exemple, quand je téléphone (de mon bureau) à mon fils (qui se trouve à la maison) lui demandant de faire ses devoirs, il me répond "oui, papa", sur le coup, en s'engageant à les faire dans les meilleurs délais. Cinq minutes plus tard sa mère lui téléphone en l'informant que je ne rentrerais pas à la maison, le soir en question. On suppose ici que sur la base de cette nouvelle information de la part de sa maman, mon fils revient sur l'engagement qu'il m'a fait et change son plan en allant voir la TV et jouer au nintendo.

Il est clair que, les modes de vie, les motivations et les préférences des protagonistes, sont sous-jacent à de tels changements. Une étude approfondie de ces notions permettra, sans aucun doute, de franchir un pas important, en direction d'un modèle informatique pour la dynamique des états mentaux.

### 3.4 Réalisation de l'objectif 4

Cette phase constitue le substrat du sujet de de recherche proposé et doit se dérouler en parallèle avec les phases précédentes. Elle vise à contribuer à l'élaboration d'un modèle computationnel qui tient compte de la dynamique des états mentaux suite à une interaction basée sur le discours. Cette logique tiendra compte des aspects de logique dynamique présentés précédemment. Un tel modèle doit rendre compte de la finitude et des capacités cognitives restreintes des locuteurs humains, aussi bien que des capacités créatives de leur compétences linguistiques. Il devrait aussi expliquer pourquoi ces locuteurs ne sont pas omniscients, ni parfaitement rationnels, même s'ils sont toujours minimalement cohérents. Des pistes comme celles de Gillet [13] et de Fagin [9] semblent prometteuses dans ce contexte. Le modèle doit particulièrement incorporer une "logique de dépendance des croyances" dans un contexte multiagents [15]. Cette logique des dépendances fait intervenir l'opérateur de dépendance  $D_{i,j}$  et l'opérateur de croyance  $B_{i,j}$ . Ce dernier est défini de la manière suivante :  $B_{i,j}\phi \stackrel{\text{def}}{=} D_{i,j}\phi \wedge B_j\phi$ . Ainsi, intuitivement  $B_{i,j}$  signifie que l'agent  $i$  croit que  $\phi$  est vraie en se fiant à  $j$ . Dans le même ordre d'idée, il faudra également faire intervenir la notion de *connaissances distribuées* (une connaissance distribuée est une connaissance obtenue en combinant les informations distribuées) si on veut par exemple tenir compte du transfert d'information d'un groupe à un individu. La relation avec les *connaissances communes* (ce que chacun connaît et sait que l'autre le sait, etc.) est également pertinente ici et il convient de la prendre en compte.

### 3.5 Réalisation de l'objectif 5

Comment utiliser la révision des états mentaux et le modèle sous-jacent pour implémenter des stratégies de communication, de coordination et plus généralement de raisonnement d'agents cognitifs au sein d'un groupe? Pour cela, il convient peut être de tenir compte : 1) des structures de planification; 2) de la nature des actes de communication entre agents; 3) des structures de coordination du groupe d'agents ( intentions partagées, plans partagés, plans communs, conflits, etc.). Il convient de déterminer enfin, les mécanismes de raisonnement qui permettent, à partir des conditions de succès et de satisfaction d'un acte donné, de réviser les états mentaux des agents intervenants dans le contexte d'une communication. Cette révision (au sens large du terme) devrait tenir compte de la crédibilité, la sincérité, la compétence et les préférences de chacun des protagonistes ainsi que de leurs modes de vie complets et l'arrière-fond de chaque jeu de langage.

## 4 Résultats escomptés

Comme ce sujet est subventionné par le CRSNG, le FCAR et le CRSH, l'accent est surtout mis sur les publications. Mr. Belloula est tenu de publier au moins un papier dans une conférence et au moins une publication dans une revue. Un autre résultat escompté est le développement d'une plateforme basé sur le modèle développé par Belloula pour la communication entre agents dans laquelle l'aspect dynamique des états mentaux est important.

## References

- [1] C. E. Alchourrón, P. Gärdenfors and D. Mackinson. On the Logic of Theory Change: Partial meet Functions for Contraction and Revision. *Journal of symbolic Logic*. 50, p. 510-530, 1985.
- [2] J. f. Allen and R. C. Perrault, Analyzing Intention in Utterances, *Artificial Intelligence*, 15(3):143-178, 1980.
- [3] D. Appelt, *Planning English Sentences*, Cambridge University Press, 1985.
- [4] M. E. Bratman. What is Intention? in M. E. Pollack, P. R. Cohen and J. L. Morgan (eds), *Intentions in Communication*, p. 15-32, MIT Press, 1990.
- [5] P. R. Cohen and H. Levesque, Intention is Choice with Commitment, *Artificial Intelligence*, 42, pp. 213–261, 1990.
- [6] D. C. Dennett, The Logical Geography of Computational Approaches: A View from the East Pole, in M. Brand and M. Harnish (eds), *The Representation of Knowledge and Belief*, pp. 59–79, Univ. of Arizona Press, Tucson, Texas, 1986.
- [7] D. C. Dennett, *Brainstorms*, Harvester Press, 1981.
- [8] R. Fagin, I. Y. Halpern, Y. Moses and M. Y. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.
- [9] R. Fagin, J. Halpern and M. Vardi. Belief, Awareness and Limited Reasoning. *Artificial Intelligence*. 34, p. 39-76, 1988.
- [10] J. Fodor, *The Language of Thought*, Crowell, NY, 1975.
- [11] J. R. Galliers. *A Theoretical Framework for Computer Models of Cooperative Dialogue*. Ph.D. thesis, Open University: UK, 1988.
- [12] P. Gärdenfors and H. Rott. Belief Revision. In D. Gabbay, C. J. Hogger and J. A. Robinson (eds), *Handbook of Logic in Artificial Intelligence and Logic Programming*. 4, Oxford Science Publication, 1995.
- [13] E. Gillet and P. Gochet. *La logique de la Connaissance. Le problème de l'omniscience logique*. Université de Liège, Séminaire de Logique et d'Épistémologie, 1990.
- [14] G. Harman, *Thought*, Princeton University Press, Princeton, NJ, 1973.
- [15] Z. Huang. *Logics for Agents with Bounded Rationality*. Ph.D. thesis, Institute for Logic, Language and Computation ILLC, Univ. van Amsterdam, 1994.
- [16] T. S. Kuhn. *The Structure of Scientific Revolutions*. The University of Chicago Press, Chicago, 1962.
- [17] I. Lakatos. *The Methodology of Scientific Research Programmes: Philosophical Papers*. vol. 1, Cambridge Univ. Press, Camb., 1978.
- [18] Y. Lesperance et al. A Logical Approach to High-Level Programming—A Progress Report. In *Control of the Physical world by Intelligent Systems*, 1994.
- [19] H. Levesque. Knowledge, Action and Ability in the Situation Calculus. overheads from invited talk at TARK 94, 1994.
- [20] D. J. Litman and J. F. Allen. A Plan Recognition Model for Subdialogues in Conversations. *em Cognitive Sciences*, 11, pp. 163-200, 1987.
- [21] D. Makinson. General Patterns in Nonmonotonic Reasoning. In D. Gabbay, C. J. Hogger and J. A. Robinson (eds), *Handbook of Logic in Artificial Intelligence and Logic Programming*. 3, Oxford Science Publication, 1995.
- [22] J. McCarthy and P. J. Hayes, Some Philosophical Problems from the Standpoint of Artificial Intelligence. In B. Meltzer and D. Mitchie (eds.), *Machine Intelligence 4*, Edinburgh University Press, 1969.

- [23] R. Moore, *Reasoning about Knowledge and Action*. Ph.D. thesis, MIT, 1980.
- [24] K. R. Popper. *The Logic of Scientific Discovery*. Princeton University Press, Princeton, New Jersey, 1974.
- [25] A. Rao and M. P. Georgeff. Modeling Rational Agents within a BDI–Architecture. in E. Sandewall and R. Fikes (eds.), *Proc. of Knowledge Representation and Reasoning*, pp. 473-484, Morgan Kaufmann Pub., 1981.
- [26] J. Searle. Conversation. In H. Parret and J. Verschueren. *(On) Searle on Conversation*. Benjamins, Amsterdam, 1992.
- [27] J. Searle and D. Vanderveken. *Foundations of Illocutionary Logic*. Cambridge Univ. Press, Camb., 1985.
- [28] Y. Shoham. Agent-Oriented Programming. *artificial Intelligence*, 60(1): 51–92, 1993.
- [29] J. Searle, *Intentionality: An Essay in the Philosophy of Mind*, Cambridge Univ. Press, NY, 1983.
- [30] S. R. Thomas. *PLACA, an Agent Oriented Programming Language*. Ph.D thesis, CS dept., Stanford University, Stanford, 1993.
- [31] S. Torrance, Philosophy and AI: Some Issues, in S. Torrance (ed.) *The Mind and the Machine: Philosophical Aspects of AI*. Ellis Horw. Lt, 1984.
- [32] A. Trognon and C. Brassac. L’enchainement conversationnel. *Cahiers de Linguistique Française* 13. p. 76-107, 1992.
- [33] D. Vanderveken. *Meaning and Speech Acts (vol. 1): Principles of Language Use*. Cambridge Univ. Press, Camb., 1990.
- [34] D. Vanderveken. *Meaning and Speech Acts (vol. 2): Formal Semantics of Success and Satisfaction*. Cambridge Univ. Press, Camb., 1990.
- [35] T. Winograd and F. Flores, *Understanding Computers and Cognition: A New Foundation for Design*. Addison–Wesley Pub. Comp. Inc., Reading, Mass., 1987.
- [36] M. Wooldridge. *The Logical Modelling of Computational Multi-Agents Systems*. Ph.D. thesis, The Manchester Metropolitan University, 1994.