THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
Institut National Polytechnique de Toulouse (INP Toulouse)

Discipline ou spécialité :

Intelligence Artificielle

Présentée et soutenue par :
Filipo Studzinski Perotto

le : vendredi 11 juin 2010

Titre :

Un Mécanisme Constructiviste d'Apprentissage Automatique d'Anticipations pour des Agents Artificiels Situés

---

JURY

Yves Demazeau
Antônio Carlos da Rocha Costa
Paulo Martins Engel
Georgi Stojanov

Ecole doctorale :
Mathématiques Informatique Télécommunications (MITT)

Unité de recherche :
Raisonnement et Décision

Directeur(s) de Thèse :
Jean-Christophe Buisson
Luís Otávio Campos Álvares

Rapporteurs :
Yves Demazeau
Antônio Carlos da Rocha Costa

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

ET

INSTITUT NATIONAL POLYTECHNIQUE
INSTITUT DE RECHERCHE EN INFORMATIQUE DE TOULOUSE

FILIPO STUDZINSKI PEROTTO

Un Mécanisme Constructiviste d'Apprentissage Automatique d'Anticipations pour des Agents Artificiels Situés

Thèse de Doctorat

Prof. Dr. Luís Otávio Campos Álvares
Directeur de Thèse (UFRGS)

Prof. Dr. Jean-Christophe Buisson
Directeur de Thèse (INPT)

Toulouse, Juin 2010.

Informations pour la Publication

Perotto, Filipo Studzinski
Un Mécanisme Constructiviste d'Apprentissage Automatique d'Anticipations pour des Agents Artificiels Situés / Filipo Studzinski Perotto: IRIT / INPT: 2010.

203 f.: Il.

Thèse (Doctorat) – L'Institut National Polytechnique de Toulouse, en coopération avec l'Universidade Federal do Rio Grande do Sul (Brésil), 2010. Directeurs de thèse: Luís Otávio ALVARES et Jean-Christophe BUISSON.

1. Intelligence Artificielle Constructiviste. 2. Intelligence Artificielle. 3. Apprentissage Automatique. 4. Agents Autonomes. 5. Induction de Concepts. 6. Développement Cognitif Artificiel. 7. Piaget. 8. Découverte de Structure dans des FPOMDP. 9. Processus de Décision Markovien (MDP). I. ALVARES, Luís Otávio Campos. II. BUISSON, Jean-Christophe. III. Titre.

Présentation

(1) La présente thèse est le résultat d'une recherche développée entre août 2004 et juin 2010, période au cours de laquelle j'ai été en doctorat dans le « Programa de Pós-Graduação em Computação » (PPGC) de l' « Universidade Federal do Rio Grande do Sul » (UFRGS), au Brésil, sous la direction du professeur Luís Otávio Campos Álvares. De même, ce travail a été réalisé en cotutelle avec l' « Institut National Polytechnique de Toulouse » (INPT), en France, sous la direction du professeur Jean-Christophe Buisson. À ces cinq années de recherche sur le thème de l'apprentissage automatique basé sur le modèle constructiviste, on peut ajouter au moins quatre années supplémentaires, puisque mon intérêt pour le sujet remonte à 2001.

(2) Pour ma licence en informatique, j'ai écrit un mémoire dans lequel j'ai analysé l'Intelligence Artificielle (IA) à travers son histoire en tant que discipline scientifique, et j'ai repris les discussions sur l'entreprise de l'IA au plan théorique et philosophique. Pendant cette période, j'ai réalisé en parallèle une incursion dans les sciences humaines, en donnant des cours dans le domaine de la philosophie, la psychologie et la pédagogie. Cette expérience n'a fait qu'aiguiser mon intérêt par rapport aux questionnements sur l'esprit et sur l'intelligence, en m'amenant finalement à la psychologie constructiviste. Après cela, ont suivies deux années de recherche dans le cadre du master, réalisé entre 2002 et 2004, sous la direction du professeur Rosa Maria Vicari. Pour le mémoire de master, j'ai développé un travail déjà relatif au thème de l'intelligence artificielle constructiviste.

(3) Ainsi, cette thèse est partie d'une trajectoire de recherche plus large. En regardant en arrière, je perçois que dans ces 10 dernières années à l'université j'ai été concentré toujours à peu près sur le même thème, dans un sujet surement fascinant. Cela ne peut que représenter, en fin de compte, un projet de vie académique. J'arrive à la fin

de la thèse, même après tout ce temps, avec l'impression qu'il y a toujours plus de questions que de réponses, et que alors ce projet est encore loin d'être épuisé.

Remerciements

(4) Je voudrais remercier les institutions qui ont financé notre recherche, notamment le Conseil National de la Recherche (CNPq) et le Conseil pour le Soutien de la Recherche (CAPES), des organismes publics brésiliens dont l'appui a été indispensable pour rendre possible la réalisation de ce travail. Je remercie également l'Université Fédérale de Rio Grande do Sul (UFRGS), à travers son Programme de Post-Graduation en Informatique (PPGC), et l'Institut National Polytechnique de Toulouse (INPT), à travers l'Institut de Recherche en Informatique (IRIT). Ces deux universités, en coopération, nous ont accueillis de manière excellente, en offrant un bon environnement et en fournissant des ressources suffisantes pour la bonne exécution du travail.

(5) Je tiens à remercier mes amis, le professeur Jean-Christophe Buisson et le professeur Luis Otávio Álvares, qui ont accepté de diriger cette recherche dans un domaine complexe mais aussi passionnant, pour leur confiance et pour leur attitude toujours correcte et encourageante.

(6) D'autre part, je tiens à remercier tous les chercheurs qui ont contribué à ce travail de façons différentes, parmi lesquels je voudrais citer les professeurs Paulo Engel, Ana Bazzan, Rosa Vicari, Cecília Flores, Dante Couto Barone, Magda Bercht, Georgi Stojanov, Jean-Luc Basille, Paulo Quaresma, Maria Alice Pimenta Parente, Maria Luiza Becker, Fernando Becker, Antônio Carlos da Rocha Costa, et Jaime Rebello.

(7) Quelques collègues et amis ont été proches du processus de construction de la thèse, en y participant parfois activement, et je souhaite vivement les remercier: Jean-Charles Quinton, Guillaume Giffard, Bruno Castro da Silva, Eduardo Wisnieski Basso, Licurgo Bennemann Almeida, Juliano Bittencourt, et Ivan Medeiros. En particulier et très spécialement je remercie Cássia Trojahn dos Santos, qui a partagé avec moi cette

bataille du doctorat, ainsi qu'une période de ma vie que je garderai toujours dans mon cœur.

(8) Je voudrais vivement exprimer ma gratitude aux personnes importantes en dehors du contexte académique, pour le soutien que j'ai reçu pendant toute la (longue) période de travail, un soutien sans lequel le succès ne serait pas possible. C'est, donc, un honneur de pouvoir remercier d'une façon ouverte les amis Fabiano Mesquita Padão et Simone Segalla de Dutra Souza, pour leur amitié sincère et présente.

(9) De la même façon je remercie les bons amis de mon séjour en France, parmi lesquels suis heureux de citer: Pierre Amilhaud, Helène Roques, Sonia Rodriguez, André de Andrade, Cláudia Santos Gai, Bernardo Cougo, Marta Ramos Oliveira, Daniel Mur, Fares Fares, Constance Le Pocher et Yoan Elgorriaga.

(10) Parmi les personnes qui ont joué un rôle remarquable dans ma vie, anciennement ou récemment, plus ou moins longtemps, académiquement ou personnellement, je voudrais mentionner: Anderson Cleiton Silveira, André Luís Nodari, Eduardo Rocha D'Ávila, Cássio Dorneles, Adriana Santana, Carina Regina Pereira, Samara Kalil, Rossana Rodrigues, Alessandra Carla Ceolin, Lidiani Käfer, Bruno Hailliot, Thaís Pessato, Cacá Prüfer, Sabrina Bertrand Coelho, Thiago Ingrassia Pereira, Josiane Faganello, Vera Mello, et Willy Ricardo Petersen Filho.

(11) À ma famille, d'abord à tous les oncles, tantes, cousins et cousines, des partenaires infaillibles. Très spécialement à mes parents, Adelino Victo Perotto et Maria Studzinski Perotto, avec qui j'ai appris la plupart des qualités nécessaires pour bien réussir à faire la thèse, et surtout pour la vie. Et aussi à mon frère, Rafaello Studzinski Perotto, et sa compagne, Kelly Lopes dos Santos, des personnes que j'admire beaucoup.

(12) Finalement je suis très heureux de pouvoir faire une dédicace, et également remercier Emanuele Carvalheira de Maupeou, ma compagne, personne et femme aux qualités innombrables, pour être à coté de moi, et pour cet amour intense qui sera toujours vivant, sur n'importe quelle partie du monde.

Table des Matières

Liste des Figures

Liste des Symboles

Æ	Système Global
д	Agent
ξ	Environnement
β	Corps
μ	Esprit
Я	Système Régulatif
Ж	Système Cognitif
Є	Extériorité

$Q = \{q_1, q_2 ...\}$	Ensemble d'États
$A = \{a_1, a_2 ...\}$	Ensemble d'Actions
$O = \{o_1, o_2 ...\}$	Ensemble d'Observations

$M = \{M_1, M_2 \ldots\}$	Variables d'Actuation (*Motrices*)
$m^{(t)} = \{m_1, m_2 \ldots\}$	Actuation à temps t
$S = \{S_1, S_2 \ldots\}$	Variables de Situation (*Sensorielles)*
$s^{(t)} = \{s_1, s_2 \ldots\}$	Situation à temps t
$C = \{C_1, C_2 \ldots\}$	Variables de Contrôle
$c^{(t)} = \{c_1, c_2 \ldots\}$	Contrôle à temps t
$X = \{X_1, X_2 \ldots\}$	Variables de Contexte
$x^{(t)} = \{x_1, x_2 \ldots\}$	Contexte à temps t
$P = \{P_1, P_2 \ldots\}$	Variables de Perception (*Observables*)
$p^{(t)} = \{p_1, p_2 \ldots\}$	Contexte Perceptif à t
$H = \{H_1, H_2 \ldots\}$	Variables Abstraites (*Cachées*)
$h^{(t)} = \{h_1, h_2 \ldots\}$	Contexte Abstrait à t

γ	Fonction d'Observation
δ	Fonction de Transition
r	Fonction de Récompense
π	Politique

$\tau = \{\tau_1, \tau_2 \ldots\}$	Fonctions de Transformation
$\sigma = \{\sigma_1, \sigma_2 \ldots\}$	Fonctions de Régularité
$л = \{л_1, л_2 \ldots\}$	Fonctions d'Évaluation

∅	Ensemble vide
t	Instant dans le temps
i	Indice d'un élément dans un ensemble
n	Taille d'un ensemble

φ	Degré de Structuration
ω	Degré d'Observabilité
∂	Degré de Déterminisme

ε	Paramètre de Curiosité
α	Paramètre de Taille Maximal

$\text{Ш} = \{\text{Ш}_1, \text{Ш}_2 \ldots\}$	Mémoires Épisodiques
$\text{М} = \{\text{м}_1, \text{м}_2 \ldots\}$	Liste de Mémoires Récentes

$\Psi = \{\Psi_1, \Psi_2 \ldots\}$	Arbres d'Anticipation
$\text{Ж} = \{\text{Ж}_1, \text{Ж}_2 \ldots\}$	Arbres de Délibération
$\text{Л}\upsilon = \{\text{Л}\upsilon_1, \text{Л}\upsilon_2 \ldots\}$	Arbres d'Évaluation

e	Anticipation (Expectative)
υ	Valeur Affective
ρ	Fiabilité

$\Lambda = \{\lambda_1, \lambda_2 \ldots\}$	Liste de Différenciateurs
$\Theta = \{\Theta_1, \Theta_2 \ldots\}$	Nœuds Intermédiaires
$\Xi = \{\Xi_1, \Xi_2 \ldots\}$	Schémas
$\text{đ} = \{\text{đ}_1, \text{đ}_2 \ldots\}$	Décideurs

prob( )	Probabilité d'un phénomène
dom( )	Domaine d'une variable
rel( )	Sous-ensemble pertinent

Liste des Définitions

Liste des Algorithmes

Un Mécanisme Constructiviste d'Apprentissage Automatique d'Anticipations pour des Agents Artificiels Situés

RÉSUMÉ

(13) Cette recherche se caractérise, premièrement, par une discussion théorique sur le concept d'agent autonome, basée sur des éléments issus des paradigmes de l'*Intelligence Artificielle Située* et de l'*Intelligence Artificielle Affective*. Ensuite, cette thèse présente le problème de *l'apprentissage de modèles du monde*, en passant en revue la littérature concernant les travaux qui s'y rapportent. À partir de ces discussions, l'architecture CAES et le mécanisme CALM sont présentés.

(14) CAES (*Coupled Agent-Environment System*) constitue une architecture pour décrire des systèmes basés sur la dichotomie agent-environnement. Il définit l'agent et l'environnement comme deux systèmes partiellement ouverts, en couplage dynamique. Dans CAES, l'agent est composé de deux sous-systèmes, l'esprit et le corps, suivant les principes de la situativité et de la motivation intrinsèque.

(15) CALM (*Constructivist Anticipatory Learning Mechanism*) est un mécanisme d'apprentissage fondé sur l'approche constructiviste de l'Intelligence Artificielle. Il permet à un agent situé de construire un modèle du monde dans des environnements partiellement observables et partiellement déterministes, sous la forme d'un *processus de décision markovien partiellement observable et factorisé* (FPOMDP). Le modèle du monde construit est ensuite utilisé pour que l'agent puisse définir une politique d'action visant à améliorer sa propre performance.

(16) **Mots-Clés:** Intelligence Artificielle Constructiviste, Intelligence Artificielle, Apprentissage Automatique, Agents Autonomes, Induction de Concepts, Développement Cognitif Artificiel, Piaget, Découverte de Structure dans des FPOMDP, Processus de Décision Markovien (MDP).

Um Mecanismo Construtivista para Aprendizagem de Antecipações em Agentes Artificiais Situados

Resumo

(17) Esta pesquisa caracteriza-se, primeiramente, pela condução de uma discussão teórica sobre o conceito de agente autônomo, baseada em elementos provenientes dos paradigmas da *Inteligência Artificial Situada* e da *Inteligência Artificial Afetiva.* A seguir, a tese apresenta o problema da *aprendizagem de modelos de mundo*, fazendo uma revisão bibliográfica a respeito de trabalhos relacionados. A partir dessas discussões, a arquitetura CAES e o mecanismo CALM são apresentados.

(18) O CAES (*Coupled Agent-Environment System*) é uma arquitetura para a descrição de sistemas baseados na dicotomia agente-ambiente. Ele define agente e ambiente como dois sistemas parcialmente abertos, em acoplamento dinâmico. No CAES, o agente é composto por dois subsistemas, mente e corpo, seguindo os princípios de situatividade e motivação intrínseca.

(19) O CALM (*Constructivist Anticipatory Learning Mechanism*) é um mecanismo de aprendizagem fundamentado na abordagem construtivista da Inteligência Artificial. Ele permite que um agente situado possa construir um modelo de mundo em ambientes parcialmente observáveis e parcialmente determinísticos, na forma de um *Processo de Decisão de Markov Parcialmente Observável e Fatorado* (FPOMDP). O modelo de mundo construído é então utilizado para que o agente defina uma política de ações a fim de melhorar seu próprio desempenho.

(20) **Palavras-Chave:** Inteligência Artificial Construtivista, Inteligência Artificial, Aprendizagem de Máquina, Agentes Autônomos, Indução de Conceitos, Desenvolvimento Cognitivo Artificial, Piaget, Descoberta de Estrutura em FPOMDP, Processo de Decisão de Markov (MDP).

A Constructivist Anticipatory Learning Mechanism for Situated Artificial Agents

ABSTRACT

(21) This research is characterized, first, by a theoretical discussion on the concept of autonomous agent, based on elements taken from the Situated AI and the Affective AI paradigms. Secondly, this thesis presents the problem of learning world models, providing a bibliographic review regarding some related works. From these discussions, the CAES architecture and the CALM mechanism are presented.

(22) The CAES (*Coupled Agent-Environment System*) is an architecture for describing systems based on the agent-environment dichotomy. It defines the agent and the environment as two partially open systems, in dynamic coupling. In CAES, the agent is composed of two sub-systems, mind and body, following the principles of situativity and intrinsic motivation.

(23) CALM (*Constructivist Learning Anticipatory Mechanism*) is based on the constructivist approach to Artificial Intelligence. It allows a situated agent to build a model of the world in environments partially deterministic and partially observable in the form of *Partially Observable and Factored Markov Decision Process* (FPOMDP). The model of the world is constructed and used for the agent to define a policy for action in order to improve its own performance.

(24) **Keywords:** Constructivist Artificial Intelligence, Artificial Intelligence, Machine Learning, Autonomous Agents, Concept Induction, Artificial Cognitive Development, Piaget, Structure Discovering in FPOMDPs, Markov Decision Processes (MDP).

1. Introduction

(25) La technologie est historiquement l'un des grands moteurs de la transformation de notre mode de vie. Bien qu'elle suscite un grand débat politique, le fait est que la technologie ouvre des nouvelles possibilités et permet graduellement à la société de produire plus, en travaillant relativement moins.

(26) De nos jours, la technologie trouve ses limites dans des domaines qui requièrent autonomie, intelligence et adaptation dynamique. Dans l'ère de l'information, ce sont précisément les tâches qui gagnent en importance et qui nécessitent encore presque intégralement le travail humain.

(27) Tant pour la robotique que pour le développement de logiciels qui peuplent des environnements virtuels, le principal défi est le même: créer des agents artificiels capables d'apprendre à agir d'une façon appropriée dans des univers complexes, en présentant de la flexibilité, de l'initiative, de la créativité et de l'entendement. Il n'est pas

possible de gérer avec des techniques d'ingénierie les nombreuses possibilités, les déroulements et les situations inattendues qui se posent dans ce type de domaine.

(28) La récente révolution technologique n'a pas seulement permis de confier aux machines les travaux bruts de la production industrielle, mais aussi les opérations de précision, l'analyse des données, le traitement de l'information, etc. Il y a de nombreux domaines et problèmes qui sont efficacement traités par l'utilisation de l'informatique. Cependant, dans presque tous les cas, l'ordinateur est utilisé comme un outil sous le contrôle humain. Même là où il y a quelque type d'intelligence artificielle, on voit peu ou pas d'autonomie de la machine, des fortes restrictions du domaine d'application du système, qui est généralement limité à des problèmes très spécifiques dans des environnements bien contrôlés. C'est précisément quand on exige un certain niveau de compréhension et de discernement que les machines actuelles échouent.

(29) Dans ce contexte, l'accent étant mis sur les technologies de l'information, il devient encore plus important que la communauté scientifique de l'Intelligence Artificielle (IA) concentre ses efforts à la recherche de systèmes plus robustes et plus autonomes. Pour trouver des réponses à ces questions, on doit, sans doute, plonger dans la recherche de base sur des mécanismes d'intelligence artificielle générale, ce qui passe par une révision paradigmatique de la discipline.

(30) Dans le cas de l'intelligence artificielle, en plus de la motivation pratique pour le développement de ces technologies, il y a aussi une question d'un plan plus théorique et philosophique. L'intelligence est l'une des caractéristiques les plus remarquables de notre identité en tant qu'êtres humains, mais elle est néanmoins un phénomène encore mal compris. Les questions sur quelle est la nature et comment marche l'intelligence sont présentes dans plusieurs disciplines sous plusieurs aspects, de la philosophie, en passant par la psychologie, les mathématiques, la biologie et la sociologie, et pour laquelle l'IA a une contribution importante à offrir. Selon les mots de McCorduck (1979): « *faire progresser l'IA en tant que science signifie aussi aider à découvrir quelque chose d'important sur nous mêmes* ».

1.1. Posture Épistémologique

(31) Puisque cette étude touche aussi des questions de nature philosophique sur l'intelligence artificielle, il nous semble approprié de déclarer dès le début notre position épistémologique, de façon à éclaircir la conception d'IA qu'on adopte et la conséquente ligne d'évolution du travail qui en découle. Elle prendre pour principe qu'il est possible de produire artificiellement des formes authentiques et véritables d'intelligence.

1.1.1. L'IA en tant que discipline scientifique

(32) Si on considère l'histoire de l'Intelligence Artificielle, il est possible d'identifier trois grandes périodes (McCORDUCK, 1979), (SIMONS, 1984), (CREVIER, 1993), (COELHO, 1996), et (RUSSELL; NORVIG, 2005): (1) à ses débuts dans les années 1940 et 1950, l'IA naît en tant que science entourée de promesses grandioses; (2) ensuite, le manque de résultats convaincants, finit par réorienter la recherche vers un discours plus réticent et moins ébloui, en essayant de remettre les pieds sur terre et de présenter des résultats concrets; (3) enfin, à la suite de ses propres progrès et aussi à cause des nouvelles découvertes provenant de ses frontières interdisciplinaires, l'IA reprend peu à peu, dans les décennies 1990 et 2000, ses idéaux et ses ambitions originelles.

(33) L'IA a été créé en tant que discipline scientifique au cours des années 1950. Il y avait, à ce moment historique, des circonstances propices pour son émergence (McCORDUCK, 1979). D'une part, le désir de comprendre le phénomène de l'intelligence, et l'inquiétude de la pensée humaine moderne. En outre, l'émergence et le développement des ordinateurs, en fournissant, dans le domaine académique, la rencontre des sciences humaines, biologiques et mathématiques, avec la naissante et prometteuse science de l'informatique.

(34) Les premiers temps d'existence de l'IA ont donné lieu à des projets ambitieux, analogues aux images créées par la science-fiction. A cette époque, les chercheurs, pris par une grande sensation d'optimisme, consécutive des premiers succès, se sont chargés d'envisager des possibilités futures, et d'établir les grands objectifs et rêves de l'IA.

(35) Dans les années qui ont suivies, les difficultés, les échecs, et l'absence de résultats convaincants, aggravés par le manque d'outils et de connaissance de la

cognition, ont frustré l'optimisme initial. L'engagement des chercheurs ne suffisait pas à éviter le fait que reproduire dans l'ordinateur le phénomène de l'intelligence était une tâche beaucoup plus difficile qu'ils ne l'avaient supposé (McCORDUCK, 1979).

(36) Dans ce contexte, un débat sur les principes fondamentaux de l'IA était devenu inévitable, et qui s'est traduit par deux différentes postures (HAUGELAND, 1985): d'un côté, les *enthousiastes*, pour qui l'avènement des ordinateurs dotés de vrais esprits n'était qu'une question de temps et, d'un autre côté, les *moqueurs*, pour lesquels une telle idée semblait irréaliste, voire ridicule.

(37) Ces deux conceptions opposées sont devenues connues sous les termes « IA Forte » et « IA Faible » (SEARLE, 1980). Du côté de l'IA Forte se sont placés les chercheurs qui estimaient que leurs programmes d'ordinateur, même rudimentaires, étaient en fait intelligents. Du côté de l'IA Faible se sont placés ceux qui disaient qu'une machine pouvait, au maximum, simuler des comportements qui semblaient intelligents de l'extérieur.

(38) Les pionniers de l'IA croyaient que les principes de l'intelligence étaient déjà présents dans leurs modèles, et qu'il faudrait tout simplement trouver les bons paramètres. Ils s'appuyaient sur les analogies du cerveau semblable à un ordinateur, et de l'esprit semblable à un logiciel. Cependant, leur insistance à affirmer que le modèle du traitement de l'information (calcul algorithmique) était en fait une bonne métaphore pour la pensée humaine et donc pour l'intelligence, a donné de la force aux critiques comme Searle (1980), Dreyfus (1972, 1992), et Penrose (1989), de postuler l'impossibilité de la réalisation d'une véritable intelligence artificielle.

(39) On peut ajouter dans cette analyse le fait que les pionniers de l'IA sont issus, dans leur majorité, des sciences mathématiques et physiques et, en général, soit ils n'avaient pas une grande culture en sciences humaines (psychologie, philosophie), soit ils sous-estimaient, consciemment ou inconsciemment, les résultats de ces disciplines. Ils n'ont donc pas pu tirer solidement parti des résultats déjà obtenus et des théories déjà formulées.

(40) Dans la pratique, la nécessité de parvenir à des résultats concrets a fait que l'IA est passée par un processus de fermeture et de spécialisation: la grande majorité des scientifiques ont tourné leurs efforts vers l'amélioration des techniques et des solutions.

Le désir original de percer les mystères généraux de l'intelligence a été remplacé par des investigations spécifiques et moins ambitieuses. Cette fermeture a atteint son sommet dans les années 1970, lorsque le pragmatisme a pris la place des rêves des années 1950 et 1960, et a conduit les chercheurs à la recherche de réalisations (COELHO, 1996).

(41) Le pragmatisme vécu par l'IA au cours de cette période a eu son côté positif, lorsque finalement les connaissances acquises pour la discipline ont commencé à être, peu à peu, utilisées dans la pratique. Le travail a progressé mais le débat a été relégué au second plan, au point qu'à partir des années 1980 les projets liés à des modèles généraux d'IA avaient même une mauvaise réputation (PENNACHIN; GOERTZEL, 2007). Comme un effet collatéral, ce pragmatisme a désarticulé l'IA en tant que grand projet de compréhension de l'intelligence, et a disposé toute une génération de scientifiques absents des discussions épistémologiques, et sceptiques par rapport aux rêves posés par les pionniers (COELHO, 1996).

(42) Cependant, le domaine de l'intelligence artificielle a vécu, pendant ces dernières décennies, un développement ininterrompu et accéléré de ses recherches, allié au surgissement de nouvelles conceptions théoriques, modèles et techniques. Tout cela ajouté à l'augmentation de la capacité et de la vitesse des ordinateurs a permis une reprise progressive de projets d'IA à des fins générales (PENNACHIN; GOERTZEL, 2007), (FRANKLIN, 2007).

1.1.2. IA Générale

(43) À partir des années 1990, les termes « IA Forte » et « IA Faible » sont tombés en désuétude. Il a été reconnu que les modèles classiques d'IA étaient limités, et donc la vieille IA ne pouvait pas arriver à être une IA forte. Le débat, toutefois, n'a pas été laissé de côté. L'ambition de construire des artefacts qui peuvent avoir une intelligence authentique a persisté, désormais appelée « IA Générale », par opposition à l' « IA Restreinte » (PENNACHIN; GOERTZEL, 2007), (FRANKLIN, 2007).

(44) Les systèmes informatiques destinés à présenter de l'intelligence appliquée à des problèmes limités et spécialisés font partie de l'IA Restreinte. Sont inclus dans cette catégorie, par exemple, des programmes qui jouent aux échecs, qui font des diagnostics, qui conduisent des dispositifs mobiles à travers des espaces cartographiés, qui

apprennent à classer des courriers électroniques, etc. Différemment, les projets d'IA générale visent à développer des mécanismes capables d'agir de façon autonome et intelligente, indépendamment du domaine ou de la tâche qui leur sera présentée.

(45) Ces deux positions épistémologiques pour l'intelligence artificielle (l'IA Générale et l'IA Restreinte), vivent côte à côte dans les laboratoires, partagent les mêmes outils théoriques et pratiques, mais divergent par rapport à leurs ambitions et présuppositions philosophiques.

(46) L'*IA Restreinte* correspond au savoir-faire scientifique de la grande majorité des chercheurs du champ, c'est-à-dire la recherche en informatique consacrée à la solution des problèmes qui exigent des méthodes intelligentes, sans forcément se préoccuper des déroulements théoriques et philosophiques par rapport à la création d'une véritable intelligence artificielle. Il s'agit donc de reproduire les stratégies de l'intelligence afin que les machines puissent résoudre certains problèmes pour lesquels ne suffit pas seulement la puissance de calcul, mais qui exigent de meilleures stratégies de résolution.

(47) D'autre part, l'*IA Générale* n'a pas d'intérêt à tout simplement mettre en œuvre des méthodes ingénieuses pour résoudre des problèmes appliqués. Les scientifiques impliqués dans cette ligne sont concernés par les questions sur l'esprit et l'intelligence. Leur recherche a comme présupposé l'idée que l'intelligence, même s'il s'agit d'un phénomène extraordinaire et complexe, est susceptible d'être comprise et systématisée. L'IA Générale soutient l'hypothèse que les machines et les systèmes artificiels peuvent, en fait, disposer de l'intelligence comme une caractéristique qui leur est propre. Sa plus grande entreprise est d'utiliser l'ordinateur comme un outil pour aider à percer les mystères de l'intelligence en général, en développant des mécanismes artificiels qui montrent un comportement intelligent comme une conséquence inhérente à son propre fonctionnement.

(48) Bien que ce défi ne soit pas nouveau, ce n'est que récemment que la communauté scientifique d'IA lui consacre des efforts continus, et chaque fois plus grands, à la poursuite de mécanismes généraux d'intelligence artificielle (PENNACHIN; GOERTZEL, 2007). Des questions fondamentales sont de nouveau mises en évidence, comme la question de savoir comment un agent informatique peut apprendre par ses propres expériences, de façon autonome, à partir d'un cycle d'interactions

sensorimotrices avec son environnement, en construisant et reconstruisant des hypothèses sur les phénomènes qu'il observe, en identifiant des régularités, en produisant son propre langage de représentation, en développant des concepts abstraits et temporels sur sa réalité, et en intégrant les connaissances construites dans un réseau de systèmes conceptuels.

(49) Jusqu'à présent, l'IA n'a pas réussi à construire un mécanisme d'intelligence générale convaincant. Néanmoins, malgré les immenses difficultés connues, il n'existe pas de preuve visant à discréditer la possibilité d'une telle réalisation. Au contraire, l'intérêt croissant qui se vérifie pour le domaine de l'IA général tend à accréditer la pertinence de ce domaine de recherche.

(50) Notre travail en particulier, s'ajoute aussi à ce mouvement. Nous croyons en la viabilité de l'entreprise d'une IA générale, et nous partageons ses suppositions et ambitions. Nous espérons que cette thèse pourra constituer une contribution, et représentera un pas en avant à la poursuite des mécanismes artificiels vraiment intelligents.

1.1.3. Objectifs et Organisation de la Thèse

(51) L'objectif premier de cette recherche est de parvenir à une définition du concept d'*agent autonome*. Il ne s'agit pas de rééditer une discussion sur des termes de sens commun, mais de reconstruire la relation entre l'agent et l'environnement selon les principes de deux paradigmes qui, pendant ces dernières années, ont gagné le respect et la reconnaissance entre les chercheurs de l'IA: l' « IA Affective » et l' « IA Située ».

(52) Ainsi, la première partie de cette thèse est consacrée à un large examen de ce sujet. Le chapitre 2 analyse la notion d'agent autonome, et présente les principales discussions promues par des travaux récents sur ce thème. En tant que contribution, le chapitre 2 décrit l'architecture CAES (*Coupled Agent-Environment System*), une architecture générale pour la description des systèmes fondés sur la dichotomie agent-environnement, en les définissant comme des systèmes partiellement ouverts et interdépendants, et qui représente une unification des visions située et affective de l'IA.

(53) Le deuxième - et plus important - but de cette thèse est de proposer une nouvelle solution au problème de l'apprentissage de modèles du monde. Il s'agit de fournir à un

agent artificiel un mécanisme capable de lui faire construire progressivement une description des régularités de l'environnement dans lequel il est inséré. Cette thèse adopte le *modèle constructiviste* comme base théorique d'un mécanisme de ce type.

(54) Ainsi, le chapitre 3 donne un petit aperçu de la théorie constructiviste du côté psychologique, et fait référence à quelques travaux en IA liés au paradigme constructiviste. Le mécanisme CALM (*Constructivist Anticipatory Learning Mechanism*) est alors présenté en tant que contribution au problème de l'apprentissage de modèles du monde, en disposant de la capacité d'apprendre des régularités sensorimotrices. Le chapitre 3 présente encore une définition détaillée des problèmes d'apprentissage, en se rapportant au cadre des processus de décision markovien (MDP) et ses variations.

(55) En particulier, le mécanisme CALM s'applique à l'apprentissage incrémental de la structure d'un MDP factorisé et partiellement observable. S'acquitter de cette tâche implique aussi de résoudre le problème de la *sélection de propriétés pertinentes*, et le *dilemme entre exploration et exploitation*. De plus, une fois construit le modèle du monde, l'agent doit pouvoir l'utiliser pour agir d'une manière adaptée à l'environnement, ce qui conduit à un problème de décision séquentielle. Tous ces sous-problèmes sont également discutés dans le chapitre 3.

(56) Toujours dans le chapitre 3, on décrit la manière dont le mécanisme CALM traite les *environnements partiellement observables*, par l'induction de nouveaux éléments de représentation. Cette fonctionnalité permet à l'agent de découvrir des propriétés cachées de l'environnement ou même de représenter des concepts abstraits. Cette restriction (l'observation partielle) rend le problème beaucoup plus difficile.

(57) Enfin, au chapitre 4, nous présentons des exemples et des résultats expérimentaux en utilisant les modèles définis dans la thèse (CAES et CALM) dans différents scénarios, en montrant la capacité de convergence, et en faisant des comparaisons avec des travaux de même type. Finalement, nous récapitulons et analysons en conclusion nos propres contributions.

1.2. Constructivisme et Cybernétique

(58) Avec l'objectif de définir l'architecture CAES et le mécanisme CALM, cette thèse aborde quelques sujets pertinents dans le champ de l'*Intelligence Artificielle*. D'une part, la recherche d'une solution constructiviste, et donc la rencontre avec d'autres travaux d'inspiration semblable, pour faire place à ce qu'on pourrait appeler le *paradigme constructiviste en IA*. De l'autre part, l'inévitable incursion à travers des modèles de l'*IA Située*, suivie d'une analyse sur la composition du rapport agent-environnement, qui remonte à des discussions cybernétiques classiques.

(59) Ce débat de caractère théorique est nécessaire pour tout travail qui se prétend impliqué dans un projet d'IA général, mais sans se perdre dans des généralités. Il est alors nécessaire de procéder à une étude en profondeur, qui, dans le cas de notre recherche, se trouve dans le domaine de l'apprentissage dans des processus de décision markoviens. Dans ce contexte, nous devons connaître l'état de l'art des travaux liés à la problématique de la construction de modèles du monde, aussi bien que ceux liés aux problèmes de décision séquentielle.

(60) L'étape décisive qui nous permettra de rendre ce travail pertinent et original, cependant, est la prise en compte du problème de l'apprentissage dans des environnements partiellement observables. Ce n'est pas seulement parce que le degré de difficulté de ce genre de problème est plus grand, mais c'est aussi parce que la capacité d'induire et de représenter des propriétés non observables de l'environnement est une façon de travailler avec des concepts abstraits, et constitue un moyen pour parvenir à des formes plus robustes d'intelligence artificielle.

1.2.1. Intelligence Artificielle Constructiviste

(61) Une des caractéristiques les plus importantes de l'intelligence est la capacité d'apprendre. Toutefois, le processus d'apprentissage est complexe et il se manifeste sous de multiples formes, et il faudra encore beaucoup de recherche dans divers domaines de la science pour qu'il soit complètement compris.

(62) La théorie psychologique constructiviste (PIAGET, 1936, 1937, 1945, 1947, 1964, 1967, 1975), (MONTANGERO; MAURICE-NAVILLE, 1994), (BODEN, 1979), (FLAVELL, 1967), (COHEN, 1983) formule des explications bien acceptées sur le

processus d'apprentissage. Selon cette théorie, le grand potentiel de l'intelligence humaine est la capacité de transformer leurs propres structures intellectuelles, en les améliorant progressivement, dans un processus de complexification, gouverné par un besoin constant d'assimiler le monde avec lequel il interagit.

(63) Le constructivisme propose la notion de *développement cognitif*, en élargissant le concept d'*apprentissage*. Le développement cognitif est le processus qui conduit le sujet à la construction de nouveaux modèles de compréhension, à la création de nouveaux outils intellectuels, à l'intensification de l'élaboration des structures de la connaissance, à l'enrichissement des formes de représentation, pour rendre possible un traitement efficace des expériences complexes.

(64) Selon le modèle constructiviste, le potentiel de l'intelligence humaine découle du passage d'une intelligence sensorimotrice à une intelligence symbolique. C'est-à-dire que le sujet surpasse la frontière de la compréhension basée uniquement sur des sensations et des mouvements, pour organiser désormais ses expériences dans un univers de relations, de concepts, et d'objets abstraits.

(65) Les problèmes d'apprentissage automatique se sont posés pour l'IA depuis son début. On dit qu'un système artificiel apprend s'il est capable de se transformer adaptativement (SIMON, 1983). Cependant, l'approche constructiviste de l'IA n'a été définitivement établie que dans les années 1990. La plus importante référence est le travail précurseur de Gary Drescher (1991), qui utilise le modèle constructiviste pour proposer un mécanisme capable d'apprendre de façon autonome à réaliser des tâches complexes.

(66) La théorie constructiviste est bien acceptée par les chercheurs dans le domaine de la psychologie du développement. Il s'agit d'un modèle philosophiquement consistant, qui fournit une description détaillée du processus cognitif, fondée sur des bases expérimentales, et qui surpasse les théories innéistes de l'intelligence ainsi que les théories empiristes. Piaget et son groupe ont étudié les processus d'apprentissage par l'observation de milliers d'enfants, de la naissance jusqu'au début de l'âge adulte.

(67) L'approche constructiviste de l' intelligence artificielle reprend le dialogue avec une théorie psychologique de grande portée, en resynchronisant la recherche en IA avec un travail mené à terme dans le champ de la psychologie.

1.2.2. Relation Agent-Environnement

(68) Une autre raison qui fait du paradigme constructiviste une proposition intéressante pour l'IA est la possibilité de réconciliation entre les processus biologiques de l'intelligence et les processus psychologiques de haut niveau, comme le raisonnement et la représentation. La théorie de Piaget établit que, en termes fonctionnels, il y a une continuité entre le biologique et le psychologique, par rapport à l'adaptation du sujet à son milieu.

(69) Plusieurs études, entre autres (VARELA et al., 1991), (BEER, 1995, 2004), (BICKHARD, 2000, 2009), (BARANDIARAN; MORENO, 2006, 2008), (RUIZ-MIRAZO; MORENO, 2000, 2004), (CLANCEY, 1997), (CLARK, 1998), (FROESE; ZIEMKE, 2009), (QUINTON et al., 2008), font valoir la nécessité d'utiliser, en IA, des architectures d'agent artificiel plus proches des modèles naturels, organiques, et, enfin, de repenser la relation entre l'agent et de l'environnement.

(70) Le problème du paradigme classique de l'*IA Symbolique* est son cognitivisme excessif, basé sur d'autres représentations que l'idée d'un agent situé dans un environnement. En général il s'agit d'un agent sans corps, qui ignore les notions de régulation dynamique et de rétroaction sensorimotrice, ce qui pose une grave question sur l'origine de la signification de ces représentations, connue comme le problème de l'ancrage des symboles (*symbol grounding problem*), (SEARLE, 1980), (HANARD, 1990).

(71) Certaines tentatives radicales pour éliminer le problème de l'ancrage des symboles ont conduit à des propositions antireprésentationnistes, où toute la cognition serait située dans le couplage entre l'agent et de l'environnement, c'est-à-dire le substrat physique et biologique de cette relation, d'où émergeraient des comportements dynamiques (BROOKS, 1991), (VAN GELDER, 1998). Toutefois, cette posture radicalement inverse est aussi fragile, dès qu'il s'agit d'expliquer les phénomènes cognitifs et comportementaux de plus haut niveau.

(72) D'un point de vue philosophique, le paradigme constructiviste de l'IA se propose de surpasser à la fois le modèle symbolique classique et le modèle antireprésentationniste. La théorie constructiviste est en ligne avec les notions de situativité (*situativity*) et d'incarnation (*embodiment*), en décrivant un mécanisme de

développement cognitif dont les structures plus abstraites ou symboliques sont construites à partir des interactions sensorimotrices plus simples, par un processus graduel de complexification de l'intelligence.

(73) Dans le même temps, l'un des plus importants défis à relever par l'IA est précisément de surmonter les limites de la perception sensorielle directe. Les phénomènes intéressants du monde, dans la plupart des cas, sont liés à des processus, des propriétés et des objets qu'on pourrait dire « macroscopiques » (THORNTON, 2003). Ainsi, un agent artificiel vraiment intelligent doit être capable de voir le monde à partir de concepts d'un niveau supérieur d'abstraction, formulés d'une manière autonome. À la limite, il pourra créer des notions et des théories à la fois plus généraux et plus adaptés pour décrire la réalité dans laquelle il vit.

1.3. Défis d'Apprentissage

(74) En intelligence artificielle, le problème général de l'apprentissage est décomposé dans une série de sous-problèmes bien délimités. Dans cette thèse, nous cherchons à en traiter simultanément quelques uns.

1.3.1. Apprentissage de Modèles du Monde

(75) Le principal problème abordé dans cette thèse est l'*apprentissage de modèles du monde*. Lorsque l'agent ne connaît pas les règles qui déterminent la dynamique de fonctionnement de son environnement, il faut qu'il les découvre peu à peu en se basant sur ses propres observations. Apprendre un modèle du monde signifie, pour un agent, construire de façon autonome une représentation interne de la dynamique d'interaction avec l'environnement à partir de son expérience. Particulièrement, dans une approche constructiviste, cette apprentissage doit être fait de façon progressive, où un modèle précédent plus grossier et moins adapté est graduellement raffiné.

(76) En général, l'entrée pour ce type de problème est un flux ininterrompu de *perceptions* successives faites par l'agent à travers ses senseurs, décrites dans l'espace défini par les propriétés de l'environnement qu'il est capable d'observer, parallèlement à un flux d'actions exécutées par l'agent à travers ses actuateurs. La tâche de l'algorithme d'apprentissage de modèles du monde est de conduire à l'induction d'une structure telle

qu'il soit possible de prédire les perceptions futures, en se basant sur les perceptions et les actions actuelles. Un modèle du monde est donc un modèle anticipatoire, qui décrit, du point de vue de l'agent, la régularité des transformations des propriétés de l'environnement au fil du temps, en fonction de l'observation qu'il fait et des actions qu'il exécute.

(77) Le problème de l'apprentissage de modèles du monde lui-même comprend un autre sous-problème crucial, qui est la *sélection des propriétés pertinentes*. Dans des environnements complexes, il y a une grande quantité de perceptions impliquées dans la description des états. Si l'environnement est bien structuré, alors seulement une petite portion de ces propriétés est pertinente pour décrire, à chaque fois, la dynamique des transformations.

1.3.2. Apprentissage des Concepts

(78) Lorsque on utilise une architecture à base d'agents, l'apprentissage des concepts peut être considéré comme la première étape vers la construction des nouveaux éléments de représentation. Le concept, au sens classique du terme, est une sorte de description généralisée des expériences. Il permet à l'agent de formuler et raisonner sur des propositions qui se rapportent à des catégories de situations, évitant ainsi des descriptions exhaustives, basées sur la référence a des états énumérés, qui sont définis à partir des combinaisons possibles des signaux de la perception.

(79) L'idée de « l'apprentissage des concepts » est bien établie dans la communauté scientifique de l'IA, ayant été déjà étudiée par plusieurs approches. L'énonciation classique du problème définit « concept » comme une description qui partitionne l'espace de caractéristiques. Il est appelé « cluster » dans le cas de l'apprentissage non-supervisé, ou « classe » dans le cas de l'apprentissage supervisé.

(80) Les clusters sont des groupes d'entités similaires, c'est-à-dire d'entités qui ont un patron de caractéristiques communes. Les clusters peuvent être considérés comme des régions continues dans un espace de caractéristiques à forte densité relative d'objets, séparés des autres clusters par des régions contenant une faible densité de points. Ainsi, le problème de clustérisation est d'analyser les patrons de caractéristiques des exemples

donnés, et leur distribution dans l'espace des caractéristiques, et de calculer une division pour cet espace.

(81) Par ailleurs, quand on parle de classes, on divise tous les échantillons qui apparaissent dans l'espace en sous-ensembles marqués avec l'étiquette de la classe. Alors que la clustérisation examine les patrons sur ses propres données d'entrée, la classification reçoit des exemples pré-classés. Le problème de la classification est d'analyser les caractéristiques de cet ensemble pré-classé d'exemples d'entraînement, puis d'induire une description généralisée pour chaque classe.

(82) Le problème est que, dans une analyse plus rigoureuse, la classification et la clustérisation en fait ne vont pas au-delà du niveau de la perception, car la description des classes et des clusters est formée en faisant directement référence aux éléments perceptifs. Ce type de « concept » est le résultat de combinaisons simples des propres caractéristiques fournies par la perception sensorielle, et en conséquence l'apprentissage de concepts dans ce sens classique ne fournit pas à l'agent la capacité de créer des éléments radicalement nouveaux de représentation.

1.3.3. Invention de Concepts Abstraits

(83) Nous distinguons le « concept » au sens classique en IA (classes et clusters), et ce que nous appelons « concept abstrait », pour parler d'une façon générale de tout ce qui est au-delà des références directement perceptives.

(84) Le problème fondamental de l'invention de concepts abstraits est la nécessité de développer des éléments de représentation radicalement nouveaux, des éléments qui désignent des entités autres que tout ce qui peut être représenté en fonction des perceptions. Les systèmes d'apprentissage ordinaires d'IA génèrent leurs concepts en utilisant des combinaisons, des spécialisations ou des généralisations des perceptions sensorielles directes (DRESCHER, 1991).

(85) En revanche, le développement cognitif exige des formes d'invention plus créatives. Il est nécessaire que l'agent puisse construire de nouveaux éléments, par le besoin d'organiser ses expériences dans des niveaux plus élevés. C'est la raison pour laquelle, dans une certaine étape du développement, les concepts ne peuvent plus être les produits de la combinaison des perceptions, mais, au contraire, ils doivent être des

éléments originaux qui réinterprètent les expériences dans des systèmes de compréhension plus abstraits.

(86) Ainsi, dans ce travail, nous traitons de ce que nous pensons être le prochain pas, au-delà de la formation de classes et de clusters, qui est l'induction de propriétés abstraites, traitée sous la forme d'une découverte d'éléments cachés. Dans un environnement partiellement observable il y a des propriétés essentielles à la modélisation de la dynamique des événements, mais qui, pourtant, ne peuvent pas être directement perçues par l'agent. Des environnements partiellement observables peuvent exhiber une dynamique apparemment arbitraire et non-déterministe en surface, même si elle est en fait déterministe en ce qui concerne le système sous-jacent et partiellement caché d'où provient la face perceptive des phénomènes (HOLMES; ISBELL, 2006).

(87) Tout type de classement des expériences basé uniquement sur la perception directe n'est pas capable de modéliser la dynamique d'un environnement partiellement observable. Pour faire face à ce genre de problème, l'agent a besoin d'un mécanisme plus robuste. Une solution est de doter l'agent d'un plus grand pouvoir de représentation, en l'autorisant à supposer l'existence d'éléments cachés (non-observables), lesquels, lorsqu'ils sont établis, lui permettront de construire un modèle du monde plus adéquat, et, par conséquent, augmenteront la capacité de l'agent à anticiper les événements.

(88) La découverte d'éléments cachés n'épuise pas le problème de l'invention de concepts abstraits, mais il est nécessaire de reconnaître que la possibilité de traiter des éléments cachés représente une avancée dans la voie entre la simple perception directe et des formes plus abstraites pour comprendre la réalité, et constituent une étape importante dans la recherche vers le développement des mécanismes d'intelligence artificielle générale.

1.3.4. Problème de Décision Séquentielle

(89) Le problème de la *décision séquentielle*, dans ce contexte, est de permettre à un agent d'utiliser son modèle du monde pour décider quelles actions il doit exécuter afin de maximiser sa performance. Cette optimisation est évaluée selon des dispositions émotionnelles et affectives internes, et dans une fenêtre de temps à long terme, ce qui

conduit l'agent à enchaîner une série d'actions visant à atteindre des objectifs qui ne sont pas immédiatement atteignables.

(90) Il s'agit de permettre à l'agent de définir une bonne politique d'actions pour l'environnement dans lequel il est inséré. Le problème de décision séquentielle est généralement traité en utilisant des algorithmes de planification et d'apprentissage par renforcement. Dans ce cas, au-delà de l'interaction normale avec l'environnement, représenté par les entrées sensorielles et les sorties motrices, l'agent reçoit un signal évaluatif qui l'informe si ce qu'il a fait a été bien ou mauvais, et il peut être modélisé comme un signal affectif interne, en indiquant des sentiments négatifs ou positifs pour l'agent.

(91) Au cours de son interaction avec le monde, l'agent, basé sur une observation continue du signal d'évaluation, a pour mission de construire une politique d'actions qui maximise la moyenne des récompenses reçues sur un horizon temporel plus grand que l'immédiat. Cette politique oriente la décision de l'agent en ce qui concerne les conduites qu'il prendra en fonction de la situation où il se trouve, envisageant non seulement de toucher des récompenses immédiates, mais aussi de réaliser une séquence de décisions planifiées pour obtenir une bonne performance à long terme.

(92) Apprendre un modèle du monde, trouver une bonne politique des actions, et en même temps, bien conduire ses activités dans l'environnement, c'est-à-dire en maximisant les signaux affectifs internes, exige d'affronter le *dilemme de l'exploration et de l'exploitation*. Dans ce cas, l'agent doit adopter une stratégie d'exploration, non seulement en planifiant des séquences d'actions pour atteindre des situations affectivement positives, mais également des actions qui le conduisent à des situations peu explorées, ou qui seraient intéressantes à explorer.

1.3.5. Processus de Décision Markoviens

(93) Dans le domaine de l'apprentissage automatique et de la planification, le problème de la décision séquentielle est généralement traité par la modélisation sous la forme d'un *Processus de Décision Markovien* (MDP), un formalisme classique, (BELLMAN, 1957), (HOWARD, 1960), et bien établi, (PUTERMAN, 1994), (SUTTON; BARTO, 1998), RUSSELL; NORVIG, 1995), (FEINBERG; SHWARTZ,

2002). Un MDP est un système qui évolue dans le temps comme un automate. À chaque instant le système est dans un état donné et il existe une certaine probabilité pour qu'il subisse une transition vers un autre état à l'instant suivant, en fonction de l'action décidée par l'agent qui a le contrôle.

(94) Quand les états de l'environnement ne sont pas directement accessibles à l'agent par sa perception, alors le système est modélisé à travers un MDP partiellement observable (POMDP), (SMALLWOOD; SONDIK, 1973), (CHRISTMAN, 1992), (SHANI et al., 2005), (KAELBLING et al., 1994, 1998). Dans ce cas, l'observation que l'agent contrôleur a du système ne le renseigne pas directement sur l'état où il se trouve.

(95) Récemment, l'orientation de la recherche s'est tournée vers les MDPs factorisés (FMDP), (BOUTILIER et al., 2000), (GUESTRIN et al., 2003), (JONSSON; BARTO, 2005), (SALLANS; HINTON, 2004), et aussi vers les POMDPs factorisés (FPOMDP), (GUESTRIN et al., 2001), (BOUTILIER; POOLE, 1996), (HANSEN; FENG, 2000), (POUPART; BOUTILIER, 2004), (POUPART, 2005), (WILIAMS, 2006), (SHANI et al., 2008), (SIM et al., 2008), où les états sont représentés dans l'espace défini par un ensemble de variables aléatoires.

(96) Bien que il y ait des algorithmes « libres de modèle » (ceux qui construisent directement une politique), le problème de la décision séquentielle est souvent considéré comme comportant le pré-problème de la construction d'un modèle du monde (RUSSELL; NORVIG, 1995). C'est-à-dire que, si l'agent n'a pas d'accès plein et direct à l'environnement et à ses règles de transformation, ni lui est fourni à l'avance un modèle de l'environnement - ce qui est habituellement le cas dans des problèmes réels – alors l'agent est obligé de construire un modèle du monde avant de calculer une politique d'actions.

(97) Les travaux liés au problème de l'apprentissage de modèles du monde ont convergé, récemment, à l'utilisation de FMDPs comme forme de représentation. Dans ce cas, construire un modèle du monde, c'est déterminer la structure et les paramètres d'un FMDP à partir de l'expérience de l'agent, comme le font (DEGRIS et al., 2006, 2008) et (STREHL et al., 2007).

1.4. Contributions

(98) Cette thèse présente deux contributions. La première est l'architecture **CAES** (*Coupled Agent-Environment System*), qui redéfinit la relation entre l'agent et l'environnement à la suite d'une discussion sur le concept d'agent autonome, sous une perspective naturaliste et basée sur la théorie des systèmes dynamiques.

(99) La seconde contribution est le mécanisme d'apprentissage **CALM** (*Constructivist Anticipatory Learning Mechanism*), qui joue le rôle de système cognitif dans un agent modélisé à travers l'architecture CAES, et qui vise à lui permettre de construire progressivement une description des régularités de l'environnement dans lequel il est inséré.

1.4.1. L'Architecture CAES

(100) L'architecture **CAES** (*Coupled Agent-Environment System*) définit l'agent et l'environnement comme deux systèmes partiellement ouverts, en interaction, et en couplage dynamique, où l'agent est autonome, situé, incarné, affectif, ainsi que mentalisé.

(101) L'architecture proposée définit un système global, dans lequel nous pouvons distinguer trois entités principales: (a) *l'environnement*, qui représente tout ce qui est en dehors de l'agent, (b) *le corps*, qui constitue l'univers intérieur de l'agent, en ayant des propriétés et une dynamique, et qui sert d'interface entre l'agent et l'environnement à travers ses senseurs et actuateurs, et enfin, (c) *l'esprit* de l'agent, qui est chargé de coordonner son comportement.

(102) De cette façon, il est établi que la seule possibilité pour l'agent est de réaliser un type de *cognition située*. D'abord, parce que l'accès que l'esprit a du corps et de l'environnement est médiatisé par des signaux partiels. L'esprit ne peut percevoir intégralement les caractéristiques extérieures, puisque la perception sensorielle est limitée, et également parce que ses actions peuvent contrôler seulement une partie des événements qui se produisent. De plus, l'agent est limité à un point de vue et à une localité.

(103) L'esprit de l'agent, dans l'architecture CAES, est composé de deux sous-systèmes: cognitif et régulatif. Le système cognitif est en fait celui qui apprend et

construit un modèle anticipatoire de la dynamique du monde qu'il observe, et est également responsable de la planification et délibération des actions. Le système régulatif, quant à lui, comprend à la fois les comportements réactifs et émotionnels, comme l'évaluation affective des expériences.

(104) La présence d'un système évaluatif, qui attribue des valeurs affectives aux événements, est un moyen d'internaliser la motivation de l'agent, en supprimant la notion de récompenses environnementales pour les actions. Le système évaluatif dessine une sorte de relief affectif sur l'espace de flux représenté au sein du système cognitif de l'agent, par son modèle du monde.

1.4.2. Le Mécanisme CALM

(105) Du point de vue psychologique, le mécanisme CALM (*Constructivist Anticipatory Learning Mechanism*) est proposé comme un *modèle systématisé de la psychologie constructiviste*. Il s'agit d'une version informatique, même si librement inspiré et non définitive, du processus de développement cognitif, tel qu'il a été décrit par Jean Piaget (1936, 1937, 1945, 1947, 1964, 1967, 1975) dans ses recherches menées entre les années 1930 et 1980.

(106) Du point de vue informatique, le mécanisme CALM se présente comme une solution au problème de *l'apprentissage de modèles du monde*, parce qu'il rend possible à un agent la construction d'une représentation de la structure et de la dynamique de son environnement, basée sur l'expérience. Le modèle du monde est représenté par un processus de décision markovien factorisé (FMDP).

(107) Le développement du mécanisme CALM a impliqué aussi l'utilisation des solutions aux problèmes de la *sélection de propriétés pertinentes*, du *dilemme de l'exploitation et de l'exploration*, et de la *décision séquentielle*.

(108) Par ailleurs, CALM est également capable de faire face à des *environnements partiellement observables* à travers l'induction de nouveaux éléments de représentation, qui peuvent être associés à des propriétés cachées de l'environnement ou même représenter des concepts abstraits. Dans ce cas, le modèle construit constitue un MDP en même temps factorisé et partiellement observable (FPOMDP).

(109) Le mécanisme fonctionne de manière *incrémentale* et *online*, puisque l'agent apprend en même temps qu'il a besoin d'interagir avec son univers. Comme restriction, il est supposé que l'environnement peut être décrit par des *fonctions discrètes*, par rapport au temps comme aux valeurs que la perception et l'action de l'agent peuvent prendre. Il est également supposé que l'environnement est *partiellement déterministe* par rapport à la fonction qui régit ses transformations.

2. CAES: Système de Couplage Agent-Environnement

(110) À partir des années 1990, il y a eu une résurgence progressive de projets en intelligence artificielle (IA) à des fins générales (PENNACHIN; GOERTZEL, 2007). Pour cela, il devenait nécessaire de réviser la notion d'agent autonome et le modèle de la relation entre l'agent et l'environnement. Des modèles simplifiés utilisés par le paradigme classique de l'IA Symbolique, excessivement cognitivistes, non-situés, se heurtent au problème de l'ancrage des symboles (*symbol grounding problem*), (SEARLE, 1980), (HANARD, 1990), au problème des *frames* (McCARTHY; HAYES, 1969), (PYLYSHYN, 1987), et encore d'autres problèmes de nature plutôt philosophique, en indiquant une certaine décalage entre les modèles et la réalité.

(111) En réponse à cela, plusieurs auteurs ont proposé de nouvelles architectures pour des agents, en utilisant des idées qui finalement ont aidé à constituer les paradigmes de de *l'Intelligence Artificielle Située* et de *l'Intelligence Artificielle Affective*. Dans cette même ligne, ce chapitre présente CAES (*Coupled Agent-Environment System*), une architecture pour le développement des agents autonomes qui essaye de composer les points forts des ces deux paradigmes.

(112) L'*IA Située* unifie la théorie des systèmes dynamiques avec les modèles organiques et naturels. Des travaux de référence tels que (BEER, 1995, 2004), (ASHBY, 1952), (VARELA et al., 1991), (BICKHARD, 2000, 2009), (FRANKLIN, 1997), (BARANDIARAN; MORENO, 2006, 2008), (CLANCEY, 1997), (CLARK, 1998), (QUICK et al., 1999), (FROESE; ZIEMKE, 2009), (ZIEMKE, 1998, 2002), (RUIZ-MIRAZO; MORENO, 2000, 2004), conçoivent l'agent comme un système dynamique partiellement ouvert, immergé dans son environnement, et en interaction avec lui à travers un flux continu d'interférences mutuelles, régies par deux facteurs principaux, *la régulation organique* et la *rétroaction sensorimotrice*.

(113) *L'AI Affective* internalise la motivation de l'agent. Des travaux comme (SINGH et al., 2004), (CAÑAMERO, 1997a, 1997b, 2001), (ALMEIDA et al., 2004), (SLOMAN et al., 1999, 2005), dépassent l'ancien modèle d'apprentissage comme conditionnement, trop comportementaliste, où c'est à l'environnement d'indiquer les signaux de renforcement par le biais de récompenses et de punitions exogènes. La nouveauté, c'est le déplacement du signal de renforcement à l'intérieur du corps de l'agent. L'idée est de modéliser des systèmes affectifs et émotionnels organiques, de sorte que la motivation à agir soit d'origine interne.

2.1. Relation entre l'Agent et l'Environnement

(114) L'architecture CAES définit l'agent et l'environnement comme deux systèmes partiellement ouverts et dynamiquement couplés, dans un cycle d'interaction continue, où l'agent est toujours une entité autonome mais inexorablement immergée dans l'environnement. Dans l'architecture CAES, l'agent est: (a) situé et incarné; (b) affectif et émotionnel (donc intrinsèquement motivé); et (c) cognitif et adaptatif. Ces qualités permettent de surpasser les modèles simplifiés qui sont habituellement utilisés par l'IA,

dans lesquels l'agent est réduit à un système de représentation et d'inférence (un « esprit détaché ») qui agit directement sur le monde.

(115) Un système construit sur l'architecture CAES présente explicitement trois types d'interaction, comme montré figure 2.1: (a) à l'extérieur, l'interaction entre l'agent et l'environnement; (b) ensuite, dans l'agent, l'interaction entre son corps et son esprit; (c) et enfin, dans l'esprit de l'agent, l'interaction entre le système cognitif et le système régulatif.

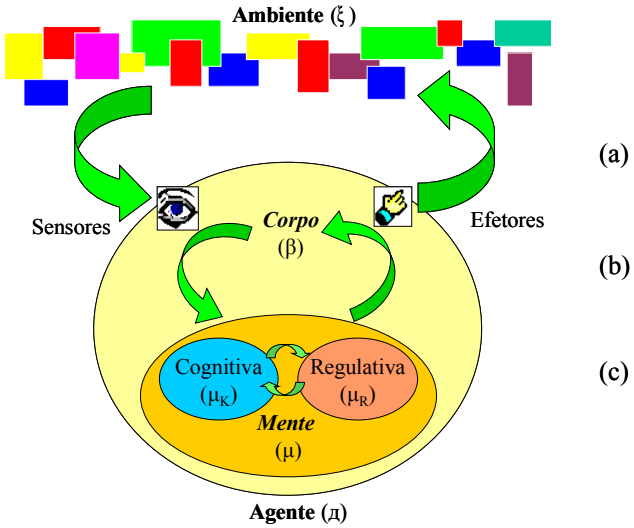


Figure 2.1: L'architecture CAES et ses trois niveaux d'interaction.

En (a) la relation agent-environnement, en (b) la relation corps-esprit, et en (c) la relation cognition-régulation.

(116) Ainsi, l'agent est vu, dans un premier niveau, comme un *objet* de l'environnement. A ce niveau, il n'y a que des phénomènes et des événements qu'on dirait d'une réalité physique. L'environnement est l'espace qui contient les objets, et qui impose à ces objets des règles d'interaction d'une nature particulière. L'agent, en tant qu'objet parmi les autres, participe également des caractéristiques, des événements et des phénomènes physiques de ce niveau de la réalité, qui sont établis au sein de ce système monde-objets.

(117) Dans un deuxième niveau, l'agent est un *organisme*, encore soumis aux phénomènes et aux événements physiques, mais en tant qu'objet spécial, l'agent possède alors un métabolisme corporel et dispose de propriétés internes. Ainsi, l'agent se configure comme un système partiellement ouvert, qui peut influencer, à partir de

processus intérieurs, la façon dont les phénomènes se produisent à l'extérieur, en modifiant localement la trajectoire du système global.

(118) Dans un troisième niveau, l'agent est une *entité mentalisée*, caractérisée par la présence d'une structure interne, l'esprit, qui régule son comportement. L'esprit reçoit des informations sensorielles de l'environnement et du corps, et coordonne des actions dirigées vers l'un et vers l'autre.

2.1.1. Interactivité, Autonomie et Situativité

(119) À partir des années 1980 l'IA a vécu l'émergence d'un nouveau paradigme, appelé *informatique basée sur des agents*. Depuis, l'idée d'« agent » est devenue populaire parmi les chercheurs du domaine (MAES, 1994), ce qui a apporté un nouveau modèle conceptuel. La définition la plus basique et de consensus pour le concept *d'agent* est celle d'une entité autonome insérée dans un environnement, qui le perçoit à travers un ensemble de senseurs, et qui agit sur lui à travers un ensemble d'actuateurs (RUSSELL; NORVIG, 1995).

(120) La dichotomie *agent-environnement* représente un modèle assez général, qui peut décrire différents types de systèmes, naturels ou artificiels, par exemple: un véhicule autonome qui se déplace sur une planète lointaine, un robot qui réalise une tâche domestique dans une maison, un programme d'ordinateur qui tourne dans un univers virtuel, un moteur de recherche intelligent qui navigue sur l'internet, un animal dans son habitat, une bactérie dans une solution chimique, une personne qui réalise une tache quotidienne dans le monde réel, etc. La dichotomie entre l'agent et l'environnement en informatique est similaire à celle entre l'organisme et l'environnement en biologie, ou celle entre je sujet et le milieu en psychologie.

(121) L'interactivité est la première caractéristique remarquable dans la relation entre l'agent et son environnement. Ainsi, le modèle d'informatique fondé sur des agents est distingué par la présence d'un cycle de fonctionnement en continu, au contraire de l'architecture traditionnelle, qui est dessinée par un flux linéaire du type « entrées, calcul et sorties », comme montré dans la figure 2.2. Pour certains auteurs, tels que (WEGNER, 1998), (COSTA; DIMURO, 2005), et (GOLDIN; WEGNER, 2008), cette

considération apparemment simple permet l'émergence d'un nouveau paradigme pour les sciences informatiques.

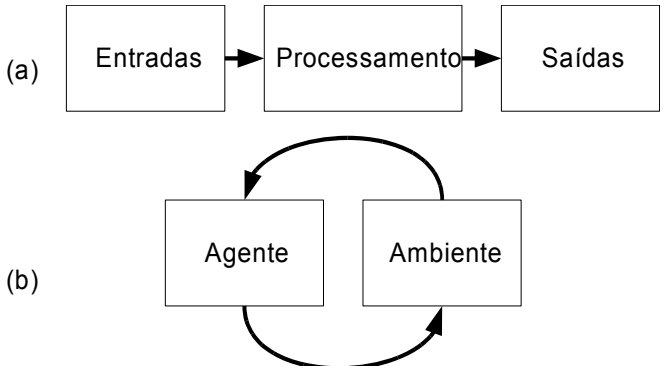


Figure 2.2: Informatique fondé sur des agents.

Différence conceptuelle entre (a) le modèle classique du calcul de données, et (b) le modèle d'agents.

(122) Le terme agent renvoie à l'idée de système partiellement ouvert. L'agent est encapsulé, il possède une structure interne, propre et finie, mais son fonctionnement dépend de l'interaction avec un environnement qui lui est extérieur (JENNINGS, 2000). Cette définition pose l'agent et l'environnement en tant que deux entités complémentaires. L'existence d'un agent implique l'existence d'un environnement corrélatif. Ce sont deux systèmes interdépendants qui constituent un système global.

(123) Dans des systèmes naturels, définir qui est l'agent et qui est l'environnement comporte un certain arbitraire de l'observateur, car il n'existe pas une frontière précise entre les deux. Cette distinction peut être fondée sur la cohésion organisationnelle de l'agent, qui constitue un système dense, relativement stable, et hautement intégré, par rapport au système global.

(124) De même, faire la distinction entre ce qui est un agent et ce qui est un objet de l'environnement implique aussi des critères subjectifs. En général, la caractéristique principale qui sert à distinguer un agent de ce qu'on pourrait appeler un simple objet est le fait que l'agent montre des comportements actifs. Il est un objet spécial de l'environnement, capable, à partir des éléments internes, de choisir une action parmi plusieurs alternatives pour une même situation (JENNINGS, 2000).

(125) Il est possible d'envisager la présence de divers agents dans un même environnement, comme le montre la figure 2.3, ce qui caractérise un système multi-agents. Ces divers agents peuvent présenter une architecture similaire ou peuvent être

très différents, formant alors une sorte d'écosystème. Du point de vue spécifique de chaque agent, les autres agents font partie de l'environnement, étant compris parmi les autres objets avec lesquels il interagit. Toutefois, c'est généralement l'existence de nombreux agents, en combinant des conduites autonomes, qui est le principal facteur d'augmentation de la complexité du système dans son ensemble.

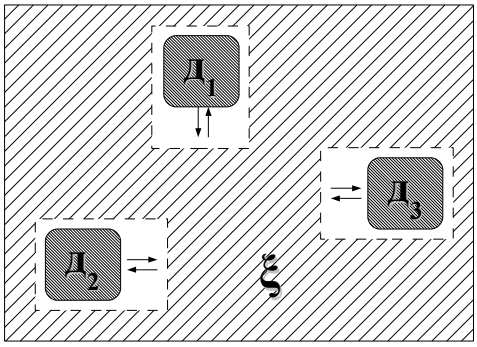


Figure 2.3: Un environnement ξ peuplé par plusieurs agents д.

(126) Un agent est une entité autonome, qui vit dans son environnement libre de contrôle externe (DAVIDSSON, 1995). Tous les processus impliqués dans les prises de décision et dans les initiatives d'action doivent avoir origine dans la propre structure interne de l'agent.

(127) Un agent autonome est une entité *située*. Ainsi, l'environnement dans lequel l'agent est inséré joue un rôle important dans la modélisation de l'agent lui-même. Les moyens par lesquels l'agent et l'environnement interagissent et provoquent des interférences mutuelles devient très important pour la caractérisation du système global (CLANCEY, 1997). L'agent est situé parce que ses décisions sont prises par rapport au contexte dans lequel il est, c'est-à-dire que les actions que l'agent effectue sont le résultat d'un processus interne de décision, mais effectué en fonction de paramètres provenant aussi (ou principalement) de l'environnement.

(128) La situativité (*situativity*) définit un modèle de cognition appelé « énaction ». La connaissance énactive est celle dérivée de l'activité, où le point de référence n'est pas un monde objectif et neutre, défini par un regard extérieur, mais est la propre structure sensorimotrice de l'agent (VARELA et al., 1991).

(129) L'interface entre l'agent et le monde est limitée. L'agent situé n'est ni *omnipotent* ni *omniscient* par rapport à l'environnement, c'est-à-dire que le monde n'est que partiellement captable par la perception sensorielle de l'agent, et que le monde n'est que

partiellement susceptible de se transformer par la force des actions de l'agent (SUCHMAN, 1987).

2.1.2. Couplage

(130) L'architecture CAES (*Coupled Agent-Environment System*) définit la relation entre l'agent et l'environnement comme le couplage de deux systèmes dynamiques, partiellement ouverts, et en interaction. Chacun de ces systèmes exerce une certaine influence sur le flux de transformation de l'autre, en le déformant continuellement. Il s'agit de perturbations mutuelles, où chaque système n'a pas de contrôle absolu sur la détermination de la trajectoire de l'autre, ni sur la sienne propre.

(131) Cette conception, appelée *modèle de couplage dynamique*, a été solidement présentée par Beer (1995, 2004), qui à son tour a repris la structure générale du *modèle cybernétique* proposé par Ashby (1952), revêtu aussi des idées issues du *modèle autopoïétique* développé par Maturana et Varela (1973, 1980).

(132) Le *modèle de couplage dynamique* représente une architecture récente et bien acceptée parmi les chercheurs, pour la modélisation de la relation entre un agent et son environnement. Cette approche est organisée autour d'un ensemble bien établi d'outils mathématiques, qui proviennent de la théorie des systèmes dynamiques. Dans ce modèle, le comportement intelligent est conçu comme une activité adaptative exercée par des agents situés.

(133) Le modèle de couplage dynamique s'inscrit dans un mouvement plus large, au sein de l'intelligence artificielle comme des sciences cognitives, (VARELA et al., 1991), (VAN GELDER, 1998), (CLARK, 1998), (MONTEBELLI et al., 2008), (QUINTON et al., 2008), qui associe les théories de la vie, de l'interaction, et de la cognition, à la théorie des systèmes dynamiques, en reprenant des notions cybernétiques, par opposition à des modèles trop symboliques, dans le sens représentationaliste classique.

2.1.2.1. Modèle Cybernétique

(134) La cybernétique, entre les années 1930 et 1960, a constitué le mouvement responsable de l'introduction de nombreux concepts importants dans la théorie de l'information et du contrôle, telles que les notions d' « équilibre homéostatique », « rétroaction » (*feedback*), et « autorégulation ». Par principe, la cybernétique se

propose de décrire une théorie générale des systèmes, qui peut être appliquée à la fois à des contextes naturels (physiques, biologiques ou même sociaux) et artificiels (machines, automates, simulations, etc.).

(135) Le concept d'homéostasie (CANNON, 1932) définit la propriété qu'à un système ouvert de réguler son environnement interne pour maintenir un état d'équilibre. Ce concept est inspiré des organismes biologiques, qu'ont des mécanismes (métaboliques ou comportementaux) pour conserver certaines variables physiologiques autour de « niveaux normaux ».

(136) La notion de rétroaction (WIENER, 1948) représente la causalité circulaire, où une entité agit sur une seconde, qui en retour agit sur la première. Le déroulement de cette idée conduit aux systèmes auto-régulés, qui modifient les conditions internes en les modulant par la rétroaction externe, afin de préserver l'équilibre homéostatique.

(137) Même si la théorie des systèmes dynamiques n'était pas encore établie à l'époque, Ashby (1952), dans un travail précurseur, a décrit la relation entre l'agent et l'environnement comme deux systèmes dynamiques en interaction (GRUSH, 1997b). Ashby affirme que l'organisme d'un animal n'est pas différent de celui d'une machine, en termes de système. L'organisme et son environnement, pris ensemble, forment un système complet. L'organisme influe sur l'environnement et l'environnement affecte l'organisme, constituant une relation de rétroaction mutuelle.

2.1.2.2. Modèle Autopoïétique

(138) Le concept d'*autopoïèse* constitue une nouvelle définition de « vie » (MATURANA; VARELA, 1980). Dans le modèle autopoïétique, fortement influencé par les notions cybernétiques, un être vivant est un système ouvert et autonome. Toutefois, l'accent est mis par le modèle sur l'« auto-production » du système. L'être vivant, en tant que système autopoïétique, non seulement change son comportement en interagissant avec l'environnement, mais aussi se transforme structurellement en raison de cette interaction. L'organisme est caractérisé par son organisation interne, qui est stable dans le temps.

(139) L'être vivant est dans un processus continu de transformation, et le cours de ce changement est modulé en partie par les interactions avec l'environnement. Souvent, des événements d'origine externe peuvent déclencher des changements dans l'organisme,

mais le sens de la transformation n'est pas déterminé par l'environnement, car il est inhérent à l'organisation interne de l'être vivant. C'est la structure de l'organisme qui établit le domaine des changements qui y ont lieu, et c'est aussi la propre structure de l'organisme qui établit les types de perturbation qui peuvent provoquer une transformation (MATURANA, 1993).

(140) La dynamique du monde exerce une pression qui tend à éloigner l'organisme de son état d'équilibre, et cette pression est une menace constante pour l'intégrité du système. L'autopoïèse représente la lutte de l'organisme pour lui permettre de continuer à vivre dans l'environnement, en compensant les perturbations extérieures à travers des changements structurels, de façon à assurer son auto-maintenance.

2.1.2.3. Systèmes Dynamiques

(141) La théorie des systèmes dynamiques fournit un outil mathématique bien structuré et largement accepté pour l'étude des systèmes déterministes dont l'état évolue au fil du temps, et qui peuvent présenter des comportements complexes (GLEICK, 1987).

(142) Un système dynamique est défini par un domaine, combiné à une fonction continûment dérivable qui décrit son évolution. À chaque instant *t*, cette fonction mappe un point du domaine à un autre dans ce même espace. Si le domaine est continu, on l'appelle un *espace de phase*, et s'il est discret, on parle d'un *espace d'états*. Quand le temps est continu, le système dynamique est représenté comme un *flux*, et quand il est discret, il est représenté comme une *application*. La fonction d'évolution définit un champ vectoriel dans l'espace de domaine, en indiquant le sens des trajectoires pour tous les points possibles.

(143) Dans un système dynamique, les *attracteurs* sont des régions de convergence, vers lesquelles le flux du système évolue lorsque il rentre dans l'aire d'attraction. Ces attracteurs peuvent avoir la forme d'un point fixe, d'un cycle limite, ou ils peuvent être complexes, délimitant une région vers laquelle le flux est attrapé, et où il reste en *fluctuation chaotique*. Si les paramètres de la fonction d'évolution du système dynamique varient au fil du temps, alors la configuration des attracteurs peut changer aussi, selon l'évolution du système.

2.2. Caractéristiques du Système Global

(144) L'architecture CAES constitue un modèle cybernétique et « naturaliste » de couplage, où l'agent et l'environnement sont deux systèmes dynamiques dans un cycle continu d'interaction, en exerçant des influences sur le flux de transformation de l'autre, et ainsi déformant continuellement ses trajectoires, mais sans avoir un contrôle absolu. Quelques paramètres de la fonction d'évolution d'un système sont des variables liées à l'état de l'autre, et donc la relation entre l'agent et l'environnement est établie par le biais de perturbations mutuelles. Le système global dans l'architecture CAES est illustré sur la figure 2.4 et il est formalisé par la définition 2.1, à la suite.

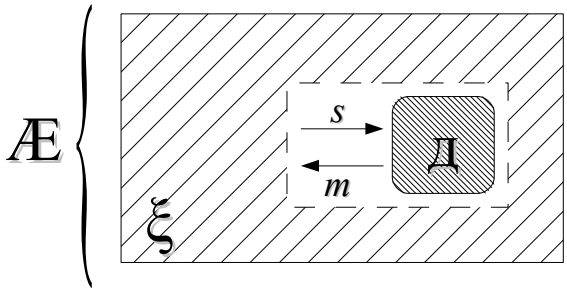


Figure 2.4: Le système global dans l'architecture CAES.

Le système global (Æ) est un système fermé, composé de deux sous-systèmes partiellement ouverts, l'agent et l'environnement, en couplage dynamique, où д est l'agent, ξ est l'environnement, m est l'actuation de l'agent sur l'environnement, et s est la situation qui s'impose à l'agent par l'environnement.

Un système couplé agent-environnement CAES (Æ) est un quadruplet:

$$Æ = \{д, ξ, s, m\}$$

où,

- д est un système dynamique partiellement ouvert *(agent)*
- ξ est un système dynamique partiellement ouvert *(environnement)*
- s est le signal qui définit l'influence de ξ sur д *(situation)*
- m est le signal qui définit l'influence de д sur ξ *(actuation)*

Définition 2.1: Système Couplé Agent-Environnement (Æ).

(145) L'environnement est représenté par un ensemble de variables $X_ξ = \{X_{ξ1}, X_{ξ2}, \ldots, X_{ξn}\}$, et par une fonction d'évolution $f_ξ$ qui établit le schéma de transformation des valeurs de ces variables. L'état de l'environnement au temps t est représenté par un vecteur $x^t_ξ \in X_ξ$. L'environnement (ξ) est tout ce qui est en dehors de l'agent, tandis que la totalité formée par *l'agent* et *l'environnement* constitue le système global (Æ).

(146) La fonction d'évolution de l'environnement ($f_ξ$) indique le sens de sa transformation dans l'avenir immédiat, selon son état actuel et l'intervention effectuée par l'agent, donc sous la forme $f_ξ : X_ξ \times M \rightarrow X_ξ$, où $M = \{M_1, M_2, ... M_n\}$ définit l'espace d'interférences possibles de l'agent sur l'environnement.

(147) De même, la fonction d'évolution de l'agent ($f_д$) indique la transformation de la situation interne de l'agent selon son état actuel et l'information provenant de l'état de l'environnement, sous la forme $f_д : X_д \times S \rightarrow X_д$, où $X_д$ définit l'espace (de phase ou des états) de l'agent en tant que système dynamique, et $S = \{S_1, S_2, ... S_n\}$ définit les possibilités d'influence de l'environnement sur l'agent.

(148) L'espace des états du système global est déterminé par le produit cartésien de l'espace de ses deux sous-systèmes, donc $X_Æ = X_ξ \times X_д$. De même, la fonction d'évolution du système global, $f_Æ : X_Æ \rightarrow X_Æ$, devient la combinaison des fonctions d'évolution de l'agent ($f_д$) et de l'environnement ($f_ξ$).

(149) Chaque action que l'agent effectue touche l'environnement d'une certaine manière. La transformation de l'environnement, à son tour, affecte l'agent, et définit ainsi un cycle permanent. Par conséquent, la fonction d'évolution d'un système, bien qu'elle soit paramétrée principalement par son propre état, est également dépendante de certains éléments relatifs à l'état de l'autre système.

2.2.1. Solipsisme Méthodologique

(150) La condition philosophique du *solipsisme méthodologique* (FODOR, 1981) affirme que l'expérience du monde est vécue de façon interne par le sujet, et donc la réalité extérieure ne peut être saisie objectivement. Elle apparaît comme une projection capturée par la perception. Il s'agit d'une affirmation similaire à celle du *phénoménisme analytique* (DANTO, 1989), (AYER, 1954), doctrine philosophique selon laquelle ne sont effectivement connaissables que les phénomènes, c'est-à-dire le contenu des perceptions spatio-temporelles, seul objet d'expérience possible.

(151) Ainsi, du point de vue de l'agent (д), l'environnement (ξ) est quelque chose d'objectivement inaccessible. Il est l'autre partie dans le rapport agent-environnement, celle qui concerne le monde. C'est le système avec lequel l'agent est en relation, sans avoir le contrôle ou l'accès direct aux facteurs qui régissent son fonctionnement. L'agent

ne peut pas connaître le monde tel qu'il est, il ne peut pas observer sa vraie constitution. L'environnement n'est accessible à l'agent que par l'intermédiaire de ses senseurs et ses effecteurs, et donc, la seule chose qui est pour l'agent connaissable c'est la relation d'interaction avec l'environnement, et non pas l'environnement lui-même.

(152) Un environnement inconnu quelconque peut être représenté d'une façon générale comme un système dynamique, constitué par un ensemble de variables (X_ξ) qui définissent ses dimensions, et par une fonction d'évolution (f_ξ) qui définit le patron de changement qui se produit dans les variables d'environnement selon son propre état, et selon les actions effectuées par l'agent, tel qu'il est défini dans l'architecture CAES.

(153) L'environnement le plus « naturel » qu'on peut imaginer est le monde réel, mais les agents informatiques peuvent également être insérés dans des mondes artificiels, où il est généralement possible pour l'expérimentateur de connaître le modèle de l'environnement et sa complexité. Toutefois, même dans un environnement simulé, où les règles et la constitution du monde sont connues par les développeurs, l'environnement est toujours objectivement inaccessible du point de vue de l'agent.

(154) Cet obstacle philosophique n'est pas une menace pour les algorithmes d'apprentissage de modèles du monde, de la même façon qu'il n'est pas un problème pour les êtres humains dans leur défi quotidien d'apprendre à interagir avec le monde réel, et d'anticiper les événements. Comme on le verra dans la section 3.1, le problème de l'apprentissage de modèles du monde peut être correctement formalisé, et les aspects pertinents de l'environnement peuvent être correctement représentés par l'agent même si la référence qu'il utilise est l'interaction entre lui et son environnement, et donc sans avoir besoin d'accéder à l'environnement « en-soi », sous-jacent à la relation.

2.2.2. Situation et Actuation

(155) L'agent intervient sur l'environnement par le biais de son actuation, un vecteur $m \in M$, généralement (mais pas uniquement) réalisé par ses actuateurs. L'actuation est dérivée de certaines propriétés de l'agent, sous la forme $m : X_д \rightarrow M$. L'autre direction de l'interférence est représentée par la situation, un vecteur $s \in S$, donné à l'agent par l'environnement, soit sous la forme de restrictions et de pressions, soit comme les

éléments capturés par les senseurs de l'agent. Ainsi, la situation est définie en fonction de certaines propriétés de l'environnement, sous la forme $s : X_{\xi} \rightarrow S$.

(156) L'agent ne perçoit le monde que de manière partielle, ce qui définit une condition de non-omniscience. Cela arrive, tout d'abord, parce que la situation que l'agent perçoit est une image appauvrie de la situation réelle, puisque les restrictions de son appareil sensoriel rendent impossible pour lui de saisir tous les aspects, les dimensions, ou les caractéristiques du monde. Le mécanisme sensoriel de l'agent définit une psychophysique, qui agit comme un filtre entre les nombreuses conditions physiques provenues de l'environnement, et les signaux sensoriels activés dans l'agent. Deuxièmement, la perception est modifiée par le point de vue de l'agent. En tant qu'une entité située, il est soumis à une sorte de filtre de localité. Par exemple, dans des environnements spatiaux, la position de l'agent par rapport aux autres objets change sa perception de la situation courante.

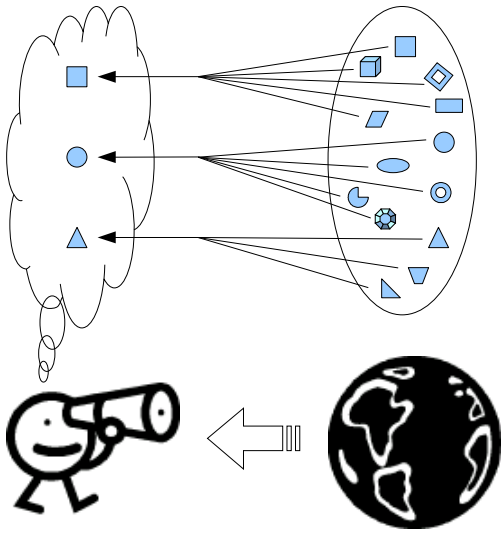


Figure 2.5: Condition de non-omniscience.

La perception comme une application surjective non réversible, dont l'ensemble de départ est formé des états physiques de l'environnement et l'ensemble d'arrivée des perceptions sensorielles de l'agent.

(157) L'agent n'a pas accès à l'identification complète de l'état de l'environnement, mais à une information qui est dérivée de cet état. Cette fonction de situation a pour ensemble de départ tous les états possibles du monde, et par ensemble d'arrivée, une collection plus petite d'observations possibles. Il s'agit d'une application surjective dont une réversibilité complète n'est pas possible, illustrée dans la figure 2.5.

(158) Ainsi, en raison de l'existence de ces limitations sensorielles, il est tout à fait possible que l'agent se trouve incapable de distinguer perceptivement entre différents états du monde qui lui apparaissent comme identiques (CROOK, HAYES, 2003). Cette confusion des états, où des situations différentes peuvent avoir une apparence indistincte, est aussi appelé *perceptual aliasing* (WHITEHEAD, BALLARD, 1991), (CHRISMAN, 1992), (WHITEHEAD, LIN, 1995), et définit des problèmes d'observabilité partielle.

(159) De même, dû au fait que les actuateurs sont également limités dans une certaine mesure, le contrôle que l'agent a sur les transformations de l'environnement n'est que partiel. Cette limitation existe, d'une part, car, s'agissant d'environnements spatiaux, le champ d'action de l'agent est restreint à un sous-espace local de l'environnement, en général à des régions qui lui sont proches. En outre, les modifications imposées sur l'environnement, bien que partiellement influencées par l'action de l'agent, sont également soumises à d'autres facteurs indépendants et externes à lui.

2.3. Caractéristiques de l'Agent

(160) Dans l'architecture CAES, l'agent est une partie dans un système de couplage dynamique entre l'agent et l'environnement. À son tour, l'agent lui-même est défini comme la composition des deux sous-systèmes, le corps et l'esprit, également couplés, comme illustre la figure 2.6, de qui est ci-après formalisé par la définition 2.2.

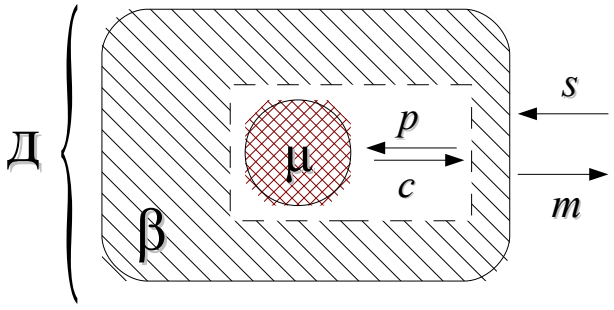


Figure 2.6: L'agent dans l'architecture CAES.

Un agent (д) est un système partiellement ouvert qui interagit avec son environnement. Il est composé de deux sous-systèmes, l'esprit (μ) et le corps (β), qui à leur tour, sont également liés en couplage dynamique.

Un agent (д) dans un système CAES (Æ) est un sextuplet:

$$\text{д} = \{\beta, \mu, c, p, s, m\}$$

où,

- β est un système dynamique partiellement ouvert *(corps)*
- μ est un système dynamique partiellement ouvert *(esprit)*
- c est l'interférence de μ sur β *(contrôle)*
- p est l'interférence de β sur μ *(perception)*
- s est un signal externe que l'agent д reçoit de l'environnement ξ *(situation)*
- m est un signal que l'agent д fournit à l'environnement ξ *(actuation)*

Définition 2.2: Agent (д).

(161) L'espace (de phase ou des états) de l'agent en tant que système dynamique est établi par le produit cartésien des espaces de ses deux sous-systèmes, $X_{\text{д}} = X_{\beta} \times X_{\mu}$. De la même façon, sa fonction d'évolution est issue de la combinaison des fonctions d'évolution du corps (f_{β}) et de l'esprit (f_{μ}), sous la forme $f_{\text{д}} : X_{\text{д}} \rightarrow X_{\text{д}}$.

(162) Ainsi, selon la définition de l'architecture CAES, le *corps* (β) de l'agent représente sa constitution physique et organique (soit réelle, soit virtuelle). Par ailleurs, l'*esprit* (μ) de l'agent est une entité spéciale, insérée dans le corps, mais chargée des fonctions cognitives, de contrôle, de régulation, et du comportement. Les signaux *s* et *m*, qui définissent respectivement la *situation* et l'*actuation* de l'agent (д), correspondent aux mêmes signaux définis dans le système global (Æ).

2.3.1. Incarnation

(163) La notion d'incarnation (*embodiment*) est présente dans les sciences cognitives et en particulier dans l'IA, depuis les années 1990, et constitue depuis lors un sujet très débattu. Certains articles de référence sont (FRANKLIN, 1997), (QUICK et al., 1999), (CLARK, 1998), (ZIEMKE, 1998, 2002), (ANDERSON, 2003), (CHRISLEY; ZIEMKE, 2002) et (BARANDIARAN; MORENO, 2008), (OVERTON et al., 2008).

(164) Bien que dans certains travaux les termes « situé » et « incarné » soient utilisés en tant que synonymes, dans cette thèse ces concepts sont clairement distincts. L'*agent situé*, comme précédemment décrit, est celui qui est en couplage dynamique avec son environnement, dans une relation d'interférences mutuelles sur ses trajectoires, mais avec une portée d'observation et d'action seulement partielle. L'*agent incarné* est celui

qui a un corps comme univers interne, composé de propriétés et de processus métaboliques propres.

(165) En fait, le modèle de couplage dynamique présuppose un minimum d'incarnation. Dire que l'environnement interfère sur les états de l'agent, tel que défini dans le concept d'agent situé, implique de dire que l'agent possède des états internes. De même, l'existence de variables essentielles, que l'agent doit essayer de maintenir dans les limites de viabilité (selon préconise le modèle cybernétique), suppose l'existence d'un corps qui les contient, en prenant le rôle d'un univers interne pour l'agent.

(166) L'accent mis par les chercheurs en intelligence artificielle sur la nécessité d'utiliser les notions d'agent situé et incarné est dû à l'épuisement des modèles traditionnels du paradigme symbolique, trop représentationnistes et abstraits, souvent éloignés des évidences découvertes par les sciences naturelles, et également incohérentes en ce qui concerne l'idée d'un agent autonome (ZIEMKE, 1998), (SHANNON, 1993). Les logiciels présentés par l'IA traditionnelle s'adaptent très peu, et ils ont des objectifs établis de l'extérieur, sans une vraie signification interne.

(167) Un mouvement similaire a eu lieu dans le domaine des sciences cognitives, qui ont rejeté les modèles purement cognitivistes, en promouvant une plus forte convergence avec des modèles plus biologiques. Au cours des années 1990, la classique opposition entre « l'esprit » et « le corps » a été dissoute. Des études menées dans les divers domaines qui composent les sciences de l'esprit ont conduit à une réconciliation entre la psychologie et la biologie (TEIXEIRA, 2000). L'évidence que le cerveau fait partie du corps, et qu'il s'est développé avec son évolution implique que les aspects mentaux de l'être humain ne peuvent pas être complètement dissociés des aspects organiques. Le cerveau et le corps se trouvent intégrés à travers des circuits neuronaux et biochimiques réciproques (DAMÁSIO, 1994).

(168) Si l'esprit d'un agent interagit directement avec l'environnement extérieur, l'agent est réduit à un automate, un système comportemental vide, analogue à un mécanisme de stimulus et réponse. Parce qu'il est incarné, l'agent devient une entité immergée dans un environnement, et dans le même temps, il reste distinct de lui. Le corps devient le médiateur entre l'environnement et l'esprit. Alors que le monde se présente à l'agent comme le « milieu extérieur », le corps constitue le « milieu intérieur ».

2.3.2. Le Corps de l'Agent

(169) Dans l'architecture CAES, le *corps* (β) d'un agent fonctionne comme médiateur entre l'esprit (μ) et l'environnement externe (ξ), puisque c'est à travers le corps que l'esprit de l'agent perçoit le monde et décide des actions à effectuer, afin de le transformer, comme l'illustre la figure 2.7. Le corps est aussi la structure qui représente l'environnement interne de l'agent, composé d'un ensemble de *propriétés internes* (X_β), et des *processus métaboliques* qui constituent la fonction d'évolution (f_β) de ce sous-système, selon la définition 2.3.

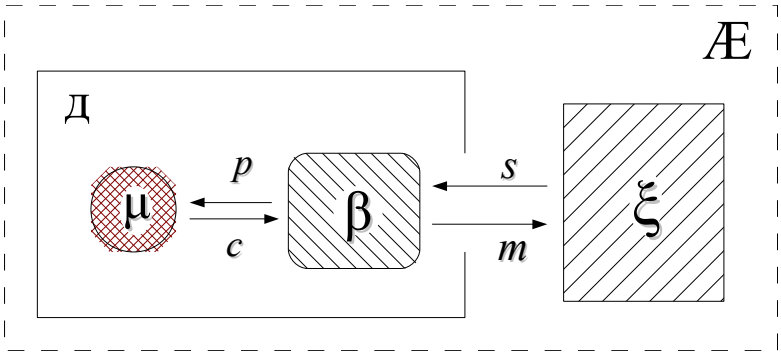


Figure 2.7: L'esprit, le corps, et l'environnement.

Le corps (β) de l'agent joue le rôle d'un médiateur entre son esprit (μ) et le monde extérieur (ξ).

Le corps (β) d'un agent (Д) est un sextuplet:

$$\beta = \{X_\beta, f_\beta, c, p, s, m\}$$

où,

$X_\beta = \{X_{\beta 1}, X_{\beta 2}, \ldots, X_{\beta n}\}$ est un ensemble fini de propriétés *(univers interne)*

$f_\beta : X_\beta \times C \times S \rightarrow X_\beta$ est la fonction d'évolution du système *(métabolisme)*

c est un signal que le corps β reçoit de l'esprit μ *(contrôle)*

p est un signal que le corps β envoie à l'esprit μ *(perception)*

s est l'interférence externe que le corps β reçoit de l'environnement ξ *(situation)*

m est l'interférence que le corps β fournit à l'environnement ξ *(actuation)*

Définition 2.3: Corps (β).

(170) Alors que le corps de l'agent communique avec l'environnement extérieur par le biais des signaux s et m, l'esprit communique avec l'organisme à travers des signaux p et c. La perception (p) est un signal informatif, produit par un ensemble de senseurs du corps de l'agent, et envoyé de ceux-là à l'esprit. De la même manière, le contrôle (c) est un signal généré par l'esprit, et envoyé aux actuateurs de l'agent, également localisés dans son corps.

(171) Les propriétés internes (X_β) définissent l'état général du corps, de forme analogue aux états chimiques, hormonaux, physiologiques, et biophysiques d'un organisme naturel. Ces propriétés ne peuvent avoir lieu que dans le corps de l'agent, parce qu'elles constituent un niveau différencié de l'esprit et aussi différencié du monde. La fonction d'évolution du corps (f_β) représente son métabolisme, constitué des processus autonomes de régulation, qui contrôlent les variations de l'état des propriétés internes du corps au long du temps.

(172) Dans la nature, le métabolisme chez un être vivant comprend tous les processus qui régissent l'activité des systèmes, des organes, des glandes, et des cellules elles-mêmes. En fonction de certaines conditions biochimiques, le métabolisme de chacune de ces composantes du corps peut promouvoir la modification de cet état biochimique, à partir, par exemple, de la production de certaines enzymes.

(173) Le métabolisme a l'état du corps comme principal paramètre. Néanmoins, comme l'organisme est un système partiellement ouvert, des événements externes peuvent provoquer des changements dans le milieu interne. Dans la nature, des circonstances extérieures qui influent sur le métabolisme sont, par exemple, l'alimentation, la variation de la température environnementale, le mouvement, le contact avec d'autres objets, etc. En plus, l'activité de l'agent interfère, elle aussi, sur l'état du corps, et finalement l'esprit prend un rôle important en tant que facteur de modulation sur les changements corporels, dès lors qu'il gère plusieurs processus métaboliques.

(174) Ainsi, la fonction d'évolution du corps (f_β) modifie l'état des propriétés internes en fonction de 3 paramètres. Tout d'abord, la condition même des propriétés internes (x_β), qui réalisent un cycle propre, corporel, de régulation métabolique. Deuxièmement, l'interférence de l'environnement, reçu comme situation (s), soit par la perception sensorielle ou d'autres types d'événements externes qui ont un impact sur le corps. Et troisièmement, l'interférence de l'esprit par le signal de contrôle (c), qui change l'état des actuateurs internes et externes.

2.3.3. Perception et Contrôle

(175) La neurophysiologie et la psychophysique (GESCHEIDER, 1997) font une distinction entre les mécanismes de la perception qui sont sensibles au changement des conditions environnementales (*extéroceptifs*), et les mécanismes qui perçoivent les conditions du corps (*intéroceptifs*), (SHERRINGTON, 1907). Chez l'être humain, les *sens* (selon la notion usuelle) sont extéroceptifs. Cette catégorie comprend la vue, l'ouïe, le toucher, l'odorat, le goût, l'équilibre, etc.

(176) Le système visuel humain, par exemple, selon (MEYER, 1997), commence par les yeux, des organes sensibles à l'incidence de la lumière provenant du milieu externe. Une image est formée sur la rétine de chaque œil par la conversion de la fréquence des ondes électromagnétiques (de la lumière) en tant que points de couleurs de différentes intensités. Cette image est transmise au lobe occipital du cerveau, qui fait une sorte de prétraitement de cette information, qui est alors convertie en signaux de mouvement, de traits et de formes.

(177) Un processus similaire se produit par rapport à l'ouïe (sensible aux vibrations sonores, qui arrivent à l'oreille par le biais de l'air et qui sont converties en signaux de fréquence et d'intensité), le goût et l'odorat (à travers les récepteurs chimiques situés sur la langue et la muqueuse nasale), le toucher (par le biais des récepteurs sensibles à différents types de pression mécanique sur la peau, et aussi à la température), et le sens de l'équilibre et de l'accélération (perçus via le mouvement des fluides dans le vestibule de l'oreille interne).

(178) Différemment, les perceptions intéroceptives sont liées à la situation du corps, en tant que sensations physiologiques (le niveau des diverses hormones et enzymes dans le sang), viscérales (qui proviennent des divers organes), kinesthésiques, la sensation de la douleur, et finalement les sensations proprioceptives (relatives à la position et au mouvement des membres du corps, à travers l'appareil musculaire et squelettique). Les sensations proprioceptives fournissent une rétroaction par rapport à l'état des actuateurs. Ce sont elles qui permettent à une personne, par exemple, de savoir quelle est la position de son propre bras à la suite d'un mouvement.

(179) De la même façon que les perceptions, les actions d'un organisme peuvent aussi être catégorisées suivant le même critère: d'une part, celles qui sont dirigées vers le monde extérieur, de l'autre, celles qui se destinent à son propre corps.

(180) Chez l'être humain, une grande part de l'activité du cerveau est destinée à la régulation du milieu corporel interne. Cependant, certains mécanismes du corps, les actuateurs, sont capables d'effectuer des interventions externes, dirigées vers l'environnement. L'appareil musculosquelettique, par exemple, est l'actuateur chargé de générer les mouvements et, par conséquent, il peut promouvoir des changements dans l'environnement au dehors du corps. Par contre, la plupart des glandes et des organes fonctionnent comme des mécanismes de régulation interne à travers la production, la transformation, et l'élimination des substances présentes dans l'organisme, et donc en modifiant son état physiologique.

(181) Parmi les mécanismes qui agissent sur l'univers corporel interne, on remarque un type spécial: les *actuateurs sensoriels*. Grâce à eux, l'esprit peut modifier les paramètres de ses propres senseurs. Chez l'être humain, un exemple typique sont les muscles qui régissent la direction du regard. Il y a plusieurs mouvements qu'on fait afin de modifier la position relative du corps, justement pour déplacer les yeux, les oreilles, le nez, la langue, les mains, en tant qu'instruments sensoriels, pour permettre une meilleure perception des objets. L'existence d'un tel contrôle de l'esprit sur des paramètres sensoriels permet la réalisation d'une *perception active* (ALOIMONOS et al., 1988), (BAJCSY, 1988), (WASSON et al., 1998).

(182) Dans l'architecture CAES, la *perception* (*p*) est un signal informatif reçu par l'esprit de l'agent, composé d'un ensemble de sensations qui représentent la situation de son corps. Dans le sens inverse, le *contrôle* (*c*) est un signal généré par l'esprit, et envoyé au corps afin de déclencher certaines actions.

(183) Comme dans les domaines naturels, dans l'architecture CAES les signaux de perception et de contrôle peuvent être définis par la combinaison de deux types de signaux: ceux relatifs à l'environnement extérieur, et ceux qui sont liés au corps, en tant qu'univers intérieur de l'agent.

(184) On divise, par conséquent, l'espace du signal de contrôle, $C = \{C_1, C_2, \ldots C_n\}$, envoyé par l'esprit au corps, en contrôle interne et contrôle externe. D'un côté, le

contrôle interne ($C_\beta \subset C$) est le moyen par lequel l'esprit interfère sur l'état du corps, entraînant des processus régulateurs. De l'autre côté, le *contrôle externe* ($C_\xi \subset C$) est le moyen par lequel l'esprit, à travers les actuateurs, interfère sur l'extérieur, en causant des changements dans l'environnement. Les actuateurs comprennent la partie des propriétés du corps (X_β) qui sont contenues dans le signal d'actuation de l'agent (M), en étant l'un des paramètres de la fonction d'évolution de l'état de l'environnement (f_ξ). Ainsi, les actuateurs constituent les instruments par lesquels l'esprit peut intervenir dans le monde extérieur. Le flux de contrôle est illustré à la figure 2.8.

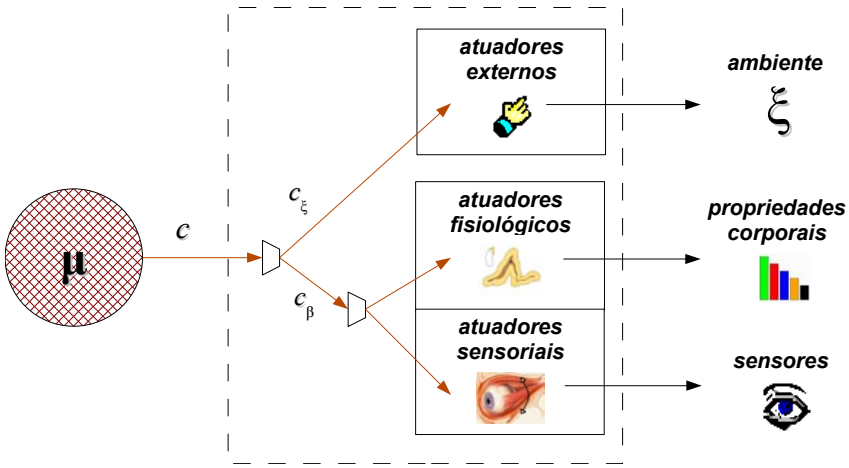


Figure 2.8: Flux de contrôle.

(185) L'espace du signal de perception, $P = \{P_1, P_2, \ldots P_n\}$, est lui aussi divisé en perception interne et perception externe. D'une part, la *perception externe* ($P_\xi \subset P$) est la partie du signal perceptif par laquelle l'esprit peut accéder au monde extérieur, parce qu'il s'agit de l'information provenant des variables du corps qui évoluent selon les conditions environnementales (senseurs). Les senseurs sont un sous-ensemble des propriétés du corps (X_β) dont la valeur est donnée par la situation (S), qui à son tour reflète les conditions environnementales. De l'autre, la *perception interne* ($P_\beta \subset P$) est constituée de variables qui prennent leurs valeurs avec les autres propriétés corporelles, en communiquant leur état à l'esprit. Les propriétés du corps et de l'environnement pour lesquelles l'esprit n'a pas d'accès perceptif par l'intermédiaire de P sont, du point de vue de l'agent, *propriétés non-observables*. Le flux de perception est illustré à la figure 2.9.

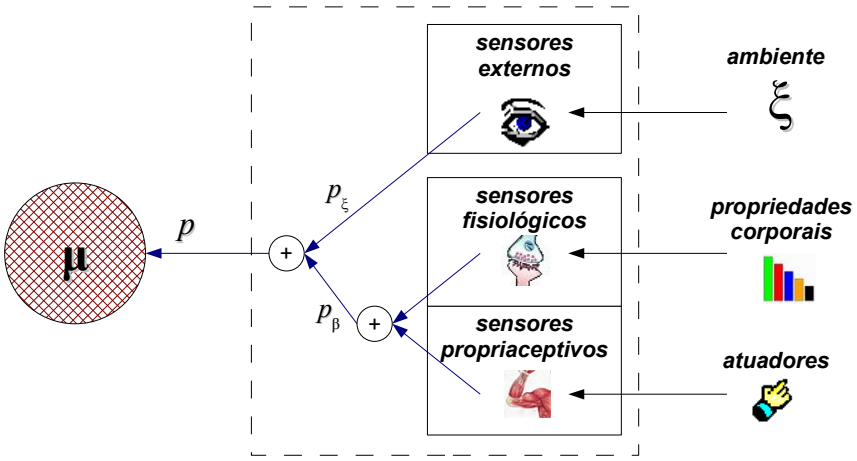


Figure 2.9: Flux de perception.

(186) En fait, ce qui s'établit entre l'esprit et tout ce qui est dehors de lui (le corps et l'environnement) est une relation *d'observation et de contrôle partiels*. L'esprit perçoit le corps à travers le signal P_β, qui ne fournit l'accès qu'à une partie des dimensions qui définissent l'état réel du corps (X_β), et les actions que l'esprit peut faire exécuter sur le corps, à travers le signal C_β ne représentent qu'une partie des paramètres de la fonction d'évolution du corps (f_β) capables d'influencer les modifications de son état.

(187) De même, l'esprit perçoit l'environnement extérieur à travers le signal P_ξ, provenant des senseurs, qui ne fournit l'accès qu'à une partie de la situation (S) dans laquelle l'agent se trouve, qui à son tour est aussi un signal partiel par rapport à l'état réel de l'environnement (X_ξ). De cette façon, comme pour le corps, l'esprit ne peut intervenir dans l'environnement que d'une manière partielle et indirecte, à travers le signal C_ξ, qui modifie l'état des actuateurs, placés dans le corps (X_β), et qui à leur tour modifient l'actuation de l'agent (M), signal qui est un des paramètres de la fonction d'évolution de l'environnement (f_ξ).

2.4. Adaptation de l'Agent à l'Environnement

(188) Dans le modèle de couplage dynamique (BEER, 1995), le critère d'adaptation est représenté d'une manière abstraite et générale comme une zone de l'espace où le flux du système doit rester. La limite d'adaptation est donc la frontière de cette région du système global (formé par le couplage de l'agent et de l'environnement), et l'agent est considéré adapté à son environnement si son activité guide la trajectoire du système

global de façon à ce qu'elle reste dans les limites de cette zone. Un exemple d'une telle trajectoire est illustré figure 2.10.

(189) Ashby (1952) avait déjà proposé des critères similaires pour définir le concept d'adaptation, en proposant que l'agent soit vu comme un système composé d'un ensemble de *variables essentielles* qui doivent rester dans des certaines *limites physiologiques normales*, ou *limites de viabilité*, pour que l'intégrité du système soit préservée et donc que la survie de l'agent soit garantie. Un certain comportement collabore à l'adaptation de l'agent s'il garantit la persistance de ces variables essentielles dans leurs limites de viabilité.

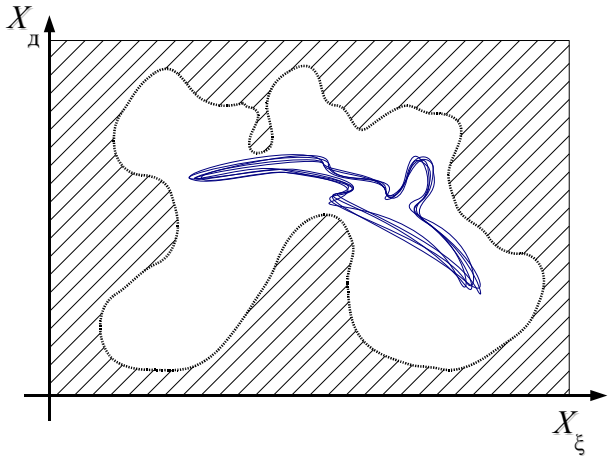


Figure 2.10: Exemple de la région d'adaptation et de la trajectoire adaptée.

L'axe horizontal représente la variation de l'état de l'environnement ($X_ξ$), et l'axe vertical représente la variation de l'état de l'agent ($X_д$).

(190) En général, soit dans la nature, soit dans les problèmes d'intelligence artificielle, l'agent et l'environnement ne se mettent pas automatiquement en parfaite harmonie. La plupart du temps, ces deux systèmes exercent des forces dans des directions opposées par rapport au flux du système global. Seulement un de ces deux pôles, l'agent, est susceptible de se désintégrer, autrement dit, de disparaître en tant qu'unité.

(191) Un couplage dynamique non destructif est réalisé dans la relation entre ces deux systèmes, quand l'agent interagit avec l'environnement de façon à assurer son auto-maintien (VARELA et al., 1991), (RUIZ-MIRAZO; MORENO, 2000).

2.4.1. Être Adapté

(192) En termes informatiques, un agent est considéré comme adapté à son environnement quand il arrive à bien atteindre ses buts. Pour des mécanismes programmés, cette mesure de réussite est généralement basée sur des facteurs exogènes. Par exemple, un bon critère d'adaptation pour un certain robot, une machine, ou un logiciel, peut être la satisfaction de l'utilisateur. Toutefois, dans la nature, la survie est ce qui balise les limites du succès. L'environnement lui-même devient juge, en condamnant à la disparition des organismes (et finalement des espèces) qui ne sont pas suffisamment adaptés.

(193) Ainsi, dans la perspective naturaliste de l'IA, un agent est considéré comme adapté à son environnement quand il réussit à survivre, pas forcément durant un temps infini, mais au moins pendant une longue période. Pour pouvoir définir une frontière entre la vie et la mort d'un agent, les modèles situés utilisent généralement le modèle cybernétique, dans lequel l'agent est défini tant que système constitué d'un ensemble de *variables essentielles*, qui doivent rester dans des certaines *limites de viabilité* de manière à ce que soit préservée l'intégrité du système et, par conséquent, la survie de l'agent.

(194) Les variables essentielles de l'agent demandent une régulation constante. En général, certaines parmi elles ont naturellement tendance à quitter les limites de viabilité, de sorte que l'agent doit être capable d'agir de façon appropriée afin d'assurer sa préservation à long terme. Parce que ces variables ne sont pas sous le contrôle absolu de l'agent, ce n'est que le flux déterminé par le couplage entre l'agent et l'environnement qui peut garantir qu'elles soient maintenues dans les limites de viabilité (BARANDIARAN; MORENO, 2008).

(195) Le comportement de l'agent devient le point clé pour assurer sa propre survie. Dans l'architecture CAES, l'esprit est le système chargé de coordonner le comportement de l'agent, afin de garantir son adaptation. L'agent est adapté à son environnement s'il sait interagir habilement avec lui, de manière à changer les conditions extérieures afin de guider le flux du système, en assurant la persistance de ses variables essentielles dans les limites de viabilité. Le maintien de l'équilibre des variables internes exige

l'adaptation du comportement aux variations de l'environnement externe (BARANDIARAN, 2004).

(196) Le paradigme de l'IA Située a inversé l'un des principes cybernétiques, en disant que les organismes, comme les agents complexes, couplés avec leurs environnements, constituent des *systèmes loin de l'équilibre thermodynamique* (*far from thermodynamical equilibrium systems*), (BICKHARD, 2000). Il s'agit d'un type spécial de système ouvert qui trouve ses limites de viabilité loin de l'équilibre du système global. Ce type de système est une structure dissipative, qui aurait naturellement tendance à la désintégration, mais qui en fait reste stable pendant une longue période de temps par le biais de la régulation de son interaction avec l'environnement.

(197) Dans cette condition, l'agent est une entité qui doit intervenir activement dans l'environnement, dans un effort pour contrer la force naturelle qui amène le système global à l'équilibre, situation qui signifierait la disparition de l'agent en tant qu'unité et organisation différenciée de l'environnement. Autrement dit, l'agent atteint son équilibre interne en maintenant le système global en constant déséquilibre.

2.4.2. Devenir Adapté

(198) La situation d'adaptation peut se réaliser dans un agent s'il est bien conçu (soit par un ingénieur, soit par le processus de sélection naturelle). Cependant, un agent peut aussi devenir adapté s'il dispose de mécanismes d'adaptation en ligne. Dans ce cas, l'agent est dit « adaptatif ». Dire qu'un agent est *adapté*, c'est dire qu'il a du succès dans ses interactions avec l'environnement. Par ailleurs, on dit qu'un agent est *adaptatif* s'il est capable de se modifier en cherchant l'adaptation. Cette caractéristique est nécessairement liée à des processus de développement et d'apprentissage.

(199) Un agent est un *système adaptatif* s'il est capable de se modifier afin de s'ajuster aux changements de l'environnement (BARANDIARAN; MORENO, 2008), c'est-à-dire, s'il est capable d'apprendre à travers l'expérience, en modifiant ses patrons de comportement afin d'augmenter la compétence dans l'exercice de ses activités (MAES, 1994). Pour le modèle de couplage dynamique, cela implique un processus de changement des structures génératrices du comportement, de sorte que, dans l'ensemble, les transformations rendent l'agent mieux adapté à l'environnement.

(200) Par exemple, une algue marine est un organisme adapté de façon innée à des conditions de vie en mer. La qualité d'adaptation de l'algue marine a été acquise à travers la sélection naturelle de l'espèce, et c'est donc un résultat d'un processus phylogénétique. Par contre, un ver de terre, s'il y était placé, finirait rapidement par cesser son existence en tant qu'organisme. Le ver de terre n'a pas la capacité d'adaptation pour faire face à un environnement aussi adverse pour lui que la mer.

(201) À l'inverse, l'être humain peut s'adapter à la mer, au moins d'une façon minimale pour y survivre. Il n'est pas possible qu'une personne transforme son métabolisme au point de pouvoir respirer l'eau salée. Toutefois, elle peut modifier son comportement en apprenant à nager, comme le résultat d'un processus ontogénique. Et nager n'est pas résister aux vagues, ni s'opposer à la dynamique de l'océan, mais c'est harmoniser le mouvement du corps à celui de la masse d'eau, c'est comprendre les rapports établis entre le corps et les vagues, en étant capable d'infliger des petits changements locaux et bien dirigés, de façon que le résultat soit la conduite du système dans un flux favorable à l'organisme (DELEUZE, 1970).

(202) Un agent artificiel peut présenter un comportement adapté sans l'avoir appris. Il peut résoudre ses problèmes intelligemment, simplement à partir d'un ensemble d'instructions préprogrammées. Toutefois, pour que ce soit possible, il est nécessaire que les stratégies de résolution soient connues du programmeur, et que la quantité de situations possibles dans l'environnement soit suffisamment petite pour que l'agent puisse être préprogrammé.

(203) Dans ce cas, créer un agent adapté à l'environnement exige que les développeurs maîtrisent les connaissances nécessaires pour résoudre les problèmes que l'agent va trouver au fil de son existence. Bien sûr, il y a de nombreux cas où il est difficile ou même impossible de prévoir et de proposer des solutions pour toutes les situations. Ces problèmes exigent que l'agent apprenne à travers l'expérience, et qu'il s'adapte à des situations imprévues.

(204) Dans la nature, l'adaptabilité peut se faire par des processus purement physiologiques. Tel est le cas des êtres simples, unicellulaires, dont l'adaptation dépend directement de leur métabolisme, lorsqu'ils sont sensibles aux changements des conditions extérieures, ou de comportements réactifs. Il y a, cependant, des processus

plus complexes d'adaptation, où sont présents des aspects cognitifs, pour lesquels on réserve le terme *apprentissage* (DENNET, 1996). L'apprentissage est l'adaptabilité menée par les êtres qui possèdent une organisation intellectuelle et des informations qui sont pertinentes à leur comportement, en général représentées en tant qu'un modèle intériorisé de la dynamique du monde.

2.5. Caractéristiques de l'Esprit

(205) Bien que certains mécanismes automatiques et certains simulateurs d'organismes simples puissent se dispenser de la notion d'« esprit », en général les modèles d'agent en intelligence artificielle présupposent un esprit. Par analogie à ce qui se passe dans la plupart des animaux, le comportement d'un agent décrit à travers l'architecture CAES est le résultat des interactions de son esprit avec l'extérieur, c'est-à-dire, avec tout ce qui est en dehors de lui, y compris l'environnement physique qui existe en dehors du corps, et le corps de l'agent lui-même en dehors de l'esprit (PARISI, 2004).

(206) L'esprit (μ) dans l'architecture CAES est une structure de contrôle composée de deux sous-systèmes: le *système régulatif* (Я) et le *système cognitif* (Ж), responsables de la gestion à la fois du métabolisme et du comportement de l'agent. L'*esprit* de l'agent communique avec le corps à travers les signaux de contrôle (*c*) et de perception (*p*), comme le montre la figure 2.11, et tel qu'il est formalisé par la définition 2.4.

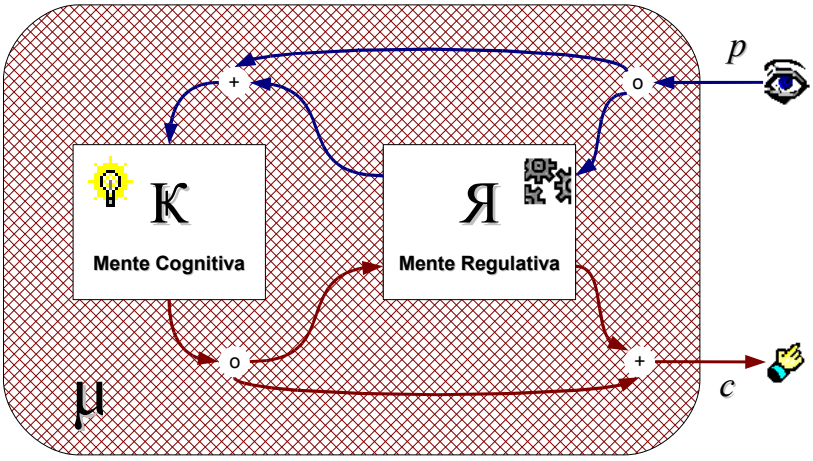


Figure 2.11: Structure interne de l'esprit (μ) dans l'architecture CAES.

Le système régulatif (Я) de l'esprit s'interpose entre le corps et le système cognitif, d'un part en ajoutant des valeurs affectives sur le signal de perception, et de l'autre part en promouvant certaines régulations du corps et de la conduite de l'agent par le biais des réactions émotionnelles et comportementales. Le système cognitif (К), à son tour, est le responsable pour la compréhension, l'apprentissage, et la prise de décision.

L'esprit (μ) d'un agent est un quadruplet:

$$\mu = \{К, Я, c, p\}$$

où,

К est le sous-système cognitif

Я est le sous-système régulatif

c est un signal que l'esprit μ envoie au corps β (contrôle)

p est un signal externe que l'esprit μ reçoit du corps β (perception)

Définition 2.4: Esprit (μ).

(207) De façon similaire à (SLOMAN et al., 2005), l'esprit contrôle le comportement de l'agent par la combinaison d'un système cognitif, chargé d'apprendre, d'interpréter, de planifier et de délibérer des actions, avec un système réactif et émotionnel.

(208) Le *système régulatif* de l'esprit est défini comme un triplet: $Я = \{Я_R, Я_E, Я_A\}$, où $Я_E$ est le *système émotionnel*, responsable des réactions métaboliques internes, $Я_R$ est le *système réactif*, chargé de mener les réactions comportementales externes, et $Я_A$ est le *système évaluatif*, qui assigne des valeurs affectives à des sensations. Le système régulatif de l'esprit (Я) est illustré à la figure 2.12.

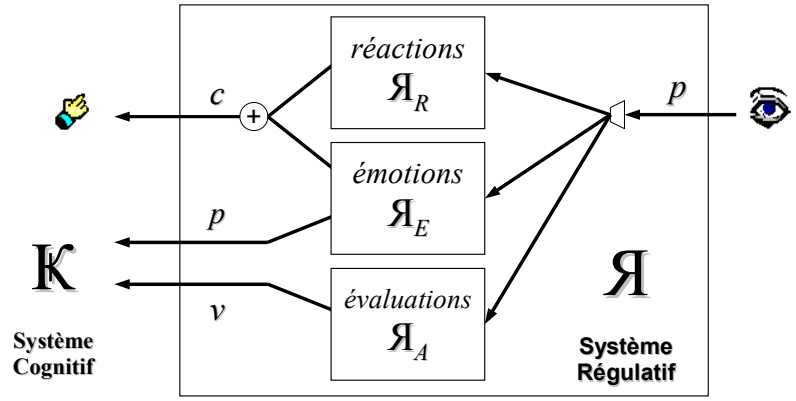


Figure 2.12: Structure interne du système régulatif (Я) dans l'architecture CAES.

2.5.1. L'Affectivité et les Émotions

(209) Un débat théorique sur la notion d'autonomie dans le contexte des agents conduit tôt ou tard à une question cruciale: la nécessité que les objectifs de l'agent soient fixés par lui-même. Un agent vraiment autonome doit agir dans le monde sous la force de *motivations intrinsèques.* Dans le paradigme de l'agent situé et incarné, en couplage dynamique avec son environnement, la motivation doit être le facteur qui entraine l'agent à agir pour le maintien de ses variables essentielles.

(210) Parallèlement, les années 1990 ont apporté de grands progrès dans les neurosciences par rapport à la compréhension du cerveau et de son fonctionnement, et cela a dissolu un autre antagonisme classique entre *la raison* et *l'émotion*. Dans les organismes complexes et intelligents, en particulier chez l'être humain, ces facteurs sont étroitement liés, et le comportement est le résultat d'un fonctionnement intégré. La raison sans l'émotion devient vide, et l'émotion sans la raison devient aveugle (ELLIS, 1962).

(211) C'est-à-dire qu'en plus de la logique il y a d'autres propriétés fondamentales qui caractérisent le comportement intelligent. L'esprit ne peut pas se confondre avec la raison, ou avec la mémoire, ou avec la conscience. En fait, l'esprit comprend tous ces aspects, et encore d'autres, qui travaillent en parallèle et de façon conjointe, y compris les aspects affectifs et émotionnels.

(212) Le comportement intelligent est une conséquence de l'intégration entre un système cognitif, qui analyse les situations, et un système émotionnel et affectif, qui évalue les scénarios et conduit à l'action (DAMÁSIO, 1994). Pour les êtres humains, il n'est pas possible d'indiquer un comportement qui soit purement cognitif, sans compter

des éléments affectifs en jeu, ni un comportement purement affectif, sans considérer des éléments cognitifs. L'affectivité est une condition nécessaire à l'existence des comportements intelligents, car elle est (au moins à l'origine) le moteur pour l'action. L'être humain agit seulement quand il est impulsé par une raison, et cela se traduit comme une forme de besoin (PIAGET, 1954, 1964, 1967).

(213) D'une part, un état d'émotion implique une distinction, et donc la capacité intellectuelle de reconnaître et de différencier les situations. D'autre part, la présence des émotions au cours des expériences vécues modifie la notion qu'on a de ces situations, en leur conférant une qualité positive ou négative (DAMÁSIO, 1994).

(214) D'une manière similaire à (CAÑAMERO, 1997a), l'architecture CAES établit une claire distinction entre « les émotions » et « l'affectivité ». L'*affectivité* est liée aux besoins de l'agent. Elle se fait à travers d'un système chargé de fournir une évaluation, positive (agréable, bien...) ou négative (désagréable, mauvais...), des signaux du corps, et éventuellement aussi de certaines conditions importantes de l'environnement. Cette information est généralement dirigée vers le système cognitif, qui, en fonction de celle-ci, effectue le processus de prise de décision.

(215) Les *émotions* sont des exécuteurs de réponses réflexes et viscérales. Ils apparaissent également sous la forme de systèmes à l'intérieur de l'agent. Cependant, différemment de l'affectivité, chaque émotion est un processus particulier qui, à partir de la reconnaissance d'événements spéciaux, promeut des réactions corporelles spécifiques, qui changent l'état interne de l'agent. Ces changements, tournés vers le corps, ont pour but de préparer l'agent à des réponses externes plus appropriées à son contexte immédiat, en potentialisant ou en inhibant certains types de comportement.

2.5.2. Système Évaluatif

(216) L'affectivité joue un rôle important dans le fonctionnement de l'intelligence. Sans elle il n'y aurait pas l'intérêt, ni le besoin, ni la motivation. Dans l'architecture CAES, un *besoin* est un déséquilibre d'une variable essentielle interne de l'agent qui doit être compensé. La motivation pour rétablir l'équilibre de ces variables provient de l'agent lui-même, par l'intermédiaire du *système évaluatif*.

(217) Le système d'évaluation est composé d'un ensemble de fonctions qui évaluent certaines propriétés du corps, et éventuellement certaines conditions perçues dans l'environnement, en fournissant des valeurs positives et négatives qui servent à indiquer des situations favorables ou défavorables dans lesquelles l'agent peut se trouver. Cette évaluation constitue une valeur affective, qui est la façon la plus simple de discerner entre ce qui est intrinsèquement bon ou mauvais pour l'agent.

(218) Par exemple, chez l'être humain, les signaux corporels de douleur, de soif, de faim, de froid ou de chaleur excessive, d'épuisement, de dégoût, entre autres, en plus de signaux externes tels que le goût et l'odeur des substances potentiellement nocives sont intrinsèquement négatifs. De même, les sensations telles que le plaisir sexuel ou le goût d'un aliment calorique, sous certaines conditions corporelles, ont une certaine valeur positive innée. Cela prend tout son sens dans le contexte de l'évolution de l'espèce humaine, et ne contredit pas le fait qu'au niveau de la pensée abstraite d'autres valeurs peuvent être agrégées à ces signaux.

(219) Dans le modèle CAES, le besoin est exprimé par l'affectivité. Lorsque une certaine variable essentielle doit être constamment préservée autour d'une certaine « valeur normale », on lui associe une relation d'*affectivité normale*. Dans ce cas, si la variable s'éloigne de la valeur normale, une sensation négative est déclenchée. D'autres variables peuvent être importantes pour l'agent, sans forcément être essentielles. Pour ces variables ou peut établir une relation d'*affectivité positive* qui correspond à un « pic » de sensation agréable quand la variable touche un certain écart de valeurs cibles. Inversement, on peut aussi établir des relations d'*affectivité négative*, comme des « vallées » de sensation désagréable.

(220) Pour être autonome, un agent doit agir et prendre des décisions en vertu de ses propres motivations ou objectifs. Dans le modèle classique de l'apprentissage par renforcement, l'agent est *extrinsèquement motivé*, c'est-à-dire qu'il est entrainé à agir en fonction des buts ou récompenses externes, provenant de l'environnement. Dans l'architecture CAES, grâce à l'existence d'un système évaluatif, l'agent devient *intrinsèquement motivé*, car il est poussé à avoir des comportements lorsqu'ils sont intrinsèquement agréables. L'idée des agents intrinsèquement motivés peut être lue en (SINGH et al., 2004) et (OUDEYER; KAPLAN, 2007).

(221) L'existence d'un système affectif artificiel constitue un moyen d'internaliser les signaux de renforcement, en remplaçant l'idée de récompense extérieure. Ainsi, la figure d'un objectif explicite disparaît, et il n'y a plus de situation prédéfinie à atteindre dans le monde; à la place, il existe une motivation pour l'action qui émerge de la relation entre la connaissance et l'affectivité.

(222) L'affectivité remplace aussi la survie en tant que paramètre d'adaptation. Si la mort du système est le seul indicateur d'inadaptation, il est impossible que l'agent puisse apprendre à s'adapter à l'environnement à travers l'expérience acquise au cours de sa vie.

(223) L'affectivité ajoute une sorte de relief évaluatif à l'espace (de phase ou d'états) du système, comme illustré sur la figure 2.13. Sans l'affectivité, cet espace est une grande plaine grise, et l'agent n'a pas de raison de préférer une situation à une autre. Ce relief évaluatif contient quelques régions plus élevées, zones positives de l'espace, qui abritent des situations éprouvées par l'agent comme plaisantes, et qui sont possiblement liées à des moments de grand avantage en termes de son auto-maintien. De même, il y a d'autres régions de l'espace où le relief est bas, en indiquant des contextes désagréables à l'agent, et qui probablement franchissent, à la fin de la descente, le seuil de la survie.

(224) Le relief évaluatif définit la motivation de l'organisme pour qu'il conserve ses variables essentielles autour des indices normaux. En conséquence, si l'organisme agit en recherchant des sensations positives et en évitant les négatives, il préserve indirectement le flux du système dans une région protégée, loin des frontières de la mort.

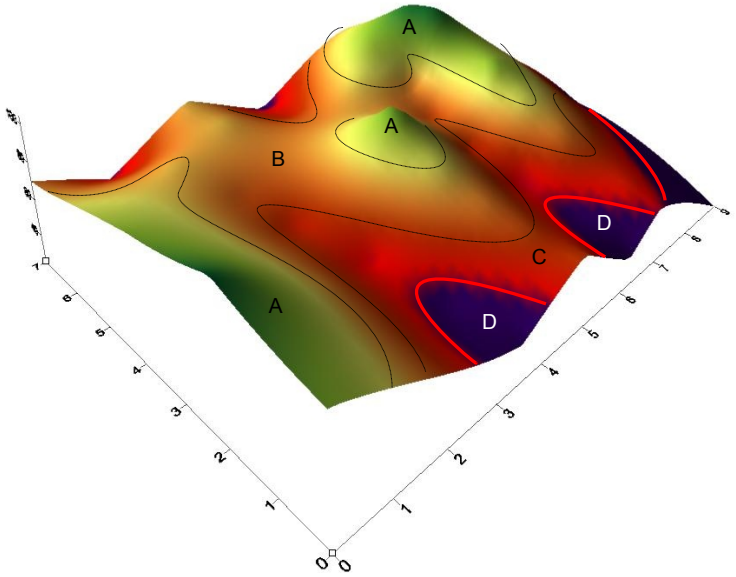


Figure 2.13: Exemple d'une surface affective.

L'espace représente le flux du système, qui fait varier l'état de l'agent et de l'environnement, et le relief représente l'évaluation affective de ces états par l'agent. Les zones hautes (A) représentent des valeurs positives, et les zones basses (C) représentent des valeurs négatives. Les dépressions sombres sur le graphique (D) sont déjà au-delà de la limite de la survie de l'agent.

(225) Par exemple, ne pas être déshydraté est une situation cruciale pour notre survie. Il aurait un coût trop grand, et il serait aussi risqué et tardif pour l'organisme, si la détection de faibles niveaux d'hydratation ne se produisait que lors de la mort des cellules. L'évolution naturelle a établi des mécanismes qui permettent d'évaluer le niveau d'hydratation, en essayant de le tenir autour de *valeurs normales*, qui assurent la santé du corps. Au moment où ce niveau s'éloigne de la zone de normalité, une sensation désagréable (la soif) commence à se faire sentir. Cet avertissement est donné bien avant le seuil de la survie, en permettant à l'agent d'agir plus tôt (dans ce cas, en buvant de l'eau) afin de compenser le déséquilibre et d'éliminer la sensation désagréable.

(226) De même, les sensations positives sont des raccourcis créés par la sélection naturelle pour que l'organisme préfère et qu'il cherche certaines situations très favorables à son auto-maintien, qui se sont révélées intéressantes au fil des générations, et qui se sont finalement incorporées dans l'organisme par le biais de ce mécanisme affectif inné. Par exemple, le plaisir qu'on éprouve quand on mange une nourriture bien calorique est un facilitateur de la survie. Ce serait beaucoup plus risqué pour l'organisme

d'être obligé de découvrir par lui-même, après être devenu sous-alimenté, que la nourriture qu'il ingère ne sert à rien.

(227) Ce modèle est biologiquement plausible. L'évolution naturelle a défini pour des organismes complexes plusieurs mécanismes pour assurer leur survie. Parmi ces mécanismes, en plus des fonctions organiques, on trouve les fonctions liées au comportement, tels que les instincts, les réactions, les émotions, et l'affectivité. Les sensations positives et négatives définies dans le système évaluatif sont précisément une manière de cartographier les limites de la survie de l'organisme, comme le montre la figure 2.14.

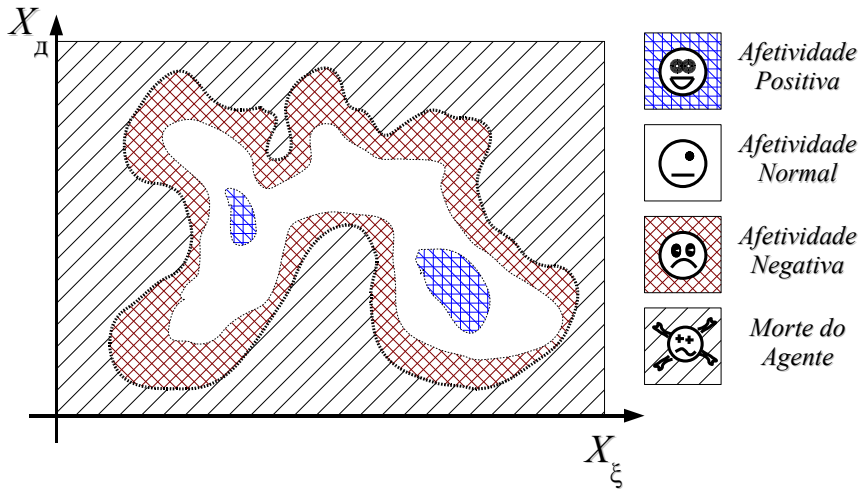


Figure 2.14: La surface affective cartographe les limites de survivance.

Les zones d'affectivité négative forment une ceinture de sécurité qui éloigne l'agent des frontières de sa propre mort, alors que les zones d'affectivité positive forment des attracteurs qui entrainent l'agent vers des situations favorables à son auto-maintien.

(228) Le *système d'évaluation* ($Я_A$) associe une valeur affective à certaines propriétés du signal de perception. Les valeurs sont mesurées sur une échelle continue et décrivent des sensations agréables ou désagréables associées à certaines conditions. Ces valeurs affectives sont envoyées au système cognitif, qui doit les utiliser comme paramètres de décision pour choisir entre des actions alternatives, afin d'amener l'agent à des situations plus confortables.

(229) Le système d'évaluation est plus précisément constitué par un ensemble de *déclencheurs évaluatifs*. Ces déclencheurs sont décrits, de façon générale, comme des fonctions qui mappent le domaine de valeurs possibles pour chaque élément P_i du signal de perception P, vers une mesure affective comprise entre -1 et +1. Ainsi, chaque

perception spécifique peut représenter une condition, soit rapportée à l'environnement, soit au corps, associée à des sensations agréables quand elles sont positives, et désagréables quand elles sont négatives. La condition d'un déclencheur évaluatif peut aussi être composée d'une combinaison de conditions simples. La composition des affections de tous les déclencheurs évaluatifs tirés à chaque instant constitue la *sensation affective générale* de l'agent. Ce type de signal d'évaluation correspond au modèle de récompenses factorisées (DEGRIS et al., 2008).

2.5.3. Système Émotionnel

(230) Dans la nature, à la fois sur les animaux et les humains, les émotions influencent le comportement, en favorisant certaines conduites au détriment des autres. Les émotions sont aussi, telle l'affectivité, un moyen d'évaluer le résultat des actions de l'agent.

(231) Les systèmes émotionnels biologiques font l'objet d'un grand intérêt pour la neuropsychologie contemporaine (LEDOUX, 2000), (DAMÁSIO, 1994), (EKMAN, 1999). De même, l'étude des émotions et la recherche sur la production des émotions synthétiques ont gagné en importance dans le domaine de l'intelligence artificielle. Quelques modèles émotionnels artificiels ont été présentés dans (CAÑAMERO, 1997a, 1997b), (VELÁSQUEZ, 1997), (SLOMAN, 1999, 2005) et (ALMEIDA et al., 2004).

(232) Selon une définition neuropsychologique, l'émotion est liée à la fois au caractère affectif et au caractère corporel. C'est la combinaison d'un processus évaluatif mental simple ou complexe, avec le déclenchement de certaines réponses, surtout dirigées vers le corps lui-même, qui entraînent un état émotionnel du corps, mais qui sont aussi dirigées vers l'esprit. À partir de l'identification ou de l'anticipation de certaines situations, un processus de modification du corps commence, en déclenchant des sensations spécifiques et en prédisposant le corps à des comportements de réponse (DAMÁSIO, 1994).

(233) Chez l'être humain il y a un ensemble d'états émotionnels innés. Chaque émotion a une configuration bien définie sur le type d'événements qui la déclenche, et sur les changements corporels et comportementaux qu'elle induit (DAMÁSIO, 1994).

L'émotion promeut des changements physiologiques dans l'organisme, afin de le préparer à réagir différemment dans ces différents états émotionnels (EKMAN, 1999).

(234) Les différentes émotions ne sont pas seulement une question d'intensité positive ou négative. Dans l'anatomie humaine, chaque émotion primaire est un sous-système particulier, qui correspond a une unité cérébrale fonctionnelle distincte. Les émotions négatives (la peur, la colère, le dégoût, la tristesse et le mépris) et les émotions positives (la joie, la fierté, la satisfaction, le soulagement et la satisfaction) correspondent à des sensations différentes. Chacune de ces émotions s'est établie dans l'espèce humaine pour avoir une valeur adaptative, c'est-à-dire, pour faciliter le traitement des tâches cruciales de la vie, en optimisant l'organisme pour répondre à la situation détectée (EKMAN, 1999).

(235) Les émotions sont dirigées vers le cerveau, en changeant l'état d'attention, en activant des réseaux associatifs, et en induisant l'évocation de souvenirs, et sont dirigées vers le corps, en provoquant des changements physiologiques et endocrinienne (l'activité hormonale, les réactions métaboliques), et en interférant avec l'activité du système nerveux autonome (les réactions corporelles, les expressions), (EKMAN; DAVIDSON, 1994).

(236) Par exemple, chez de nombreux animaux une émotion de peur peut être déclenché lorsque un bruit fort est entendu. L'activation de la peur déclenche des modifications corporelles dans l'organisme, en augmentant le niveau de certaines substances telles que l'adrénaline dans le sang. L'hormone sert, ensuite, de déclencheur pour d'autres processus physiologiques et métaboliques, comme l'augmentation de la fréquence cardiaque, et en conséquence l'augmentation de l'oxygénation des muscles, produisant, dans le corps, les conditions et la propension à une conduite plus appropriée, comme la fuite.

(237) Dans l'architecture CAES, l'agent a des émotions par le biais d'un modèle synthétique qui définit des propriétés émotionnelles. Ces variables, en plus d'influencer les processus métaboliques internes du corps, influencent le comportement ultérieur de l'agent par le changement de certains paramètres impliqués dans le processus d'évaluation affective et de prise de décision.

(238) Le *système émotionnel* ($Я_E$) est formé par un ensemble d'*émotions artificielles*, chacune sensible à certaines perceptions de l'agent (qu'elles soient externes, provenant du milieu, ou internes, provenant de son propre corps ou même de son propre esprit), semblable aux émotions primaires chez l'être humain (DAMÁSIO, 1994). Une émotion est activée à partir de la détection d'un contexte spécifique, et son activation entraîne la réalisation de modifications corporelles, représentées par un ensemble d'effets transmis de l'esprit à l'organisme par le signal de contrôle interne (C_β). Ces changements, dirigés vers le corps, ont pour but de préparer l'agent à des réponses externes plus appropriées à son contexte immédiat, en potentialisant ou en inhibant certains types de comportement.

(239) Dans le CAES, tel que dans (CAÑAMERO, 2001) et (SLOMAN et al., 2005), le système émotionnel ne reproduit pas directement les effets superficiels connus des émotions sur le comportement, mais implémente les processus sous-jacents de l'émotion à travers lesquels les effets comportementaux peuvent émerger. Le système émotionnel est formé par un ensemble d'émotions primaires. Chaque émotion est activée par la détection de certains contextes spécifiques, vérifiés à partir de la perception interne et externe, en déclenchant des réactions métaboliques qui modifient les valeurs des propriétés internes du corps de l'agent. Les émotions peuvent aussi envoyer des signaux directement au système cognitif, à travers la perception interne, et peuvent être associées à des valeurs affectives dans le système évaluatif.

2.5.4. Système Réactif

(240) Le système réactif de l'esprit est un exécuteur de réponses réflexes. Chaque réaction est un processus particulier, qui, à partir de la reconnaissance d'événements spéciaux, effectuent des actions dirigées. Dans la nature, la sélection naturelle a créé ce type de mécanisme chez les organismes car ils lui apportent des avantages adaptatifs.

(241) Dans l'architecture CAES, le principal régulateur du comportement est le système cognitif de l'agent. Cependant, de façon similaire à (SLOMAN, 1999), il est possible de combiner un système de délibération, coordonné par la cognition, avec un système réactif.

(242) Dans la nature, même les animaux qui ont développé des mécanismes cognitifs sophistiqués (comme l'être humain) gardent dans leur structure certains comportements

réactifs. Ces mécanismes réactifs ne réalisent aucune réflexion sur les situations, mais ils sont utiles pour déclencher des réponses rapides de l'organisme. Ils exécutent des comportements simples mais spécifiques, vers l'environnement, en réponse à la détection de certains stimuli (SLOMAN, 1999).

(243) Dans l'architecture CAES, le système réactif ($Я_R$) implémente les pulsions innées de l'agent, également activées à partir de la détection de certains contextes. Ces impulsions ont comme résultat le déclenchement d'actions externes. La différence par rapport au système émotionnel est que les réactions sont transmises de l'esprit au corps par le signal de contrôle externe (C_ξ), et donc généralement induisent des comportements qui modifient l'environnement.

(244) De la même façon que dans l'esprit humain (LEDOUX, 1996, 2000), (ÖHMAN, 2005), dans l'esprit de l'agent, les pulsions réactives peuvent entrer en conflit avec les signaux de contrôle générés par le système cognitif. Un mécanisme d'inhibition doit être utilisé par un système ou par l'autre, afin d'assurer la priorité de son signal. Le système cognitif, se renseigne sur l'action qui est effectivement réalisé par le biais de la rétroaction proprioceptive, puisque une partie du signal de perception interne (P_β) informe l'esprit sur l'état des actuateurs du corps.

2.5.5. Cognition et Apprentissage

(245) D'un côté, les interprétations trop radicales sur l'idée de couplage dynamique finissent, comme un effet collatéral, par négliger les modèles cognitifs en IA, en réduisant les agents à des automates réactifs. De l'autre côté, les modèles d'intelligence trop cognitivistes ont tendance à ignorer certains mécanismes d'adaptation qui sont présents dans la nature, tel que les instincts, les réactions et les comportements associatifs. Ce genre de réflexion peut être trouvé dans (BROOKS, 1991), (STEWART et al., 2008), (VAN GELDER, 1998), (CHEMERO, 2000), (WILSON; CLARK, 2008), (FRENCH; THOMAS, 1998), (CLARK, 1998), (NOLFI, 2002), (GRUSH, 1997a, 1997b, 2004), (SYMONS, 2001), (BICKHARD, 2000, 2009), et (BARANDIARAN; MORENO, 2006, 2008).

(246) Un *agent réactif* est un mécanisme du genre stimulus-réponse. Il n'organise pas un souvenir explicite des expériences, ni coordonne ses actions dans le temps. Il ne crée

pas des éléments conceptuels pour interpréter ou comprendre son monde. Sa seule capacité est d'associer des entrées sensorielles à des sorties motrices. Pour un agent réactif il n'y a que le présent, et les perceptions sensorielles instantanées. Différemment, un *agent cognitif* fait une représentation de la dynamique de son interaction avec l'environnement, sous la forme d'un modèle interne, en étant capable d'organiser l'expérience vécue, de réaliser des anticipations, et de planifier des actions futures.

(247) L'existence des mécanismes cognitifs ouvre un horizon infini à l'agent, à la limite, en lui permettant de réaliser la construction de concepts abstraits et de théories, la pensée par le biais de raisonnements hypothéticodéductifs, la constitution de relativisations et d'analogies, et la compréhension de l'univers au travers des relations causales et spatiotemporelles.

(248) Plusieurs théories psychologiques, parmi lesquels le constructivisme (PIAGET, 1936, 1937, 1964), suggèrent que le développement de l'intelligence chez l'être humain commence par l'apprentissage de coordinations sensorimotrices, et ces connaissances pratiques donnent origine, au fur et à mesure, à des modèles mentaux chaque fois plus conceptuels et abstraits pour interpréter, anticiper, et interagir avec la réalité.

(249) Dans le cas d'un agent cognitif, l'apprentissage se réalise par la réorganisation du modèle du monde qu'il représente dans ses structures intellectuelles internes, pendant qu'il vit des nouvelles expériences. L'agent s'adapte à l'environnement à travers l'amélioration de son modèle interne, ce qui le conduit, par conséquent, à modifier son comportement.

2.5.6. Système Cognitif

(250) CAES se propose en tant qu'architecture générale pour la modélisation d'agents autonomes, et ainsi la constitution exacte du système cognitif est laissée ouverte. De cette façon, il est possible de l'utiliser avec différents mécanismes d'intelligence artificielle.

(251) Le *système cognitif* (𝕂) d'un esprit (μ) peut être un simple mécanisme stimulus-réponse, qui est limité à la gestion d'associations, ou il peut être constitué d'un mécanisme intellectuel hautement complexe, implémenté sous forme de différents niveaux de contrôle et d'interprétation des événements.

(252) Le système cognitif de l'agent effectue les tâches d'apprentissage et de délibération. C'est à lui que revient la charge de construire la connaissance sur le monde et sur lui-même, et aussi de planifier les actions et de prendre des décisions. Il est clair que le système cognitif est la partie la plus importante et la plus complexe de l'agent en tant que système intelligent, ce qui est l'objet du prochain chapitre, consacré à la définition du CALM, un mécanisme que nous avons développé pour jouer le rôle de système cognitif pour CAES.

(253) La structure définie par l'architecture CAES impose l'utilisation de mécanismes qui réalisent une cognition située. Plus précisément, il faut considérer l'existence d'une intelligence sensorimotrice qui précède l'intelligence symbolique, dans un processus qui évolue de l'une vers l'autre. L'abstrait est construit à partir du sensorimoteur (PIAGET, 1945). La cognition située concilie l'utilisation des contextes perceptifs réels avec l'utilisation de représentations et de couches conceptuelles qui servent à interpréter les événements du monde. Le développement de l'intelligence commence par l'apprentissage des coordinations sensorimotrices, et ces connaissances pratiques donnent naissance, peu à peu, à des modèles de plus en plus conceptuels et abstraits, capables de correctement interpréter, anticiper et faire face à une réalité complexe.

3. CALM: Mécanisme d'Apprentissage Constructiviste

(254) Dans ce chapitre, nous présentons le mécanisme CALM (*Constructivist Anticipatory Learning Mechanism*), conçu pour jouer le rôle de système cognitif dans des agents définis à travers l'architecture CAES (présentée dans le chapitre précédent). CALM effectue l'apprentissage des modèles du monde, et aussi la construction des politiques d'action pour la prise de décision.

(255) Le problème de l'apprentissage des modèles du monde est équivalent au problème de la découverte de la structure dans un processus de décision markovien factorisé et partiellement observable (FPOMDP). Par ailleurs, le problème de la construction d'une politique d'actions, dans la littérature scientifique, est identique au problème de la résolution d'un processus de décision. C'est ce qui est montré dans la première section de ce chapitre, consacrée à la caractérisation, la formalisation et

l'analyse des problèmes d'apprentissage. Ensuite, la deuxième section du chapitre présente en détails le mécanisme CALM, envisagé pour résoudre ces deux problèmes.

3.1. Définition des Problèmes d'Apprentissage

(256) Tout au long de cette section, nous montrerons comment les problèmes d'apprentissage se posent pour un agent défini à travers l'architecture CAES (présentée dans le chapitre précédent). En général, l'agent est confronté à deux problèmes majeurs d'apprentissage: (a) procéder à la *construction d'un modèle du monde* qui décrit la dynamique des relations dans le système global, et, à partir de celui-là, (b) entreprendre la *construction d'une politique d'actions* visant à atteindre ses objectifs, c'est-à-dire, la maximisation des situations affectivement positives au cours de sa vie.

3.1.1. Apprentissage de Modèles du Monde

(257) Par *apprentissage de modèle du monde*, nous nous référons au processus dans lequel un agent, inséré dans un environnement particulier qui est nouveau pour lui, construit une représentation descriptive des événements qu'il observe au cours de son expérience.

(258) L'entrée d'un algorithme d'apprentissage de modèles du monde dans l'architecture CAES est un flux continu du type $\{p^{(0)}c^{(0)}, p^{(1)}c^{(1)}, \ldots, p^{(t)}c^{(t)}\}$. Les éléments p sont des signaux perceptifs, qui correspondent à des observations successives faites par l'esprit de l'agent à travers ses senseurs, et les éléments c sont des signaux de contrôle de l'agent, qui correspondent à des actions mises en œuvre par le biais de ses actuateurs. Si on considère une représentation discrète en ce qui concerne le temps, la paire $p^{(i)}c^{(i)}$ indique l'observation réalisée au temps $t=i$ et l'action effectuée immédiatement après. La tâche de l'algorithme d'apprentissage de modèles du monde est d'induire, à partir de l'expérience, une structure Ψ telle que, à partir d'une nouvelle situation quelconque $p^{(t)}c^{(t)}$, il soit possible de prévoir $p^{(t+1)}$.

(259) Le signal perceptif p représente la situation observée à un moment donné. Le domaine de p est factorisé en $|P|$ propriétés, ce qui correspond à l'espace $P = P_1 \times P_2 \times \ldots \times P_{|P|}$. L'agent induit des transformations dans l'état du système à travers le signal

de contrôle *c*, qui représente ses actions. À son tour, le domaine de *c* est factorisé en $|C|$ propriétés, ce qui correspond à l'espace $C = C_1 \times C_2 \times ... \times C_{|C|}$.

(260) La représentation bâtie par l'agent ne modélise pas le monde lui-même, mais l'interaction entre l'agent et l'environnement. Cela est dû au fait que l'agent expérimente l'environnement par rapport à son propre point de vue, et qu'il est limité par les restrictions de ses senseurs et actuateurs. En effet, ce que l'agent apprend est le modèle de la dynamique des interactions entre lui et le monde, ce n'est pas un modèle direct du monde.

(261) Le modèle du monde Ψ possible est donc un modèle anticipatoire, qui décrit, du point de vue de l'agent, la régularité des transformations dans les propriétés de l'environnement (et de son propre corps) au fil du temps, en fonction de l'observation qu'il fait et des actions qu'il exécute.

3.1.2. Construction de Politiques d'Actions

(262) Le problème de la *construction d'une politique d'actions* est connu dans la littérature d'apprentissage automatique sous le nom de problème de *décision séquentielle*. Il s'agit de donner les moyens à l'agent d'apprendre une bonne politique des actions par l'expérimentation dans un environnement quelconque.

(263) Contrairement à l'apprentissage supervisé, dans lequel l'agent dispose d'une source d'exemples qui désignent les comportements corrects ou désirables, dans l'apprentissage par renforcement l'agent doit apprendre par sa propre expérience, par essais et erreurs, et avec une notion inexacte et possiblement tardive de la vraie utilité de ses actions.

(264) Dans la définition classique, (KAELBLING et al., 1996), (SUTTON; BARTO, 1998), (RUSSELL; NORVIG, 1995), l'agent reçoit un signal de renforcement (une valeur positive ou négative) souvent sporadique, qui représente des récompenses ou des punitions, de différentes intensités, données à l'agent selon les actions qu'il effectue. Basé sur l'observation continue des récompenses reçues au cours de l'interaction avec le monde, l'agent a pour mission de construire une politique d'actions qui maximise la valeur moyenne des récompenses au fil du temps.

(265) Bien qu'il existe des méthodes d'apprentissage par renforcement « sans modèle » (QUINLAN, 1986, 1993), qui apprennent une politique d'actions à partir des interactions immédiates, dans ce travail nous considérons le problème de la construction d'une politique d'actions lié au problème de la construction d'un modèle du monde. Dans ce cas, l'agent apprend un modèle du monde, et calcule une politique à partir de lui.

(266) Dans l'architecture CAES, les signaux p de la perception sont associés à des signaux d'évaluation donnés par les fonctions л, qui indiquent la valeur immédiate de certaines situations pour l'agent. Dans ce cas, les entrées d'un algorithme d'apprentissage de politique d'actions sont le modèle du monde courant $\Psi^{(t)}$, et le modèle de récompenses courant $\text{Л}\mathfrak{b}^{(t)}$. Le modèle du monde $\Psi : P \times C \rightarrow P'$ est une fonction anticipatoire, où p représente une observation, c représente une action exécutée par l'agent, et p' représente l'observation attendue au moment suivant (une anticipation). Le modèle de récompenses $\text{Л}\mathfrak{b} : P \rightarrow \Re$ représente la valeur attendue du signal évaluatif л quand p est observé.

(267) La tâche de l'algorithme est de définir une politique d'actions à partir du modèle du monde (Ψ) et du modèle d'évaluation (л). La politique d'actions est représentée comme une correspondance entre les situations possibles de l'environnement et les meilleures actions à prendre lorsque l'agent y est ($\pi : P \rightarrow C$). La politique optimale est celle qui définit des séquences d'actions qui maximisent la somme décomptée des signaux d'évaluation sur un horizon de temps à long terme (c'est-à-dire quelque chose de plus grand que l'immédiat). Cependant, pour le genre de problème posé lorsqu'on utilise l'architecture CAES, la solution (autrement dit, une bonne politique) peut être basé sur la stabilité de la dynamique plutôt que sur son optimalité. Des solutions « suffisamment bonnes » sont valables et peuvent permettre à l'agent de se maintenir très longtemps dans les limites de viabilité.

3.1.3. Apprentissage Actif, sur Horizon Infini, et Incrémental

(268) En suivant les prémisses adoptées par ce travail, en particulier l'utilisation de l'architecture CAES pour définir le rapport entre l'agent et l'environnement, les problèmes d'apprentissage sont posés à l'agent en tant que situations d'*apprentissage actif*, *sur un horizon infini*, et *incrémental*.

3.1.3.1. Apprentissage Actif

(269) Dans un système agent-environnement tel que CAES, l'agent interfère avec le flux du système global à travers ses actions. L'agent perturbe l'environnement et observe les résultats de cette perturbation. Quand l'agent est plus qu'un simple observateur, c'est-à-dire quand c'est lui qui modifie l'état de l'environnement par le biais de ses actions, et quand il doit en même temps apprendre un modèle du monde à partir de ces interactions, alors on parle d'apprentissage actif (RON; RUBINFELD, 1997).

(270) Contrairement au cas de l'apprentissage passif, dans lequel l'agent est seulement un observateur, sans avoir de contrôle sur le flux du processus, le cas de l'apprentissage actif lie le problème d'apprentissage de modèle du monde aux problèmes de décision et de contrôle. Dans l'apprentissage actif c'est l'agent qui choisit les actions à prendre, et, donc, pour qu'il puisse apprendre un modèle du monde adéquat, il a besoin de résoudre le dilemme entre *explorer ou exploiter*. Plus précisément, l'agent doit adopter une stratégie pour choisir entre des actions qui visent à découvrir de nouvelles choses de l'environnement (*exploration*), ou des actions qui utilisent les connaissances déjà acquises afin de conduire l'agent à satisfaire ses objectifs (*exploitation*).

(271) S'agissant de l'apprentissage de politiques d'action, s'il n'y a pas une stratégie d'exploration, l'algorithme finit par converger vers un *maximum local*, c'est-à-dire une politique qui n'est bonne que si on la compare à ses voisines proches dans l'espace de politiques.

(272) Bien que l'apprentissage actif implique cette difficulté supplémentaire, c'est-à-dire, de trouver une heuristique appropriée pour résoudre le dilemme entre explorer et exploiter, lorsque le problème est bien résolu, la capacité d'apprentissage de l'agent augmente, car il peut effectuer des « expériences » dans l'environnement, en élaborant des stratégies qui visent délibérément à explorer les chemins mal connus. L'apprentissage passif est, en fait, plus restrictif, parce que s'il n'y a pas de contrôle sur les actions, il faut assurer statistiquement que des séquences aléatoires peuvent explorer de manière adéquate l'environnement (RON; RUBINFELD, 1997).

3.1.3.2. Apprentissage sur un Horizon Infini

(273) Dans toutes les architectures inspirées par des modèles naturalistes (telle que CAES), l'agent est inséré dans l'environnement pour expérimenter le monde pendant le

temps où il est « en vie », de la même façon que le fait un organisme vivant. Si l'agent a besoin d'apprendre un modèle du monde dans ces conditions, alors on parle d'apprentissage sur un horizon infini.

(274) Plusieurs problèmes en IA sont modélisés sur un horizon épisodique. C'est-à-dire que, en général, il y a des états finaux précis et, lorsqu'ils sont atteints, le système redémarre, repositionnant l'agent dans l'état initial. Dans certains cas, l'agent a lui-même accès à une action de réinitialisation (*reset*), qui permet de replacer le système dans son état initial. Dans d'autres cas, il y a un nombre maximum de cycles par épisode.

(275) L'apprentissage sur un horizon infini est plus difficile, surtout s'agissant de problèmes partiellement observables, car il n'y a pas de ressource à réinitialiser, et donc il n'y a pas moyen de revenir directement à un état connu. L'agent doit apprendre à prévoir les prochaines observations en se basant seulement sur une « marche » unique et continue dans l'environnement.

(276) Le mécanisme de réinitialisation permet à l'agent d'être sûr, au moins, d'un état initial, à partir duquel il commence toujours. Certains algorithmes d'apprentissage se sont basés sur ce point pour construire leurs modèles (RON; RUBINFELD, 1997). Sans cette ressource, la certitude d'avoir retrouvé un état déjà visité disparaît, et les algorithmes d'apprentissage fondés sur ce principe deviennent non-viables.

3.1.3.3. Apprentissage Incrémental

(277) L'architecture CAES impose à l'agent un mode d'apprentissage incrémental. L'agent doit apprendre peu à peu, tout en effectuant en même temps ses activités. L'agent doit construire le modèle du monde d'une façon progressive, et chaque expérience particulière est alors une nouvelle information qui peut être utilisée pour le raffiner.

(278) L'apprentissage incrémental se fait en ligne, c'est-à-dire qu'il n'y a pas de séparation entre le temps d'apprendre et le temps d'agir. L'agent ne dispose pas d'un ensemble d'exemples d'entrainement à priori. Avoir une base de données de cas disponible dès le début du processus d'apprentissage permettrait de construire d'un coup une hypothèse après avoir examiné tous les exemples. Dans le mode incrémental, toutefois, la construction du modèle est à la merci d'un mauvais départ, induit par des exemples initiaux non-représentatifs, et l'algorithme d'apprentissage doit tenir compte

de cette possibilité, c'est-à-dire que le mécanisme doit être capable de défaire certaines choix prises pendant la construction du modèle.

3.1.4. Représentation du Système par un Processus de Décision Markovien

(279) De nombreux algorithmes proposés pour les problèmes de décision et d'apprentissage utilisent des représentations de l'environnement basées sur l'énumération des états, souvent définies comme un *Processus de Décision Markovien* (MDP).

(280) Un MDP fournit à l'agent l'information complète sur l'état où il se trouve, c'est-à-dire que l'agent perçoit les états du système directement. Cependant, dans de nombreux problèmes cette information complète n'est pas possible ou n'est pas disponible. Dans ce cas, un environnement est typiquement représenté par un *Processus de Décision Markovien Partiellement Observable* (POMDP), dans lequel l'ensemble des états de l'environnement n'est que partiellement observable par l'agent. Dans un POMDP, l'agent a l'accès à une information qui n'indique que de manière partielle et indirecte l'état actuel du système.

(281) L'inconvénient de ce type de représentation qui fait l'énumération des états, tels que le MDP et le POMDP, c'est que la taille du problème croît exponentiellement par rapport au nombre d'attributs considérés, ce qui rend de telles solutions non-viables pour des grands problèmes. Ainsi, la communauté d'IA utilise, de plus en plus, des représentations factorisées, qui n'énumèrent pas les états, mais qui traitent directement les propriétés.

(282) L'utilisation de Processus de Décision Markovien Factorisés (FMDP) permet d'explorer la structure de l'environnement, qui peut donc être représentée sous une forme compacte, même si le MDP correspondant est exponentiellement grand (GUESTRIN et al., 2003). En plus, ce formalisme peut également être étendu afin de représenter l'observabilité partielle, en constituant alors un *Processus de Décision Markovien Factorisé et Partiellement Observable* (FPOMDP).

3.1.4.1. Processus de Décision Markovien

(283) Le Processus de Décision Markovien a été introduit par (BELLMAN, 1957) et (HOWARD, 1960), et est devenu populaire parmi la communauté académique dans les années 1990, (PUTERMAN, 1994), (SUTTON; BARTO, 1998), (RUSSELL; NORVIG,

1995), (FEINBERG; SHWARTZ, 2002). Un MDP représente l'environnement à travers l'énumération des états, d'une façon similaire à une *machine d'états* dont la fonction de transition peut être non-déterministe. Dans le MDP, le flux du système dépend des actions exécutées par l'agent, et il y a un signal de récompense pour certaines transitions, ce qui caractérise un problème de contrôle.

(284) Dans le cadre des agents, un MDP peut être formalisé par un quadruplet $\{Q, A, \delta, r\}$, selon la définition 3.1, où Q est l'ensemble fini et non vide des états du système, A est l'ensemble des actions de l'agent, δ est la fonction de transition du système, et r est la fonction de récompense. Si le système est non-déterministe, les fonctions de transition et de récompense constituent des matrices de distribution de probabilités. Il peut y avoir une fonction δ^0 pour définir l'état initial.

Un MDP est un quadruplet:

$\in = \{Q, A, \delta, r\}$

où

$Q = \{q_1, q_2, \ldots, q_{|Q|}\}$ est un ensemble fini et non-vide d'*états*

$A = \{a_1, a_2, \ldots, a_{|A|}\}$ est un ensemble fini et non-vide d'*actions* de l'agent

$\delta : Q \times A \rightarrow \Pi(Q)$ est une fonction probabiliste de *transition d'états*

définie par une matrice $\delta = prob(q' \mid q, a)$

$r : Q \times A \rightarrow \Pi(\Re)$ est une fonction probabiliste de *récompense*

définie par une matrice $r = prob(r' \mid q, a)$

en plus

$\delta^0 : \rightarrow \Pi(Q)$ est une fonction probabiliste d'état initial

définie par une matrice $\delta^0 = prob(q^{(0)})$

Définition 3.1: Processus de Décision Markovien.

(285) La représentation de l'environnement sous la forme d'un MDP suppose implicitement que le système global est en accord avec quatre restrictions (BOUTILIER et al., 2000), (McALLESTER; SINGH, 1999). Initialement, un MDP est un modèle discret. Il discrétise l'espace de phase du système en un nombre fini d'états, et il discrétise aussi le temps, en le fragmentant à des instants bien définis (des cycles). Autrement dit, il n'y a pas d'états intermédiaires entre deux états quelconques q_i et q_j, et de même, aucun événement n'est possible dans le système entre deux instants du temps successifs t et t'.

(286) Deuxièmement, un MDP décrit les transitions entre les états par des fonctions probabilistes markoviennes de premier ordre. L'hypothèse de Markov suppose que la probabilité d'occurrence d'une transition entre les états q et q' ne dépend que de l'état actuel du système (q) et de la dernière action prise par l'agent (a), et donc il n'est pas nécessaire de considérer une séquence historique d'états passés ($q^{(t-1)}$, $q^{(t-2)}$, ...). Cela signifie que l'histoire n'a aucune influence sur le déroulement du processus quand le présent est complètement spécifié, ou, par ailleurs, que l'histoire pertinente est déjà intégrée dans la représentation des états.

(287) Troisièmement, un MDP représente un système stationnaire, dont la fonction de transition entre les états ne change pas au fil du temps, c'est-à-dire que les probabilités d'occurrence d'une transition sont indépendantes de l'instant particulier dans lequel elles se produisent, et donc pour deux instants du temps t_i et t_j, $\delta^{(i)} = \delta^{(j)}$.

(288) Enfin, un MDP définit un environnement complètement observable, où toute l'information nécessaire pour décrire la fonction d'évolution du système est à la disposition de l'agent. Ainsi, la représentation par MDP suppose que la perception de l'agent l'informe de façon directe et sans équivoque de l'état actuel (q) dans lequel le système se trouve.

3.1.4.2. Représentation de l'Observation Partielle

(289) Dans de nombreux problèmes, surtout ceux du monde réel, l'information complète sur l'état actuel de l'environnement n'est pas disponible. Les problèmes de ce type sont généralement représentés par des *Processus de Décision Markovien Partiellement Observables* (POMDPs), selon (MEULEAU et al., 1999), (SHANI et al., 2005) et (HOLMES; ISBELL, 2006). La définition du POMDP a été initialement proposée par (SMALLWOOD; SONDIK, 1973), en devenant populaire après les travaux réalisés par (CHRISTMAN, 1992) et (KAELBLING et al., 1994, 1998).

(290) Le POMDP, selon la définition 3.2, est une extension du MDP dans laquelle on ajoute un ensemble d'observations et une fonction d'observation. Dans cette représentation, l'état q de l'environnement n'est pas accessible à l'agent. L'agent n'a accès qu'à une observation incomplète donnée en fonction de l'état sous-jacent de l'environnement.

Un POMDP est un sextuplet:

$$\epsilon = \{Q, A, O, \delta, \gamma, r\}$$

où

$Q = \{q_1, q_2, \ldots, q_{|Q|}\}$ est un ensemble fini et non-vide d'*états*

$A = \{a_1, a_2, \ldots, a_{|A|}\}$ est un ensemble fini et non-vide d'*actions* de l'agent

$O = \{o_1, o_2, \ldots, o_{|O|}\}$ est un ensemble fini et non-vide d'*observations* de l'agent

$\delta : Q \times A \rightarrow \Pi(Q)$ est une fonction probabiliste de *transition d'états*

définie par une matrice $\delta = prob(q' \mid q, a)$

$\gamma : Q \times A \rightarrow \Pi(O)$ est une fonction probabiliste d'*observation*

définie par une matrice $\gamma = prob(o' \mid q, a)$

$r : Q \times A \rightarrow \Pi(\Re)$ est une fonction probabiliste de *récompense*

définie par une matrice $r = prob(r' \mid q, a)$

en plus

$\delta^0 : \rightarrow \Pi(Q)$ est une fonction probabiliste d'état initial

définie par une matrice $\delta^0 = prob(q^{(0)})$

Définition 3.2: Processus de Décision Markovien Partiellement Observable.

(291) Un POMDP définit des problèmes de décision et d'apprentissage d'une manière plus réaliste qu'un MDP complètement observable, pourtant la difficulté de résoudre ces problèmes augmente considérablement. La complexité résultant de l'absence d'informations sur l'état limite l'application des POMDPs à des petits problèmes, ou à des domaines avec des dimensions très restreintes (MEULEAU et al., 1999).

(292) La solution d'un POMDP est beaucoup plus coûteuse que celle de son correspondant MDP, et l'utilisation de représentations « plates » (*flat*), basées sur l'énumération des états et des actions, est le principal facteur limitant. Plusieurs études récentes sont en train de mettre l'accent sur des représentations structurées, en utilisant des états et des actions factorisés, en tant que moyen de surpasser le problème de la extensibilité pour des problèmes d'apprentissage et de décision dans des environnements partiellement observables (POUPART; BOUTILIER, 2004), (SHANI et al., 2008).

3.1.4.3. Question d'Extensibilité

(293) Une question d'extensibilité se pose dans les représentations basées sur l'énumération d'états, telles que les MDPs et les POMDPs. La quantité d'exemples nécessaires pour faire apprendre un certain concept augmente de façon exponentielle

avec le nombre d'attributs utilisés pour les représenter (VALIANT, 1984). Ce type de représentation n'est viable que s'il existe déjà un modèle, c'est-à-dire s'il y a une connaissance sur le domaine qui simplifie l'ensemble des caractéristiques à considérer dans l'environnement, en rendant possible de le représenter à travers un nombre limité d'états pertinents.

(294) Si un agent est inséré dans un environnement qui lui est inconnu (ce qui est plus typique), alors le nombre d'états provient de la combinaison croisée de chacune de ses perceptions. Par exemple, dans un univers composé par n variables binaires, le nombre d'états énumérés est $|Q| = 2^n$. Ainsi, si l'on considère que chaque perception décrit une propriété de l'environnement, alors le problème est que le nombre d'états augmente de façon exponentielle par rapport au nombre de propriétés considérées. Ce phénomène est appelé *le fléau de la dimension* (BELLMAN, 1961).

(295) En plus, un algorithme d'apprentissage de modèles du monde basé sur une représentation énumérée des états affronterait un espace de recherche de taille exponentielle par rapport au nombre d'états, qui est déjà lui-même exponentiel par rapport au nombre de propriétés. C'est-à-dire qu'il peut exister $|Q|^n$ environnements différents représentables à partir de n propriétés. Par conséquent, dans des environnements complexes, tel que le monde réel, un problème représenté à travers l'énumération d'états devient intraitable (DEGRIS et al., 2006).

(296) L'utilisation d'une représentation factorisée peut garantir que la taille du problème ne dépasse pas un ordre polynomial acceptable (BOUTILIER et al., 2000). Cependant, pour que ce soit vrai, l'environnement en question doit être bien structuré, comme on le verra dans la suite.

3.1.4.4. Factorisation des Ensembles Énumérés

(297) Pour éviter l'énumération des états, on utilise des représentations fondées directement sur les caractéristiques de l'environnement, c'est-à-dire que, au lieu d'énumérer une liste d'états atomiques, l'environnement est décrit à travers un ensemble de propriétés ou d'attributs. Le formalisme utilisé récemment à cette fin est le *Processus de Décision Markovien Factorisé* (FMDP), initialement présenté par (BOUTILIER et al., 1995).

(298) Un FMDP peut être formalisé comme un quadruplet $\{X, C, \tau, л\}$, selon la définition 3.3, où X est l'ensemble des propriétés (qui définissent l'état de l'environnement), C est l'ensemble des variables de contrôle (qui définissent les actions de l'agent), τ définit un ensemble de fonctions de transformation, de sorte que τ_i décrive l'évolution de la propriété i, et $л$ est un ensemble de fonctions de récompense, de telle façon que $л_i$ soit la fonction de récompense factorisée pour la propriété x_i. Si le FMDP est non-déterministe, alors les fonctions définissent des distributions de probabilité.

Un FMDP est un quadruplet:

$$Є = \{X, C, \tau, л\}$$

où

$X = \{X_1, X_2, ..., X_{|X|}\}$ est un ensemble de *propriétés*

$C = \{C_1, C_2, ..., C_{|C|}\}$ est un ensemble de *variables de contrôle*

$\tau = \{\tau_1, \tau_2, ..., \tau_{|X|}\}$ est un ensemble de fonctions de *transformation*

tel que $\tau_i : X \times C \rightarrow \Pi(X_i)$

$л = \{л_1, л_2, ..., л_{|X|}\}$ est un ensemble de fonctions d'*évaluation*

tel que $л_i : X_i \rightarrow \Pi(\Re)$

Définition 3.3: Processus de Décision Markovien Factorisé.

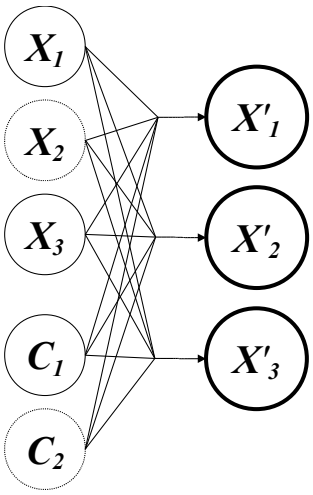
(299) L'utilisation des FMDPs rend possible qu'on prenne profit de la structure de l'environnement, permettant de le représenter d'une façon compacte, même si le MDP correspondant est exponentiellement grand. Dans cette formalisation, les états, les actions, et les récompenses, sont factorisés en propriétés, qui sont représentées par des variables aléatoires. Les avantages de la factorisation sont discutés dans (GUESTRIN et al., 2003), (BOUTILIER et al., 2000), (JONSSON; BARTO, 2005), (DEGRIS et al., 2006, 2008), et (TRIVIÑO; MORALES, 2000).

(300) Vu qu'il n'est plus nécessaire d'énumérer tous les états possibles de l'espace, définis par la combinaison des dimensions fournies par les propriétés, alors un FMDP peut représenter la fonction de transformation de chaque propriété de façon indépendante. La fonction d'évolution du système global équivaut à la combinaison de ces fonctions de transformation particulières.

(301) L'ensemble des fonctions de transformation peut être représenté par un *Réseau Bayésien Dynamique* (DBN) spécial, tel qu'un graphe des deux couches, acyclique et orienté, selon (DEAN; KANAZAWA, 1989) et (DARWICHE; GOLDSZMIDT, 1994).

Les nœuds de la première couche représentent les propriétés et les variables de contrôle ($X \cup C$), qui indiquent respectivement l'état de l'environnement et les actions qui peuvent être effectuées par l'agent au temps *t*. La deuxième couche représente seulement les propriétés du système (X'), en indiquant ses valeurs à l'instant suivant $t + 1$.

(302) Dans le cas d'un DBN complet, chaque nœud de la deuxième couche est associé à une distribution de probabilité conditionnelle, qui décrit le comportement de la propriété en fonction de la valeur des nœuds de la première couche, comme cela est montré sur la figure 3.1.



$$\tau_i : X \times C \rightarrow \Pi(X_i)$$

Figure 3.1: Exemple d'un réseau bayésien dynamique (DBN).

Un DBN de deux couches, complètement connecté, acyclique et orienté, représente les fonctions de transformation su système. L'exemple est composé de 3 propriétés et de 2 variables de contrôles. À gauche on représente l'état du système à temps *t*, et à droite, la transformation qu'il subit à *t+1*.

(303) Simplement pour bien remarquer l'indépendance entre chaque fonction de transformation τ_i (de chaque propriété), il serait formellement équivalent de les représenter chacune par un DBN particulier. Dans ce cas, la deuxième couche du graphe présente un seul nœud, qui indique la valeur d'une certaine propriété spécifique de l'environnement (X'_i) à l'instant suivant $t + 1$, comme le montre la figure 3.2. Le fait que chaque transformation peut être décrite indépendamment des autres est important parce que cela permet l'implémentation d'un mécanisme d'apprentissage que parallélise le processus de construction du modèle du monde.

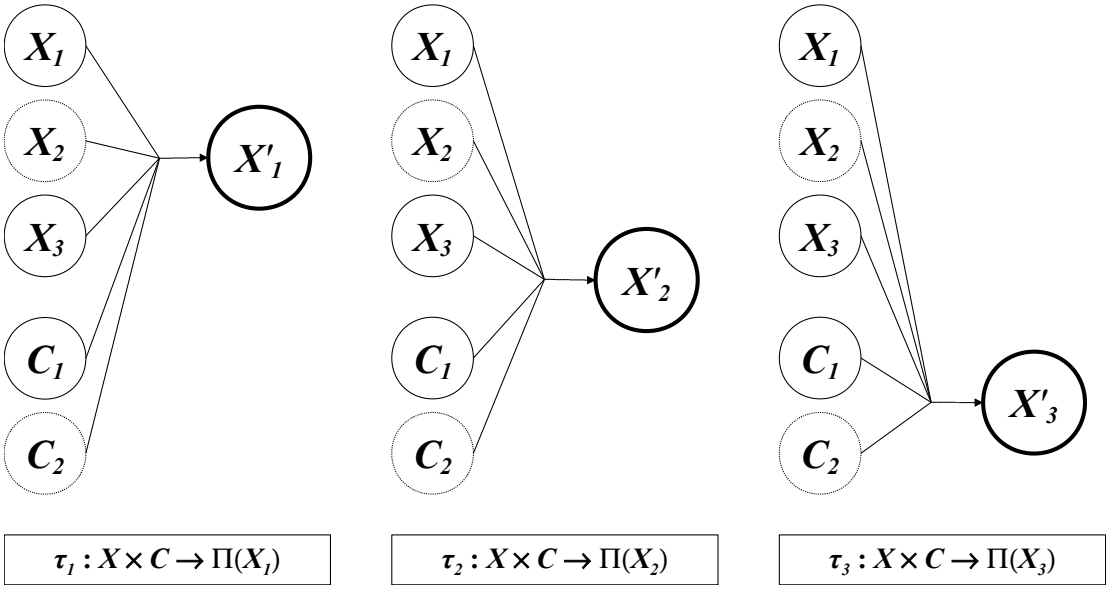


Figure 3.2: Exemple de DBNs complets mais indépendants.

3.1.5. Environnements Structurés

(304) En général, les environnements complexes sont décrits par un très grand nombre de propriétés. Toutefois, si l'environnement est structuré, alors les événements ont des causes précises et limitées. En d'autres termes, dans des *environnements structurés*, pour expliquer la dynamique d'une certaine propriété X_i, généralement il suffit de faire attention à un petit ensemble de *variables pertinentes* à sa fonction de transformation, cet ensemble étant défini par $rel(\tau_i) \subset (X \cup C)$.

(305) C'est pour cette raison qu'un DBN peut représenter la fonction de transformation d'une propriété d'une façon compacte, si on suppose que l'ensemble de propriétés pertinentes est connu. Dans la première couche, il y a seulement ce sous-ensemble de nœuds pertinents $rel(\tau_i)$, au lieu de représenter le domaine complet $X \times C$. Dans la deuxième couche, il n'y a que le nœud X_i' relatif à la propriété anticipée. La figure 3.3 montre un exemple des fonctions de transformation représentées par DBNs dans un environnement hypothétique. Seules les variables pertinentes dans la première couche du graphe sont connectées avec le nœud d'anticipation dans la deuxième couche.

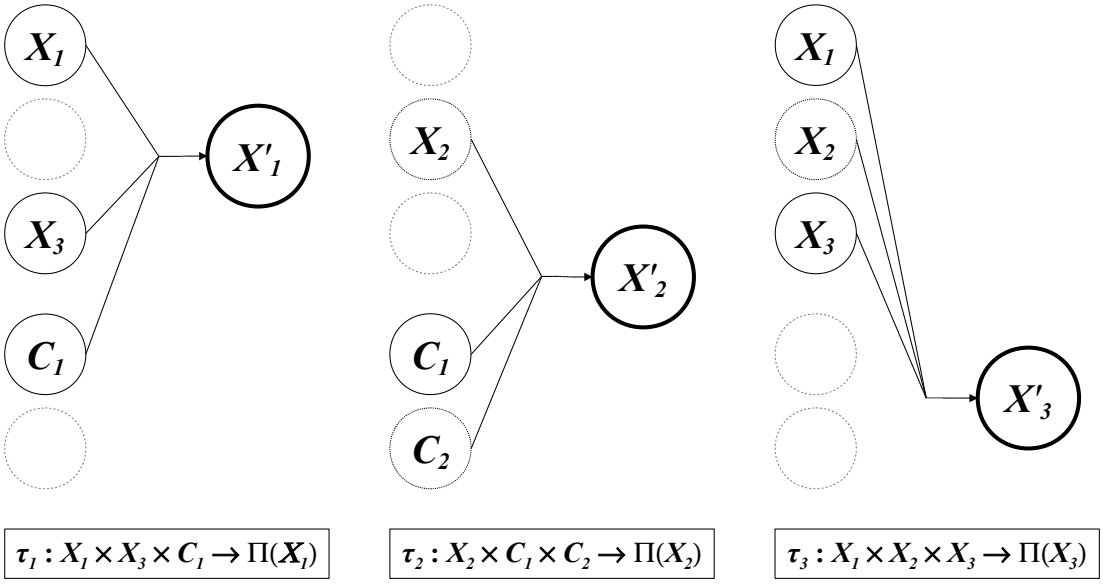


Figure 3.3: Exemple de DBNs indépendants et compacts.

Les DBNs sont compactes parce que la première couche contient seulement les variables pertinentes à la description de la transformation.

3.1.5.1. Pertinence

(306) La notion de pertinence est liée à celle de causalité. Les variables pertinentes pour décrire la dynamique d'une propriété donnée sont, en principe, les causes déterminantes pour la transformation ou pour la permanence de son état. L'ensemble de variables pertinentes pour la fonction de transformation spécifique d'une certaine propriété est défini par $rel(\tau_i)$. L'ensemble des propriétés pertinentes pour le système global, $rel(\tau)$, est donné par l'union de ces ensembles particuliers, de façon que $rel(\tau) = rel(\tau_1) \cup rel(\tau_2) \cup \ldots \cup rel(\tau_{|X|})$.

(307) Une propriété est pertinente s'il existe au moins un phénomène qui n'est explicable qu'en fonction de celle-là. Autrement dit, X_i est une propriété pertinente à la transformation τ_j s'il y a une paire d'expériences exp_1 et exp_2 (du type $\{x, c, x'_j\}$ correspondant respectivement à l'état du système global, à l'action de l'agent, et à la transformation de la propriété X_j vérifiée à l'instant suivant) pour lesquelles x'_j soit divergent et $\{x, c\}$ soit différentiable seulement par la composante x_i (BLUM; LANGLEY, 1997). C'est-à-dire que la valeur de x'_j, qui est régie par la fonction de transformation τ_j, est conditionnellement dépendante de la valeur de x_i.

3.1.5.2. Degré de Structuration d'un Environnement

(308) Un FMDP peut être classé selon son *degré de structuration* (φ). L'idée de cette mesure est la suivante: plus réduite est le nombre de causes qui déterminent les phénomènes, plus l'environnement est structuré. Autrement dit, le degré de structuration indique à quel niveau l'information nécessaire à la description de sa dynamique est concentrée. Plus le nombre moyen de propriétés pertinentes pour décrire les transformations est petit, plus l'environnement est structuré. Donc, le degré de structuration indique à quel point les fonctions de transformation peuvent être représentées sous une forme compacte.

(309) Le degré de structuration φ_i d'une certaine propriété X_i est calculé par rapport à la fonction de transformation correspondante τ_i. Il est donné par la raison inverse entre le nombre de propriétés pertinentes pour décrire la transformation, $rel(\tau_i)$, et le nombre de propriétés existantes dans le domaine complet $X \cup C$, selon l'équation 3.1.

$$\varphi_i = 1 - \frac{|rel(\tau_i)|}{|X \cup C|} \quad \textit{(eq. 3.1)}$$

(310) Le degré de structuration total φ de l'environnement est déterminé par la moyenne du degré de structuration des fonctions de transformation particulières, selon l'équation 3.2.

$$\varphi = \frac{\sum_{i=1}^{|X|} \varphi_i}{|X|} \quad \textit{(eq. 3.2)}$$

(311) Si on imagine φ comme un axe, dans une direction de cet axe on retrouve des environnements dont les propriétés concentrent de plus en plus l'information sur le fonctionnement du système, et dans l'autre direction, on retrouve des environnements dont les propriétés ont peu de signification toutes seules, en ce qui concerne la détermination de la dynamique de l'environnement.

(312) Dans un cas extrême, si $rel(\tau_i) = \varnothing$, alors $|rel(\tau_i)| = 0$, et donc $\varphi_i = 1$. Dans ce cas, la transformation τ_i est conditionnellement indépendante, car il est possible de définir la valeur de X'_i sans faire référence à aucune autre propriété ou signal de contrôle. À l'inverse, quand $rel(\tau_i) = X \cup C$, alors $\varphi_i = 0$. Dans ce cas, la fonction de transformation τ_i, qui définit la valeur de X'_i, est conditionnellement dépendante de l'ensemble des

propriétés dans X et C. En effet, si l'environnement n'est pas trop structuré, la représentation factorisée peut devenir plus coûteuse que l'énumération complète des états.

(313) Ainsi, la réduction de la complexité de la représentation d'un FMDP par rapport à son MDP corrélatif est directement proportionnelle à son degré de structuration. Plus la valeur de φ est élevée, plus petit est le nombre moyen de variables pertinentes pour décrire la dynamique du système, et plus compacte est une représentation basée sur la factorisation des états en propriétés.

(314) Dans le cas contraire, plus la valeur de φ diminue, plus l'information nécessaire pour décrire les transformations est distribuée entre les propriétés, et moins il est intéressant de faire une représentation factorisée par rapport à la construction d'un modèle directement fondé sur l'énumération des états.

(315) De retour à l'exemple illustré dans la figure 3.3, on calcule le degré de structuration de la propriété X_1 par la proportion entre les propriétés pertinentes pour décrire la fonction de transformation, $|rel(\tau_1)| = 3$, et le nombre total de propriétés qui pouvaient être utilisées, $|X \cup C| = 5$, résultant en un taux de $\varphi_1 = 0{,}6$.

(316) Si on utilise la formule proposée pour la mesure de structuration, on peut définir un environnement bien structuré comme celui dont le nombre moyen de propriétés pertinentes pour les transformations reste dans un ordre maximal logarithmique par rapport à la quantité totale de propriétés, c'est-à-dire $|rel(\tau)| \leq log_b(|X \cup C|)$.

3.1.5.3. Des Implications pour l'Apprentissage

(317) La figure 3.4 illustre la raison pour laquelle les algorithmes classiques d'apprentissage sont inefficaces. Ils sont ancrés dans l'énumération des états, ce qui est exponentiel par rapport au nombre de variables du problème. Différemment, les algorithmes basés sur une représentation factorisée, et qui cherchent à profiter de la structure de l'environnement, présentent une stratégie quasi-polynomial.

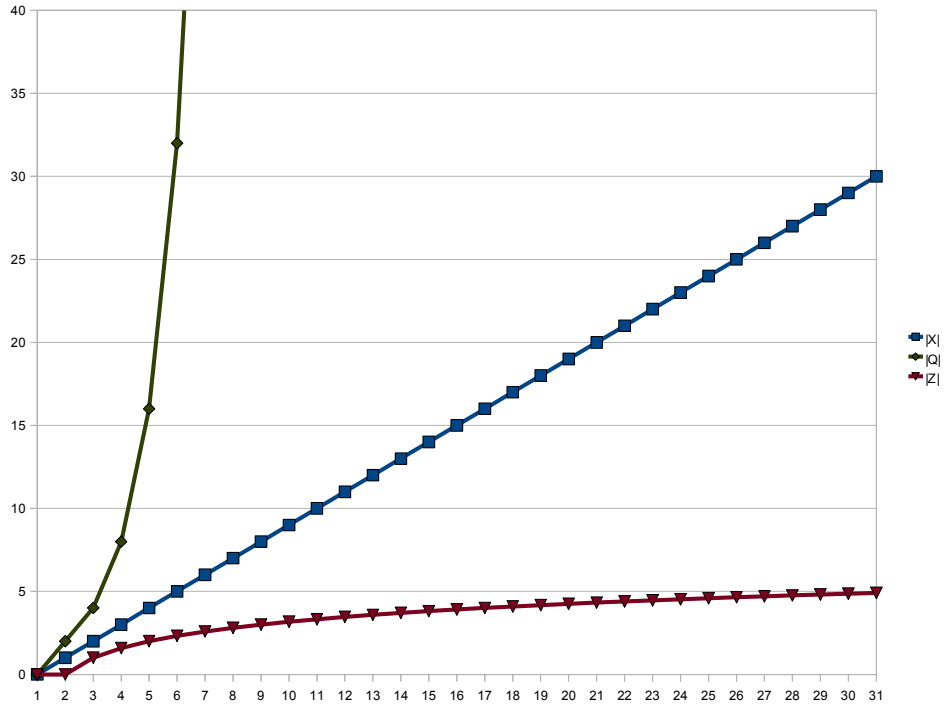


Figure 3.4: Relation de complexité entre les propriétés, les états, et la pertinence.

Sur le graphique, la ligne |X| représente la quantité de propriétés du problème (croissance linéaire), la ligne |Q| représente la quantité d'états nécessaire pour représenter de façon plate les combinaisons des propriétés (exponentielle), et la ligne |rel| représente la moyenne de propriétés pertinentes pour décrire les transformations dans un environnement bien structuré (logarithmique). On considère un univers de variables binaires.

(318) En ce qui concerne le problème de l'apprentissage de modèles du monde, la question est que ce ne sont pas toutes les propriétés de l'environnement qui sont pertinentes pour décrire les fonctions de transformation. L'existence d'une quantité importante de propriétés non pertinentes peut compliquer considérablement la tâche de l'algorithme d'apprentissage. En général, cette situation implique la tâche supplémentaire d'identifier les propriétés qui sont pertinentes dans l'ensemble total des propriétés, problème qui n'a pas une solution simple. Quelques références sur le sujet sont (GUYON et al. 2006), (LIU; MOTODA, 2007), (BLUM; LANGLEY, 1997), (GUYON; ELISSEEFF, 2003), (SIEDLECKI, SKLANSKY, 1993).

(319) Potentiellement, l'ensemble des propriétés pertinentes à une fonction de transformation τ_i donnée est un sous-espace parmi les combinaisons possibles de l'espace $X \times C$. Ces domaines possibles de la fonction τ_i constituent une hiérarchie partiellement ordonnée par l'inclusion d'un domaine dans l'autre. Si on considère des

propriétés binaires, on a $2^{|X \times C|}$ combinaisons d'espaces possibles. La figure 3.5 illustre cet espace de combinaisons possibles de pertinence, dans un exemple composé par 3 variables binaires, dont deux propriétés, et une variable de contrôle.

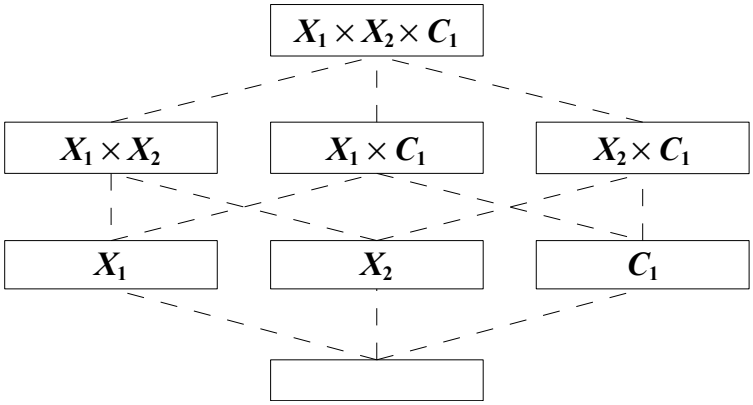


Figure 3.5: Hiérarchie partiellement ordonné des domaines possibles.

Chaque fonction de régularité a un ensemble de variables pertinentes, qui compose son domaine. Pourtant, l'espace des domaines possibles totalise $2^{|X \times C|}$ combinaisons.

(320) Par exemple, le domaine total $X \times C$, défini pour $\varphi = 0$, contient tous les autres. En outre, le domaine vide, défini pour $\varphi = 1$, représente le plus petit ensemble de propriétés, et il est contenu dans tous les autres. On établit que l'ensemble $rel(\tau_i)$ des propriétés pertinentes à la fonction τ_i est le plus petit sous-ensemble contenu dans $X \cup C$ qui soit suffisant pour décrire la transformation de la propriété X_i sans perte de précision.

3.1.6. Processus Factorisé et Partiellement Observable

(321) Un *Processus de Décision Markovien Factorisé et Partiellement Observable* (FPOMDP) peut donc être considéré comme une extension du FMDP capable de représenter l'observabilité partielle, ou alors comme une représentation factorisée du POMDP. Il s'agit de la jonction du FMDP et du POMDP, ce qui compose un MDP factorisé et partiellement observable. Des propositions à cet effet sont présentées dans (BOUTILIER; POOLE, 1996), (HANSEN; FENG, 2000), (GUESTRIN et al., 2001), (POUPART; BOUTILIER, 2004), (POUPART, 2005), (SHANI et al., 2008), et (SIM et al., 2008).

(322) Un FPOMDP est formalisé comme un quintuplet $\{P, H, C, \tau, л\}$, selon la définition 3.4, où P est l'ensemble des propriétés observables (accessibles à travers la perception), et H est l'ensemble des propriétés cachées (non observables directement par

la perception), C est l'ensemble des variables de contrôle (qui définissent les possibilités d'action de l'agent), τ est la fonction d'évolution factorisée du système, et $л$ est la fonction factorisée de récompense.

Un FPOMDP est un quintuplet:

$$Є = \{P, H, C, \tau, л\}$$

où

$P = \{P_1, P_2, ..., P_{|P|}\}$ est un ensemble de *propriétés observables*

$H = \{H_1, H_2, ..., H_{|H|}\}$ est un ensemble de *propriétés cachées*

$C = \{C_1, C_2, ..., C_{|C|}\}$ est un ensemble de *variables de contrôle*

$\tau = \{\tau_1, \tau_2, ..., \tau_{|X|}\}$ est un ensemble de fonctions de *transformation*

tel que $\tau_i : X \times C \rightarrow \Pi(X_i)$

$л = \{л_1, л_2, ..., л_{|X|}\}$ est un ensemble de fonctions d'*évaluation*

tel que $л_i : P_i \rightarrow \Pi(\Re)$

en plus

$$X = P \cup H$$

Définition 3.4: Processus de Décision Markovien Factorisé et Partiellement Observable.

(323) Un environnement modélisé en tant qu'un FPOMDP pose deux difficultés au problème d'apprentissage. D'une part, parmi les propriétés observables de l'environnement, toutes ne sont pas pertinentes pour décrire les fonctions de transformation, et cela implique le sous-problème de la *sélection des propriétés pertinentes*. D'autre part, parmi les propriétés pertinentes, toutes ne sont pas observables, ce qui pose le sous-problème de la *découverte des propriétés non-observables*.

3.1.6.1. CAES établit un FPOMDP pour l'esprit

(324) Dans l'architecture CAES (définie dans le chapitre 2), le problème de la construction de modèles du monde équivaut au problème de la découverte de la structure d'un FPOMDP. Dans ce modèle, le système cognitif de l'esprit de l'agent reçoit un signal de perception *p*, et induit des changements dans l'état du système via un signal de contrôle *c*. Ces signaux sont factorisés, composés donc par diverses propriétés qui forment les espaces $P = (P_1 \times P_2 \times ... \times P_{|P|})$ et $C = (C_1 \times C_2 \times ... \times C_{|C|})$. La fonction d'évolution de la perception peut être également prise avec $|P|$ fonctions de transformation τ_i pour chacune des perceptions P_i.

(325) Ainsi, la relation entre l'agent et l'environnement définie à travers des propriétés (au lieu d'une énumération d'états) devient une métaphore naturelle pour l'architecture CAES, car les signaux de perception et de contrôle sont définis comme des vecteurs, où chaque élément est associé à un senseur ou à un actuateur.

(326) Du point de vue du système cognitif (Ж), le FPOMDP (Є) à construire représente tout ce qui lui est extérieur. Cette extériorité est constituée d'abord par l'environnement externe (ξ) à l'agent, mais aussi par le corps (β), qui est en dehors de l'esprit, et par le système régulatif de l'esprit (Я), qui est en dehors du système cognitif. Par conséquent, l'univers extérieur, Є, celui qui doit être appris par le système cognitif à partir des expériences données par la succession des signaux *p* et *c* au fil du temps, est une espèce reflet de la composition de $Я \times \beta \times \xi$.

(327) L'autre relation importante, l'observabilité partielle, résulte, dans l'architecture CAES, de la façon dont la perception est générée. La fonction d'évolution de l'environnement ($f_\xi : X_\xi \times M \rightarrow X_\xi$), ainsi que la fonction d'évolution du corps ($f_\beta : X_\beta \times C \times S \rightarrow X_\beta$) sont inconnues de l'agent. Le signal perceptif (p) est définie en fonction de l'état du corps (x_β), qui à son tour est influencé par un signal de situation (s), dérivé de l'état de l'environnement (x_ξ).

(328) Il est possible qu'une certaine propriété $X_{\xi i}$ qui ne fait pas partie de la composition du signal S soit pertinente pour la détermination d'autres propriétés de l'environnement. Si elle ne fait pas partie du signal S, elle n'est donc pas accessible à l'esprit par l'intermédiaire de P, et par conséquent elle ne peut pas être utilisée directement par le mécanisme d'apprentissage pour décrire les transformations perceptives.

(329) Pour cette raison, il est tout à fait possible que la fonction de l'évolution (f_ξ) d'un système soit complètement déterministe par rapport aux états sous-jacents de l'environnement (X_ξ) et pourtant se présenter à l'agent comme un système non-déterministe en ce qui concerne les propriétés de la perception (P). De façon analogue, une propriété du corps qui constitue un paramètre important pour l'évolution de l'état du corps peut être inaccessible à l'esprit si elle ne fait pas partie de la composition du signal de perception. Ces propriétés pertinentes et inaccessibles deviennent, du point de vue de l'esprit de l'agent, des propriétés non-observables (H).

(330) Ainsi, la perception révèle à l'esprit seulement une partie de l'état du corps, et encore indirectement, et seulement une partie de l'état du monde, comme l'illustre la figure 3.6. C'est sur la base de cette information incomplète que l'esprit de l'agent doit essayer d'organiser un ensemble de connaissances capable de prédire la succession des événements.

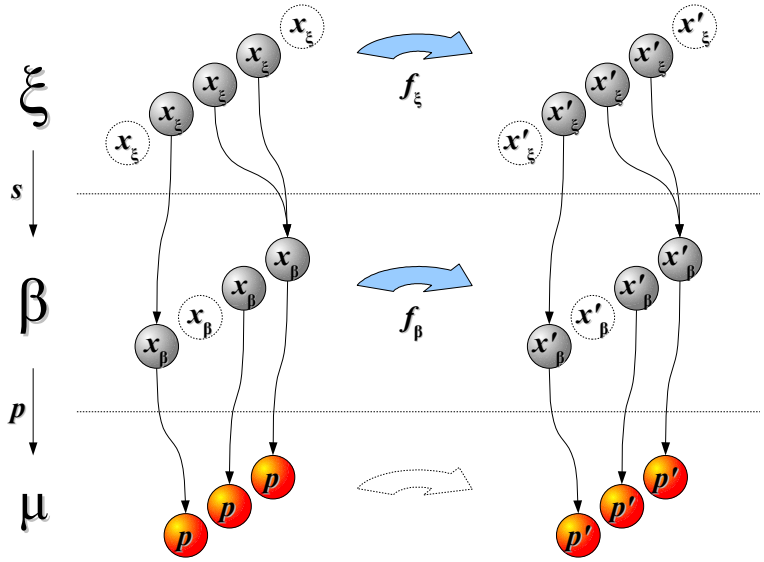


Figure 3.6: Transition d'état du système, d'un instant à l'autre. Perception indirecte et partielle.

(331) Ainsi, le problème de la construction d'un modèle du monde se pose, pour l'esprit de l'agent, comme une construction progressive basée sur l'expérience. Le problème est d'apprendre un ensemble τ de transformations du type $\tau_i : rel(\tau_i) \rightarrow X_i$ qui anticipent l'état de chaque propriété X'_i en fonction des autres propriétés et des signaux de contrôle pertinents, où $rel(\tau_i) \subset (X \times C)$.

(332) Pourtant, l'agent ne connait pas à priori les paramètres de ces fonctions. L'ensemble X est initialement équivalent à P, et H est initialisé comme un ensemble vide. L'une des tâches est donc de créer de nouveaux éléments dans l'ensemble H, et par conséquent dans X et τ, lorsque les expériences indiquent la présence d'une variable pertinente qui n'est pas observable, et ensuite, en plus de l'utiliser comme entrée pour les fonctions de transformation pour lesquelles cette nouvelle variable est pertinente, il est également nécessaire d'apprendre sa fonction de transformation elle-même. En outre, l'ensemble des variables pertinentes, $rel(\tau_i) \subset (X \times C)$, n'est pas donné a priori non plus, et c'est une tâche pour le mécanisme d'apprentissage de le découvrir.

3.1.6.2. Degré d'Accessibilité Perceptive à l'Environnement

(333) Un FPOMDP peut être analysé par rapport à l'accessibilité perceptive de l'agent à l'environnement. Le degré d'accessibilité perceptive ω_i pour la dynamique d'une certaine propriété X_i est la proportion entre la quantité de variables pertinentes observables et la quantité totale de variables pertinentes pour la description complète et exacte de sa fonction de transformation τ_i, tel que décrit par l'équation 3.3.

$$\omega_i = \frac{|rel(\tau_i) \cap (P \times C)|}{|rel(\tau_i)|} \quad \textit{(eq. 3.3)}$$

(334) Le degré d'accessibilité perceptive ω à l'environnement dans son ensemble, à son tour, est donné par la moyenne du degré d'accessibilité particulier de ses propriétés, telle que décrite par l'équation 3.4.

$$\omega = \frac{\sum_{i=0}^{|P|} \omega_i}{|P|} \quad \textit{(eq. 3.4)}$$

(335) Lorsque $\omega = 1$, alors l'environnement est totalement observable. Dans ce cas, les senseurs de l'agent perçoivent toutes les propriétés de l'environnement pertinentes à la détermination de la dynamique, c'est-à-dire $P = X$ e $H = \varnothing$, et le FPOMDP devient équivalent à un FMDP.

(336) Si $0 < \omega < 1$, alors l'environnement est partiellement observable. Dans ce cas, des états différents du système peuvent être observées par l'agent comme similaires (*perceptual aliasing*), puisque des situations équivalentes par rapport à X peuvent générer des signaux identiques sur P. Lorsque des propriétés pertinentes ne sont pas perçues, c'est-à-dire que $rel(\tau) \not\subset P$, il devient beaucoup plus difficile à l'agent d'anticiper la séquence des événements.

(337) À mesure que ω diminue, en s'approchant de 0, la proportion des propriétés cachées par rapport aux propriétés observables devient plus grande, si on considère l'ensemble des propriétés pertinentes. Si $\omega = 0$, alors aucune information nécessaire pour décrire les transformations n'est directement disponible par le biais de la perception.

(338) La figure 3.7 montre un exemple dans lequel la perception est composée de deux variables, $p = \{p_1, p_2\}$, le signal de contrôle est également composé de deux variables,

$c = \{c_1, c_2\}$, mais le système compte une propriété non-observable, $h = \{h_1\}$. La fonction de transformation τ_1 de la perception p_1, en particulier, est conditionnellement dépendante de p_1, h_1 et c_1, donc il n'y a accès qu'à 2 des 3 variables pertinentes, ce qui signifie un degré d'accessibilité $\omega_{P1} = 0{,}667$, selon l'équation 3.3. Pour les autres propriétés on trouve $\omega_{P2} = 1{,}0$ e $\omega_{H1} = 0{,}667$, ce qui définit un degré d'accessibilité global $\omega = 0{,}778$, d'après l'équation 3.4.

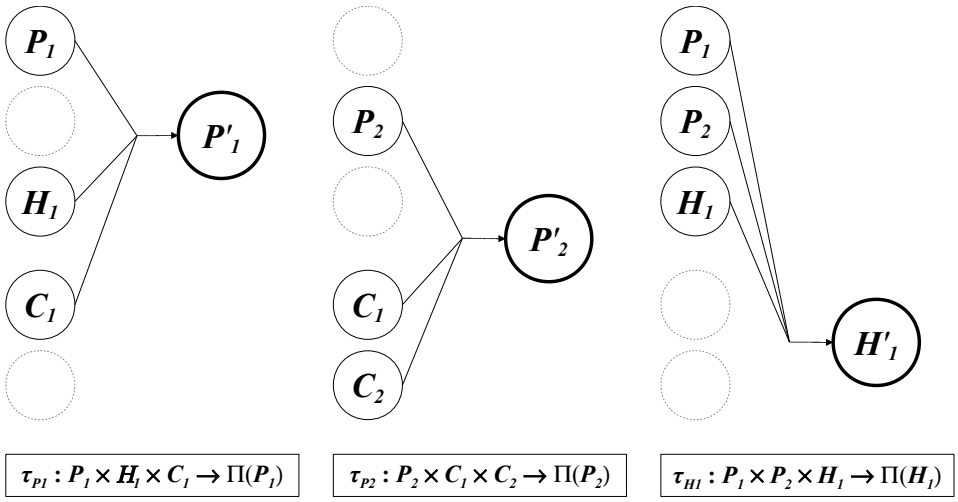


Figure 3.7: Exemple des DBNs avec des propriétés cachées.

3.1.6.3. Comparaison entre le FPOMDP et le POMDP

(339) Il faut remarquer que le degré d'accessibilité perceptive dans un FPOMDP est calculé par rapport au nombre de propriétés non-observables, et non par rapports au nombre d'états non-observables. Par exemple, dans un FPOMDP complètement observable décrit par des propriétés binaires, puisque que les états sont énumérés à partir des propriétés, l'insertion d'une seule variable non-observable double le nombre d'états dans le système, dans lequel la moitié sera non-observable dans le POMDP correspondent.

(340) Plus précisément, dans un univers constitué des propriétés binaires, si l'agent a l'accès à $|P|$ propriétés observables, mais qu'il existe encore $|H|$ propriétés pertinentes non-observables dans le système, alors il y aura $2^{|P|}$ états observables sur un total de $2^{|P|} \times 2^{|H|}$ états possibles. De cette façon, il y aura $2^{|H|}$ états sous-jacents possibles pour chaque état observé par l'agent, ce qui signifie que le nombre d'états non-observables s'accroit de façon exponentielle par rapport au nombre de propriétés non-observables.

3.1.7. Déterminisme

(341) Une méthode classique pour aborder la complexité d'un problème en intelligence artificielle est de réduire le nombre de variables considérées, mais en retour les traiter comme non-déterministes. Pour cette raison, MDPs, POMDPs et FMDPs décrivent la dynamique d'un environnement par le biais de distributions de probabilité. Toutefois, de nombreux problèmes peuvent être modélisés de façon déterministe, en utilisant respectivement les formalismes D-MDP, D-POMDPs, ou D-FMDPs.

(342) Un D-MDP est un cas particulier de MDP dans lequel seulement une transition possède tout la masse de probabilité, tandis que toutes les autres ont une probabilité nulle. C'est l'équivalent d'une machine d'états (MOORE, 1956), c'est-à-dire, un graphe orienté dont les flèches représentent des transitions déterministes d'un état à l'autre selon une action spécifique. Le problème de la découverte de la structure d'un D-MDP est équivalent au problème de l'identification incrémentale de la machine d'états par expérimentation, pour lequel la solution est intuitive. Chaque observation de l'agent peut être directement apprise comme une régularité du système, parce que tous les états sont visibles et les transitions sont déterministes, le seul problème restant étant de définir une stratégie d'exploration.

(343) Par ailleurs, le calcul de la politique optimale pour un D-MDP n'est pas une tâche triviale, qui a fait l'objet de nombres travaux et algorithmes. La politique optimale dans un D-MDP est le cycle contenu dans le graphe qui maximise la moyenne des récompenses (AHUJA et al., 1993). Le premier algorithme efficient à été suggéré par (KARP, 1978), dont la complexité informatique est de l'ordre $O(|M|\times|S|)$. D'autres algorithmes ont été proposés par (PAPADIMITRIOU; TSITSIKLIS; 1987), (YOUNG et al., 1991), (HARTMANN; ORLIN, 1993), (DASDAN et al., 1999), (MADANI, 2002), (ANDERSSON; VOROBYOV, 2006), (ORTNER, 2008), (MADANI et al., 2009), qui peuvent être plus rapides sous certaines conditions.

(344) En revanche, la découverte de la structure d'un D-POMDP est une tâche beaucoup plus difficile par rapport au D-MDP. Il s'agit d'un problème équivalent à l'induction active d'*automates finis déterministes* (DFAs), tel que présenté dans les travaux de (LANG, 1992), (LANG et al., 1998), (JUILLÉ; POLLACK, 1998), (PENA; OLIVEIRA, 1998), (LANG, 1999), (CICCHELLO; KREMER, 2003).

3.1.7.1. Déterminisme Partiel

(345) Dans ce travail, nous nous référons à des environnements partiellement déterministes, représentés à travers des *processus de décision markoviens factorisés, partiellement observables et partiellement déterministes* (PD-FPOMDP).

(346) Dans la définition d'un PD-FPOMDP, l'ensemble $\tau = \{\tau_1, \tau_2, \dots, \tau_{|X|}\}$ des fonctions de transformation est remplacé par un ensemble $\sigma = \{\sigma_1, \sigma_2, \dots, \sigma_{|X|}\}$ des fonctions de régularité. La *régularité* est une représentation partiellement déterministe de la transformation. Alors que la transformation est décrite comme $\tau_i : rel(\tau_i) \rightarrow \Pi(X_i)$, la régularité se présente en tant que $\sigma_i : rel(\tau_i) \rightarrow X_i \cup \{\#\}$.

(347) Plus précisément, dans les situations où une transformation τ_j donnée est déterministe, cas dans lequel $\exists\ x_i \in dom(X_i) \mid prob(\tau_j \rightarrow x_i) = 1$, alors la régularité indique directement la valeur x_i comme le résultat de la fonction. Sinon, quand la transformation est stochastique ou aléatoire, lorsque $\forall\ x_i \in dom(X_i) \mid prob(\tau_j \rightarrow x_i) < 1$, alors la fonction de régularité retourne le symbole spécial #, ce qui indique justement le non-déterminisme de la transformation dans la situation donnée.

(348) Le PD-FPOMDP est un modèle moins général que le FPOMDP car il ne représente pas la distribution de probabilités pour les transformations non-déterministes, mais il est un modèle plus général que le D-POMDP parce qu'il permet au moins d'indiquer leur existence. Autrement dit, un PD-FPOMDP représente la partie déterministe d'un FPOMDP.

3.1.7.2. Degré de Déterminisme d'un Environnement

(349) Un PD-FPOMDP peut être classé selon le degré de déterminisme (∂) de sa dynamique. Pour une propriété donnée X_i, le degré de déterminisme ∂_i de sa fonction de transformation τ_i est l'équivalent du nombre de cas déterministes qu'il présente sur le nombre total de cas, selon l'équation 3.5. On note par convention $|\sigma_i|$ le nombre de cas déterministes, pour lesquels $\sigma_i \neq \#$, dans un univers de $|\tau_i|$ cas de transformation, total donné par la combinaison croisée des propriétés pertinentes $rel(\tau_i)$.

$$\partial_i = \frac{|\sigma_i|}{|\tau_i|} \quad \textit{(eq. 3.5)}$$

(350) Le degré de déterminisme global ∂ de l'environnement est calculé par la moyenne du degré de déterminisme des fonctions de transformation de chaque propriété particulière, selon l'équation 3.6.

$$\partial = \frac{\sum_{i=0}^{|X|} \partial_i}{|X|} \quad (eq.\ 3.6)$$

(351) Lorsque $\partial = 0$, les fonctions de transformation τ_i de chaque propriété X_i produisent une distribution de probabilités telle qu'il n'y a aucun cas où l'évolution de x_i à x_i' puisse être mappée de façon directe et unique en fonction x de c. Dans ce cas, l'environnement est complètement non-déterministe, et les fonctions de régularité retournent constamment $\sigma_i = \#$. Inversement, lorsque $\partial = 1$, alors toutes les transformations peuvent être représentées sans perte en tant que fonctions de régularité, puisque x' a un mappage direct d'après x et c. Dans ce cas, l'environnement est totalement déterministe, l'équivalent d'un D-FPOMDP. Enfin, un environnement est partiellement déterministe s'il est situé entre ces deux extrêmes, c'est-à-dire quand $0 < \partial < 1$, comme l'illustre la figure 3.8.

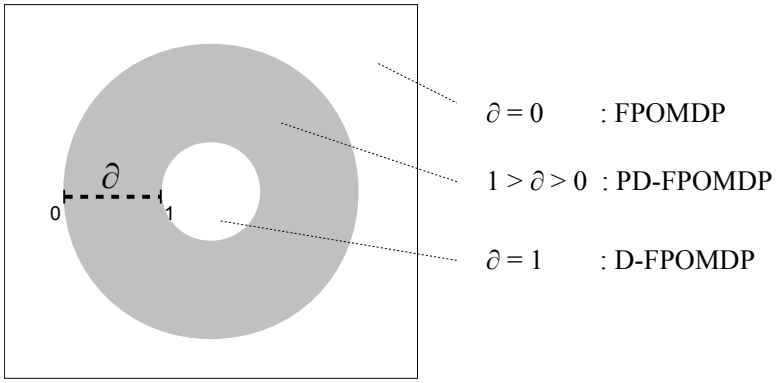
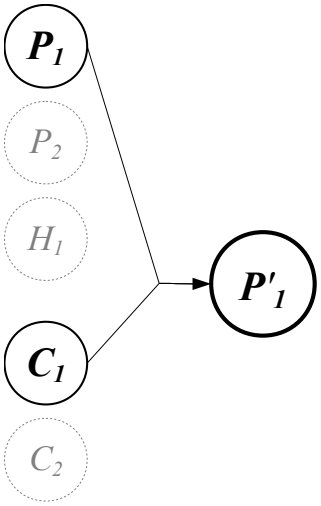


Figure 3.8: Relation de déterminisme partiel.

(352) Dans l'exemple illustré à la figure 3.9, il y a un DBN qui décrit la transformation de la perception P_1 en fonction de ses propriétés pertinentes P_1 et C_1. Dans l'exemple, on remarque les transformations déterministes. Donc on peut calculer le degré de déterminisme $\partial_1 = 0.5$ pour cette perception, qui résulte de la proportion de transformations déterministes, $|\sigma_1| = 2$, sur le nombre total de cas de transformation considérés, $|\tau_1| = 4$.

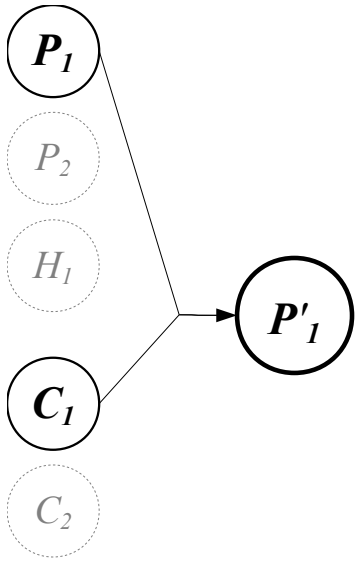


P_1	C_1	P'_1	$\sim P'_1$
0	0	0	1
	1	0,3	0,7
1	0	1	0
	1	0,5	0,5

$$\tau_1 : P_1 \times C_1 \rightarrow \Pi(P_1)$$

Figure 3.9: Exemple d'un DBN avec ses probabilités de transformation.

(353) Le degré de déterminisme de l'environnement a une influence directe sur la façon et sur la difficulté d'en construire un modèle. L'apprentissage des modèles du monde dont les transformations sont stochastiques (non-déterministes) constitue un problème d'induction de distributions, qui est un problème beaucoup plus difficile que dans le cas où les transformations sont déterministes. L'apprentissage dans le cas partiellement déterministe est plus facile que le cas stochastique, mais plus difficile que le cas complètement déterministe. Dans la figure 3.10 on montre la même transformation, représentée sous forme d'une régularité.



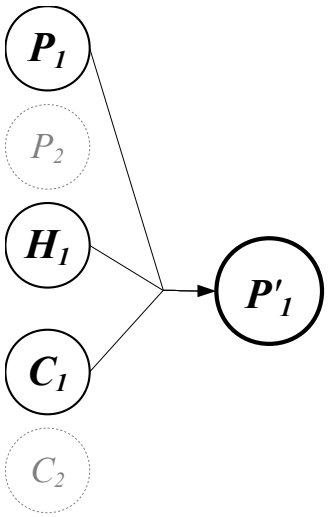
P_1	C_1	P'_1
0	0	0
	1	#
1	0	1
	1	#

$$\sigma_1 : P_1 \times C_1 \rightarrow P_1$$

Figure 3.10: Exemple de la transformation représentée en tant que régularité.

(354) D'un côté, pour l'apprentissage dans le cadre partiellement déterministe on n'a pas besoin d'induire les probabilités, et ainsi il est plus facile de couper des liens de causalité entre des variables. D'un autre côté, l'apprentissage partiellement déterministe

nécessite de trouver par soi même si une situation de transformation est déterministe ou si elle ne l'est pas, une difficulté qui ne se pose pas dans le cadre déterministe. Finalement, la force donnée à la recherche pour les causes des transformations peut compenser l'absence d'une analyse probabiliste. Dans la figure 3.11, en utilisant le même cas des figures précédentes, on montre comment la découverte d'une propriété non-observable, mais pertinente, peut rendre le modèle plus intéressant.



H_1	P_1	C_1	P'_1
0	0	0	0
		1	0
	1	0	1
		1	0
1	0	0	0
		1	#
	1	0	1
		1	#

$$\sigma_1 : P_1 \times H_1 \times C_1 \rightarrow P_1$$

Figure 3.11: Exemple de régularité, en considérant des propriétés cachées.

3.1.7.3. Comparaison entre le PD-FPOMDP et le PD-POMDP

(355) Il faut remarquer que la fonction de transition d'états (δ) dans un MDP (ou POMDP) équivaut à $|X|$ fonctions de transformation τ_i dans son FMDP (ou FPOMDP) correspondant. Ainsi, une transition peut être non-déterministe en δ mais présenter des composants déterministes en τ. De cette façon, il est possible qu'on trouve un sous-ensemble de cas de transformations déterministes dans un FMDP, dont la probabilité est égale à 1, même si le MDP correspondant est complètement non-déterministe, ne possédant aucune transition de probabilité 1. Un exemple est donné figure 3.12.

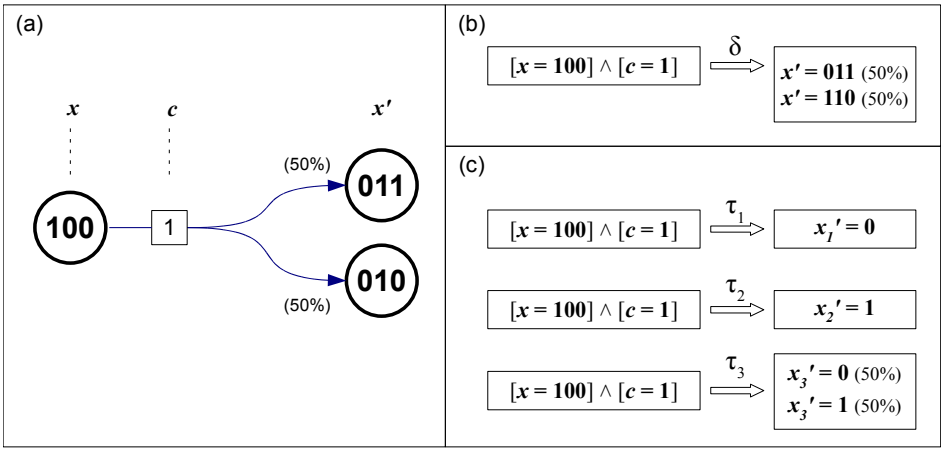


Figure 3.12: Exemple d'une situation ambigüe.

En (a) on voit une partie d'un MDP exemple, où une transition à partir de l'état [100] avec l'action [1] conduit de façon non-déterministe aux états [010] et [011]. En (b) on a la représentation de cette partie de la fonction de transition du MDP. En (c) on a la même transition, alors représentée de manière factorisée, où chaque propriété est déterminée par une fonction de transformation indépendante. On peut observer que, bien que la fonction de transition soit non-déterministe dans cette situation, deux des trois fonctions de transformation dérivées peuvent être représentées de façon déterministe.

3.1.8. Le Monde Réel comme un Environnement pour Apprendre

(356) Le monde réel est un environnement de complexité élevée, principalement en raison de l'énorme quantité d'éléments qui le composent. L'idée d'essayer de modéliser un problème du monde réel d'une façon plate est impraticable, car une telle représentation nécessiterait d'être dimensionnée par un nombre immense d'états.

(357) Cependant, bien que en général il s'agisse de problèmes de grande dimension, le monde réel dans a des caractéristiques qui facilitent le processus d'apprentissage par des agents situés. L'argument le plus fort dans la défense de cette affirmation est le fait que les êtres humains, immergés dans cet univers hautement complexe, arrivent, d'une manière ou d'une autre, à construire un modèle relativement adéquat pour décrire l'environnement immédiat, et pour anticiper les événements pertinents pour les tâches et les besoins quotidiens. En termes plus spécifiques, nous croyons que cela s'explique par les hypothèses suivantes: Un problème du monde réel présente (1) plusieurs événements déterministes ($\partial >> 0$); (2) est largement accessible par la perception ($\omega >> 0$); et (3) est bien structuré ($\varphi >> 0$).

(358) La première hypothèse est dérivée de la théorie du *déterminisme causal*, qui stipule que les événements futurs sont nécessairement et précisément déterminés par les

faits passés et présents, combinés avec les lois naturelles qui gouvernent l'univers (SUPPES, 1993). Selon le déterminisme causal, tous les phénomènes constitutifs de la réalité sont soumis à un système de causes et d'effets nécessaires. Il s'agit d'un sujet de grande importance dans la philosophie, et qui suscite beaucoup de controverses. Quelques références sur le thème sont (DOOB, 1988), (BOBZIEN, 1998) et (BUNGE, 1959).

(359) Le déterminisme causal a un rapport direct avec la prévisibilité du monde. La possibilité d'anticiper les phénomènes est conditionnée au fait qu'ils soient causalement déterminés. Ainsi, en principe, si un sujet pouvait connaître l'état de tous les éléments qui constituent l'univers, ainsi que toutes les lois générales de la nature, alors il serait capable de déduire avec une précision absolue la séquence des événements du présent au futur.

(360) Dans le monde réel, l'incapacité des sujets de prédire de façon exacte les événements n'est pas due au manque de déterminisme du monde, mais à l'impossibilité d'avoir toutes les informations sur son état. Le sujet est limité par son point de vue local, par l'excès d'information, par la complexité du monde, et par ses capacités de connaissance et de compréhension.

(361) Dans le cas d'un agent $д$ inséré dans un environnement ξ, moins l'agent a d'informations sensorielles, plus il va observer dans l'environnement de phénomènes apparemment non-déterministes. Toutefois, s'il est capable de considérer l'existence de ces éléments non-observables, le monde représenté dans sa connaissance sera de plus en plus déterministe.

(362) Dans un environnement partiellement observable il y a des propriétés cruciales pour la modélisation de la dynamique des événements qui ne sont pas directement perçues par l'agent. Les environnements partiellement observables peuvent présenter une dynamique apparemment arbitraire et non-déterministe en surface, même si elle est, en fait, déterministe par rapport au système sous-jacent et partiellement caché qui donne lieu à la face perceptive des phénomènes (HOLMES; ISBELL, 2006).

(363) Même si l'hypothèse du déterminisme causal n'est pas vérifiable jusqu'à ses conséquences ultimes, nous croyons à l'affirmation que, en ce qui concerne le degré de déterminisme des transformations du système (∂), même en acceptant l'existence

d'événements chaotiques et probabilistes, le monde réel est un environnement partiellement mais hautement déterministe, où la plupart des événements peuvent être représentées de façon déterministe, étant donné qu'ils sont bien contextualisés en fonction de leurs causes, que celles-ci soient directement observables par l'agent ou non.

(364) Le monde réel est aussi un environnement riche en caractéristiques, et un agent situé, en général, selon notre deuxième hypothèse, a une quantité raisonnable de senseurs. C'est-à-dire que, dans les problèmes du monde réel, l'accessibilité perceptive (ω) aux propriétés du monde peut être relativement élevée, et les situations peuvent être identifiées, pour la plupart, en s'appuyant sur les éléments observables. En d'autres termes, la première difficulté (même si elle est aussi importante) n'est pas le manque d'information, mais son excès.

(365) En revanche, d'après notre troisième hypothèse, le monde réel est un environnement dont les phénomènes ont des causes spécifiques et bien définies , donc le degré de structuration (φ) est élevé, et alors la quantité de facteurs pertinents pour les transformations reste limitée à un petit ensemble de variables qui peuvent être apprises par un agent.

(366) Dans cette perspective, deux défis se posent à un agent situé qui cherche à construire un modèle dans des problèmes du monde réel. Premièrement, la nécessité de trouver des représentations généralisées pour les situations, étant donné qu'il y a une très grande quantité d'informations sensorielles, ce qui exige la capacité d'identifier efficacement les caractéristiques pertinentes impliquées dans les événements observés. En même temps, l'agent doit faire face à des ambiguïtés perceptives, c'est-à-dire, des situations qui ne peuvent pas être identifiées seulement par la perception sensorielle immédiate. Dans ce cas, il est nécessaire d'induire et d'utiliser d'autres informations provenant de l'environnement, qui ne sont pas directement observables, afin de construire un modèle du monde cohérent.

(367) En analyse ultime, et conformément aux observations de la psychologie expérimentale de Piaget (1937), un agent (humain ou artificiel), inséré dans le monde réel, a la tâche initiale d'organiser ses perceptions et ses sensations dans un modèle qui puisse rendre l'environnement intelligible. Si d'un côté beaucoup de régularités se cantonnent au niveau sensorimoteur, par ailleurs il restera toujours certains phénomènes

qui nécessiteront une nouvelle couche de concepts abstraits pour devenir compréhensibles. En général, une propriété abstraite est une variable non-observable induite par l'agent, et l'induction de ce genre de propriété représente donc, à notre avis, le début de la pensée symbolique et abstraite.

(368) Le mécanisme d'apprentissage proposé dans cette thèse, décrit dans la section suivante, tire parti des caractéristiques que nous jugeons être courantes par rapport au monde réel. Tout d'abord, il traite l'univers en le décomposant par propriétés au lieu de le représenter en tant qu'un ensemble plat (combinatoire) d'états, puis en essayant de construire un modèle du monde partiellement déterministe, basé d'une part sur l'information apportée par la perception, et d'autre part sur la recherche des éléments non-observables pertinents.

3.2. Le Mécanisme d'Apprentissage CALM

(369) Depuis cette section, nous décrivons le mécanisme CALM (*Constructivist Anticipatory Learning Mechanism*), qui est conçu pour jouer le rôle du système cognitif dans l'esprit d'un agent défini selon l'architecture CAES (présentée dans le chapitre 2). Le mécanisme CALM implémente une méthode d'apprentissage basée sur l'approche constructiviste de l'IA, afin de doter l'agent de la capacité d'inférer un modèle du monde qui représente les régularités observées lors de ses interactions avec l'environnement, et de lui permettre d'utiliser ce modèle pour améliorer ses comportements.

3.2.1. Idée Générale du Mécanisme

(370) Le mécanisme CALM exécute les tâches (a) *d'apprentissage d'un modèle du monde* et (b) *de construction d'une politique d'actions*, de façon progressive, et à partir de l'expérience. Le modèle du monde est représenté par un ensemble d'*arbres d'anticipation*, et la politique par un ensemble d'*arbres de délibération*, comme ce sera expliqué plus loin. Ces arbres ont pour rôle de partitionner l'espace de situations de façon optimisée, et ils sont construits par le biais de différenciations progressives.

(371) Comme cela a été montré dans la section 3.1, apprendre un modèle du monde signifie découvrir la structure d'un *processus de décision markovien partiellement observable et partiellement déterministe* (PD-FPOMDP). Il s'agit d'un formalisme

générique pour décrire les régularités du système du point de vue de l'agent. De même, la construction d'une politique d'actions signifie définir un ensemble de règles de décision qui doivent être prises par l'agent pour tout état possible du PD-FPOMDP, ce qui établit un patron de comportement visant à maximiser la moyenne du signal d'évaluation (les récompenses reçues) dans une fenêtre de temps à long terme.

(372) Le mécanisme CALM est fondé sur trois concepts issus de la psychologie constructiviste (PIAGET, 1936, 1937): le « schéma », l'« assimilation » et l'« accommodation ». Le *schéma* est la structure élémentaire, qui représente une unité primaire de connaissance. *L'assimilation* est le processus qui incorpore une nouvelle situation à la connaissance déjà acquise par l'agent, c'est-à-dire que l'assimilation est la force conservatrice qui soumet les nouvelles expériences aux schémas existants. Pour compléter ce dispositif, *l'accommodation* est le processus qui transforme la connaissance en vertu d'une situation de déséquilibre. Quand une expérience donnée n'est pas assimilable de la même façon que les expériences précédentes, alors l'accommodation change les schémas à fin que le nouvel ensemble puisse faire face à la nouvelle.

3.2.1.1. Caractéristiques du Mécanisme

(373) Le Mécanisme CALM a été conçu pour apprendre les régularités d'un environnement, pris en tant que système discret et partiellement déterministe. Cela signifie que l'agent doit identifier, parmi l'ensemble total des transformations du FPOMDP, le sous-ensemble des transformations déterministes (σ), ce qui constitue alors un PD-FPOMDP. Ainsi, d'une part le mécanisme ne nécessite pas que le système soit complètement déterministe, d'autre part il ignore les régularités non-déterministes, en s'exonérant de toute tentative de décrire une distribution de probabilité. Par conséquent, la solution construite par CALM est plus intéressante lorsque l'environnement présente de nombreux événements déterministes ($0 << \partial < 1$).

(374) L'efficacité du mécanisme CALM est directement liée au degré de structuration du système. Plus le nombre moyen de variables nécessaires pour décrire les régularités du PD-FPOMDP est petit, plus l'univers est structuré ($\varphi >> 0$), par conséquent, plus le modèle du monde construit sera compact, et plus la convergence de l'algorithme d'apprentissage sera rapide. Cela implique, cependant, la tâche supplémentaire de

trouver l'ensemble des propriétés pertinentes, *rel*(σ_i), pour décrire les transformations, dans l'ensemble total des propriétés du système.

(375) Enfin, le mécanisme CALM permet à un agent artificiel d'apprendre un modèle du monde, même si l'environnement n'est que partiellement observable. Dans ce cas, les propriétés pertinentes pour décrire la dynamique d'interaction entre l'agent et le monde ne sont pas toutes directement accessibles via les capteurs ($0 < \omega < 1$). Cela entraine la tâche supplémentaire de créer un ensemble d'éléments de synthèse (*H*) qui puissent représenter des propriétés non-observables de l'environnement dans son modèle du monde.

3.2.1.2. Observabilité Partielle x Déterminisme Partiel

(376) À première vue, le fait d'échapper à la tentative de calculer les distributions de probabilité des transformations peut donner l'impression que CALM renonce à trouver une partie importante des relations de causalité du monde. En général, dans l'IA, on suppose que représenter les transitions de façon non-déterministe est la meilleure option pour traiter des environnements bruyants et aussi l'absence d'observabilité complète.

(377) Toutefois, la stratégie adoptée par le mécanisme CALM est différente. L'idée est de simplifier la représentation des régularités, et aussi le processus pour les découvrir. Savoir quand une transformation n'est pas déterministe est facile, parce qu'il faut tout simplement que deux épisodes similaires conduisent à des résultats différents. En parallèle, CALM est capable de découvrir l'existence de propriétés non-observables, qui permettent une meilleure contextualisation des transformations, jusqu'au point où elles peuvent être représentées d'une manière déterministe.

(378) Dans un modèle non-déterministe classique, l'observation d'un événement qui infirme une prédiction conduit à une révision des probabilités. En revanche, dans la stratégie du CALM, cette discordance conduit à la nécessité de spécialiser encore plus le contexte de la situation.

(379) Par exemple, imaginons un robot qui, jour après jour, appuie sur un bouton et remarque qu'à chaque fois, à ce moment, une lampe s'allume. Cette régularité finira par être saisie, et sera représentée dans le modèle du monde du robot. Supposons qu'un jour, en raison d'une panne d'électricité, la lampe ne s'allume pas lorsque le robot appuie sur le bouton. Si on accepte trop vite le non-déterminisme comme explication, cet

événement ne provoque pas de changement dans la structure de la connaissance. Il sera traité comme une nuisance ou comme un événement rare, ce qui conduira à une simple correction des probabilités. Dans CALM, l'attitude est différente: l'événement qui à causé le déséquilibre va déclencher la recherche d'une explication, qui est, dans ce cas, une nouvelle condition qui peut différencier la situation déséquilibrée.

(380) Ainsi, la stratégie est de toujours chercher à spécialiser suffisamment les conditions (contexte et action) d'une situation afin que la transformation décrite par le régime reste déterministe. Ces conditions peuvent être liées à des propriétés non-observables de l'environnement, représentées par CALM sous la forme d'éléments abstraits. Dans l'exemple de la lampe, un robot, doué du mécanisme CALM, serait amené à supposer l'existence d'une condition qui n'est pas directement observable (dans ce cas, la fourniture d'électricité), et l'inclusion de cette condition dans la représentation de la transformation permet qu'elle reste déterministe. Précisément, quand il y a l'électricité, et que le robot appuie sur le bouton, alors la lampe s'allume. Il peut que la condition ne puisse pas être vérifiée de façon déterministe, mais de toute façon la transformation sera effectivement déterministe.

3.2.1.3. Représentation de la Connaissance et Fonctionnement du Mécanisme

(381) Dans le mécanisme CALM on trouve 3 grandes structures de représentation de la connaissance: (1) un ensemble Ш de *mémoires épisodiques généralisées*; (2) un ensemble Ψ d'*arbres d'anticipation*; et (3) un ensemble Ж d'*arbres de délibération*.

(382) Pour chaque variable x_i' dont CALM décide de construire un modèle anticipatoire il existe une mémoire épisodique généralisée $Ш_i$ et un arbre d'anticipation $Ψ_i$. La mémoire épisodique est une sorte de souvenir des situations réelles vécues par l'agent. Elle est essentielle au processus d'apprentissage parce que l'arbre d'anticipation est créé et modifié à partir de ses informations. L'arbre d'anticipation est la structure qui sert à prévoir les transformations de la variable associée. Il essaye de partitionner l'espace des situations de façon optimisée, en utilisant seulement les variables apparemment pertinentes. Ainsi, chaque arbre $Ψ_i$ envisage de représenter d'une manière compacte la fonction de régularité $σ_i$ qui décrit la dynamique de la propriété x_i. Les arbres de délibération sont construits à partir des arbres d'anticipation, et ils constituent la politique d'actions de l'agent.

(383) Pour l'introduire de façon brève, le cycle de fonctionnement de l'algorithme a 7 étapes, répétées à l'infini: (1) recevoir le signal perceptif, provenant de l'extérieur; (2) actualiser les variables internes, qui représentent le contexte abstrait; (3) actualiser la mémoire épisodique généralisée; (4) dans le cas où cette situation crée un déséquilibre, réaliser le processus d'apprentissage afin de corriger le modèle du monde; (5) actualiser la politique d'actions; (6) étant donnée la situation courante et les modèles actuels, prendre une décision; et (7) envoyer le signal de contrôle, afin d'effectuer les actions choisies. L'algorithme 3.1 précise ce cycle de fonctionnement de base de CALM.

CALM – Méthode PRINCIPALE

```
SOIENT:
  p   un vecteur représentant le signal de perception
  c   un vecteur représentant le signal de contrôle
  h   un vecteur représentant les éléments synthétiques
  p'  un vecteur représentant la perception suivante
  Ш   la structure qui représente la mémoire épisodique généralisée
  Ψ   la structure qui représente le modèle du monde
  Ж   la structure qui représente la politique d'actions

COMMENCEMENT:
  Initialiser_Structures;          //Initialisation structures de connaissance
  RÉPÉTER(indéfiniment):
    p ← Recevoir_Perception;       //Réception du signal perceptif venant des senseurs
    h ← Déduire_Abstrait;          //Déduction des propriétés non-observables
    Ш ← Actualiser_Mémoire;        //Actualisation de la mémoire épisodique
    Ψ ← Actualiser_Modèle;         //Apprentissage du modèle du monde
    Ж ← Actualiser_Politique;      //Apprentissage de la politique d'actions
    c ← Prendre_Décision;          //Prise de décision pour la situation actuelle
    Envoyer_Contrôle(c);           //Envoi du signal de contrôle aux actuateurs
FIN;
```

Algorithme 3.1: Méthode principale de CALM, qui décrit le cycle de base du mécanisme.

3.2.2. Mémoire Épisodique Généralisée

(384) L'apprentissage dans CALM est fait de façon incrémentale, car il ne lui est pas donné un ensemble *hors-ligne* de situations pré-classées dans une base de données. L'agent apprend en interagissant directement avec l'environnement. Toutefois, pour qu'il puisse construire ses arbres d'anticipation, en différenciant efficacement les situations, il doit y avoir une forme de mémoire, au-delà du modèle anticipatoire lui-même et au-delà de la simple observation instantanée, qui puisse servir de référence pour CALM à la découverte des éléments possiblement pertinents pour expliquer une situation génératrice de déséquilibre. L'absence complète de cette information ferait du processus d'apprentissage une poursuite aveugle dans l'espace de différenciations. Il est donc

nécessaire de garder, sous une forme ou sous une autre, le souvenir des épisodes passés, et c'est le rôle de la mémoire épisodique généralisée (Ш).

(385) La mémoire épisodique généralisée est un ensemble de sous-mémoires $Ш = \{Ш_1, Ш_2, \dots, Ш_{|X|}\}$, une par chaque transformation qu'on désire décrire la dynamique, et chacune des sous-mémoires est, à son tour, un ensemble des tables $Ш_i = \{Ш_{i0}, Ш_{i1}, \dots, Ш_{|i\alpha|}\}$. Chacune de ces tables $Ш_{ij}$ garde le souvenir des observations réalisées d'une propriété spécifique *i* en considérant une dépendance conditionnelle à *j* propriétés. Ainsi, dans une table donnée $Ш_{ij}$ il y aura une entrée pour chacune des situations possibles qu'elle peut identifier. Chaque colonne représente une combinaison de variables, groupées *j* par *j*, qui sont les combinaisons candidates pour les propriétés pertinentes. Chaque colonne possède un nombre de lignes équivalant à la combinaison des valeurs de ses *j* variables.

(386) Les entrées (cellules) de cette table prennent leurs valeurs selon l'observation de la propriété x_i. La table est complètement initialisée avec la valeur spéciale « ? », qui signale l'absence d'observation. La fonction de *mémorisation* marque les cas avec des valeurs existantes dans $dom(X_i)$ selon ses expériences (par exemple « 0 » ou « 1 » s'il s'agit de l'anticipation d'une propriété binaire), et finalement avec la valeur spéciale « # » pour les cas ou il n'y a pas d'accord entre les observations.

(387) Quand il s'agit de problèmes complexes, il n'est pas viable d'implémenter une mémoire « exhaustive et photographique ». Autrement dit, la mémoire ne peut pas stocker la liste complète de toutes les situations passées, ni se souvenir des événements dans tous leurs détails. Pour cette raison il faut établir un degré maximal de précision, qui va déterminer la force de généralisation minimale à être appliquée aux épisodes.

(388) Le paramètre $\alpha_Ш$ définit le nombre maximal de dépendances conditionnelles observées simultanément par la mémoire. Lorsque $\alpha_Ш = 0$, la généralisation est extrême, et la mémoire ne peut considérer aucune dépendance causale entre la propriété anticipée, et les propriétés qui contextualisent la situation. Au niveau $\alpha_Ш = 1$, la mémoire peut considérer une dépendance à conditionner l'anticipation (cas connu comme « naïf »). Au niveau $\alpha_Ш = 2$, alors une interdépendance conditionnelle est considérée, donc la mémoire change en prenant en compte les combinaisons des deux de variables, et ainsi progressivement, en augmentant la valeur de $\alpha_Ш$.

(389) La gestion de la mémoire épisodique généralisée est une clé pour assurer la convergence de l'algorithme d'apprentissage, ainsi que pour préserver son extensibilité par rapport au temps et à l'espace.

(390) La fonction de *mémorisation* a le rôle d'inclure une nouvelle situation expérimentée dans la représentation de l'ensemble des situations connues. L'algorithme 3.2 décrit la méthode de mise à jour de la mémoire épisodique, et la figure 3.13 illustre un exemple de situations, et de comment elles sont stockées dans la mémoire.

Méthode CALM – ACTUALISE MÉMOIRE

```
SOIENT:
  Wij  une table de la mémoire épisodique généralisée
  p    le dernier contexte perceptif
  c    le dernier contrôle
  h    le dernier contexte abstrait
  p'   la perception suivante

COMMENCEMENT:

  POUR chaque table Wij de la mémoire FAIRE:

    POUR chaque combinaison k de j propriétés de {p,c,h} FAIRE:
      SI Wij(k) = "?" ALORS
        Wij(k) ← pi'
      SINON
        SI Wij(k) ≠ pi' ALORS
          Wij(k) ← "#";

FIN;
```

Algorithme 3.2: Méthode d'actualisation de la mémoire épisodique généralisée.

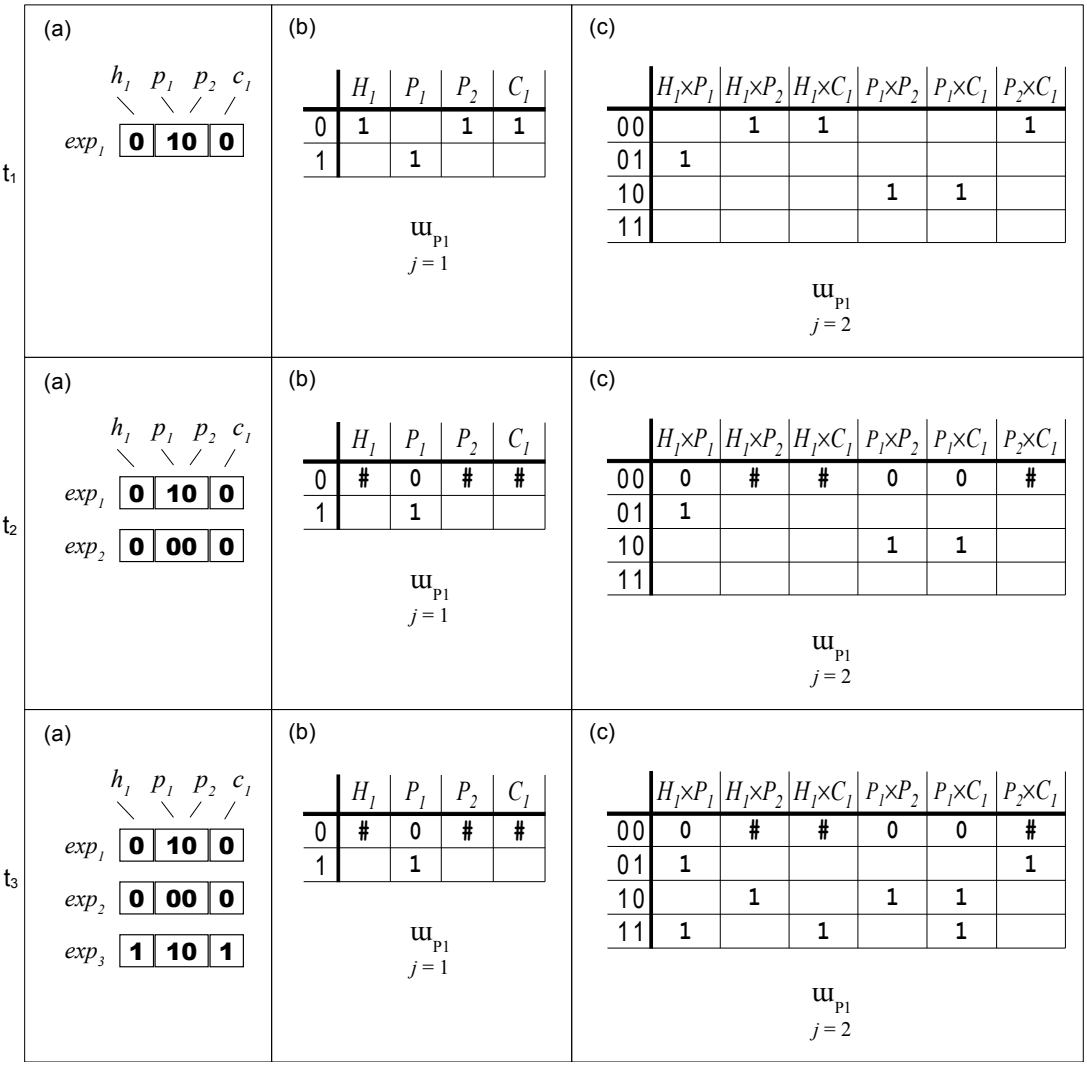


Figure 3.13: Exemple de la formation de la mémoire épisodique généralisée.

En (*a*) ce sont les situations expérimentées, en (b) la mémoire de niveau $j = 1$, où une seule condition est considérée comme possible, et en (c) la mémoire de niveau $j = 2$, où deux dépendances sont considérées.

3.2.3. Sélection des Propriétés Pertinentes

(391) Dans des environnements complexes, le nombre de caractéristiques que l'agent perçoit peut être très grand, mais en général, si l'environnement est bien structuré, seul un petit nombre de ces propriétés sont importantes pour caractériser une situation. Plus précisément, un environnement bien structuré est tel que le nombre moyen de propriétés pertinentes pour la description d'une transformation est d'ordre logarithmique par rapport au nombre total de propriétés, donc $|rel(\tau)| \approx log_b(|P|+|H|+|C|)$, tel que cela a été défini dans la section 3.1.

(392) Le problème de la découverte des propriétés pertinentes est difficile, et la réussite de cette tâche dépend de la faisabilité du calcul de ce type d'algorithme d'apprentissage (BLUM; LANGLEY, 1997), (MURPHY; McCRAW, 1991). Il convient de noter que, dans le cas de l'apprentissage incrémental, l'ensemble des propriétés qui semblent pertinentes à un instant donné, fondé sur les expériences passées, peut changer au fil du temps, lorsque l'agent expérimente de nouvelles situations. Cela signifie qu'une propriété calculée comme pertinente au moment t, peut cesser de l'être à $t+n$, ou l'inverse.

(393) Le mécanisme CALM déduit les propriétés qui sont pertinentes pour décrire la transformation d'une propriété donnée en analysant l'état de la mémoire épisodique généralisée. On dit qu'une propriété est pertinente lorsque la description correcte de la transformation dépend, de façon irremplaçable, de la valeur de cette propriété. Donc, pour décrire la transformation de la propriété X_i, le mécanisme utilise la liste de différenciateurs Λ associée à Ψ_i, qui sont en principe les propriétés pertinentes extraites de la mémoire épisodique.

(394) La sélection des propriétés pertinentes marche comme une recherche heuristique dans un espace défini par les combinaisons possibles des propriétés considérées. Cet espace est partiellement ordonné, et on peut naviguer à travers lui en ajoutant ou en supprimant une variable de la liste des différenciateurs. La figure 3.14, similaire à celle présentée dans (BLUM; LANGLEY, 1997), illustre l'espace de recherche de pertinence et son rapport avec la mémoire épisodique.

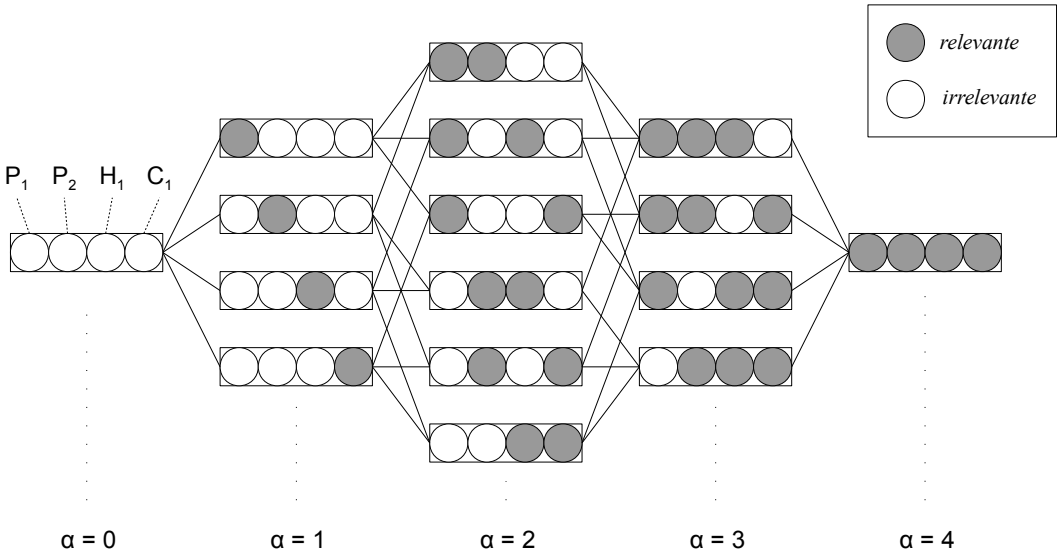


Figure 3.14: Exemple de l'espace de recherche de la pertinence.

Un problème formé par 4 propriétés, dont deux sont le contexte perceptif (P_1, P_2), une est le contexte abstrait (H_1), et la dernière est une action (C_1). L'espace est formé par les combinaisons des propriétés. Dans le dessin, les points sombres représentent l'inclusion de la propriété dans la liste de pertinence.

(395) Pour trouver l'ensemble des variables pertinentes pour décrire la transformation d'une propriété X_i, le mécanisme CALM calcule le degré de déterminisme pour chaque colonne de chaque table de la mémoire épisodique. Ce degré est mis à jour chaque fois qu'une valeur de la colonne est modifiée. Enfin, la colonne qui a le degré le plus élevé de déterminisme, dans l'une des tables liées à la propriété en question, indique la liste des variables pertinentes (Λ_i).

(396) Quand il y a égalité entre les ensembles candidats par rapport au degré de déterminisme, le mécanisme fait un choix basé sur les heuristiques suivantes (par ordre de priorité): (1) choisir l'ensemble qui a le moins de variables; (2) choisir l'ensemble qui inclut l'élément qui est anticipé; (3) choisir l'ensemble qui fournit le moins de variables abstraites; (4) choisir l'ensemble qui fournit le plus de variables de contrôle; (5) choisir l'ensemble qui a le plus d'éléments en commun avec les listes de pertinence des autres propriétés; (6) sinon, le choix est fait au hasard.

(397) La première heuristique s'explique par le désir de créer une description compacte de la transformation, associé au fait que si deux ensembles ont le même degré de déterminisme, il est probable que le plus gros d'entre eux contient des propriétés qui ne sont pas pertinentes. La seconde heuristique repose sur l'hypothèse que la transformation d'une propriété dépend probablement de son propre état. La troisième et

la quatrième heuristiques incitent l'algorithme à préférer, en cas d'égalité, les explications liées aux éléments perceptifs et aux actions de l'agent. Le cinquième heuristique favorise une tendance à utiliser des causes similaires pour expliquer des différentes transformations, ce qui conduit à la construction d'un modèle final plus compact. Il faut noter que la liste de différenciateurs n'aura jamais une taille supérieure au nombre maximum de dépendances conditionnelles établi pour la mémoire épisodique généralisée α. La méthode est définie dans l'algorithme 3.1.

CALM – Méthode SÉLECTIONNER DIFFÉRENCIATEUR:

```
SOIT
  Λk une nouvelle séquence possible de différenciateurs;

COMMENCEMENT;
  Pour chaque table j dans la mémoire épisodique, de 0 à α, faire:
    Recalculer le degré de déterminisme des colonnes modifiées
  Sélectionner la colonne Λk qui a le meilleur degré;
  Si a trouvé plus d'une séquence Λk candidate, alors:
    Choisir Λk selon heuristiques;
FIN;
```

Algorithme 3.3: Méthode pour la sélection des tests de différenciation (différenciateurs).

3.2.4. Arbre d'Anticipation

(398) En ce qui concerne les problèmes de construction de modèles du monde et les problèmes de décision, des travaux récents basés sur les *Processus de Décision Markoviens Factorisés* (FMDPs) représentent le système en utilisant des Réseaux *Bayésiens Dynamiques* (DBNs), (DEAN; KANAZAWA, 1989), qui à leur tour peuvent représenter les fonctions de transformation par le biais des arbres (BOUTILIER et al., 2000), (GUESTRIN et al., 2003).

(399) Dans le mécanisme CALM, le modèle du monde se compose d'un ensemble *d'arbres d'anticipation*. Un arbre d'anticipation Ψ_i partitionne de façon généralisée, exacte et complète, l'espace des situations qui conditionnent la transformation d'une variable X_i donnée. Il est composé d'un ensemble de *nœuds intermédiaires*, $\Theta = \{\Theta_1, \Theta_2, \ldots, \Theta_{|\Theta|}\}$, distribués dans ses niveaux supérieurs, et d'un ensemble de *schémas*, $\Xi = \{\Xi_1, \Xi_2, \ldots, \Xi_{|\Xi|}\}$, qui sont les nœuds terminaux. La topologie de l'arbre est donnée selon un ensemble de *différenciateurs*, $\Lambda = \{\Lambda_1, \Lambda_2, \ldots, \Lambda_{|\Lambda|}\}$.

(400) CALM, en tant que mécanisme cognitif de l'esprit, est chargé de la construction et du maintien d'un arbre d'anticipation pour chaque propriété importante du système. Il

est attendu qu'après une période suffisante d'expérimentation l'agent arrive à définir l'ensemble de tous ces arbres, $\Psi = \{\Psi_1, \Psi_2, \ldots, \Psi_{|\Psi|}\}$, en tant que description correcte des régularités du système, $\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_{|X|}\}$, qui à leur tour décrivent la dynamique des propriétés $X = \{P_1, P_2, \ldots, P_{|P|}\} \cup \{H_1, H_2, \ldots, H_{|H|}\}$. Ainsi, chaque arbre Ψ_i représente une fonction de régularité σ_i qui décrit la dynamique d'une propriété X_i donnée.

(401) Chaque nœud de l'arbre est identifié par la concaténation de ses vecteurs, sous la forme $\Theta_{[p][h][c]}$, ou $\Xi_{[p][h][c]}$ pour les schémas. Le nœud racine $\Theta_{[*][*][*]}$ représente la situation la plus générale. De la racine vers les feuilles, les nœuds représentent des situations de plus en plus spécialisées. Chaque niveau de l'arbre est plus spécialisé que le niveau immédiatement précédent par l'inclusion d'un *élément différenciateur,* λ, qu'on peut appeler aussi descripteur. L'arbre d'anticipation garde une liste ordonnée de ces éléments, $\Lambda \subseteq (P \cup H \cup C)$, dont le nombre d'éléments indique le nombre de niveaux dans l'arbre. Chaque différenciateur ajouté à l'arbre (et donc chaque niveau) permet la spécialisation d'un élément dans l'identificateur des schémas, en rendant possible que les valeurs indéfinies « * » soient remplacées par les différentes valeurs possibles du domaine de chaque propriété différenciatrice.

(402) Les nœuds terminaux de l'arbre sont les schémas. Chacun est identifié par une situation suffisamment spécialisée, par le biais de la restriction du contexte auquel le schéma s'applique, et par la définition minimale des actions à prendre. Un schéma est donc un nœud spécial de l'arbre qui anticipe une certaine transformation dans une situation précis de contexte et d'actuation. La topologie d'un arbre d'anticipation (Ψ_i) est illustrée figure 3.15, et sa définition formelle (3.5) est donnée ensuite.

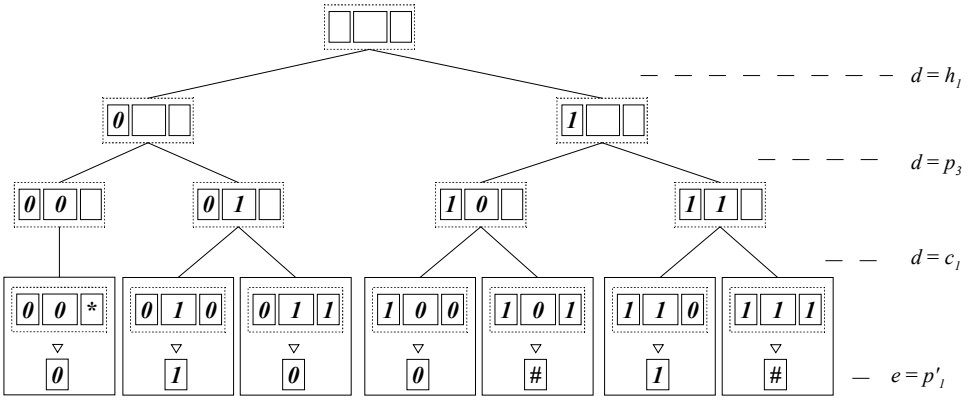


Figure 3.15: Exemple d'un arbre d'anticipation.

Un arbre d'anticipation (Ψ_i) est topologiquement défini selon une liste ordonnée de différenciateurs (Λ). La racine de l'arbre représente la situation complètement généralisée, celle qui mappe entièrement l'espace de situations. Ensuite on a les nœuds intermédiaires, et les nœuds terminaux de l'arbre sont les schémas.

Une arbre d'anticipation (Ψ_i) est un quadruplet:

$\Psi_i = \{\Xi, \Lambda, \Theta, e\}$

où

$\Lambda = \{\Lambda_1, \Lambda_2, \ldots, \Lambda_{|\Lambda|}\}$ est une liste ordonnée de différenciateurs

$\Theta = \{\Theta_1, \Theta_2, \ldots, \Theta_{|\Theta|}\}$ est un ensemble de nœuds intermédiaires

$\Xi = \{\Xi_1, \Xi_2, \ldots, \Xi_{|\Xi|}\}$ est un ensemble de schémas (feuilles de l'arbre)

e est la propriété à être anticipée x_i', où $X_i \in (P \cup H)$

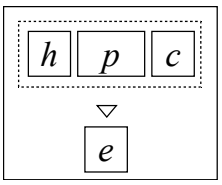
Définition 3.5: Arbre d'anticipation.

(403) Dans CALM, les arbres d'anticipation utilisent une représentation factorisée des états et des actions à travers un ensemble de variables définies par les signaux vectoriels de la perception (P), du contrôle (C), et par les éléments synthétiques (H). Cependant, chacun de ces arbres est construit en fonction de la prévision d'un seul élément de la perception (P_i) ou d'un seul élément synthétique (H_i). Par conséquent, le mécanisme CALM garde, au maximum, $|P| + |H|$ arbres d'anticipation, mais ce chiffre est, dans la pratique, optimisé, de façon que CALM maintienne seulement les arbres qui anticipent des propriétés importantes pour l'agent.

3.2.5. Schéma

(404) Le schéma est la structure anticipatoire élémentaire, qui représente une régularité observée par l'agent lors de son interaction avec le monde. Chaque schéma Ξ fait une déclaration du type $(p \wedge h \wedge c) \rightarrow e$, où p est le contexte perceptif minimal du schéma, h

est le contexte abstrait minimal, *c* est l'action de contrôle minimal, et *e* est l'anticipation de transformation régulière pour la propriété x_i'. Les domaines *P* et *C* sont définis par l'architecture de l'agent, en indiquant la façon dont les senseurs et effecteurs sont connectés à l'esprit. *H* est un ensemble de propriétés abstraites définies par le système cognitif lui-même. Chaque schéma fait partie d'un arbre d'anticipation selon la propriété qu'il doit prévoir, donc, les schémas qui anticipent la valeur de la propriété x_i' appartiennent à l'arbre Ψ_i. Un schéma est illustré figure 3.16, et ensuite formalisé dans la définition 3.6.



h = contexte abstrait
p = contexte perceptif
c = contrôle
e = anticipation (x'_i)

Figure 3.16: Représentation d'un schéma.

Un schéma (Ξ) est un sextuplet:

$\Xi = \{p, h, c, e, \upsilon, \rho\}$

où

$p = \{p_1, p_2, \ldots, p_{|\Lambda p|}\}$ est le contexte perceptif du schéma
tel que p_i est un élément lié au signal de perception de l'agent (*P*)

$h = \{h_1, h_2, \ldots, h_{|\Lambda h|}\}$ est le contexte abstrait du schéma
tel que h_i est un élément synthétique (*H*)

$c = \{c_1, c_2, \ldots, c_{|\Lambda c|}\}$ est l'action proposée par le schéma
tel que c_i est un élément lié au signal de contrôle (*C*)

e est l'anticipation d'une propriété x_i contenue dans $(P \cup H)$

υ est la valeur affective associée au schéma

ρ est la fiabilité associée au schéma

Définition 3.6: Schéma.

(405) Comme CALM envisage de découvrir uniquement les transformations déterministes dans un environnement partiellement déterministe, alors l'anticipation représente une valeur absolue, et non une distribution de probabilités. C'est-à-dire que chaque schéma doit correspondre à l'une des trois possibilités: (1) soit il présente une anticipation inexistante, « ? », par rapport à une situation inconnue, jamais expérimentée; (2) ou bien il représente une régularité, avec une anticipation précisément définie parmi $dom(X_i)$, laquelle anticipe la transformation de la propriété X_i'; (3) ou

l'anticipation est indéfinie, « # », lorsque l'agent ne trouve pas de déterminisme dans la transformation.

(406) L'anticipation *e* d'un schéma donné est justement la valeur d'une propriété spécifique x_i' que le schéma cherche à anticiper, afin de caractériser une régularité, où $X_i \in (P \cup H)$. Elle est constituée d'un seul élément *e*, qui est associé à une propriété p_i' ou h_i'. L'ensemble des valeurs que l'anticipation peut assumer est formé par le domaine de valeurs de la propriété respective, auquel on ajoute trois nouveaux éléments: l'*anticipation indéterminée*, « # », quand elle représente une transformation non-déterministe; l'*anticipation inexistante*, « ? », cas que se produit lorsque le schéma est nouveau et n'a jamais été mis en exécution; et l'*anticipation de préservation*, « ≈ », quand il est prévu que la valeur reste inchangée après l'application du schéma. Alors, $e \in (dom(X_i) \cup \{\#, ?, \approx\})$.

(407) Il faut noter qu'un schéma n'est pas une règle, du type « *contexte* → *action* », parce que la connaissance représentée n'a pas un caractère impératif, mais plutôt anticipatoire, du type « *contexte* ∧ *action* → *anticipation* ». Le schéma ne définit pas un comportement qui va nécessairement être adopté par l'agent à la rencontre d'une situation donnée, mais indique plutôt les transformations que l'agent doit s'attendre à observer s'il effectue l'action proposée par le schéma. L'anticipation d'un schéma est une prévision pour l'instant qui suit son activation. En outre, pour rendre possible le processus d'apprentissage, le schéma doit garder d'autres informations telles que sa fiabilité (ρ) et sa valeur affective (υ).

3.2.5.1. Généralisation

(408) L'ensemble des schémas d'un arbre constitue un partitionnement encore plus généralisé que celui mis en œuvre par la mémoire épisodique. Un schéma Ξ est un classeur qui décrit de façon généralisée un ensemble particulier de situations qui sont équivalentes à des fins d'anticipation.

(409) Les vecteurs *p*, *h*, *c* des arbres d'anticipation en général sont sous-ensembles des vecteurs *p*, *h*, *c* de l'esprit de l'agent. Les signaux reçus par l'esprit de façon continuelle sont ensuite filtrés par la liste Λ de chaque arbre Ψ_i, qui garde les propriétés considérées comme pertinentes. C'est-à-dire que $P_\Lambda \subseteq P_\mu$, $C_\Lambda \subseteq C_\mu$, et $H_\Lambda \subseteq H_\mu$. En plus, dans chaque arbre d'anticipation Ψ_i, les vecteurs qui identifient un schéma peuvent prendre une

valeur indéfinie « * », qui est une marque de généralisation. Ainsi, $p_{\Xi} \in (dom(P_{\Lambda}) \cup \{*\})$, $c_{\Xi} \in (dom(C_{\Lambda}) \cup \{*\})$, et $h_{\Xi} \in (dom(H_{\Lambda}) \cup \{*\})$.

(410) Par exemple, supposons que P_i soit une perception binaire d'un agent hypothétique, définie par le domaine $dom(P_i) = \{0, 1\}$. Dans un schéma, l'élément du vecteur de contexte lié à cette perception peut donc prendre l'une des trois valeurs possibles: $p_{i\,\Xi} \in \{0, 1, *\}$. Un autre exemple serait $dom(P_j) = \{\spadesuit, \clubsuit, \heartsuit, \diamondsuit\}$, ce qui donnerait $p_{j\,\Xi} \in \{\spadesuit, \clubsuit, \heartsuit, \diamondsuit, *\}$. Cette façon de représenter la généralisation est similaire à celle popularisée par (HOLLAND et al., 1986), où la valeur indéfinie généralise le schéma car il permet d'ignorer quelques propriétés pour représenter des ensembles de situations. Par exemple, un schéma Ξ qui a son vecteur de contexte perceptif défini comme p_{Ξ} = [100*] assimile les situations [1000] et [1001].

(411) L'utilisation de valeurs indéfinies pour certains éléments des vecteurs qui décrivent les situations, et par conséquent la capacité de créer des représentations généralisées, est, en effet, un moyen de regrouper un ensemble d'états du monde sous un identifiant unique. Dans le scénario idéal, un schéma ne doit utiliser que les propriétés pertinentes pour la transformation qu'il décrit. La représentation généralisée des situations permet d'éviter l'énumération (combinatoire) des états, en réduisant le nombre de schémas nécessaires pour décrire la dynamique de l'environnement.

3.2.6. Analyse d'Extensibilité

(412) Si les arbres d'anticipation étaient construits par un algorithme naïf, alors l'espace nécessaire pour les stocker, ainsi que le temps nécessaire pour les traiter, subiraient une explosion combinatoire avec l'augmentation de la taille du problème. Si on laissait l'arbre d'anticipation pousser sans discernement, il deviendrait rapidement ingérable, puisque l'inclusion de chaque différenciateur provoquerait une multiplication du nombre de nœuds. En outre, un arbre trop grand fragmente excessivement l'espace de situations, en créant des schémas trop spécialisés, qui sont donc rarement activés, ce qui augmente de façon exponentielle la quantité d'expériences nécessaire pour la convergence de l'algorithme.

(413) Cependant, dans le mécanisme CALM, les arbres d'anticipation sont construits en se basant sur une liste de différenciateurs pré-sélectionnés de la mémoire épisodique.

Cela simplifie la complexité de la construction de l'arbre, mais exige de l'efficacité dans le traitement de la mémoire épisodique, puisque celle-ci stocke explicitement des combinaisons multiples d'observations.

(414) C'est le paramètre $\alpha_{\text{Ш}}$ qui empêche la croissance exponentielle de la mémoire épisodique, parce que c'est lui qui limite le nombre de dépendances conditionnelles qu'elle prendra en compte. Par conséquent, le paramètre $\alpha_{\text{Ш}}$ indique aussi la limite de profondeur de l'arbre d'anticipation, puisque le nombre de dépendances conditionnelles observées détermine la quantité maximale de différenciateurs dans Λ.

(415) Le nombre de nœuds de l'arbre, ainsi que le nombre de cellules dans la mémoire épisodique, est d'ordre exponentiel par rapport à $\alpha_{\text{Ш}}$, donc $O(n^{\alpha})$, où $n = |P \times H \times C|$. On remarque que n n'est pas le nombre d'états, cardinalité typiquement indiqué pour les MDPs, mais le nombre de variables (perceptives, abstraites et de contrôle) du FMDP. Ainsi, la faisabilité de calcul de cette représentation, et donc de l'algorithme, incombe au paramètre $\alpha_{\text{Ш}}$.

(416) Si $\alpha_{\text{Ш}}$ est une valeur constante et petite, alors la quantité de mémoire nécessaire à CALM pour résoudre les problèmes d'apprentissage augmente de façon polynomiale, donc gérable. Toutefois, si $\alpha_{\text{Ш}}$ devenait trop petit, la mémoire épisodique deviendrait trop généralisée, ce qui empêcherait le mécanisme de découvrir les régularités nécessaires pour décrire un modèle du monde correct. Tel que cela a été mentionné précédemment, si l'agent est situé dans un environnement bien structuré, alors le nombre de propriétés pertinentes grandit de façon logarithmique par rapport à la taille du problème, et grâce à cela, il suffit que le paramètre $\alpha_{\text{Ш}}$ soit d'une magnitude similaire à $log_b(n)$ pour que CALM puisse travailler correctement, en gardant une complexité quasi-polynomiale, $O(n^{log(n)})$.

(417) La taille totale maximale $|\text{Ш}|$ de la mémoire épisodique généralisée est donnée par l'équation 3.7, qui est la somme de la taille de ses tables. Pour cela $n = |P \times H|$, alors il peut y avoir jusqu'à $(n).(\alpha_{\text{Ш}}+1)$ tables Ш_{ij} dans la mémoire. L'index i varie de 1 à n (ce qui correspond à chaque propriété x_i) et l'index j varie de 0 à $\alpha_{\text{Ш}}$ (ce qui correspond à chaque niveau de dépendance considéré).

$$|\text{Ш}| = \sum_{i=1}^{n} \sum_{j=0}^{\alpha} |\text{ш}_{ij}| \quad \text{(eq. 3.7)}$$

(418) La taille d'une table donnée $|\text{Ш}_{ij}|$, comme indiqué dans l'équation 3.8, est le produit cartésien du nombre de colonnes par le nombre de lignes. Le nombre de colonnes est équivalent à la quantité de sous-ensembles dans l'ensemble des n propriétés, combinés j à j, donc C_n^j. Le nombre de lignes est la combinaison des valeurs des propriétés considérées dans la colonne. En supposant que chaque propriété puisse prendre b valeurs différentes, alors le nombre de lignes sera équivalant à b à la puissance j.

$$|\text{ш}_{ij}| = b^j \cdot \frac{n!}{j! \cdot (n-j)!} \quad \textit{(eq. 3.8)}$$

(419) Par exemple, en imaginant un problème qui présente n = 30 propriétés binaires, et en établissent le paramètre $\alpha_{\text{Ш}}$ = 5, alors, par le biais des équations on trouve que CALM dimensionne la mémoire épisodique généralisée avec, au maximum, un peu plus de 5 millions d'entrées. Une trentaine de propriétés peut paraitre peu par rapport aux problèmes du monde réel, cependant, elles représentent plus de 1 milliard d'états dans une représentation plate de l'environnement, une quantité intraitable pour des algorithmes traditionnels d'apprentissage. La figure 3.17 montre que la croissance de la mémoire épisodique généralisée est d'un ordre plus petit que l'exponentiel, $O(2^n)$, et même plus petit que le quasi-polynomial, $O(n^{log(n)})$.

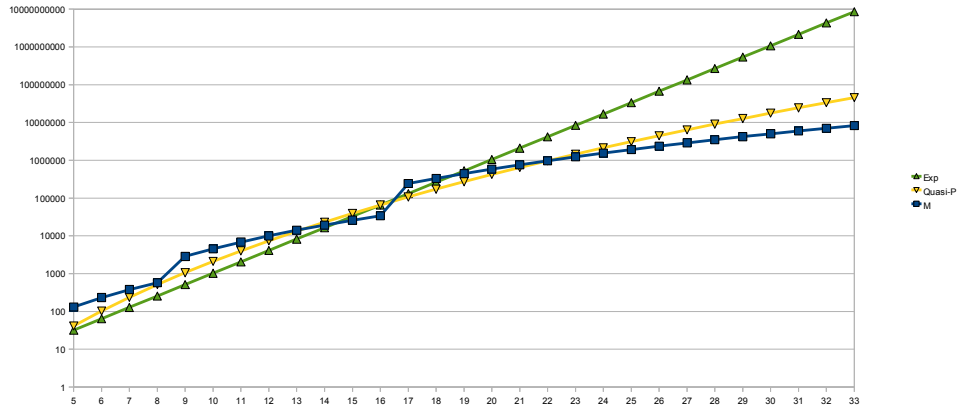


Figure 3.17: Croissance de la mémoire épisodique généralisée (en échelle logarithmique)

L'axe vertical compte le nombre d'états, et l'axe horizontal compte le nombre de variables. La complexité (en taille) de la mémoire épisodique généralisée est, au pire, selon on augmente le nombre de variables du problème, mieux que les ordres exponentiel et quasi-polynomial. On fait augmenter le paramètre $\alpha_{\text{Ш}}$ en échelle logarithmique (d'après la définition d'un environnement bien structuré), selon on augmente le nombre de propriétés. La discontinuité de la ligne de la mémoire est conséquence de l'arrondi du paramètre $\alpha_{\text{Ш}}$.

3.2.7. Actualisation de l'Arbre d'Anticipation

(420) L'affrontement des nouvelles situations en général occasionne des changements dans la mémoire épisodique. Quand ces modifications rendent nécessaire la restructuration d'un arbre d'anticipation, alors il s'agit d'une *expérience déséquilibrante*, qui nécessite une accommodation, au sens de (PIAGET, 1936). Dans ce cas, le mécanisme peut: (a) corriger le schéma inadapté; (b) restructurer l'arbre d'anticipation; ou (c) décider de créer un nouvel élément synthétique.

(421) D'une façon générale, la construction de chaque arbre d'anticipation suit l'évolution suivante: lors d'un premier moment, à partir d'un schéma initial général, l'arbre a une tendance à s'élargir, par *différenciation (split)*, à travers l'inclusion de nouveaux différenciateurs, et la création consécutive de schémas chaque fois plus spécialisés. Si les propriétés perceptives ne sont pas suffisantes, la différenciation peut exiger l'utilisation ou la *création d'éléments synthétiques* pour représenter des propriétés non-observables de l'environnement. Ensuite, cette tendance s'inverse. Quand il n'est plus possible de partitionner l'espace, ou quand il est déjà excessivement partitionné, les situations de déséquilibre cognitif conduisent à l'indétermination, par *ajustement*, des anticipations des schémas déséquilibrés, en les considérant comme des transformations non-déterministes. Enfin, peu à peu, la combinaison d'ajustements successifs révèle des différenciations inutiles, et ainsi progressivement les branches redondantes de l'arbre sont unifiées par *intégration* (*join*), ce qui réduit sa taille.

(422) Dans des environnements complexes, il est intéressant de partitionner l'espace à travers une stratégie *top-down*, c'est-à-dire en commençant par un schéma simple et général, qui donne progressivement naissance à des schémas plus spécialisés, selon se posent les expériences déséquilibrantes, par la recherche et l'inclusion des propriétés pertinentes dans les vecteurs qui définissent les schémas. Une fois démarré, le processus qui se produit alors est celui d'une différenciation et spécialisation progressive des vecteurs de contexte et d'action des schémas. Inversement, l'évolution des anticipations suit une stratégie *bottom-up*, c'est-à-dire, en allant du particulier au général.

(423) Dans cette approche, l'algorithme devient naturellement incrémental, car l'agent a un modèle du monde général et complet à n'importe quel instant t de l'exécution du programme (et donc de la vie de l'agent), par rapport aux expériences déjà vécues,

même si au début, quand elles sont encore peu nombreuses, ce modèle est souvent imparfait. L'algorithme 3.4 explique le code de base de la méthode d'apprentissage de modèle du monde.

Méthode CALM – ACTUALISE MODÈLE

```
SOIENT:
  Ψ  l'ensemble des arbres d'anticipation
  Ψi un des arbres d'anticipation dans l'ensemble
  Ξx le schéma activé de l'arbre en question;
  Ωi la mémoire épisodique généralisée de l'arbre
  p,h,c  la situation vécue
  p'     la perception suivante

COMMENCEMENT:

  POUR chaque arbre Ψi de Ψ FAIRE:

    Actualise_Mémoire(Ωi,x,p');

    SI Pertinentes(Ωi) ≠ Λi ALORS:
      Ψi ← Réorganiser(Ψi,Ωi);
      Ψi ← Intégration_Générale(Ψi);

    SINON:

      SI Nouveau(Ξx) ALORS
        Détermine(Ξx,x);

      SI Déséquilibré(Ξx,x) ALORS
        Ajuste(Ξx);

      SI Modifié(Ξx) ALORS
        Intégration_Locale(Ξx);

FIM;
```

Algorithme 3.4: Méthode d'apprentissage de modèles du monde, qui actualise les arbres d'anticipation.

3.2.7.1. Description Maximale et Minimale

(424) Selon les termes utilisés par (BUCHANAN; WILKINS, 1993), l'identité d'un schéma Ξ_i donné, composé par ses vecteurs *p*, *h*, et *c*, est équivalente à une *description discriminante minimale*, car elle spécifie le nombre minimum de différenciateurs qui sont nécessaires pour distinguer ce schéma des autres en le gardant encore cohérent avec les situations déjà expérimentées. De l'autre côté, la mémoire épisodique généralisée constitue une *description caractéristique maximale* pour les situations déjà expérimentées, dans la limite établie par le paramètre α_Ω.

(425) D'une façon similaire à un espace de versions (MITCHELL, 1982), la vraie régularité de l'environnement, celle que le schéma essaye de représenter, se situe entre les limites fixées par l'identificateur du schéma (généralisation consistante générale maximale) et les limites posées par la mémoire épisodique (généralisation consistante spécifique maximale), comme cela est montré figure 3.18.

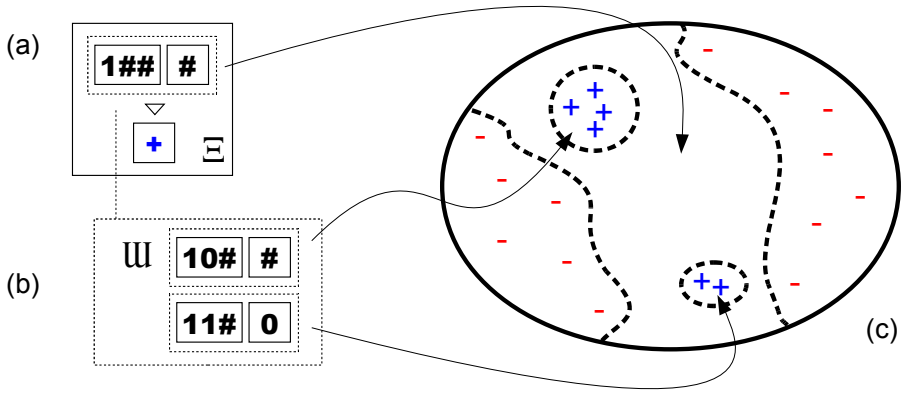


Figure 3.18: Relation général / particulier, entre le schéma et la mémoire.

Un schéma (a) représente la “généralisation consistante générale maximale”, et la mémoire épisodique associée (b) représente les "généralisations consistantes spécifiques maximales".

3.2.7.2. Les Schémas Initiaux et La Construction Progressive de l'Arbre

(426) Initialement, chaque arbre d'anticipation Ψ_i a un seul schéma, qui est aussi la racine de l'arbre. Ce schéma-racine a un identificateur complètement général, ce qui signifie qu'il n'existe encore aucune différenciation, et que toutes les situations (c'est-à-dire n'importe quel contexte combiné avec n'importe quelle action) sont assimilées par ce schéma unique. L'anticipation de ce schéma-racine est initialisée dans un état spécial d'indéfinition. Donc nous avons d'abord $\Xi_{[*][*][*]}$ = « ? », selon l'algorithme 3.5, détaillé ci-après.

CALM: Méthode INITIALISE STRUCTURES

```
SOIENT:
  Ψ  : l'ensemble des arbres d'anticipation
  Ψi : une des arbres d'anticipation dans l'ensemble
  Λ  : la liste des différenciateurs de l'arbre
  Θ  : la liste des nœuds intermédiaires de l'arbre
  Ξ  : la liste des schémas de l'arbre
  Ξ0 : un schéma initial racine
  p  : le vecteur de contexte perceptif du schéma
  c  : le vecteur de contrôle du schéma
  h  : le vecteur de contexte non-observable du schéma
  e  : l'anticipation du schéma
  b  : la cardinalité de la propriété de l'anticipation
  Ш  : la mémoire épisodique généralisée de l'arbre
  α  : le nombre de dépendances conditionnelles gardées dans la mémoire
  ν  : la valeur affective du schéma
  ρ  : la fiabilité du schéma
COMMENCEMENT:
   POUR chaque propriété xi en (P ∪ H) FAIRE:
      Créer Ψi:
         Λ ← ∅;
         Θ ← ∅;
         Ш ← {?,?, … ,?};
         Créer Ξ0:
            p ← '*'; h ← '*'; c ← '*';
            e ← '?';
            ν ← Sensation_Affective(e);
            ρ ← 0.0;
         Ξ ← {Ξ0};
   Ψ ← Ψ ∪ {Ψi};
FIN;
```

Algorithme 3.5: Méthode d'initialisation des structures de la connaissance.

(427) En supposant que l'arbre Ψ_i anticipe une propriété perceptive, alors la première fois que le schéma initial (unique) est activé, son anticipation est remplacée par l'observation directe du résultat trouvé immédiatement après, par conséquent $\Xi' = \{\Xi_{[*][*][*]} = p_i'\}$. En fait, tout schéma nouveau, après la première exécution, fait définir son anticipation en accord avec le premier résultat observé. Dans le cas des arbres qui anticipent des propriétés non-observables, le processus est différent et il sera détaillé ultérieurement. L'algorithme 3.6 décrit la méthode de détermination initiale des anticipations.

Méthode DÉTERMINER ANTICIPATION

```
SOIENT:
  Ξ   un schéma activé dans le cycle précédent et en déséquilibre
  x_i' la valeur de la propriété X_i dans la situation actuelle
  e   l'anticipation du schéma pour x_i'

COMMENCEMENT:

   SI e = "?" ALORS:
      e ← x_i';

FIN;
```

Algorithme 3.6: Détermination initiale des anticipations.

3.2.7.3. Différenciation

(428) Lorsqu'un déséquilibre se produit, la première hypothèse qu'envisage le mécanisme, c'est que le schéma déséquilibré est encore trop général, et par conséquent englobe un domaine d'application trop large. La solution est alors de créer de nouveaux schémas plus spécialisés, qui puissent représenter correctement la régularité de la transformation par le partitionnement de son domaine.

(429) Le principe qui régit cette procédure c'est que si le schéma ne fonctionne pas bien, c'est qu'il est probable que l'ensemble de situations assimilées par lui est trop grand. La méthode partage alors son champ d'assimilation par la création de nouveaux schémas plus spécialisés, dans une stratégie du genre « diviser pour régner ». Le schéma instable est remplacé par un sous-arbre de deux niveaux. Il reste en tant que nœud parent de ce sous-arbre, en ayant pour fils les nouveaux schémas. Ces nouveaux schémas ont leurs identificateurs plus spécialisés d'un niveau, par rapport à celui de leur père. Le principe de la différenciation est illustré figure 3.19.

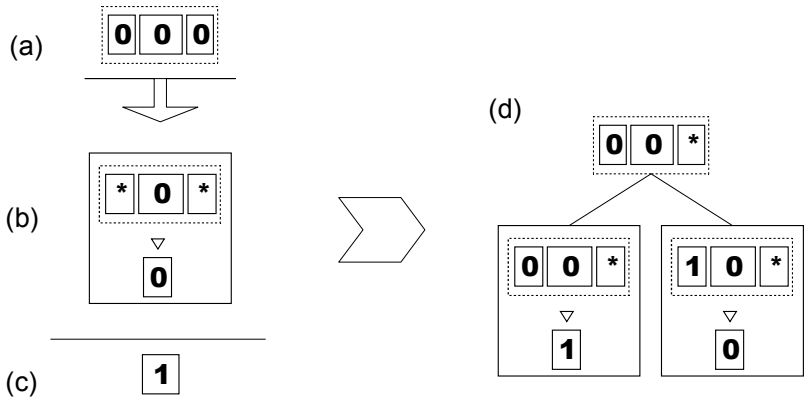


Figure 3.19: Exemple de différenciation.

(a) contexte et action expérimentés; (b) schéma activé; (c) résultat réel observé; (d) sous-arbre généré.

3.2.7.4. Ajustement

(430) L'ajustement transforme une anticipation déterminé en une anticipation indéterminé. La méthode se met en marche quand il n'est plus possible de différencier un schéma déséquilibré. Dans ce cas, l'ajustement annule l'anticipation du schéma instable, en changeant sa valeur pour le symbole du non-déterminisme « # ». Dans CALM, une anticipation indéterminée est consistante avec n'importe quelle expérience, mais sans rien anticiper.

(431) En général, au fil du processus d'apprentissage, les arbres d'anticipation commencent par être optimistes, en essayant de trouver des anticipations déterminées pour toutes les situations. Peu à peu, en raison des ajustements, les anticipations des schémas se réduisent jusqu'à ce qu'il ne reste que celles qui représentent les vraies régularités déterministes de l'environnement.

(432) Les différenciations sont limitées par le paramètre α_Λ, qui indique le nombre maximal de différenciateurs dans un arbre d'anticipation, qui est donc sa profondeur maximale. Lorsque cette limite est atteinte, les déséquilibres sont résolus par ajustement. La figure 3.20 illustre le principe d'ajustement.

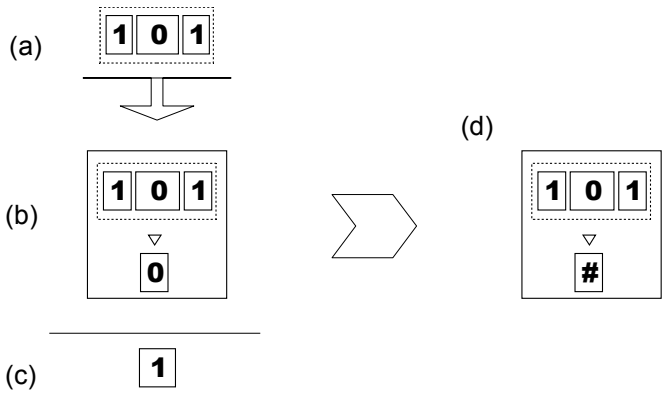


Figure 3.20: Exemple d'ajustement.

(a) contexte et action expérimentés; (b) schéma activé; (c) résultat réel observé; (d) schéma après l'ajustement, qui a réduit son anticipation.

3.2.7.5. Intégration

(433) Enfin, des ajustements successifs sur les anticipations de plusieurs schémas finissent par dévoiler des différenciations inutiles, qui peuvent alors être retirés de l'arbre. L'intégration cherche des sous-arbres qui ont des anticipations identiques. L'idée

est de réunir ces sous-arbres en un seul. Le principe d'intégration, montré à l'algorithme 3.7, est illustré figure 3.21.

CALM – Méthode INTEGRATION:

```
SOIENT:
  Ξ_A   un schéma dont l'anticipation a été modifiée
  S     un ensemble de sous-arbres frères
  S_U   une unification possible entre les sous-arbres dans S
  Θ_i   un nœud de l'arbre
  d     un élément différenciateur au niveau de Θ_i
  x'_s  l'ensemble d'anticipations des sous-arbres S
BEGIN:
  POUR chaque nœud Θ_i, à partir du père de Ξ_A vers la racine, FAIRE:
    S ← sous-arbre(Θ_i);
    SI Compatibles(x'_s) ALORS:   //sous-arbres ont des anticipations similaires
      S_U ← Unifier(S);           //unifie les schémas correspondants des sous-arbres
      d ← '*';                    //généralise le différenciateur
      FAIRE de S_U un fils unique du nœud Θ_i;
    SI pour un nœud au niveau Θ_i, d = *, ALORS:
      Λ ← Λ − {d};                //élimine le différenciateur
END;
```

Algorithme 3.7: Méthode d'intégration de sous-arbres.

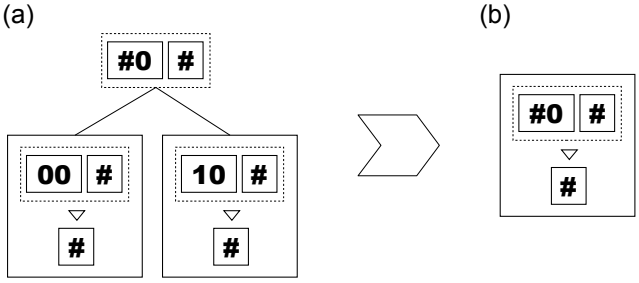


Figure 3.21: Intégration entre schémas frères.

(a) sous-arbre après un ajustement; (b) un schéma intégré remplace le sous-arbre.

(434) Il faut noter que la méthode d'intégration analyse l'ensemble de sous-arbres frères à partir du schéma modifié. Alors l'intégration peut se produire entre des schémas qui ne sont pas directement frères dans l'arbre, mais cousins. La figure 3.22 montre ce cas d'intégration, où la différenciation éliminée n'est pas immédiatement au-dessus du schéma modifié.

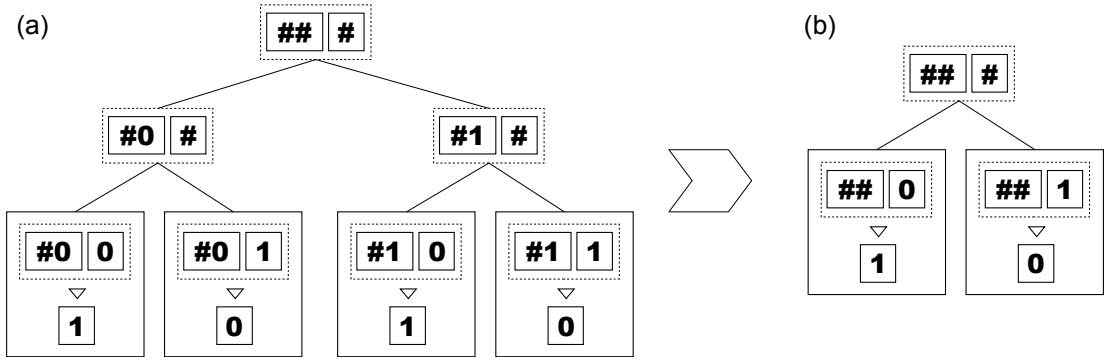


Figure 3.22: Intégration entre schémas cousins.

3.2.8. Propriétés Non-Observables

(435) Dans le cas partiellement observable se pose le problème de la *découverte des propriétés non-observables*. L'agent doit induire l'existence de caractéristiques cachées ou abstraites de l'environnement, et les inclure dans le vocabulaire de représentation de son modèle du monde. CALM traite les propriétés non-observables de l'univers à travers la *création d'éléments synthétiques*, dans l'ensemble H, qui sont ajoutés aux éléments de perception P et de contrôle C pour décrire les situations.

(436) Dans des environnements partiellement observables, l'existence d'aliasing perceptif (des situations qui se confondent) impose une difficulté importante pour les algorithmes d'apprentissage (CROOK; HAYES, 2003). En fait, plus le nombre moyen de propriétés pertinentes non-observables est grand ($1 >> \omega > 0$), plus le processus d'apprentissage est difficile.

(437) De façon similaire à (DRESCHER, 1991) et (LANG, 1999), pour traiter l'observation partielle, CALM utilise des éléments synthétiques. ils sont destinés à postuler l'existence de quelque chose au-delà de la perception sensorielle qui soit capable d'expliquer ou de différencier les situations ambiguës. Ces nouveaux éléments représentent des variables qui ne sont pas directement observables par l'agent, mais qui peuvent avoir leurs valeurs déduites par l'analyse de la chaîne historique des contextes sensorimoteurs, et ils peuvent donc contribuer à la définition de nouvelles anticipations.

3.2.8.1. Découverte des Propriétés Abstraites

(438) Dans le cas complètement déterministe ($\partial = 1$) mais partiellement observable ($\omega < 1$), les transformations peuvent être représentées en tant que fonctions déterministes sous la forme $\tau_i : P \times H \times C \rightarrow X_i$. Mais l'agent ne perçoit pas

sensoriellement les variables de l'ensemble H, qui sont des propriétés pertinentes mais cachées de l'environnement. L'espace de situations qui l'agent considère initialement est limité à $P \times C$. Par conséquent, même si l'environnement a une dynamique totalement déterministe, il pourra paraitre non-déterministe lorsqu'il est analysé uniquement dans sa face observable.

(439) Par contre, dans ce cas, lorsque deux situations ne sont pas distinguables par leurs caractéristiques perceptives, il est possible de les distinguer par leurs « signatures » (RON; RUBINFELD, 1997), (FREUND et al., 1997). La signature identifie la situation par son voisinage temporel, c'est-à-dire, par les transitions qui se produisent dans les cycles immédiatement avant et immédiatement après. Cette identification est illustrée figure 3.23.

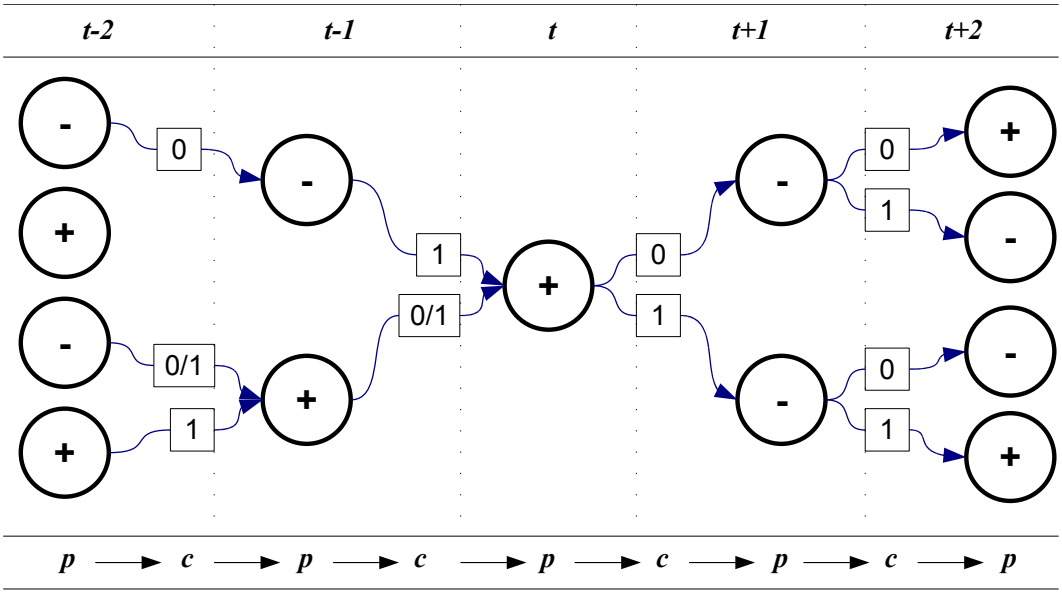


Figure 3.23: Exemple de la signature d'une situation à partir des observations.

L'état sous-jacent ne peut pas être correctement déterminé uniquement par l'observation actuelle, représentée au centre de la figure, mais il devient identifiable lors de l'analyse des séquences d'observation immédiates, passées et futures, conduisant à cet état. Si une séquence suffisante est considérée, alors la signature de chaque état se révèle unique.

(440) Lors d'une expérience déséquilibrante qui ne semble pas avoir d'explication au niveau sensorimoteur, l'agent peut supposer l'existence d'une propriété cachée de l'environnement, nécessaire pour distinguer les situations confondues. Les donnés que l'agent utilise pour découvrir l'occurrence de cette ambiguïté sont les séquences d'observations postérieures à la signature (les observations qui ont suivi le passage par une situation possiblement ambigüe), comme illustré figure 3.24.

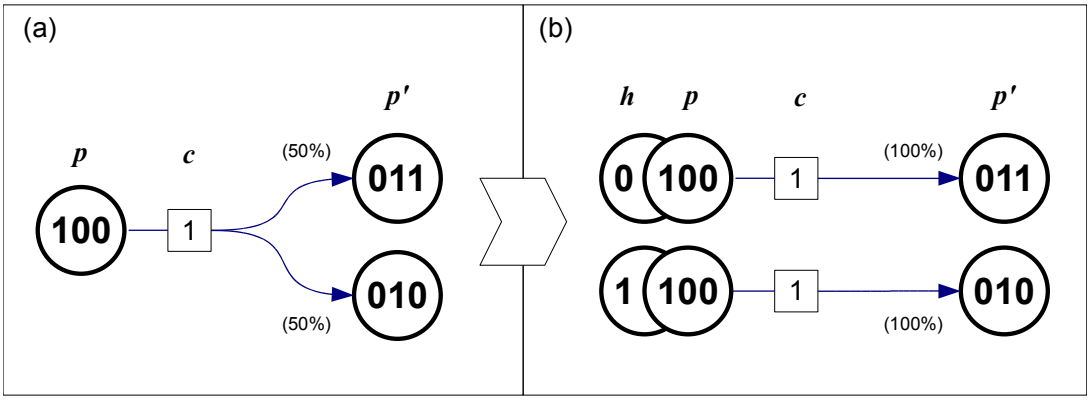


Figure 3.24: Désambigüisation.

À gauche (a), la perception directe ne suffit pas pour savoir si l'observation fait référence à un même état sous-jacent; à droite (b), l'incohérence entre les résultats des transitions permet de déduire, dans le cas déterministe, qu'il s'agissait de deux états différents.

(441) Ainsi, si l'environnement est déterministe et bien structuré, l'agent peut, sans crainte d'erreurs, créer un élément synthétique chaque fois qu'il trouve une expérience déséquilibrante, pour laquelle il ne dispose pas d'éléments suffisants dans son répertoire actuel pour la différencier. Par l'inclusion d'éléments synthétiques dans l'ensemble H, afin de représenter des propriétés cachées, l'espace initial de représentation des transformations $P \times C$ sera progressivement étendu, jusqu'à se stabiliser, quand toutes les propriétés pertinentes de l'environnement seront présentes dans $P \times H \times C$. À ce moment, l'agent sera en mesure de représenter correctement toutes les transformations.

(442) Pour CALM le problème n'est pas aussi simple que cela, parce que l'environnement n'est pas complètement déterministe, donc $\partial < 1$, et en même temps, il est partiellement observable, $\omega < 1$. Dans ce cas, il n'est pas facile de définir si une transformation a été observée comme non-déterministe car elle l'est vraiment, ou car il y a encore une propriété cachée non représentée. La figure 3.25 illustre les deux façons alternatives d'expliquer une situation ambigüe.

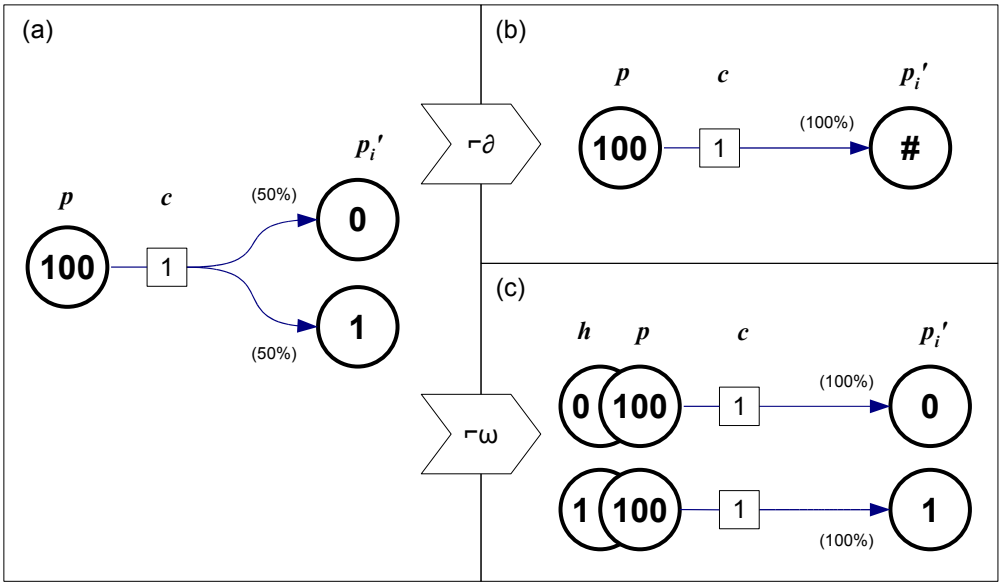


Figure 3.25: L'observabilité partielle peut expliquer l'apparence non-déterministe.

À gauche (a), on illustre l'observation des deux transformations observées à partir d'une situation identique, mais qui, cependant, conduisent à des résultats différents. À droite les deux solutions possibles sont présentées: (b) supposer que la transformation est non-déterministe, et alors cesser d'essayer de l'anticiper; (c) supposer l'existence d'une propriété cachée qui explique la transformation d'une façon déterministe, et alors introduire un élément synthétique.

(443) Il faut empêcher que l'agent tombe dans le piège de créer de nouveaux éléments synthétiques indéfiniment dans l'espoir de trouver du déterminisme dans une transformation non-déterministe. Dans ce cas, il n'arrivera jamais à construire un modèle stable. En raison de cela, un paramètre α_H est utilisé pour établir le nombre maximal des éléments synthétiques que le mécanisme peut créer.

3.2.8.2. Anticipation de la valeur d'une propriété cachée

(444) La seule découverte de l'existence de propriétés non-observables, et l'attribution des éléments synthétiques correspondantes, n'est pas suffisante pour construire un modèle du monde en mesure de bien anticiper les transformations. L'agent doit également connaître les valeurs des éléments synthétiques au fil du temps, et puisqu'il s'agit de propriétés qui ne sont pas directement observables, il n'y a aucun moyen de saisir leurs valeurs directement par la perception.

(445) Autrement dit, quand un élément synthétique h_j est conçu pour résoudre les difficultés d'anticipation d'une propriété particulière x_i de l'arbre Ψ_i, il est nécessaire de créer un nouvel arbre d'anticipation Ψ_j pour modéliser la transformation des valeurs de h_j.

(446) Supposons qu'une seule perception p = [0] est la face observable de deux situations sous-jacentes différentes. L'ambigüité de la situation p = [0] est remarqué parce qu'elle conduit l'agent de fois à p' = [0] et de fois à p' = [1]. Ainsi, on peut supposer que cette transformation est régulé par une propriété non-observable, représentée par un élément synthétique, alors en différenciant les situations hp = [00] et hp = [10]. Supposons un agent observe la séquence $p^{(0)}$ = [0], $p^{(1)}$ = [0], et $p^{(2)}$ = [1], respectivement dans les cycles d'exécution t = 0, t = 1, e t = 2. Dans les instants t = 0 et t = 1 l'agent ne voit pas d'ambiguïté, en ignorant l'existence d'une propriété cachée. Cependant, lorsqu'il atteint $p^{(2)}$ = [1], l'agent pourra: (1) supposer l'existence d'une propriété non-observable qui est pertinente pour décrire la transformation de p, et qui sera alors représentée par h; (2) déduire qu'il était précédemment dans la situation hp = [00] qui est différente de hp = [10]; et (3) inclure l'anticipation de l'élément synthétique h à partir de p, comme le montre la figure 3.26.

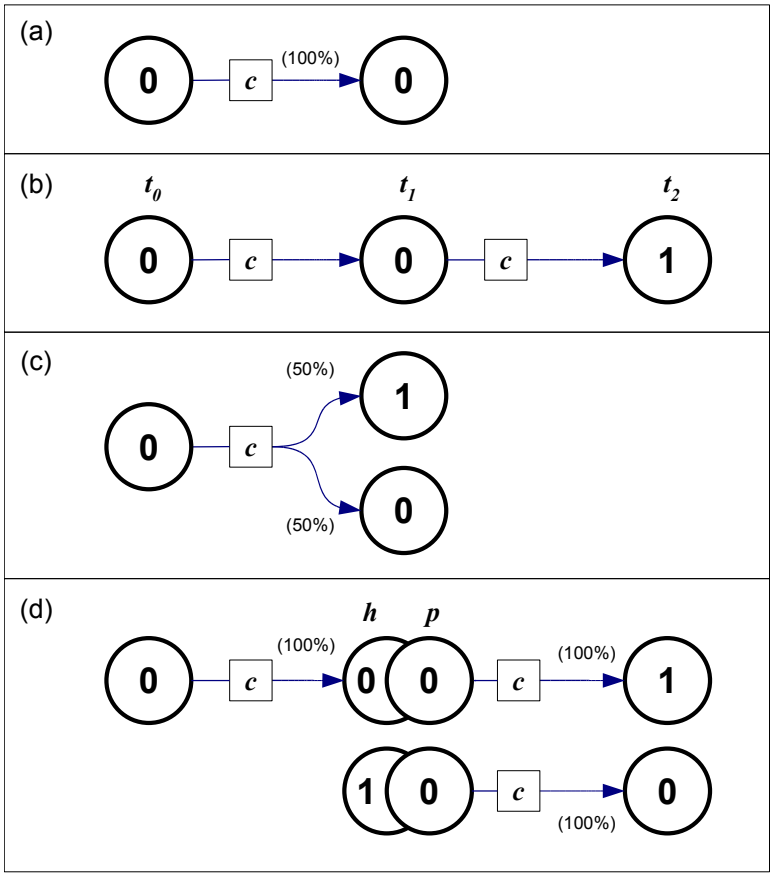


Figure 3.26: Induction de l'existence de propriétés non-observables.

Illustration d'une séquence de 3 états perceptifs atteints par l'agent à partir d'une certaine action. En (a) la connaissance initiale. En (b) l'agent enchaine une expérience deséquilibrante. En (c) on représente cette contradiction en tant que non-déterminisme. En (d) on montre la solution CALM: le mécanisme utilise l'information de la séquence de transformation, en créant un élément synthétique nécessaire à distinguer entre deux situations ambigües.

3.2.8.3. Rétropropagation de l'élément synthétique

(447) Tant que la supposition d'une nouvelle propriété non-observable n'est pas nécessaire, le mécanisme va tenter de lever toute ambiguïté en utilisant les éléments synthétiques déjà créés. Éventuellement, comme c'est le mécanisme lui-même qui induit les valeurs des éléments synthétiques, une erreur lors de l'anticipation d'une propriété sensorielle dont la prévision dépend de la valeur des éléments synthétiques peut signifier que c'est l'anticipation de la valeur non-observable qui n'est pas précise.

(448) Ainsi, lorsque cette possibilité existe, avant d'essayer de créer de nouveaux éléments synthétiques, le mécanisme doit observer s'il y a d'autres schémas précédents au déséquilibre, qui peuvent être différenciés en fonction des éléments synthétiques déjà existants. Chaque arbre d'anticipation Ψ_i possède à cet effet, une *mémoire récente* $\mathcal{M}_i$ qui

garde le souvenir d'une séquence finie des schémas récemment activés. Le nombre de schémas retenus par cette mémoire est limitée par le paramètre α_M. Dans le cas d'une prédiction qui a échoué, la première alternative que le mécanisme essaye est d'exécuter une « différenciation retardée » dans un schéma activé dans le passé, qui était potentiellement responsable du déséquilibre manifesté dans l'instant présent.

3.2.9. Processus de Décision

(449) Pour réaliser le *processus de prise de décision*, le mécanisme CALM utilise les connaissances dont il dispose (son modèle du monde) pour interpréter les situations et pour planifier ses actions. Le problème est équivalent à la recherche d'une bonne politique d'actions dans un PD-FPOMDP. À partir de l'ensemble d'arbres d'anticipation $\Psi = \{\Psi_1, \Psi_2, \ldots, \Psi_{|\Psi|}\}$, CALM construit un ensemble d'arbres de délibération $Ж = \{Ж_1, Ж_2, \ldots, Ж_{|Ж|}\}$, qui est la première étape vers la constitution d'une politique d'actions. Un arbre de délibération est formalisé dans la définition 3.7.

Un arbre de délibération ($Ж_i$) est un quadruplet:

$Ж_i = \{đ, \Theta, \Lambda, E\}$

où

$đ = \{đ_1, đ_2, \ldots, đ_{|đ|}\}$ est un ensemble de décideurs (feuilles de l'arbre)

$\Theta = \{\Theta_1, \Theta_2, \ldots, \Theta_{|\Theta|}\}$ est un ensemble de nœuds intermédiaires

Λ est une liste ordonnée de différenciateurs

E est l'ensemble de propriétés à anticiper, où $E \subset (P \cup H)$

Définition 3.7: Arbre de Délibération.

(450) Un arbre de décision est construit à partir de l'union d'un ou de plusieurs arbres d'anticipation. Ainsi, alors que chaque arbre d'anticipation est responsable de la prévision d'une seule propriété, les arbres de délibération peuvent anticiper une combinaison d'entre elles. Une fois constitués, chacun des arbres présentera un ensemble de décideurs, qui sont alors évalués, en fournissant les paramètres pour la construction de la politique d'actions.

3.2.9.1. Excitation et Activation des Décideurs

(451) Dans le mécanisme CALM, le processus de décision marche de la façon suivante: à chaque instant du temps *t*, ce qui équivaut à un cycle d'exécution, le contexte dans lequel l'agent se trouve, perçu par l'esprit à travers des signaux *p* et *h*, excite les

décideurs compatibles dans chacun des arbres de délibération. Chaque décideur excité đ$_j$ est un candidat pour l'activation dans l'arbre Ж$_i$, en présentant une alternative d'action à effectuer par le biais du signal de contrôle *c*.

(452) Le décideur est compatible avec une situation donnée lorsque les éléments définis dans son vecteur perceptif p_d sont équivalents à ceux de la perception actuelle p_μ reçus par l'esprit de l'agent, et les éléments synthétiques h_d sont compatibles avec les valeurs actuelles de h_μ, qui sont induites par le mécanisme lui-même. Les décideurs anticipent les conséquences possibles de leur activation, et à partir de cette prévision le processus d'activation choisi l'un d'eux pour être activé.

(453) Les différenciateurs dans l'arbre de délibération sont organisés comme suit: d'abord, plus proche de la racine, les différenciateurs du type *P*; en suite, les différenciateurs du type *H*; et enfin, avant d'atteindre les décideurs, il y a les différenciateurs de type *C*. Le processus d'excitation des décideurs commence par la racine de l'arbre de délibération, et se propage à travers les branches, pour atteindre les décideurs (les nœuds terminaux).

(454) Au cours du processus d'excitation de l'arbre de délibération, dans la première partie du chemin, parmi des différenciateurs du type *F*, une seule branche du nœud est excitée, celle qui est compatible avec la perception actuelle. Puis, dans la deuxième partie du chemin, le flux d'excitation dépend des inférences que le mécanisme a été capable de faire par rapport aux valeurs des éléments de l'ensemble *H*. Si la valeur d'un élément h_i différenciateur a été inférée par le système cognitif, alors l'excitation suit cette branche uniquement, sinon toutes les branches sont excitées, puisque la valeur du différenciateur n'est pas connue. Enfin, la dernière partie du chemin présente les différenciateurs du type *C*, où l'excitation se propage de manière inconditionnelle à tous les branches. À la fin, seulement les décideurs compatibles avec la situation actuelle seront excités, en présentant à la fois les options d'action (pour *C*) et éventuellement quelques doutes par rapport au véritable état de l'environnement (en *H*).

(455) L'étape suivante consiste à choisir un décideur à activer parmi les décideurs excités. Ce choix est fait sur la base des valeurs affectives attribuées à leurs anticipations respectives. CALM reçoit un signal de récompense factorisé, provenant du système affectif de l'esprit. Pour chaque élément p_i du signal de perception *P* qui soit

une variable essentielle de l'organisme, il y aura un signal d'évaluation v_i correspondant. Enfin, l'action proposée par le décideur activé est effectivement réalisée par l'agent. Quand un décideur est activé, son action c_d est envoyée comme signal de contrôle c_μ de l'esprit au corps, en mobilisant les actuateurs de l'agent.

(456) La figure 3.27 illustre une décision. Dans l'exemple, l'agent perçoit la situation [01], qui excitera deux décideurs alternatifs $đ_{[*1][*][0]}$ et $đ_{[*1][*][1]}$. Le système évaluatif fournit une valeur affective pour chaque décideur, en fonction de leurs anticipations. Enfin, le schéma de plus grande valeur affective (+0,8) est activé, puis l'action [1], proposée par lui, est envoyée au signal de contrôle.

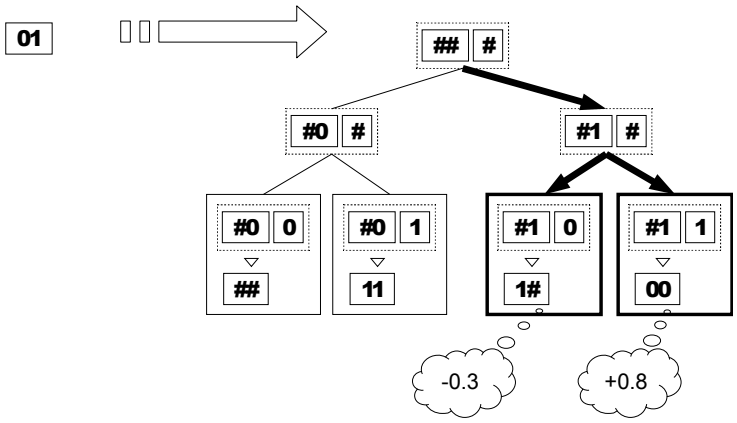


Figure 3.27: Exemple d'une prise de décision.

À gauche on a la situation perçue, et à droite, l'arbre de délibération de l'agent au moment de la décision.

3.2.9.2. Exploration et Exploitation

(457) Tout algorithme d'apprentissage actif et incrémental affronte le *dilemme entre l'exploration et l'exploitation*. Pour construire le modèle du monde et la politique d'actions, il est nécessaire que l'agent explore explicitement l'environnement (KAELBLING et al., 1996), en essayant des actions alternatives, ou en cherchent à vivre des situations nouvelles. Mais faire des explorations sans jamais tirer profit des connaissances acquises n'a pas beaucoup de sens pour un agent situé. Les connaissances acquises doivent être utilisées pour maximiser l'utilité des actions, et par conséquent pour augmenter la moyenne des retours affectifs reçus au cours du temps. Toutefois, si l'agent décide trop rapidement d'abandonner l'exploration d'alternatives, il prend le risque d'accepter un modèle du monde insuffisant, et d'avoir une politique d'actions

maintenue dans un maximum local, une fois que la politique est dérivée du modèle, en générant un comportement insatisfaisant (RUSSELL; NORVIG, 1995).

(458) Pour cette raison, CALM a un *paramètre de curiosité* ε qui caractérise la tendance de l'agent à explorer ce qui n'est pas bien connu, ou, au contraire, à chercher les situations affectivement plus agréables. Le paramètre ε indique la probabilité que la décision soit donnée en fonction de l'évaluation affective υ des décideurs, ou en raison de leur fiabilité ρ.

(459) Quand l'agent est en train d'exploiter, le choix du décideur à être activé est basé sur la valeur affective attribuée aux anticipations des décideurs excités. Celles qui mènent l'agent à des situations plus « agréables » sont préférables. Inversement, lorsque l'agent est en train d'explorer, le choix du décideur à être activé est fondé sur sa « stabilité ». Les décideurs à faible stabilité, qui n'ont pas été beaucoup testés, ou qui ont été récemment modifiés, sont préférés dans ce cas, en supposant que ce sont précisément ceux qui doivent être encore plus expérimentés, afin qu'ils puissent atteindre la stabilité. Un décideur stable est celui qui a vu ses anticipations satisfaites après plusieurs activations, et donc la stabilité d'un décideur indique également la confiance que l'agent a par rapport à ses prévisions.

3.2.9.3. Définition de l'Ensemble des Arbres de Délibération

(460) La première étape dans la construction des arbres de délibération est d'exécuter un filtrage sur l'ensemble des arbres d'anticipation, en ne sélectionnant que le sous-ensemble des arbres qui anticipent des *variables importantes*. Une variable est considérée comme importante si elle est affectivement positive ou négative, ou si elle conditionne l'anticipation d'une autre variable importante. Par exemple, supposons un problème composé de 7 variables $\{p_1, p_2, p_3, p_4, h_1, c_1, c_2\}$, où les variables de perception $\{p_1, p_2, p_3, p_4\}$ sont associées respectivement à des valeurs affectives $\{0, +0.7, -0.3, 0\}$. Dans ce cas, l'ensemble des variables importantes commence par le sous-ensemble $\{p_2, p_3\}$ des variables affectivement non-neutres.

(461) L'étape suivante consiste à analyser quelles sont les autres variables de P et H qui déterminent la transformation de ces propriétés importantes. Si, par exemple, h_1 est pertinente à la dynamique de p_2, alors h_1 sera également importante, ce qui élargit l'ensemble à $\{p_2, p_3, h_1\}$, et ainsi de suite jusqu'à ce que l'ensemble des propriétés

importantes soit clos. Supposons que, dans l'exemple, p_4 soit un facteur causal pour h_1, alors l'ensemble complet des propriétés importantes sera $\{p_2, p_3, p_4, h_1\}$.

(462) Avant de passer à la construction des arbres de décision, il est nécessaire de filtrer dans l'ensemble des propriétés importantes les variables qui ne sont soumises à aucun type de contrôle de l'agent. Imaginons, dans le même exemple, que la transformation de h_1 soit dépendante de l'action c_1. Une fois que la transformation de p_2 dépend de h_1, le contrôle réalisé par c_1, qui interfère sur h_1, va interférer aussi sur p_2, indirectement. Par conséquent, h_1 et p_2 restent dans l'ensemble. Si p_3 dépend de c_2, alors p_3 y reste aussi. Par contre p_4 sera exclue parce qu'elle n'est pas liée, pas même indirectement, à une variable de contrôle.

(463) Après que soit défini cet ensemble de variables importantes et contrôlables, dans l'exemple équivalant à $\{p_2, p_3, h_1\}$, il faut définir combien d'arbres de délibération sont nécessaires. Les arbres de délibération sont construits à partir de l'union des arbres d'anticipation qui dépendent des mêmes éléments de contrôle. Dans le cas le plus simple, chaque propriété de l'ensemble génère un arbre de délibération séparé. À l'inverse, si toutes les variables importantes sont dépendantes des mêmes décisions de contrôle, alors il est nécessaire de les regrouper toutes dans un arbre de délibération unique. Dans l'exemple, nous avons vu que p_2 et h_1 dépendent du même contrôle c_1, tandis que p_3 dépend de c_2, et dans ce cas il est nécessaire de construire deux arbres.

3.2.9.4. Construction des Arbres de Délibération

(464) La construction d'un arbre de délibération commence par la concaténation des arbres d'anticipation. L'ensemble des éléments différenciateurs (descripteurs) dans l'arbre de délibération $Ж_i$ est défini par l'union des différenciateurs de chaque arbre d'anticipation $\{\Psi_j, \Psi_k, \ldots, \Psi_m\}$ auquel il est associé, c'est-à-dire, $\Lambda_i = \Lambda_j \cup \Lambda_k \cup \ldots \cup \Lambda_m$. Ensuite, ces différenciateurs sont ordonnés de façon que ceux liés à la perception restent à proximité de la racine, puis des différenciateurs synthétiques, puis des différenciateurs liés au contrôle. En d'autres termes, *P* avant *H*, et *H* avant *C*.

(465) Ensuite, à partir de la définition des différenciateurs, la structure de l'arbre est construite, et le vecteur d'anticipations de chaque décideur (nœud terminal) est déterminé par la concaténation des anticipations unitaires des schémas dans les arbres d'anticipation. L'arbre de délibération est construit initialement comme un arbre

complet, où tous les nœuds de l'arbre sont différentiés, en incluant dans cette différenciation une branche pour la valeur indéfinie. Dans le cas des propriétés binaires, les descripteurs définissent 3 branches: {0, 1, *}.

(466) Cette première construction de l'arbre va conduire à une énumération exhaustive des décideurs. Ensuite un processus d'intégration est exécuté, qui unifie les décideurs qui ont une même anticipation. Enfin, chaque décideur est connecté à au moins un décideur successeur, selon l'anticipation qu'il fait. Les décideurs qui ne sont pas connectés peuvent être éliminés.

(467) Chaque arbre d'anticipation décrit une partie de la dynamique d'interaction entre l'agent et l'environnement, ce qui n'est qu'un morceau du modèle du monde. D'autre part, l'esprit de l'agent dispose d'un système d'évaluation qui attribue des valeurs d'affect positif ou négatif à ces anticipations. L'arbre de délibération est la structure où tous ces éléments sont combinés. D'une part, les éléments qui étaient traités comme des prévisions séparées pour les arbres d'anticipation, se mettent maintenant ensemble pour anticiper la transformation combinée de certaines propriétés. D'autre part, les situations anticipées peuvent désormais être mises en balance par le système d'évaluation, permettant au mécanisme de construire une politique d'actions appropriée, et ainsi permettre à l'agent d'adopter un comportement adapté.

3.2.9.5. Construction de la Politique des Actions

(468) Une fois que l'arbre est construit, il faut qu'il puisse évaluer les décideurs. Quatre paramètres sont calculés pour chaque décideur, selon la situation à laquelle il s'attend à conduire l'agent: (1) la valeur affective immédiate; (2) l'utilité affective, qui est une mesure de la capacité du décideur à apporter des retours affectifs futurs; (3) la valeur d'exploration immédiate, liée à la curiosité; et (4) l'utilité d'exploration, qui mesure la capacité du décideur à mener l'agent vers des situations intéressantes et inconnues à long terme.

(469) La valeur affective immédiate $\nu_{\text{Я}}$ d'un décideur est la composition (prise ici comme une somme) des affectivités impliquées dans chaque élément de son anticipation. Comme défini dans le chapitre 2, l'esprit reçoit un signal d'évaluation affective, qui est une sorte de récompense factorisée. Chaque propriété est indépendamment associée à une fonction de retour évaluatif, et la valeur de la situation

dans son ensemble peut être donnée par la simple somme des retours évaluatifs particuliers. La valeur d'exploration immédiate $v_К$ d'un décideur traduit la confiance du système pour son anticipation. Il fonctionne comme une sorte de *fitness* (BOOKER, et al., 1989), augmentant chaque fois que le décideur confirme ses anticipations, et revenant à zéro quand un élément de l'anticipation est modifié.

(470) Toutefois, il n'est pas possible de définir une politique d'actions qui maximise les gains dans une perspective à long terme, simplement en se basant sur des récompenses instantanées. Parfois, il est nécessaire de subir des petites pertes immédiates afin de réaliser de grandes récompenses un peu plus tard dans l'avenir. C'est le rôle de la fonction d'utilité, qui représente généralement la somme décomptée des gains attendus au cours du temps (SUTTON; BARTO, 1998). À partir de $v_Я$ et de $v_К$ il est possible de calculer, respectivement, l'utilité affective $u_Я$ et l'utilité exploratoire $u_К$ d'un décideur. L'utilité est définie par la formule suivante, basée sur l'équation de Bellman (1957): $u(đ) = v(đ) + \gamma \cdot u(đ')$.

(471) Toutefois, la question clé pour le calcul des utilités c'est de savoir qui est đ', c'est-à-dire, qui est le décideur qui sera activé après l'activation de đ. Le mécanisme choisit deux successeurs: $đ'_Я$ parmi les décideurs compatibles avec l'anticipation de đ, qui est le décideur pour lequel *c* maximise l'utilité affective, et $đ'_К$ parmi les décideurs dont *c* maximise l'utilité exploratoire.

4. Résultats Expérimentaux

(472) Dans ce chapitre, nous présentons une série d'expériences où l'on utilise un agent modélisé selon l'architecture CAES (telle qu'elle a été définie au chapitre 2) et qui implémente le mécanisme CALM (tel qu'il a été défini au chapitre 3) en tant que système cognitif. Les résultats expérimentaux obtenus sont ensuite comparés à ceux de travaux similaires. Deux problèmes sont présentés: (1) *wepp*, et (2) *flip*. Les expériences visent à démontrer les conséquences et les résultats de l'utilisation de l'architecture CAES et du mécanisme CALM sur ces problèmes.

(473) Le mécanisme CALM et l'architecture CAES sont des contributions de cette thèse, pourtant chacun d'eux a une finalité spécifique et différente, et l'analyse des résultats de l'expérimentation doit en tenir compte. L'architecture CAES est évaluée sur sa capacité à modéliser des problèmes d'une façon située, en montrant la factorisation du système à travers l'utilisation de senseurs et d'actuateurs vectorisés, en intériorisant la motivation de l'agent grâce à un système affectif corporel, et en définissant des problèmes d'observation partielle par la propre nature de la communication indirecte des signaux entre l'esprit et le corps, et entre le corps et l'environnement. Le mécanisme CALM est évalué sur sa capacité à apprendre un modèle du monde et une politique d'actions pour les scénarios expérimentaux présentés. Dans ce cas, les expériences ont

pour but de tester et de démontrer l'efficacité, la stabilité et la qualité des solutions obtenues pour chaque problème.

4.1. Problème Wepp

(474) Nous avons construit une première expérience, appelée *wepp* (acronyme pour *walking-exhaustion-pain-pleasure*) dans laquelle l'agent doit apprendre à se déplacer dans un environnement qui lui est inconnu, en cherchant à développer un mode de comportement qui puisse augmenter les sensations de plaisir et réduire les sensations de douleur et d'épuisement.

4.1.1. Définition du Problème

(475) Dans le problème *wepp*, l'agent peut marcher en avant ou il peut pivoter sur son propre axe dans un espace bidimensionnel. Chaque fois que l'agent marche, il reçoit une sensation de plaisir qui est affectivement positive. Toutefois, le dilemme c'est que s'il marche trop, il finit par devenir épuisé, ce qui est affectivement négatif. En outre, s'il se heurte à un obstacle, il reçoit une sensation de douleur, qui est aussi négative. Il est donc nécessaire que l'agent coordonne ses actions selon les contextes qu'il observe afin d'optimiser son comportement. Un aperçu du problème wepp est montré dans la figure 4.1, et les variables qui le concernent, ainsi que leurs relations, sont montrées figure 4.2.

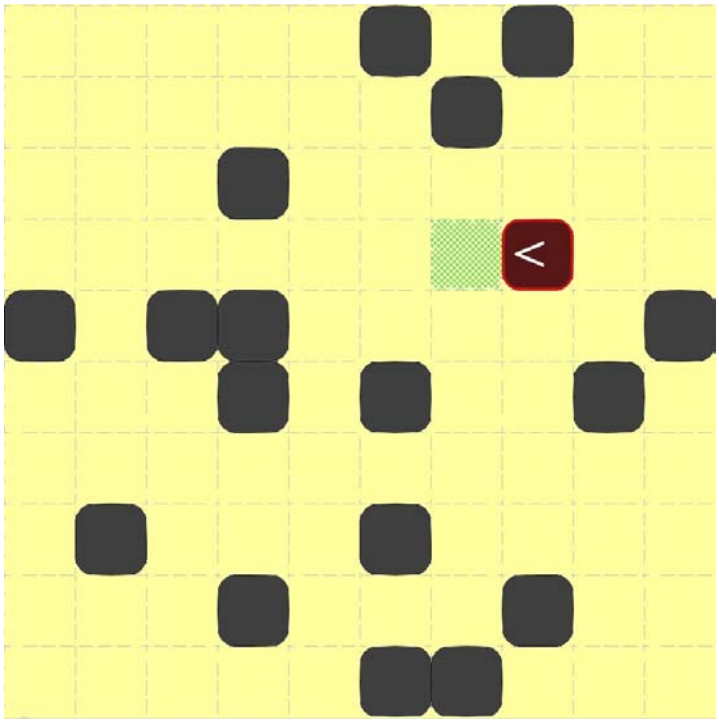


Figure 4.1: Aperçu du problème wepp.

L'objet remarqué est l'agent, tourné vers la gauche. Il regarde une cellule devant lui. Les autres objets sont les obstacles.

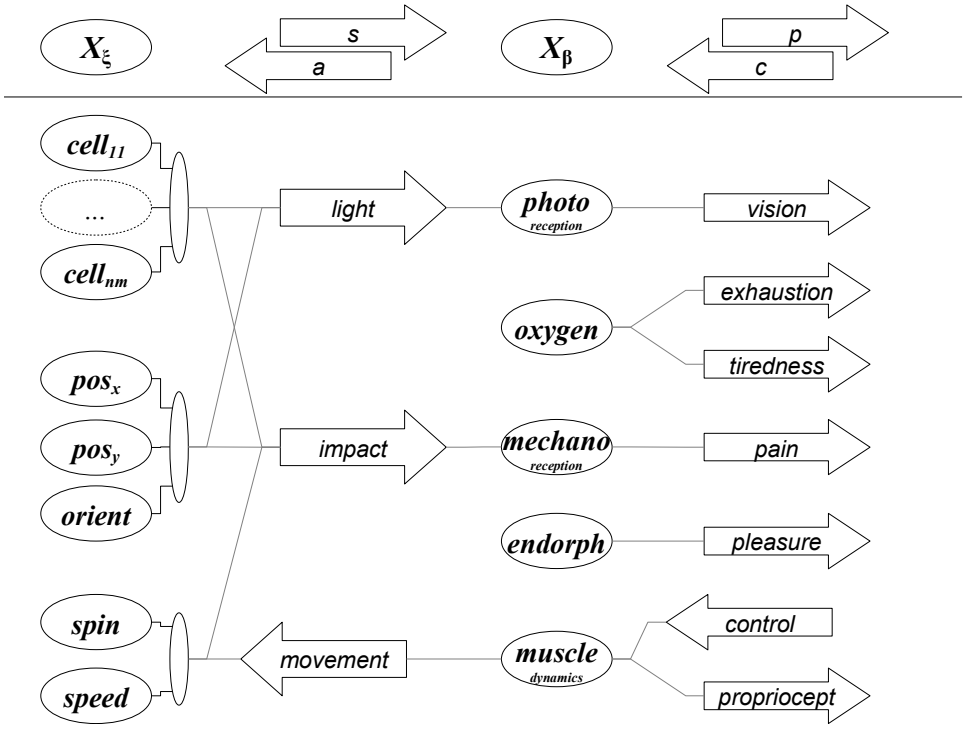


Figure 4.2: Fonction d'évolution de l'environnement dans le problème wepp.

4.1.1.1. L'Univers Wepp

(476) L'environnement de cette expérience est un espace discrétisé où l'agent se déplace. Cet espace a la forme d'un plateau (chaque objet occupe une cellule) toroïdal (les bords opposés sont connectés). Il n'existe que deux types d'objets dans cet univers: l'agent et les obstacles. À chaque cycle d'exécution, l'agent se trouve à une certaine position sur le plateau. Il peut circuler librement à travers les espaces vides, mais s'il essaye d'aller contre un obstacle, il souffrira de la collision, et son déplacement sera empêché.

(477) Dans ce problème, l'agent a un seul actuateur, qui réalise ses actions de déplacement. À chaque cycle d'exécution, en utilisant cet actuateur, l'agent peut effectuer l'une des deux actions: *marcher*, en faisant un pas en avant, ou *pivoter*, en se tournant sur lui-même à 90 degrés, à droite ou à gauche. L'agent a également un seul senseur externe, qui informe ce que l'agent voit dans la cellule devant lui.

(478) Lorsqu'on utilise l'architecture CAES, un plateau quadrillé de n par m cellules peut être représenté dans la structure de l'environnement ξ par une matrice de n par m variables. Il faut également ajouter à l'environnement un vecteur de 5 autres variables qui indiquent l'orientation actuelle de l'agent, sa position x et y, et sa vitesse de pivotage et de déplacement. Chaque cellule (variables $cell_{ij}$) du plateau peut prendre l'une des 2 valeurs possibles: {*vide*, *obstacle*}. La variable d'orientation (*orientation*) peut avoir une des 4 valeurs possibles: {*nord*, *sud*, *est*, *ouest*}. Les variables de position (pos_x et pos_y) peuvent prendre les valeurs des domaines: {*1*, *2*, … , *n*} et {*1*, *2*, … , *m*}. Les variables de pivotage (*spin*) et de déplacement (*speed*) codent de façon discrétisée la vitesse des mouvements de l'agent, respectivement sous les formes {-1, 0, 1} et {0, 1}.

(479) Dans l'architecture CAES, l'agent et l'environnement sont des systèmes indépendants et partiellement ouverts. Il y a un canal d'interférence mutuelle entre eux formé par les signaux d'*actuation* (*m*) et de *situation* (*s*). L'agent agit sur la dynamique de l'environnement par son actuation, qui dans ce cas consiste en un seul signal de mouvement (variable *movement*), qui prend une des deux valeurs: {*marcher*, *pivoter*}. La fonction de transformation de l'environnement f_ξ dépend d'abord de son l'état actuel (c'est-à-dire, la position de l'agent, et l'emplacement des obstacles), mais dépend aussi de l'actuation de l'agent.

(480) D'autre part, le signal de situation représente l'interférence de l'environnement sur l'évolution interne de l'agent. Dans cette expérience, la situation est composée de deux variables. La première variable (*light*) représente la cellule qui est devant l'agent sur le plateau, dont la valeur est donnée en fonction de la position et de l'orientation de l'agent, en indiquant la luminosité de ce qui est immédiatement en face de lui, dans le domaine {*clair*, *sombre*}. La seconde variable (*impact*) représente l'occurrence ou l'absence de collision lors des déplacements de l'agent, à cause d'un obstacle sur le chemin.

(481) La définition 4.1 montre la composition de l'ensemble de propriétés de l'environnement, et aussi la composition des signaux de situation et d'actuation, ce qui définit le problème *wepp* au niveau de l'environnement. La figure 4.3, à la suite, décrit la fonction d'évolution de l'environnement.

Ensembles de propriétés de l'environnement dans le problème wepp:

$X_\xi = \{cell_{(1,1)}, \ldots, cell_{(n,m)}, pos_x, pos_y, orientation, spin, speed\}$

$M = \{movement\}$

$S = \{light, impact\}$

Définition 4.1: Variables du problème wepp, au niveau de l'environnement.

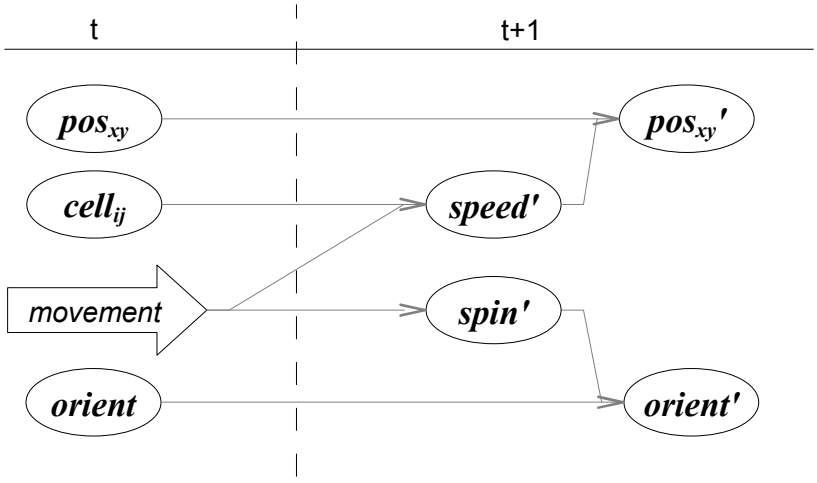


Figure 4.3: Fonction d'évolution de l'environnement dans le problème wepp.

4.1.1.2. Le Corps de l'Agent Wepp

(482) Le corps de l'agent est composé de 5 propriétés internes. Bien que ce soit un corps virtuel, on associe des noms biologiques à ces variables, afin de caractériser la fonction de chacune dans l'organisme de l'agent. La première variable (*endorphin*) est liée à la sensation de plaisir de l'agent. La deuxième variable (*mechanoreception*) est

liée à l'impact. Une troisième variable (*oxygen*) représente l'énergie actuelle de l'agent. Il y a une autre variable (*muscledynamics)* qui contrôle l'actuateur de mouvement de l'agent. La dernière variable (*photoreception*), relative à son senseur externe, retourne la perception de luminosité, en indiquant ainsi la présence ou l'absence des obstacles à l'avant. La plupart de ces propriétés sont des variables binaires, à l'exception de l'oxygène, défini dans une échelle discrète de 0 à 20.

(483) La fonction d'évolution du corps f_β est définie par les règles suivantes. (1) La présence d'*endorphine* dans le corps de l'agent ne dépend que de son *mouvement*, c'est-à-dire que si l'agent marche, l'endorphine est produite dans l'organisme, mais pas si l'agent tourne. (2) La *mécanoréception* devient active lorsque l'agent entre en collision avec un obstacle, c'est-à-dire qu'elle arrive au moment où l'agent reçoit un *impact*, et elle disparaît l'instant suivant. (3) À chaque pas que fait l'agent, il y a une dépense d'énergie, ce qui diminue progressivement l'*oxygénation* du corps. En outre, l'action de pivoter ne dépense pas d'énergie, en permettant au corps d'augmenter son oxygénation. (4) La *musculodynamique* est déterminée par un signal de contrôle provenant de l'esprit. (5) La *photoréception* est définie par le signal de situation *light*.

(484) Le corps de l'agent communique avec l'esprit à travers les signaux de *perception* (*p*) et de *contrôle* (*c*). Le signal de perception est composé de 6 variables. La première correspond à la sensation de douleur (variable *pain*), et est liée à la mécanoréception dans le corps, qui à son tour est déterminée par l'occurrence d'impact entre l'agent et des obstacles dans l'environnement. La deuxième variable (*pleasure*) correspond à la sensation de plaisir, et informe l'esprit sur la présence ou l'absence d'endorphine dans le corps. La *fatigue* (variable *tiredness*) est un signal déclenché lorsque le niveau d'oxygène dans l'organisme est inférieur à 7 sur une échelle de 0 à 20, c'est-à-dire 35 %. Une autre sensation (variable *épuisement*), également liée à la condition d'oxygénation du corps, est déclenché lorsque le niveau est inférieur à 6. Une cinquième variable (*vision*) communique à l'esprit le signal de photoréception, qui à son tour indique la luminosité de ce qui est en face de l'agent dans l'environnement. La dernière variable de la perception (*proprioception)* donne le feedback du mouvement, en indiquant l'état de l'unique actuateur de l'agent. L'esprit peut commander une seule signal de contrôle (variable *control*), qui détermine la musculodynamique du corps, et qui à son tour induit des changements dans l'environnement.

(485) La définition 4.2 montre la composition de l'ensemble des propriétés du corps de l'agent, et aussi la composition des signaux de perception et de contrôle. La définition 4.3, ci-après, représente la fonction d'évolution du corps.

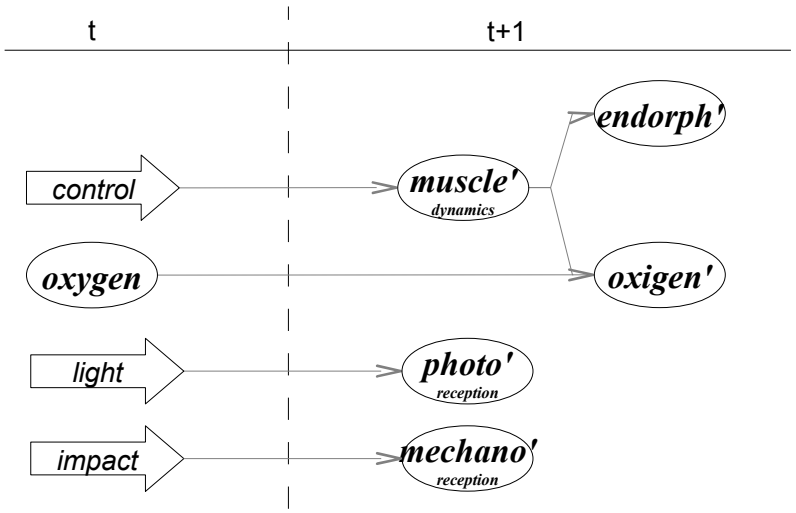
Ensembles de propriétés du corps de l'agent dans le problème wepp:

X_β = {*endorphin, mechanoreception, oxygen, muscledynamics*, *photoreception*}

P = {*pain*, *pleasure*, *tiredness*, *exhaustion*, *vision, proprioception*}

C = {*control*}

Définition 4.2: Ensembles de propriétés du corps de l'agent dans le problème wepp.



Définition 4.3: Fonction d'évolution du corps dans le problème wepp.

(486) La dynamique des signaux de l'esprit, en fonction de l'évolution du corps, peut être résumée comme suit: lorsque l'agent effectue à plusieurs reprises l'acte de *marcher*, la *fatigue* se produit, et dans ce cas, il suffit alors que l'agent fasse un seul pas de plus pour que la sensation d'épuisement soit activée. Ces sensations disparaissent lorsque l'oxygénation du corps reprend des niveaux normaux. En bref, l'action de marcher peut amener l'agent à la fatigue parce que chaque pas en avant diminue son oxygénation. Par contre pivoter éloigne l'agent de l'état de fatigue parce que cette action permet au métabolisme de l'organisme d'augmenter l'oxygénation.

(487) Le problème *wepp* est intéressant parce qu'il crée un conflit à être résolu par l'agent: il veut toujours marcher une fois que comme ça il éprouve le plaisir, mais il ne peut pas le faire d'une façon inconséquente parce que sinon il finirait par éprouver aussi des conséquences désagréables comme l'épuisement et de la douleur.

4.1.1.3. Questions de Situativité

(488) C'est important de remarquer que dans le problème *wepp*, en utilisant l'architecture CAES, la position spatiale de l'agent constitue, à travers les variables pos_x et pos_y, une paire de propriétés de l'environnement, et non des variables internes de l'agent. L'agent peut interférer sur la valeur de ces variables par le biais du signal d'actuation, dans ce cas formé par la variable de *mouvement*. D'autre part, l'observation que l'agent fait de l'environnement passe par les senseurs de *mécanoréception* et de *photoréception*, qui sont des propriétés internes de l'agent, et qui par conséquent ne font pas partie de l'environnement, même si c'est l'environnement qui les détermine, au travers du signal de situation, composé ici des variables *light* et *impact*.

(489) Dans les expériences, le problème *wepp* a été représenté de deux façons distinctes: d'une part en utilisant la configuration classique des problèmes de labyrinthe dans le domaine de l'apprentissage automatique, dans laquelle l'agent reçoit comme entrées sa position et orientation sur le plateau, et d'autre part l'architecture CAES, où l'agent perçoit l'environnement à travers une fenêtre visuelle. Les résultats expérimentaux montrent que, en raison de la nature située de CAES, l'agent peut résoudre le problème sans cartographier explicitement l'environnement. Autrement dit, il peut apprendre un patron de comportement qui maximise le retour affectif, en se basant uniquement sur son actuation et son point de vue locaux, et sans avoir nécessairement un modèle du monde qui forme une carte avec la position de tous les obstacles.

(490) Deux avantages sont remarquables dans l'approche située. Tout d'abord, la difficulté de résoudre le problème n'augmente pas avec l'augmentation de la taille du plateau. Deuxièmement, la solution trouvée par l'agent est indépendante de la position des obstacles, et ainsi l'agent peut transposer la solution directement d'un scénario à l'autre.

(491) Il n'y a pas une « position-cible » sur le plateau, et il n'y a donc pas le concept d'état final, en tant que motivation extrinsèque. Dans le problème *wepp*, les buts de l'agent sont établis indirectement par l'intermédiaire du système régulatif de l'esprit, qui attribue des valeurs affectives négatives pour la douleur et pour l'épuisement, et une valeur affective positive pour le plaisir. Ces sensations affectives sont équivalentes à un

signal de renforcement intériorisé. Dans les expériences, on utilise les valeurs suivantes: *douleur = -1.0*; *épuisement = -0.7*; *plaisir = +0.4*.

4.1.2. Résultats et Considérations

(492) On a réalisé deux implémentations du problème *wepp*. Dans la première on a utilisé une représentation classique, dans laquelle l'agent perçoit directement l'environnement, de façon omnisciente, une fois qu'il est informé de sa position et de son orientation. Dans la seconde, nous avons utilisé l'architecture CAES en tant que référence pour la modélisation, dans laquelle l'agent est situé.

(493) Deux mécanismes d'apprentissage différents ont été exécutés. Le premier, utilisé comme référence de comparaison, est une implémentation typique de l'algorithme *Q-Learning* (WATKINS; DAYAN, 1992), où l'apprentissage est réalisé en se basant sur une représentation plate des états, c'est-à-dire dans un espace formé par la combinaison des valeurs de toutes les variables considérées. Le deuxième est l'implémentation du mécanisme CALM.

(494) 9 différentes configurations de scénario ont été utilisées, en augmentant progressivement la taille du plateau, ainsi que la quantité relative d'obstacles. Dans la configuration la plus simple, le plateau a une dimension 5 *x* 5 (25 cellules), la seconde configuration établit un plateau 25 *x* 25, et la configuration la plus grande est celle d'un plateau 125 *x* 125 (15.625 cellules). Les obstacles sont aléatoirement placés, en faisant varier la quantité d'obstacles de 10% à 20% puis à 30%.

(495) Pour chacune de configurations on a défini 10 scénarios randomisés, et l'expérience a été répétée 10 fois pour chacun d'eux, en totalisant donc 900 simulations. Ces neuf configurations sont exemplifiées figure 4.4, où les cellules sombres sont les obstacles, les cellules claires sont les espaces vides, et l'objet circulaire est l'agent. Lorsque l'emplacement aléatoire des objets à laissé l'agent enfermé dans un sous-espace trop petit, le scénario a été écarté et une nouvelle randomisation a été réalisée.

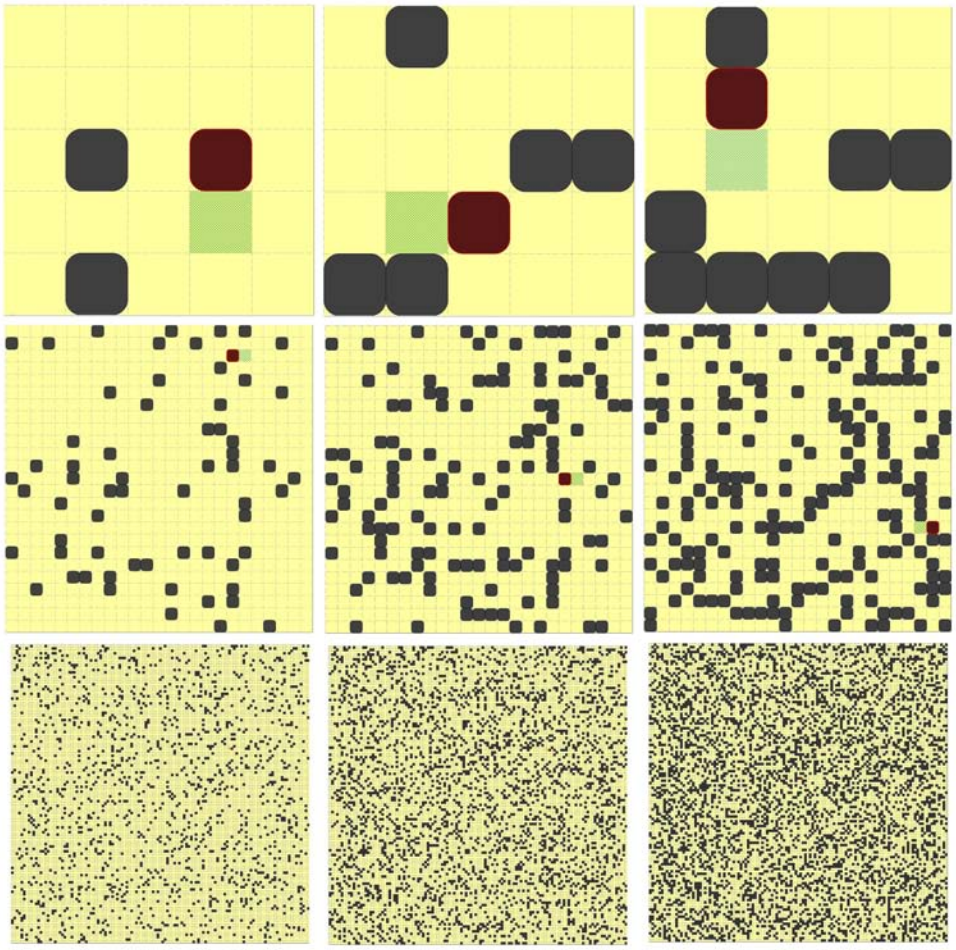


Figure 4.4: Configurations de simulation utilisées pour les expériences du problème *wepp*.

Les configurations montrées dans la ligne d'en haut montrent des plateaux de dimensions 5 x 5 qui présentent, de gauche à droite, respectivement, 10%, 20%, ou 30% des cellules occupées par des obstacles. Au milieu, on montre les configurations 25 x 25, et en bas, les configurations 125 x 125.

4.1.2.1. Évaluation du CAES: les effets de la situativité

(496) Les graphiques des figures 4.5, 4.6 et 4.7 montrent la série d'expériences dans les différents scénarios, en utilisant, dans cette première étape, uniquement l'algorithme *Q-Learning*. Les résultats confirment que l'augmentation de la taille du plateau n'influence pas le temps de convergence ni la qualité de la solution, lorsque l'agent est implémenté avec CAES. En revanche, l'implémentation classique, rattachée à la carte du plateau, subit une augmentation exponentielle dans le temps requis pour la convergence, en fonction de la taille du problème.

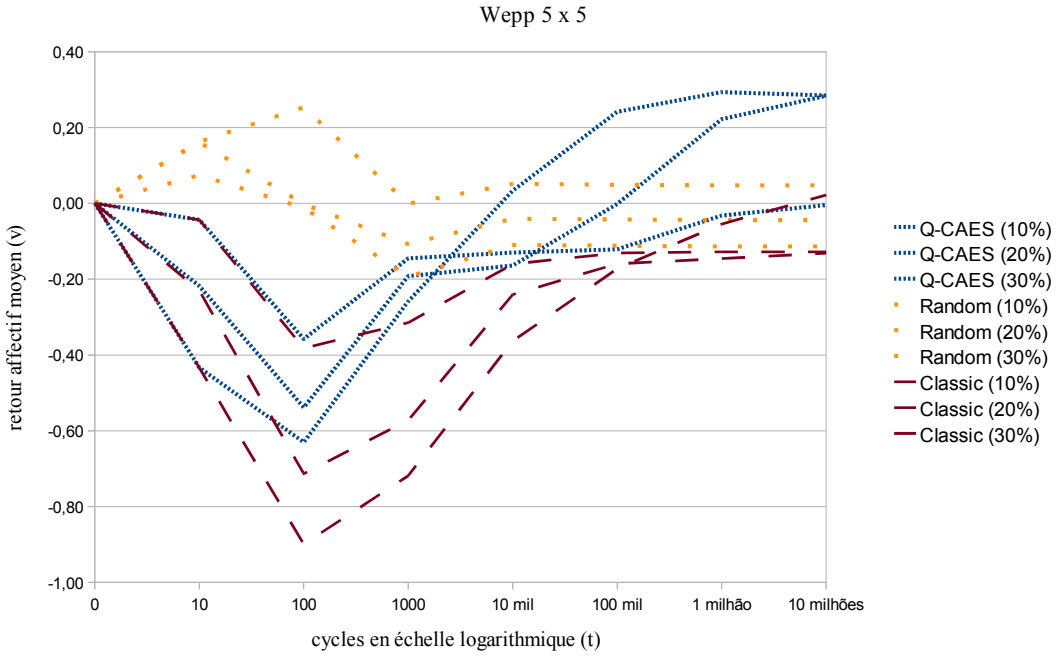


Figure 4.5: Courbes de convergence pour le plateau 5 x 5 (25 cellules).

L'implémentation située (CAES) converge un peu plus rapidement, et la solution est atteinte avant 10 mille cycles. L'implémentation classique se montre stable seulement vers 100 mille cycles. L'algorithme d'apprentissage utilisé à ce moment est Q-Learning.

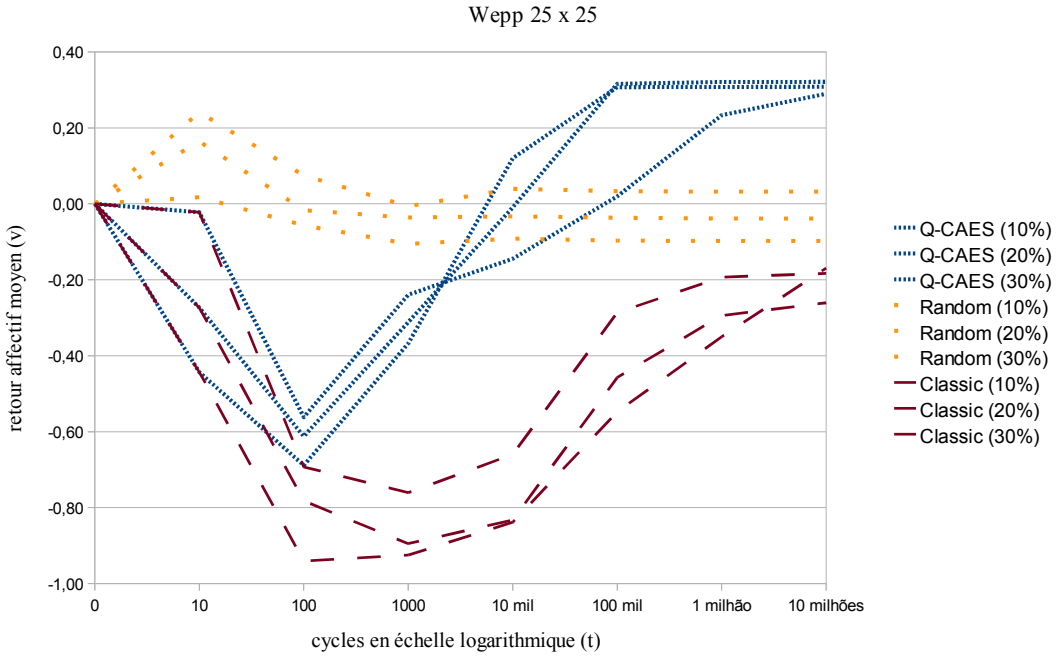


Figure 4.6: Courbes de convergence pour le plateau 25 x 25 (625 cellules).

L'implémentation située (CAES) ne se dégrade pas par rapport au temps de convergence, même si le plateau est plus grand, alors que l'implémentation classique passe de 100 mille à 1 million de cycles.

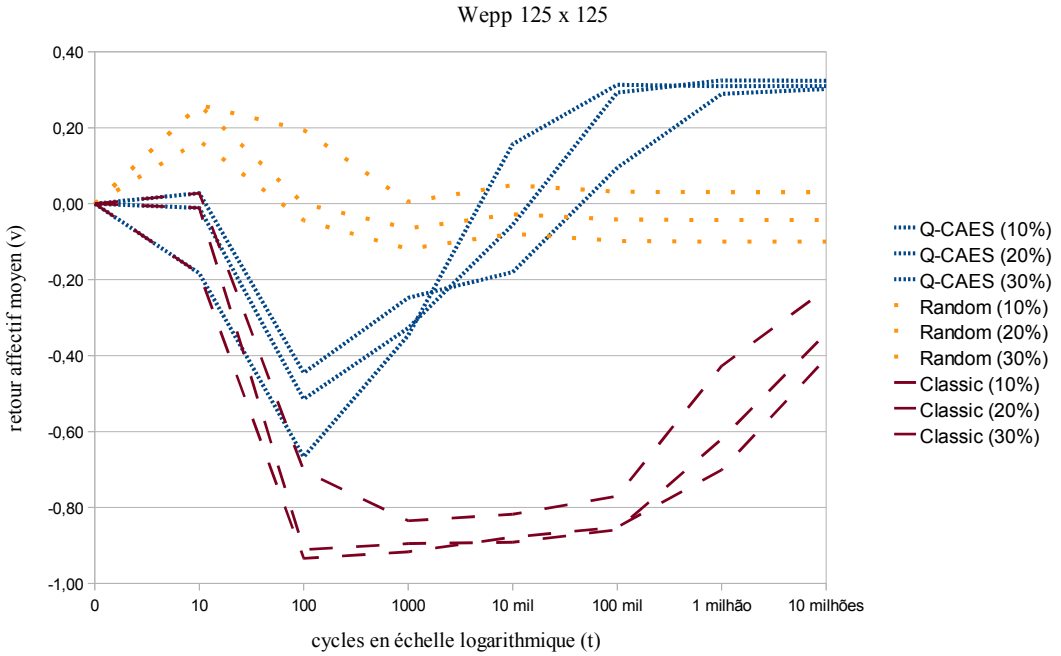


Figure 4.7: Courbes de convergence pour le plateau 125 x 125 (15625 cellules).

L'implémentation située (CAES) converge dans 10 mille cycles, pendant que l'implémentation classique met 10 millions.

(497) Quand l'agent est situé (CAES), la solution n'est pas affectée par l'augmentation de la complexité du plateau, car elle n'est pas basée sur une cartographie de l'environnement, mais sur la découverte des régularités observées dans l'interaction entre agent et environnement. Différemment, l'implémentation classique soufre avec ça, selon on peut voir figure 4.8. Autrement dit, pour toutes les configurations testées, à la fois la solution et le temps de convergence ont été similaires. Cela est dû à l'utilisation du modèle situé, défini dans l'architecture CAES, lorsque l'agent n'a pas la perception omnisciente du plateau, mais un point de vue centré sur lui, selon sa position spatiale et son orientation.

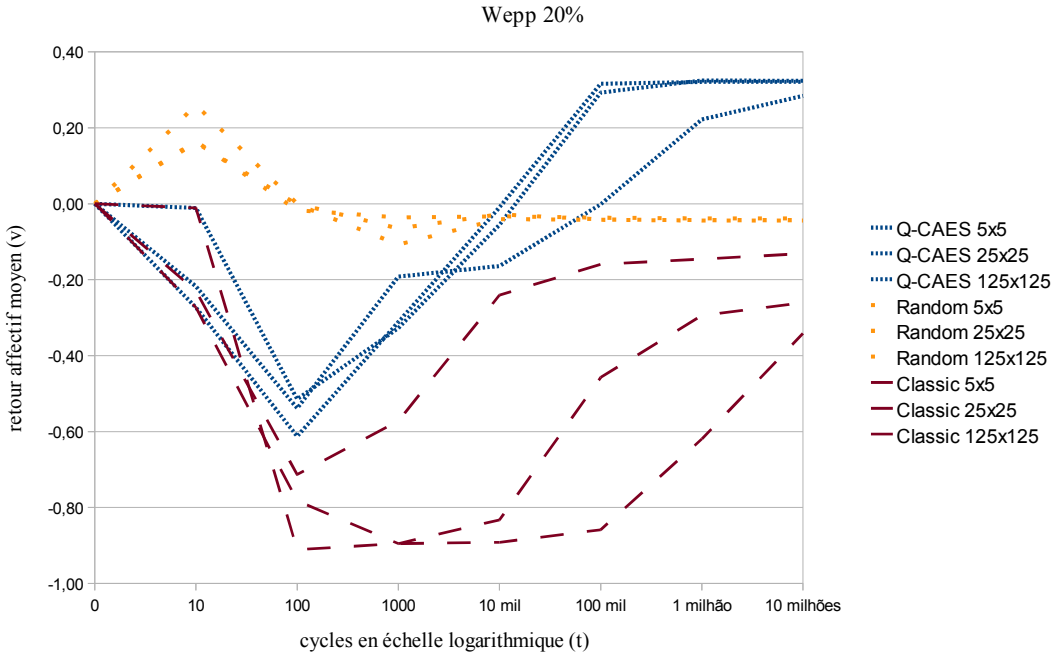


Figure 4.8: Courbes de convergence pour 20% d'obstacles.

On peut facilement observer l'impact de la taille du plateau sur l'implémentation classique (non-située), ce qui n'arrive pas quand on utilise l'architecture CAES.

4.1.2.2. Évaluation du CALM: solution constructiviste

(498) Au début de la simulation, l'agent ne sait rien sur la façon dont se déroule la dynamique d'interaction avec l'environnement, ni ce qui cause ses sensations. Il ne distingue pas les obstacles des chemins libres, et il ne connait pas les conséquences impliquées par ses actions. Dans ces conditions, le mécanisme CALM a été capable de converger de façon constante vers la solution attendue, en construisant un modèle du monde adéquat pour représenter les régularités de l'environnement, les régularités de ses sensations corporelles, ainsi que pour représenter l'influence régulière de ses actions sur les deux.

(499) L'agent apprend sur les conséquences de ses actions dans différentes situations, qui sont représentées par un nombre réduit de schémas bien généraux. À partir d'eux le mécanisme peut construire une politique d'actions qui lui permet d'éviter les situations affectivement négatives et de chercher celles qui sont affectivement positives. Cette solution arrive à décrire précisément toutes les régularités que l'agent peut percevoir sans construire un plan complet de l'environnement. Le modèle du monde construit par

CALM est montré à travers ses arbres d'anticipation dans les figures 4.9, 4.10, 4.11, 4.12, 4.13 et 4.14.

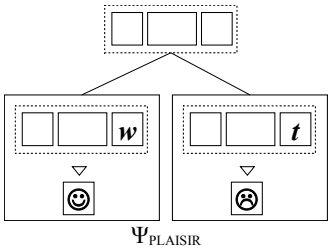


Figure 4.9: Arbre d'anticipation du plaisir.

Les schémas indiquent (de gauche à droite): “si marcher alors plaisir”; et “si pivoter alors non-plaisir”.

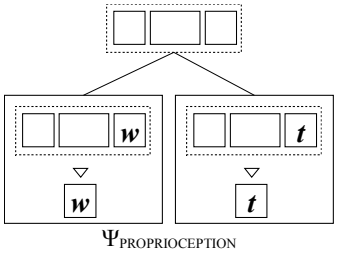


Figure 4.10: Arbre d'anticipation de la proprioception.

“Si marche, perçoit la marche”; “si pivote, perçoit le pivotage”

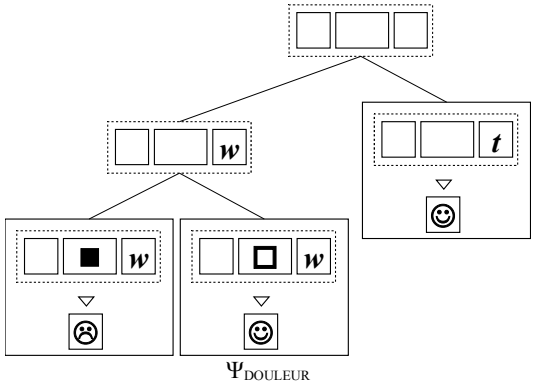


Figure 4.11: Arbre d'anticipation de la douleur.

“Si vision du mur et marcher alors douleur”; “si vision du vide et marcher alors non-douleur”; et “si pivoter alors non-douleur”.

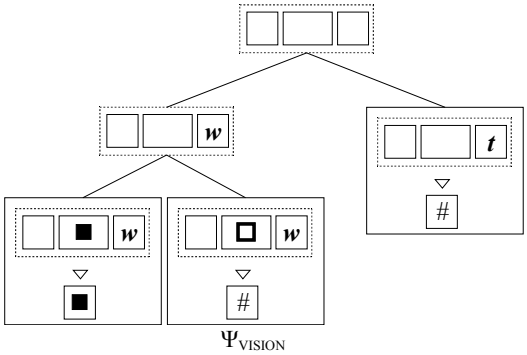


Figure 4.12: Arbre d'anticipation de la vision.

“Si vision du mur et marcher alors vision du mur”; “si vision du vide et marcher alors vision indéterminée”, et “si pivoter alors vision indéterminée”.

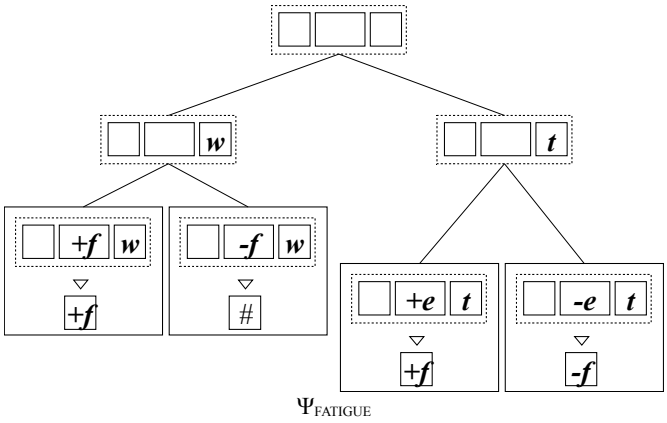


Figure 4.13: Arbre d'anticipation de la fatigue.

“Si fatigué et marcher alors fatigué”, “si non-fatigué et marcher alors fatigue indéterminée”; “si épuisé et pivoter alors fatigué”, et “si non-épuisé et pivoter alors non-fatigué”.

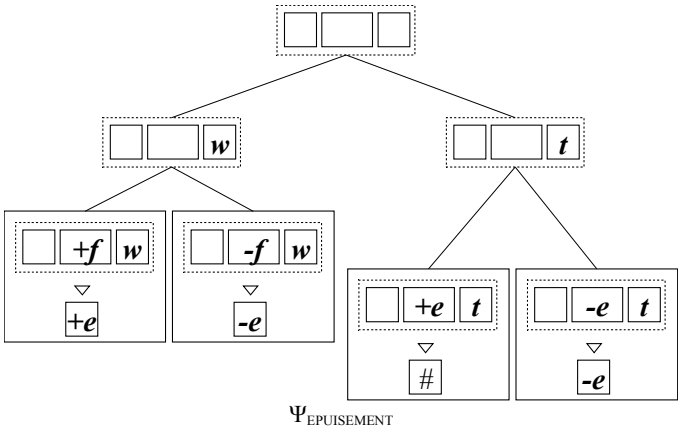


Figure 4.14: Arbre d'anticipation de l'épuisement.

“Si fatigué et marcher alors épuisé”; “si non-fatigué et marcher alors non-épuisé”; “si épuisé et pivoter alors épuisement indéterminé”; et “si non-épuisé et pivoter alors non-épuisé”.

(500) La figure 4.15 montre l'évolution du comportement présenté par un agent CALM pendant une exécution typique de l'expérience *wepp*, dans un scénario de taille 25 *x* 25, où 20% des cellules sont occupées par des obstacles. Les trois premiers graphiques

(a, b, c) montrent respectivement le plaisir, la douleur, et l'épuisement de l'agent. Les deux graphiques suivants (d, e) représentent les transformations cognitives que l'agent traverse avec le temps, relatives à la mémoire épisodique et aux arbres d'anticipation respectivement. L'axe *x* des graphiques compte 4000 pas d'exécution, en marquant un point tous les 100 cycles. Il faut noter que le comportement optimal est atteint après 2000 pas d'exécution, mais le modèle du monde est déjà stable à partir de 500 cycles, comme on peut voir dans l'avant-dernier graphique (f), qui décrit la moyenne des retours affectifs. Quelques événements négatifs se produisent même après que l'esprit soit bien stabilisé en raison du comportement exploratoire de l'agent, illustré dans le dernier graphique (g), qui montre les moments où les actions ont été choisies par curiosité.

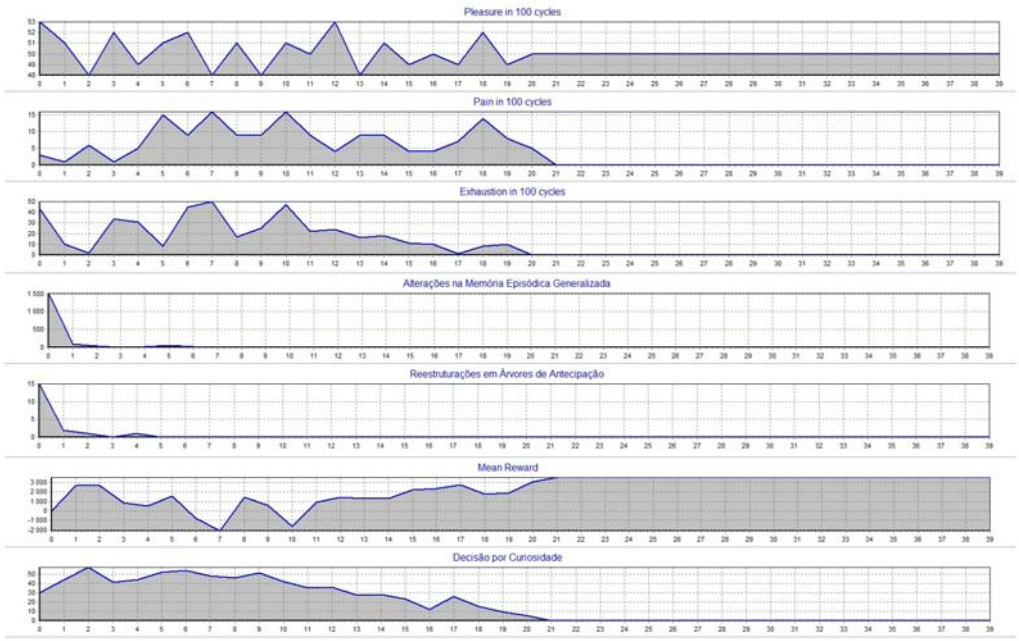


Figure 4.15: Résultats de la simulation (cas typique).

Chaque point sur l'axe x des graphiques représente le nombre d'événements tous les 100 pas d'exécution. À partir du graphique en haut on présente: a) le plaisir, b) la douleur, et c) l'épuisement; d) les changements dans la mémoire épisodique, et e) les transformations des arbres d'anticipation; f) la récompense moyenne, et g) les décisions exploratoires (prises par curiosité).

(501) Quand on compare les solutions présentées par l'algorithme Q-Learning et le mécanisme CALM, tous les deux implémentés en tant qu'agents de type CAES, on remarque que le mécanisme CALM converge beaucoup plus rapidement. Les figures 4.16, 4.17, et 4.18 montrent, respectivement aux plateaux de dimensions 5x5, 25x25 et 125x125, la comparaison entre CALM et Q-Learning avec l'architecture CAES (située),

ainsi que la performance du Q-Learning avec un agent non-situé classique (où chaque position du plateau est représentée par un état), et un agent de comportement aléatoire.

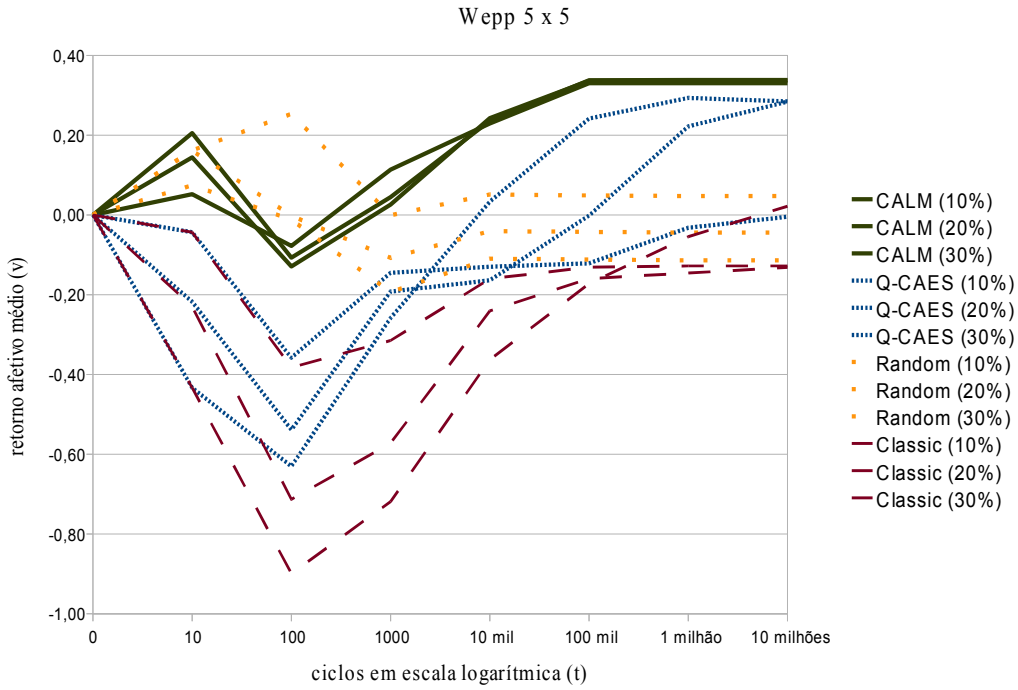


Figure 4.16: Wepp 5 x 5: Q-Learning x CALM.

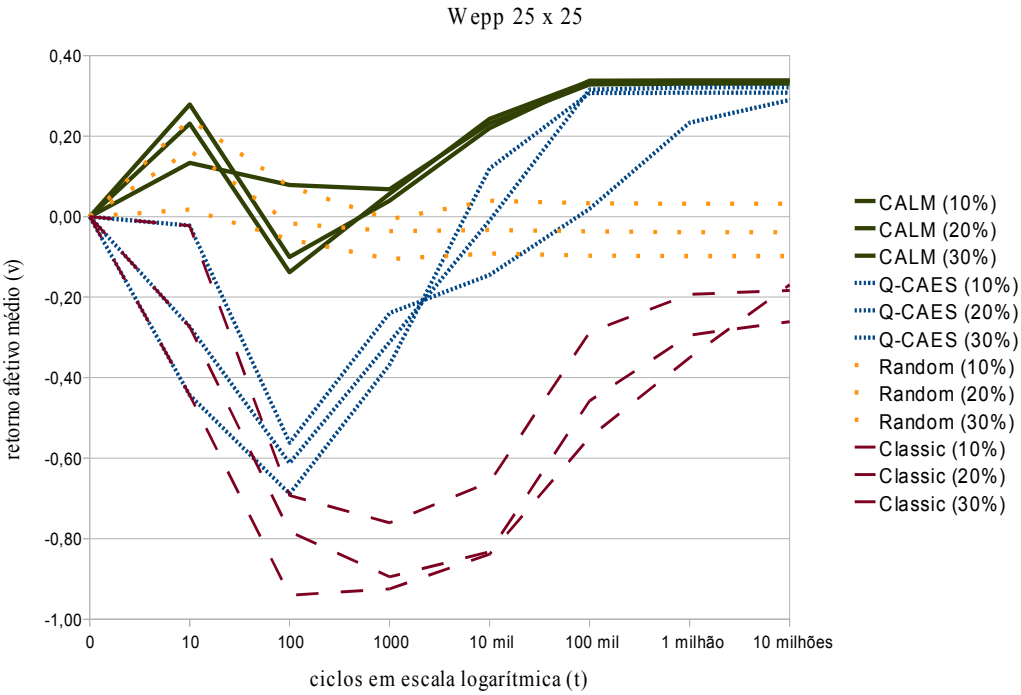


Figure 4.17: Wepp 25 x 25: Q-Learning x CALM.

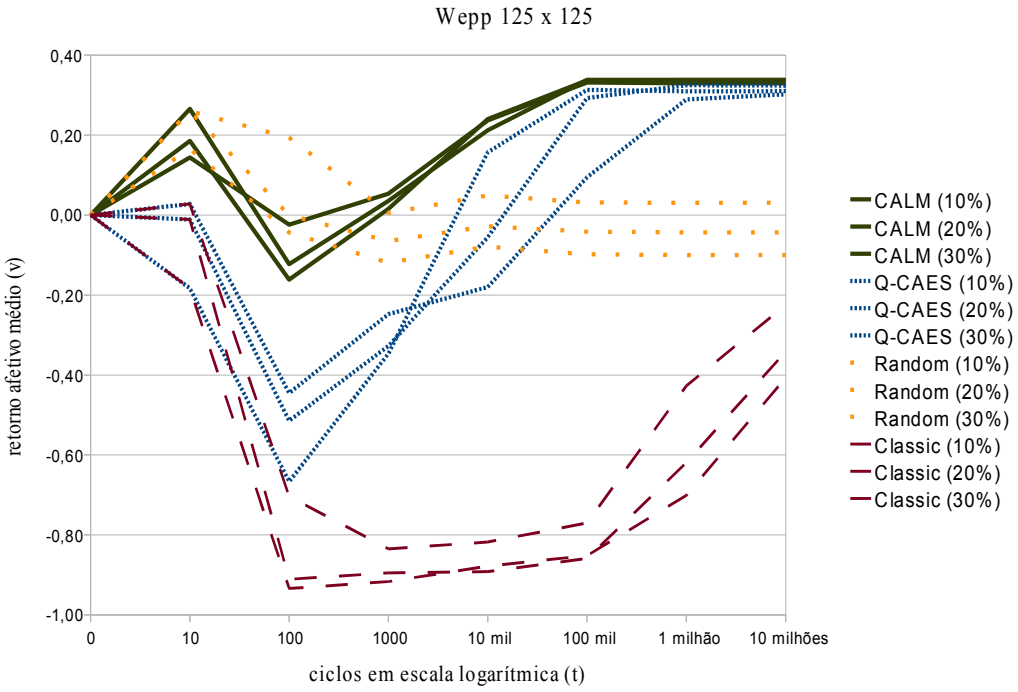


Figure 4.18: Wepp 125 x 125: Q-Learning x CALM.

4.1.2.3. Extensibilité

(502) Ensuite, afin de tester l'extensibilité du mécanisme, une autre série d'expériences a été réalisée, toujours en utilisant la structure du problème wepp. Le nombre de variables du problème a été augmenté, par l'insertion de propriétés servant seulement à bruiter le signal perceptif de l'agent. Ainsi le problème continue à être essentiellement le même, avec $|P|$ augmenté de propriétés non pertinentes et de dynamique aléatoire.

(503) Deux types d'agents ont été soumis à 4 scénarios différents, en répétant l'expérience 10 fois pour chacun. D'un côté, on a utilisé un agent implémentant le mécanisme CALM, et de l'autre un agent Q-Learning classique, qui représente les situations de façon plate, où l'espace est créé par la combinaison des valeurs des propriétés perceptives. Le premier scénario est le problème wepp original, sur un plateau de dimensions 25 x 25 avec 20% d'obstacles, et dans les scénarios suivants le nombre de variables $|P|$ est augmenté de 6 à 14, 22, puis 30, en insérant 8 nouvelles variables aléatoires à chaque fois. Dans la représentation classique, le dernier scénario correspond à un problème de plus d'un milliard d'états.

(504) La figure 4.19 montre la performance de CALM, où on peut voir que la moyenne du nombre de cycles nécessaires pour atteindre la stabilité de la solution

change peu avec l'introduction des variables aléatoires. Cela est du au fait que CALM anticipe rapidement leur comportement non-déterministe, ainsi que leur manque de pertinence pour la dynamique du système. À l'inverse, Q-Learning classique subit une importante perte de performance avec l'introduction des variables non-pertinentes.

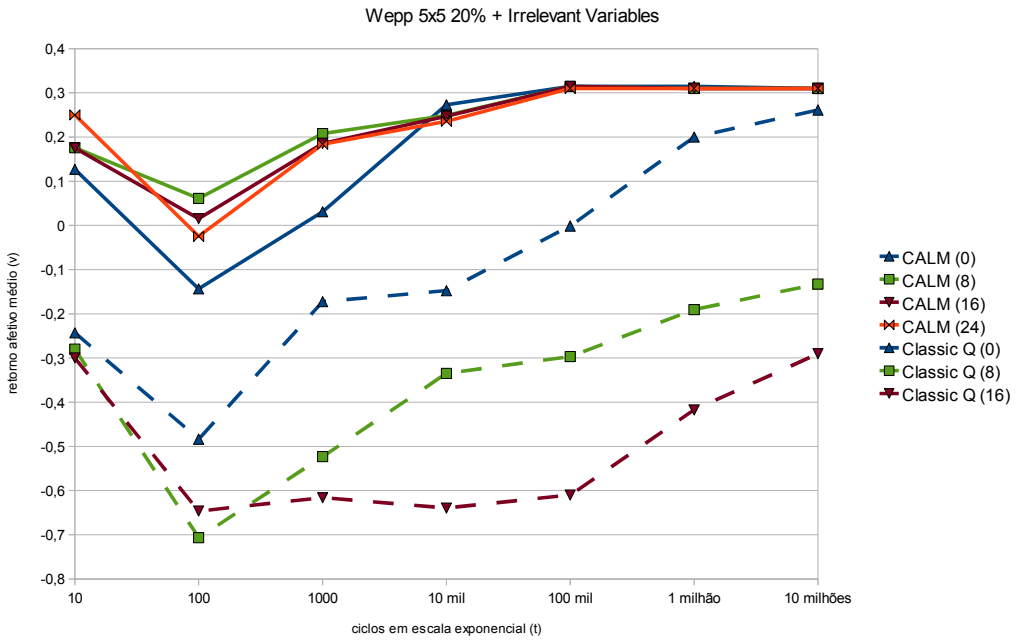


Figure 4.19: Analyse d'extensibilité.

Résultats lorsque on augmente le nombre de propriétés non-pertinentes de 0 à 8, 16, jusqu'à 24.

4.2. Problème Flip

(505) Le problème *wepp*, présenté précédemment, n'exploite pas la capacité qu'a le mécanisme CALM à traiter le cas d'observabilité partielle. Pour tester explicitement cette fonctionnalité, on a utilisé le problème *flip*, proposé dans la littérature par (SINGH et al., 2003) et récemment utilisé par (HOLMES; ISBELL, 2006). Il s'agit d'un agent vivant dans un monde composé de deux états $\{R, L\}$, qui lui sont cachés. Il a un actuateur qui permet de réaliser trois actions $\{l, r, u\}$, et il a la perception de deux valeurs possibles $\{0, 1\}$. L'agent observe '1' lors d'un changement de l'état sous-jacent, et '0' s'il reste le même. L'action u ne fait rien, l'action l déclenche une transition déterministe vers l'état L (à gauche), et l'action r de la même manière vers l'état R (à droite), selon illustré figure 4.20.

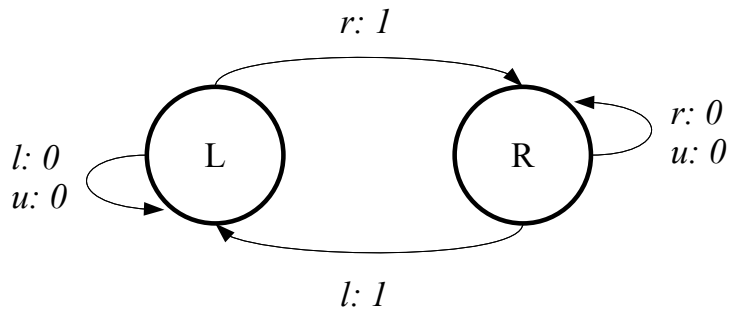


Figure 4.20: Problème *flip*, montré sous la forme d'une machine d'états.

(506) Le problème *flip* peut être décrit par un D-FPOMDP qui possède une seule variable perceptive $P_1 = \{0, 1\}$, en représentant la perception du changement de l'état, une seule variable de contrôle $C_1 = \{l, r, u\}$, et une seule variable non-observable $H_1 = \{R, L\}$. L'anticipation de p_1' et de h_1' peut se faire en fonction de h_1 et c_1, sous la forme $(H_1 \times C_1 \rightarrow P_1)$, et $(H_1 \times C_1 \rightarrow H_1)$.

4.2.1. Construction de la Solution par CALM

(507) Lorsqu'il est confronté au problème *flip*, le mécanisme CALM commence la construction de son modèle en découvrant les régularités qui sont basées seulement sur les propriétés observables (ensemble *P*) ou sur les actions prises par l'agent (ensemble *C*). Dans cette première étape, le modèle se stabilise avec une seule régularité, représentée par le schéma $[u \rightarrow 0]$. Dans la figure 4.21 on observe cette première stabilisation du modèle, quand CALM ne possède que l'arbre d'anticipation Ψ_{P1}, qui essaye de décrire la dynamique de la propriété P_1. À ce moment, l'arbre a 3 schémas, différenciés par la variable de contrôle C_1. Les schémas qui assimilent les actions "*r*" et "*l*", pendant cette première étape d'apprentissage, finissent par avoir des anticipations indéterminées, $[r \rightarrow \#]$ et $[l \rightarrow \#]$, parce qu'il n'est pas possible de prévoir l'observation suivante uniquement à partir de l'action et de l'observation précédentes.

(508) Quand le modèle se stabilise, c'est-à-dire quand les schémas ne se modifient plus pendant une longue période, la mémoire épisodique peut être nettoyé, en éliminant les tables qui ne gardent plus aucune information intéressante. Sur la figure 4.22 on montre le contenu de la mémoire épisodique du problème après l'élimination de ses tables inutiles.

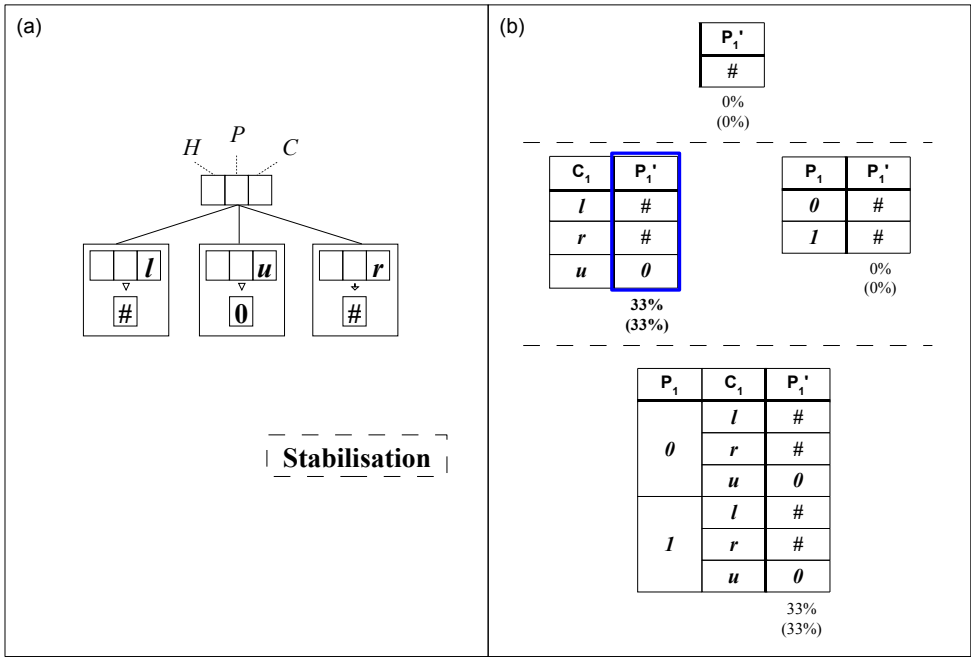


Figure 4.21: Début de la construction de la solution CALM pour le problème *flip*.

Dans (a), l'arbre d'anticipation de la variable P_1, qui a la variable C_1 comme différenciateur. Dans (b), le contenu de la mémoire épisodique à la fin de cette première étape, quand le modèle se stabilise sans l'utilisation des éléments synthétiques. La table en évidence est celle que donne origine à l'arbre.

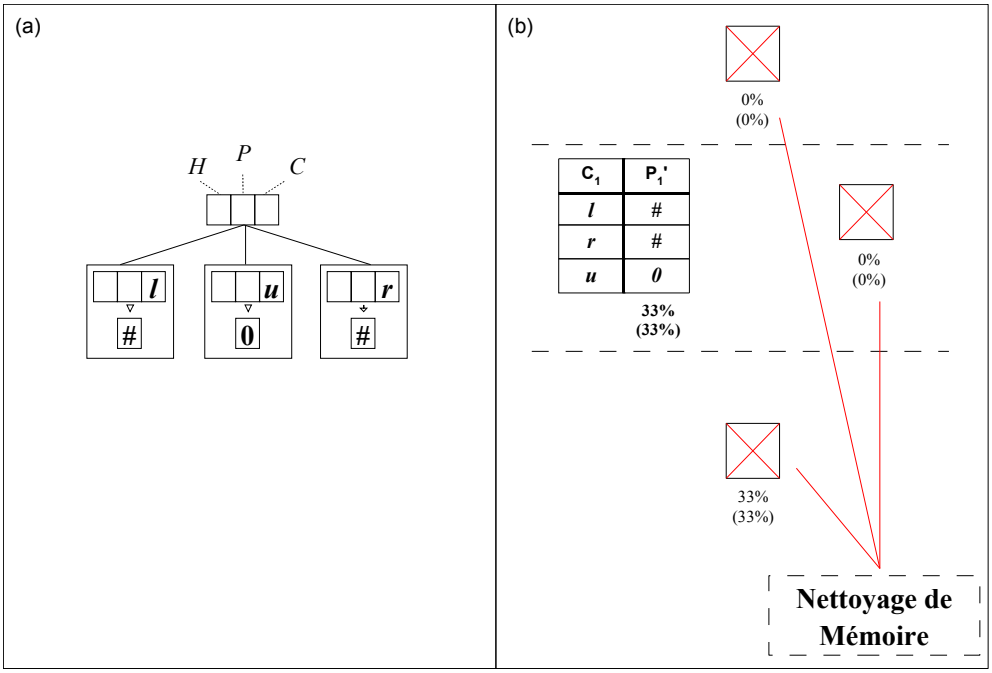


Figure 4.22: Première stabilisation de la solution, encore sans compter des éléments synthétiques.

Dans (a), l'arbre d'anticipation de la variable P_1. Dans (b), après la stabilisation de la connaissance, les tables inutiles de la mémoire épisodique sont éliminées, soit parce qu'elles contiennent moins d'information que la table en évidence, soit parce qu'elles contiennent la même information en occupant plus de place.

(509) Sur la figure 4.23, à la suite de la première stabilisation, CALM crée un élément synthétique, $H_1 = \{\clubsuit, \blacklozenge\}$ envisageant d'améliorer l'anticipation de P_1, et en supposant, pour cela, l'existence d'une propriété non-observable capable de préciser les anticipations indéterminées dans l'arbre d'anticipation actuel. L'élément synthétique H_1 est utilisé pour différencier le schéma [$r \rightarrow$ #], qui désormais sera spécialisé comme [$\clubsuit$ $r \rightarrow 1$] et [$\blacklozenge$ $r \rightarrow 0$]. Ces valeurs sont placées dans la table correspondante de la mémoire épisodique, dont les cellules respectives sont appelées « situations-ancre », parce qu'elles sont les situations qui déterminent, à posteriori, l'état de la condition non-observable.

(510) Dans le problème, faire r au temps t et observer 1 à $t+1$ signifie que l'agent était en $\clubsuit$ à t, et de la même façon, faire l au temps t et observer 0 à $t+1$ signifie qu'il était en $\blacklozenge$ à t. L'insertion de ce nouveau élément synthétique, H_1, dans le modèle du monde de l'agent provoque deux modifications principales. Premièrement, une nouvelle table est insérée dans la mémoire épisodique Ш_{P1}, qui est la mémoire responsable pour les observations pour l'anticipation de P_1. Cette nouvelle table est une copie de celle qui jusqu'à présent donnait origine à l'arbre Ψ_{P1}, c'est-à-dire la table qui avait le plus grand degré de déterminisme observé, agrandi avec le nouveau élément synthétique. Cette nouvelle table hérite les valeurs déterminées de la table déjà stable, et donc, dans le problème, la cellule de mémoire [$u \rightarrow 0$] définira [$\clubsuit$ $u \rightarrow 0$] et [$\blacklozenge$ $u \rightarrow 0$]. Deuxièmement, un nouveau arbre d'anticipation, Ψ_{H1}, est créé pour réaliser l'anticipation de H_1, et elle est initialisée avec un schéma unique. Également, toute un nouvel ensemble de tables est créé, composant la mémoire épisodique Ш_{H1}.

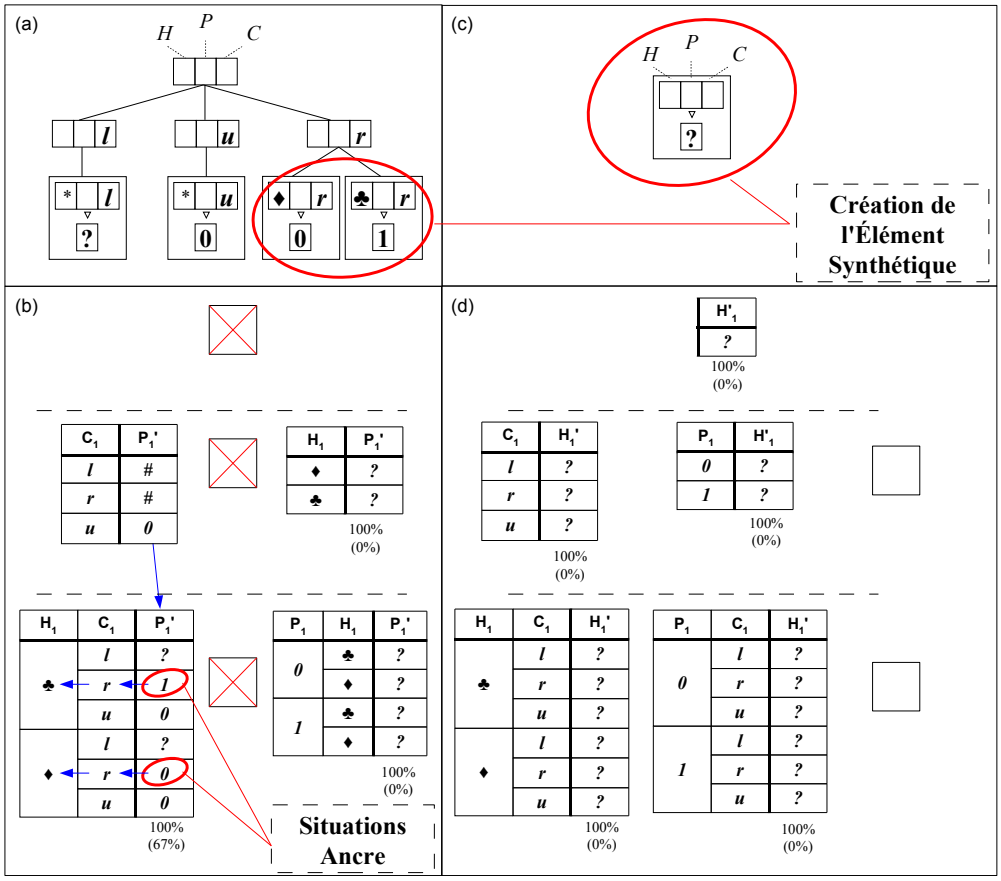


Figure 4.23: Création de l'élément synthétique.

Une différenciation abstraite est créée à partir du schéma “r” dans (a), et un nouvel arbre dans (c) pour essayer d'anticiper la dynamique de cette propriété non-observable. Les mémoires épisodiques liées à l'anticipation de P_1 sont montrées dans (b), où une nouvelle table est insérée, avec les situations-ancre, qui définissent l'état de la condition non-observable. Les mémoires épisodiques liées à l'anticipation de H_1 sont montrées dans (d).

(511) Dans la figure 4.24 est montré l'état du modèle au moment d'une deuxième stabilisation, lorsqu'il utilise déjà l'élément synthétique. Le schéma unique de l'arbre Ψ_{HI} est différencié par la variable C_1, représentant l'information contenue dans la table de mémoire mis en évidence. Dans cette phase, l'anticipation de l'élément synthétique H_1 est encore partielle parce que les mémoires épisodiques reliées à son anticipation sont enregistrées seulement quand sa valeur est confirmée par l'occurrence d'une situation ancre en $\mathfrak{U}_{P1}$.

(512) Par exemple, quand l'agent exécute une séquence d'actions t_1 : [r] et t_2 : [r], selon la définition du problème, l'observation à t_2 sera toujours [0], et une fois que t_2 : [$r \rightarrow 0$] est justement une des situations-ancre, cela permet de déduire que l'élément synthétique à t_2 est constamment [♦], c'est-à-dire que t_2 : [♦ $r \rightarrow 0$]. À partir de cette information il

est donc possible d'inférer que [$r \rightarrow$ ♦], et d'ajouter cette régularité au modèle. On a le cas équivalent t_1 : [l] et t_2 : [r], qui finira constamment avec l'observation de la situation ancre t_2 : [♣ $r \rightarrow$ 1], qui conduit à la découverte de la régularité [$l \rightarrow$ ♣].

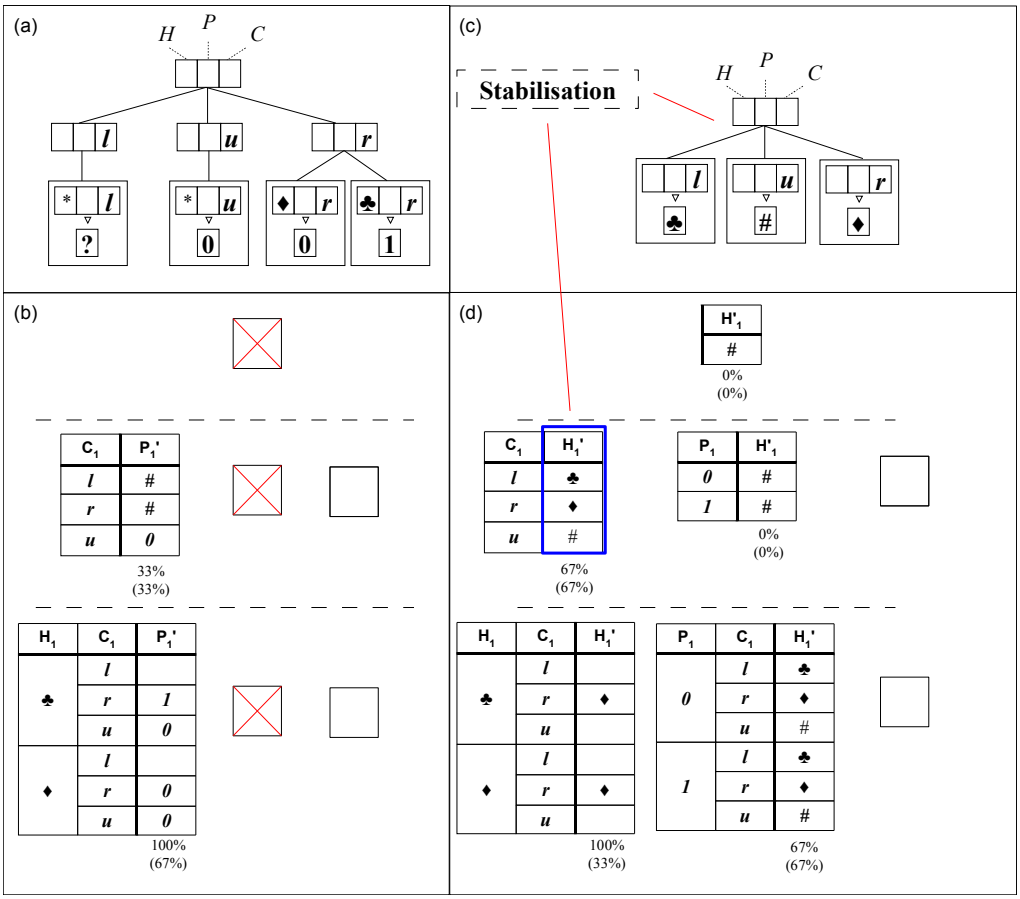


Figure 4.24: Deuxième stabilisation de la solution.

Sur ce point, l'anticipation de l'élément synthétique commence a être construite.

(513) De façon similaire, les tables de la mémoire dans lesquelles l'élément synthétique apparait en tant que condition ne peuvent être actualisées que si l'agent vit une situation ancre. Par exemple, supposons une occurrence de t_1 : [1 $r \rightarrow$ 1] et t_2 : [1 $r \rightarrow$ 0], d'où on déduit que t_1 : [♣ 1 $r \rightarrow$ ♦ 1] et t_2 : [♦ 1 $r \rightarrow$ 0]. La situation t_1 permettra l'actualisation des cellules [♣ 1 $\rightarrow$ ♦] et [♣ $\rightarrow$ ♦] dans ses tables respectives.

(514) L'occurrence ultérieure de t_3 : [♣ 1 $l \rightarrow$ ♣ 0] et t_4 : [♣ 0 $r \rightarrow$ 1] déclenche l'indétermination des observations des cellules, [♣ 1 $\rightarrow$ #] et [♣ $\rightarrow$ #]. Des situations similaires conduiront les autres cellules des tables à [♣ 0 $\rightarrow$ #], [♦ 0 $\rightarrow$ #], [♦ 1 $\rightarrow$ #], et [♦ $\rightarrow$ #], en rendant ces tables inutiles et, par conséquence, éliminables, comme montré figure 4.25. À ce moment, certaines régularités observées dans des tables qui ne

possèdent pas H_1 comme condition peuvent être transmises aux autres tables qui ont l'élément synthétique dans la condition, ce qui est aussi illustré sur la figure.

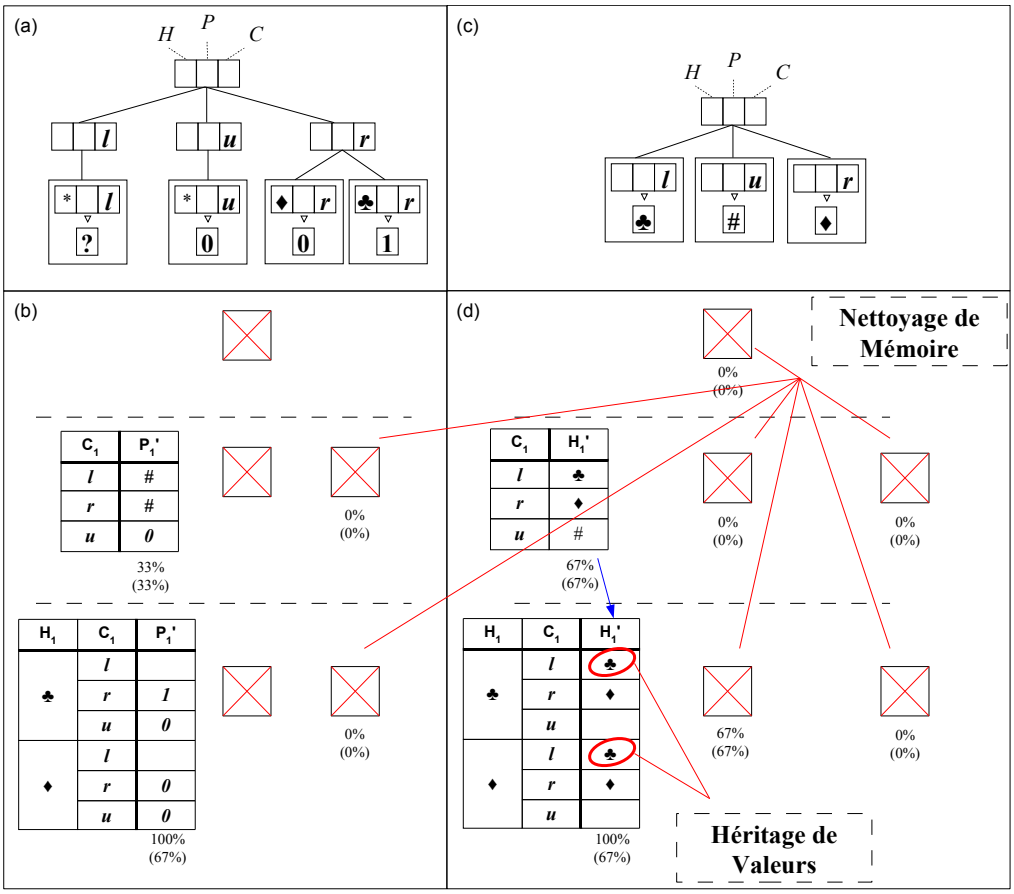


Figure 4.25: La nouvelle table de la mémoire épisodique hérite les valeurs de sa correspondante.

Dans ce processus il y a la transmission des valeurs définis des tables qui n'ont pas l'élément synthétique comme condition, aux tables qui en ont.

(515) La figure 4.26 illustre les dernières modifications dans l'arbre Ψ_{HI}, qui se stabilise en définissant l'anticipation pour le schéma lié à l'action $[u]$, sous la forme $[u \rightarrow \approx]$, ce qui signifie que la valeur de H_1 n'est pas modifiée par l'exécution de l'action $[u]$, donc $[\blacklozenge\ u \rightarrow \blacklozenge]$ et $[\clubsuit\ u \rightarrow \clubsuit]$.

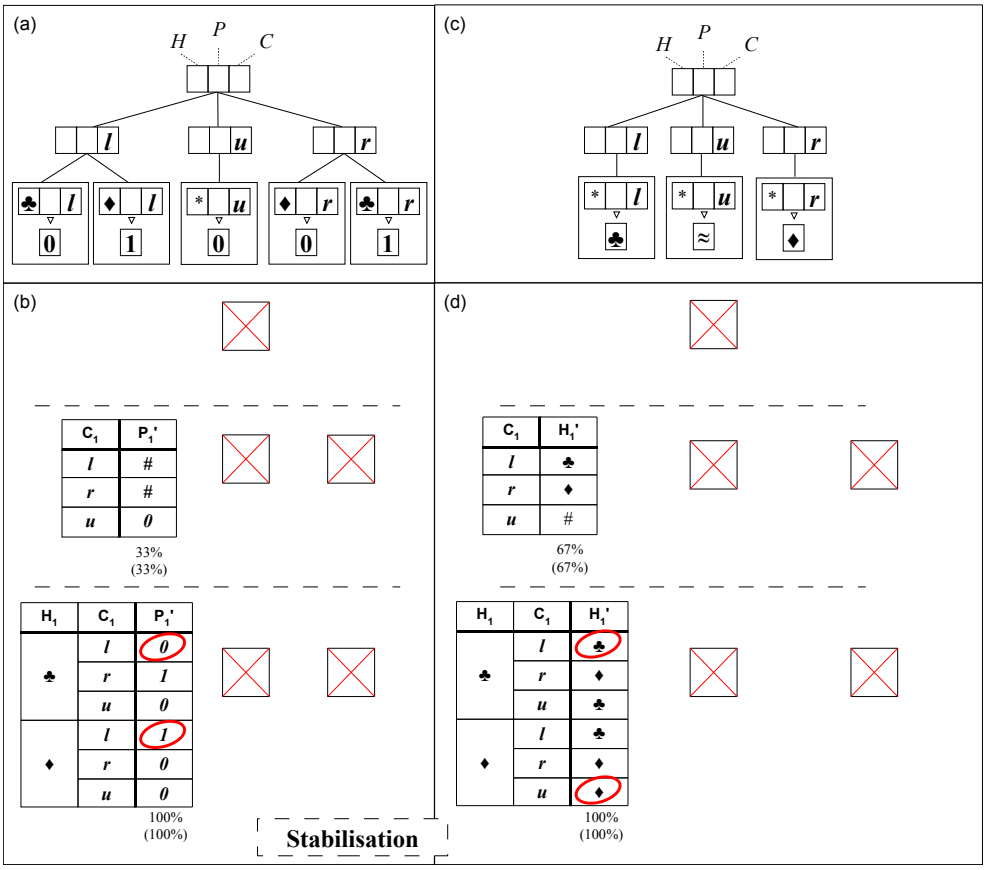


Figure 4.26: Dernière stabilisation, avec la solution finale.

4.2.2. Comparaison des Solutions

(516) CALM est capable de résoudre le problème *flip* à travers la création d'un nouvel élément synthétique qui représente les états sous-jacents gauche (♣) et droite (♦). La figure 4.27 montre les arbres d'anticipation finaux construits par le mécanisme. Le problème *flip*, même s'il est apparemment simple, n'est pas trivial pour plusieurs algorithmes d'apprentissage (SINGH et al., 2003) car le cycle créé par l'action *u* peut se répéter indéfiniment sans laisser d'évidences historiques observables.

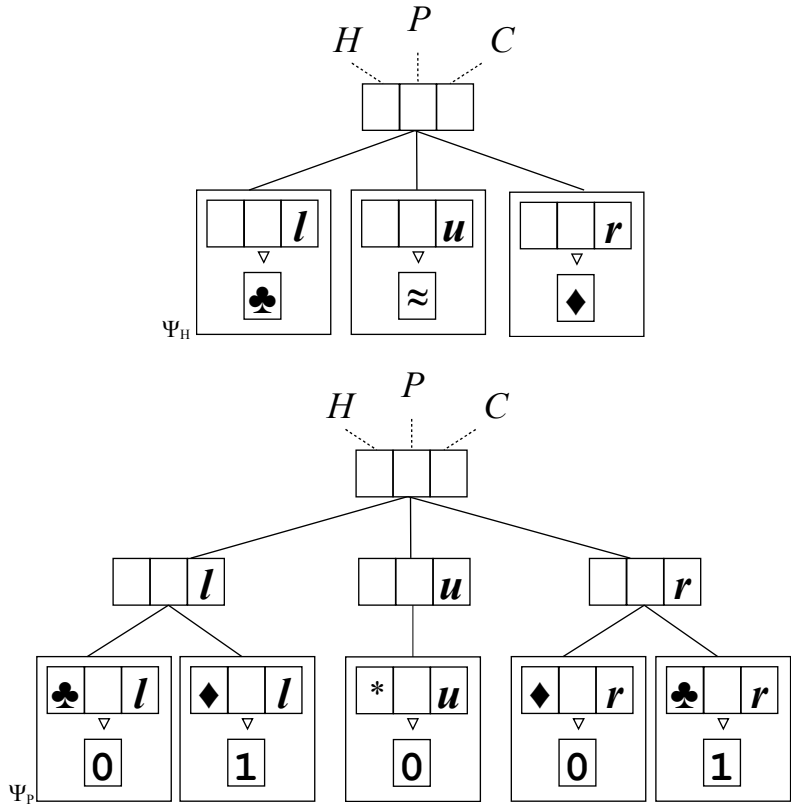


Figure 4.27: Arbres d'anticipation construits par CALM pour l'expérience *flip*.

(517) Afin d'adapter le problème *flip* au mécanisme CALM, les perceptions ont reçu des valeurs affectives, de manière que l'agent reçoive une valeur positive quand il observe 1, et une valeur négative quand il observe 0. CALM calcule l'utilité des schémas dans l'arbre de délibération, soit pour l'exploration, soit pour l'exploitation, ce qui rend possible la découverte du modèle complet du problème. L'arbre de délibération est montré figure 4.28.

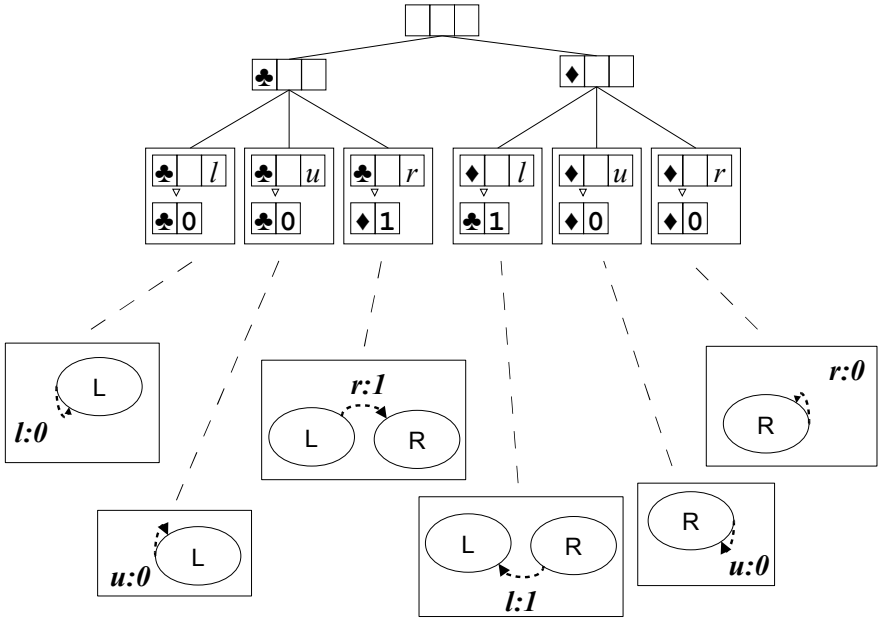


Figure 4.28: Arbre de délibération construit par CALM pour le problème *flip*.

(518) Holmes et Isbell (2006) on proposé avec succès un algorithme pour résoudre le problème *flip*, à travers l'apprentissage d'arbres de préfixes (PST). Leur méthode, même si elle est efficace, présente deux désavantages par rapport à CALM. Premièrement, un PST ne constitue pas un modèle du monde, c'est-à-dire qu'il n'essaye pas de représenter les variables non-observables de l'environnement. La représentation des situations dans le PST est directement basée sur les séquences historiques des observations et des actions précédentes. Deuxièmement, la méthode n'est pas incrémentale, et il est nécessaire de présenter tout l'ensemble des observations d'un seul coup, quand le processus commence. Le PST qui résoudre le problème *flip* est montré figure 4.29.

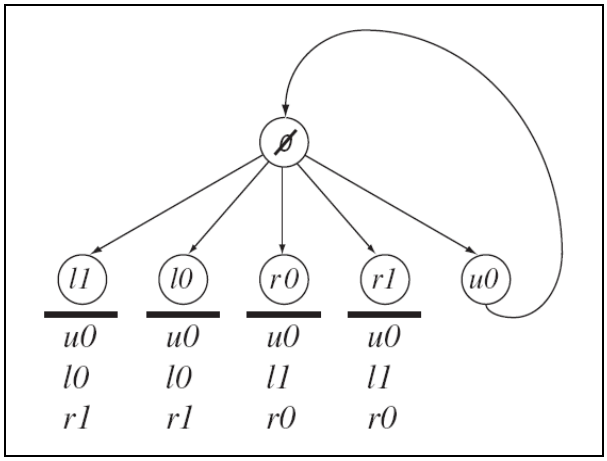


Figure 4.29: PST pour le problème *flip*.

Le nœud racine est un nœud vide, et chaque branche représente une séquence d'interaction possible. Au dessous de ces nœuds terminaux se présentent les anticipations. Par exemple, la première branche, à gauche, indique que « faire l'action 'l' et observer '1' », permet d'anticiper que « en exécutant 'u' juste de suite, on observe '0' », « en exécutant 'l', on observe '0' aussi », et « en exécutant 'r', on observe '1' ». La branche à droite montre le cycle créé pae l'action 'u'.

5. CONCLUSIONS

(519) L'intelligence artificielle est probablement la plus grande promesse de révolution technologique de ce siècle. Peu à peu des systèmes intelligents, qu'ils soient robotiques ou virtuels, surgissent et s'insinuent dans divers domaines de la vie, soit dans le quotidien domestique, soit dans l'industrie, soit sur internet, etc. Pour cette raison la recherche fondamentale gagne en importance, celle qui vise à repenser la discipline et à comprendre les mécanismes généraux de l'intelligence, particulièrement par rapport à l'apprentissage autonome, en définissant de nouvelles stratégies et de nouvelles méthodes qui pourront ensuite être utilisées pour développer de nouvelles technologies, et pour résoudre les problèmes qui, jusqu'à présent, ne peuvent pas être résolus par les ordinateurs intelligemment, sans une tutelle très étroite des opérateurs et des programmeurs.

(520) Une forme de représentation actuellement bien utilisée pour des problèmes de décision est le *Processus de Décision Markovien* (MDP), en particulier dans la version factorisée (FMDP) et dans la version partiellement observable (POMDP). Bien que plusieurs études ont proposé des mécanismes pour le calcul de la politique d'actions à partir d'un FMDP donné, ou à partir d'un POMDP donné, peu d'attention a été portée sur le problème de l'apprentissage de la structure du processus markovien à travers l'observation. Il s'agit, pourtant, d'un problème fondamental, car un agent autonome

inséré dans un environnement inconnu a besoin de découvrir la structure de cet environnement à travers ses observations et ses interactions.

(521) Dans ce contexte, la réalisation de cette thèse a permis de produire les résultats suivants: (1) la proposition du mécanisme d'apprentissage CALM (*Constructivist Anticipatory Learning Mechanism*), conçu précisément pour la construction de modèles du monde à travers l'interaction; et (2) la proposition de l'architecture CAES (*Coupled Agent-Environment System*), qui structure le rapport qu'un agent autonome et situé entretient avec son milieu.

5.1. Considérations sur le Mécanisme CALM

(522) La principale contribution de cette thèse est la proposition d'un mécanisme général d'apprentissage basé sur la théorie constructiviste, CALM, présenté dans la section 3.2, et conçu pour résoudre des problèmes importants d'apprentissage automatique, tels que ceux définis dans la section 3.1.

(523) CALM est un mécanisme original pour la construction de modèles du monde par des agents artificiels à partir de l'interaction avec l'environnement, en utilisant un processus qui se déroule de façon autonome. CALM a été conçu pour être appliqué à des problèmes d'apprentissage actif et continuel associés à la décision et à la planification. Ce sont des problèmes dans lesquels l'agent doit construire son modèle du monde d'une manière incrémentale, en étant en interaction ininterrompue avec son environnement, et tout en définissant en même temps une politique d'action qui optimise son comportement.

(524) Le modèle du monde que CALM construit décrit les régularités déterministes de l'environnement où l'agent est situé, même s'il y a des phénomènes non-déterministes dans cet univers. Comme cela a été argumenté dans le chapitre 3, dans des univers bien structurés, même ceux qui sont complexes comme le monde réel, la majorité des phénomènes se présentent comme des transformations régulières et déterministes si les conditions causales sont bien identifiées.

(525) CALM est également capable de découvrir des régularités, même quand elles sont dépendantes des propriétés non-observables de l'environnement. La stratégie est d'augmenter le vocabulaire de représentation de l'agent en ajoutant des éléments

synthétiques qui peuvent être associés à des propriétés cachées, des conditions séquentielles, ou des conditions abstraites présentes dans les situations vécues par l'agent.

(526) Le mécanisme CALM permet à l'agent de construire un modèle du monde sous la forme d'arbres d'anticipation, ce qui a nécessité de traiter le problème de la fragmentation de la connaissance. Dans des problèmes complexes, l'augmentation de la profondeur des arbres crée des branches de telle façon que le nombre de nœuds peut croître exponentiellement. Ainsi, une partie importante du défi posé par ce travail a été d'assurer que les arbres aient une taille contrôlée, en maintenant le problème à des niveaux de calcul acceptables. Le principal moyen d'y réussir est de sélectionner les variables utilisées pour composer les arbres, en estimant la pertinence de chaque variable pour chaque transformation modélisée.

(527) Dans CALM, la sélection des propriétés pertinentes pendant la construction de chaque arbre d'anticipation est traitée grâce à l'analyse de la mémoire épisodique généralisée associée à l'arbre. Cette structure nécessite une grande quantité d'espace de stockage, mais sa croissance est contrôlée par le mécanisme. La mémoire épisodique garde un souvenir général des situations vécues par l'agent, liées à une transformation spécifique dont l'anticipation est en train de se construire. Sa taille est gérable précisément parce qu'elle est généralisée, en pouvant observer simultanément un nombre limité de conditions.

(528) L'efficacité du mécanisme est basée sur l'hypothèse que l'environnement est bien structuré, c'est-à-dire que le nombre de propriétés pertinentes pour décrire chacune des transformations régulières doit être, au plus, d'ordre logarithmique par rapport au nombre total de propriétés du problème. Si cela est garanti, alors l'environnement traité par CALM peut être partiellement observable et partiellement déterministe.

(529) Un avantage de la représentation utilisée par CALM est que les arbres sont construits chacun de façon indépendante, et qu'ils peuvent donc être traités en parallèle. La parallélisation de l'apprentissage est importante car elle permet de casser la complexité de calcul du problème.

(530) Le mécanisme a été développé pour faire face à des objectifs à long terme, dans des problèmes de décision séquentielle, où l'agent doit construire une politique

d'actions qui vise à maximiser l'efficacité de son comportement. Il a donc été mis en œuvre une méthode de calcul de l'utilité des actions, basée sur les équations de Bellman, qui utilise le modèle du monde construit par l'agent, pour ensuite construire une politique d'actions sous la forme d'arbres de délibération.

(531) Des expériences menées sur les problèmes *wepp* et *flip* (dans le chapitre 4) ont montré qu'un agent artificiel qui implémente le mécanisme CALM est capable de converger de manière satisfaisante vers les solutions attendues. Ces résultats confirment la capacité du mécanisme à découvrir des régularités dans l'interaction avec l'environnement, en les représentant dans ses arbres d'anticipation, et à utiliser ces connaissances pour construire une bonne politique d'actions, rendant ainsi l'agent mieux adapté à son monde.

5.2. Considérations sur l'Architecture CAES

(532) Afin de mettre en œuvre un modèle constructiviste de l'intelligence artificielle, il a été nécessaire de réaliser une réflexion plus fondamentale sur la conception des agents autonomes. Ainsi, l'autre contribution de cette thèse, présentée dans le chapitre 2, concerne le débat théorique sur l'idée d'un agent autonome, et les discussions autour des notions de situativité, d'incarnation, et de motivation intrinsèque, qui a conduit à proposer l'architecture CAES, utilisée pour la définition des expériences.

(533) L'architecture CAES définit l'agent et l'environnement comme étant deux systèmes partiellement ouverts et couplés l'un à l'autre, où l'agent est situé et doit coordonner son comportement avec la dynamique du système global. Dans CAES il y a trois types fondamentaux d'interactions: à l'extérieur, l'interaction entre l'agent et l'environnement; puis, dans l'agent, l'interaction entre son corps et son esprit; et enfin, dans l'esprit, l'interaction entre le système cognitif et le système régulatif.

(534) L'architecture CAES définit un modèle formel, qui est innovant en ce qu'il intègre des concepts qui proviennent de deux paradigmes distincts: l'IA Affective et l'IA Située. L'agent gagne un corps, qui est un univers différencié à la fois du monde (extérieur) et de l'esprit (intérieur). L'esprit, donc, devient le système qui interagit avec le corps, et par lui avec l'extérieur.

(535) Dans l'esprit de l'agent, l'architecture CAES établit, d'une part, le fonctionnement d'un système régulatif, qui gère les comportements émotionnels et réactifs, et de l'autre, le fonctionnement d'un système cognitif, capable d'apprendre un modèle du monde et une politique d'actions. Le système régulatif est également responsable des signaux affectifs. Dans CAES, les objectifs de l'agent ne sont pas fournis sous la forme d'états-cibles ou de récompenses externes. C'est le système affectif qui évalue d'une façon factorisée la transformation des signaux perceptifs de l'agent, en le motivant à agir par la qualité agréable ou désagréable des sensations.

5.3. Propriétés Non-Observables et Abstraction

(536) La majorité des phénomènes intéressants de notre monde ne peut être exprimée que par des concepts de haut niveau. Donc, un des grands défis qui se pose pour l'IA est de trouver des moyens pour surmonter les limites de la perception sensorielle directe, et d'atteindre des modes de pensée et de représentation plus abstraits.

(537) Les algorithmes de classification traditionnels sont capables de généraliser des ensembles de situations, et ainsi réalisent un premier éloignement de la perception sensorielle directe. Toutefois, la description des classes se fait à travers des combinaisons de propriétés perceptives, ce qui implique que, d'une certaine manière, les classifications restent au niveau sensoriel, sans vraiment constituer des éléments radicalement nouveaux de représentation.

(538) CALM implémente une méthode pour la découverte des propriétés non-observables de l'environnement, basée sur l'analyse des informations historiques sur l'expérience sensorielle de l'agent. Supposer l'existence de ces variables cachées est une manière de discerner des situations sensoriellement ambigües dans des environnements partiellement observables. Cette découverte passe par la création d'éléments synthétiques de représentation, qui peuvent être considérés comme étant un pas au-delà du niveau sensoriel. Le mécanisme a la capacité d'apprendre à anticiper les valeurs de ces propriétés cachées, qui peuvent alors être utilisées dans la construction de la politique d'actions au même titre que les éléments sensoriels.

(539) Dans ce sens, CALM est une contribution car il offre une solution différente pour ce problème, qui est tout à fait pertinente dans le domaine de l'apprentissage

automatique. La stratégie utilisée par CALM est inspiré par l'idée proposée dans (DRESCHER, 1991), qui prévoyait déjà l'utilisation d'éléments synthétiques pour représenter les propriétés non-observables de l'environnement. Toutefois, la méthode d'apprentissage utilisées par Drescher est fondée une approche statistique qui présente des coûts de calcul prohibitifs (CHAPUT, 2004).

(540) Parmi les travaux récents, la méthode utilisée par CALM pour différencier les états ambigus peut être comparée à celle présentée par (HOLMES; ISBELL, 2006), qui effectue des anticipations dans des environnements représentés en tant que D-POMDPs. Cependant, leur solution est différente de CALM, d'abord parce qu'elle est basée sur la représentation plate des états, ensuite parce qu'elle fonctionne en mode hors-ligne, et il exige que l'environnement sous-jacent, même s'il est partiellement observable, soit complètement déterministe.

(541) CALM restreint le problème en ne cherchant à apprendre que les régularités déterministes de l'environnement, mais il n'exige pas que l'environnement soit complètement déterministe, c'est-à-dire que le mécanisme est applicable à des univers partiellement déterministes. Cela lui permet de mettre en œuvre une méthode d'apprentissage directement inductive, constituant une approche qui, à la connaissance des auteurs, n'a été utilisée dans aucun autre travail.

(542) Fait intéressant, une propriété non-observable peut représenter divers types de conditions abstraites d'un environnement. Trois cas permettent de bien illustrer ce point: (a) les environnements répartis en sous-environnements; (b) les environnements non-stationnaires discrets; et (c) les effets produits par des actions enchaînées.

(543) Le premier cas (a) se produit quand un agent est inséré dans un environnement qui a une fonction de transformation qui varie en fonction d'une propriété cachée. Par exemple, si l'environnement est divisé en sous-environnements d'apparences perceptives similaires pour l'agent, mais où les régularités ne sont pas équivalentes. Dans ce cas, des éléments synthétiques peuvent être utilisés pour identifier le sous-environnement dans lequel l'agent est à chaque instant. De même, le second cas (b) se produit lorsque l'environnement est non-stationnaire de forme discrète, c'est-à-dire, quand les régularités de l'environnement subissent des changements périodiques fixes. Dans ce cas, les

périodes durant lesquelles l'environnement présente des régularités spécifiques peuvent aussi être identifiées par des éléments synthétiques.

(544) Le cas de l'enchaînement d'actions (c) concerne les situations où l'agent doit effectuer une séquence spécifique d'étapes pour voir se produire un résultat, mais où, toutefois, les étapes intermédiaires n'ont aucun effet apparent, le résultat ne venant qu'après l'exécution de toute la séquence. Les éléments synthétiques peuvent alors représenter, dans l'esprit de l'agent, l'exécution des étapes intermédiaires.

(545) Bien que la construction d'éléments synthétiques ne constitue pas une forme de pensée abstraite, proprement dit, puisqu'une fois créés, les éléments synthétiques sont ensuite traités au même niveau que les éléments sensoriels, on peut quand même dire que ce processus est légitimement une forme d'invention de concepts, car l'agent a construit des éléments de représentation nouveaux qui permettent de désigner des entités différentes de tout ce qui peut être représenté à partir des perceptions directes.

(546) Ainsi, la possibilité de traiter des propriétés non-observables représente un pas en avant dans le chemin entre la perception simple et directe, vers des formes plus abstraites de compréhension du monde.

5.4. Limitations et Travaux Futurs

(547) La fin de cette thèse, comme c'est souvent le cas pour de nombreuses recherches de ce genre, a laissée plus d'hypothèses ouvertes, plus de questions sans réponses, plus d'idées et de promesses, que de conclusions pertinentes ou de résultats remarquables. Néanmoins, on peut penser que le travail réalisé constitue un pas vers la bonne direction, même s'il s'agit d'un pas très modeste, pour l'avenir de la recherche en intelligence artificielle et en apprentissage automatique.

(548) Une des principales limitations de l'approche utilisée dans cette thèse est la discrétisation des représentations, soit des signaux reçus et transmis par l'agent, soit du temps lui-même, qui a été considéré en tant que succession de cycles. Plusieurs problèmes du monde réel ne peuvent être correctement abordés qu'au travers de représentations continues. L'architecture CAES est décrite de façon suffisamment générale, pour qu'elle soit applicable indifféremment à des représentations discrètes ou à

des représentations continues. Par contre le mécanisme d'apprentissage CALM marche en étant étroitement ancré à la discrétisation des problèmes.

(549) Ainsi, le besoin de discrétisation est une première grande limitation de CALM, surtout quand l'idée est de créer un agent qui puisse affronter des problèmes réels au niveau sensorimoteur. Cependant, cela ne veux pas dire que le mécanisme ne peut pas être adapté à des univers continus, et ce défi constitue une possibilité pour un travail futur.

(550) Il est aussi nécessaire d'avouer que les problèmes utilisés dans la thèse, même s'ils sont adéquats pour démontrer les caractéristiques désirées de l'architecture CAES et du mécanisme CALM, restent encore très simples. Il est nécessaire de réaliser des expériences plus élaborées pour montrer que CALM est capable de découvrir des propriétés non-observables et pertinentes de l'environnement, et de les utiliser en tant qu'éléments synthétiques dans son modèle du monde, lorsqu'il affronte des problèmes plus complexes.

(551) Un autre grand terrain, explicitement laissé sans exploration par la nécessité de rendre possible la réalisation de la thèse, est le problème du non-déterminisme. Bien que, à notre avis, la définition et l'exploration des problèmes partiellement déterministes s'est révélée une contribution intéressante du travail, le traitement des situations qui sont dans la pratique mieux représentées par des modèles stochastiques est une autre possibilité de travail futur.

(552) Plus techniquement, certains choix réalisés dans les méthodes qui constituent le mécanisme CALM ont besoin de plus d'attention, car il peut y avoir des moyens d'optimiser les algorithmes, en particulier en ce qui concerne la gestion de la mémoire épisodique et les arbres d'anticipation et de délibération.

(553) Enfin, plusieurs choix faits dans la définition des modèles ont été davantage motivés par l'aspect pratique que par conviction philosophique. L'achèvement d'un ouvrage de ce type, même avec des ambitions d'IA générale, requiert tôt ou tard un certain nombre de simplifications pour parvenir à un résultat plus concret. Le système de motivation du modèle, même si on a fait un effort pour construire un agent intrinsèquement motivé, et cohérent dans la perspective d'une IA incarnée, est encore nettement trop utilitariste, et ignore certaines évidences établies par les théories

constructiviste et interactiviste de l'intelligence, que la motivation est aussi liée à l'activité elle-même. Boire de l'eau parce qu'on a soif est un genre de comportement qui peut être facilement ancré à une explication utilitariste biologique, mais d'autres comportements comme jouer aux dames, écouter de la musique, faire une thèse, ne seraient guère expliqués par une simple allusion à des besoins physiologiques.

(554) Un autre aspect innovant de CALM est relatif aux notions de curiosité et de comportement exploratoire. Dans les modèles traditionnels l'exploration est généralement synonyme de faire des actions aléatoires. Dans CALM il y a une mesure d'utilité d'exploration qui permet à l'agent de planifier ses actions afin d'arriver, à long terme, à des nouvelles découvertes, c'est-à-dire, à des nouvelles connaissances qui renforceront son modèle du monde. Cependant, CALM sépare encore au hasard une partie du temps pour l'exploration et l'autre pour l'exploitation, ce qui est encore une façon *ad-hoc* de modéliser la curiosité.

(555) En fait, plusieurs processus qui sont observés chez l'être humain en tant qu'agent intelligent sont encore très loin d'être modélisés de façon convaincante par l'intelligence artificielle, et CALM n'est pas différent. Plusieurs mécanismes clairement importants sont ignorés dans nos modèles, non par négligence, mais à cause de la nécessité de s'adapter à ce qui est possible de faire d'après les connaissances actuelles, aussi bien sur l'intelligence en général, que sur l'IA. Et bien sûr, il fallait limiter aussi nos ambitions à un projet faisable durant un doctorat.

(556) Dans le cas de cette thèse, le point le plus frappant est que notamment nos efforts pour traiter les propriétés non-observables comme des éléments abstraits, même si cela représente une avancée importante sur les mécanismes traditionnels d'IA, ont aboutis à quelque chose qui est encore très loin de ce qui le constructivisme a défini comme la « pensée symbolique ».

(557) On a consacré des efforts sérieux et assidus dans cette recherche, et quand même nous sommes restés bloqués entre deux grands murs: d'un côté, la complexité des problèmes sensorimoteurs, qui nécessitent des modèles continus, capables de réaliser une adaptation cybernétique et interactive, traitement d'imprécision, raffinement de compétences, etc.; de l'autre côté, le problème de la construction de symboles qui représentent des entités et des processus abstraits, et qui mènent l'agent en fait vers une

sorte de pensée de plus haut niveau. Cette thèse n'a réussi à dépasser aucun des deux murs, mais, disons, essaye d'aider l'IA à mettre un pied de chaque côté, pour commencer une longue escalade.

Publications

PEROTTO, F.S.; ÁLVARES, L.O.; BUISSON, J.-C. Un Mecanismo Constructivista para el Aprendizaje de Anticipaciones en Sistemas Acoplados Agente-Ambiente. In: Latin-American Informatics Conference, CLEI, 35th, 2009 Pelotas, RS, Brazil. Proceedings… Pelotas: UFPel, 2009.

PEROTTO, F.S.; ÁLVARES, L.O.; BUISSON, J.-C. *Um Mecanismo Construtivista para a Aprendizagem de Estrutura de MDPs Fatorados e Parcialmente Observáveis*. In: Encontro Nacional de Inteligência Artificial, ENIA, 7th, 2009, Bento Gonçalves, RS, Brazil. **Proceedings**… Porto Alegre: SBC, 2009. p.1029-1038. **ISSN: 2175-2761.**

QUINTON, J.-C.; PEROTTO, F.S.; BUISSON, J.-C. *Anticipative Coordinated Cognitive Processes for Interactivist and Piagetian Theories*. In: Conference on Artificial General Intelligence, AGI, 1st, 2008, Memphis, TN, USA. **Proceedings**… v.171, Amsterdam: IOS Press, 2008. p.287-298. **ISBN: 978-1-58603-833-5.**

PEROTTO, F.S.; ÁLVARES, L.O.; BUISSON, J.-C. *Constructivist Anticipatory Learning Mechanism (CALM): dealing with partially deterministic and partially observable environments*. In: International Conference on Epigenetic Robotics, EpiRob, 7th, 2007, Piscataway, NJ, USA. **Proceedings**… New Jersey: Lund University, 2007. p.117-127. **ISBN: 91-974741-8-5.**

PEROTTO, F.S.; ÁLVARES, L.O. *Incremental Inductive Learning in a Constructivist Agent*. In: International Conference on Innovative Techniques and Applications of Artificial Intelligence, SGAI, 26th, 2006, Cambridge, UK. **Research and Development in Intelligent Systems**, v.23. London: Springer-Verlag, 2007. p.129-144. **ISBN: 97-818462866-5-0.**

PEROTTO, F.S.; ÁLVARES, L.O. *Learning World Models with a Constructivist Agent*. In: Workshop on MSc dissertation and PhD thesis in Artificial Intelligence, WTDIA, 3rd, Ribeirão Preto, SP, Brazil. **Proceedings**… Ribeirão Preto: SBC, 2006. **ISBN 978-85-87837-11-7.**

PEROTTO, F.S.; ÁLVARES, L.O. *Learning Environment Regularities with a Constructivist Agent*. In: International Joint Conference on Autonomous Agents and Multi-Agent Systems, AAMAS, 5th, 2006, Hakodate, Japan. **Proceedings**… New York: ACM, 2006. p.807-809. **ISBN 978-1-59593-303-4.**

SILVA, B.C.; BASSO, E.W.; PEROTTO, F.S. *Reinforcement Learning with Context Detection*. In: International Joint Conference on Autonomous Agents and Multi-Agent Systems, AAMAS, 5th, 2006, Hakodate, Japan. **Proceedings**… New York: ACM, 2006. p.810-812. **ISBN 978-1-59593-303-4.**

PEROTTO, F.S.; ÁLVARES, L.O. *A Alternativa Construtivista em Inteligência Artificial*. In: Latin-American Informatics Conference, CLEI, 31st, 2005, Cali, Colombia. **Proceedings**… v.1, Cali: Feriva, 2005. **ISBN: 978-958-670-422-x.**

PEROTTO, F.S.; VICARI, R.M.; ÁLVARES, L.O. *An Autonomous Intelligent Agent based on Constructivist AI*. In: Artificial Intelligence Applications and Innovations, AIAI, 1st, 2004, Toulouse, France. **Proceedings**… Norwell, MA: Kluwer, 2004. p.103-115. **ISBN: 1-4020-8150-2.**

Références

AHUJA, R.K.; MAGNANTI, T.L.; ORLIN, J.B. **Network Flows**: theory, algorithms, and applications. Englewood Cliffs, NJ: Prentice Hall, 1993.

ALMEIDA, L.B.; SILVA, B.C.; BAZZAN, A.L.C. *Towards a physiological model of emotions: first steps*. In: Spring Symposium, AAAI, 2004, Palo Alto, CA, USA. **Architectures for Modeling Emotion**: cross-disciplinary foundations, Menlo Park: AAAI Press, v.1, 2004. p.1-4.

ALOIMONOS, J.Y.; WEISS, I.; BANDOPADHAY, A. *Active Vision*. **Computer Vision**, Kluwer, v.1, n.4. p.333-356, 1987.

ANDERSON, M. *Embodied Cognition: a field guide.* **Artificial Intelligence**, Elsevier, v.149, n.1, p.91-130, 2003.

ANDERSSON, D.; VOROBYOV, S. *Fast algorithms for monotonic discounted linear programs with two variables per inequality.* **Technical Report**, Cambridge, UK: Isaac Newton Institute, 2006.

ASHBY, W.R. **Design for a Brain**. London: Chapman and Hall, 1952.

AYER, A.J. *Phenomenalism*. In: **Philosophical Essays**. London: Macmillan, 1954. p.142-162.

BAJCSY, R. *Active Perception*. **IEEE Proceedings**, v.76, n.8, p.996-1006, 1988.

BALLARD, D. Animate Vision. **Artificial Intelligence**, Elsevier, v.48, n.1, p.1-27, 1991.

BARANDIARAN, X. *Behavioral Adaptive Autonomy: a milestone on the ALife route to AI?*. In: International Conference on Artificial Life, ALIFE, 9th, 2004, Boston, MA, USA. **Proceedings…** Cambridge, MA: MIT Press, 2004. p.514-521.

BARANDIARAN, X.; MORENO, A. *On what makes certain dynamical systems cognitive*. **Adaptive Behavior**, SAGE, v.14, n.2, p.171-185, 2006.

BARANDIARAN, X.; MORENO, A. *Adaptivity: From Metabolism to Behavior.* **Adaptive Behavior**, SAGE, v.16, n.5, p.325-344, 2008.

BEER, R.D. *A dynamical systems perspective on agent-environment interactions*. **Artificial Intelligence**, Elsevier, v.72, p.173-215, 1995.

BEER, R.D. *Autopoiesis and Cognition in the Game of Life*. **Artificial Life**, MIT Press, v.10, p.10-309, 2004.

BELLMAN, R. *A Markovian Decision Process*. **Journal of Mathematics and Mechanics**, Bloomington: Indiana University Press, v.6. p.679-684, 1957.

BELLMAN, R.E. **Adaptive Control Processes**: a guided tour. New Jersey: Princeton University Press, 1961.

BICKHARD, M.H. *The Interactivist Model*. **Synthese**, Springer, v.166, n.3, p.547-591, 2009.

BICKHARD, M.H. *Emergence*. In: ANDERSEN, P.B. et al. (eds.). **Downward Causation**. Aarhus, Denmark: University of Aarhus Press, 2000. p.322-348.

BICKHARD, M.H.; TERVEEN, L. **Foundational Issues in Artificial Intelligence and Cognitive Science**: impasse and solution. Amsterdam: Elsevier Scientific, 1995.

BLUM, A.; LANGLEY, P. *Selection of relevant features and examples in machine learning*. **Artificial Intelligence**, Elsevier, v.97, p.245-271, 1997.

BOBZIEN, S. **Determinism and Freedom in Stoic Philosophy**. New York: Oxford University Press, 1998.

BODEN, M. **Piaget**. Glasgow: Fontana Paperbacks, 1979.

BOOKER, L.; GOLDBERG, D; HOLLAND, J. *Classifier Systems and Genetic Algorithms*. **Artificial Intelligence**, Elsevier, v.40, p.235-282, 1989.

BOUTILIER, C.; DEARDEN, R.; GOLDSZMIDT, M. *Exploiting Structure in Policy Construction*. In: International Joint Conference on Artificial Intelligence, IJCAI, 14th, 1995, Montreal, Canada. **Proceedings…** Morgan-Kaufmann, v.2, 1995. p.1104-1113.

BOUTILIER, C.; DEARDEN, R.; GOLDSZMIDT, M. *Stochastic dynamic programming with factored representations*. **Artificial Intelligence**, Elsevier, v.121, n.1-2, p.49-107, 2000.

BOUTILIER, C.; POOLE, D. *Computing optimal policies for partially observable decision processes using compact representations*. In: National Conference on Artificial Intelligence, AAAI, 13rd, 1996, Portland, OR, USA. **Proceedings…** Portland: AAAI Press, v.2, 1996. p.1168-1175.

BROOKS, R.A. *Intelligence Without Representation*. **Artificial Intelligence**, Elsevier, v.47, p.139-159, 1991.

BUCHANAN, B.G.; WILKINS, D.C. (eds.). **Readings in Knowledge Acquisition and Learning**, San Mateo, CA: Morgan Kaufmann, 1993.

BUNGE, M. **Causality**: the place of the causal principle in modern science. Cambridge: Harvard University Press, 1959.

CAÑAMERO, D. *A hormonal model of emotions for behavior control*. In: European Conference on Artificial Life, ECAL, 4th. **Proceedings…** Brussels, Belgium: Vrije Universiteit, 1997a. p.1-10.

CAÑAMERO, L. *Modeling motivations and emotions as a basis for intelligent behavior.* In: International Conference on Autonomous Agents, 1st, 1997. **Proceedings…** New York, NY: ACM, 1997b. p.148-155.

CAÑAMERO, L. *Emotions and Adaptation in Autonomous Agents*: a design perspective. **Cybernetics and Systems**, Taylor and Francis, v.32, n.5, p.507-529, 2001.

CANNON, W.B. **The Wisdom of the Body**. New York: W.W. Norton, 1932.

CHAPUT, H. **The Constructivist Learning Architecture**. 2004. Thesis (PhD) – University of Texas.

CHEMERO, A. *Anti-Representationalism and the Dynamical Stance.* **Philosophy of Science**, v.67, n.4, p.625-647, 2000.

CHRISLEY, R.L.; ZIEMKE, T. *Embodiment.* In: NADEL, L. (ed.). **Encyclopedia of Cognitive Science**. London: Macmillan, 2002. p.1102-1108.

CHRISMAN, L. *Reinforcement Learning with Perceptual Aliasing: the perceptual distinctions approach*. In: National Conference on Artificial Intelligence, AAAI, 10th, 1992, San Jose, CA, USA. **Proceedings…** AAAI Press, 1992. p.183-188.

CICCHELLO, O.; KREMER, S. *Inducing grammars from sparse data sets*: a survey of algorithms and results. **Journal of Machine Learning Research**, MIT Press, v.4, p.603-632, 2003.

CLANCEY, W.J. **Situated Cognition**: on human knowledge and computer representation. New York: Cambridge University Press, 1997.

CLARK, A. **Being There**: putting brain, body, and world together again. Cambridge, MA: MIT Press, 1998.

COELHO, H. (1996). **Sonho e Razão**: ao lado do artificial. 2ed. Lisboa: Relógio d'Água, 1999.

COHEN, D. **Piaget**: critique and reassessment. London: Croom Helm, 1983.

COSTA, A.C.R.; DIMURO, G.P. *Interactive Computation*: stepping stone in the pathway from classical to developmental computation. In: Workshop on the Foundations of Interactive Computation, FinCo, 2005, Edinburgh, Scotland. **Electronic Notes in Theoretical Computer Science**, Elsevier, v.141, n.5, 2005. p.5-31.

CREVIER, D. **AI**: the tumultuous search for artificial intelligence. New York: BasicBooks, 1993.

CROOK, P.; HAYES, G. *Learning in a State of Confusion*: perceptual aliasing in grid world navigation. In: Towards Intelligent Mobile Robots, TIMR, 2003, Bristol, UK. **Proceedings…** Bristol: UWE, 2003.

DAMÁSIO, A. **Descartes' Error**: emotion, reason, and the human brain. New York: Putnam Publishing, 1994.

DANTO, A. **Connections to the World**. New York: Harper and Row, 1989.

DARWICHE, A.; GOLDSZMIDT, M. *Action networks*: a framework for reasoning about actions and change under uncertainty. In: International Conference on Uncertainty in Artificial Intelligence, UAI, 10th, 1994, Seattle, WA, USA. **Proceedings…** Morgan-Kaufmann, 1994. p.136-144.

DASDAN, A.; IRANI, S.S.; GUPTA, R.K. *Efficient algorithms for optimum cycle mean and optimum cost to time ratio problems*. In: Design Automation Conference, DAC, 36th, 1999, New Orleans, LA, USA. **Proceedings…** New York: ACM Press, 1999. p.37-42.

DAVIDSSON, P. *On the concept of concept in the context of autonomous agents*. In: World Conference on the Fundamentals of Artificial Intelligence, WOCFAI, 2nd, 1995, Paris, France. **Proceedings…** Paris: Angkor Press, 1995. p.85-96.

DEAN, T.; KANAZAWA, K.A. *Model for reasoning about persistence and causation.* **Computational Intelligence**, Wiley-Blackwell, v.5, n.3, p.142-150, 1989.

DEGRIS, T.; SIGAUD, O.; WUILLEMIN, P-H. *Learning the Structure of Factored Markov Decision Processes in Reinforcement Learning Problems*. In: International Conference on Machine Learning, ICML, 23th, 2006, Pittsburg, PA, USA. **Proceedings…** ACM, 2006. p.257-264.

DEGRIS, T.; SIGAUD, O.; WUILLEMIN, P-H. Exploiting Additive Structure in Factored MDPs for Reinforcement Learning. In: European Workshop or Reinforcement Learning, EWRL, 2008, Villeneuve d'Ascq, France. **Recent Advances in Reinforcement Learning**, Berlin: Springer, 2008. p.15-26.

DELEUZE, G. (1970). **Spinoza**: philosophie pratique. Paris: De Minuit, 1981.

DENNETT, D. **Kinds of Minds**. NewYork: Basic Books, 1996.

DOOB, L.W. **Inevitability:** determinism, fatalism, and destiny. New York: Greenwood Press, 1988.

DRESCHER, G.L. **Made-Up Minds**: a constructivist approach to artificial intelligence. Cambridge: MIT Press, 1991.

DREYFUS, H. **What Computers Can't Do**: the limits of artificial intelligence. New York: Harper and Row, 1972.

DREYFUS, H. **What Computers Still Can't Do**: a critique of artificial reason. Cambridge, MA: MIT Press, 1992.

EKMAN, P. *Basic Emotions*. In: DALGLEISH, T.; POWER, M. (Eds.). **Handbook of Cognition and Emotion**. Sussex, UK: Wiley, 1999. p.45-60.

EKMAN, P.; DAVIDSON, R. **The Nature of Emotion**: fundamental questions. Oxford University Press, 1994.

ELLIS, A. **Reason and Emotion in Psychotherapy**. New York: Lyle Stuart, 1962.

FEINBERG, E.A.; SHWARTZ, A. **Handbook of Markov Decision Processes**: methods and applications. Norwell: Kluwer, 2002.

FLAVELL, J. **The Developmental Psychology of Jean Piaget**. New York: Van Nostrand, 1967.

FODOR, J.A. *Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology.* In: **Mind Design**: philosophy, psychology, artificial intelligence. Cambridge: Bradford Books, 1981. p.307-338.

FRANKLIN, S. *Autonomous Agents as Embodied AI*. **Cybernetics and Systems**, Taylor & Francis, v.28, n.6, p.499-520, 1997.

FRANKLIN, S. *A foundational architecture for artificial general intelligence*. In: Artificial General Intelligence Workshop, 2006, Washington. **Advances in Artificial General Intelligence**: concepts, architectures and algorithms, Amsterdam: IOS Press, 2007. p.36-54.

FRENCH, R.; THOMAS, E. *The dynamical hypothesis: One Battle Behind.* **Behavior and Brain Sciences**, Cambridge Journals, v.21, n.5, p.640-641, 1998.

FREUND, Y.; KEARNS, M.; RON, D.; RUBINFELD, R.; SCHAPIRE, R.E.; SELLIE L. *Efficient learning of typical finite automata from random walks*. Information and Computation, Elsevier, v.138, n.1, p.23-48, 1997.

FROESE, T.; ZIEMKE, T. *Enactive Artificial Intelligence: investigating the systemic organization of life and mind*. **Artificial Intelligence**, Elesvier, v.173, n.3-4, p.466-500, 2009.

GESCHEIDER, G.A. **Psychophysics**: the fundamentals. 3.ed. Mahwah, NJ: Lawrence Erlbaum, 1997.

GLEICK, J. **Chaos**: making a new science. New York: Viking Penguin, 1987.

GOLDIN, D.; WEGNER, P. *Refuting the Strong Church-Turing Thesis: the interactive nature of computing*. **Minds and Machines**, Kluwer, v.18, n.1, p.17-38, 2008.

GRUSH, R. *The Architecture of Representation*. **Philosophical Psychology**, Routledge, v.10, n.1, p.5-23, 1997a.

GRUSH, R. *Yet another design for a brain?*: "Mind as Motion" book review. **Philosophical Psychology**, Routledge, v.10, n.2. p.233-242, 1997b.

GRUSH, R. *The emulation theory of representation: motor control, imagery, and perception*. **Behavioral and Brain Sciences**, Cambridge Journals, v.27. p.377-442, 2004.

GUESTRIN, C.; KOLLER, D.; PARR, R. *Solving Factored POMDPs with Linear Value Functions*. In: Workshop on Planning under Uncertainty and Incomplete Information, 2001, Seattle, WA. **Proceedings…** 2001. p.67-75.

GUESTRIN, C.; KOLLER, D.; PARR, R.; VENKATARAMAN, S. *Efficient Solution Algorithms for Factored MDPs*. **Journal of Artificial Intelligence Research**, AAAI Press, v.19, p.399-468, 2003.

GUYON, I.; ELISSEEFF, A. *An Introduction to Variable and Feature Selection*. In: **Journal of Machine Learning Research**, v.3, p.1157-1182, 2003.

GUYON, I.; GUNN, S.; NIKRAVESH, M.; ZADEH, L. (eds.). **Feature Extraction**: foundations and applications. Heidelberg: Springer, 2006.

HANARD, S. *The Symbol Grounding Problem*. **Physica D**, Elsevier, v.42, p.335-346, 1990.

HANSEN, E.A.; FENG, Z. *Dynamic programming for POMDPs using a factored state representation*. In: International Conference on Artificial Intelligence, Planning and Scheduling, AIPS, 5th, 2000, Breckenridge, CO, USA. **Proceedings…** AAAI, 2000. p.130-139.

HARTMANN, M.; ORLIN, J.B. *Finding minimum cost to time ratio cycles with small integral transit times*. **Networks**, Wiley, v.23, n.6, p.567-574, 1993.

HAUGELAND, J. **Artificial Intelligence**: the very idea. Cambridge: MIT Press, 1985.

HOLLAND, J.; HOLYOAK, K.; NISBETT, R.; THAGARD, P. **Induction**: processes of inference, learning, and discovery, MIT Press, Cambridge, 1986.

HOLMES, M.P.; ISBELL, C.L. *Looping Suffix Tree-Based Inference of Partially Observable Hidden State*. In: International Conference on Machine Learning, ICML, 23th, 2006, Pittsburg, PA, USA. **Proceedings…** ACM, 2006. p.409-416.

HOWARD, R.A. **Dynamic Programming and Markov Processes**. Cambridge: MIT Press, 1960.

JENNINGS, N.R. *On agent-based software engineering*. **Artificial Intelligence**, Elsevier, v.117, p.277-296, 2000.

JONSSON, A.; BARTO, A. *A Causal Approach to Hierarchical Decomposition of Factored MDPs*. In: International Conference on Machine Learning, ICML, 22nd, 2005, Bonn, Germany. **Proceedings…** ACM, 2005. p.401-408.

JUILLÉ, H.; POLLACK, J.B. *SAGE: a sampling based heuristic for tree search*. In: International Colloquium on Grammatical Inference, ICGI, 4th, 1998, Ames, IA, USA. **Proceedings…** Springer-Verlag, 1998. p.126-137. (LNAI 1433).

KAELBLING, L.P.; LITTMAN, M.L.; CASSANDRA, A.R. *Acting optimally in partially observable stochastic domains.* In: National Conference on Artificial Intelligence, AAAI, 12th, 1994, Seattle, WA, USA. **Proceedings…** AAAI Press, 1994. p.1023-1028.

KAELBLING, L.P.; LITTMAN, M.L.; CASSANDRA, A.R. *Planning and acting in partially observable stochastic domains.* **Artificial Intelligence**, Elsevier, v.101, p.99-134, 1998.

KAELBLING, L.P.; LITTMAN, M.L.; MOORE, A. *Reinforcement Learning: a survey.* **Journal of Artificial Intelligence Research**, AAAI Press, v.4, p.237-285, 1996.

KARP, R.M. *A characterization of the minimum cycle mean in a digraph*. **Discrete Mathematics**, Elsevier, v.23, n.3, p.309-311, 1978.

LANG, K.J. *Random DFA's can be Approximately Learned from Sparse Uniform Examples*. ACM Workshop on Computational Learning Theory, COLT, 15th, 1992, Pittsburgh, PA, USA. **Proceedings…** ACM, 1992. p.45-52.

LANG, K.J. *Faster algorithms for finding minimal consistent DFAs*. **Technical Report**. Princetown: NEC, 1999.

LANG, K.J.; PEARLMUTTER, B.; PRICE, R. *Results of the Abbadingo One DFA Learning Competition and a New Evidence Driven State Merging Algorithm*. In: International Colloquium on Grammatical Inference, 4th, ICGI, 1998, Ames, IA, USA. **Proceedings…** Springer-Verlag, 1998. p.1-12. (LNAI 1433).

LEDOUX, J.E. **The Emotional Brain**: the mysterious underpinnings of emotional life. New York: Simon and Schuster, 1996.

LEDOUX, J.E. *Emotion Circuits in the Brain*. **Annual Review of Neuroscience**, Annual Reviews, v.23, n.1, p.155-184, 2000.

LIU, H.; MOTODA, H. (eds.). **Computational Methods of Feature Selection**. Boca Raton, FL: Chapman and Hall, 2007.

MADANI, O. *Polynomial value iteration algorithms for deterministic MDPs*. In: International Conference on Uncertainty in Artificial Intelligence, UAI, 18th,

2002, Alberta, Canada. **Proceedings…** San Francisco: Morgan Kaufmann, 2002. p.311-318.

MADANI, O.; THORUP, M.; ZWICK, U. *Discounted deterministic Markov decision processes and discounted all-pairs shortest paths*. In: Annual Symposium on Discrete Algorithms, SODA, 2009, New York, NY, USA. **Proceedings…** New York: ACM-SIAM, 2009. p.958-967.

MAES, P. *Modeling Adaptive Autonomous Agents*. **Artificial Life**, MIT Press, v.1, p.135-162, 1994.

MATURANA, H.R. **Desde la Biologia a la Psicologia**. Santiago: Synthesis, 1993.

MATURANA, H.; VARELA, F. **De Máquinas y Seres Vivos**: una caracterización de la organización biológica. Santiago, Chile: Editorial Universitaria, 1973.

MATURANA, H.; VARELA, F. *Autopoiesis: the realization of the living*. In: MATURANA, H.; VARELA, F. (eds.), **Autopoiesis and Cognition**: the realization of the living. Dordrecht, Holland: D. Reidel Publishing, 1980. p.73-138.

McALLESTER, D.A.; SINGH, S.P. *Approximate Planning for Factored POMDPs using Belief State Simplification*. In: International Conference on Uncertainty in Artificial Intelligence, UAI, 15th, 1999, Stockholm, Sweden. **Proceedings…** San Francisco, CA: Morgan Kaufmann, 1999. p.409-416.

McCARTHY, J.; HAYES, P.J. Some Philosophical Problems from the Standpoint of Artificial Intelligence. **Machine Intelligence**, Edinburgh University Press, v.4, p.463-502, 1969.

McCORDUCK, P. **Machines Who Think**. San Francisco: Freeman, 1979.

MEULEAU, N.; KIM, K-E.; KAELBLING, L.P.; CASSANDRA, A.R. *Solving POMDPs by Searching the Space of Finite Policies*. In: International Conference on Uncertainty in Artificial Intelligence, UAI, 15th, 1999, Stockholm, Sweden. **Proceedings…** San Francisco, CA: Morgan Kaufmann, 1999. p.427-443.

MEYER, P. **L'œil et le Cerveau**: biophilosophie de la perception visuelle. Paris: Odile Jacob, 1997.

MITCHELL, T. *Generalization as Search*. **Artificial Intelligence**, Elsevier, v.18, p.203-226, 1982.

MONTANGERO, J.; MAURICE-NAVILLE, D. **Piaget**: l'intelligence en marche. Liège, Belgium: Mardaga, 1994.

MONTEBELLI, A.; HERRERA, C.P.; ZIEMKE, T. *On Cognition as Dynamical Coupling: an analysis of behavioral attractor dynamics*. **Adaptive Behavior**, v.16, n.2-3, p.182-195, 2008.

MOORE, E.F. *Gedanken-experiments on Sequential Machines*. **Annals of Mathematical Studies**, Princeton, NJ: Princeton University Press, v.34, p.129-153, 1956.

MURPHY, O.J.; McCRAW, R.L. *Designing storage efficient decision trees*. **IEEE Transactions on Computers**, v.40, n.3, p.315-319, 1991.

NOLFI, S. *Power and limits of reactive agents*. **Neurocomputing**, Elsevier, v.49, p.119-145, 2002.

ÖHMAN, A. *The role of the amygdala in human fear: automatic detection of threat*. **Journal of Psychoneuroendocrinology**, Oxford, UK: Elsevier, v.30, n.10, p.953-958, 2005.

ORTNER, R. *Online Regret Bounds for Markov Decision Processes with Deterministic Transitions*. In: International Conference on Algorithmic Learning Theory, ALT, 19th, 2008, Budapest, Hungary. **Proceedings…** Springer, 2008. p.123-137. (LNAI 5254).

OUDEYER, P.; KAPLAN, F. *What is Intrinsic Motivation? a typology of computational approaches*. **Frontiers in Neurorobotics**, v.1, n.6, 2007.

OVERTON, W.; MÜLLER, U.; NEWMAN, J. (eds.). **Developmental Perspectives on Embodiment and Consciousness**. New York: Lawrence Erlbaum, 2008.

PAPADIMITRIOU, C.H.; TSITSIKLIS, J.N. *The complexity of Markov decision processes*. **Mathematics of Operations Research**, Informs, v.12, n.3, p.441-450, 1987.

PARISI, D. *Internal Robotics*. **Connection Science**, Taylor and Francis, v.16, n.4, p.325-338, 2004.

PENA, J.M.; e OLIVEIRA, A.L. *A new algorithm for the reduction of incompletely specified finite state machines*. In: International Conference on Computer Aided Design, ICCAD, 1998, San Jose, CA, USA. **Proceedings…** ACM, 1998. p.482-489.

PENNACHIN, C.; GOERTZEL, B. *Contemporary Approaches to Artificial General Intelligence.* In: GOERTZEL, B.; PENNACHIN, C. (eds.). **Artificial General Intelligence**. New York: Springer, 2007.

PENROSE, R. **The Emperor's New Mind**: concerning computers, minds, and the laws of physics. New York: Oxford University Press, 1989.

PIAGET, J. **La Naissance de l'Intelligence Chez l'Enfant**. Neuchatel: Delachaux et Niestlé, 1936.

PIAGET, J. **La Construction du Reél Chez l'Enfant**. Neuchatel: Delachaux et Niestlé, 1937.

PIAGET, J. **La Formation du Symbole Chez l'Enfant**: imitation, jeu et revê, image et représentation. Neuchatel: Delachaux et Niestlé, 1945.

PIAGET, J. **La Psychologie de l'Intelligence**. Paris: Armand Colin, 1947.

PIAGET, J. (1954). *The relation of affetivity to intelligence in the mental development of the child*. **Bulletin of the Menninger Clinic**, Guilford Press, v.26, n.3, 1962.

PIAGET, J. **Biologie et Connaissence**: essai sur lês relations entre lês régulations organiques et lês processus cognitifs. Paris: Éditions de la Pléïade, 1967.

PIAGET, J. **L'Équilibration des Structures Cognitives**: problème central du développement. Paris: PUF, 1975.

PIAGET, J. *Le Developpement Mental de l'Enfant*. In: PIAGET, J. **Six Études de Psychologie**. Paris: Gonthier, 1964.

POUPART, P.; BOUTILIER, C. *VDCBPI: an approximate scalable algorithm for large scale POMDPs*. In: Advances in Neural Information Processing Systems, NIPS, 17th, 2004, Vancouver, Canada. **Proceedings…** Cambridge: MIT Press, 2004. p.1081-1088.

POUPART, P. **Exploiting Structure to Efficiently Solve Large Scale Partially Observable Markov Decision Processes**. 2005. Thesis (PhD on Computer Sciences) – University of Toronto, Canada.

PUTERMAN, M.L. **Markov Decision Processes**: discrete stochastic dynamic programming. New York: Wiley, 1994.

PYLYSHYN, Z.W. (ed.) **The Robot's Dilemma**: the frame problem in artificial intelligence, Norwood: Ablex, 1987.

QUICK, T.; DAUTENHAHN, K.; NEHANIV, C.L.; ROBERTS, G. *The Essence of Embodiment*: a framework for understanding and exploiting structural coupling between system and environment. In: International Conference on Computing Anticipatory Systems, CASYS, 3rd, 1999, Liège, Belgium. **Proceedings…** Liège: Chaos, 1999. p.649-660.

QUINLAN, J.R. *Induction of Decision Trees*. **Machine Learning**, Springer, v.1, n.1, p.81-106, 1986.

QUINLAN, J.R. ***C4.5***: programs for machine learning. San Mateo, CA: Morgan Kaufmann, 1993.

QUINTON, J.-C.; PEROTTO, F.S.; BUISSON, J.-C. *Anticipative Coordinated Cognitive Processes for Interactivist and Piagetian Theories*. In: Conference on Artificial General Intelligence, AGI, 1st, 2008, Memphis, TN, USA. **Proceedings**… v.171, Amsterdam: IOS Press, 2008. p.287-298.

RON, D.; RUBINFELD, R. *Exactly Learning Automata of Small Cover Time*. **Machine Learning**, Springer, v.27, n.1, p.69-96, 1997.

RUIZ-MIRAZO, K.; MORENO, A. *Searching for the Roots of Autonomy: the natural and artificial paradigms revisited*. **Artificial Intelligence**, Elsevier, v.17, n.3-4, p.209-228, 2000.

RUIZ-MIRAZO, K.; MORENO, A. *Basic autonomy as a fundamental step in the synthesis of life*. **Artificial Life**, MIT Press, v.10, n.3, p.235-259, 2004.

RUSSELL, S.; NORVING, P. (1995). **Artificial Intelligence**: a modern approach. 2.ed. New Jersey: Prentice-Hall, 2003.

SALLANS, B.; HINTON, G.E. *Reinforcement Learning with Factored States and Actions*. **Journal of Machine Learning Research**, MIT Press, v.5, p.1063-1088, 2004.

SEARLE, J. *Minds, Brains and Programs*. **Behavioral and Brain Sciences**, Cambridge Journals, v.3, n.3. p.417-457, 1980.

SHANI, G.; BRAFMAN, R.I.; SHIMONY, S.E. *Model-Based Online Learning of POMDPs.* In: European Conference on Machine Learning, ECML, 16th, 2005,

Porto, Portugal. **Proceedings…** Berlin: Springer-Verlag, 2005. p.353-364. (LNCS 3720).

SHANI, G.; POUPART, P.; BRAFMAN, R.I.; SHIMONY, S.E. *Efficient ADD Operations for Point-Based Algorithms*. In: International Conference on Automated Planning and Scheduling, ICAPS, 8*th*, 2008, Sydney, Australia. *Proceedings*... AAAI Press, 2008. p.330-337

SHANON, B. **The Representational and the Presentational**: an essay on cognition and the study of the mind. Hemel Hempstead, UK: Harvester Wheatsheaf, 1993.

SHERRINGTON, C.S. *On the proprioceptive system, especially in its reflex aspect*. **Brain Journal**, Oxford Journals, v.29, n.4, p.467-482, 1907.

SIEDLECKI, W.; SKLANSKY, J. *On Automatic Feature Selection*. In: CHEN, C.H.; PAU, L.F.; WANG, P.S.P. (eds.). **Handbook of Pattern Recognition and Computer Vision**, River Edge, NJ: World Scientific Publishing, 1993. p.63-87.

SIM, H.S.; KIM, K.-E.; KIM, J.H.; CHANG, D.-S.; KOO, M.-W. *Symbolic Heuristic Search Value Iteration for Factored POMDPs*. In: National Conference on Artificial Intelligence, AAAI, 23rd, 2008, Chicago, IL, USA. **Proceedings…** AAAI Press, 2008. p.1088-1093.

SIMON, H.A. *Why Should Machines Learn?*. In: MICHALSKI, R.S.; CARBONELL, J.G.; MITCHELL, T.M. (eds.). **Machine Learning**: an artificial intelligence approach. Palo Alto, CA: Tioga, 1983. p.25-38.

SIMONS, G.L. **Introducing Artificial Intelligence**. New York: Halsted, 1984.

SINGH, S.; LITTMAN, M.; JONG, N.; PARDOE, D; STONE, P. *Learning Predictive State Representations*. In: International Conference on Machine Learning, ICML, 20th, 2003, Washington, DC, USA. **Proceedings…** AAAI Press, 2003. p.712-719.

SINGH, S.; BARTO, A.G.; CHENTANEZ, N. *Intrinsically Motivated Reinforcement Learning*. In: Advances in Neural Information Processing Systems, NIPS, 17th, 2004, Vancouver, Canada. **Proceedings…** Cambridge, MA: MIT Press, 2004.

SLOMAN, A. *Architectural Requirements for Human-like Agents Both Natural and Artificial: what sorts of machine can love?*. In: DAUTENHAHN, K. (Ed.). **Human Cognition and Social Agent Technology**. London: John Benjamins, 1999.

SLOMAN, A.; CHRISLEY, R.; SCHEUTZ, M. *The architectural basis of affective states and processes*. In: FELLOUS; ARBIB (eds.). **Who Needs Emotions?** the brain meets the machine. Oxford, UK: Oxford University Press, 2005. p.201-244.

SMALLWOOD, R.D.; SONDIK, E.J. *The optimal control of partially observable Markov decision processes over a finite horizon*. **Operations Research**, Informs, v.21, p.1071-1088, 1973.

STEWART, J.; GAPENNE, O.; DI PAOLO, E. **Enaction**: a new paradigm for cognitive science. Cambridge, MA: MIT Press, 2008.

STREHL, A.L., DIUK, C., LITTMAN, M.L. *Efficient Structure Learning in Factored-State MDPs*. In: National Conference on Artificial Intelligence, AAAI, 22nd, 2007, Vancouver, Canada. **Proceedings…** AAAI Press, 2007. p.645-650.

SUCHMAN, L.A. **Plans and Situated Actions**. Cambridge: Cambridge University Press, 1987.

SUPPES, P. *The Transcendental Character of Determinism.* **Midwest Studies in Philosophy**, Wiley-Blackwell, v.18, p.242-257, 1993.

SUTTON, R.S.; BARTO, A.G. **Reinforcement Learning**: an introduction. Cambridge, MA: MIT Press, 1998.

SYMONS, J. *Explanation, Representation and the Dynamical Hypothesis*. **Minds and Machines**, Springer, v.11, n.4, p.521-541, 2001.

TEIXEIRA, J.F. **Mente, Cérebro e Cognição**. Petrópolis: Vozes, 2000.

THORNTON, C. *Indirect sensing through abstractive learning*. **Intelligent Data Analysis**, IOS Press, v.7, n.3, p.1-16, 2003.

TRIVIÑO, J.; MORALES, R. *Multiattribute Prediction Suffix Graphs*: a unified view of Markov chains and decision trees. In: **Tendencias de la Minería de Datos en España**. Madrid: REMD, 2004.

VALIANT L.G. *A Theory of the Learnable*. **Communications of the ACM**, v.27, n.11, p.1134-1142, 1984.

VAN GELDER, T.J. *The dynamical hypothesis in cognitive science.* **Behavioral and Brain Sciences**, Cambridge Journals, v.21. p.615-628, 1998.

VARELA, F.; THOMPSON, E.; ROSCH, E. **The Embodied Mind**: cognitive science and human experience. Cambridge, MA: MIT Press, 1991.

VELÁSQUEZ, J. *Modeling Emotions and Other Motivations in Synthetic Agents*. In: National Conference on Artificial Intelligence, AAAI, 14th, 1997, Providence, RI. **Proceedings…** AAAI Press, 1997. p.10-15.

WASSON, G.; KORTENKAMP, D.; HUBER, E. *Integrating Active Perception with an Autonomous Robot Architecture*. **Robotics and Autonomous Systems**, v.29, n.2-3, p.175-186, 1999.

WATKINS, C.J.; DAYAN, P. *Q-Learning*. **Machine Learning**, Springer, v.8, n.3, p.279-292, 1992.

WEGNER, P. *Interactive Foundations of Computing*. **Theoretical Computer Science**, Elsevier, v.192, p.315-351, 1998.

WHITEHEAD, S.D.; BALLARD, D.H. *Learning to perceive and act by trial and error.* **Machine Learning**, Springer, v.7, n.1, p.45-83, 1991.

WHITEHEAD, S.D.; LIN, L.J. *Reinforcement Learning of Non-Markov Decision Processes*. **Artificial Intelligence**, Elsevier, v.73, n.1-2, p.271-306, 1995.

WIENER, N. **Cybernetics**: or the control and communication in the animal and the machine. Cambridge, MA: MIT Press, 1948.

WILIAMS, J.D.; **Partially Observable Markov Decision Processes for Spoken Dialogue Management**. Ph.D. Thesis, Cambridge University, 2006.

WILSON, R.; CLARK, A. *How to Situate Cognition: Letting Nature Take its Course.* In: AYDEDE, M.; ROBBINS, P. (eds.). **Cambridge Handbook of Situated Cognition**. New York: Cambridge University Press, 2008.

YOUNG, N.E.; TARJAN, R.E.; ORLIN, J.B. *Faster parametric shortest path and minimum-balance algorithms*. **Networks**, Wiley-Blackwell, v.21. p.205-221, 1991.

ZIEMKE, T. *Adaptive Behavior in Autonomous Agents*. **Presence**, MIT Press, v.7, n.6. p.564-587, 1998.

ZIEMKE, T. *What's that thing called Embodiment?* In: Annual Meeting of the Cognitive Science Society, 25th, 2002, Boston, MA, USA. **Proceedings**… Lawrence Erlbaum, 2003. p.1305-1310.