

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

INSTITUT NATIONAL POLYTECHNIQUE  
INSTITUT DE RECHERCHE EN INFORMATIQUE DE TOULOUSE

FILIPO STUDZINSKI PEROTTO

**Um Mecanismo Construtivista para  
Aprendizagem de Antecipações em  
Agentes Artificiais Situados**

Tese de Doutorado

Prof. Dr. Luís Otávio Campos Álvares  
Orientador (UFRGS)

Prof. Dr. Jean-Christophe Buisson  
Orientador (INPT)

Porto Alegre, Junho de 2010.

## CATALOGAÇÃO NA PUBLICAÇÃO

**Perotto, Filipo Studzinski**

Um Mecanismo Construtivista para Aprendizagem de Antecipações em Agentes Artificiais Situados / Filipo Studzinski Perotto: PPGC / UFRGS, 2010.

197 f.: Il.

Tese (Doutorado) – Universidade Federal do Rio Grande do Sul, Programa de Pós Graduação em Computação, Porto Alegre, RS, Brasil, em cooperação com o Institut National Polytechnique de Toulouse, França, 2010. Orientadores: Luís Otávio Álvares e Jean-Christophe Buisson.

1. Inteligência Artificial Construtivista. 2. Inteligência Artificial. 3. Aprendizagem de Máquina. 4. Agentes Autônomos. 5. Indução de Conceitos. 6. Desenvolvimento Cognitivo Artificial. 7. Piaget. 8. Descoberta de Estrutura em FPOMDP. 9. Processo de Decisão de Markov (MDP). I. ALVARES, Luís Otávio Campos. II. BUISSON, Jean-Christophe. III. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Vice-Reitor: Prof. Rui Vicente Oppermann

Pró-Reitor de Pós-Graduação: Prof. Aldo Bolten Lucion

Diretor do Instituto de Informática: Prof. Flávio Rech Wagner

Coordenador do PPGC: Prof. Álvaro Freitas Moreira

Bibliotecária-Chefe do Instituto de Informática: Beatriz Regina Bastos Haro

INSTITUT NATIONAL POLYTECHNIQUE DE TOULOUSE (INPT)

Président: Prof. Gilbert Casamatta

Directeur de l'IRIT: Luis Farinas del Cerro

Directeur Adjoint à l'ENSEEIH: Michel Dayde

Président de l'école doctorale MITT: Louis Féraud

## APRESENTAÇÃO E AGRADECIMENTOS

- (1) A presente tese é resultado da pesquisa de doutorado realizada por mim, Filipo Studzinski Perotto, enquanto aluno do Programa de Pós-Graduação em Computação (PPGC) da Universidade Federal do Rio Grande do Sul (UFRGS), sob a orientação do professor Luís Otávio Campos Álvares, no período de agosto de 2004 a junho de 2010. O trabalho foi executado em cotutela com o Instituto Nacional Politécnico de Toulouse (INPT), na França, sob a coorientação do professor Jean-Christophe Buisson. A estes cinco anos de pesquisa abordando o tema da aprendizagem de máquina no modelo construtivista somam-se ainda pelo menos mais quatro, pois de fato, o interesse pelo assunto remonta ao ano de 2001.
- (2) Como trabalho de conclusão do curso de graduação em computação, escrevi uma monografia analisando a Inteligência Artificial através de sua história enquanto disciplina científica, e resgatando as discussões sobre o empreendimento da IA de um ponto de vista teórico-filosófico. Paralelamente ao trabalho de diplomação, realizei uma incursão às disciplinas do Instituto de Filosofia e Ciências Humanas, de Psicologia, e de Educação, o que aguçou minha inquietação em relação às questões sobre a mente e a inteligência, levando-me finalmente a conhecer e interessar-me pela Psicologia Construtivista. A partir daí seguiram-se mais dois anos de pesquisa durante o mestrado, entre 2002 e 2004, sob orientação da professora Rosa Maria Vicari, onde realizamos um trabalho já relacionado ao tema da Inteligência Artificial Construtivista.
- (3) Enfim, esta tese é parte de uma trajetória de pesquisa maior. Olhando para trás, percebo que nos últimos 10 anos de universidade estive focado mais ou menos constantemente sobre o mesmo assunto, um tema que é sem dúvida fascinante. Tanta dedicação não pode representar outra coisa que não um projeto de vida acadêmica. E, apesar de todo esse tempo, chego ao final da tese com a impressão de que ainda há

muito mais perguntas do que respostas, significando, portanto, que esse projeto está muito longe de se esgotar.

(4) Gostaria de agradecer às instituições que fomentaram nossa pesquisa, notadamente ao Conselho Nacional de Pesquisa (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), órgãos nacionais públicos cujo apoio foi imprescindível para a realização do trabalho. Agradeço igualmente à Universidade Federal do Rio Grande do Sul (UFRGS), por intermédio do Programa de Pós-Graduação em Computação (PPGC) do Instituto de Informática, e ao Instituto Nacional Politécnico de Toulouse (INPT), através do Instituto de Pesquisas em Informática (IRIT). Estas duas universidades, em cooperação, nos acolheram de forma excepcional, oferecendo o ambiente e os recursos adequados para a boa realização do trabalho.

(5) Gostaria muito de agradecer aos amigos, o professor Jean-Christophe Buisson e o professor Luís Otávio Álvares, que aceitaram orientar essa pesquisa no terreno imprevisível de um tema complexo e ao mesmo tempo apaixonante, por sua confiança e por sua atitude sempre correta e encorajadora.

(6) De outra parte gostaria de agradecer a todos os pesquisadores que contribuíram de diversas formas neste trabalho, entre os quais estão os professores 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, e Jaime Rebello.

(7) Alguns colegas também estiveram próximos do processo de construção da tese, participando algumas vezes ativamente, e a estes amigos, como Jean-Charles Quinton, Guillaume Giffard, Bruno Castro da Silva, Eduardo Wisnieski Basso, Licurgo Bennemann Almeida, Juliano Bittencourt, e Ivan Medeiros, meu agradecimento. Em particular e muito especialmente a Cássia Trojahn dos Santos, que compartilhou comigo esta batalha travada durante o doutorado, assim como um período da minha vida que eu guardarei sempre no coração.

(8) Eu gostaria de expressar vivamente minha gratidão àquelas pessoas importantes fora do contexto acadêmico, pelo apoio recebido durante todo o (longo) período de trabalho, sem o qual o sucesso não seria possível. Muito me honra poder agradecer

abertamente a Fabiano Mesquita Padão e a Simone Segalla Dutra de Souza, por sua amizade sincera e presente.

(9) Do mesmo jeito agradeço aos bons amigos do meu tempo na França, entre os quais eu gostaria muito de citar: Pierre Amilhaud, Helène Roques, Sonia Rodriguez, André de Andrade, Cláudia Santos Gai, Bernardo Cougo, Marta Ramos Oliveira, Daniel Mur, Fares Fares, Constance Le Pocher e Yoan Elgorriaga.

(10) Entre as tantas pessoas que passaram de forma marcante pela minha vida, a mais ou a menos tempo, rápida ou longamente, acadêmica ou pessoalmente, gostaria de mencionar: 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 Hailiot, Thaís Pessato, Cacá Prüfer, Sabrina Bertrand Coelho, Thiago Ingrassia Pereira, Josiane Faganello, Vera Mello, e Willy Ricardo Petersen Filho.

(11) À minha família, primeiramente a todos os muitos tios e primos, parceiros infalíveis. Muito especialmente aos meus pais, Adelino Victo Perotto e Maria Studzinski Perotto, com quem aprendi a maior parte das qualidades necessárias para ter sucesso nesta tese e na vida. E também a meu irmão, Rafaello Studzinski Perotto, e sua noiva, Kelly Lopes dos Santos, pessoas por quem tenho enorme carinho e admiração.

(12) Finalmente fico muito feliz de poder dedicar esse trabalho, e também agradecer a Emanuele Carvalheira de Maupeou, minha companheira, pessoa e mulher de tantos adjetivos, por estar ao meu lado, e por este amor tão intenso que viverá sempre, em qualquer tempo e em qualquer lugar.

## SUMÁRIO

LISTA DE FIGURAS.....	8
LISTA DE SÍMBOLOS E ACRÔNIMOS.....	11
LISTA DE DEFINIÇÕES.....	13
LISTA DE ALGORITMOS.....	14
<b>1. INTRODUÇÃO.....</b>	<b>18</b>
1.1. Postura Epistemológica.....	19
1.1.1. A IA Como Disciplina Científica.....	19
1.1.2. IA Geral.....	22
1.1.3. Objetivos e Organização da Tese.....	24
1.2. Construtivismo e Cibernética.....	25
1.2.1. Inteligência Artificial Construtivista.....	26
1.2.2. Relação Agente-Ambiente.....	27
1.3. Desafios da Aprendizagem.....	28
1.3.1. Aprendizagem de Modelos de Mundo.....	29
1.3.2. Aprendizagem de Conceitos.....	29
1.3.3. Invenção de Conceitos Abstratos.....	30
1.3.4. Problema da Decisão Sequencial.....	32
1.3.5. Processos de Decisão de Markov.....	33
1.4. Contribuições.....	34
1.4.1. A Arquitetura CAES.....	34
1.4.2. O Mecanismo CALM.....	35
<b>2. CAES: SISTEMA DE ACOPLAMENTO AGENTE-AMBIENTE.....</b>	<b>37</b>
2.1. Relação entre Agente e Ambiente.....	38
2.1.1. Interatividade, Autonomia e Situatividade.....	40
2.1.2. Acoplamento.....	43
2.2. Características do Sistema Global.....	45
2.2.1. Solipsismo Metodológico.....	47
2.2.2. Situação e Atuação.....	48
2.3. Características do Agente.....	50
2.3.1. Corporificação.....	51
2.3.2. O Corpo do Agente.....	53
2.3.3. Percepção e Controle.....	54
2.4. Adaptação do Agente ao Ambiente.....	58
2.4.1. Estar Adaptado.....	59

2.4.2. Ser Adaptativo.....	61
2.5. Características da Mente.....	62
2.5.1. Afetividade e Emoções.....	64
2.5.2. Sistema Avaliativo.....	65
2.5.3. Sistema Emocional.....	70
2.5.4. Sistema Reativo.....	72
2.5.5. Cognição e Aprendizagem.....	73
2.5.6. Sistema Cognitivo.....	74
3. CALM: MECANISMO DE APRENDIZAGEM CONSTRUTIVISTA.....	76
3.1. Definição dos Problemas de Aprendizagem.....	76
3.1.1. Aprendizagem de Modelos de Mundo.....	77
3.1.2. Construção de Política de Ações.....	77
3.1.3. Aprendizagem Ativa, em Horizonte Infinito, e Incremental.....	79
3.1.4. Representando o Sistema como Processo de Decisão de Markov.....	81
3.1.5. Ambientes Estruturados.....	89
3.1.6. Processo Fatorado e Parcialmente Observável.....	94
3.1.7. Determinismo.....	99
3.1.8. O mundo real como ambiente a ser aprendido.....	105
3.2. O Mecanismo de Aprendizagem CALM .....	108
3.2.1. Ideia Geral do Mecanismo.....	108
3.2.2. Memória Episódica Generalizada.....	112
3.2.3. Seleção de Propriedades Relevantes.....	115
3.2.4. Árvore de Antecipação.....	118
3.2.5. Esquema.....	120
3.2.6. Análise de Escalabilidade.....	123
3.2.7. Atualização da Árvore de Antecipação .....	125
3.2.8. Propriedades Não-Observáveis.....	133
3.2.9. O Processo de Decisão.....	139
4. RESULTADOS EXPERIMENTAIS.....	146
4.1. Problema Wepp.....	146
4.1.1. Definição do Problema.....	147
4.1.2. Resultados e Considerações.....	153
4.2. Problema Flip.....	164
4.2.1. Construção da Solução pelo CALM.....	165
4.2.2. Comparação das Soluções.....	171
5. CONCLUSÃO.....	174
5.1. Considerações sobre o Mecanismo CALM.....	175
5.2. Considerações sobre a Arquitetura CAES.....	177
5.3. Propriedades Não-Observáveis e Abstração.....	177
5.4. Limitações e Trabalhos Futuros.....	180
PUBLICAÇÕES.....	183
REFERÊNCIAS.....	185

## LISTA DE FIGURAS

Figura 2.1: Arquitetura CAES e seus três níveis de interação.....	39
Figura 2.2: Computação baseada em agentes.....	41
Figura 2.3: Um ambiente $\xi$ povoado por diversos agentes $d$ .....	42
Figura 2.4: O sistema global na arquitetura CAES.....	46
Figura 2.5: Condição de não-onisciência.....	49
Figura 2.6: O agente na arquitetura CAES.....	50
Figura 2.7: A mente, o corpo, e o ambiente.....	53
Figura 2.8: Fluxo de controle. ....	57
Figura 2.9: Fluxo de percepção. ....	57
Figura 2.10: Exemplo de região de adaptação e trajetória adaptada.....	58
Figura 2.11: Estrutura interna da mente ( $\mu$ ) na arquitetura CAES.....	63
Figura 2.12: Estrutura interna do sistema regulativo ( $\mathcal{R}$ ) na arquitetura CAES. ....	64
Figura 2.13: Exemplo de uma superfície afetiva. ....	68
Figura 2.14: A superfície afetiva mapeando o limiar de sobrevivência.....	69
Figura 3.1: Exemplo de uma rede bayesiana dinâmica (DBN).....	88
Figura 3.2: Exemplo de DBNs completos mas independentes.....	89
Figura 3.3: Exemplo de DBNs independentes e compactos.....	90
Figura 3.4: Relação de complexidade entre propriedades, estados, e relevância.....	93
Figura 3.5: Hierarquia parcialmente ordenada de domínios possíveis.....	94
Figura 3.6: Transição do estado do sistema, de um instante para outro. Percepção indireta e parcial.....	97
Figura 3.7: Exemplo de DBNs explicitando as propriedades ocultas.....	99
Figura 3.8: Relação de determinismo parcial.....	102
Figura 3.9: Exemplo de um DBN e de suas probabilidades de transformação.....	103
Figura 3.10: Exemplo da transformação representada como regularidade.....	103
Figura 3.11: Exemplo de regularidade, considerando propriedades ocultas.....	104
Figura 3.12: Exemplo de uma situação ambígua.....	105
Figura 3.13: Exemplo de formação da memória episódica generalizada.....	115

Figura 3.14: Exemplo de espaço de busca da relevância.....	117
Figura 3.15: Exemplo de uma árvore de antecipação.....	120
Figura 3.16: Representação de um esquema.....	121
Figura 3.17: Crescimento da memória episódica generalizada (numa escala logarítmica).....	125
Figura 3.18: Relação geral / especializado, entre o esquema e a memória.....	128
Figura 3.19: Exemplo de diferenciação. ....	131
Figura 3.20: Exemplo de ajuste. ....	132
Figura 3.21: Integração entre irmãos.....	133
Figura 3.22: Integração entre esquemas primos.....	133
Figura 3.23: Exemplo de assinatura de uma situação a partir das observações. ....	135
Figura 3.24: Desambiguação.....	135
Figura 3.25: Observabilidade parcial pode explicar o aparente não-determinismo.....	136
Figura 3.26: Indução de propriedades não-observáveis.....	138
Figura 3.27: Exemplo de uma tomada de decisão.....	141
Figura 4.1: Retrato do problema wepp. ....	147
Figura 4.2: Função de evolução do ambiente no problema wepp.....	148
Figura 4.3: Função de evolução do ambiente no problema wepp.....	150
Figura 4.4: Configurações de simulação utilizadas nos experimentos com o problema wepp. .....	155
Figura 4.5: Curvas de convergência no tabuleiro 5 x 5 (25 células).....	156
Figura 4.6: Curvas de convergência no tabuleiro 25 x 25 (625 células).....	156
Figura 4.7: Curvas de convergência no tabuleiro 125 x 125 (15625 células).....	157
Figura 4.8: Curvas de convergência para 20% de obstáculos.....	158
Figura 4.9: Árvore de antecipação do prazer.....	159
Figura 4.10: Árvore de antecipação da propriocepção.....	159
Figura 4.11: Árvore de antecipação da dor.....	159
Figura 4.12: Árvore de antecipação da visão.....	159
Figura 4.13: Árvore de antecipação da fadiga.....	160
Figura 4.14: Árvore de antecipação da exaustão.....	160
Figura 4.15: Resultados da simulação (caso típico).....	161
Figura 4.16: Wepp 5 x 5: Q-Learning x CALM.....	162
Figura 4.17: Wepp 25 x 25: Q-Learning x CALM.....	162
Figura 4.18: Wepp 125 x 125: Q-Learning x CALM.....	163
Figura 4.19: Análise de escalabilidade.....	164
Figura 4.20: Problema flip, mostrado como uma máquina de estados.....	165

Figura 4.21: Começo da construção da solução CALM para o problema flip.....	166
Figura 4.22: Primeira estabilização da solução, ainda sem elementos sintéticos.....	166
Figura 4.23: Criação do elemento sintético.....	168
Figura 4.24: Segunda estabilização da solução. ....	169
Figura 4.25: Nova tabela da memória episódica herda valores de sua correspondente.....	170
Figura 4.26: Última estabilização, com a solução final. ....	171
Figura 4.27: Árvores de antecipação construídas pelo CALM no experimento flip.....	172
Figura 4.28: Árvore de deliberação construída pelo CALM para o problema flip. ....	172
Figura 4.29: PST para o problema flip.....	173

## LISTA DE SÍMBOLOS E ACRÔNIMOS

$\mathcal{A}$	Sistema Global	$M = \{M_1, M_2 \dots\}$ Variáveis de Atuação ( <i>Motoras</i> )
$\mathcal{A}$	Agente	$m^{(t)} = \{m_1, m_2 \dots\}$ Atuação no tempo $t$
$\xi$	Ambiente	$S = \{S_1, S_2 \dots\}$ Variáveis de Situação ( <i>Sensoriais</i> )
$\beta$	Corpo	$s^{(t)} = \{s_1, s_2 \dots\}$ Situação no tempo $t$
$\mu$	Mente	$C = \{C_1, C_2 \dots\}$ Variáveis de Controle
$\mathcal{R}$	Sistema Regulativo	$c^{(t)} = \{c_1, c_2 \dots\}$ Controle no tempo $t$
$\mathcal{K}$	Sistema Cognitivo	$X = \{X_1, X_2 \dots\}$ Variáveis de Contexto
$\mathcal{E}$	Exterioridade	$x^{(t)} = \{x_1, x_2 \dots\}$ Contexto no tempo $t$
		$P = \{P_1, P_2 \dots\}$ Variáveis de Percepção ( <i>Observáveis</i> )
$Q = \{q_1, q_2 \dots\}$	Conjunto de Estados	$p^{(t)} = \{p_1, p_2 \dots\}$ Contexto Perceptivo em $t$
$A = \{a_1, a_2 \dots\}$	Conjunto de Ações	$H = \{H_1, H_2 \dots\}$ Variáveis Abstratas ( <i>Ocultas</i> )
$O = \{o_1, o_2 \dots\}$	Conjunto de Observações	$h^{(t)} = \{h_1, h_2 \dots\}$ Contexto Abstrato em $t$
$\gamma$	Função de Observação	$\tau = \{\tau_1, \tau_2 \dots\}$ Funções de Transformação
$\delta$	Função de Transição	$\sigma = \{\sigma_1, \sigma_2 \dots\}$ Funções de Regularidade
$r$	Função de Recompensa	$\pi = \{\pi_1, \pi_2 \dots\}$ Funções de Avaliação
$\pi$	Política	
		$\phi$ Taxa de Estruturação
$\emptyset$	Conjunto vazio	$\omega$ Taxa de Observabilidade
$t$	Instante no tempo	$\partial$ Taxa de Determinismo
$i$	Índice de um elemento num conjunto	
$n$	Tamanho de um conjunto	

$\varepsilon$	Parâmetro de Curiosidade	$prob()$	Probabilidade de um fenômeno
$\alpha$	Parâmetro de Tamanho Máximo	$dom()$	Domínio de uma variável
		$rel()$	Subconjunto relevante
$\Psi = \{\Psi_1, \Psi_2 \dots\}$	Árvores de Antecipação		
$\mathcal{K} = \{\mathcal{K}_1, \mathcal{K}_2 \dots\}$	Árvores de Deliberação	IA	Inteligência Artificial
$\mathcal{L} = \{\mathcal{L}_1, \mathcal{L}_2 \dots\}$	Árvores de Avaliação		
		MDP	Processo de Decisão de Markov
$\Lambda = \{\Lambda_1, \Lambda_2 \dots\}$	Lista de Diferenciadores	POMDP	MDP Parcialmente Observável
$\Theta = \{\Theta_1, \Theta_2 \dots\}$	Nós Intermediários	FMDP	MDP Fatorado
$\Xi = \{\Xi_1, \Xi_2 \dots\}$	Esquemas	FPOMDP	MDP Parcialmente Observável e Fatorado
$\mathfrak{d} = \{\mathfrak{d}_1, \mathfrak{d}_2 \dots\}$	Decisores	DBN	Rede Bayesiana Dinâmica
		D-	Determinístico
$\mathbb{W} = \{\mathbb{W}_1, \mathbb{W}_2 \dots\}$	Memórias Episódicas	PD-	Parcialmente Determinístico
$\mathbb{M} = \{\mathbb{M}_1, \mathbb{M}_2 \dots\}$	Lista de Memórias Recentes		
$e$	Expectativa		
$v$	Valor Afetivo		
$\rho$	Confiabilidade		

## LISTA DE DEFINIÇÕES

Definição 2.1: Sistema Acoplado Agente-Ambiente ( $\mathcal{AE}$ ).....	46
Definição 2.2: Agente ( $\mathcal{A}$ ).....	51
Definição 2.3: Corpo ( $\beta$ ).....	53
Definição 2.4: Mente ( $\mu$ ).....	63
Definição 3.1: Processo de Decisão de Markov.....	83
Definição 3.2: Processo de Decisão de Markov Parcialmente Observável.....	85
Definição 3.3: Processo de Decisão de Markov Fatorado.....	87
Definição 3.4: Processo de Decisão de Markov Fatorado e Parcialmente Observável.....	95
Definição 3.5: Árvore de Antecipação.....	120
Definição 3.6: Esquema.....	121
Definição 3.7: Árvore de Deliberação.....	139
Definição 4.1: Variáveis do problema wepp, no nível do ambiente.....	149
Definição 4.2: Conjuntos de propriedades corporais do agente no problema wepp.....	151
Definição 4.3: Função de evolução do corpo no problema wepp.....	152

## LISTA DE ALGORITMOS

Algoritmo 3.1: Método principal do CALM, que descreve o funcionamento básico do mecanismo.....	112
Algoritmo 3.2: Método de atualização da memória episódica generalizada.....	114
Algoritmo 3.3: Método para seleção dos testes de diferenciação (diferenciadores).....	118
Algoritmo 3.4: Método de aprendizagem de modelo de mundo, que atualiza as árvores de antecipação.....	127
Algoritmo 3.5: Método de inicialização das estruturas do conhecimento.....	129
Algoritmo 3.6: Determinação inicial das expectativas.....	130
Algoritmo 3.7: Método de integração de subárvores.....	132

## Um Mecanismo Construtivista para Aprendizagem de Antecipações em Agentes Artificiais Situados

### RESUMO

(13) 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. 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. O agente, por sua vez, é composto por dois subsistemas, mente e corpo, seguindo os princípios de situatividade e motivação intrínseca. 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.

(14) **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).

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

### RÉSUMÉ

(15) 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. 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. L'agent, à son tour, est composé de deux sous-systèmes, l'esprit et le corps, suivant les principes de la situativité et de la motivation intrinsèque. 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).

## A Constructivist Anticipatory Learning Mechanism for Situated Artificial Agents

### ABSTRACT

(17) 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. 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. The agent is composed of two sub-systems, mind and body, following the principles of situativity and intrinsic motivation. 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.

(18) **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. INTRODUÇÃO

(19) A tecnologia é historicamente um dos grandes motores da transformação do nosso modo de vida. Embora engendre uma grande discussão política, é fato que a tecnologia abre possibilidades e permite que gradativamente a sociedade possa produzir mais, trabalhando relativamente menos.

(20) No momento presente, a tecnologia encontra seus limites nos domínios que exigem autonomia, inteligência e adaptação dinâmica. Na era da informação, são justamente essas as tarefas que mais ganham importância e que ainda exigem quase na integralidade o trabalho humano.

(21) Tanto para a robótica quanto para o desenvolvimento de softwares que habitam ambientes virtuais, o desafio principal é o mesmo: criar agentes artificiais capazes de aprender a agir adequadamente em universos complexos, apresentando flexibilidade, iniciativa, criatividade e entendimento. Não é possível tratar no nível de engenharia as inúmeras possibilidades, desdobramentos e situações inesperadas que podem surgir nesse tipo de domínio.

(22) A revolução tecnológica recente permitiu confiar às máquinas não só o trabalho bruto da produção industrial, mas também operações de precisão, análise de dados, processamento de informação, etc. É possível citar incontáveis domínios e problemas que atualmente são tratados através do uso da computação. Entretanto, na quase totalidade dos casos, o computador é utilizado como uma ferramenta sob controle humano. Mesmo onde existe inteligência artificial, observa-se pouca ou nenhuma autonomia da máquina, e uma forte restrição do domínio de aplicação do sistema, que geralmente limita-se a problemas bem específicos em ambientes bem controlados. Justamente quando se exige algum nível de compreensão e discernimento, então as máquinas atuais fracassam.

(23) Neste contexto de ênfase às tecnologias da informação, torna-se ainda mais importante que a comunidade científica da Inteligência Artificial (IA) dedique esforços na busca de sistemas mais robustos e autônomos. Encontrar respostas para tais anseios necessita sem dúvida da pesquisa de base sobre mecanismos de inteligência artificial geral, que é indissociável de uma revisão paradigmática da disciplina.

(24) No caso da Inteligência Artificial, além da motivação prática para o desenvolvimento deste tipo de tecnologia, existe também um motivo de caráter mais teórico e filosófico. A inteligência é uma das características mais marcantes da nossa identidade enquanto seres humanos, e é no entanto um fenômeno ainda pouco compreendido. O questionamento sobre o que é e como funciona a inteligência está presente em várias disciplinas sob diversos aspectos, desde a filosofia, passando pela psicologia, matemática, biologia e sociologia, e para o qual a IA tem uma contribuição importante a oferecer. Nas palavras de McCorduck (1979): *'fazer avançar a IA enquanto ciência significa também ajudar a descobrir algo importante sobre nós mesmos'*.

## **1.1. Postura Epistemológica**

(25) Uma vez que esta pesquisa também aborda questões de natureza filosófica da Inteligência Artificial, julgamos ser conveniente declarar desde o início nossa postura epistemológica, esclarecendo assim a concepção adotada sobre a disciplina e a consequente linha de desenvolvimento do trabalho, que está ligado ao princípio de que a IA pode produzir formas de inteligência verdadeira.

### **1.1.1. A IA Como Disciplina Científica**

(26) Ao se retratar a história da Inteligência Artificial é possível identificar três períodos principais (McCorduck, 1979), (Simons, 1984), (Crevier, 1993), (Coelho, 1996), e (Russell; Norvig, 1995): (1) em seus primórdios, nas décadas de 1940 e 1950, a IA nasce como uma ciência envolvida em promessas grandiosas; (2) depois, pela falta de resultados contundentes, a pesquisa acaba reorientando-se por um discurso reticente e menos deslumbrado, buscando recolocar os pés no chão e apresentar resultados práticos; e finalmente, (3) como consequência de seus próprios avanços e também de novas descobertas provenientes de suas fronteiras interdisciplinares, a IA

volta pouco a pouco, nas décadas de 1990 e 2000, a resgatar seus ideais e ambições originais.

(27) A IA foi estabelecida enquanto disciplina científica durante a década de 1950. Havia naquele momento histórico um contexto propício para seu florescimento (McCORDUCK, 1979). De um lado a vontade de compreender o fenômeno da inteligência, inquietação contínua do pensamento humano moderno. De outro lado o surgimento e desenvolvimento dos computadores, propiciando, no campo acadêmico, o encontro das ciências humanas, biológicas e matemáticas, com a incipiente e promissora ciência da computação.

(28) Os primeiros tempos de existência da IA davam lugar a projetos ambiciosos, semelhantes às imagens criadas pela ficção científica. Os pesquisadores, contagiados por uma grande sensação de otimismo proveniente de êxitos iniciais, encarregaram-se de vislumbrar possibilidades e estabelecer as grandes metas e sonhos da IA.

(29) Nos anos que se seguiram, as dificuldades, os fracassos e a falta de resultados convincentes, agravado pela falta de ferramentas e conhecimentos sobre cognição, frustraram aquele otimismo inicial. O empenho dos pesquisadores não foi suficiente para evitar a constatação de que reproduzir no computador o fenômeno da inteligência era uma tarefa muitíssimo mais árdua do que se imaginava (McCORDUCK, 1979).

(30) Nesse panorama, um debate a respeito dos fundamentos da IA tornou-se inevitável, dando origem a duas correntes (HAUGELAND, 1985): de um lado os *entusiastas*, para quem o advento dos computadores mentalizados era apenas uma questão de tempo, e de outro lado os *zombadores*, para quem uma tal ideia parecia inviável ou mesmo ridícula.

(31) Essas duas concepções antagônicas ficaram conhecidas como “IA Forte” e “IA Fraca” (SEARLE, 1980). Do lado da *IA Forte* estavam os pesquisadores que acreditavam que seus programas de computador, mesmo rudimentares, eram de fato inteligentes. Do lado da *IA Fraca* ficaram aqueles que afirmavam que uma máquina poderia no máximo simular comportamentos inteligentes.

(32) Os pioneiros da IA acreditavam que os princípios da inteligência já estariam presentes nos seus modelos, e que era questão apenas de encontrar os parâmetros corretos. Davam força às analogias do cérebro como um computador, e da mente como

um programa. Entretanto, sua insistência em afirmar que o modelo do processamento de informações (computação algorítmica) era de fato uma boa metáfora para o pensamento humano, e portanto para a inteligência, deu força para que críticos como Searle (1980), Dreyfus (1972, 1992), e Penrose (1989), pudessem postular a impossibilidade da realização de uma inteligência artificial verdadeira.

(33) Pode-se acrescentar a esta análise o fato que os pioneiros da IA proveem, na sua maioria, das ciências matemáticas e físicas e, em geral, ou não tinham uma grande cultura em ciências humanas (psicologia, filosofia), ou tomavam de maneira equivocada, consciente ou inconscientemente, os resultados dessas disciplinas. Eles não puderam, portanto, tirar solidamente proveito dos resultados já obtidos e das teorias já formuladas.

(34) Na prática, a necessidade de obter resultados concretos fez com que a IA passasse por um processo de fechamento e especialização: a grande maioria dos cientistas voltou seus esforços para o aperfeiçoamento de técnicas e soluções. A velha vontade de desvendar os mistérios gerais da inteligência foi sendo substituída por pesquisas específicas e menos ambiciosas. Esse fechamento teve seu ápice na década de 1970, quando o pragmatismo tomou o lugar dos sonhos da década de 1950 e 1960, e levou os pesquisadores a uma busca por realizações (COELHO, 1996).

(35) O pragmatismo vivido pela IA nesse período teve seu lado positivo, pois finalmente os conhecimentos adquiridos pela disciplina passavam a ser, pouco a pouco, utilizados na prática. O trabalho continuou e o debate foi deixado em segundo plano, a tal ponto que, a partir da década de 1980, projetos relacionados a modelos gerais de IA chegaram mesmo a ter uma reputação negativa (PENNACHIN; GOERTZEL, 2007). Como efeito colateral, esse pragmatismo desarticulou a IA enquanto um grande projeto de compreensão da inteligência, e concebeu toda uma geração de cientistas ausentes das discussões epistemológicas e céticos quanto aos sonhos colocados por seu passado (COELHO, 1996).

(36) Entretanto, o campo da Inteligência Artificial vivenciou, nas décadas recentes, um ininterrupto e acelerado desenvolvimento de suas pesquisas, aliado ao aparecimento de novas concepções, modelos e técnicas. Tudo isso somado ao aumento da capacidade

e velocidade dos computadores tem propiciado um gradativo ressurgimento de projetos de IA com propósitos gerais (PENNACHIN; GOERTZEL, 2007), (FRANKLIN, 2007).

### 1.1.2. IA Geral

(37) A partir da década de 1990, caíram em desuso os termos “IA Forte” e “IA Fraca”. Reconheceu-se que os modelos clássicos da IA são limitados e portanto a velha IA não poderia mesmo chegar a ser uma IA forte. O debate, entretanto, não foi deixado de lado. Permanece viva a ambição de construir artefatos artificiais que apresentem inteligência genuína, então chamada “IA Geral”, em oposição à “IA Restrita” (*Narrow*), (PENNACHIN; GOERTZEL, 2007), (FRANKLIN, 2007).

(38) Os sistemas computacionais projetados para apresentar inteligência em problemas limitados e especializados fazem parte da IA Restrita. Incluem-se nessa categoria, por exemplo, programas que jogam xadrez, fazem diagnósticos, guiam dispositivos móveis através de espaços mapeados, aprendem a classificar e-mails, etc. Diferentemente, os projetos de IA Geral procuram desenvolver mecanismos capazes de agir de forma inteligente e autônoma, independentemente do domínio ou tarefa a que vão ser submetidos.

(39) Na Inteligência Artificial, estas duas posturas epistemológicas (IA Geral e IA Restrita) estão lado a lado nos laboratórios, compartilham o mesmo ferramental teórico e prático, mas divergem porque têm ambições e pressupostos filosóficos distintos.

(40) A *IA Restrita* corresponde ao fazer científico da grande maioria dos pesquisadores da área. São as pesquisas em ciência da computação voltadas à solução de problemas que exigem métodos inteligentes, sem estar necessariamente preocupadas com desdobramentos em termos filosófico-teóricos a respeito da realização de uma verdadeira inteligência artificial. Trata-se então de reproduzir estratégias da inteligência no intuito de que as máquinas possam resolver determinados problemas para os quais não basta apenas poder computacional, mas que exigem boas estratégias de solução.

(41) Do outro lado, a *IA Geral* não tem interesse em simplesmente implementar métodos engenhosos para resolver problemas aplicados. Os cientistas ligados a esta linha inquietam-se com as questões sobre a mente e a inteligência. Suas pesquisas têm como pressuposto a ideia de que a inteligência, embora sendo um fenômeno

extraordinário e complexo, é passível de ser compreendida e sistematizada. A IA Geral sustenta a hipótese de que as máquinas e os sistemas artificiais podem de fato apresentar inteligência como uma característica própria. Sua ambição maior é utilizar o computador como ferramenta para ajudar a desvendar os mistérios da inteligência em geral, desenvolvendo mecanismos artificiais que apresentem comportamentos inteligentes como resultado inerente ao seu próprio funcionamento.

(42) Embora o desafio seja antigo, apenas recentemente a comunidade científica da IA tem dedicado esforços continuados e cada vez maiores na busca de mecanismos gerais de inteligência artificial (PENNACHIN; GOERTZEL, 2007). Questões fundamentais estão novamente em evidência, por exemplo, de que forma um agente computacional pode aprender com as próprias experiências, de forma autônoma, a partir de um ciclo de interação sensório-motora com seu ambiente, construindo e reconstruindo hipóteses sobre os fenômenos que observa, identificando regularidades, formulando sua própria linguagem de representação, desenvolvendo noções abstratas e temporais sobre sua realidade, e integrando o conhecimento construído em uma rede de sistemas conceituais.

(43) Até o presente momento a IA não conseguiu construir um mecanismo de inteligência geral convincente, entretanto, apesar da sabida imensa dificuldade, não existe nenhuma evidência que desacredite a possibilidade de tal realização. Pelo contrário, o interesse cada vez maior que se verifica no domínio da IA Geral reforça a relevância deste campo de pesquisa.

(44) Nosso trabalho em particular também se soma a esse movimento. Acreditamos na viabilidade do empreendimento da IA Geral, e compartilhamos de seus pressupostos e ambições. Esperamos que esta tese possa constituir uma contribuição, representando um passo adiante na busca por mecanismos artificiais realmente inteligentes.

### 1.1.3. Objetivos e Organização da Tese

- (45) O primeiro objetivo desta pesquisa é realizar uma definição do conceito de *agente autônomo*. Não se trata de re-editar uma discussão sobre termos do senso comum, mas sim de reconstruir a relação entre agente e ambiente sob os princípios de dois paradigmas que nos últimos anos ganharam respeito e reconhecimento entre os pesquisadores da IA: são a *IA Situada* e a *IA Afetiva*.
- (46) Assim, a primeira parte desta tese é dedicada a uma ampla revisão sobre o tema. O capítulo 2 promove uma análise sobre o conceito de agente autônomo, e apresenta as principais discussões promovidas por trabalhos recentes relacionados ao assunto. Como contribuição, o capítulo 2 descreve a arquitetura CAES (*Coupled Agent-Environment System*), uma arquitetura geral para a descrição de sistemas baseados na dicotomia agente-ambiente, definindo-os enquanto sistemas parcialmente abertos e interdependentes, promovendo assim uma unificação das visões situada e afetiva da IA.
- (47) O segundo – e maior – objetivo desta tese é propor uma nova solução para o problema da *aprendizagem de modelos de mundo*. Trata-se de dotar um agente artificial de um mecanismo capaz de fazê-lo construir gradativamente uma descrição das regularidades do ambiente em que está inserido. Esta tese adota o *modelo construtivista* como base para a definição de um mecanismo deste tipo.
- (48) Assim, o capítulo 3 traz uma breve síntese da teoria psicológica construtivista, e faz referência a alguns trabalhos da IA ligados ao paradigma construtivista. O mecanismo CALM (*Constructivist Anticipatory Learning Mechanism*) é apresentado, então, como contribuição para o problema da aprendizagem de modelos de mundo, com a capacidade de aprender regularidades sensório-motoras. Ainda no capítulo 3, uma definição detalhada dos problemas de aprendizagem é apresentada, relacionando-os aos *Processos de Decisão de Markov* (MDP) e suas variantes.
- (49) Em particular, o mecanismo CALM se aplica à aprendizagem incremental da estrutura de um MDP fatorado e parcialmente observável. Realizar essa tarefa implica resolver também o problema da *seleção de propriedades relevantes*, e o *dilema da exploração e desempenho*. De outro lado, uma vez construído o modelo de mundo, é preciso que o agente possa usá-lo para de fato agir de forma adaptada no ambiente, o

que o leva ao *problema da decisão sequencial*. Todos estes subproblemas são discutidos igualmente no capítulo 3.

(50) Ainda no capítulo 3 descrevemos como o mecanismo CALM trata *ambientes parcialmente observáveis* através da indução de novos elementos de representação. Essa funcionalidade permite ao agente descobrir propriedades ocultas do ambiente, ou mesmo representar conceitos abstratos. Essa restrição (observabilidade parcial) torna o problema de aprendizagem muito mais difícil.

(51) Finalmente, no capítulo 4, apresentamos exemplos e resultados experimentais dos modelos definidos nesta tese (CAES e CALM) em diversos cenários, mostrando sua capacidade de convergência, e fazendo comparações com trabalhos relacionados. A tese é encerrada com as conclusões, onde fazemos a análise das contribuições do trabalho.

## 1.2. Construtivismo e Cibernética

(52) Tendo o objetivo de definir a arquitetura CAES e o mecanismo CALM, esta tese permeia alguns temas de relevância da área de *Inteligência Artificial*. De um lado, a busca de uma solução construtivista, e portanto o encontro com outros trabalhos da área cuja inspiração é a mesma, dando forma ao que poderia ser chamado de *paradigma construtivista da IA*. De outro lado, a inevitável incursão pelos modelos da *IA Situada*, e então uma análise sobre a composição da relação agente-ambiente, que remonta a discussões cibernéticas clássicas.

(53) Essa discussão de caráter teórico é necessária a qualquer trabalho que se deseje envolvido em um projeto de IA Geral, mas ela, sozinha, perde-se em abrangência. Ao lado dela, portanto, é preciso conduzir um trabalho em profundidade, que no caso desta pesquisa, recai sobre o domínio de aprendizagem em *Processos de Decisão de Markov*. Nesse contexto, é preciso chegar ao estado da arte dos trabalhos relacionados ao problema da construção de modelos de mundo, bem como das soluções para o problema da decisão sequencial.

(54) O passo determinante para poder caracterizar este trabalho como relevante e original, entretanto, é a abordagem do problema da aprendizagem em ambientes parcialmente observáveis. Isto não somente porque o grau de dificuldade implicado neste tipo de problema é maior, mas também porque a capacidade de induzir e

representar propriedades não-observáveis do ambiente é uma forma de começar a trabalhar com conceitos abstratos, constituindo portanto um caminho para se alcançar formas mais robustas de inteligência artificial.

### 1.2.1. Inteligência Artificial Construtivista

(55) Uma das características mais importantes da inteligência é a capacidade de aprender. Entretanto, o processo de aprendizagem é complexo, manifesta-se em múltiplas formas, e ainda vai exigir um grande esforço de pesquisa em diversas áreas da ciência até que seja completamente compreendido.

(56) A teoria psicológica construtivista (PIAGET, 1936, 1937, 1945, 1947, 1964, 1967, 1975), (MONTANGERO; MAURICE-NAVILLE, 1994), (BODEN, 1979), (FLAVELL, 1967), (COHEN, 1983) formula explicações bem aceitas a respeito do processo de aprendizagem. Segundo esta teoria, o grande potencial da inteligência humana é a capacidade de transformar suas próprias estruturas intelectuais, aprimorando-as gradativamente, num processo de complexificação, pela constante necessidade de compreender o mundo com o qual interage.

(57) O construtivismo propõe a noção de *desenvolvimento cognitivo*, ampliando o conceito de aprendizagem. O desenvolvimento cognitivo é o processo que leva o sujeito a construir novos modelos de compreensão, criar novas ferramentas intelectuais, aumentar a elaboração das estruturas do conhecimento, enriquecer as formas de representação, para então poder tratar de modo mais eficiente a complexidade das experiências.

(58) Segundo o modelo construtivista, o potencial da inteligência humana advém da passagem da inteligência sensório-motora à inteligência simbólica. O sujeito ultrapassa a fronteira da compreensão que se baseia apenas em sensações e movimentos, e passa então a organizar suas experiências em um universo formado por relações, conceitos, e objetos abstratos.

(59) Problemas de *aprendizagem de máquina* estão colocados à IA desde seus primórdios. Um sistema artificial apresenta aprendizagem se ele é capaz de se transformar adaptativamente (SIMON, 1983). Entretanto, a abordagem construtivista da IA somente veio a ser definitivamente estabelecida na década de 1990. A referência

mais importante é o trabalho precursor de Gary Drescher (1991), que utiliza o modelo construtivista para propor um mecanismo capaz de aprender de forma autônoma a realizar tarefas complexas.

(60) A teoria construtivista é bem aceita pelos pesquisadores da área de psicologia do desenvolvimento. Trata-se de um modelo filosoficamente consistente, aportando uma descrição detalhada do processo cognitivo, fundada em base experimental, e superando tanto teorias inatistas quanto empiristas da inteligência. Piaget e seu grupo pesquisaram o processo de aprendizagem através da observação de milhares de crianças, desde seu nascimento até o começo da idade adulta.

(61) A abordagem construtivista da inteligência artificial retoma o diálogo com uma teoria psicológica de grande porte, resintonizando a pesquisa em IA com o trabalho levado adiante no domínio da psicologia.

### **1.2.2. Relação Agente-Ambiente**

(62) Outra razão que faz do paradigma construtivista uma proposta interessante para a IA é a possibilidade de reconciliação entre os processos biológicos da inteligência e os processos psicológicos de alto nível, como o raciocínio e a representação. A teoria de Piaget estabelece que, em termos funcionais, há uma continuidade entre o biológico e o psicológico, uma vez que são duas dimensões que procuram tornar o sujeito adaptado a seu meio.

(63) Diversos trabalhos como (VARELA et al., 1991), (BEER, 1995, 2004), (BICKHARD; TERVEEN, 1995), (BICKHARD, 2000, 2009), (BARANDIARAN; MORENO, 2006, 2008), (RUIZ-MIRAZO; MORENO, 2000, 2004), (CLANCEY, 1997), (CLARK, 1998), (FROESE; ZIEMKE, 2009), (QUINTON et al., 2008) afirmam a necessidade de se utilizar, na IA, arquiteturas de agente artificial mais próximas de modelos naturais, orgânicos, e finalmente, repensando as relações entre agente e ambiente.

(64) O problema do paradigma clássico da IA Simbólica é seu excessivo cognitivismo, baseado em representações alheias à ideia de um agente situado em um ambiente. Um agente sem corpo, ignorando a noção de regulação dinâmica e feedback sensorio-motor, gera um questionamento sério sobre de onde vem a significação dessas

representações, situação conhecida como problema do ancoramento dos símbolos (*symbol grounding problem*), (SEARLE, 1980), (HANARD, 1990).

(65) Algumas tentativas radicais de eliminar o problema do ancoramento dos símbolos levaram a propostas *anti-representacionistas*, onde toda a cognição estaria localizada no acoplamento entre agente e ambiente, no substrato físico e biológico dessa relação, e na emergência de comportamentos dinâmicos (BROOKS, 1991), (VAN GELDER, 1998). No entanto, esta postura radicalmente inversa é também frágil, encontrando dificuldade para explicar os fenômenos cognitivos e comportamentais de mais alto-nível.

(66) De um ponto de vista filosófico, o paradigma construtivista da IA se propõe a superar tanto o modelo simbólico clássico, quanto o modelo anti-representacionista. A teoria construtivista está em sintonia com noções de situatividade e corporificação, descrevendo um mecanismo de desenvolvimento cognitivo onde as estruturas mais abstratas ou simbólicas são construídas a partir das interações sensório-motoras mais simples, num processo de gradativa complexificação da inteligência.

(67) Ao mesmo tempo, um dos desafios mais importantes a ser enfrentado pela IA é justamente a superação dos limites da percepção sensorial direta. Os fenômenos interessantes do mundo, em sua maior parte estão relacionados a processos, propriedades e objetos "macroscópicos" (THORNTON, 2003). Assim um agente artificial realmente inteligente deve ser capaz de perceber o mundo a partir de conceitos de um nível mais alto de abstração, formulados de maneira autônoma, indo ao limite em que ele possa de fato criar noções e teorias ao mesmo tempo mais gerais e mais adaptadas para descrever a realidade em que vive.

### **1.3. Desafios da Aprendizagem**

(68) Em inteligência artificial, o problema geral da aprendizagem está decomposto em uma série de subproblemas bem delimitados. Nesta tese, procuramos tratar simultaneamente alguns deles.

### 1.3.1. Aprendizagem de Modelos de Mundo

(69) O principal problema tratado nesta tese é o da *aprendizagem de modelos de mundo*. Quando o agente não conhece as regras que determinam a dinâmica de funcionamento do seu ambiente, então ele precisa descobri-las pouco a pouco baseado em suas próprias observações. Aprender um modelo de mundo significa, para um agente, construir de forma autônoma uma representação interna da dinâmica de interação com o ambiente a partir da sua experiência. Especialmente numa abordagem construtivista, esta aprendizagem deve ser feita de forma progressiva, onde um modelo mais grosseiro e menos adaptado é gradativamente aprimorado.

(70) De uma forma geral, a entrada para este tipo de problema é um fluxo ininterrupto de *percepções* sucessivas feitas pelo agente através de seus sensores, descritas no espaço definido pelas propriedades do ambiente que ele é capaz de observar, paralelo a um fluxo de *ações* executadas pelo agente através de seus atuadores. A tarefa do algoritmo de aprendizagem de modelos de mundo é induzir uma estrutura tal que seja possível prever as percepções futuras, com base nas ações e percepções atuais. Um modelo de mundo é, portanto, um modelo antecipatório, que descreve, a partir do ponto de vista do agente, a regularidade das transformações das propriedades do ambiente ao longo do tempo, segundo a observação que faz e as ações que executa.

(71) O problema da aprendizagem de modelos de mundo engloba em si outro subproblema crucial, que é o da *seleção de propriedades relevantes*. Em ambientes complexos, existe um número muito grande de percepções implicadas na descrição dos estados. Se o ambiente é bem estruturado, então apenas uma pequena parcela dessas propriedades é relevante para descrever, a cada vez, a dinâmica das transformações.

### 1.3.2. Aprendizagem de Conceitos

(72) Quando se está utilizando uma arquitetura agentificada, a aprendizagem de conceitos pode ser considerada como uma primeira etapa no sentido de construir novos elementos de representação. O conceito, nesse sentido clássico, é uma forma de descrição generalizada das experiências. Ele permite que o agente formule e processe asserções em relação a classes de situações, escapando assim de descrições exaustivas,

baseadas na referência a estados enumerados, definidos a partir das possíveis combinações de valores dos sinais da percepção.

(73) A ideia de "aprendizagem de conceitos" está bem estabelecida na comunidade científica da IA, já tendo sido pesquisada através de diversas abordagens. A proposição clássica do problema define "conceito" como uma descrição que particiona o espaço de características. Ele é chamado de "agrupamento" (*cluster*), no caso da aprendizagem não-supervisionada, ou de "classe", no caso de aprendizagem supervisionada.

(74) Os agrupamentos são conjuntos de entidades similares, ou seja, entidades que possuem um padrão de características comuns. Os agrupamentos podem ser vistos como regiões contínuas de um espaço de características contendo uma alta densidade relativa de pontos, separada de outras regiões como esta por regiões contendo baixa densidade relativa de pontos. Assim, o problema de aprender agrupamentos (clusterização) significa analisar os padrões das características dos exemplos e sua distribuição no espaço de características, e induzir uma divisão deste espaço.

(75) Diferentemente, as classes dividem a totalidade das instâncias que aparecem no espaço em subconjuntos marcados com o rótulo da classe. Enquanto a clusterização analisa padrões dos próprios dados de entrada, a classificação recebe exemplos pré-classificados. O problema da classificação significa analisar as características desse conjunto pré-classificado de exemplos de treinamento e induzir uma descrição generalizada para cada classe.

(76) O problema é que numa análise mais rigorosa, classificação e clusterização não superam de fato o nível perceptivo, pois a descrição das classes e dos clusters é formada ainda com referência direta aos elementos perceptivos. É um tipo de "conceito" derivado da combinação das próprias características fornecidas pela percepção sensorial. Nessa perspectiva, a aprendizagem de conceitos no sentido clássico não fornece capacidade ao agente para criar elementos radicalmente novos de representação.

### **1.3.3. Invenção de Conceitos Abstratos**

(77) Fazemos a distinção entre o "conceito" no sentido clássico da IA (classes e agrupamentos), e aquilo que chamamos "conceito abstrato", para falar de maneira geral de tudo o que está além das referências diretamente perceptivas.

(78) O problema fundamental da invenção de conceitos abstratos é a necessidade de definir elementos de representação radicalmente novos, elementos que designem entidades diferentes de tudo aquilo que pode ser representado em função das percepções. Os sistemas de aprendizagem convencionais da IA geram seus conceitos utilizando combinações, especializações ou generalizações das percepções sensoriais diretas (DRESCHER, 1991).

(79) Em contraste, o desenvolvimento cognitivo exige formas de invenção mais criativas. É necessário que o agente possa construir elementos novos pela necessidade de organizar suas experiências em níveis mais altos. É por isso que, em algum passo do desenvolvimento, os conceitos não podem mais ser produtos da combinação de percepções, mas ao contrário, devem ser elementos originais que reinterpretem as experiências em sistemas de compreensão mais abstratos.

(80) Assim, neste trabalho, tratamos com o que acreditamos ser o próximo passo, além da formação de classes e agrupamentos, que é a indução de propriedades abstratas, tratada sob a forma de descoberta de elementos ocultos. Em um ambiente parcialmente observável existem propriedades determinantes para a modelagem da dinâmica dos acontecimentos, as quais, porém, não podem ser diretamente percebidas pelo agente. Ambientes parcialmente observáveis podem exibir dinâmicas aparentemente arbitrarias e não-determinísticas na superfície, mesmo que sejam de fato determinísticos com relação ao sistema subjacente e parcialmente oculto que dá origem a face perceptiva dos fenômenos (HOLMES; ISBELL, 2006).

(81) Qualquer tipo de classificação das experiências baseada apenas nas percepções diretas não é capaz de modelar a dinâmica de um ambiente parcialmente observável. Para tratar este tipo de problema, o agente precisa de um mecanismo mais robusto. Uma solução é dotar o agente com um poder maior de representação, capacitando-o a suportar a existência de elementos ocultos (não-observáveis), os quais, uma vez criados, permitam a construção de um modelo de mundo adequado e, conseqüentemente, ampliem a capacidade do agente para antecipar os acontecimentos.

(82) A descoberta de elementos ocultos não encerra o problema da invenção de conceitos abstratos, porém faz-se necessário reconhecer que a possibilidade de se tratar elementos ocultos representa um avanço no caminho entre a mera percepção direta e

formas mais abstratas de compreensão da realidade, constituindo um passo de pesquisa importante para o desenvolvimento de mecanismos de inteligência artificial geral.

#### 1.3.4. Problema da Decisão Sequencial

(83) O problema da *decisão sequencial*, nesse contexto, significa fazer com que um agente utilize seu modelo de mundo para decidir quais ações executar visando a maximizar seu desempenho. Esta otimização de desempenho é avaliada segundo disposições emocionais e afetivas internas, e numa janela de tempo de longo-prazo, levando o agente a encadear uma sequência de ações para atingir objetivos que não são imediatamente acessíveis.

(84) Trata-se de fazer com que o agente defina uma boa política de ações para o ambiente em que está inserido. O problema da decisão sequencial é comumente tratado através de algoritmos de planejamento ou de aprendizagem por reforço. Neste caso, além da interação normal com o ambiente, representada pelas entradas sensoriais e saídas motoras, o agente recebe um sinal avaliativo informando se o que ele fez foi bom ou ruim, que pode ser modelado como um sinal afetivo interno, indicando sensações negativas ou positivas para o agente.

(85) Durante sua interação com o mundo, baseando-se na observação continuada desse sinal avaliativo, a tarefa do agente é construir uma política de ações que maximize o valor médio das recompensas recebidas sobre algum horizonte de tempo maior que o imediato. Essa política orienta a decisão do agente em relação às ações que ele vai tomar dependendo da situação em que se encontra, visando não apenas alcançar recompensas imediatas, mas realizar uma sequência de decisões planejadas para obter um bom desempenho de longo-prazo.

(86) Aprender um modelo de mundo, encontrar uma boa política de ações, e, ao mesmo tempo, ter que desempenhar bem sua atividade no ambiente, maximizando os sinais afetivos internos, exige enfrentar o *dilema da exploração e do desempenho*. Neste caso, o agente deve adotar alguma estratégia de exploração, planejando não apenas sequências de ações para atingir situações positivas, mas também ações que o conduzam para situações pouco exploradas, ou em que haja interesse de se explorar.

### 1.3.5. Processos de Decisão de Markov

- (87) No domínio de aprendizagem de máquina e de planejamento, o problema da decisão sequencial é geralmente tratado através de modelagens sobre *Processos de Decisão de Markov* (MDP), um formalismo clássico, (BELLMAN, 1957), (HOWARD, 1960), e bem estabelecido, (PUTERMAN, 1994), (SUTTON; BARTO, 1998), (RUSSELL; NORVIG, 1995), (FEINBERG; SHWARTZ, 2002). Um MDP é um sistema que evolui no tempo como um autômato, portanto, a cada instante o sistema se encontra em um determinado estado, e existe uma certa probabilidade de que ele sofra uma transição para algum outro estado no instante seguinte, em função da ação tomada por um agente que está no controle.
- (88) Quando os estados do ambiente não são diretamente acessíveis ao agente através de sua percepção, então o sistema é modelado através de um MDP parcialmente observável (POMDP), (SMALLWOOD; SONDIK, 1973), (CHRISTMAN, 1992), (SHANI et al., 2005), (KAELBLING et al., 1994, 1998), neste caso a observação que o agente controlador tem do sistema não lhe informa diretamente o estado em que ele se encontra.
- (89) Recentemente o foco das pesquisas voltou-se para os MDPs fatorados (FMDP) (BOUTILIER et al., 2000), (GUESTRIN et al., 2003), (JONSSON; BARTO, 2005), (SALLANS; HINTON, 2004), e também POMDPs fatorados (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), onde os estados são representados no espaço definido por um conjunto de variáveis aleatórias.
- (90) Embora existam algoritmos “livres de modelo”, o problema da decisão sequencial é frequentemente considerado como contendo em si o pré-problema da construção de um modelo de mundo (RUSSELL; NORVIG, 1995). Ou seja, se o agente não tem acesso integral e direto ao ambiente e às suas regras de transformação, nem tampouco é fornecido de antemão um modelo deste ambiente – que é geralmente o caso em problemas reais – então o agente é obrigado a construir um modelo de mundo antes de calcular uma política de ações.

(91) Trabalhos ligados ao problema da aprendizagem de modelos de mundo têm convergido, recentemente, para o uso de FMDPs como forma de representação. Neste caso, construir um modelo de mundo significa determinar a estrutura e os parâmetros de um FMDP a partir da experiência do agente, como fazem (DEGRIS et al., 2006, 2008) e (STREHL et al., 2007).

## 1.4. Contribuições

(92) Esta tese apresenta duas contribuições. A primeira é a arquitetura **CAES** (*Coupled Agent-Environment System*), que redefine a relação entre agente e ambiente, como resultado da discussão sobre o conceito de agente autônomo, sob uma perspectiva naturalista e baseada na teoria dos sistemas dinâmicos.

(93) A segunda contribuição é o mecanismo de aprendizagem **CALM** (*Constructivist Anticipatory Learning Mechanism*), que tem a finalidade de permitir que um agente artificial construa gradativamente uma descrição das regularidades do ambiente em que está inserido.

### 1.4.1. A Arquitetura CAES

(94) A arquitetura **CAES** (*Coupled Agent-Environment System*) define agente e ambiente como dois sistemas parcialmente abertos, interativos, e em acoplamento dinâmico, onde o agente é autônomo, situado, corporificado, afetivo, e mentalizado.

(95) A arquitetura proposta define um sistema global, onde podemos distinguir três grandes entidades: (a) o *ambiente*, que representa tudo o que está fora do agente, (b) o *corpo*, que forma o universo interior do agente, possuindo propriedades e metabolismos próprios, e servindo como interface entre agente e ambiente através de sensores e atuadores, e finalmente, (c) a *mente* do agente, responsável por coordenar seu comportamento.

(96) Dessa forma, fica estabelecida como única alternativa para o agente, realizar um tipo de *cognição situada*. Primeiramente porque o acesso que a mente tem ao corpo e ao ambiente é mediado através de sinais parciais. A mente não pode perceber na íntegra as características exteriores, pois a percepção sensorial é limitada, assim como suas ações

também só têm controle sobre uma parte dos eventos que ocorrem. Além disso, o agente está restrito também a um ponto de vista e a um alcance local.

(97) A mente do agente, na arquitetura CAES, é composta por dois subsistemas: cognitivo e regulativo. O sistema cognitivo é de fato aquele que aprende e constrói um modelo antecipatório da dinâmica do mundo que o agente observa, e também é responsável por planejar e deliberar as ações. Do outro lado, o sistema regulativo inclui tanto os comportamentos reativos e emocionais, como a avaliação afetiva das experiências.

(98) A presença de um sistema avaliativo, que atribui valor afetivo aos eventos, é uma forma de internalização da motivação do agente, eliminando a noção de recompensas ambientais para as ações. O sistema avaliativo desenha uma espécie de relevo afetivo no espaço de fluxo representado pelo sistema cognitivo do agente, em seu modelo de mundo.

#### 1.4.2. O Mecanismo CALM

(99) Do ponto de vista psicológico, o mecanismo CALM (*Constructivist Anticipatory Learning Mechanism*) é proposto como um *modelo sistematizado para a psicologia construtivista*. Trata-se de uma versão computacional, ainda que livremente inspirada e não definitiva, do processo de desenvolvimento cognitivo, conforme descrito por Jean Piaget (1936, 1937, 1945, 1947, 1964, 1967, 1975) em sua pesquisa realizada entre as décadas de 1930 e 1980.

(100) Do ponto de vista computacional, o mecanismo CALM apresenta-se como solução para o problema da *aprendizagem de modelos de mundo*, ou seja, permite que um agente construa uma representação da estrutura e da dinâmica de seu ambiente, baseado na experiência. O modelo de mundo é representado através de um *Processo de Decisão de Markov Fatorado* (FMDP).

(101) O desenvolvimento do mecanismo CALM envolve também a adaptação de soluções para os problemas da *seleção de propriedades relevantes*, do *dilema da exploração e desempenho*, e da *decisão sequencial*.

(102) Finalmente, o CALM também é capaz de enfrentar *ambientes parcialmente observáveis* através da indução de novos elementos de representação, os quais podem

estar associados a propriedades ocultas do ambiente ou mesmo representar conceitos abstratos. Neste caso, o modelo construído constitui um MDP ao mesmo tempo fatorado e parcialmente observável (FPOMDP).

(103) O mecanismo opera de forma *incremental* e *online*, então o agente aprende ao mesmo tempo em que precisa interagir com seu universo. Como restrição, é suposto que o ambiente pode ser descrito através de *funções discretas*, tanto no tempo quanto nos valores que a percepção e a ação do agente podem assumir. É também suposto que o ambiente é *parcialmente determinístico* em relação à função que governa suas transformações.

## 2. CAES: SISTEMA DE ACOPLAMENTO AGENTE-AMBIENTE

(104) A partir da década de 1990, houve um gradativo ressurgimento de projetos de Inteligência Artificial (IA) com propósitos gerais (PENNACHIN; GOERTZEL, 2007). Para tanto, fez-se necessário revisar o conceito de agente autônomo e o modelo de relação entre agente e ambiente. Modelos simplificados utilizados pelo paradigma clássico da *IA Simbólica*, excessivamente cognitivistas, não-situados, incorrem no problema do ancoramento dos símbolos (*symbol grounding problem*), (SEARLE, 1980), (HANARD, 1990), no problema dos *frames* (McCARTHY; HAYES, 1969), (PYLYSHYN, 1987), e em outros problemas de natureza mais filosófica, que indicam uma certa incompatibilidade entre os modelos e a realidade.

(105) Em resposta a isso, diversos autores propuseram novas arquiteturas para agentes, utilizando ideias que ajudaram a constituir os paradigmas da *Inteligência Artificial Situada* e da *Inteligência Artificial Afetiva*. Nessa mesma linha, o presente capítulo define o CAES (*Coupled Agent-Environment System*), uma arquitetura para desenvolvimento de agentes autônomos que tentam unificar os pontos fortes destes dois paradigmas.

(106) A *IA Situada* une a teoria dos sistemas dinâmicos com modelos orgânicos e naturais. Trabalhos referenciais como (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), concebem o agente como um sistema dinâmico parcialmente aberto, imerso em seu ambiente, e interagindo com este através de um fluxo contínuo de interferências mútuas, governadas por dois fatores principais: *regulação orgânica e feedback sensoriomotor*.

(107) A *IA Afetiva* internaliza a motivação do agente. Trabalhos como (SINGH et al., 2004), (CAÑAMERO, 1997a, 1997b, 2001), (ALMEIDA et al., 2004) e (SLOMAN et al., 1999, 2005), derrubam o velho modelo do treinamento, excessivamente comportamentalista, onde o ambiente indica ações corretas através de recompensas e punições exógenas, e trazem o sinal de reforço para dentro do próprio corpo do agente. A ideia é modelar sistemas afetivos e emocionais orgânicos, fazendo com que a motivação para agir passe a ter origem interna.

## 2.1. Relação entre Agente e Ambiente

(108) A arquitetura CAES define agente e ambiente como dois sistemas parcialmente abertos e em acoplamento dinâmico, onde o agente é uma entidade autônoma, mas que está inescapavelmente imerso no ambiente. Na arquitetura CAES, o agente é: (a) situado e corporificado; (b) afetivo e emocional (portanto intrinsecamente motivado); e (c) cognitivo e adaptativo. Essas características permitem superar os modelos simplificados que a IA costuma utilizar, onde o agente reduz-se a um sistema de representação e inferência (uma “mente solta”) que atua diretamente no mundo.

(109) Um sistema construído através da arquitetura CAES apresenta de forma explícita três tipos de interação, conforme ilustra a figura 2.1: (a) mais externamente, a interação entre agente e ambiente; (b) a seguir, dentro do agente, a interação entre seu corpo e sua mente; (c) e por fim, dentro da mente, a interação entre o sistema cognitivo e o sistema regulativo.

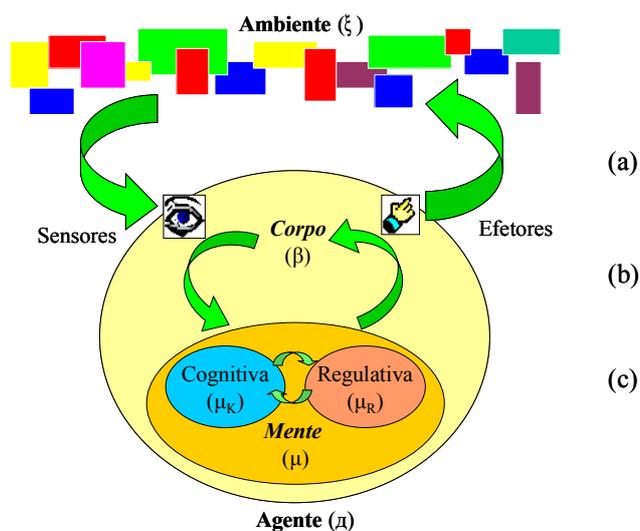


Figura 2.1: Arquitetura CAES e seus três níveis de interação.

Em (a) a relação agente-ambiente, em (b) a relação corpo-mente, e em (c) a relação cognitivo-regulativa.

(110) Assim, o agente é visto, num primeiro nível, como um *objeto* do ambiente. Nesse nível existem apenas os fenômenos e eventos que se diria de uma realidade física. O ambiente é o espaço que contém os objetos, e que impõe a estes objetos padrões de interação de uma determinada natureza. O agente, enquanto um objeto entre os outros, participa igualmente das propriedades, eventos e fenômenos físicos desse nível da realidade, estabelecidos nesse sistema mundo-objetos.

(111) Num segundo nível o agente é um *organismo*, sujeito ainda aos fenômenos e eventos físicos, mas, como um objeto especial, possui um metabolismo corporal próprio e um universo de propriedades internas. Assim o agente se configura como um sistema parcialmente aberto, podendo influenciar, a partir de processos internos, na maneira com que ocorrem os fenômenos exteriores, alterando localmente a trajetória do sistema global.

(112) Num terceiro nível, um agente é uma *entidade mentalizada*, caracterizado por possuir uma estrutura interna diferenciada, a mente, que regula seu comportamento. A mente recebe informações sensoriais do ambiente e do corpo, e coordena ações dirigidas sobre um e sobre outro.

### 2.1.1. Interatividade, Autonomia e Situatividade

- (113) A partir da década de 1980 a IA viu o surgimento de um novo paradigma, chamado de *computação baseada em agentes*. Desde então, a ideia de “agente” passou a se popularizar entre os pesquisadores da área (MAES, 1994), trazendo consigo um novo modelo conceitual. A definição mais consensual e simples para o conceito de *agente* é a de uma entidade inserida em um ambiente, percebendo-o através de um conjunto de sensores, e agindo sobre ele através de um conjunto de atuadores (RUSSELL; NORVIG, 1995).
- (114) A dicotomia *agente-ambiente* representa um modelo bastante geral, podendo descrever diversos tipos de sistema, naturais ou artificiais, como por exemplo: um veículo autônomo se deslocando em um planeta distante, um robô realizando uma tarefa doméstica em casa, um programa de computador rodando em um universo virtual, um mecanismo de busca inteligente navegando na internet, um animal em seu habitat, uma bactéria em uma solução química, uma pessoa realizando uma tarefa cotidiana no mundo real, etc. A dicotomia entre agente e ambiente em computação é equivalente àquela entre organismo e ambiente na biologia, ou sujeito e meio na psicologia.
- (115) A interatividade é a primeira propriedade a ser destacada na relação entre agente e ambiente. Assim, o modelo de computação baseado em agentes caracteriza-se por um ciclo de funcionamento ininterrupto, em lugar da arquitetura tradicional caracterizada como um fluxo linear de entrada, processamento e saída, conforme ilustra a figura 2.2. Para alguns autores, como (WEGNER, 1998), (COSTA; DIMURO, 2005), e (GOLDIN; WEGNER, 2008), esta consideração aparentemente simples propicia a emergência de um novo paradigma computacional.

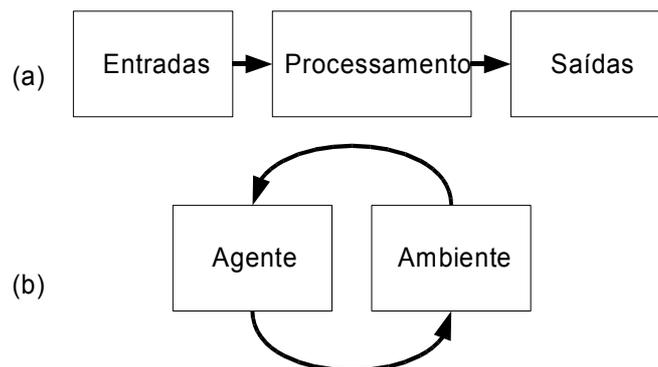


Figura 2.2: Computação baseada em agentes.

Diferença conceitual entre (a) o modelo clássico do processamento de dados e (b) o modelo de agentes.

(116) A noção de agente remete à ideia de sistema parcialmente aberto. O agente é um sistema encapsulado, possuindo uma estrutura interna, própria e finita, mas seu funcionamento depende da interação com um ambiente que lhe é exterior (JENNINGS, 2000). Essa definição coloca agente e ambiente como entidades complementares. A existência de um agente implica a existência de um ambiente correlato. São dois sistemas interdependentes constituindo um sistema global.

(117) Em sistemas naturais, definir quem é o agente e quem é o ambiente envolve uma certa arbitrariedade do observador, uma vez que pode não haver uma fronteira precisa entre ambos. Esta distinção pode ser baseada na coesão organizacional do agente, que constitui um sistema denso, relativamente estável e altamente integrado, em relação ao sistema total.

(118) Da mesma forma, distinguir entre o que é um agente e o que é um objeto do ambiente também envolve critérios subjetivos. Geralmente, a principal característica usada para diferenciar um agente de um mero objeto é o fato de que o agente apresenta comportamentos ativos, ou seja, ele é um objeto especial do ambiente, capaz de escolher, a partir de elementos internos, entre ações alternativas para uma mesma situação (JENNINGS, 2000).

(119) É possível considerar a presença de diversos agentes num mesmo ambiente, conforme ilustra a figura 2.3, caracterizando um sistema multiagente. Estes diversos agentes podem ter uma arquitetura semelhante ou podem ser muito diferentes, formando então um tipo de ecossistema. Do ponto de vista de cada agente específico, os outros agentes fazem parte do ambiente, incluídos entre os demais objetos com os quais ele

interage. Entretanto, é geralmente a existência de muitos agentes, combinando atuações autônomas, o principal fator gerador de aumento da complexidade do sistema como um todo.

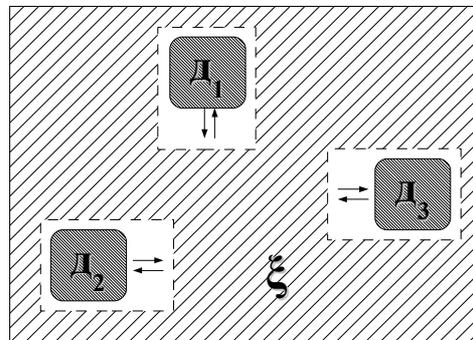


Figura 2.3: Um ambiente  $\xi$  povoado por diversos agentes  $A$ .

(120) Um agente é uma entidade autônoma, que atua em seu ambiente livre de qualquer controle externo (DAVIDSSON, 1995). Todos os processos envolvidos nas tomadas de decisão e nas iniciativas para agir devem ter origem na própria estrutura interna do agente.

(121) Um agente autônomo é uma entidade *situada*. Nessa condição, o ambiente em que ele está inserido complementa a definição do próprio agente. Os meios pelos quais agente e ambiente se relacionam e provocam interferências recíprocas (CLANCEY, 1997) tornam-se importantes para a caracterização do sistema global. O agente é situado porque suas decisões são tomadas com base no contexto em que ele se encontra, ou seja, as ações que o agente executa são resultado de um processo interno de decisão, mas em função de parâmetros provenientes também (ou principalmente) do ambiente.

(122) A situatividade (*situativity*) define um modelo de cognição chamado de “enação”. O conhecimento enativo é aquele proveniente da atividade, onde o ponto de referência não é um mundo objetivo e neutro, definido por um olhar externo, mas sim a estrutura sensório-motora do agente (VARELA et al., 1991).

(123) A interface entre mundo e agente é limitada. O agente situado não é nem *onisciente* nem *onipotente* com relação ao ambiente, ou seja, o mundo é apenas parcialmente capturável pela sua percepção sensorial, e é apenas parcialmente passível de ser transformado por suas ações (SUCHMAN, 1987).

### 2.1.2. Acoplamento

(124) A arquitetura CAES (*Coupled Agent-Environment System*) define a relação entre agente e ambiente como o acoplamento de dois sistemas dinâmicos, parcialmente abertos, e em interação. Cada um destes dois sistemas exerce uma determinada influência sobre o fluxo de transformação do outro, deformando-o continuamente. Trata-se de uma perturbação mútua, onde cada sistema não possui controle absoluto sobre a determinação da trajetória do outro, nem de sua própria.

(125) Tal concepção, denominada *modelo de acoplamento dinâmico*, foi solidamente apresentada por Beer (1995, 2004), que por sua vez resgatou a estrutura geral do *modelo cibernético* proposto por Ashby (1952), revestido de algumas ideias do *modelo autopoiético*, desenvolvido por Maturana e Varela (1973, 1980).

(126) O *modelo de acoplamento dinâmico* representa uma arquitetura recente e de boa aceitação, para a modelagem da relação entre um agente e seu ambiente. Essa abordagem organiza-se em torno de um conjunto bem estabelecido de ferramentas matemáticas, provenientes da teoria dos sistemas dinâmicos. Nesse modelo, o comportamento inteligente é concebendo-o como uma atividade adaptativa realizada por agentes situados.

(127) O modelo de acoplamento dinâmico faz parte de um movimento maior, na inteligência artificial e nas ciências cognitivas, (VARELA et al., 1991), (VAN GELDER, 1998), (CLARK, 1998), (MONTEBELLI et al., 2008), (QUINTON et al., 2008), de aproximação das teorias da vida, da interação, e da cognição, com a teoria dos sistemas dinâmicos, e o resgate de algumas noções cibernéticas, em oposição à modelos excessivamente simbólicos, no sentido representacionista clássico.

#### 2.1.2.1. Modelo Cibernético

(128) A cibernética, entre as décadas de 1930 e 1960, foi o movimento responsável pela introdução de diversos conceitos importantes para a teoria do controle e da informação, como por exemplo as noções de “equilíbrio homeostático”, “feedback” (retroalimentação), e “auto-regulação”. Por princípio, buscava-se a descrição de uma teoria geral de sistemas que poderia se aplicar tanto em contextos naturais (físicos, biológicos ou mesmo sociais), como artificiais (máquinas, autômatos, simulações, etc.).

(129) O conceito de *homeostase* (CANNON, 1932) refere-se à propriedade de um sistema aberto de regular seu ambiente interno para preservar uma determinada condição de equilíbrio. Tal conceito é inspirado nos organismos biológicos, que possuem uma série de mecanismos (metabólicos ou comportamentais) com a finalidade de manter determinadas variáveis fisiológicas em torno de “níveis normais”.

(130) O conceito de *feedback* (WIENER, 1948) representa a causalidade circular, onde uma entidade age sobre outra, que em retorno age sobre a primeira. O desdobramento dessa ideia leva aos sistemas auto-regulados, que alteram condições internas modulando-as pelo feedback externo, a fim de preservar o equilíbrio homeostático.

(131) Embora a teoria dos sistemas dinâmicos não estivesse ainda estabelecida na época, Ashby (1952) de forma pioneira, descreve a relação entre agente e ambiente como dois sistemas dinâmicos em interação (GRUSH, 1997b). Ashby afirma que um organismo animal não é diferente de uma máquina quando definido como sistema. O organismo e seu ambiente, tomados juntos, formam um sistema absoluto. O organismo afeta o ambiente e o ambiente afeta o organismo, numa relação de feedback mútuo.

#### 2.1.2.2. Modelo Autopoiético

(132) O conceito de *autopoiese* constitui uma nova definição para “vida” (MATURANA; VARELA, 1980). No modelo autopoiético, fortemente influenciado pelas noções cibernéticas, um ser vivo é um sistema aberto e autônomo. Entretanto, a ênfase do modelo está na “auto-produção” do sistema. O ser vivo, sendo um sistema autopoiético, não apenas altera o seu comportamento enquanto interage com o ambiente, mas também transforma-se estruturalmente em virtude desta interação. O organismo é caracterizado por sua organização interna, estável ao longo do tempo.

(133) O ser vivo está num processo de contínua transformação, e o curso que segue esta mudança é modulado em parte pelas interações com o ambiente. Muitas vezes eventos de origem externa podem desencadear alterações no organismo, mas o sentido dessas mudanças não é determinado pelo ambiente, pois é inerente à própria organização interna do ser vivo. É a estrutura do organismo que estabelece o domínio de mudanças que podem ter lugar nele, e também o domínio das perturbações que podem disparar nele uma mudança (MATURANA, 1993).

(134) A dinâmica do mundo exerce uma pressão que tende a afastar o organismo do seu estado de equilíbrio, e essa pressão é uma ameaça constante à integridade do sistema. A autopoiese representa a luta do organismo para que lhe seja possível continuar a viver no ambiente, compensando as perturbações externas por transformar-se estruturalmente, garantindo assim sua auto-manutenção.

### 2.1.2.3. *Sistemas Dinâmicos*

(135) A teoria dos sistemas dinâmicos fornece um ferramental matemático bem estruturado e amplamente aceito para o estudo de sistemas determinísticos que evoluem seu estado ao longo do tempo, e que podem exibir comportamentos complexos (GLEICK, 1987).

(136) Um sistema dinâmico é definido por um espaço de domínio, combinado com uma função suave que descreve a evolução do sistema. Para qualquer instante do tempo, esta função mapeia um ponto do espaço de domínio de volta para o próprio espaço. Quando o domínio é contínuo, então trata-se de um *espaço de fase*, e quando é discreto, trata-se de um *espaço de estados*. Quando o tempo é contínuo o sistema dinâmico corresponde a um *fluxo*, e quando é discreto, corresponde a um *mapa*. A função de evolução define um campo vetorial no espaço de domínio, indicando o sentido das trajetórias para todos os pontos possíveis.

(137) Num sistema dinâmico, os *atratores* são regiões de convergência, para onde evolui o fluxo do sistema quando ele entra na área de atração. Estes atratores podem ter a forma de um ponto fixo, de um limite cíclico, ou podem ser complexos, delimitando uma região para onde o fluxo é atraído, e onde ele permanece numa *flutuação caótica*. Se os parâmetros da função de evolução do sistema dinâmico variam no tempo, então a própria configuração dos atratores pode mudar também, conforme o sistema evolui.

## 2.2. Características do Sistema Global

(138) A arquitetura CAES constitui um modelo cibernético e “naturalista” de acoplamento, onde agente e ambiente são dois sistemas dinâmicos em um ciclo de interação ininterrupta, que exercem influência um sobre o fluxo de transformação do outro, deformando continuamente suas trajetórias, mas sem exercer controle absoluto. Alguns dos parâmetros da função de evolução de um sistema são variáveis relacionadas

ao estado do outro, e portanto a relação entre agente e ambiente se estabelece através de perturbações mútuas. O sistema global na arquitetura CAES é ilustrado na figura 2.4, e formalizado na definição 2.1, na sequência.

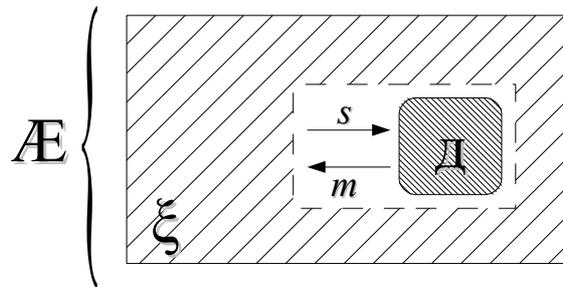


Figura 2.4: O sistema global na arquitetura CAES.

O sistema global ( $\mathcal{A}$ ) é um sistema fechado, formado por dois sub-sistemas parcialmente abertos, agente e ambiente, em acoplamento dinâmico, onde  $\mathcal{A}$  é o agente,  $\xi$  é o ambiente,  $m$  é a atuação do agente sobre o ambiente, e  $s$  é a situação que se impõe ao agente pelo ambiente.

Um sistema acoplado agente-ambiente CAES ( $\mathcal{A}$ ) é uma 4-upla:

$$\mathcal{A} = \{\mathcal{A}, \xi, s, m\}$$

Onde,

$\mathcal{A}$  é um sistema dinâmico parcialmente aberto (*agente*)

$\xi$  é um sistema dinâmico parcialmente aberto (*ambiente*)

$s$  é o sinal que define a interferência do ambiente  $\xi$  sobre o agente  $\mathcal{A}$  (*situação*)

$m$  é o sinal que define a interferência do agente  $\mathcal{A}$  sobre o ambiente  $\xi$  (*atuação*)

Definição 2.1: Sistema Acoplado Agente-Ambiente ( $\mathcal{A}$ ).

(139) O ambiente é representado por um conjunto de variáveis  $X_\xi = \{X_{\xi_1}, X_{\xi_2}, \dots, X_{\xi_n}\}$ , e uma função de evolução  $f_\xi$  que estabelece o padrão de transformações dos valores dessas variáveis. O estado do ambiente num dado instante  $t$  é representado por um vetor  $x_\xi^t \in X_\xi$ . O ambiente ( $\xi$ ) é tudo aquilo que está fora do agente, e a totalidade *agente* + *ambiente* forma o sistema global ( $\mathcal{A}$ ).

(140) A função de evolução do ambiente ( $f_\xi$ ) indica o sentido de sua transformação no futuro imediato, baseada em seu estado atual e na intervenção realizada pelo agente, portanto na forma  $f_\xi: X_\xi \times M \rightarrow X_\xi$ , onde  $M = \{M_1, M_2, \dots, M_n\}$  define o espaço das possíveis interferências do agente sobre o ambiente.

(141) Correlatamente, a função de evolução do agente ( $f_a$ ) indica a transformação da situação interna do agente de acordo com seu estado atual e com uma informação proveniente do estado do ambiente, na forma  $f_a : X_a \times S \rightarrow X_a$ , onde  $X_a$  define o espaço (de fase ou de estados) do agente enquanto sistema dinâmico, e  $S = \{S_1, S_2, \dots, S_n\}$  define toda a janela de interferência do ambiente sobre o agente.

(142) O espaço de estados do sistema global é estabelecido pelo produto cartesiano dos espaços de seus dois sub-sistemas, portanto  $X_E = X_\xi \times X_a$ . Da mesma forma, a função de evolução do sistema global,  $f_E : X_E \rightarrow X_E$ , é a combinação das funções de evolução do agente ( $f_a$ ) e do ambiente ( $f_\xi$ ).

(143) Cada ação que o agente executa afeta o ambiente de alguma maneira. A transformação do ambiente, por sua vez, afeta o agente, e isso define um ciclo permanente. Portanto, a função de evolução de um sistema, embora parametrizada principalmente por seu próprio estado, é também dependente de alguns elementos relativos ao estado do outro sistema.

### 2.2.1. Solipsismo Metodológico

(144) A condição filosófica do *solipsismo metodológico* (FODOR, 1981) afirma que a experiência do mundo é vivida internamente pelo sujeito, e então a realidade exterior não pode ser apreendida objetivamente. Ela surge como uma projeção capturada pela percepção. Trata-se de uma afirmação similar à do *fenomenalismo analítico* (DANTO, 1989), (AYER, 1954), doutrina filosófica segundo a qual só se pode conhecer efetivamente os fenômenos, ou seja, o conteúdo espaço-temporal das percepções sensoriais, único objeto possível de experiência.

(145) Assim, do ponto de vista do agente ( $\pi$ ), o ambiente ( $\xi$ ) é algo objetivamente inacessível. Ele é a outra parte na relação agente-ambiente, aquela que diz respeito ao mundo. É o sistema com o qual o agente se relaciona, sem ter controle ou acesso direto aos fatores que regem seu funcionamento. O agente não pode conhecer o mundo como ele é, não pode observar sua constituição de fato. O ambiente só é acessível para o agente por intermédio dos seus sensores e atuadores, e portanto, a única coisa que é possível para o agente conhecer é a relação de interação com o ambiente, e não o ambiente em si.

(146) Um ambiente desconhecido qualquer pode ser representado de maneira geral como um sistema dinâmico, possuindo um conjunto de variáveis ( $X_\xi$ ) que define suas dimensões, e uma função de evolução ( $f_\xi$ ) que define o padrão de transformações que ocorre nessas variáveis segundo o próprio estado do ambiente, e também em função das ações do agente, tal como é definido na arquitetura CAES.

(147) O ambiente mais “natural” que podemos imaginar é o próprio mundo real, mas agentes computacionais podem ser também inseridos em mundos artificiais, onde geralmente é possível para o experimentador conhecer o modelo do ambiente e sua complexidade. Entretanto, mesmo num ambiente simulado, onde as regras e a constituição do mundo são conhecidas por seus programadores, ainda assim do ponto de vista do agente esse ambiente é objetivamente inacessível.

(148) Tal empecilho filosófico não constitui ameaça aos algoritmos de aprendizagem de modelos de mundo, assim como não representa um problema para os seres humanos no seu desafio cotidiano de aprender a interagir com o mundo real, e a antecipar os acontecimentos. Como será visto na seção 3.1, o problema da aprendizagem de modelos de mundo pode ser adequadamente formalizado, e os aspectos pertinentes do ambiente podem ser corretamente representados pelo agente mesmo quando a referência que ele utiliza é a sua interação, e não o ambiente de fato subjacente à relação.

### 2.2.2. Situação e Atuação

(149) O agente interfere sobre o ambiente através de sua atuação, um vetor  $m \in M$ , geralmente (mas não unicamente) realizada através dos seus atuadores. A atuação é dada em função de certas propriedades do agente, na forma  $m : X_a \rightarrow M$ . O outro sentido da interferência é representado pela situação, um vetor  $s \in S$ , dada ao agente pelo ambiente, seja sob a forma de restrições e pressões ambientais, seja como elementos capturados pelos sensores do agente. Portanto a situação é definida em função de determinadas propriedades do ambiente, na forma  $s : X_\xi \rightarrow S$ .

(150) O agente percebe o mundo apenas de uma maneira parcial, o que define uma condição de não-onisciência. Isso acontece, primeiramente, porque a situação que o agente percebe é um retrato empobrecido da situação real, uma vez que as restrições do seu aparato sensorial impossibilitam a captura de todos os aspectos, dimensões, ou

características do mundo. O mecanismo de sensoriamento do agente define uma psicofísica, que funciona como um filtro entre as inúmeras condições físicas provenientes do ambiente, e os sinais sensoriais ativados no agente. Em segundo lugar, a percepção é modificada pelo ponto de vista do agente. O agente enquanto entidade situada está sujeito a uma espécie de filtro de localidade. Por exemplo, em ambientes espaciais, a posição do agente em relação aos outros objetos modifica sua percepção da situação corrente.

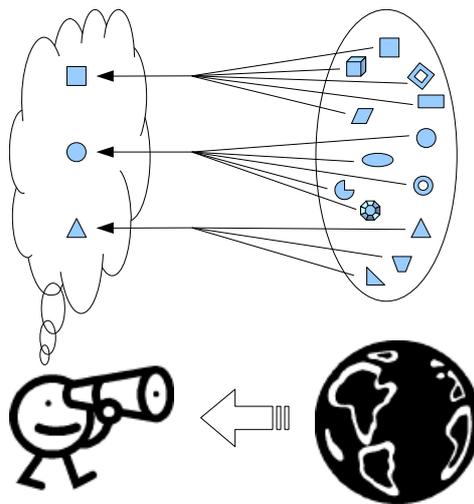


Figura 2.5: Condição de não-onisciência.

A percepção como uma função sobrejetora não-reversível, cujo domínio são os estados físicos do ambiente e o contra-domínio são as percepções sensoriais do agente.

(151) O agente não tem acesso à identificação completa do estado do ambiente, mas sim à uma informação que é derivada deste estado. Esta função de situação tem como domínio todos os possíveis estados do mundo, e como contra-domínio um conjunto (menor) de observações possíveis. Trata-se de uma função sobrejetora em que uma reversibilidade completa não é possível, ilustrada na figura 2.5.

(152) Assim, devido a existência dessas limitações sensoriais, torna-se inteiramente possível que o agente se encontre incapaz de distinguir perceptivamente entre diferentes estados do mundo que aparecem para ele como idênticos (CROOK; HAYES, 2003). Esta confusão de estados, onde situações diferentes podem ter aparência indistinta, é também chamada de *perceptual aliasing* (WHITEHEAD; BALLARD, 1991), (CHRISMAN, 1992), (WHITEHEAD; LIN, 1995), e define problemas de observabilidade parcial.

(153) Da mesma forma, devido ao fato dos atuadores também serem limitados de alguma maneira, o controle que o agente tem sobre as transformações do ambiente é apenas parcial. Esta limitação existe, de um lado, porque, em ambientes espaciais, o alcance de atuação do agente está restrito a um sub-espaco local do ambiente, geralmente às regiões próximas a ele. De outro lado, porque as transformações impostas ao ambiente, embora parcialmente influenciadas pela ação do agente, também estão condicionadas a outros fatores independentes e exteriores a ele.

### 2.3. Características do Agente

(154) Na arquitetura CAES, o agente é uma das partes num sistema de acoplamento dinâmico entre agente e ambiente. Por sua vez, o agente é ele próprio definido como a composição de dois subsistemas, corpo e mente, conforme ilustrado na figura 2.6, e formalizado na definição 2.2, a seguir.

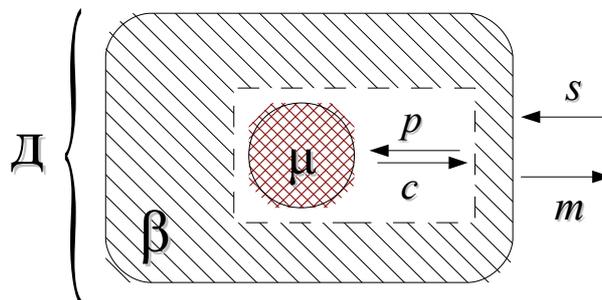


Figura 2.6: O agente na arquitetura CAES.

Um agente ( $\mathcal{A}$ ) é um sistema parcialmente aberto que interage com seu ambiente. Ele é formado por dois subsistemas, mente ( $\mu$ ) e corpo ( $\beta$ ), que por sua vez estão também relacionados em acoplamento dinâmico.

Um agente ( $\mathcal{A}$ ) em um sistema CAES é uma 6-upla:

$$\mathcal{A} = \{\beta, \mu, c, p, s, m\}$$

Onde,

$\beta$  é um sistema dinâmico parcialmente aberto (*corpo*)

$\mu$  é um sistema dinâmico parcialmente aberto (*mente*)

$c$  é a interferência de  $\mu$  sobre  $\beta$  (*controle*)

$p$  é a interferência de  $\beta$  sobre  $\mu$  (*percepção*)

$s$  é um sinal externo que o agente  $\mathcal{A}$  recebe do ambiente  $\xi$  (*situação*)

$m$  é um sinal que o agente  $\mathcal{A}$  fornece ao ambiente  $\xi$  (*atuação*)

Definição 2.2: Agente ( $\mathcal{A}$ ).

(155) O espaço (de fase ou de estados) do agente enquanto sistema dinâmico é estabelecido pelo produto cartesiano dos espaços de seus dois subsistemas,  $X_{\mathcal{A}} = X_{\beta} \times X_{\mu}$ . Da mesma forma, sua função de evolução é derivada da combinação das funções de evolução do corpo ( $f_{\beta}$ ) e da mente ( $f_{\mu}$ ), na forma  $f_{\mathcal{A}} : X_{\mathcal{A}} \rightarrow X_{\mathcal{A}}$ .

(156) Assim, segundo a definição da arquitetura CAES, o *corpo* ( $\beta$ ) do agente representa sua constituição física e orgânica (seja ela real ou virtual). Já a *mente* ( $\mu$ ) do agente é uma entidade especial, inserida nesse corpo, porém encarregada das funções cognitivas, de controle, de regulação e de comportamento. Os sinais  $s$  e  $m$ , que definem respectivamente a *situação* e a *atuação* do agente ( $\mathcal{A}$ ), são os mesmos sinais definidos no sistema global ( $\mathcal{E}$ ).

### 2.3.1. Corporificação

(157) A noção de corporificação (*embodiment*) está presente nas ciências cognitivas e em especial na IA, desde a década de 1990, tendo sido um tema imensamente debatido. Alguns artigos de referência são (FRANKLIN, 1997), (QUICK et al., 1999), (CLARK, 1998), (ZIEMKE, 1998, 2002), (ANDERSON, 2003), (CHRISLEY; ZIEMKE, 2002) e (BARANDIARAN; MORENO, 2008), (OVERTON et al., 2008).

(158) Embora em alguns trabalhos os termos “situado” e “corporificado” sejam usados como sinônimos, nesta tese esses conceitos são claramente distintos. O *agente situado* é aquele que está em acoplamento dinâmico com seu ambiente, interferindo-se mutuamente em suas trajetórias, mas tendo um alcance de observação e de ação apenas

parcial. Já o *agente corporificado* é aquele que possui um corpo como universo interno, composto de propriedades e processos metabólicos próprios.

(159) De fato, o modelo de acoplamento dinâmico pressupõe um nível mínimo de corporificação. Dizer que o ambiente interfere nos estados do agente, como define o conceito de agente situado, implica dizer que o agente possui estados internos. Da mesma maneira, a existência de variáveis essenciais, às quais o agente deve procurar manter dentro de limites de viabilidade (conforme preconiza o modelo cibernético), supõe a existência de um corpo que as contenha, fazendo o papel de um universo interno para o agente.

(160) A ênfase dada pelos pesquisadores em Inteligência Artificial às noções de agente situado e corporificado se deveu ao esgotamento dos modelos tradicionais do paradigma simbólico, excessivamente representacionais e abstratos, muitas vezes distantes de certas evidências provenientes das ciências naturais, e também incoerentes com relação à ideia de agente autônomo (ZIEMKE, 1998), (SHANNON, 1993). Os programas apresentados pela IA tradicional se adaptam muito pouco, e tem seus objetivos estabelecidos à partir do exterior, sem ter propriamente significação interna.

(161) Este movimento ocorreu de forma similar nas ciências cognitivas, rejeitando os modelos puramente cognitivistas, e promovendo uma forte aproximação com modelos mais biológicos. Durante a década de 1990, a clássica oposição entre “mente” e “corpo” foi derrubada. Os estudos nas diversas áreas que compõe as ciências da mente conduziram a pesquisa na direção de uma reconciliação entre a psicologia e a biologia (TEIXEIRA, 2000). A evidência de que o cérebro faz parte do corpo, e desenvolveu-se junto com a evolução do corpo, implica que os aspectos mentais do ser humano não podem ser completamente dissociados dos aspectos orgânicos. O cérebro e o corpo encontram-se integrados através de circuitos bioquímicos e neuronais recíprocos dirigidos um ao outro (DAMÁSIO, 1994).

(162) Se a interação da mente de um agente se dá diretamente com o ambiente externo, o agente fica reduzido a um autômato, um sistema comportamental vazio do tipo estímulo e resposta. Por ser corporificado, o agente se torna uma entidade imersa em um ambiente e ao mesmo tempo distinta dele. O corpo torna-se assim o mediador entre o

ambiente e a mente. Enquanto o mundo se apresenta para o agente como o "ambiente externo", o corpo constitui o "ambiente interno".

### 2.3.2. O Corpo do Agente

(163)

Na arquitetura CAES, o *corpo* ( $\beta$ ) de um agente funciona como mediador entre a mente ( $\mu$ ) e o meio externo ( $\xi$ ), uma vez que é através do corpo que a mente do agente percebe o mundo e delibera ações a fim de transformá-lo, conforme ilustra a figura 2.7. O corpo também é a estrutura que representa o ambiente interno do agente, composto por um conjunto de *propriedades internas* ( $X_\beta$ ), e *processos metabólicos* que constituem a função de evolução ( $f_\beta$ ) desse subsistema, conforme a definição 2.3.

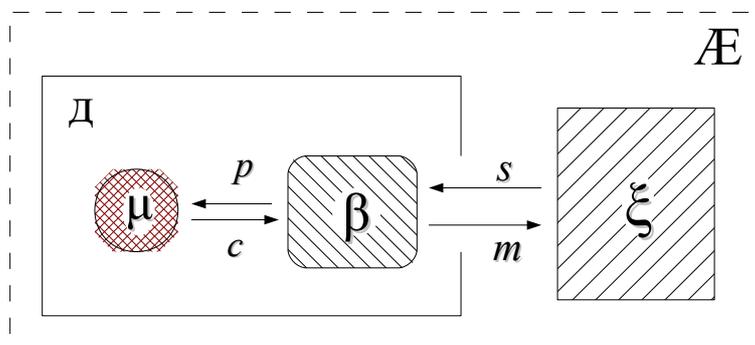


Figura 2.7: A mente, o corpo, e o ambiente.

O corpo ( $\beta$ ) do agente faz o papel de mediador entre sua mente ( $\mu$ ) e o mundo exterior ( $\xi$ ).

O corpo ( $\beta$ ) de um agente ( $\mathcal{A}$ ) é uma 6-upla:

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

Onde,

$X_\beta = \{X_{\beta 1}, X_{\beta 2}, \dots, X_{\beta n}\}$  é um conjunto finito de propriedades (*universo interno*)

$f_\beta : X_\beta \times C \times S \rightarrow X_\beta$  é a função de evolução do sistema (*metabolismo*)

$c$  é um sinal que o corpo  $\beta$  recebe da mente  $\mu$  (*controle*)

$p$  é um sinal que o corpo  $\beta$  envia para a mente  $\mu$  (*percepção*)

$s$  é a interferência externa que o corpo  $\beta$  recebe do ambiente  $\xi$  (*situação*)

$m$  é a interferência que o corpo  $\beta$  fornece ao ambiente  $\xi$  (*atuação*)

Definição 2.3: Corpo ( $\beta$ ).

(164)

Enquanto o corpo do agente se comunica com o ambiente externo através dos sinais  $s$  e  $m$ , a mente se comunica com o corpo através dos sinais  $p$  e  $c$ . A percepção ( $p$ ) é um sinal informativo, gerado por um conjunto de sensores do corpo do agente, e enviado destes à mente. Da mesma maneira, o controle ( $c$ ) é um sinal gerado pela mente, e enviado aos atuadores do agente, igualmente localizados no seu corpo.

(165) As propriedades internas ( $X_{\beta}$ ) definem o estado geral do corpo, de forma análoga ao que seriam os estados químicos, hormonais, fisiológicos, e biofísicos de um organismo natural. Essas propriedades só podem ter lugar no corpo do agente, pois constituem um nível diferenciado da mente e também diferenciado do mundo. A função de evolução do corpo ( $f_{\beta}$ ) representa seu metabolismo, constituído de processos regulatórios autônomos, que controlam a variação do estado das propriedades internas ao longo do tempo.

(166) Na natureza, o metabolismo de um ser vivo engloba os processos que regulam a atividade dos sistemas, órgãos, glândulas, e mesmo das próprias células. Em função de certas condições bioquímicas, o metabolismo de cada um desses componentes do corpo pode promover a alteração desse estado bioquímico, a partir, por exemplo, da produção de determinadas enzimas.

(167) O metabolismo é parametrizado principalmente pelo próprio estado do corpo. Entretanto, como o organismo é um sistema parcialmente aberto, então eventos externos podem ocasionar alterações no meio interno. Na natureza, situações exteriores que afetam o metabolismo são, por exemplo, a alimentação, a variação da temperatura ambiente, o movimento, o contato com outros objetos, etc. Além disso, a atividade do agente também interfere no estado do corpo, e por fim a mente tem um papel importante enquanto fator de modulação das mudanças corporais, uma vez que gerencia muitos dos processos metabólicos.

(168) Assim, a função de evolução do corpo ( $f_{\beta}$ ) modifica o estado das propriedades internas em função de 3 parâmetros. Primeiro, a própria condição das propriedades internas ( $x_{\beta}$ ), realizando um ciclo próprio, corporal, de regulação metabólica. Segundo, a interferência do ambiente, recebida como situação ( $s$ ), seja através da percepção sensorial ou então de outros tipos de eventos externos que impactem sobre o corpo. E terceiro, a interferência da mente através do sinal de controle ( $c$ ), alterando o estado de atuadores internos e externos.

### 2.3.3. Percepção e Controle

(169) A neurofisiologia e a psicofísica (GESCHIEDER, 1997) distinguem os mecanismos de percepção sensíveis à alteração de condições do ambiente

(*exteroceptivas*), dos mecanismos que monitoram condições do corpo (*interoceptivas*), (SHERRINGTON, 1907). No ser humano, os *sentidos* (conforme o conceito usual) são exteroceptivos. Entram nessa categoria visão, audição, tato, paladar, olfato, equilíbrio, etc.

(170) O sistema visual humano, por exemplo, conforme (MEYER, 1997), começa pelos olhos, órgãos sensíveis à incidência da luz proveniente do meio externo. Uma imagem bidimensional é formada na retina de cada olho pela conversão da frequência das ondas eletromagnéticas (de luz) para a forma de cores e intensidades. Essa imagem é comunicada ao lobo occipital do cérebro, onde ocorre um tipo de pré-processamento dessa informação, que então é convertida em sinais de movimento, traços e formas.

(171) Processo semelhante se dá com a audição (sensível às vibrações sonoras, chegadas ao ouvido através do ar e convertidas em sinais de frequência e intensidade), paladar e olfato (através de receptores químicos localizados na língua e mucosa nasal), tato (através de receptores sensíveis a diferentes tipos de pressão mecânica sobre a pele, e também à temperatura), e sensação de equilíbrio e aceleração (percebidas através do movimento dos fluidos do vestíbulo auricular).

(172) Diferentemente, as percepções interoceptivas dizem respeito à situação do corpo, como é o caso das sensações fisiológicas (registro do nível de diversos elementos hormonais e enzimáticos na corrente sanguínea), viscerais (provenientes dos diversos órgãos), cinestésicas, sensações de dor, e ainda sensações propioceptivas (relativas à posição e movimento dos membros do corpo, através do aparato esqueleto-muscular). As sensações propioceptivas fornecem o feedback sobre o estado dos atuadores. São elas que fazem, por exemplo, a pessoa ter a sensação da posição em que está o seu próprio braço após um movimento.

(173) Assim como as percepções, as ações de um organismo também podem ser tipificadas pelo mesmo critério: de um lado aquelas voltadas para o mundo exterior, e de outro lado aquelas dirigidas ao próprio corpo.

(174) No ser humano, grande parte da atividade cerebral é destinada à regulação do próprio meio corporal interno. Entretanto, certos mecanismos do corpo, os atuadores, são capazes de realizar intervenções externas, dirigidas ao ambiente. O aparato esqueleto-muscular, por exemplo, é o atuador responsável por gerar os movimentos, e,

portanto, pode promover alterações no ambiente fora do corpo. Já as glândulas e a maioria dos órgãos funcionam como mecanismos de regulação interna, produzindo, eliminando e transformando as substâncias presentes no organismo, e portanto modificando seu estado fisiológico.

(175) Entre os mecanismos que atuam sobre o universo corporal interno, encontra-se um tipo especial, que chamaremos *atuadores sensoriais*. Através deles, a mente pode atuar modificando parâmetros dos próprios sensores. No ser humano, o exemplo típico são os músculos que regulam a direção do olhar. De fato, existem vários movimentos que nós fazemos a fim de modificar a posição relativa do corpo, justamente para deslocar os olhos, os ouvidos, o nariz, a língua, as mãos, enquanto instrumentos sensoriais, para permitir uma melhor percepção das coisas. A existência deste tipo de controle da mente sobre parâmetros sensoriais possibilita a realização de *percepção ativa* (ALOIMONOS et al., 1988), (BAJCSY, 1988), (WASSON et al., 1998).

(176) Na arquitetura CAES, a *percepção* ( $p$ ) é um sinal informativo recebido pela mente do agente, composto por um conjunto de sensações que representam a situação de seu corpo. No sentido inverso, o *controle* ( $c$ ) é um sinal gerado pela mente, e enviado ao corpo a fim de desencadear determinadas ações.

(177) Assim como ocorre nos domínios naturais, na arquitetura CAES os sinais de percepção e controle podem ser definidos como a combinação de dois tipos de sinais: aqueles relacionados ao ambiente externo, e aqueles relacionados ao próprio corpo, enquanto universo interno do agente.

(178) Divide-se, portanto, o espaço do sinal de controle,  $C = \{C_1, C_2, \dots, C_n\}$ , enviado pela mente ao corpo, em controle interno e externo. De um lado, o *controle interno* ( $C_\beta \subset C$ ) é o meio pelo qual a mente interfere no próprio estado do corpo, conduzindo processos regulatórios. De outro lado, o *controle externo* ( $C_\xi \subset C$ ) é o meio pelo qual a mente, através dos atuadores, interfere e produz alterações no ambiente. Os atuadores são a parte das propriedades do corpo ( $X_\beta$ ) que participam do sinal de atuação do agente ( $M$ ), sendo um dos parâmetros da função de evolução do estado do ambiente ( $f_\xi$ ). Portanto, os atuadores constituem-se como os instrumentos através dos quais a mente pode intervir no mundo exterior. O fluxo de controle é ilustrado na figura 2.8.

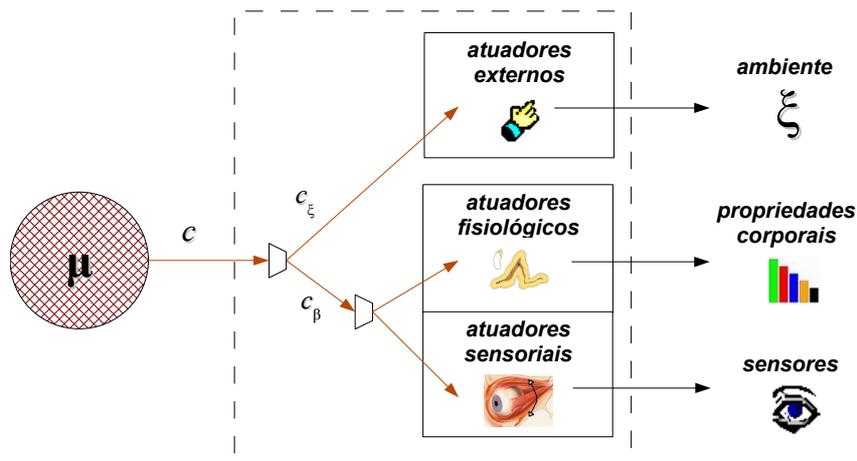


Figura 2.8: Fluxo de controle.

(179) O espaço da percepção,  $P = \{P_1, P_2, \dots, P_n\}$ , é igualmente dividido em percepção externa e interna. De um lado, a *percepção externa* ( $P_\xi \subset P$ ) é a parte do sinal perceptivo através do qual a mente tem acesso ao mundo exterior, pois monitora variáveis do corpo que evoluem em função de condições do ambiente (sensores). Os sensores são um subconjunto das propriedades do corpo ( $X_\beta$ ) cujo valor é dado pela situação ( $S$ ), que por sua vez retrata as condições do ambiente. De outro lado, a *percepção interna* ( $P_\beta \subset P$ ) é composta por variáveis que monitoram valores de outras propriedades corporais, comunicando seu estado à mente. As propriedades do corpo e do ambiente para as quais a mente não possui acesso perceptivo através de  $P$  são, do ponto de vista do agente, *propriedades não-observáveis*. O fluxo perceptivo é ilustrado na figura 2.9.

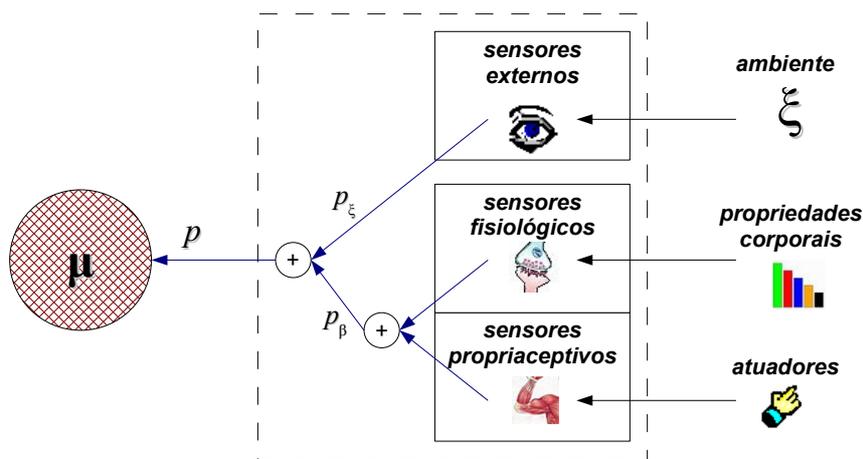


Figura 2.9: Fluxo de percepção.

(180) De fato, o que se estabelece entre mente e aquilo que está fora dela (o corpo e o ambiente) é uma relação de *observação e controle parciais*. A percepção da mente sobre o corpo, através do sinal  $P_\beta$  fornece acesso a apenas uma parcela das dimensões

que definem o estado real do corpo ( $X_\beta$ ), e as ações que a mente pode executar sobre o corpo, através do sinal  $C_\beta$  representam apenas uma parte dos parâmetros da função de evolução do corpo ( $f_\beta$ ) capazes de influenciar nas modificações sobre seu estado.

(181) Da mesma forma, a percepção que a mente tem do ambiente exterior, através do sinal  $P_\xi$  proveniente dos sensores, fornece acesso a apenas uma parte da situação ( $S$ ) em que se encontra o agente, sinal que por sua vez já é parcial em relação ao estado real do ambiente ( $X_\xi$ ). Assim também, a mente pode interferir sobre o ambiente apenas de forma parcial e indireta, através do sinal  $C_\xi$ , que altera o estado dos atuadores, localizados no corpo ( $X_\beta$ ), e que por sua vez modificam a atuação do agente ( $M$ ), sinal este que é um dos parâmetros da função de evolução do ambiente ( $f_\xi$ ).

## 2.4. Adaptação do Agente ao Ambiente

(182) No modelo de acoplamento dinâmico (BEER, 1995) o critério de adaptação é representado de maneira abstrata e geral como uma área do espaço onde o fluxo do sistema deve permanecer. O limiar de adaptação é portanto a fronteira dessa região do sistema global (formado pelo acoplamento entre agente e ambiente), e o agente é considerado adaptado ao ambiente se sua atividade conduzir a trajetória do sistema global de forma que ela permaneça restrita aos limites dessa área. Um exemplo é ilustrado na figura 2.10.

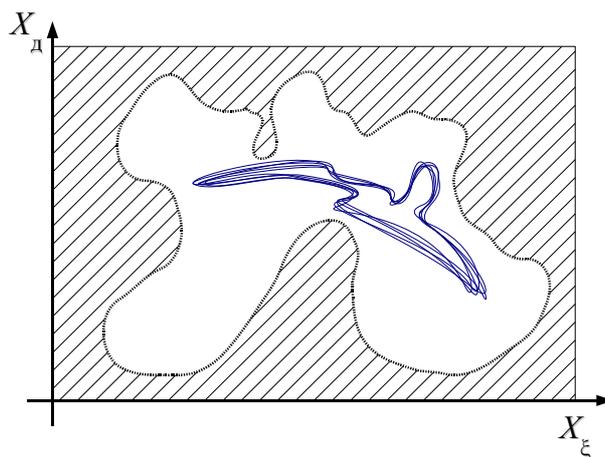


Figura 2.10: Exemplo de região de adaptação e trajetória adaptada.

O eixo horizontal representa a variação do estado do ambiente ( $X_\xi$ ), e o vertical a variação do estado do agente ( $X_\pi$ ).

(183) Ashby (1952) já havia sugerido critério semelhante para definir o conceito de adaptação, propondo que o agente é um sistema dotado de um conjunto de variáveis essenciais que devem permanecer dentro de certos limites fisiológicos normais, ou limites de viabilidade, para que seja preservada a integridade do sistema, e conseqüentemente a sobrevivência do agente. Um determinado comportamento colabora para a adaptação do agente se ele garante a persistência dessas variáveis essenciais dentro de seus limites de viabilidade.

(184) Em geral, seja na natureza, seja nos problemas de inteligência artificial, agente e ambiente não se estabelecem automaticamente em perfeita harmonia. Na maior parte do tempo, estes dois sistemas exercem forças em direções contrárias no que diz respeito ao fluxo do sistema global. Destes dois polos, apenas um deles, o agente, corre o risco de se desintegrar, ou seja, de desaparecer enquanto unidade.

(185) Um acoplamento dinâmico não-destrutivo é alcançado na relação entre os dois sistemas, quando o agente interage com o ambiente de forma a garantir sua auto-manutenção (VARELA et al., 1991), (RUIZ-MIRAZO; MORENO, 2000).

#### **2.4.1. Estar Adaptado**

(186) Em termos computacionais, um agente é considerado adaptado ao seu ambiente quando ele é bem sucedido na realização de seus objetivos. Em mecanismos programados, esta medida de sucesso é geralmente baseada em fatores exógenos. Por exemplo, um bom critério de adaptação para um determinado robô, máquina, ou programa, pode ser a satisfação do usuário. Entretanto, na natureza, a sobrevivência é quem demarca os limites do sucesso. O próprio ambiente torna-se juiz, condenando ao desaparecimento aqueles organismos (e finalmente as espécies) que não estavam suficientemente adaptados.

(187) Assim, na perspectiva naturalista da IA, um agente é considerado *adaptado* ao seu ambiente quando ele é bem-sucedido na tarefa de sobreviver, não necessariamente por um tempo infinito, mas pelo menos numa janela de longo-prazo. Para poder definir uma fronteira entre a vida e a morte de um agente, os modelos situados geralmente utilizam o modelo cibernético, estabelecendo o agente como um sistema dotado de um conjunto de *variáveis essenciais*, as quais devem permanecer dentro de certos *limites de*

*viabilidade* para que seja preservada a integridade do sistema, e conseqüentemente a sobrevivência do agente.

(188) As variáveis essenciais do agente exigem constante regulação. Em geral, algumas delas tendem naturalmente a sair dos limites de viabilidade, portanto o agente deve ser capaz de agir adequadamente a fim de garantir sua preservação a longo prazo. Como essas variáveis não estão sob controle absoluto do agente, então somente o fluxo determinado pelo acoplamento entre agente e ambiente pode garantir que elas sejam mantidas dentro dos limites de viabilidade (BARANDIARAN; MORENO, 2008).

(189) O comportamento do agente torna-se, então, o ponto chave para garantir sua própria sobrevivência. Na arquitetura CAES, a mente é o sistema responsável por coordenar o comportamento do agente, visando garantir sua adaptação. O agente estará adaptado ao seu ambiente se souber interagir habilmente de forma a coordenar a modificação de condições exteriores com a finalidade de guiar o fluxo global do sistema, garantindo a persistência de suas variáveis essenciais dentro dos limites de viabilidade. A manutenção do equilíbrio das variáveis internas exige a adaptação do comportamento às variações do ambiente externo (BARANDIARAN, 2004).

(190) O paradigma da IA Situada inverteu um dos princípios cibernéticos, afirmando que organismos ou agentes complexos, acoplados a seus ambientes, constituem sistemas distantes do equilíbrio (*far from equilibrium systems*), (BICKHARD, 2000). Trata-se de um tipo especial de sistema aberto que encontra seus limites de viabilidade longe do equilíbrio do sistema global. Esse tipo de sistema é uma estrutura dissipativa, que tenderia naturalmente a desintegração, mas que entretanto permanece estável por um longo período de tempo através da regulação da sua interação com o ambiente.

(191) Nesta condição, o agente é uma entidade que deve interferir ativamente no ambiente, num esforço para contrariar a força natural que induz o sistema global ao equilíbrio, situação que implicaria o desaparecimento do agente enquanto unidade, enquanto uma organização diferenciada do ambiente. É mantendo o sistema global em constante desequilíbrio que o agente alcança seu equilíbrio interno.

### 2.4.2. Ser Adaptativo

(192) A situação de adaptação pode ser alcançada por um agente porque ele foi bem projetado (seja por um engenheiro, seja pelo processo de seleção natural), ou porque ele possui mecanismos de adaptação *online*. Nesse caso o agente é adaptativo. Enquanto *estar adaptado* refere-se à situação de ser bem sucedido nas interações com o ambiente, *ser adaptativo* é a propriedade que determinados sistemas têm de se modificarem em busca da adaptação. Essa característica está necessariamente vinculada a processos de desenvolvimento e aprendizagem.

(193) Um agente constitui um *sistema adaptativo* se ele é capaz de modificar-se a fim de se tornar ajustado às mudanças do ambiente (BARANDIARAN; MORENO, 2008), ou seja, se ele é capaz de aprender através da experiência, modificando seus padrões de comportamento no intuito de aumentar a competência no desempenho de suas atividades (MAES, 1994). No caso do modelo de acoplamento dinâmico, isto significa mudar as estruturas geradoras do comportamento, a fim de que, como um todo, elas tornem o agente mais adaptado ao ambiente.

(194) Por exemplo, uma alga marinha é um organismo inatamente adaptado às condições normais de vida no mar. A qualidade de adaptação da alga marinha foi adquirida através da seleção natural da espécie, portanto é resultado de um processo filogenético. Já uma minhoca, se ali colocada, termina rapidamente por cessar sua existência enquanto organismo. A minhoca não tem capacidade adaptativa para tratar de um ambiente tão adverso para ela quanto é o mar.

(195) O ser humano, diferentemente, pode adaptar-se ao mar, pelo menos de uma forma mínima para sobreviver a ele. Não é possível que uma pessoa transforme seu metabolismo a ponto de poder respirar água salgada. Entretanto, é possível modificar o comportamento aprendendo a nadar, como resultado de um processo ontogenético. E nadar não é resistir às ondas, nem opor-se à dinâmica do oceano, mas sim harmonizar o movimento do corpo ao da água, compreender as relações que se estabelecem entre o corpo e as ondas, infligindo pequenas mudanças locais bem dirigidas de modo que o resultado é a condução do sistema por um fluxo favorável ao organismo (DELEUZE, 1970).

(196) Um agente artificial pode apresentar um comportamento adaptado sem aprendizagem, mostrando soluções inteligentes para seus problemas somente a partir de um conjunto de instruções pré-programadas. No entanto, para que isto seja possível, é preciso que as estratégias de solução sejam conhecidas pelo programador, e que o número de situações existentes no ambiente seja pequeno o suficiente para permitir que o agente seja pré-programado.

(197) Neste caso, construir um agente adaptado ao ambiente exige que os programadores dominem o conhecimento necessário para resolver os problemas que serão enfrentados pelo agente ao longo de sua existência. Evidentemente, existem muitos casos em que é difícil ou mesmo impossível prever e fornecer soluções para todas as situações. Para estes problemas é preciso que o agente aprenda com a experiência e se adapte às situações imprevistas.

(198) Na natureza, a adaptatividade pode ocorrer através de processos puramente fisiológicos, como é o caso dos seres simples, unicelulares, cuja adaptação depende diretamente do seu metabolismo, enquanto sensível a mudanças nas condições externas, ou então de comportamentos reativos. Existem, entretanto, processos de adaptação mais complexos, que envolvem aspectos cognitivos, para os quais reservamos o termo *aprendizagem* (DENNET, 1996). A aprendizagem é a adaptatividade que se realiza nos seres mentalizados, que dispõem de uma organização intelectual, e de informações que são relevantes para o seu comportamento, em geral representado sob a forma de um modelo interiorizado da dinâmica do mundo.

## 2.5. Características da Mente

(199) Embora certos mecanismos automáticos e simuladores de organismos simples possam dispensar o conceito de “mente”, em geral os modelos de agentes em inteligência artificial são mentalizados. Em analogia ao que acontece na maioria dos animais, o comportamento de um agente descrito através da arquitetura CAES é resultado das interações de sua mente, com a exterioridade, ou seja, com tudo o que está fora dela, incluindo o ambiente físico que existe fora do corpo, e o corpo do agente fora da mente (PARISI, 2004).

(200)

A mente ( $\mu$ ) na arquitetura CAES é uma estrutura de controle composta por dois subsistemas: *regulativo* ( $\mathcal{R}$ ) e *cognitivo* ( $\mathcal{K}$ ), encarregados de gerenciar tanto o metabolismo quanto o comportamento do agente. A *mente* do agente se comunica com o corpo através de sinais de controle ( $c$ ) e de percepção ( $p$ ), conforme ilustrado na figura 2.11, e formalizado na definição 2.4, a seguir.

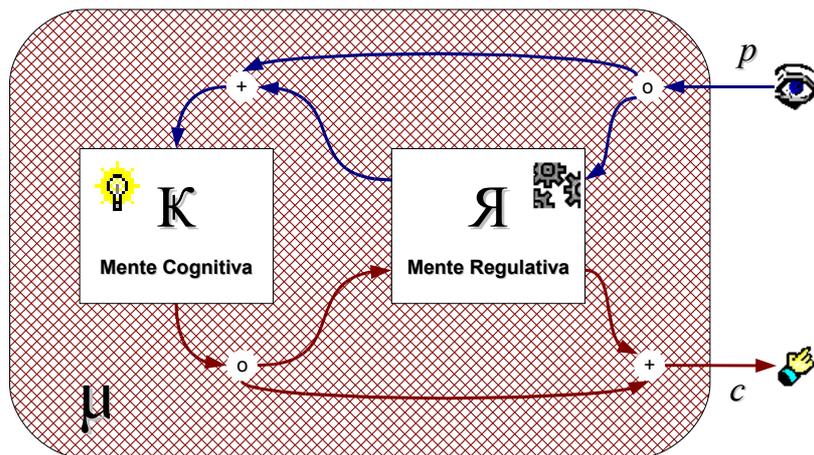


Figura 2.11: Estrutura interna da mente ( $\mu$ ) na arquitetura CAES.

O sistema regulativo ( $\mathcal{R}$ ) da mente se interpõe entre o corpo e o sistema cognitivo, de um lado introduzindo valores afetivos sobre o sinal de percepção, e de outro lado promovendo certas regulações do corpo e da conduta do agente através de reações emocionais e comportamentais. O sistema cognitivo ( $\mathcal{K}$ ), por sua vez, é responsável pela compreensão, aprendizagem, e tomada de decisão.

A mente ( $\mu$ ) de um agente é uma 4-upla:

$$\mu = \{\mathcal{K}, \mathcal{R}, c, p\}$$

Onde,

$\mathcal{K}$  é o subsistema cognitivo

$\mathcal{R}$  é o subsistema regulativo

$c$  é um sinal que a mente  $\mu$  envia para o corpo  $\beta$  (controle)

$p$  é um sinal externo que a mente  $\mu$  recebe do corpo  $\beta$  (percepção)

Definição 2.4: Mente ( $\mu$ ).

(201)

De forma similar a (SLOMAN et al., 2005), a mente regula o comportamento do agente combinando um sistema cognitivo, responsável por aprender, interpretar, planejar e deliberar as ações, com um sistema reativo e emocional.

(202)

O *sistema regulativo* da mente é definido como uma 3-upla:  $\mathcal{R} = \{\mathcal{R}_R, \mathcal{R}_E, \mathcal{R}_A\}$ , onde  $\mathcal{R}_E$  é o *sistema emocional*, responsável por realizar reações metabólicas internas,  $\mathcal{R}_R$  é o *sistema reativo*, responsável por realizar reações comportamentais externas, e  $\mathcal{R}_A$

é o *sistema avaliativo*, que atribui valor afetivo às sensações. O sistema regulativo da mente ( $\mathcal{R}$ ) é ilustrado na figura 2.12.

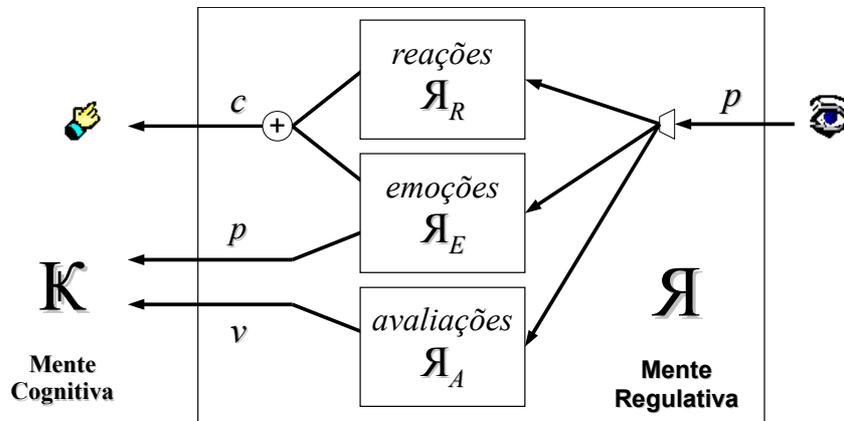


Figura 2.12: Estrutura interna do sistema regulativo ( $\mathcal{R}$ ) na arquitetura CAES.

### 2.5.1. Afetividade e Emoções

(203) A discussão teórica sobre o conceito de *autonomia* em agentes conduziria mais cedo ou mais tarde a uma questão crucial: a necessidade de que os objetivos do agente sejam estabelecidos por ele mesmo. Um agente verdadeiramente autônomo deve atuar no mundo sob a força de *motivações intrínsecas*. No paradigma do agente situado e corporificado, dinamicamente acoplado ao seu ambiente, essa motivação deve ser o fator que leva o agente a agir em favor da manutenção de suas variáveis essenciais.

(204) Paralelamente, a década de 1990 apresentou grandes avanços da neurociência quanto à compreensão do cérebro e de seu funcionamento, e isso dissolveu o antagonismo clássico entre *razão* e *emoção*. Nos organismos inteligentes complexos, em especial no ser humano, estes fatores estão entrelaçados, e o comportamento é resultado de seu funcionamento integrado. Razão sem emoção é vazia, emoção sem razão é cega (ELLIS, 1962).

(205) Quer dizer que além do raciocínio lógico existem outras propriedades fundamentais que caracterizam o comportamento inteligente. A mente não pode ser confundida com a razão, a memória, ou a consciência. De fato, a mente são todos esses aspectos, e ainda outros, trabalhando em paralelo e conjuntamente, aí inclusos os aspectos afetivos e emocionais.

(206) O comportamento inteligente é uma consequência da integração entre um sistema cognitivo, que analisa os cenários, e um sistema emocional e afetivo, que os

avalia e induz à ação (DAMÁSIO, 1994). No caso humano, não é possível apontar um comportamento que seja puramente cognitivo, sem elementos afetivos envolvidos, nem um comportamento puramente afetivo, sem elementos cognitivos. A afetividade é uma condição necessária para a existência de comportamento inteligente, pois é (ao menos na origem) o motor para a ação. O ser humano só age quando impulsionado por um motivo, e este se traduz sob a forma de uma necessidade (PIAGET, 1954, 1964, 1967).

(207) De um lado, um estado de emoção supõe uma discriminação, e portanto a capacidade intelectual de reconhecer e diferenciar as situações. De outro lado, a presença das emoções durante as experiências vividas altera a noção que temos dessas situações, conferindo a elas uma determinada qualidade positiva ou negativa (DAMÁSIO, 1994).

(208) De forma similar à (CAÑAMERO, 1997a), a arquitetura CAES estabelece uma clara distinção entre “emoção” e “afetividade”. A *afetividade* está relacionada às necessidades do agente. Ela se efetua através de um sistema responsável por fornecer uma avaliação, positiva (agradável, boa...) ou negativa (desagradável, ruim...) dos sinais do corpo, e eventualmente também de algumas condições importantes do ambiente. Essa informação avaliativa é geralmente dirigida ao sistema cognitivo, que, baseado nela, realiza o processo de tomada de decisão.

(209) As *emoções* são executores de respostas reflexas e viscerais. Elas também se realizam através de sistemas interiores ao agente. Entretanto, diferentemente da afetividade, cada emoção é um processo particular, que, a partir do reconhecimento de eventos especiais, promove reações corporais específicas, alterando o estado interno do agente. Essas alterações, dirigidas ao corpo, têm a finalidade de preparar o agente para repostas externas mais adequadas ao seu contexto imediato, potencializando ou inibindo certos tipos de comportamento.

### 2.5.2. Sistema Avaliativo

(210) O afeto desempenha um papel importante no funcionamento da inteligência. Sem afeto não haveria interesse, nem necessidade, nem motivação. Na arquitetura CAES, uma *necessidade* é o desequilíbrio de alguma variável interna essencial do

agente que precisa ser compensado. A motivação para restabelecer o equilíbrio dessas variáveis tem origem no próprio agente através de seu *sistema avaliativo*.

(211) O sistema avaliativo é composto por um conjunto de funções que avaliam determinadas propriedades do corpo, e eventualmente certas condições percebidas do ambiente, fornecendo valores positivos e negativos que servem como indicação de situações favoráveis ou desfavoráveis em que o agente pode se encontrar. Essa avaliação constitui um valor afetivo, que é a forma mais básica de discernimento entre o que é inerentemente bom ou ruim para o agente.

(212) Por exemplo, no ser humano, os sinais corporais de dor, sede, fome, frio ou calor excessivos, exaustão, enjoo, entre outros, além de sinais externos como o gosto e o cheiro de substâncias potencialmente nocivas, são, em algum ponto, inerentemente negativos. Da mesma forma, sensações como o prazer sexual ou o sabor de um alimento calórico, sob certas condições corporais, possuem valor inatamente positivo. Isso faz sentido no contexto evolutivo da espécie humana, e não contradiz o fato de que no nível do pensamento abstrato outros valores possam ser agregados a estes sinais.

(213) No modelo CAES, a necessidade se expressa através da afetividade. Quando uma certa variável essencial precisa ser mantida constantemente em torno de um determinado “valor normal”, associa-se a ela uma relação de *afetividade normal*. Neste caso, o afastamento dos valores normais gera uma sensação negativa. Outras variáveis podem ser importantes para o agente, sendo ou não essenciais. Para essas variáveis é estabelecida uma relação de *afetividade positiva* correspondente a um “pico” de sensação agradável quando elas atingem uma determinada faixa de valores alvo. Inversamente, também podem ser estabelecidas relações de *afetividade negativa* como “vales” de sensação desagradável.

(214) Para ser autônomo, um agente deve agir e tomar decisões por força de motivações ou objetivos próprios. No modelo clássico de aprendizagem por reforço, o agente é *extrinsecamente motivado*, ou seja, ele é levado a agir em função de algum objetivo ou recompensa externa, vinda do ambiente. Na arquitetura CAES, pela existência do sistema avaliativo, o agente torna-se *intrinsecamente motivado*, pois ele é levado a fazer alguma coisa porque ela é inerentemente agradável. A ideia de agentes

intrinsecamente motivados pode ser lida em (SINGH et al., 2004) e (OUDEYER; KAPLAN, 2007).

(215) A existência de um sistema afetivo artificial constitui uma forma de internalizar os sinais de reforço, substituindo a ideia de recompensa externa. Desaparece assim a figura de um objetivo explícito, como uma situação a ser alcançada no mundo, e em seu lugar existe uma motivação para a ação, que emerge da relação entre conhecimento e afetividade.

(216) A afetividade também substitui a sobrevivência como parâmetro de adaptação. Se somente a morte do sistema é o indicador de inadaptação, torna-se impossível que um agente aprenda a se adaptar ao ambiente através da experiência durante o tempo de sua vida, pois assim que descobrir que fez a escolha errada, já terá morrido.

(217) A afetividade adiciona um tipo de *relevo avaliativo* ao espaço (de fase ou de estados) do sistema, como ilustrado na figura 2.13. Sem afetividade, esse espaço é uma grande planície cinza, e o agente não tem como preferir uma situação a outra. Esse relevo avaliativo contém algumas regiões mais altas, positivas do espaço, que abrigam situações sentidas pelo agente como prazerosas, e que possivelmente estão associadas a momentos de grande benefício em termos de sua auto-manutenção. Da mesma forma, há outras regiões do espaço onde o relevo é baixo, indicando contextos desagradáveis para o agente, e que muito provavelmente guardam, no final da descida, a fronteira do limiar de sobrevivência.

(218) O relevo avaliativo define a motivação do organismo para que ele mantenha suas variáveis essenciais em torno dos índices normais. Como consequência, se o organismo age buscando as sensações positivas e evitando as negativas, ele indiretamente estará mantendo o fluxo do sistema numa região protegida, longe das fronteiras da morte.

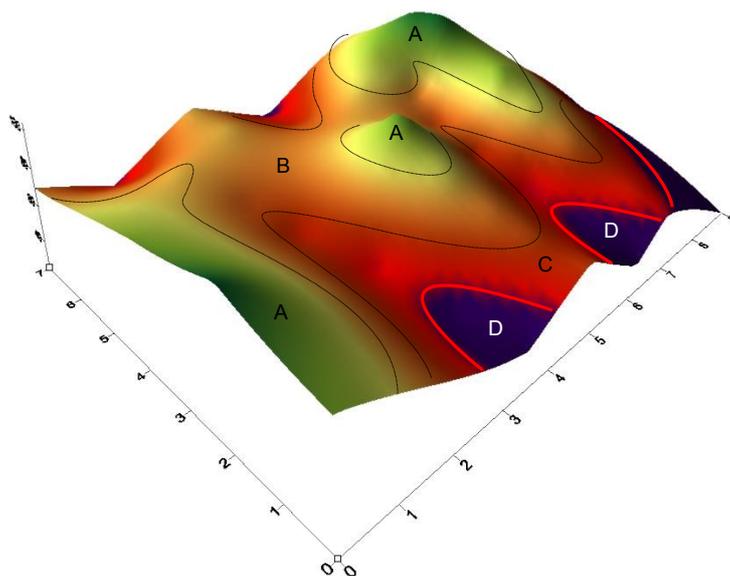


Figura 2.13: Exemplo de uma superfície afetiva.

O espaço representa o fluxo do sistema, variando o estado do agente e do ambiente, e o relevo representa a avaliação afetiva desses estados pelo agente. As áreas altas (A) representam valores positivos, e as zonas baixas (C) representam valores negativos. As depressões escuras do gráfico (D) encontram-se já além do limite de sobrevivência do agente.

(219) Por exemplo, manter-se hidratado é uma situação crucial para a nossa sobrevivência. Seria bastante custoso, arriscado e tardio para o organismo se a detecção de baixos níveis de hidratação só ocorresse no limiar da sobrevivência, quando as células começassem a morrer. A evolução natural fez muito bem em estabelecer mecanismos para avaliar o nível de hidratação, procurando mantê-lo em torno de *valores normais*, onde garante-se que o organismo fica saudável. No momento em que esse nível se afasta dessa região de normalidade, uma sensação desagradável (de sede) começa a surgir. Esse alerta é dado bem antes do limiar de sobrevivência, garantindo que o agente vá atuar mais cedo (no caso bebendo água) a fim de compensar o desequilíbrio e eliminar a sensação desagradável.

(220) Da mesma forma, as sensações positivas são atalhos criados pela seleção natural para que o organismo prefira e busque certas situações muito favoráveis a sua auto-manutenção, que foram reveladas ao longo das gerações, e que acabaram incorporadas no organismo através desse mecanismo afetivo inato. Por exemplo, o prazer gerado pela ingestão de alimentos altamente calóricos é um facilitador de sobrevivência. Seria bem mais arriscado para o organismo ter que descobrir sozinho, após ficar desnutrido, que a comida que ele está ingerindo não serve pra nada.

(221) Tal modelo é biologicamente plausível. A evolução natural definiu nos organismos complexos uma série de mecanismos para garantir sua sobrevivência. Entre esses mecanismos, além das funções orgânicas, estão funções ligadas ao comportamento, como os instintos, as reações, as emoções, e a afetividade. As sensações positivas e negativas definidas no sistema avaliativo são justamente uma forma de mapear os limites de sobrevivência do organismo, conforme ilustra a figura 2.14.

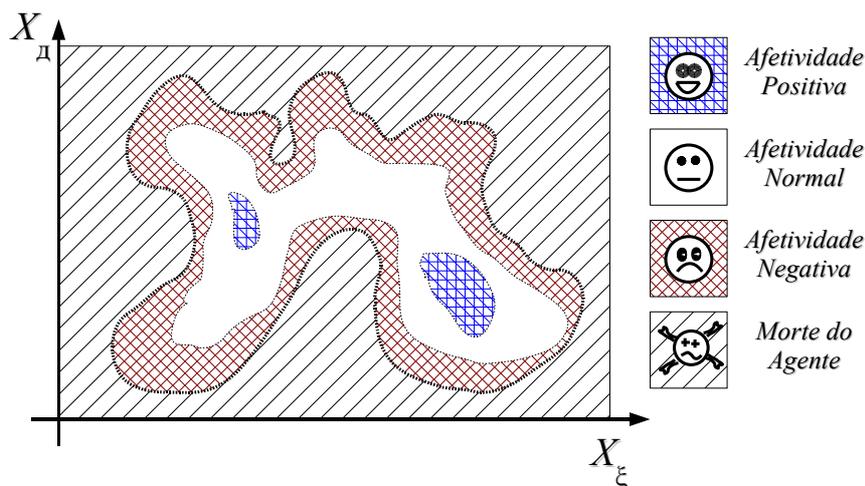


Figura 2.14: A superfície afetiva mapeando o limiar de sobrevivência.

As zonas de afetividade negativa formam uma faixa de segurança que afasta o agente das fronteiras da sua própria morte, enquanto que as zonas de afetividade positiva funcionam como atratores que levam o agente a situações vantajosas para sua auto-manutenção.

(222) O sistema avaliativo ( $\mathfrak{R}_A$ ) associa um valor afetivo a determinadas propriedades do sinal de percepção. Os valores estão numa escala contínua e descrevem sensações agradáveis ou desagradáveis associadas a determinadas condições. Estes valores afetivos são enviados ao sistema cognitivo, que deve utilizá-los como parâmetro de decisão entre ações alternativas, a fim de levar o agente a situações mais confortáveis.

(223) O sistema avaliativo pode ser mais precisamente definido como um conjunto de *gatilhos avaliativos*. Esses gatilhos são descritos, de forma geral, como funções que mapeiam o domínio dos valores possíveis de cada elemento  $P_i$  do sinal de percepção  $P$  para um contra-domínio de avaliação escalado entre  $-1$  e  $+1$ . Assim cada percepção específica pode representar uma condição, seja ela reportada ao ambiente, seja ao corpo, associada a sensações agradáveis quando positivas e desagradáveis quando negativas. A condição de um gatilho avaliativo também pode ser composta pela combinação de condições simples. A composição das afetividades de todos os gatilhos avaliativos

disparados em cada instante forma a *sensação afetiva geral* do agente. Esse tipo de sinal avaliativo corresponde ao modelo de recompensas fatoradas (DEGRIS et al., 2008).

### 2.5.3. Sistema Emocional

(224) Na natureza, tanto nos animais como nos seres humanos, as emoções influenciam o comportamento, privilegiando determinadas condutas em detrimento de outras. Assim como a afetividade, as emoções também são uma forma de avaliar o resultado das ações do agente.

(225) Os sistemas emocionais biológicos são tema de grande interesse da neuropsicologia contemporânea (LEDOUX, 2000), (DAMÁSIO, 1994), (EKMAN, 1999). Da mesma forma, o estudo das emoções e a pesquisa relativa à produção de emoções sintéticas, ganharam espaço no campo da inteligência artificial. Alguns modelos emocionais artificiais interessantes foram apresentados em (CAÑAMERO, 1997a, 1997b), (VELÁSQUEZ, 1997), (SLOMAN, 1999, 2005) e (ALMEIDA et al., 2004).

(226) Segundo uma definição neuropsicológica, a emoção está ligada tanto ao caráter afetivo quanto ao caráter corporal. Ela é a combinação de um processo avaliatório mental, simples ou complexo, com o desencadeamento de respostas dispositivas, em sua maioria dirigidas ao próprio corpo, resultando num estado emocional do corpo, mas também dirigidas à mente. A partir da identificação ou antecipação de determinadas situações, inicia-se um processo de alteração corporal, que desencadeia sensações específicas e predispõe o organismo para comportamentos de resposta (DAMÁSIO, 1994).

(227) No ser humano existe um conjunto de estados emocionais inatamente organizados. Cada emoção possui uma configuração bem definida, relativa ao tipo de eventos que irão desencadeá-la e às alterações corporais e comportamentos induzidos por ela (DAMÁSIO, 1994). A emoção promove mudanças fisiológicas no organismo, preparando-o para responder diferentemente em estados emocionais diferentes (EKMAN, 1999).

(228) Diferentes emoções não são apenas uma questão de intensidade positiva ou negativa. Na anatomia humana, cada emoção básica é um subsistema particular,

correspondendo a uma unidade funcional distinta no cérebro. Tanto as emoções negativas (medo, raiva, nojo, tristeza e desprezo) como as emoções positivas (alegria, orgulho, satisfação, alívio e contentamento) correspondem a sensações diferentes. Cada uma dessas emoções se estabeleceu na espécie humana pelo seu valor adaptativo, por facilitar o tratamento de tarefas fundamentais da vida, preparando o organismo para responder à situação detectada (EKMAN, 1999).

(229) As emoções são dirigidas ao cérebro, alterando estados de atenção, ativando redes associativas, e induzindo a evocação de memórias, e são dirigidas ao corpo, provocando alterações fisiológicas e endócrinas (atividade hormonal, reações metabólicas), e interferindo na atividade do sistema nervoso autônomo (reações corporais, expressões), (EKMAN; DAVIDSON, 1994).

(230) Por exemplo, em muitos animais uma emoção de medo pode ser disparada quando se ouve um barulho forte e ruidoso. A ativação do medo desencadeia alterações corporais no organismo, aumentando o nível de substâncias como a adrenalina na corrente sanguínea. No caso o hormônio serve como desencadeador de outros processos metabólicos e fisiológicos, como o aumento do batimento cardíaco, e consequente aumento da oxigenação dos músculos, produzindo no organismo as condições e uma propensão para comportamentos mais adaptados, como os de fuga.

(231) Na arquitetura CAES, o agente possui emoções através de um modelo sintético que define propriedades emocionais. Estas variáveis, além de interferir nos processos metabólicos internos do corpo, também influenciam os comportamentos subsequentes do agente porque modificam certos parâmetros implicados nos processos de avaliação afetiva e de tomada de decisão.

(232) O *sistema emocional* ( $\mathcal{R}_E$ ) é formado por um conjunto de *emoções artificiais*, cada uma delas sensível a determinadas percepções do agente (sejam elas externas, provenientes do meio, ou internas, provenientes do próprio corpo ou mesmo da própria mente), de forma análoga às emoções primárias do ser humano (DAMASIO, 1994). Uma emoção é ativada a partir da detecção de um determinado contexto, e sua ativação leva à realização de alterações corporais, representadas por um conjunto de efeitos transmitidos da mente ao corpo pelo sinal de controle interno ( $C_\beta$ ). Essas alterações, dirigidas ao corpo, têm a finalidade de preparar o agente para repostas externas mais

adequadas ao seu contexto imediato, potencializando ou inibindo certos tipos de comportamento.

- (233) No CAES, como em (CAÑAMERO, 2001) e (SLOMAN et al., 2005), o sistema emocional não reproduz diretamente efeitos superficiais conhecidos de emoções sobre o comportamento, mas implementa processos subjacentes da emoção através dos quais os efeitos comportamentais podem emergir. O sistema emocional é formado por um conjunto de emoções primárias. Cada emoção é ativada pela detecção de determinados contextos específicos, extraídos das percepções internas e externas, disparando reações metabólicas que iram alterar os valores das propriedades internas do corpo do agente. As emoções podem também enviar sinais diretamente ao sistema cognitivo, através da percepção interna, e podem estar associadas a valores afetivos no sistema avaliativo.

#### 2.5.4. Sistema Reativo

- (234) O *sistema reativo* da mente é um executor de respostas reflexas. Cada reação é um processo particular, que, a partir do reconhecimento de eventos especiais, promove ações dirigidas. Na natureza, a seleção natural provê tais mecanismos aos organismos pois eles trazem vantagens adaptativas.

- (235) Na arquitetura CAES, o principal regulador do comportamento é o sistema cognitivo do agente. Entretanto, de forma similar ao proposto em (SLOMAN, 1999), é possível combinar um sistema deliberativo, coordenado pela cognição, com um sistema reativo.

- (236) Na natureza, mesmo os animais que desenvolveram mecanismos cognitivos sofisticados (como o ser humano) preservaram em sua estrutura comportamentos reativos. Estes mecanismos reativos não fazem nenhuma reflexão sobre as situações, mas são muito úteis para desencadear respostas rápidas do organismo. Eles executam determinados comportamentos simples, voltados para o ambiente, em resposta à detecção de certos estímulos (SLOMAN, 1999).

- (237) Na arquitetura CAES, o *sistema reativo* ( $\mathcal{R}$ ) implementa os impulsos inatos do agente, também ativados a partir da detecção de certos contextos. Estes impulsos têm como consequência o desencadeamento de ações externas. A diferença para o sistema

emocional é que essas reações são transmitidas da mente ao corpo pelo sinal de controle externo ( $C_{\xi}$ ), e portanto geralmente induzem comportamentos que alteram o ambiente.

(238) Tal como acontece na mente humana (LEDOUX, 1996, 2000), (ÖHMAN, 2005), na mente do agente, os impulsos reativos podem entrar em conflito com os sinais de controle gerados pelo sistema cognitivo. Um mecanismo de inibição deve ser utilizado por um ou outro sistema, garantindo a prioridade de seu sinal. O sistema cognitivo informa-se sobre a ação que é de fato realizada através do feedback proprioceptivo, pois uma parte do sinal perceptivo interno ( $P_{\beta}$ ) comunica à mente sobre o estado dos atuadores do corpo.

### 2.5.5. Cognição e Aprendizagem

(239) Quando se interpreta a ideia de acoplamento dinâmico de forma muito radical, o efeito colateral é acabar negligenciando os aspectos cognitivos dos agentes, reduzindo-os a agentes reativos. Por outro lado, modelos de inteligência demasiadamente cognitivistas tendem a ignorar certos mecanismos de adaptação presentes na natureza, como é o caso dos instintos, das reações e dos comportamentos associativos. Discussões sobre estas duas ênfases podem ser colhidas em (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), e (BARANDIARAN; MORENO, 2006, 2008).

(240) Um *agente reativo* é um mecanismo do tipo estímulo-resposta. Ele não organiza uma memória explícita das suas experiências, e não coordena suas próprias ações no tempo. Não cria elementos conceituais para interpretar ou compreender o seu mundo. A única capacidade deste tipo de agente é associar entradas sensoriais à saídas motoras. Para um agente reativo só existe o presente, e as percepções sensoriais instantâneas. Diferentemente, um *agente cognitivo* faz uma representação da dinâmica de sua interação com o ambiente, sendo capaz de organizar a experiência passada, realizar antecipações, e planejar ações futuras.

(241) A existência de mecanismos cognitivos abre um horizonte infinito para o agente, no limite, permitindo que ele realize a construção de conceitos abstratos e teorias, o

pensamento através de raciocínios hipotético-dedutivos, a constituição de relativizações e analogias, e a compreensão do mundo através de relações causais e espaço-temporais.

(242) Diversas teorias psicológicas, entre as quais o construtivismo (PIAGET, 1936, 1937, 1964), sugerem que o desenvolvimento da inteligência no ser humano começa pela aprendizagem de coordenações sensório-motoras, e esses conhecimentos práticos vão dando origem, pouco a pouco, a modelos mentais cada vez mais conceituais e abstratos para interpretar, antecipar, e interagir com a realidade.

(243) No caso do agente cognitivo, a aprendizagem ocorre pela reorganização do modelo de mundo que ele representa nas suas estruturas intelectuais internas, na medida em que passa por novas experiências. Neste caso, o agente pode se adaptar ao ambiente através da melhora de seu modelo interno, que leva, conseqüentemente, a alterações no seu comportamento.

### 2.5.6. Sistema Cognitivo

(244) Uma vez que o CAES se propõe a ser uma arquitetura geral para modelar agentes autônomos, a exata constituição do sistema cognitivo é deixada em aberto. Dessa forma, é possível utilizá-la com diversos mecanismos de inteligência artificial.

(245) O *sistema cognitivo* ( $\mathbb{K}$ ) de uma mente ( $\mu$ ) na arquitetura CAES pode ser desde um simples mecanismo de estímulo e resposta, que se limite a gerenciar associações, até um mecanismo intelectual altamente complexo, que implemente níveis diferenciados de controle e interpretação dos eventos.

(246) O sistema cognitivo do agente exerce as funções de aprendizagem e deliberação. Cabe a ele construir o conhecimento do agente a respeito do mundo e de si mesmo, e também planejar ações e tomar decisões. Evidentemente é o sistema cognitivo a parte mais complexa e importante do agente enquanto um sistema inteligente, e será portanto objeto do próximo capítulo, dedicado à definição do CALM, um mecanismo desenvolvido para exercer o papel de sistema cognitivo na arquitetura CAES.

(247) A estrutura definida pela arquitetura CAES impõe a utilização de mecanismos que realizem cognição situada, ou seja, considerando a existência de uma inteligência sensório-motora antes da inteligência simbólica, e um processo que evolui de uma para a outra. O abstrato constrói-se à partir do sensoriomotor (PIAGET, 1945). A cognição

situada concilia o uso dos contextos perceptivos reais, com o uso de representações e camadas conceituais que interpretam os eventos do mundo. O desenvolvimento da inteligência começa pela aprendizagem de coordenações sensório-motoras, e esses conhecimentos práticos vão dando origem, pouco a pouco, a modelos mentais cada vez mais conceituais e abstratos, capazes de interpretar, antecipar, e lidar com uma realidade complexa.

### 3. CALM: MECANISMO DE APRENDIZAGEM CONSTRUTIVISTA

(248) Neste capítulo apresentamos o mecanismo CALM (*Constructivist Anticipatory Learning Mechanism*), desenvolvido para exercer o papel de sistema cognitivo em agentes definidos através da arquitetura CAES (apresentada no capítulo precedente), realizando tanto a aprendizagem de modelos de mundo, quanto a construção de políticas de ação para a tomada de decisão.

(249) O problema da aprendizagem de modelos de mundo é equivalente ao problema da descoberta da estrutura de um processo de decisão markoviano fatorado e parcialmente observável (FPOMDP). Já o problema de construir uma política de ações, na literatura científica, é sinônimo de “solucionar” um processo de decisão, propriamente. Isto é demonstrado ao longo da primeira seção deste capítulo, dedicada à caracterização, formalização e análise dos problemas de aprendizagem enfrentados. A seguir, a segunda seção do capítulo apresenta em detalhe o mecanismo CALM, projetado para resolver estes dois problemas.

#### 3.1. Definição dos Problemas de Aprendizagem

(250) Ao longo desta seção, mostraremos como os problemas de aprendizagem se colocam para um agente definido segundo a arquitetura CAES (apresentada no capítulo anterior). De uma forma geral, o agente enfrenta dois grandes problemas de aprendizagem: (a) realizar a *construção de um modelo de mundo* que descreva a dinâmica das relações no sistema global, e a partir dele (b) realizar a *construção de uma política de ações* visando a obtenção de seus objetivos, ou seja, a maximização das situações afetivamente positivas ao longo do curso de sua vida.

### 3.1.1. Aprendizagem de Modelos de Mundo

(251) Por *aprendizagem de modelo de mundo*, nos referimos ao processo em que um agente, inserido em um determinado ambiente que para ele é novo, constrói uma representação descritiva para os eventos que observa durante sua experiência.

(252) A entrada para um algoritmo de aprendizagem de modelos de mundo na arquitetura CAES é um fluxo contínuo do tipo  $\{p^{(0)}c^{(0)}, p^{(1)}c^{(1)}, \dots, p^{(i)}c^{(i)}\}$ . Os elementos  $p$  são sinais perceptivos, que correspondem a observações sucessivas feitas pela mente do agente através de seus sensores, e os elementos  $c$  são sinais de controle do agente, que correspondem a ações executadas através de seus atuadores. Considerando-se uma representação discreta com relação ao tempo, o par  $p^{(i)}c^{(i)}$  indica a observação realizada no instante  $t=i$  e a ação realizada a seguir. A tarefa do algoritmo de aprendizagem de modelo de mundo é induzir, a partir das experiências, uma estrutura  $\Psi$  tal que, dada uma nova instância  $p^{(i)}c^{(i)}$ , seja possível prever  $p^{(i+1)}$ .

(253) O sinal perceptivo  $p$  representa a situação observada num dado momento. O domínio de  $p$  é fatorado em  $|P|$  propriedades, correspondendo ao espaço  $P = P_1 \times P_2 \times \dots \times P_{|P|}$ . O agente induz mudanças no estado do sistema através do sinal de controle  $c$ , que representa suas ações. Por sua vez, o domínio de  $c$  é fatorado em  $|C|$  propriedades, correspondendo ao espaço  $C = C_1 \times C_2 \times \dots \times C_{|C|}$ .

(254) A representação construída pelo agente não modela o mundo propriamente dito, mas sim a interação entre o agente e o ambiente. Isto porque o agente experimenta o ambiente centrado sob seu próprio ponto de vista, e limitado pelas restrições de seus sensores e atuadores. De fato, o que o agente aprende é o modelo da dinâmica de interações entre ele e o mundo, e não um modelo do mundo diretamente.

(255) O modelo de mundo  $\Psi$  possível é portanto um modelo antecipatório, que descreve, a partir do ponto de vista do agente, a regularidade das transformações das propriedades do ambiente (e do seu próprio corpo) ao longo do tempo, segundo a observação que faz e as ações que executa.

### 3.1.2. Construção de Política de Ações

(256) O problema da *construção de política de ações* é chamado na literatura de aprendizagem de máquina como problema de *decisão sequencial*. Trata-se de fazer com

que o agente aprenda uma boa política de ações através da experimentação em um dado ambiente.

(257) Diferentemente da aprendizagem supervisionada, na qual o agente tem acesso a uma fonte de exemplos a respeito dos comportamentos corretos ou desejáveis, na aprendizagem por reforço o agente precisa aprender a partir de sua própria experiência, por tentativa e erro, e tendo apenas uma noção inexata e possivelmente tardia de quanto suas ações estão sendo realmente úteis.

(258) Na definição clássica, (KAELBLING et al., 1996), (SUTTON; BARTO, 1998), (RUSSELL; NORVIG, 1995), o agente recebe um sinal de reforço (um valor positivo ou negativo), muitas vezes esporádico, que representa recompensas ou punições de diferentes intensidades dadas ao agente em função das ações que ele faz. Com base na observação continuada das recompensas recebidas durante sua interação com o mundo, a tarefa do agente é construir uma política de ações que maximize o valor médio das recompensas ao longo do tempo.

(259) Embora existam métodos de aprendizagem por reforço “livres de modelo” (QUINLAN, 1986, 1993), que aprendem uma política de ações a partir das interações imediatas, neste trabalho consideramos o problema da construção de uma política de ações atrelado ao problema da construção de um modelo de mundo. Neste caso, o agente aprende um modelo de mundo, e a partir dele calcula uma política.

(260) Na arquitetura CAES, os sinais  $p$  da percepção estão associados a sinais de avaliação, dados pelas funções  $\pi$ , que indicam o valor imediato de determinadas situações para o agente. Neste caso, a entrada para um algoritmo de aprendizagem de política de ações é o modelo de mundo corrente  $\Psi^{(t)}$ , e o modelo de recompensas corrente  $\mathcal{U}^{(t)}$ . O modelo de mundo  $\Psi : P \times C \rightarrow P'$  é uma função antecipatória, onde  $p$  representa uma observação,  $c$  representa uma ação executada pelo agente, e  $p'$  representa a expectativa da observação para o instante seguinte. O modelo de recompensas  $\mathcal{U} : P \rightarrow \mathfrak{R}$  representa o valor esperado do sinal avaliativo  $\pi$  quando  $p$  é observado.

(261) A tarefa do algoritmo é definir uma política de ações a partir dos modelos de mundo ( $\Psi$ ) e de avaliação ( $\mathcal{U}$ ) dados. A política de ações é representada como um mapeamento entre as situações possíveis do ambiente e as ações a serem tomadas

quando o agente está nelas ( $\pi : P \rightarrow C$ ). A política ótima é aquela que define sequências de ações que maximizam a soma descontada dos sinais de avaliação num horizonte de tempo de longo-prazo (ou seja, um intervalo maior do que o imediato). Entretanto, para o tipo de problema colocado quando se utiliza a arquitetura CAES, a solução (ou seja, uma boa política) pode se basear na estabilidade da dinâmica e não necessariamente na sua otimalidade. Soluções « suficientemente boas » são válidas e podem permitir ao agente manter-se por muito tempo dentro dos limites de viabilidade.

### 3.1.3. Aprendizagem Ativa, em Horizonte Infinito, e Incremental

(262) Conforme as premissas adotadas por este trabalho, sobretudo pela utilização da arquitetura CAES como definição da relação agente-ambiente, os problemas de aprendizagem são colocados para o agente como casos de *aprendizagem ativa*, sob um *horizonte infinito, e incremental*.

#### 3.1.3.1. Aprendizagem Ativa

(263) Num sistema agente-ambiente como o CAES, o agente interfere no fluxo do sistema global através de suas ações. O agente perturba o ambiente e observa os resultados desta perturbação. Quando o agente é mais do que um mero observador, ou seja, quando é ele quem modifica o estado do ambiente através de suas ações, e quando ele precisa, ao mesmo tempo, aprender um modelo de mundo à partir dessa interação, então fica caracterizado o caso de aprendizagem ativa (RON; RUBINFELD, 1997).

(264) Opondo-se ao modo de aprendizagem passiva, onde o agente é apenas um observador, sem controle sobre o fluxo do processo, o modo de aprendizagem ativa atrela o problema da aprendizagem do modelo de mundo aos problemas de decisão e controle. Na aprendizagem ativa é o agente quem escolhe as ações a tomar, e, portanto, para que possa aprender um modelo de mundo adequado, precisa resolver o dilema entre explorar e usufruir (*exploration x exploitation*). Ou seja, o agente precisa adotar alguma estratégia para ponderar entre a escolha de ações que visam descobrir novas coisas do ambiente (*exploration*), ou ações que utilizem o conhecimento já adquirido com o fim de conduzir o agente a atingir seus objetivos (*exploitation*).

(265) Em se tratando da aprendizagem de políticas de ação, se não há uma estratégia de exploração, o algoritmo acaba convergindo para um *máximo local*, ou seja, uma

política que é boa apenas se comparada à sua vizinhança próxima no espaço de busca das políticas.

(266) Se por um lado o caso de aprendizagem ativa implica esta dificuldade adicional, ou seja, encontrar uma heurística adequada que resolva o dilema entre explorar e desempenhar, por um lado, quando o problema é bem resolvido, a capacidade de aprendizagem do agente aumenta, pois ele pode realizar “experiências” no ambiente, traçando estratégias para deliberadamente explorar caminhos pouco conhecidos. A aprendizagem passiva é, neste aspecto, mais restritiva, pois se não há controle sobre as ações, é preciso garantir estatisticamente que sequências aleatórias poderão explorar adequadamente o ambiente (RON; RUBINFELD, 1997).

### 3.1.3.2. *Aprendizagem em Horizonte Infinito*

(267) Em qualquer arquitetura inspirada em modelos naturalistas (como é o CAES), o agente é inserido no ambiente para experimentar o mundo no “tempo de sua vida”, assim como faz um organismo vivo. Se o agente precisa aprender um modelo de mundo nessa condição, então fica caracterizado o caso de aprendizagem em horizonte infinito.

(268) Muitos problemas em IA são modelados em horizontes episódicos. Em geral significa que existem estados finais específicos que, uma vez atingidos, reiniciam o sistema, reposicionando o agente no estado inicial. Em alguns casos o agente tem ele próprio acesso a uma ação de reinicialização (*reset*), que permite recolocar o sistema no estado inicial. Em outros existe um número máximo de ciclos por episódio, após o que sucede-se um novo episódio, reiniciando o estado do sistema.

(269) A aprendizagem em horizonte infinito é mais difícil, sobretudo em problemas parcialmente observáveis, porque não há recurso a reinicialização, e portanto, não há um meio de retornar diretamente a um estado conhecido. O agente precisa aprender a prever as próximas observações somente com base em uma única e ininterrupta “caminhada” sobre o ambiente.

(270) O mecanismo de reinicialização permite que o agente tenha certeza sobre ao menos um estado, o inicial, de onde as caminhadas sempre começam. Muitos algoritmos de aprendizagem baseiam-se nisto para construir seus modelos (RON; RUBINFELD, 1997). Sem este recurso, a certeza de ter reencontrado um estado já

visitado anteriormente desaparece, e algoritmos de aprendizagem apoiados nessa premissa tornam-se inviáveis.

### 3.1.3.3. *Aprendizagem Incremental*

(271) A arquitetura CAES impõe ao agente um modo de aprendizagem incremental. O agente precisa aprender pouco a pouco, ao mesmo tempo em que desempenha suas atividades. O agente deve construir o modelo de mundo gradativamente, e então cada experiência particular é uma nova informação que pode ser utilizada para refiná-lo.

(272) A aprendizagem incremental é *online*, ou seja, não há separação entre o tempo de aprender e o tempo de desempenhar. O agente não dispõe de um conjunto de exemplos de treinamento a priori. Ter um banco de dados de casos à disposição desde o início do processo de aprendizagem permitiria que a construção da hipótese fosse realizada depois de serem analisados todos os exemplos. No modo incremental, pelo contrário, a construção do modelo está sujeita a um começo equivocado, induzido por exemplos iniciais não representativos, e o algoritmo de aprendizagem precisa levar em conta essa possibilidade, em outras palavras, precisa ser capaz de voltar atrás em certas escolhas tomadas ao longo da construção do modelo.

### 3.1.4. **Representando o Sistema como Processo de Decisão de Markov**

(273) Muitos algoritmos propostos para problemas de decisão e aprendizagem utilizam representações do ambiente baseadas na enumeração de estados, frequentemente definidos como um *Processo de Decisão de Markov* (MDP).

(274) Um MDP fornece ao agente a informação completa do estado em que ele se encontra, ou seja, o agente percebe diretamente os estados do sistema. Entretanto, em muitos problemas esta informação completa não é possível ou não está disponível. Neste caso, um ambiente é tipicamente representado através de um *Processo de Decisão de Markov Parcialmente Observável* (POMDP), onde o conjunto de estados do ambiente é apenas parcialmente observável pelo agente. Num POMDP, o agente tem acesso a uma informação que indica apenas de maneira parcial e indireta o estado atual do sistema.

(275) O problema é que em representações baseadas na enumeração de estados, tais como MDP e POMDP, o tamanho do problema cresce exponencialmente com o número

de atributos considerados, tornando tais soluções inviáveis para problemas grandes. Em vista disso, a comunidade de IA tem utilizado, cada vez mais, representações fatoradas, onde não se enumeram os estados, mas tratam-se diretamente as propriedades.

(276) O uso de *Processos de Decisão de Markov Fatorados* (FMDP) possibilita explorar a estrutura do ambiente, permitindo sua representação de forma compacta, mesmo no caso em que o MDP correspondente é exponencialmente grande (GUESTRIN et al., 2003). Além disso, o formalismo de FMDP também pode ser estendido para representar observabilidade parcial, constituindo então um *Processo de Decisão de Markov Fatorado e Parcialmente Observável* (FPOMDP).

#### 3.1.4.1. *Processo de Decisão de Markov*

(277) O *Processo de Decisão de Markov* foi introduzido por (BELLMAN, 1957) e (HOWARD, 1960), e popularizou-se entre a comunidade acadêmica na década de 1990, (PUTERMAN, 1994), (SUTTON; BARTO, 1998), (RUSSELL; NORVIG, 1995), (FEINBERG; SHWARTZ, 2002). Um MDP representa o ambiente através da enumeração de seus estados, de forma semelhante a uma *máquina de estados* cuja função de transição pode ser não-determinística. No MDP, o fluxo do sistema depende das ações executadas pelo agente, havendo um sinal de recompensa para determinadas transições, o que caracteriza um problema de controle.

(278) No contexto de agentes, um MDP pode ser formalizado como sendo uma 4-upla  $\{Q, A, \delta, r\}$ , de acordo com a definição 3.1, onde  $Q$  é o conjunto finito e não-vazio de estados do sistema,  $A$  é o conjunto de ações do agente,  $\delta$  é a função de transição do sistema, e  $r$  é a função de recompensa. Quando o sistema é não-determinístico, então as funções de transição e recompensa são matrizes de distribuição de probabilidades. Pode-se estabelecer uma função  $\delta^0$  para definir o estado inicial.

Um MDP é uma 4-upla:

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

onde

$Q = \{q_1, q_2, \dots, q_{|Q|}\}$  é um conjunto finito e não-vazio de *estados*

$A = \{a_1, a_2, \dots, a_{|A|}\}$  é um conjunto finito e não-vazio de *ações* do agente

$\delta : Q \times A \rightarrow \Pi(Q)$  é uma função probabilística de *transição de estados*  
definida pela matriz  $\delta = \text{prob}(q' | q, a)$

$r : Q \times A \rightarrow \Pi(\mathfrak{R})$  é uma função probabilística de *recompensa*  
definida pela matriz  $r = \text{prob}(r' | q, a)$

adicionalmente

$\delta^0 : \rightarrow \Pi(Q)$  é uma função probabilística de estado inicial  
definida pela matriz  $\delta^0 = \text{prob}(q^{(0)})$

Definição 3.1: Processo de Decisão de Markov.

(279) A representação do ambiente através de um MDP supõe implicitamente que o sistema global está de acordo com quatro restrições (BOUTILIER et al., 2000), (McALLESTER; SINGH, 1999). Primeiramente, um MDP é um modelo discreto. Ele discretiza o espaço de fase do sistema em um número finito de estados, e também discretiza o tempo, fragmentando-o em instantes bem determinados (ciclos). Em outras palavras, não há estados intermediários entre dois estados quaisquer  $q_i$  e  $q_j$ , assim como não há acontecimentos no sistema entre dois instantes de tempo sucessivos  $t$  e  $t'$ .

(280) Em segundo lugar, um MDP descreve as transições entre os estados como funções probabilísticas markovianas de primeira ordem. A hipótese de Markov supõe que a probabilidade da ocorrência de uma transição entre os estados  $q$  e  $q'$  depende apenas do estado atual do sistema ( $q$ ) e da última ação tomada pelo agente ( $a$ ), não sendo preciso considerar uma sequência histórica de estados passados ( $q^{(t-1)}$ ,  $q^{(t-2)}$ , ...). Significa que a história não tem influência na evolução futura do processo se o presente está completamente especificado, ou então, de outro modo, que a história relevante já está incorporada na representação dos estados.

(281) Em terceiro lugar, um MDP representa um sistema estacionário, onde a função de transição entre os estados não se modifica no decorrer do tempo, ou seja, um sistema em que as probabilidades relativas à ocorrência de uma transição são independentes do

instante particular em que elas se dão, e portanto, para quaisquer dois instantes de tempo  $t_i$  e  $t_j$ ,  $\delta^{(i)} = \delta^{(j)}$ .

(282) Finalmente, um MDP define um ambiente completamente observável, onde toda a informação necessária para descrever a função de evolução do sistema está disponível para o agente. Assim, a representação através de MDP considera que a percepção do agente lhe informa direta e inequivocamente o estado atual ( $q$ ) em que o sistema se encontra.

#### 3.1.4.2. Representando a Observação Parcial

(283) Em muitos problemas, sobretudo aqueles do mundo real, a informação completa sobre o estado atual do ambiente não está disponível. Problemas desse tipo são em geral representados através de *Processos de Decisão de Markov Parcialmente Observáveis* (POMDPs), conforme (MEULEAU et al., 1999), (SHANI et al., 2005) e (HOLMES; ISBELL, 2006). A definição do POMDP foi inicialmente proposta por (SMALLWOOD; SONDIK, 1973), e popularizada em trabalhos como (CHRISTMAN, 1992) e (KAELBLING et al., 1994, 1998).

(284) O POMDP, conforme a definição 3.2, é uma extensão do MDP, onde se especifica um conjunto de observações e uma função de observação. Nessa representação, o estado subjacente  $q$  do ambiente não é acessível ao agente, que só tem acesso a uma observação dada em função dele.

Um POMDP é uma 6-upla:

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

onde

$Q = \{q_1, q_2, \dots, q_{|Q|}\}$  é um conjunto finito e não-vazio de *estados*

$A = \{a_1, a_2, \dots, a_{|A|}\}$  é um conjunto finito e não-vazio de *ações* do agente

$O = \{o_1, o_2, \dots, o_{|O|}\}$  é o conjunto finito e não-vazio de *observações* do agente

$\delta : Q \times A \rightarrow \Pi(Q)$  é uma função probabilística de *transição de estados*

definida pela matriz  $\delta = \text{prob}(q' | q, a)$

$\gamma : Q \times A \rightarrow \Pi(O)$  é uma função probabilística de *observação*

definida pela matriz  $\gamma = \text{prob}(o' | q, a)$

$r : Q \times A \rightarrow \Pi(\mathcal{R})$  é uma função probabilística de *recompensa*

definida pela matriz  $r = \text{prob}(r' | q, a)$

adicionalmente

$\delta^0 : \rightarrow \Pi(Q)$  define uma função de probabilidade para o estado inicial

definida pela matriz  $\delta^0 = \text{prob}(q^{(0)})$

Definição 3.2: Processo de Decisão de Markov Parcialmente Observável.

(285) POMDPs definem problemas de decisão e de aprendizagem de maneira mais realista que os MDPs completamente observáveis, entretanto a dificuldade de resolver tais problemas aumenta consideravelmente. A complexidade resultante da falta de informação sobre o estado limita a aplicação de POMDPs a problemas pequenos, ou a domínios com características bastante restritas (MEULEAU et al., 1999).

(286) A solução de um POMDP é muito mais custosa do que seria a de seu correspondente MDP, e a utilização de representações “rasas” (*flat*), baseadas na enumeração de estados e ações, é o grande fator limitante. Muitos trabalhos recentes estão apostando na representação estruturada, usando estados e ações fatorados, como forma de superar o problema de escalabilidade em problemas de aprendizagem e decisão em ambientes parcialmente observáveis (POUPART; BOUTILIER, 2004), (SHANI et al., 2008).

### 3.1.4.3. Problema de Escalabilidade

(287) Existe um problema de escalabilidade em qualquer representação baseada em enumeração de estados, como é o caso dos MDPs e POMDPs. É que o número de exemplos necessários para se aprender um certo conceito cresce exponencialmente de

acordo com o número de atributos utilizados para sua representação (VALIANT, 1984). Este tipo de representação só é viável se existe de antemão um modelo, ou seja, algum conhecimento inerente ao domínio que simplifique o conjunto de características a serem consideradas no ambiente, possibilitando que o sistema seja representado por um número limitado de estados relevantes.

(288) Se um agente é inserido em um ambiente desconhecido para ele (caso mais típico), então o número de estados advém da combinação cruzada de cada uma das suas percepções. Por exemplo, num universo composto de  $n$  variáveis binárias, o número de estados enumerados é  $|Q| = 2^n$ . Assim, se considerarmos que cada percepção descreve uma propriedade do ambiente, então o problema é que o número de estados aumenta exponencialmente com o número de propriedades consideradas. Esse fenômeno é chamado de *maldição da dimensionalidade* (BELLMAN, 1961).

(289) Para piorar, um algoritmo de aprendizagem de modelos de mundo baseado na representação enumerada dos estados enfrentaria um espaço de busca de tamanho exponencial sobre o número de estados, que já é exponencial com relação ao número de propriedades. Quer dizer que podem existir  $|Q|^n$  diferentes ambientes representáveis a partir de  $n$  propriedades. Portanto, em ambientes complexos, como o mundo real, um problema representado através da enumeração dos estados torna-se inerentemente intratável (DEGRIS et al., 2006).

(290) O uso de representações fatoradas pode manter o tamanho do problema numa ordem polinomial aceitável (BOUTILIER et al., 2000). Entretanto, para que isto seja verdadeiro, o ambiente tratado precisa ser bem estruturado, como sera visto adiante.

#### 3.1.4.4. Fatorando os Conjuntos Enumerados

(291) Para evitar a enumeração dos estados, são utilizadas representações baseadas diretamente nas características do ambiente, ou seja, em vez de enumerar uma lista de estados atômicos, o ambiente passa a ser descrito através de um conjunto de propriedades ou atributos. O formalismo utilizado recentemente com esta finalidade é o *Processo de Decisão de Markov Fatorado* (FMDP), inicialmente apresentado em (BOUTILIER et al., 1995).

(292) Um FMDP pode ser formalizado como uma 4-upla  $\{X, C, \tau, \pi\}$ , conforme a definição 3.3, onde  $X$  é o conjunto de propriedades (que definem o estado do ambiente),

$C$  é o conjunto de variáveis de controle (que definem as ações do agente),  $\tau$  é um conjunto de funções de transformação, de tal forma que  $\tau_i$  descreve a evolução da propriedade  $i$ , e  $\pi$  é um conjunto de funções de recompensa, de tal forma que  $\pi_i$  descreve a função de recompensa fatorada para a propriedade  $x_i$ . O FMDP é um modelo não-determinístico, então as funções definem distribuições probabilísticas.

Um FMDP é uma 4-upla:

$$\mathcal{C} = \{X, C, \tau, \pi\}$$

onde

$X = \{X_1, X_2, \dots, X_{|X|}\}$  é um conjunto de *propriedades*

$C = \{C_1, C_2, \dots, C_{|C|}\}$  é um conjunto de *variáveis de controle*

$\tau = \{\tau_1, \tau_2, \dots, \tau_{|X|}\}$  é um conjunto de funções de *transformação*

$$\text{tal que } \tau_i : X \times C \rightarrow \Pi(X_i)$$

$\pi = \{\pi_1, \pi_2, \dots, \pi_{|X|}\}$  é um conjunto de funções de *avaliação*

$$\text{tal que } \pi_i : X_i \rightarrow \Pi(\mathcal{R})$$

Definição 3.3: Processo de Decisão de Markov Fatorado.

(293) O uso de FMDPs possibilita explorar a estrutura do ambiente, permitindo sua representação de forma compacta, mesmo no caso em que o MDP correspondente é exponencialmente grande. Nessa formalização, estados, ações, e recompensas, estão fatorados em propriedades, representadas por variáveis aleatórias. As vantagens da fatoração são discutidas em (GUESTRIN et al., 2003), (BOUTILIER et al., 2000), (JONSSON; BARTO, 2005), (DEGRIS et al., 2006, 2008), e (TRIVIÑO; MORALES, 2000).

(294) Como não há necessidade de enumerar todos os possíveis estados do espaço, definidos pela combinação das dimensões fornecidas pelas propriedades, então um FMDP pode representar a função de transformação de cada propriedade de forma independente. A função de evolução do sistema global é equivalente à combinação dessas funções de transformação particulares.

(295) O conjunto de funções de transformação pode ser representado através de uma *Rede Bayesiana Dinâmica* (DBN) especial, como um grafo de duas camadas, acíclico e orientado, apresentado em (DEAN; KANAZAWA, 1989) e (DARWICHE; GOLDSZMIDT, 1994). Os nós da primeira camada representam as propriedades e as variáveis de controle ( $X \cup C$ ), indicando respectivamente o estado do ambiente e as

ações que podem ser realizadas pelo agente no tempo  $t$ . A segunda camada representa apenas as propriedades do sistema ( $X'$ ), indicando seu valor no instante seguinte  $t + 1$ .

(296) No caso de um DBN completo, cada nó da segunda camada está associado a uma distribuição de probabilidade condicional, que descreve o comportamento da propriedade em função do valor dos nós da primeira camada, conforme ilustra a figura 3.1.

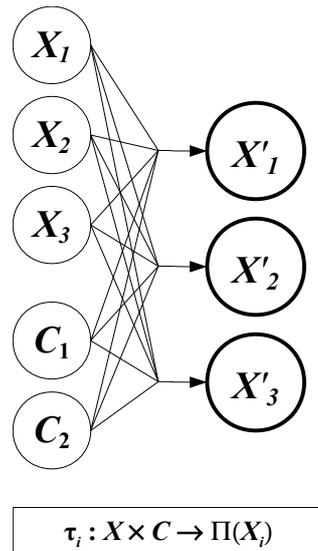


Figura 3.1: Exemplo de uma rede bayesiana dinâmica (DBN).

Um DBN de duas camadas, completamente conectado, acíclico e orientado, representa as funções de transformação do sistema. O exemplo é composto por 3 propriedades e 2 variáveis de controle. À esquerda representa-se o estado do sistema em  $t$ , e à direita, sua transformação em  $t+1$ .

(297) Simplesmente para bem demonstrar a independência entre cada função de transformação  $\tau_i$  (de cada propriedade), seria formalmente equivalente representar cada uma delas através de um DBN particular. Nesse caso, a segunda camada do grafo possuiria um único nó, indicando o valor de uma determinada propriedade específica do ambiente ( $X'_i$ ) no instante seguinte  $t + 1$ , como mostra a figura 3.2. O fato de que cada função de transformação pode ser descrita independentemente das outras é importante pois permite a implementação de um mecanismo de aprendizagem que paralelize o processo de construção do modelo de mundo.

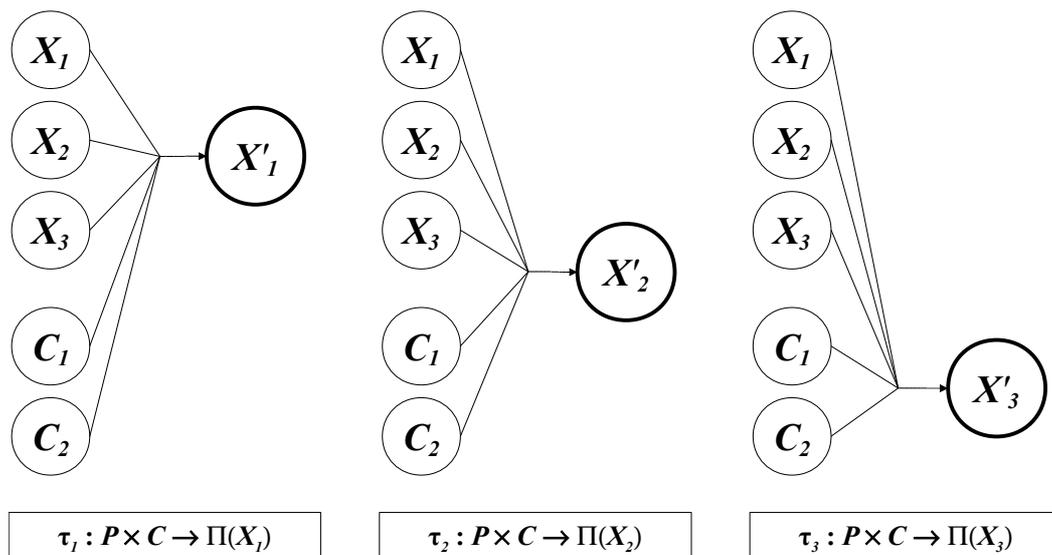


Figura 3.2: Exemplo de DBNs completos mas independentes.

### 3.1.5. Ambientes Estruturados

(298)

Em geral, ambientes complexos são descritos através de um número muito grande de propriedades. Entretanto, se o ambiente é estruturado, então os eventos tem causas precisas e limitadas. Dito de outro modo, em *ambientes estruturados*, para explicar a dinâmica de uma determinada propriedade  $X_i$ , em geral basta prestar atenção a um pequeno conjunto de *variáveis relevantes* para a sua função de transformação, conjunto este definido como  $rel(\tau_i) \subset (X \cup C)$ .

(299)

É por isso que um DBN pode representar a função de transformação de uma propriedade de maneira compacta, desde que o conjunto de propriedades relevantes seja conhecido. Na primeira camada coloca-se apenas esse subconjunto de nós relevantes  $rel(\tau_i)$ , em vez de representar o domínio completo  $X \times C$ . Na segunda camada coloca-se apenas o nó  $X'_i$  da propriedade a ser antecipada. A figura 3.3 mostra um exemplo de funções de transformação representadas por DBNs em um ambiente hipotético. Note-se que apenas as variáveis relevantes na primeira camada do grafo estão conectadas com o nó de antecipação na segunda camada.

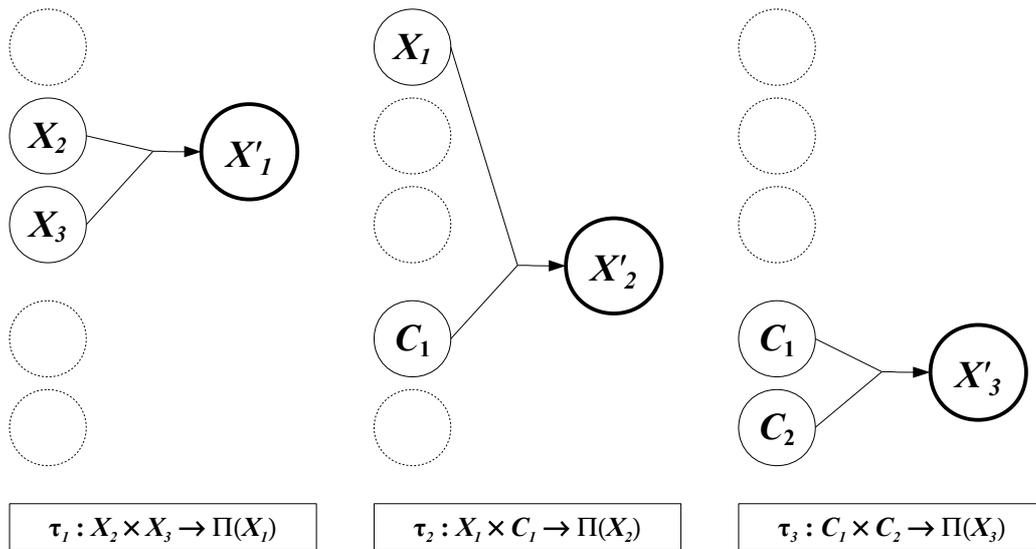


Figura 3.3: Exemplo de DBNs independentes e compactos.

Os DBNs são compactos porque a primeira camada contém apenas as variáveis relevantes para descrever a transformação.

### 3.1.5.1. Relevância

(300) O conceito de relevância está relacionado à causalidade. As variáveis relevantes para descrever a dinâmica de uma dada propriedade são, em princípio, as causas determinadoras da transformação ou da permanência de seu estado. O conjunto de variáveis relevantes para a função de transformação específica de uma determinada propriedade é definido como  $rel(\tau_i)$ . O conjunto das propriedades relevantes em relação ao sistema como um todo,  $rel(\tau)$ , é dado pela união desses conjuntos particulares, na forma  $rel(\tau) = rel(\tau_1) \cup rel(\tau_2) \cup \dots \cup rel(\tau_{|X|})$ .

(301) Uma propriedade é relevante se existe ao menos um fenômeno que só se torna explicável em função dela. Dito de outro modo, uma propriedade  $X_i$  é relevante para uma transformação  $\tau_j$  se existe um par de experiências  $exp_1$  e  $exp_2$  (do tipo  $\{x, c, x'_j\}$  correspondendo respectivamente ao estado do sistema global, à ação do agente, e à transformação da propriedade  $X_j$  verificada no instante seguinte) para as quais  $x'_j$  seja divergente e  $\{x, c\}$  seja diferenciável apenas pelo elemento  $x_i$  (BLUM; LANGLEY, 1997). Significa que o valor de  $x'_j$ , que é regulado pela função de transformação  $\tau_j$ , é condicionalmente dependente do valor de  $x_i$ .

### 3.1.5.2. Grau de Estruturação de um Ambiente

(302) Um FMDP pode ser classificado segundo seu *grau de estruturação* ( $\varphi$ ). A ideia por trás dessa medida é a seguinte: quanto mais reduzido é o número de causas que determinam os fenômenos, mais estruturado o ambiente é. Em outras palavras, o grau de estruturação indica o quanto está concentrada ou está diluída a informação necessária para a descrição da dinâmica do mundo. Quanto menor é o número médio de propriedades relevantes para descrever as transformações, então mais estruturado é o ambiente. Conseqüentemente, o grau de estruturação indica o quanto as funções de transformações podem ser representadas de maneira compacta.

(303) O grau de estruturação  $\varphi_i$  de uma determinada propriedade  $X_i$  é calculado com relação à correspondente função de transformação  $\tau_i$ . Ele é dado pela razão inversa entre a quantidade de propriedades relevantes para descrever a transformação,  $rel(\tau_i)$ , e a quantidade de propriedades existentes no domínio completo  $X \cup C$ , conforme a equação 3.1.

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

(304) Finalmente, o grau de estruturação total  $\varphi$  do ambiente é determinado pela média do grau de estruturação das funções de transformação particulares, conforme a equação 3.2.

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

(305) Imaginando  $\varphi$  como um eixo, em um sentido encontramos ambientes em que as propriedades concentram cada vez mais a informação sobre o funcionamento do sistema, e no outro sentido, encontramos ambientes em que as propriedades têm cada vez menos significado sozinhas, no que diz respeito à determinação da dinâmica do ambiente.

(306) Num caso extremo, se  $rel(\tau_i) = \emptyset$ , então  $|rel(\tau_i)| = 0$ , e portanto  $\varphi_i = 1$ . Nesse caso a transformação  $\tau_i$  é condicionalmente independente, e então é possível definir o valor de  $X_i'$  sem fazer referência a nenhuma outra propriedade ou sinal de controle. Inversamente, quando  $rel(\tau_i) = X \cup C$ , então  $\varphi_i = 0$ . Nesse caso a função de

transformação  $\tau_i$ , que define o valor de  $X_i'$ , é condicionalmente dependente de todas as propriedades de  $X$  e  $C$ . De fato, quando o ambiente é muito desestruturado, a representação fatorada pode ser mais custosa do que a enumeração completa dos estados.

(307) Assim, a redução da complexidade da representação de um FMDP frente a seu correlativo MDP é diretamente proporcional a seu grau de estruturação. Quanto maior é o valor de  $\phi$ , menor é o número médio de variáveis relevantes necessárias para descrever a dinâmica do sistema, e mais compacta é uma representação baseada na fatoração dos estados em propriedades.

(308) No caso oposto, quanto menor é o valor de  $\phi$ , mais a informação sobre as transformações está distribuída entre as propriedades, e menos interessante se torna a representação fatorada quando comparada diretamente a construção de um modelo baseado na enumeração de estados.

(309) Retornando ao exemplo ilustrado na figura 3.3, calcula-se o grau de estruturação da propriedade  $X_l$  pela proporção entre as propriedades relevantes para a descrição da função de transformação,  $|rel(\tau_l)| = 2$ , em relação ao número total de propriedades que poderiam ser utilizadas,  $|X \cup C| = 5$ , resultando numa taxa de  $\phi_l = 0,6$ .

(310) Utilizando o cálculo de estruturação proposto, definiremos um ambiente bem estruturado como sendo aquele onde o número médio de propriedades relevantes para definir as transformações é de ordem, no máximo, logarítmica com relação ao número total de propriedades, ou seja,  $|rel(\tau)| \leq \log_b(|X \cup C|)$ .

### 3.1.5.3. Implicações para a Aprendizagem

(311) A figura 3.4 ilustra o motivo pelo qual os algoritmos clássicos de aprendizagem são inerentemente ineficientes. Eles ancoram-se na enumeração dos estados, que é exponencial em relação ao número de variáveis do problema. Já algoritmos baseados numa representação fatorada, e que procuram tirar proveito da estruturação do ambiente, apresentam uma estratégia quase-polinomial.

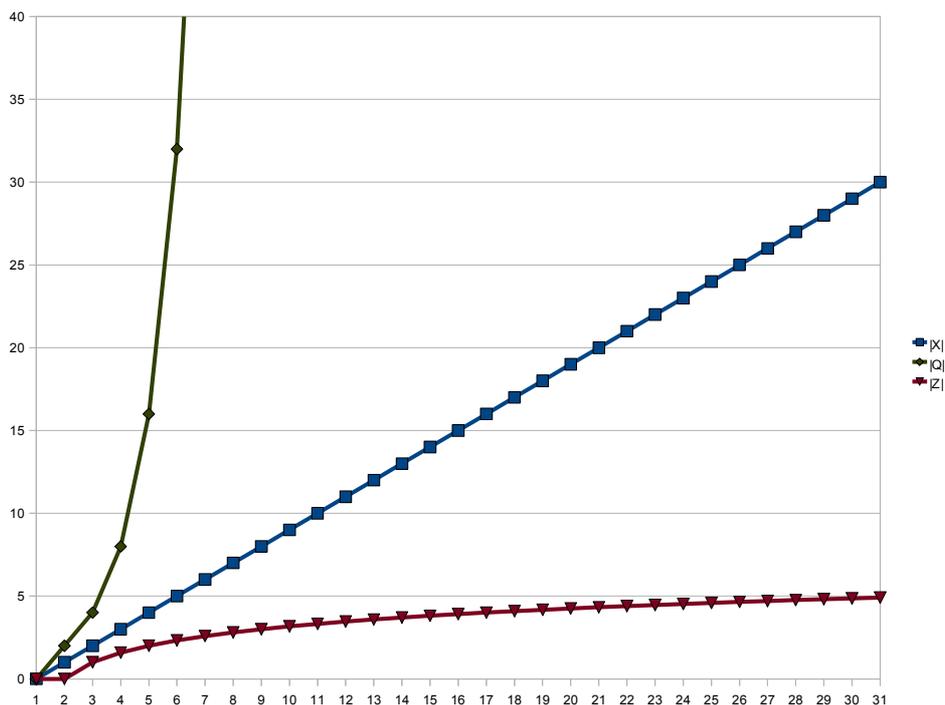


Figura 3.4: Relação de complexidade entre propriedades, estados, e relevância.

No gráfico, a linha  $|X|$  representa o número total de propriedades do problema (crescimento linear), a linha  $|Q|$  representa a quantidade de estados necessária para representar de maneira rasa a combinação dessas propriedades (exponencial), e a linha  $|rel|$  representa o número médio máximo de propriedades relevantes para a descrição das transformações em um ambiente bem estruturado (logarítmica). Considera-se um universo de variáveis binárias.

(312) No que toca o problema de aprendizagem de modelos de mundo, a questão é que nem todas as propriedades do ambiente são relevantes para descrever as funções de transformação. A existência de uma quantidade significativa de propriedades irrelevantes pode complicar bastante a tarefa do algoritmo de aprendizagem. Em geral, essa situação implica a tarefa adicional de identificar quais são as propriedades relevantes dentro do conjunto total de propriedades, problema que não tem uma solução óbvia. Algumas referências sobre o tema são (GUYON et al. 2006), (LIU; MOTODA, 2007), (BLUM; LANGLEY, 1997), (GUYON; ELISSEFF, 2003), (SIEDLECKI, SKLANSKY, 1993).

(313) Potencialmente, o conjunto de propriedades relevantes para uma determinada função de transformação  $\tau_i$  é qualquer subespaço dentro das possíveis combinações do espaço  $X \times C$ . Esses possíveis domínios da função  $\tau_i$  constituem uma hierarquia parcialmente ordenada pela inclusão de um domínio em outro. Se considerarmos

propriedades binárias, teremos  $2^{X \times C_1}$  combinações de espaços possíveis. A figura 3.5 ilustra esse espaço de possíveis combinações de relevância, num exemplo composto de 3 variáveis binárias, sendo duas propriedades, e uma variável de controle.

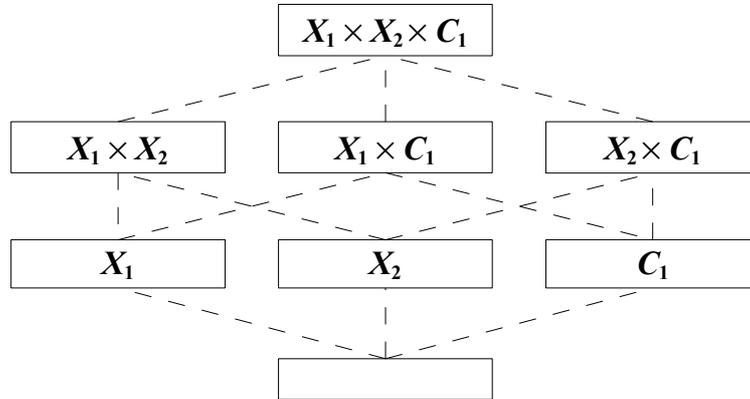


Figura 3.5: Hierarquia parcialmente ordenada de domínios possíveis.

Cada função de regularidade possui um conjunto de variáveis relevantes, que forma o seu domínio. Entretanto, o espaço de domínios possíveis totaliza  $2^{X \times C_1}$  combinações.

(314) Por exemplo, o domínio total  $X \times C$ , definido para  $\varphi = 0$ , contém todos os outros. De outro lado, o domínio vazio, definido quando  $\varphi = 1$ , representa o menor conjunto de propriedades, e está contido em todos os outros. Definimos o conjunto  $rel(\tau_i)$  das propriedades relevantes para a função  $\tau_i$  como o menor subconjunto contido em  $X \cup C$  suficiente para descrever a transformação da propriedade  $X_i$  sem perda de precisão.

### 3.1.6. Processo Fatorado e Parcialmente Observável

(315) Um *Processo de Decisão de Markov Fatorado e Parcialmente Observável* (FPOMDP) pode então ser visto como uma extensão do FMDP capaz de representar observabilidade parcial, ou como uma representação fatorada do POMDP. Trata-se da junção do FMDP e do POMDP compondo um MDP fatorado e parcialmente observável. Propostas nesse sentido aparecem em (BOUTILIER; POOLE, 1996), (HANSEN; FENG, 2000), (GUESTRIN et al., 2001), (POUPART; BOUTILIER, 2004), (POUPART, 2005), (SHANI et al., 2008), e (SIM et al., 2008).

(316) Um FPOMDP é formalizado como uma 5-upla  $\{P, H, C, \tau, \pi\}$ , conforme a definição 3.4, onde  $P$  é o conjunto das propriedades observáveis (acessíveis através da percepção), e  $H$  é o conjunto das propriedades ocultas (não-observáveis diretamente pela percepção),  $C$  é o conjunto de variáveis de controle (que definem as possibilidades

de atuação do agente),  $\tau$  é a função de evolução fatorada do sistema, e  $\pi$  é a função de recompensa fatorada.

Um FPOMDP é uma 5-upla:

$$\mathcal{C} = \{P, H, C, \tau, \pi\}$$

onde

$P = \{P_1, P_2, \dots, P_{|P|}\}$  é o conjunto de *propriedades observáveis*

$H = \{H_1, H_2, \dots, H_{|H|}\}$  é o conjunto de *propriedades ocultas*

$C = \{C_1, C_2, \dots, C_{|C|}\}$  é o conjunto de *variáveis de controle*

$\tau = \{\tau_1, \tau_2, \dots, \tau_{|X|}\}$  é um conjunto de funções de *transformação*

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

$\pi = \{\pi_1, \pi_2, \dots, \pi_{|P|}\}$  é um conjunto de funções de *avaliação*

tal que  $\pi_i : P_i \rightarrow \Pi(\mathfrak{R})$

adicionalmente

$$X = P \cup H$$

Definição 3.4: Processo de Decisão de Markov Fatorado e Parcialmente Observável.

(317) Um ambiente modelado como um FPOMDP coloca duas sutilezas para o problema de aprendizagem. De um lado, entre as propriedades observáveis do ambiente, nem todas são relevantes para descrever as funções de transformação, e isso implica o subproblema da *seleção de propriedades relevantes*. De outro lado, entre as propriedades relevantes, nem todas são observáveis, e isso implica no subproblema da *descoberta de propriedades não-observáveis*.

### 3.1.6.1. CAES estabelece um FPOMDP para a mente

(318) Na arquitetura CAES (definida no capítulo 2), o problema da construção de modelos de mundo equivale ao problema da descoberta da estrutura de um FPOMDP. Nesse modelo, o sistema cognitivo da mente do agente recebe um sinal perceptivo  $p$ , e induz mudanças no estado do sistema através de um sinal de controle  $c$ . Esses sinais são fatorados, portanto compondo-se de diversas propriedades que formam os espaços  $P = (P_1 \times P_2 \times \dots \times P_{|P|})$  e  $C = (C_1 \times C_2 \times \dots \times C_{|C|})$ . A função de evolução da percepção pode ser igualmente fatorada em  $|P|$  funções de transformação  $\tau_i$  para cada uma das percepções  $P_i$ .

(319) Assim, a relação entre agente e ambiente definida através de propriedades (e não da enumeração dos estados) é uma metáfora natural na arquitetura CAES, pois os sinais

de percepção e controle são definidos como vetores, onde cada elemento é associado a um sensor ou atuador.

(320) Da perspectiva do sistema cognitivo ( $K$ ), o FPOMDP ( $C$ ) a ser construído representa tudo o que é lhe exterior. Essa exterioridade é constituída, além do ambiente externo ( $\xi$ ) ao agente, também pelo corpo ( $\beta$ ), que está fora da mente, e pelo sistema regulativo da mente ( $\mathcal{A}$ ), que está fora do sistema cognitivo. Portanto, o universo exterior,  $C$ , esse que deve ser aprendido pelo sistema cognitivo a partir das experiências dadas pela sucessão dos sinais  $p$  e  $c$  ao longo do tempo, é uma espécie de reflexo da composição de  $\mathcal{A} \times \beta \times \xi$ .

(321) A outra relação importante, de observabilidade parcial, advém, na arquitetura CAES, da forma com que a percepção é gerada. A função de evolução do ambiente ( $f_\xi: X_\xi \times M \rightarrow X_\xi$ ) e a função de evolução do corpo ( $f_\beta: X_\beta \times C \times S \rightarrow X_\beta$ ) são desconhecidas do agente. O sinal perceptivo ( $p$ ) é definido em função do estado do corpo ( $x_\beta$ ), que por sua vez é influenciado por um sinal de situação ( $s$ ), derivado do estado do ambiente ( $x_\xi$ ).

(322) É possível que uma determinada propriedade  $X_{\xi_i}$  que não faz parte da composição do sinal  $S$  seja relevante para a determinação de outras propriedades do ambiente. Se ela não faz parte do sinal  $S$ , não é portanto acessível à mente através de  $P$ , e conseqüentemente não pode ser usada diretamente pelo mecanismo de aprendizagem para descrever as transformações perceptivas.

(323) Por essa razão, é inteiramente possível que a função de evolução ( $f_\xi$ ) de um sistema seja completamente determinística em relação aos estados subjacentes do ambiente ( $X_\xi$ ) e mesmo assim aparacer para o agente como um sistema não-determinístico com relação às propriedades da percepção ( $P$ ). Coisa similar pode ocorrer a uma propriedade do corpo, importante como parâmetro da evolução do estado do corpo, mas inacessível para a mente por não fazer parte da composição do sinal perceptivo. Essas propriedades relevantes e inacessíveis tornam-se, do ponto de vista da mente do agente, propriedades não-observáveis ( $H$ ).

(324) Assim, a percepção revela à mente apenas uma parte do estado do corpo, e ainda indiretamente, apenas uma parte do estado do mundo, como ilustra a figura 3.6. É a partir dessa informação incompleta que a mente do agente deverá tentar organizar um conjunto de conhecimentos capaz de prever a sucessão dos eventos.

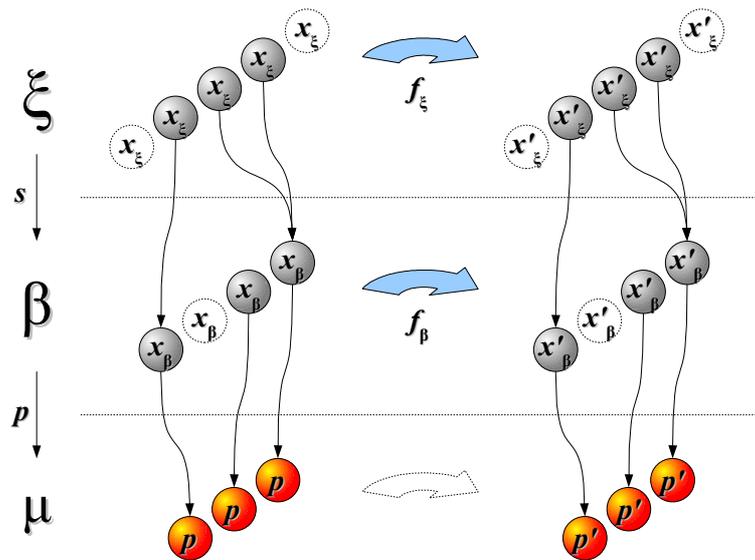


Figura 3.6: Transição do estado do sistema, de um instante para outro. Percepção indireta e parcial.

(325) Assim, o problema da construção de um modelo de mundo se coloca, para a mente do agente, como uma construção progressiva baseada na experiência. O problema é aprender um conjunto  $\tau$  de transformações do tipo  $\tau_i : rel(\tau_i) \rightarrow X_i$  que antecipem o estado de cada propriedade  $X_i'$  em função de outras propriedades e sinais de controle relevantes, onde  $rel(\tau_i) \subset (X \times C)$ .

(326) Entretanto, o agente não conhece à priori os parâmetros dessas funções. O conjunto  $X$  é inicialmente equivalente a  $P$ , e o conjunto  $H$  é iniciado vazio. Uma das tarefas é portanto criar novos elementos no conjunto  $H$ , e conseqüentemente em  $X$  e em  $\tau$ , quando as experiências indicarem a presença de uma variável relevante que não é observável, e então, além de utilizá-la como entrada das funções de transformação para as quais essa nova variável é relevante, torna-se preciso também aprender sua própria função de transformação. Do mesmo modo, o conjunto de variáveis relevantes,  $rel(\tau_i) \subset (X \times C)$ , não é dado à priori, e é tarefa do mecanismo de aprendizagem descobri-las.

### 3.1.6.2. Grau de Acessibilidade Perceptiva do Ambiente

(327) Um FPOMDP pode ser classificado quanto à acessibilidade perceptiva que o agente tem sobre o ambiente. O grau de acessibilidade perceptiva  $\omega_i$  para a dinâmica de uma dada propriedade  $X_i$  é dado pela proporção entre a quantidade de variáveis observáveis relevantes e a quantidade total de variáveis relevantes para a descrição completa e precisa de sua função de transformação  $\tau_i$ , conforme descrito pela equação 3.3.

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

(328) O grau de acessibilidade perceptiva  $\omega$  ao ambiente como um todo, por sua vez, é dado pela média do grau de acessibilidade particular de suas propriedades, conforme descrito pela equação 3.4.

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

(329) Quando  $\omega = 1$ , então o ambiente é completamente observável. Neste caso, os sensores do agente percebem todas as propriedades do ambiente relevantes para a determinação da dinâmica, ou seja  $P = X$  e  $H = \emptyset$ , e o FPOMDP equivale a um FMDP.

(330) Quando  $0 < \omega < 1$ , então o ambiente é parcialmente observável. Nesse caso, estados diferentes do sistema podem ser observados pelo agente como semelhantes (*perceptual aliasing*), ou seja, situações não equivalentes em  $X$  podem gerar sinais idênticos em  $P$ . Quando propriedades relevantes não estão sendo observadas, ou seja, quando  $rel(\tau) \not\subset P$ , o agente tem muito mais dificuldade em antecipar a sequência dos eventos.

(331) Conforme  $\omega$  diminui, aproximando-se de 0, maior fica a proporção de propriedades ocultas em relação às propriedades observáveis, dentro do conjunto de propriedades relevantes. Se  $\omega = 0$ , então nenhuma informação necessária para descrever as transformações está diretamente disponível através da percepção.

(332) A figura 3.7 apresenta um exemplo onde a percepção é composta de duas variáveis,  $p = \{p_1, p_2\}$ , o sinal de controle é igualmente composto de duas variáveis,  $c = \{c_1, c_2\}$ , mas o sistema conta com uma propriedade não-observável,  $h = \{h_1\}$ . A função de transformação  $\tau_1$  da percepção  $p_1$ , em particular, é condicionalmente dependente de  $p_1$ ,  $h_1$  e  $c_1$ , portanto, tendo acesso a 2 de 3 variáveis relevantes, o que significa uma taxa de acessibilidade  $\omega_{p_1} = 0,667$ , pela equação 3.3. Para as outras propriedades encontraremos  $\omega_{p_2} = 1,0$  e  $\omega_{h_1} = 0,667$ , resultando numa taxa de acessibilidade global  $\omega = 0,778$ , conforme a equação 3.4.

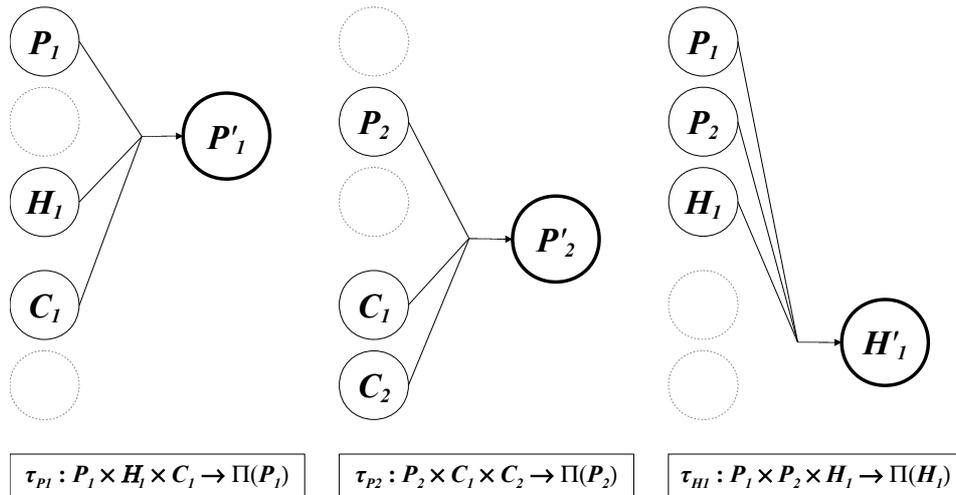


Figura 3.7: Exemplo de DBNs explicitando as propriedades ocultas.

### 3.1.6.3. Comparando o FPOMDP e o POMDP

(333) Note-se que o grau de acessibilidade perceptiva num FPOMDP é computado com relação ao número de propriedades não-observáveis, e não ao número de estados não-observáveis. Por exemplo, num FPOMDP completamente observável descrito através de propriedades binárias, uma vez que os estados são enumerados à partir das propriedades, a inserção de uma única variável não-observável dobra o número de estados no sistema, dos quais metade será não-observável no POMDP correspondente.

(334) Mais precisamente, no caso de um universo definido por variáveis binárias, se o agente tem acesso a  $|P|$  propriedades observáveis, mas existam ainda  $|H|$  propriedades relevantes não-observáveis no sistema, então existirão  $2^{|P|}$  estados observáveis para um total de  $2^{|P|} \times 2^{|H|}$  estados possíveis. Dessa maneira haverá  $2^{|H|}$  estados subjacentes possíveis para cada estado observado pelo agente, o que significa que o número de estados não-observáveis cresce exponencialmente com o número de propriedades não-observáveis.

### 3.1.7. Determinismo

(335) Uma fórmula clássica para enfrentar a complexidade de um problema em inteligência artificial é reduzir o número de variáveis consideradas, mas em compensação tratá-lo como um sistema não-determinístico. Por essa razão, MDPs, POMDPs e FMDPs descrevem a dinâmica de um ambiente através de distribuições de probabilidade. Entretanto, há muitos problemas que poderiam ser modelados de forma determinística, utilizando então, respectivamente, D-MDPs, D-POMDPs, ou D-FMDPs.

(336) Um D-MDP é um caso particular do MDP onde apenas uma transição detém toda a probabilidade, e todas as demais têm probabilidade nula. Ele é equivalente a uma máquina de estados (MOORE, 1956), ou seja, um grafo orientado onde os vértices representam transições determinísticas de um estado a outro segundo uma determinada ação. O problema da descoberta da estrutura de um D-MDP é equivalente ao problema da identificação incremental da máquina de estados por experimentação, para o qual a solução é intuitiva. Cada observação do agente pode ser diretamente aprendida como uma regularidade do sistema, uma vez que todos os estados são visíveis e as transições são determinísticas, restando apenas a preocupação de definir uma estratégia de exploração.

(337) Já o cálculo da política ótima para um D-MDP não é uma tarefa trivial, sendo objeto de diversos trabalhos e algoritmos. A política ótima em um D-MDP é o ciclo contido no grafo que maximize a média das recompensas (AHUJA et al., 1993). O primeiro algoritmo eficiente foi sugerido por (KARP, 1978), cuja complexidade computacional é de ordem  $O(|M| \times |S|)$ . Outros algoritmos foram propostos por (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), que podem ser mais rápidos segundo determinadas condições.

(338) Entretanto, a descoberta de estrutura em um D-POMDP é uma tarefa muito mais difícil do que em um D-MDP. Este problema é equivalente à indução ativa de *autômatos finitos determinísticos* (DFAs), tal como apresentado nos trabalhos de (LANG, 1992), (LANG et al., 1998), (JUILLÉ; POLLACK, 1998), (PENA; OLIVEIRA, 1998), (LANG, 1999), (CICCHELLO; KREMER, 2003).

### 3.1.7.1. Determinismo Parcial

(339) Neste trabalho, fazemos referência a ambientes parcialmente determinísticos, representados através de *Processos de Decisão de Markov Fatorados, Parcialmente Observáveis e Parcialmente Determinísticos* (PD-FPOMDP).

(340) Na definição de um PD-FPOMDP, o conjunto  $\tau = \{\tau_1, \tau_2, \dots, \tau_{|X|}\}$  das funções de transformação é substituído por um conjunto  $\sigma = \{\sigma_1, \sigma_2, \dots, \sigma_{|X|}\}$  de funções de regularidade. A *regularidade* é uma representação parcialmente determinística da

transformação. Enquanto a transformação é descrita como  $\tau_i : rel(\tau_i) \rightarrow \Pi(X_i)$ , a regularidade se apresenta na forma  $\sigma_i : rel(\tau_i) \rightarrow X_i \cup \{\#\}$ .

(341) Mais precisamente, nas situações onde uma dada transformação  $\tau_j$  é determinística, caso em que  $\exists x_i \in dom(X_i) \mid prob(\tau_j \rightarrow x_i) = 1$ , então a regularidade indica diretamente o valor  $x_i$  como resultado da função. Senão, quando a transformação é estocástica ou aleatória, caso em que  $\forall x_i \in dom(X_i) \mid prob(\tau_j \rightarrow x_i) < 1$ , então a função de regularidade retorna o simbolo especial #, indicando justamente o não-determinismo da transformação na situação dada.

(342) O PD-FPOMDP é um modelo menos geral que o FPOMDP porque não representa a distribuição de probabilidades para as transformações não-determinísticas, mas é um modelo mais geral que o D-POMDP pois ao menos permite indicar a existência delas. Em outras palavras, um PD-FPOMDP representa a parte determinística contida em um FPOMDP.

### 3.1.7.2. Grau de Determinismo de um Ambiente

(343) Um PD-FPOMDP pode ser classificado segundo o grau de determinismo ( $\partial$ ) da sua dinâmica. Para uma determinada propriedade  $X_i$ , o grau de determinismo  $\partial_i$  de sua função de transformação  $\tau_i$  é equivalente ao número de casos determinísticos que ela apresenta sobre o número total de casos, conforme a equação 3.5. Convencionamos que  $|\sigma_i|$  é o número de casos determinísticos, onde  $\sigma_i \neq \#$ , num universo de  $|\tau_i|$  casos de transformação, total que é dado pela combinação cruzada das propriedades relevantes  $rel(\tau_i)$ .

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

(344) O grau de determinismo global  $\partial$  do ambiente é calculado pela média do grau de determinismo das funções de transformações de cada propriedade particular, conforme a equação 3.6.

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

(345) Quando  $\partial = 0$ , as funções de transformação  $\tau_i$  de cada propriedade  $X_i$  produzem uma distribuição de probabilidades onde em nenhum caso a evolução de  $x_i$  para  $x_i'$  pode ser mapeada de forma direta e única em função de  $x$  e  $c$ . Neste caso, o ambiente é completamente não-determinístico, e as funções de regularidade retornam constantemente  $\sigma_i = \#$ . Inversamente, quando  $\partial = 1$ , então todas as transformações podem ser representadas sem perda como funções de regularidade, pois  $x'$  possui um mapeamento direto em função de  $x$  e  $c$ . Neste caso, o ambiente é completamente determinístico, equivalente a um D-FPOMDP. Um ambiente é parcialmente determinístico se está situado entre um extremo e outro desse eixo, ou seja, quando  $0 < \partial < 1$ , conforme ilustrado na figura 3.8.

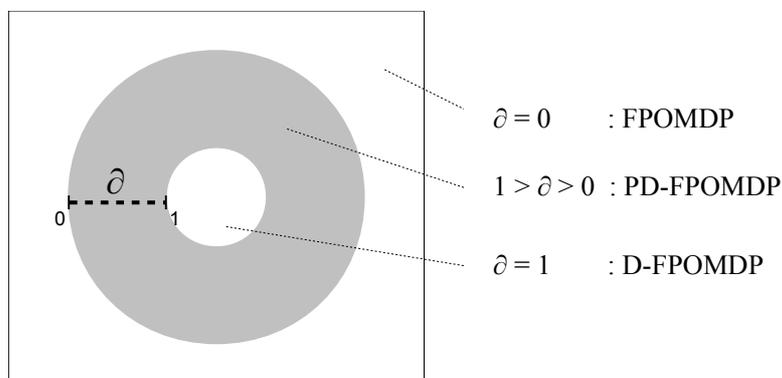


Figura 3.8: Relação de determinismo parcial.

(346) No exemplo ilustrado na figura 3.9, mostra-se o DBN que descreve a transformação da percepção  $P_1$  segundo suas propriedades relevantes  $P_1$  e  $C_1$ . No exemplo estão em destaque as transformações determinísticas. Para esta percepção, portanto, o grau de determinismo é  $\partial_1 = 0.5$ , calculado pela proporção de transformações determinísticas,  $|\sigma_1| = 2$ , em relação ao número total de transformações consideradas,  $|\tau_1| = 4$ .

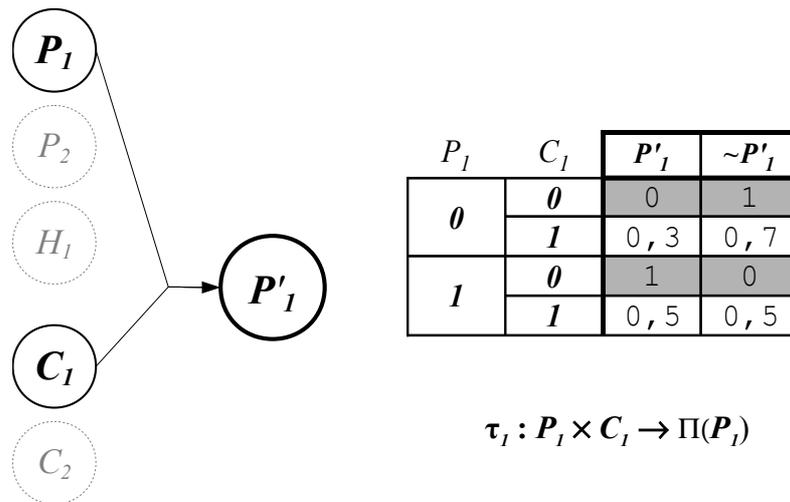


Figura 3.9: Exemplo de um DBN e de suas probabilidades de transformação.

(347)

A taxa de determinismo de um ambiente tem influência direta na maneira e na dificuldade implicada em sua aprendizagem. A aprendizagem de modelos de mundo onde as transformações são estocásticas (não-determinísticas) constitui um problema de indução de distribuições, que é um problema muito mais difícil se comparado ao caso em que as transformações são sabidamente determinísticas. A aprendizagem no caso parcialmente determinístico é mais fácil que no caso estocástico, mas mais difícil que no caso completamente determinístico. Na figura 3.10 mostra-se a mesma transformação, agora representada como uma regularidade.

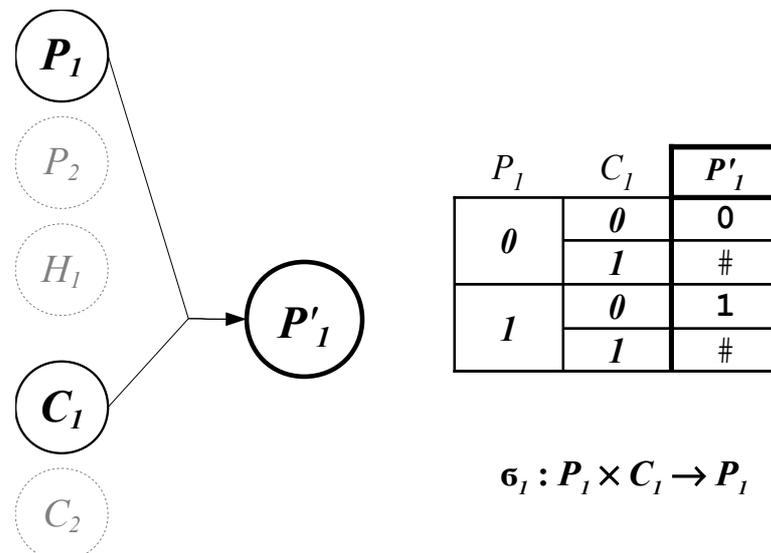


Figura 3.10: Exemplo da transformação representada como regularidade.

(348)

De um lado, a aprendizagem no caso parcialmente determinístico não precisa induzir probabilidades, e dessa forma é mais fácil cortar laços de causalidade entre variáveis. Por outro lado, a aprendizagem parcialmente determinística precisa descobrir

por si só se uma situação de transformação é ou não determinística, uma dificuldade que não está presente no caso determinístico. Finalmente, a ênfase dada à procura pelas causas das transformações pode compensar a falta da análise probabilística sobre elas. Na figura 3.11, utilizando o mesmo caso das figuras anteriores, mostramos como a descoberta de uma propriedade não-observável, porém relevante, pode tornar o modelo mais interessante.

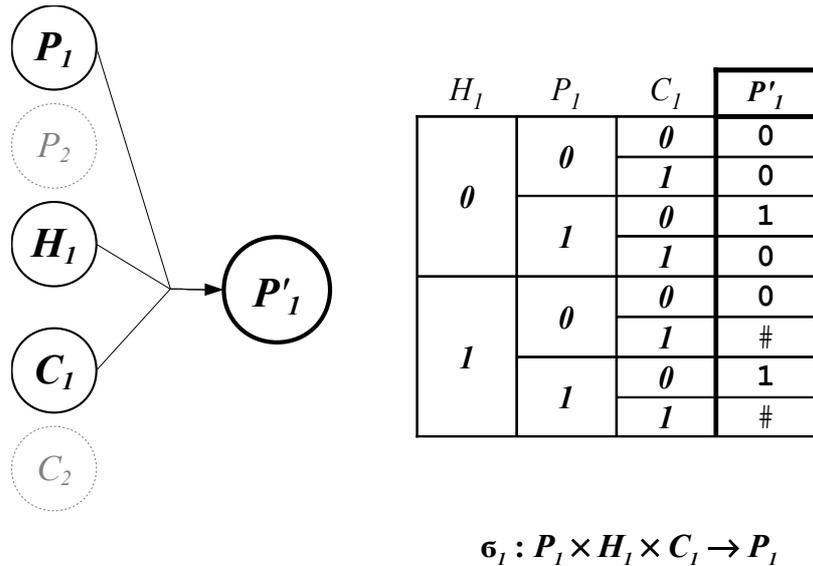


Figura 3.11: Exemplo de regularidade, considerando propriedades ocultas.

3.1.7.3. Comparando o PD-FPOMDP e o PD-POMDP

(349)

Note-se que a função de transição de estados ( $\delta$ ) em um MDP (ou POMDP) equivale a  $|X|$  funções de transformação  $\tau_i$  em seu FMDP (ou FPOMDP) correspondente. Assim, uma transição pode ser não-determinística em  $\delta$  mas apresentar componentes determinísticas em  $\tau$ . Desse modo, é possível encontrar um subconjunto de casos de transformação determinísticas em um FMDP, onde a probabilidade é igual a 1, mesmo quando o MDP correspondente é completamente não-determinístico, não possuindo nenhum caso de transição com probabilidade 1. Um exemplo é ilustrado na figura 3.12.

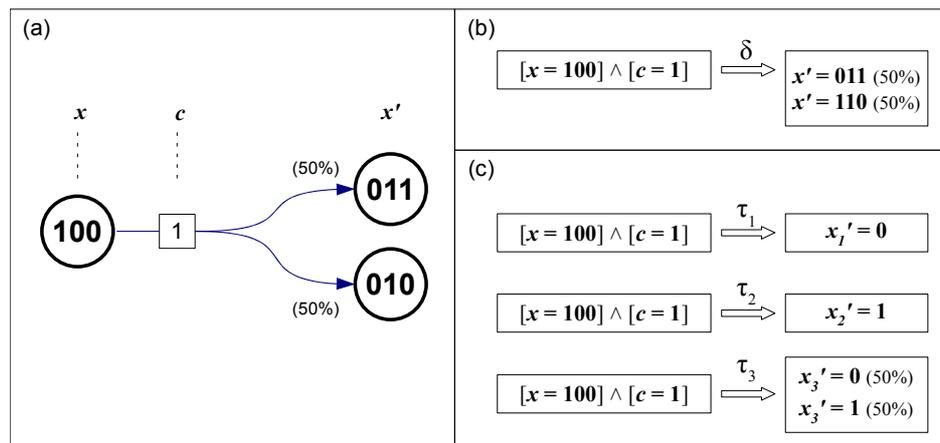


Figura 3.12: Exemplo de uma situação ambígua.

Em (a) vê-se um segmento de um MDP exemplo, onde uma transição a partir do estado [100] com a ação [1] leva de forma não-determinística aos estados [010] e [011]. Em (b) temos a representação dessa parcela da função de transição do MDP. Em (c) temos a mesma transição, agora representada de maneira fatorada, onde cada propriedade é determinada por uma função de transformação independente. O que se pode observar é que, embora a função de transição seja não-determinística nessa situação, duas das três funções de transformação derivadas dela podem ser representadas de maneira determinística.

### 3.1.8. O mundo real como ambiente a ser aprendido

(350) O mundo real é um ambiente de alta complexidade, por conta principalmente da enorme quantidade de elementos que o compõem. A ideia de tentar modelar de forma rasa problemas do mundo real é impraticável, pois uma tal representação necessitaria ser dimensionada num número incontável de estados.

(351) Entretanto, apesar de geralmente apresentar problemas de dimensão elevada, o mundo real possui algumas características que facilitam o processo de aprendizagem por agentes situados. O argumento mais forte em defesa de uma tal afirmação é o fato de que os seres humanos, imersos nesse universo altamente complexo, conseguem, de uma forma ou de outra, construir um modelo relativamente adequado para descrever o ambiente próximo, e antecipar os eventos relevantes para as tarefas e necessidades quotidianas. Em termos mais específicos, acreditamos que isso se explica pelas seguintes hipóteses: um problema do mundo real apresenta (1) muitos eventos determinísticos, portanto  $\hat{d} \gg 0$ ; (2) é largamente acessível pela percepção, portanto  $\omega \gg 0$ ; e (3) é bem estruturado, portanto  $\varphi \gg 0$ .

(352) A primeira hipótese é derivada da teoria do *determinismo causal*, que afirma que os eventos futuros são necessária e precisamente determinados pelos fatos passados e

presentes, combinados com as leis naturais que governam o universo (SUPPES, 1993). Segundo o determinismo causal, todos os fenômenos constitutivos da realidade estão submetidos a um sistema de causas e efeitos necessários. Este é um tema de grande importância e de controvérsia na filosofia. Algumas referências para o assunto são (DOOB, 1988), (BOBZIEN, 1998) e (BUNGE, 1959).

(353) O determinismo causal tem relação direta com a previsibilidade do mundo. A possibilidade de antecipar os fenômenos está condicionada ao fato deles serem causalmente determinados. Assim, em princípio, se um sujeito pudesse conhecer o estado de todos os elementos constituintes do universo, e também todas as leis gerais da natureza, então este sujeito seria capaz de deduzir com absoluta precisão a sequência dos acontecimentos do presente para o futuro.

(354) No mundo real, a incapacidade dos sujeitos de prever de forma exata os acontecimentos não é consequência de falta de determinismo do mundo, mas sim da impossibilidade de se possuir toda a informação sobre seu estado. O sujeito é limitado pelo seu ponto de vista local, e pelo excesso tanto de informação quanto de complexidade do mundo, frente à sua capacidade de conhecimento e compreensão.

(355) No caso de um agente  $\pi$  inserido em um ambiente  $\xi$ , quanto menos informação sensorial o agente possuir, maior será a quantidade de fenômenos aparentemente não determinísticos que ele vai observar no ambiente. Entretanto, se ele for capaz de considerar a existência desses elementos não-observáveis, o mundo representado em seu conhecimento parecerá cada vez mais determinístico.

(356) Em um ambiente parcialmente observável existem propriedades determinantes para a modelagem da dinâmica dos acontecimentos que não são diretamente percebidas pelo agente. Ambientes parcialmente observáveis podem exibir dinâmicas aparentemente arbitrárias e não-determinísticas na superfície, mesmo que sejam de fato determinísticos com relação ao sistema subjacente e parcialmente oculto que dá origem a face perceptiva dos fenômenos (HOLMES; ISBELL, 2006).

(357) Mesmo que a hipótese do determinismo causal não seja verificável em suas últimas consequências, acreditamos na afirmação de que, com relação ao grau de determinismo das transformações do sistema ( $\partial$ ), mesmo aceitando a existência de eventos caóticos e probabilísticos, o mundo real é um ambiente parcial porém altamente

determinístico, onde a maior parte dos eventos pode ser representada de maneira determinística, desde que eles estejam bem contextualizados em função de suas causas, sejam elas diretamente observáveis para o agente ou não.

(358) O mundo real também constitui um ambiente rico em características, e um agente situado, em geral, conforme nossa segunda hipótese, possui uma quantidade razoável de sensores. Ou seja, em problemas de mundo real, a acessibilidade perceptiva ( $\omega$ ) às propriedades do mundo pode ser relativamente alta, e as situações podem ser identificadas, em sua maior parte, com base nesses elementos observáveis. Em outras palavras, a primeira dificuldade (mesmo se ela é também importante) não é a falta de informação, mas sim o excesso.

(359) Em compensação, pela nossa terceira hipótese, o mundo real é um ambiente cujos fenômenos tem causas específicas e bem definidas, ou seja a taxa de estruturação ( $\varphi$ ) é alta, e então a quantidade de fatores relevantes para as transformações está restrita a um pequeno conjunto de variáveis, passíveis de serem aprendidas por um agente.

(360) Nessa perspectiva, dois desafios se colocam para um agente situado tentando aprender um modelo, em problemas do mundo real. Primeiro, a necessidade de encontrar representações generalizadas para as situações, uma vez que há uma quantidade muito grande de informação sensorial, o que exige a capacidade de identificar eficientemente as características relevantes implicadas nos eventos observados. Ao mesmo tempo, o agente precisa tratar ambiguidades perceptivas, ou seja, situações que não podem ser identificadas somente pela percepção sensorial imediata. Nesse caso é preciso induzir e utilizar outras informações do ambiente, que não são diretamente observáveis, para poder construir um modelo de mundo consistente.

(361) Em última análise, e em concordância com as observações da psicologia experimental de Piaget (1937), um agente (humano ou artificial), inserido no mundo real, tem a tarefa inicial de organizar suas percepções e sensações em um modelo que torne o ambiente inteligível. Se por um lado existem muitas regularidades que se encerram no próprio nível sensório-motor, de outro lado sempre restarão determinados fenômenos a exigir uma nova camada de conceitos abstratos para se tornarem compreensíveis. De uma forma geral, uma propriedade abstrata é uma variável não-

observável induzida pelo agente, e a indução deste tipo de propriedade representa, portanto, segundo nossa visão, o começo do pensamento simbólico e abstrato.

(362) O mecanismo de aprendizagem proposto nesta tese, descrito na próxima seção, tira proveito das características que julgamos serem correntes no mundo real. Primeiramente, tratando um universo decomposto em propriedades em vez de representá-lo como um conjunto raso (combinatório) de estados, e então buscando construir um modelo de mundo parcialmente determinístico, baseado de um lado na informação trazida pela percepção, e de outro lado na procura por elementos não-observáveis relevantes.

## 3.2. O Mecanismo de Aprendizagem CALM

(363) A partir desta seção, descreveremos o mecanismo CALM (*Constructivist Anticipatory Learning Mechanism*), que é desenvolvido para exercer o papel de sistema cognitivo na mente de um agente definido pela arquitetura CAES (apresentada no capítulo 2). O mecanismo CALM implementa um método de aprendizagem baseado na abordagem construtivista da IA, com o objetivo de dotar o agente de capacidade para inferir um modelo de mundo que represente as regularidades observadas durante sua interação com o ambiente, e então utilizar esse modelo para aprimorar seus padrões de comportamento.

### 3.2.1. Ideia Geral do Mecanismo

(364) O mecanismo CALM realiza as tarefas de (a) *aprendizagem de modelos de mundo* e (b) *construção de políticas de ação*, de modo incremental, e a partir da experimentação. O modelo de mundo é representado por um conjunto de *árvores de antecipação*, e a política por um conjunto de *árvores de deliberação*, como será explicado adiante. Essas árvores tem o papel de particionar de maneira otimizada o espaço de situações, sendo construídas através de diferenciações sucessivas.

(365) Conforme descrito na seção 3.1, aprender um modelo de mundo significa descobrir a estrutura de um *processo de decisão markoviano fatorado parcialmente observável e parcialmente determinístico* (PD-FPOMDP). Trata-se de um formalismo genérico para descrever as regularidades do sistema do ponto de vista do agente. Do

mesmo modo, construir uma política de ações significa definir um conjunto de regras de decisão que devem ser tomadas pelo agente para qualquer estado possível do PD-FPOMDP, definindo um padrão de comportamento que visa maximizar a média dos sinais avaliativos (recompensas recebidas) numa janela de tempo de longo-prazo.

(366) O mecanismo CALM fundamenta-se em três conceitos trazidos da psicologia construtivista (PIAGET, 1936, 1937): “esquema”, “assimilação” e “acomodação”. O *esquema* é a estrutura elementar que representa uma unidade básica de conhecimento. A *assimilação* é o processo que incorpora uma nova situação ao conhecimento que o agente já possui, ou seja, é a força conservadora que submete as novidades vivenciadas aos esquemas já existentes. De forma complementar, a *acomodação* é o processo que transforma o conhecimento em virtude de uma situação desequilibrante. Quando uma dada experiência não se mostra adequada à tentativa do agente de tratá-la como outras já conhecidas, então a acomodação modifica os esquemas para poder dar conta dessa novidade.

#### 3.2.1.1. Características do Mecanismo

(367) O mecanismo CALM foi desenvolvido para aprender as regularidades de um ambiente, tomado enquanto um sistema discreto e parcialmente determinístico. Isto significa que o agente deve identificar, dentro do conjunto total de transformações do FPOMDP, o subconjunto de transformações determinísticas ( $\sigma$ ), constituindo então um PD-FPOMDP. Assim, se por um lado, o mecanismo não requer que o sistema seja completamente determinístico, por outro lado, ele ignora regularidades não-determinísticas, isentando-se de tentar descrever uma distribuição de probabilidades. Por essa razão, a solução construída pelo CALM é mais interessante quando o ambiente apresenta muitos eventos determinísticos ( $0 \ll \partial < 1$ ).

(368) A eficiência do mecanismo CALM está diretamente ligada ao grau de estruturação do sistema. Quanto menor for o número médio de variáveis necessárias para a descrição das regularidades do PD-FPOMDP, mais estruturado será o universo ( $\varphi \gg 0$ ), portanto mais compacto será o modelo de mundo construído, e mais rápida será a convergência do algoritmo de aprendizagem. Isso implica, entretanto, na tarefa adicional de encontrar o conjunto de propriedades relevantes,  $rel(\sigma_i)$ , para a descrição das transformações, dentro do conjunto total de propriedades do sistema.

(369) Finalmente, o mecanismo CALM permite que um agente artificial aprenda um modelo de mundo, mesmo quando o ambiente é apenas parcialmente observável. Neste caso, nem todas as propriedades relevantes para a descrição da dinâmica de interação entre o agente e o mundo são diretamente acessíveis através dos sensores ( $0 < \omega < 1$ ). Isso implica na tarefa adicional de criar um conjunto de elementos sintéticos ( $H$ ) para representar propriedades não-observáveis do ambiente no seu modelo de mundo.

### 3.2.1.2. *Observabilidade Parcial x Determinismo Parcial*

(370) À primeira vista, o fato de se esquivar da tentativa de calcular as distribuições de probabilidade das transformações pode dar a impressão de que o CALM fica alheio a uma parte importante das relações causais do mundo. Geralmente, em IA, supõem-se que representar transições de forma não-determinística é a melhor opção para tratar ambientes ruidosos e falta de observabilidade completa.

(371) Entretanto, a estratégia adotada pelo mecanismo CALM é outra. A ideia é simplificar a representação das regularidades, e também o processo de descoberta delas. Descobrir quando uma transformação não é determinística é simples, pois basta que dois episódios equivalentes levem a resultados diferentes. Em compensação, o CALM é capaz de descobrir a existência de propriedades não-observáveis, que permitem uma melhor contextualização das transformações, ao ponto de elas poderem ser representadas de forma determinística.

(372) Num modelo não-determinista clássico, a observação de um evento que desconfirma uma previsão leva à uma revisão de probabilidades. Diferentemente, na estratégia do CALM, essa discordância leva à necessidade de especializar ainda mais o contexto da situação.

(373) Por exemplo, imaginemos um robô que diariamente liga um interruptor, e toda a vez observa que no mesmo momento uma lâmpada acende. Essa regularidade acabará sendo aprendida, e estará representada no modelo de mundo do robô. Entretanto num determinado dia, por conta de uma inédita falta de energia elétrica, o robô liga o interruptor e a lâmpada não acende. Quando se aceita precipitadamente o não-determinismo como explicação, esse evento não ocasiona nenhum tipo de mudança na estrutura do conhecimento. Será tratado como um ruído, ou como um evento raro, o que levará apenas a uma correção das probabilidades. No CALM, a postura é outra: o

evento desequilibrante provocará a busca por uma explicação, que nesse caso é uma nova condição que possa diferenciar a situação desequilibrada.

(374) Assim, a estratégia é procurar sempre especializar suficientemente as condições (contexto e ação) de uma situação para que a transformação descrita pelo esquema continue determinística. Essas condições podem estar relacionadas a propriedades não-observáveis do ambiente, induzidas pelo CALM como elementos abstratos. No exemplo da lâmpada, um robô, dotado do mecanismo CALM, seria levado a supor a existência de uma condição que não é diretamente observável (no caso, o fornecimento de energia elétrica), e a inclusão dessa condição na representação da transformação faz com que ela permaneça determinística. Precisamente, quando há energia elétrica, e o robô liga o interruptor, então, aí sim, a lâmpada acende. Talvez a condição não possa ser verificada de maneira determinística, mas a transformação sim, pode ser de fato determinística dentro das condições.

### 3.2.1.3. Representação do Conhecimento e Funcionamento do Mecanismo

(375) O mecanismo CALM possui 3 grandes estruturas de representação do conhecimento: (1) um conjunto  $\mathbb{M}$  de *memórias episódicas generalizadas*; (2) um conjunto  $\Psi$  de *árvores de antecipação*; e (3) um conjunto  $\mathcal{K}$  de *árvores de deliberação*.

(376) Para cada variável  $x_i'$  que o CALM decide construir um modelo antecipatório haverá uma respectiva memória episódica generalizada  $\mathbb{M}_i$  e uma respectiva árvore de antecipação  $\Psi_i$ . A memória episódica é um tipo de lembrança das situações reais vividas pelo agente. Ela é essencial ao processo de aprendizagem porque a árvore de antecipação é criada e modificada à partir de suas informações. A árvore de antecipação é propriamente a estrutura que serve para prever as transformações da variável em questão. Ela tenta particionar o espaço de situações de modo otimizado, utilizando apenas as variáveis aparentemente relevantes. Assim, cada árvore  $\Psi_i$  tenta representar de maneira compacta a função de regularidade  $\sigma_i$  que descreve a dinâmica da propriedade  $x_i$ . As árvores de deliberação são construídas à partir das árvores de antecipação, e constituem a política de ações do agente.

(377) Resumidamente, o ciclo básico de funcionamento do algoritmo tem 7 etapas, repetidas indefinidamente: (1) receber o sinal perceptivo, proveniente do exterior; (2) atualizar as variáveis internas, que representam o contexto abstrato (3) atualizar a

memória episódica generalizada (4) no caso dessa situação gerar um desequilíbrio, então realizar o processo de aprendizagem, corrigindo o modelo de mundo (5) atualizar a política de ações; (6) dada a situação atual e os modelos correntes, tomar uma decisão; e (7) enviar o sinal de controle, a fim de executar as ações decididas. O algoritmo 3.1 precisa esse ciclo de funcionamento básico do CALM.

```

Método CALM – PRINCIPAL

SEJAM:
  p um vetor representando o sinal de percepção
  c um vetor representado o sinal de controle
  h um vetor representado elementos sintéticos
  p' um vetor representado a percepção seguinte
  W a estrutura que representa a memória episódica generalizada
  Ψ a estrutura que representa o modelo de mundo
  X a estrutura que representa a política de ações

INÍCIO:

  Inicializar_Estruturas;           //Inicialização as estruturas de conhecimento
  REPETIR(indefinidamente):
    p ← Receber_Percepção;         //Recepção do sinal de percepção dos sensores
    h ← Dedução_Abstrata;         //Dedução do estado das propriedades não-observáveis
    W ← Atualizar_Memória;        //Atualização da memória episódica
    Ψ ← Atualizar_Modelo;        //Aprendizagem de modelo de mundo
    X ← Atualizar_Política;      //Aprendizagem de política de ações
    c ← Tomar_Decisão;           //Tomada de decisão para a situação atual
    Enviar_Controlo(c);           //Envio do sinal de controle para os atuadores

FIM;

```

Algoritmo 3.1: Método principal do CALM, que descreve o funcionamento básico do mecanismo.

### 3.2.2. Memória Episódica Generalizada

(378)

A aprendizagem no CALM é feita de forma incremental, pois não lhe é dado um conjunto *off-line* de situações pré-classificadas num banco de dados. O agente aprende interagindo diretamente com o ambiente. Entretanto, para que ele possa construir suas árvores de antecipação, diferenciando com eficiência as situações, é preciso haver algum tipo de memória, além do próprio modelo antecipatório e além da simples observação instantânea, que possa servir como referência para que o CALM descubra quais são os elementos possivelmente relevantes para explicar uma situação geradora de desequilíbrio. A ausência completa dessa informação transformaria o processo de aprendizagem numa busca cega pelo espaço de diferenciações. É necessário, portanto, guardar de alguma forma os episódios passados, e esse é o papel da memória episódica generalizada.

(379) A memória episódica generalizada é um conjunto de sub-memórias  $\mathbb{W} = \{\mathbb{W}_1, \mathbb{W}_2, \dots, \mathbb{W}_{|X|}\}$ , uma para cada transformação que se deseja descrever a dinâmica, e cada uma das sub-memórias é, por sua vez, um conjunto de tabelas  $\mathbb{W}_i = \{\mathbb{W}_{i0}, \mathbb{W}_{i1}, \dots, \mathbb{W}_{i|A|}\}$ . Cada uma dessas tabelas  $\mathbb{W}_{ij}$  guarda a lembrança das observações realizadas para uma propriedade específica  $i$  considerando uma dependência condicional a  $j$  propriedades. Assim, numa dada tabela  $\mathbb{W}_{ij}$  existirá uma entrada para cada uma das possíveis situações que ela puder identificar. Cada coluna representa uma combinação de variáveis agrupadas  $j$  a  $j$ , que são as combinações candidatas à propriedades relevantes. Cada coluna possui um número de linhas equivalente à combinação dos valores de suas  $j$  variáveis.

(380) As entradas (células) dessa tabela assumem seus valores segundo a observação da propriedade  $x_i$ . A tabela é completamente inicializada com o valor especial “?”, que assinala a ausência de observação. A função de *memorização* marca os casos com valores dentro de  $dom(X_i)$  de acordo com a experiência (por exemplo “0” ou “1” quando se trata de uma propriedade binária), e finalmente com o valor especial “#” para os casos em que não há acordo entre as observações.

(381) Em se tratando de problemas complexos, não é viável implementar uma memória “exaustiva e fotográfica”. Ou seja, a memória não pode guardar a lista completa de todos as situações passadas, nem pode lembrar dos eventos na plenitude dos seus detalhes. Por isso é preciso estabelecer um nível de precisão máximo, que justamente determinará a força da generalização mínima a ser aplicada aos episódios.

(382) O parâmetro  $\alpha_{\mathbb{W}}$  define o número máximo de dependências condicionais observadas simultaneamente pela memória. Quando  $\alpha_{\mathbb{W}} = 0$ , então a generalização é extrema, e a memória não pode considerar nenhuma dependência causal entre a propriedade antecipada e as propriedades que contextualizam a situação. No nível  $\alpha_{\mathbb{W}} = 1$ , a memória pode considerar a existência de uma dependência condicionando a antecipação (caso conhecido como “ingênuo”). Se  $\alpha_{\mathbb{W}} = 2$ , então uma interdependência condicional é considerada, e a memória pode levar em conta as combinações entre duas variáveis, e assim sucessivamente, conforme aumenta-se o valor de  $\alpha_{\mathbb{W}}$ .

(383) O gerenciamento da memória episódica generalizada é peça chave para garantir a convergência do algoritmo de aprendizagem, bem como para preservar sua escalabilidade em tempo e espaço.

(384) A função de *memorização* tem o papel de incluir uma nova situação experimentada na representação do conjunto de situações conhecidas. O algoritmo 3.2 descreve o método de atualização da memória episódica, e a figura 3.13 ilustra um exemplo de situações, e de como elas são armazenadas na memória.

#### Método CALM – ATUALIZA MEMÓRIA

SEJAM:

$\mathbf{w}_{i,j}$  uma tabela da memória episódica generalizada  
 $\mathbf{p}$  o último contexto perceptivo  
 $\mathbf{c}$  o último controle  
 $\mathbf{h}$  o último contexto abstrato  
 $\mathbf{p}'$  a percepção subsequente

INÍCIO:

PARA cada tabela  $\mathbf{w}_{i,j}$  da memória FAÇA:

PARA cada combinação  $\mathbf{k}$  de  $j$  propriedades de  $\{\mathbf{p}, \mathbf{c}, \mathbf{h}\}$  FAÇA:

SE  $\mathbf{w}_{i,j}(\mathbf{k}) = \text{"?"}$  ENTÃO

$\mathbf{w}_{i,j}(\mathbf{k}) \leftarrow \mathbf{p}'$

SENÃO

SE  $\mathbf{w}_{i,j}(\mathbf{k}) \neq \mathbf{p}'$  ENTÃO

$\mathbf{w}_{i,j}(\mathbf{k}) \leftarrow \text{"#"};$

FIM;

Algoritmo 3.2: Método de atualização da memória episódica generalizada.

$t_1$	<p>(a)</p>	<p>(b)</p> <table border="1"> <thead> <tr> <th></th> <th><math>H_1</math></th> <th><math>P_1</math></th> <th><math>P_2</math></th> <th><math>C_1</math></th> </tr> </thead> <tbody> <tr> <th>0</th> <td>1</td> <td></td> <td>1</td> <td>1</td> </tr> <tr> <th>1</th> <td></td> <td>1</td> <td></td> <td></td> </tr> </tbody> </table> <p><math>w_{p_1}</math> <math>j = 1</math></p>		$H_1$	$P_1$	$P_2$	$C_1$	0	1		1	1	1		1			<p>(c)</p> <table border="1"> <thead> <tr> <th></th> <th><math>H_1 \times P_1</math></th> <th><math>H_1 \times P_2</math></th> <th><math>H_1 \times C_1</math></th> <th><math>P_1 \times P_2</math></th> <th><math>P_1 \times C_1</math></th> <th><math>P_2 \times C_1</math></th> </tr> </thead> <tbody> <tr> <th>00</th> <td></td> <td>1</td> <td>1</td> <td></td> <td></td> <td>1</td> </tr> <tr> <th>01</th> <td>1</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <th>10</th> <td></td> <td></td> <td></td> <td>1</td> <td>1</td> <td></td> </tr> <tr> <th>11</th> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p><math>w_{p_1}</math> <math>j = 2</math></p>		$H_1 \times P_1$	$H_1 \times P_2$	$H_1 \times C_1$	$P_1 \times P_2$	$P_1 \times C_1$	$P_2 \times C_1$	00		1	1			1	01	1						10				1	1		11						
	$H_1$	$P_1$	$P_2$	$C_1$																																																	
0	1		1	1																																																	
1		1																																																			
	$H_1 \times P_1$	$H_1 \times P_2$	$H_1 \times C_1$	$P_1 \times P_2$	$P_1 \times C_1$	$P_2 \times C_1$																																															
00		1	1			1																																															
01	1																																																				
10				1	1																																																
11																																																					
$t_2$	<p>(a)</p>	<p>(b)</p> <table border="1"> <thead> <tr> <th></th> <th><math>H_1</math></th> <th><math>P_1</math></th> <th><math>P_2</math></th> <th><math>C_1</math></th> </tr> </thead> <tbody> <tr> <th>0</th> <td>#</td> <td>0</td> <td>#</td> <td>#</td> </tr> <tr> <th>1</th> <td></td> <td>1</td> <td></td> <td></td> </tr> </tbody> </table> <p><math>w_{p_1}</math> <math>j = 1</math></p>		$H_1$	$P_1$	$P_2$	$C_1$	0	#	0	#	#	1		1			<p>(c)</p> <table border="1"> <thead> <tr> <th></th> <th><math>H_1 \times P_1</math></th> <th><math>H_1 \times P_2</math></th> <th><math>H_1 \times C_1</math></th> <th><math>P_1 \times P_2</math></th> <th><math>P_1 \times C_1</math></th> <th><math>P_2 \times C_1</math></th> </tr> </thead> <tbody> <tr> <th>00</th> <td>0</td> <td>#</td> <td>#</td> <td>0</td> <td>0</td> <td>#</td> </tr> <tr> <th>01</th> <td>1</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <th>10</th> <td></td> <td></td> <td></td> <td>1</td> <td>1</td> <td></td> </tr> <tr> <th>11</th> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p><math>w_{p_1}</math> <math>j = 2</math></p>		$H_1 \times P_1$	$H_1 \times P_2$	$H_1 \times C_1$	$P_1 \times P_2$	$P_1 \times C_1$	$P_2 \times C_1$	00	0	#	#	0	0	#	01	1						10				1	1		11						
	$H_1$	$P_1$	$P_2$	$C_1$																																																	
0	#	0	#	#																																																	
1		1																																																			
	$H_1 \times P_1$	$H_1 \times P_2$	$H_1 \times C_1$	$P_1 \times P_2$	$P_1 \times C_1$	$P_2 \times C_1$																																															
00	0	#	#	0	0	#																																															
01	1																																																				
10				1	1																																																
11																																																					
$t_3$	<p>(a)</p>	<p>(b)</p> <table border="1"> <thead> <tr> <th></th> <th><math>H_1</math></th> <th><math>P_1</math></th> <th><math>P_2</math></th> <th><math>C_1</math></th> </tr> </thead> <tbody> <tr> <th>0</th> <td>#</td> <td>0</td> <td>#</td> <td>#</td> </tr> <tr> <th>1</th> <td></td> <td>1</td> <td></td> <td></td> </tr> </tbody> </table> <p><math>w_{p_1}</math> <math>j = 1</math></p>		$H_1$	$P_1$	$P_2$	$C_1$	0	#	0	#	#	1		1			<p>(c)</p> <table border="1"> <thead> <tr> <th></th> <th><math>H_1 \times P_1</math></th> <th><math>H_1 \times P_2</math></th> <th><math>H_1 \times C_1</math></th> <th><math>P_1 \times P_2</math></th> <th><math>P_1 \times C_1</math></th> <th><math>P_2 \times C_1</math></th> </tr> </thead> <tbody> <tr> <th>00</th> <td>0</td> <td>#</td> <td>#</td> <td>0</td> <td>0</td> <td>#</td> </tr> <tr> <th>01</th> <td>1</td> <td></td> <td></td> <td></td> <td></td> <td>1</td> </tr> <tr> <th>10</th> <td></td> <td>1</td> <td></td> <td>1</td> <td>1</td> <td></td> </tr> <tr> <th>11</th> <td>1</td> <td></td> <td>1</td> <td></td> <td>1</td> <td></td> </tr> </tbody> </table> <p><math>w_{p_1}</math> <math>j = 2</math></p>		$H_1 \times P_1$	$H_1 \times P_2$	$H_1 \times C_1$	$P_1 \times P_2$	$P_1 \times C_1$	$P_2 \times C_1$	00	0	#	#	0	0	#	01	1					1	10		1		1	1		11	1		1		1	
	$H_1$	$P_1$	$P_2$	$C_1$																																																	
0	#	0	#	#																																																	
1		1																																																			
	$H_1 \times P_1$	$H_1 \times P_2$	$H_1 \times C_1$	$P_1 \times P_2$	$P_1 \times C_1$	$P_2 \times C_1$																																															
00	0	#	#	0	0	#																																															
01	1					1																																															
10		1		1	1																																																
11	1		1		1																																																

Figura 3.13: Exemplo de formação da memória episódica generalizada.

Em (a) estão as situações experimentadas, em (b) a memória de nível  $j = 1$ , onde uma única condição é considerada possível, e em (c) a memória de nível  $j = 2$ , onde duas dependências são consideradas.

### 3.2.3. Seleção de Propriedades Relevantes

(385) Em ambientes complexos, o número de características que o agente percebe pode ser muito grande, mas em geral, se o ambiente é bem estruturado, apenas algumas poucas dessas propriedades são importantes para caracterizar uma situação. Mais especificamente, um ambiente bem estruturado é aquele em que o número médio de propriedades relevantes para a descrição de uma transformação é de ordem logarítmica em relação ao número total de propriedades, portanto  $|rel(\tau)| \approx \log_b(|P|+|H|+|C|)$ , conforme definido na seção 3.1.

(386) O problema da descoberta de relevância não é trivial, e da boa realização dessa tarefa depende a viabilidade computacional desse tipo de algoritmo de aprendizagem

(BLUM; LANGLEY, 1997), (MURPHY; McCRAW, 1991). Note-se ainda que, no caso de aprendizagem incremental, o conjunto de propriedades que parecem relevantes num dado momento, baseado no viés das experiências passadas, pode se alterar no futuro, conforme o agente experimentar novas situações. Quer dizer que uma propriedade calculada como relevante no tempo  $t$ , pode deixar de sê-lo em  $t+n$ , ou o contrário, em função de novas observações realizadas.

(387) O mecanismo CALM infere quais são as propriedades relevantes para descrever a transformação de uma determinada propriedade analisando o estado da memória episódica generalizada. Uma propriedade é dita relevante quando a correta descrição da transformação depende insubstituivelmente do valor dessa propriedade. Assim para descrever a transformação da propriedade  $X_i$ , o mecanismo utiliza a lista de diferenciadores  $\Lambda$  associada a  $\Psi_i$ , que são as propriedades supostamente relevantes extraídas da memória episódica.

(388) A seleção de propriedades relevantes funciona como uma busca heurística num espaço definido pelas possíveis combinações das propriedades consideradas. Este espaço é parcialmente ordenado, e pode-se navegar por ele através da inclusão ou remoção de uma variável da lista de diferenciadores. A figura 3.14, semelhante àquela apresentada em (BLUM; LANGLEY, 1997), ilustra o espaço de busca da relevância e sua relação com a memória episódica.

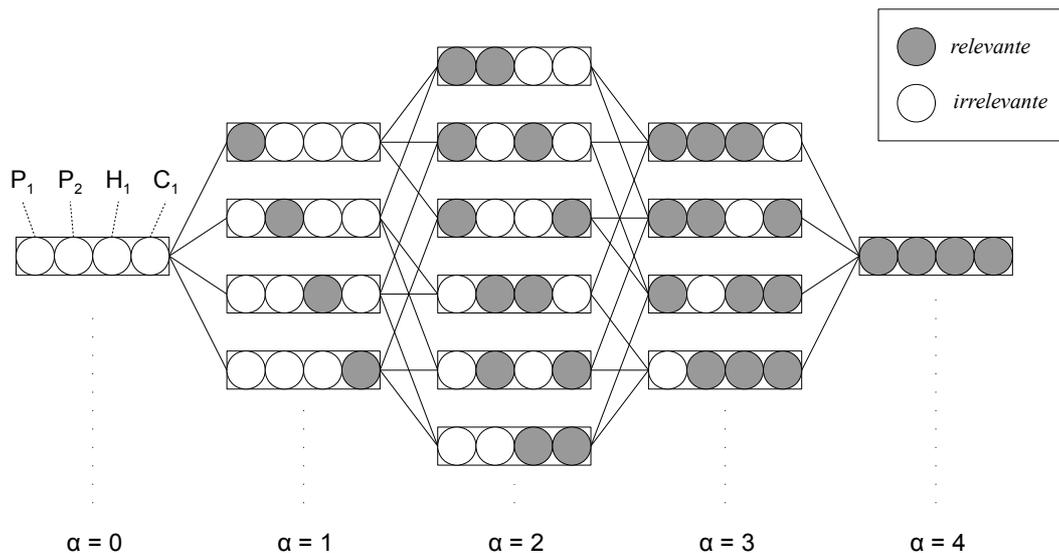


Figura 3.14: Exemplo de espaço de busca da relevância.

Um problema formado por 4 propriedades, sendo duas de contexto perceptivo ( $P_1$ ,  $P_2$ ), uma de contexto abstrato ( $H_1$ ), e uma de ação ( $C_1$ ). O espaço se forma a partir das combinações de propriedades. No desenho, os pontos escuros representam a inclusão da propriedade na lista de relevância.

(389) Para encontrar o conjunto de variáveis relevantes para a descrição da transformação de uma propriedade  $X_i$ , o mecanismo CALM calcula a taxa de determinismo para cada coluna de cada tabela da memória episódica. Essa taxa é atualizada sempre que algum valor da coluna é alterado. Finalmente, a coluna que tiver a maior taxa de determinismo, em qualquer uma das tabelas ligadas à propriedade em questão, indica a lista de variáveis relevantes ( $\Lambda_i$ ).

(390) Quando há um empate entre conjuntos candidatos, o mecanismo escolhe um deles baseado na seguinte prioridade de heurísticas: (1) escolher o conjunto que possui menos variáveis; (2) escolher o conjunto que incluir o próprio elemento que está sendo antecipado; (3) escolher o conjunto que apresente menos variáveis abstratas; (4) escolher o conjunto que apresente mais variáveis de controle; (5) escolher o conjunto que tem mais elementos em comum com as listas de relevância de outras propriedades; (6) senão, a escolha é feita de maneira aleatória.

(391) A primeira heurística explica-se pelo desejo de criar uma descrição compacta da transformação, associada ao fato de que se dois conjuntos possuem a mesma taxa de determinismo, provavelmente o maior deles possui propriedades irrelevantes. A segunda heurística baseia-se na suposição de que provavelmente a transformação de uma propriedade depende do estado dela mesma. A terceira e a quarta heurísticas

induzem o algoritmo a preferir, em caso de empate, explicações ligadas aos elementos perceptivos e às ações do agente. A quinta heurística promove uma tendência a utilizar causas semelhantes para explicar transformações diferentes, propiciando também um modelo final mais compacto. Note-se que a lista de diferenciadores nunca será maior do que o limite máximo de dependências estabelecido para a memória episódica generalizada  $\alpha$ . O método é definido no algoritmo 3.3.

**CALM – Método SELECIONA DIFERENCIADORES:**

SEJA  
 $\Lambda_k$  uma possível nova sequência de diferenciadores;

INÍCIO;  
 Para cada tabela  $j$  da memória episódica, indo de  $0$  a  $\alpha$ , faça:  
 Recalcula a taxa de determinismo das colunas modificadas  
 Seleciona a coluna  $\Lambda_k$  que teve a melhor taxa;  
 Se encontrou mais de uma sequência  $\Lambda_k$  candidata (empate), então:  
 Escolhe  $\Lambda_k$  segundo heurísticas;

FIM;

Algoritmo 3.3: Método para seleção dos testes de diferenciação (diferenciadores).

### 3.2.4. Árvore de Antecipação

(392) No que diz respeito aos problemas de construção de modelos de mundo e problemas de decisão, trabalhos recentes focados em *Processos de Decisão de Markov Fatorados* (FMDPs) utilizam uma representação do sistema baseada em *Redes Bayesianas Dinâmicas* (DBNs), (DEAN; KANAZAWA, 1989), que por sua vez podem representar as funções de transformação através de árvores, (BOUTILIER et al., 2000), (GUESTRIN et al., 2003).

(393) No mecanismo CALM, o modelo de mundo é constituído por um conjunto de *árvores de antecipação*. Uma árvore de antecipação  $\Psi_i$  particiona, de forma generalizada, exata e completa, o espaço de situações que condicionam a transformação de uma dada variável  $X_i$ . Ela é constituída por um conjunto de *nós intermediários*,  $\Theta = \{\Theta_1, \Theta_2, \dots, \Theta_{|\Theta|}\}$ , distribuídos através de seus níveis superiores, e um conjunto de *esquemas*,  $\Xi = \{\Xi_1, \Xi_2, \dots, \Xi_{|\Xi|}\}$ , que são os nós terminais. A topologia da árvore é dada pelo conjunto de *diferenciadores*,  $\Lambda = \{\Lambda_1, \Lambda_2, \dots, \Lambda_{|\Lambda|}\}$ .

(394) O CALM, enquanto mecanismo cognitivo da mente, é encarregado de construir e manter uma árvore de antecipação para cada propriedade importante do sistema. Espera-se que após um período suficiente de experimentação o agente tenha definido o

conjunto de todas essas árvores,  $\Psi = \{\Psi_1, \Psi_2, \dots, \Psi_{|\Psi|}\}$ , como uma descrição correta das regularidades do sistema,  $\sigma = \{\sigma_1, \sigma_2, \dots, \sigma_{|X|}\}$ , que por sua vez descrevem a dinâmica das propriedades  $X = \{P_1, P_2, \dots, P_{|P|}\} \cup \{H_1, H_2, \dots, H_{|H|}\}$ . Assim, cada árvore  $\Psi_i$  representa uma função de regularidade  $\sigma_i$  que descreve a dinâmica de uma determinada propriedade  $X_i$ .

(395) Cada nó da árvore é identificado pela concatenação de seus vetores, na forma  $\Theta_{[p][h][c]}$ , ou  $\Xi_{[p][h][c]}$  para os esquemas. O nó raiz  $\Theta_{[*][*][*]}$  representa a situação mais geral. A partir dele em direção às folhas, os nós passam a representar situações cada vez mais especializadas. Cada nível da árvore é mais especializado que o nível imediatamente anterior pela inclusão de um *elemento diferenciador*,  $\Lambda$ , também chamado de descritor. A árvore de antecipação guarda, assim, uma lista ordenada desses elementos,  $\Lambda \subseteq (P \cup H \cup C)$ , cuja extensão indica o número de níveis da árvore. Cada diferenciador adicionado à árvore (e conseqüentemente cada nível) possibilita a especialização de um elemento no identificador dos esquemas, permitindo que valores indefinidos '\*' sejam substituídos pelos diferentes valores possíveis de acordo com o domínio de cada propriedade diferenciadora.

(396) Os nós terminais da árvore são os esquemas. Cada um deles é identificado por uma situação suficientemente especializada, pela restrição do contexto ao qual o esquema é aplicável, e pelo detalhamento do tipo de ação a ser executada. Um esquema é portanto um nó especial da árvore que antecipa uma determinada transformação numa situação precisa de contexto e de atuação. A topologia de uma árvore de antecipação ( $\Psi_i$ ) é ilustrada na figura 3.15, e sua definição formal (3.5) é dada a seguir.

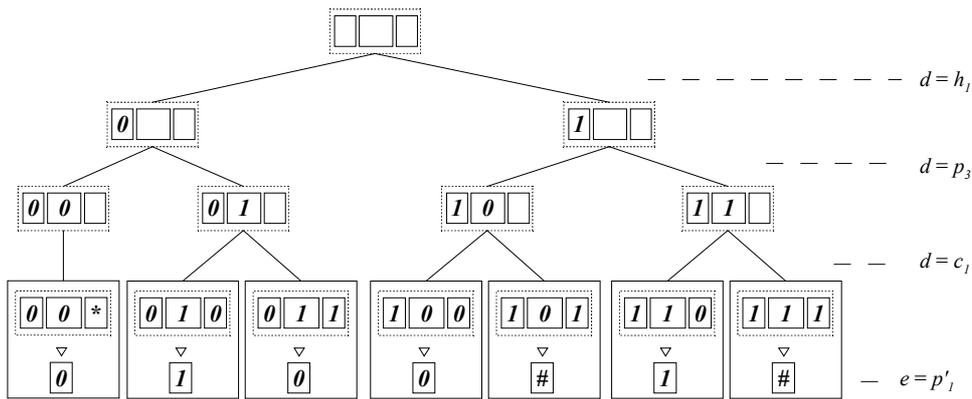


Figura 3.15: Exemplo de uma árvore de antecipação.

Uma árvore de antecipação ( $\Psi_i$ ) é topologicamente definida segundo uma lista ordenada de diferenciadores ( $\Lambda$ ). A raiz da árvore representa a situação completamente generalizada, que mapeia todo o espaço de situações. A seguir estão os nós intermediários, e os nós terminais da árvore são os esquemas.

Uma árvore de antecipação ( $\Psi_i$ ) é uma 4-upla:

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

onde

- $\Lambda = \{\Lambda_1, \Lambda_2, \dots, \Lambda_{|\Lambda|}\}$  é uma lista ordenada de diferenciadores
- $\Theta = \{\Theta_1, \Theta_2, \dots, \Theta_{|\Theta|}\}$  é um conjunto de nós intermediários
- $\Xi = \{\Xi_1, \Xi_2, \dots, \Xi_{|\Xi|}\}$  é um conjunto de esquemas (folhas da árvore)
- $e$  é a propriedade a ser antecipada (expectativa)  $x'_i$ , onde  $X_i \in (P \cup H)$

Definição 3.5: Árvore de Antecipação.

(397)

No CALM, as árvores de antecipação utilizam uma representação fatorada dos estados e das ações através de um conjunto de variáveis definidas pelos sinais vetoriais de percepção ( $P$ ), de controle ( $C$ ), e pelos elementos sintéticos ( $H$ ). Entretanto, cada uma dessas árvores é construída em função da previsão de apenas um elemento da percepção ( $P_i$ ) ou um elemento sintético ( $H_i$ ). Portanto, o mecanismo CALM mantém, no máximo,  $|P| + |H|$  árvores de antecipação, mas esse número, na prática, é otimizado de forma que o CALM mantém apenas as árvores que antecipam propriedades importantes para o agente.

### 3.2.5. Esquema

(398)

O esquema é a estrutura antecipatória elementar, que representa uma regularidade observada pelo agente durante sua interação com o mundo. Cada esquema  $\Xi$  faz uma afirmação do tipo  $(p \wedge h \wedge c) \rightarrow e$ , onde  $p$  é o contexto perceptivo mínimo do

esquema,  $h$  é o contexto abstrato mínimo,  $c$  é a ação de controle mínima, e  $e$  é a expectativa de transformação regular para a propriedade  $x_i'$ . Os domínios  $P$  e  $C$  são definidos pela arquitetura do agente, indicando a forma com que os sensores e atuadores estão conectados à mente.  $H$  é um conjunto de propriedades abstratas definidas pelo próprio sistema cognitivo. Cada esquema faz parte de uma árvore de antecipação de acordo com a propriedade que ele deve prever, assim, os esquemas que antecipam o valor da propriedade  $x_i'$  pertencem à árvore  $\Psi_i$ . Um esquema é ilustrado na figura 3.16, e formalizado a seguir, na definição 3.6.

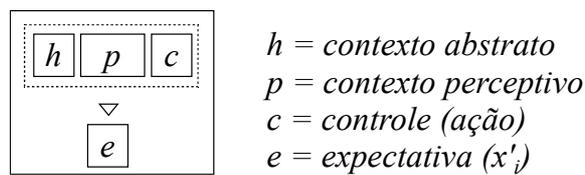


Figura 3.16: Representação de um esquema.

Um esquema ( $\Xi$ ) é uma 6-upla:

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

onde

$p = \{p_1, p_2, \dots, p_{|\Delta p|}\}$  é o contexto perceptivo do esquema

tal que  $p_i$  é um elemento ligado ao sinal de percepção do agente ( $P$ )

$h = \{h_1, h_2, \dots, h_{|\Delta h|}\}$  é o contexto abstrato do esquema

tal que  $h_i$  é um elemento sintético ( $H$ )

$c = \{c_1, c_2, \dots, c_{|\Delta c|}\}$  é a ação proposta pelo esquema

tal que  $c_i$  é um elemento ligado ao sinal de controle ( $C$ )

$e$  é a antecipação de uma propriedade  $x_i$  contida em  $(P \cup H)$

$v$  é o valor afetivo associado ao esquema

$\rho$  é a confiabilidade associada ao esquema

Definição 3.6: Esquema.

(399)

Como o CALM visa descobrir apenas as transformações determinísticas em um ambiente parcialmente determinístico, então a expectativa representa um valor absoluto, e não uma distribuição de probabilidades. Ou seja, cada esquema deve corresponder a uma de três possibilidades: (1) ou ele apresenta uma expectativa inexistente ("?"), relativa a uma situação desconhecida, jamais experimentada; (2) ou ele define uma regularidade, tendo a expectativa definida como um determinado valor de  $dom(X_i)$ , que antecipa a transformação da propriedade  $X_i$ ; (3) ou a expectativa será indeterminada ("#"), no caso de o agente não descobrir determinismo na transformação.

(400) A expectativa  $e$  de um dado esquema é justamente o valor de alguma propriedade específica  $x_i'$  que o esquema busca antecipar, caracterizando uma regularidade, onde  $X_i \in (P \cup H)$ . Ela é formada por um único elemento  $e$ , associado a uma propriedade  $p_i'$  ou  $h_i'$ . O conjunto de valores que a expectativa pode assumir é formado pelo domínio de valores da respectiva propriedade, adicionado de três novos elementos: *expectativa indeterminada*, '#', quando trata-se de uma transformação não-determinística; *expectativa inexistente*, '?', caso que ocorre quando o esquema é novo e nunca foi colocado em execução; e expectativa de *preservação*, '≈', quando espera-se que o valor permaneça inalterado após a aplicação do esquema. Portanto,  $e \in (dom(X_i) \cup \{\#, ?, \approx\})$ .

(401) Note-se que um esquema não é uma regra, do tipo “*contexto*  $\rightarrow$  *ação*”, pois o conhecimento representado não tem caráter imperativo, mas sim antecipatório, do tipo “*contexto*  $\wedge$  *ação*  $\rightarrow$  *expectativa*”. O esquema não define um comportamento a ser necessariamente adotado pelo agente no caso de encontrar-se numa determinada situação, mas sim, indica quais transformações o agente deve esperar observar se executar a ação proposta pelo esquema. A expectativa de um esquema é uma previsão para o instante seguinte à sua ativação. Além disso, para possibilitar o processo de aprendizagem, o esquema deve guardar outras informações, como confiabilidade ( $\rho$ ) e valor afetivo ( $v$ ).

### 3.2.5.1. Generalização

(402) O conjunto de esquemas de uma árvore constitui um particionamento ainda mais generalizado do que aquele realizado pela memória episódica. Um esquema  $\Xi$  é um classificador que descreve de forma generalizada um determinado conjunto de situações que são equivalentes para fins de antecipação.

(403) Os vetores  $p$ ,  $h$ ,  $c$  das árvores de antecipação em geral são sub-conjuntos dos vetores  $p$ ,  $h$ ,  $c$  da mente do agente. Os sinais recebidos pela mente de forma contínua são a seguir filtrados de acordo com a lista  $\Lambda$  de cada árvore  $\Psi_i$ , que guarda as propriedades consideradas relevantes. Quer dizer que  $P_\Lambda \subseteq P_\mu$ ,  $C_\Lambda \subseteq C_\mu$ , e  $H_\Lambda \subseteq H_\mu$ . Além disso, dentro de cada árvore de antecipação  $\Psi_i$ , os vetores que identificam um esquema podem assumir um *valor indefinido* '\*', que é uma marca de generalização. Dessa forma,  $p_\Xi \in (dom(P_\Lambda) \cup \{*\})$ ,  $c_\Xi \in (dom(C_\Lambda) \cup \{*\})$ , e  $h_\Xi \in (dom(H_\Lambda) \cup \{*\})$ .

(404) Por exemplo, suponhamos uma percepção binária  $P_i$  de um agente hipotético, que é definida pelo domínio  $dom(P_i) = \{0, 1\}$ . Num esquema, o elemento do vetor de contexto referente à esta percepção poderá, portanto, assumir um de três valores possíveis:  $p_{i\Xi} \in \{0, 1, *\}$ . Um outro exemplo seria  $dom(P_j) = \{\spadesuit, \clubsuit, \heartsuit, \diamondsuit\}$ , que resultaria em  $p_{j\Xi} \in \{\spadesuit, \clubsuit, \heartsuit, \diamondsuit, *\}$ . Essa forma de representar a generalização é semelhante àquela popularizada por (HOLLAND et al., 1986), onde o valor indefinido generaliza o esquema porque permite ignorar algumas propriedades para representar conjuntos de situações. Por exemplo, um esquema  $\Xi$  que tenha o seu vetor de contexto perceptivo definido como  $p_{\Xi} = [100*]$  assimila as situações  $[1000]$  e  $[1001]$ .

(405) O uso de valores indefinidos para alguns elementos dos vetores que descrevem as situações, e a conseqüente capacidade de criar representações generalizadas, é de fato uma maneira de agrupar um conjunto de estados do mundo sob um mesmo identificador. No cenário ideal, um esquema deve utilizar apenas as propriedades relevantes para a transformação que ele descreve. A representação generalizada das situações evita a enumeração (combinatória) dos estados, reduzindo o número de esquemas necessários para descrever a dinâmica do ambiente.

### 3.2.6. Análise de Escalabilidade

(406) Se as árvores de antecipação fossem construídas através de um algoritmo ingênuo, então o espaço requerido para armazená-las, bem como o tempo para seu processamento, sofreriam uma explosão combinatória com o aumento do tamanho do problema. Se deixássemos a árvore de antecipação crescer indiscriminadamente, rapidamente ela se tornaria ingerenciável, uma vez que a inclusão de cada diferenciador multiplica o número de nós. Além disso, uma árvore muito grande fragmenta demais o espaço de situações, criando esquemas excessivamente especializados que, por isso, são ativados raramente, aumentando assim exponencialmente a quantidade de experiências necessária para a convergência do algoritmo.

(407) Entretanto, no mecanismo CALM, as árvores de antecipação são construídas com base numa lista de diferenciadores pré-selecionados a partir da memória episódica. Isso simplifica a complexidade de construção da árvore, mas exige eficiência no tratamento da memória episódica, pois ela armazena explicitamente múltiplas combinações das observações.

(408) É o parâmetro  $\alpha_{\mathbb{W}}$  que impede o crescimento exponencial da memória episódica, pois é ele quem limita o número de dependências condicionais que ela vai considerar. Em consequência disso, o parâmetro  $\alpha_{\mathbb{W}}$  indica também o limite da profundidade da árvore de antecipação, uma vez que o número de dependências condicionais observadas determina a quantidade máxima de diferenciadores presentes em  $\Lambda$ .

(409) O número de nós na árvore, bem como o número de células na memória episódica, é de ordem exponencial em relação à  $\alpha_{\mathbb{W}}$ , portanto  $O(n^\alpha)$ , onde  $n = |P \times H \times C|$ . Note-se que  $n$  não é o número de estados, cardinalidade indicada tipicamente em MDPs, mas sim o número de variáveis (perceptivas, abstratas e de controle) do FMDP. Dessa maneira, a viabilidade computacional dessa representação, e portanto, do algoritmo, recai sobre o parâmetro  $\alpha_{\mathbb{W}}$ .

(410) Se  $\alpha_{\mathbb{W}}$  é um valor constante e pequeno, então a quantidade de memória que o CALM necessita para resolver os problemas de aprendizagem cresce numa ordem polinomial tratável. Entretanto, se  $\alpha_{\mathbb{W}}$  é excessivamente pequeno, a memória episódica fica muito generalizada, impossibilitando que o mecanismo descubra as regularidades necessárias para descrever um modelo de mundo correto. Como já visto, se o ambiente em que o agente está inserido é bem estruturado, então o número de propriedades relevantes cresce numa ordem logarítmica com relação ao tamanho do problema, e graças a isso, basta que o parâmetro  $\alpha_{\mathbb{W}}$  tenha magnitude semelhante a  $\log_b(n)$  para que o CALM possa funcionar corretamente, mantendo-se numa complexidade quase-polinomial,  $O(n^{\log(n)})$ .

(411) O tamanho total máximo  $|\mathbb{W}|$  da memória episódica generalizada é dado pela equação 3.7, que é o somatório do tamanho de todas as suas tabelas. Sendo  $n = |P \times H|$ , então podem existir até  $(n) \cdot (\alpha_{\mathbb{W}} + 1)$  tabelas  $\mathbb{W}_{ij}$  na memória, fazendo variar  $i$  de 1 a  $n$  (correspondendo a cada propriedade  $x_i$ ) e fazendo variar  $j$  de 0 a  $\alpha_{\mathbb{W}}$  (correspondendo a cada nível de dependência considerado).

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

(412) O tamanho de uma dada tabela  $|\mathbb{W}_{ij}|$ , conforme indicado da equação 3.8, é o produto cartesiano do seu número de colunas pelo seu número de linhas. O número de colunas é equivalente à quantidade de subconjuntos presentes no conjunto das  $n$

propriedades, combinados  $j$  a  $j$ , ou seja  $C_n^j$ . O número de linhas é a combinação dos valores das propriedades consideradas na coluna. Supondo que cada propriedade possa assumir  $b$  valores diferentes, então o número de linhas será equivalente a  $b$  na potência  $j$ .

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

(413) Por exemplo, imaginando-se um problema que apresente  $n = 30$  propriedades binárias, e estabelecendo o parâmetro  $\alpha_{\omega} = 5$ , então aplicando-se as equações descobrimos que o CALM dimensiona a memória episódica generalizada com, no máximo, pouco mais de 5 milhões de entradas. Trinta propriedades podem parecer pouco para problemas de mundo real, entretanto, elas constituem mais de 1 bilhão de estados numa representação rasa do ambiente, uma quantidade intratável para algoritmos de aprendizagem convencionais. A figura 3.17 mostra que o crescimento da memória episódica é de ordem menor do que exponencial,  $O(2^n)$ , e mesmo menor do que quase-polinomial,  $O(n^{\log(n)})$ .

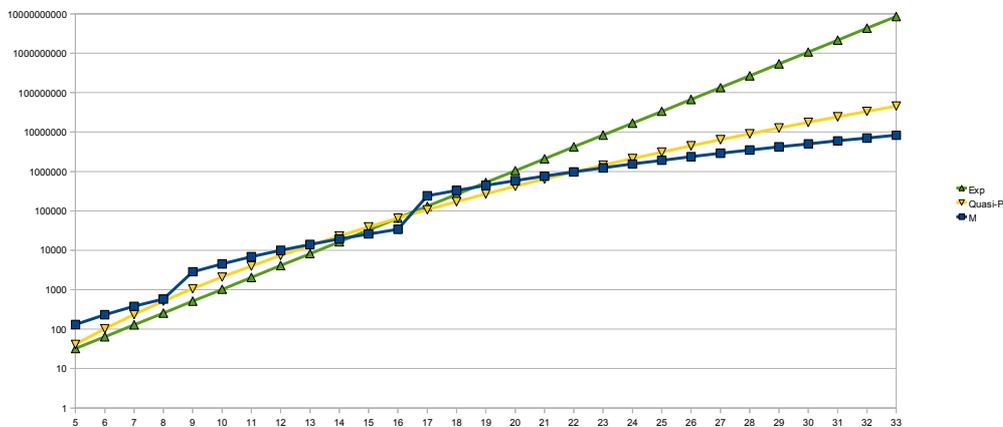


Figura 3.17: Crescimento da memória episódica generalizada (numa escala logarítmica).

O eixo vertical conta o número de estados, e o horizontal, o número de variáveis. A complexidade (em tamanho) da memória episódica generalizada é, no pior caso, conforme se aumenta o número de variáveis do problema, menor que as ordens exponencial e quase-polinomial. Faz-se aumentar o parâmetro  $\alpha_{\omega}$  de forma logarítmica (segundo a definição de um ambiente bem estruturado), conforme aumenta-se do número de propriedades. A descontinuidade na linha da memória é consequência do arredondamento do parâmetro  $\alpha_{\omega}$ .

### 3.2.7. Atualização da Árvore de Antecipação

(414) O enfrentamento de situações novas geralmente ocasiona alterações na memória episódica. Quando essas alterações revelam a necessidade de reestruturar alguma árvore

de antecipação, então se trata de uma *experiência desequilibrante*, que requer uma acomodação, no sentido definido por (PIAGET, 1936). Nesse caso, o mecanismo pode: (a) corrigir o esquema falho; (b) reestruturar a árvore de antecipação; ou (c) decidir pela criação de um novo elemento sintético.

(415) De uma forma geral, a construção de cada árvore de antecipação segue o seguinte caminho: a partir de um esquema inicial geral, numa primeira época, a árvore tende a se expandir, por *diferenciação (split)*, através da inclusão de novos diferenciadores, e a conseqüente criação de esquemas cada vez mais especializados. Se as propriedades perceptivas não são suficientes, a diferenciação pode requerer a utilização ou a *criação de elementos sintéticos* para representar propriedades não-observáveis do ambiente. Depois essa tendência se inverte. Quando não é mais possível particionar o espaço, ou quando ele já estará excessivamente particionado, as situações de desequilíbrio cognitivo passam a provocar a indeterminação, por *ajuste*, das expectativas dos esquemas desequilibrados, considerando-os como transformações não-determinísticas. Enfim, pouco a pouco, a combinação de sucessivos ajustes revela diferenciações desnecessárias, e assim gradualmente os ramos redundantes da árvore serão unificados por *integração (join)*, reduzindo a árvore.

(416) Em ambientes complexos, é interessante construir o particionamento do espaço através de uma estratégia *top-down*, ou seja, começando com um esquema simples e geral, que gradativamente dará origem a esquemas mais especializados, conforme surgirem experiências desequilibrantes, através da busca e inclusão de propriedades relevantes nos vetores que descrevem os esquemas. Uma vez iniciado, o processo que ocorre então é uma gradativa diferenciação e especialização dos vetores de contexto e ação dos esquemas. Já a evolução das expectativas segue uma estratégia *bottom-up*, ou seja, vai do específico para o geral.

(417) Nessa abordagem, o algoritmo torna-se naturalmente incremental, pois o agente possui um modelo de mundo geral e completo em qualquer instante  $t$  da execução do programa (e portanto da vida do agente), no que diz respeito às experiências já vividas, ainda que no começo, quando elas são poucas, esse modelo possa ser imperfeito. O algoritmo 3.4 explicita o código base do método de aprendizagem de modelo de mundo.

```

Método CALM – ATUALIZA MODELO

SEJAM:
   $\Psi$  o conjunto de árvores de antecipação
   $\Psi_i$  uma árvore de antecipação
   $\Xi$  o esquema ativo na árvore analisada
   $\mathbb{W}_i$  a memória episódica generalizada
   $p, h, c$  a situação vivida
   $p'$  a percepção subsequente

INÍCIO:

  PARA cada árvore  $\Psi_i$  em  $\Psi$  FAÇA:

    Atualiza_Memória( $\mathbb{W}_i, p, h, c, p'$ ); //Assimila a situação atual na memória episódica

    SE Pertinentes( $\mathbb{W}_i$ )  $\neq$   $\Lambda_i$  ENTÃO: //Se a memoria sugere outra estrutura...
       $\Psi_i \leftarrow$  Reorganizar( $\Psi_i, \mathbb{W}_i$ ); //Reorganiza a Arvore
       $\Psi_i \leftarrow$  Integração_Geral( $\Psi_i$ ); //Enxuga a Arvore

    SENÃO: //Se a memoria sugere outra estrutura...

      SE Novato( $\Xi_x$ ) ENTÃO //Se E tinha expectativa inexistente...
        Determina( $\Xi_x, x$ ); //Determina sua expectativa

      SE Desequilíbrio( $\Xi_x, x$ ) ENTÃO //Se E falhou na previsão... (desequilíbrio)
        Ajusta( $\Xi_x$ ); //Reduz a expectativa por ajuste

      SE Modificado( $\Xi_x$ ) ENTÃO //Se modificou o esquema ...
        Integração_Local( $\Xi_x$ ); //Tentar integrar, para reduzir a árvore

  FIM;

```

Algoritmo 3.4: Método de aprendizagem de modelo de mundo, que atualiza as árvores de antecipação.

### 3.2.7.1. Descrição Máxima e Mínima

(418) Nos termos usados por (BUCHANAN; WILKINS, 1993), a identidade de um dado esquema  $\Xi_i$ , composta por seus vetores  $p$ ,  $h$ , e  $c$ , é equivalente a uma *descrição discriminante mínima*, pois ela especifica o número mínimo de diferenciadores necessário para distinguir este esquema dos outros mantendo-o ainda consistente com as situações já experimentadas. Por outro lado, a memória episódica generalizada constitui uma *descrição característica máxima* para as situações já experimentadas, no limite dado pelo parâmetro  $\alpha_{\mathbb{W}}$ .

(419) De forma similar a um espaço de versões (MITCHELL, 1982), a verdadeira regularidade do ambiente, que o esquema está tentando representar, se encontra em algum lugar entre os limites colocados pelo identificador do esquema e os limites colocados pela memória episódica, conforme ilustra a figura 3.18.

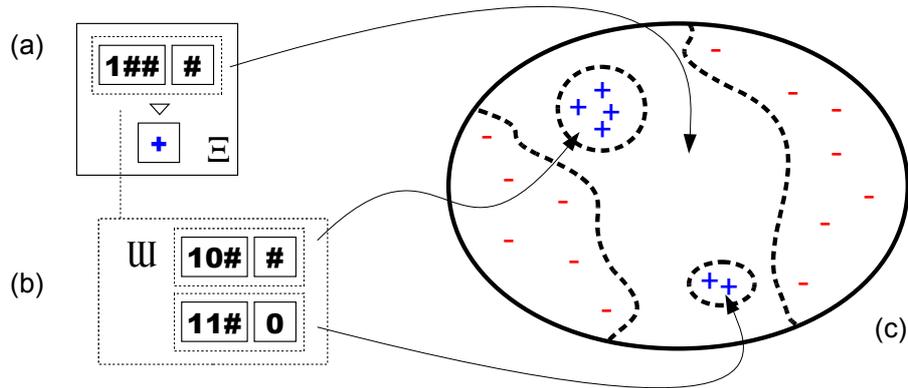


Figura 3.18: Relação geral / especializado, entre o esquema e a memória.

Um esquema (a) representa a "generalização consistente maximamente geral", e sua memória (b) representa "generalizações consistentes maximamente específicas".

### 3.2.7.2. Esquemas Iniciais e Construção Gradativa da Árvore

(420)

Inicialmente, cada árvore de antecipação  $\Psi_i$  possui apenas um esquema, que é também raiz da árvore. Este esquema raiz tem um identificador completamente geral, o que significa que ainda não há qualquer diferenciação, e que todas as situações (ou seja, qualquer contexto combinado com qualquer ação) são assimiladas por este esquema único. A expectativa desse esquema raiz é iniciada num estado especial de indefinição. Assim temos inicialmente  $\Xi_{[*][*][*]} = \text{"?"}$ , conforme o algoritmo 3.5, abaixo.

**CALM: Método INICIA ESTRUTURAS**

SEJAM:

$\Psi$  : o conjunto de árvores de antecipação  
 $\Psi_i$  : uma das árvores de antecipação do conjunto  
 $\Lambda$  : a lista de diferenciadores da árvore  
 $\Theta$  : a lista de nos intermediários da árvore  
 $\Xi$  : a lista de esquemas da árvore  
 $\Xi_0$  : um esquema inicial raiz  
 $p$  : o vetor de contexto perceptivo do esquema  
 $c$  : o vetor de controle do esquema  
 $h$  : o vetor de contexto não-observável do esquema  
 $e$  : a expectativa do esquema  
 $b$  : a cardinalidade da propriedade da expectativa  
 $\mathbb{W}$  : a memória episódica generalizada  
 $\alpha$  : o número de dependências condicionais registradas na memória  
 $v$  : o valor afetivo do esquema  
 $\rho$  : a confiabilidade do esquema

INÍCIO:

PARA cada propriedade  $x_i$  em  $(P \cup H)$  FAÇA:  Cria  $\Psi_i$ :     $\Lambda \leftarrow \emptyset$ ;     $\Theta \leftarrow \emptyset$ ;     $\mathbb{W} \leftarrow \{?, ?, \dots, ?\}$ ;  Cria  $\Xi_0$ :     $p \leftarrow '*'$ ;  $h \leftarrow '*'$ ;  $c \leftarrow '*'$ ;     $e \leftarrow '?'$ ;     $v \leftarrow \text{Sensação\_Afetiva}(e)$ ;     $\rho \leftarrow 0.0$ ;     $\Xi \leftarrow \{\Xi_0\}$ ;   $\Psi \leftarrow \Psi \cup \{\Psi_i\}$ ;

FIM;

Algoritmo 3.5: Método de inicialização das estruturas do conhecimento.

(421)

Supondo que a árvore  $\Psi_i$  antecipe uma propriedade perceptiva, então na primeira vez que o esquema inicial (único) é ativado, sua expectativa é substituída pela observação direta do resultado encontrado imediatamente a seguir, portanto  $\Xi' = \{\Xi_{[*][*]} = p_i\}$ . De fato, qualquer novo esquema, após a primeira execução, tem sua expectativa definida de acordo com o primeiro resultado observado. No caso de árvores que antecipam propriedades não-observáveis, o processo é diferente e será devidamente explicado mais adiante. O algoritmo 3.6 descreve o método de determinação inicial de expectativas.

<p><b>CALM: Método DETERMINA EXPECTATIVA</b></p> <p>SEJAM:</p> <ul style="list-style-type: none"> <li><b>e</b> um esquema ativado no ciclo anterior e em desequilíbrio</li> <li><b><math>x_i</math></b> o valor da propriedade <math>X_i</math> na situação atual</li> <li><b>e</b> a expectativa do esquema para <math>x_i</math>'</li> </ul> <p>INÍCIO:</p> <p>SE <b>e</b> = '?' ENTÃO:</p> <ul style="list-style-type: none"> <li><b>e</b> ← <math>x_i</math>';</li> </ul> <p>FIM;</p>
---

Algoritmo 3.6: Determinação inicial das expectativas.

### 3.2.7.3. Diferenciação

(422) Quando ocorre um desequilíbrio, a primeira hipótese que o mecanismo faz é a de que o esquema desequilibrado é ainda muito geral, e conseqüentemente abarca um domínio de aplicação amplo demais. Assim, a solução é criar novos esquemas mais especializados, que possam representar corretamente a regularidade da transformação através do particionamento do seu domínio.

(423) O princípio que fundamenta esse procedimento é o seguinte: se o esquema não está funcionando bem, talvez o problema seja que o conjunto de situações assimiladas por ele ainda esteja muito abrangente. O método então partilha seu domínio de assimilação, criando novos esquemas mais especializados, numa estratégia do tipo “dividir para conquistar”. O esquema instável é substituído por uma sub-árvore de dois níveis, transformado-se em nó pai dessa sub-árvore, e tendo como filhos os novos esquemas. Estes novos esquemas terão seus identificadores um nível mais especializados do que o de seu pai. O princípio da diferenciação está ilustrado na figura 3.19.

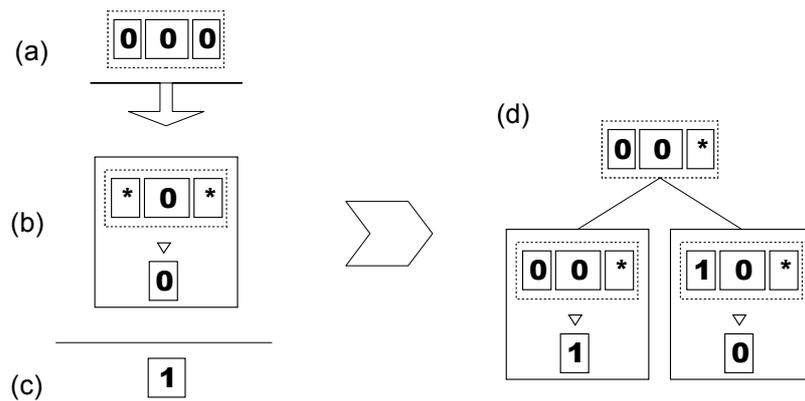


Figura 3.19: Exemplo de diferenciação.

Em (a) contexto e ação experimentados; em (b) esquema ativado; em (c) resultado real observado; e em (d) sub-árvore gerada pela diferenciação.

#### 3.2.7.4. Ajuste

(424) O ajuste transforma uma expectativa determinada em uma expectativa indeterminada. O método entra em ação quando não é mais possível diferenciar um esquema desequilibrado. Nesse caso o ajuste reduz a expectativa do esquema instável, trocando seu valor pela marca de não-determinismo “#”. No CALM, uma expectativa indeterminada é consistente com qualquer experiência, embora não antecipe nada.

(425) De forma geral, ao longo do processo de aprendizagem, as árvores de antecipação começam otimistas, tentando encontrar expectativas determinadas para todas as situações. No decorrer do tempo, pouco a pouco, por causa dos ajustes, as expectativas dos esquemas são reduzidas até o ponto em que permanecem definidas somente aquelas que representam as verdadeiras regularidades determinísticas do ambiente.

(426) As diferenciações estão limitadas pelo parâmetro  $\alpha_\Lambda$ , que indica o número máximo de diferenciadores numa árvore de antecipação, ou seja, sua profundidade máxima. Quando esse limite é atingido, então os desequilíbrios se resolvem através de ajuste. A figura 3.20 ilustra o princípio de ajuste.

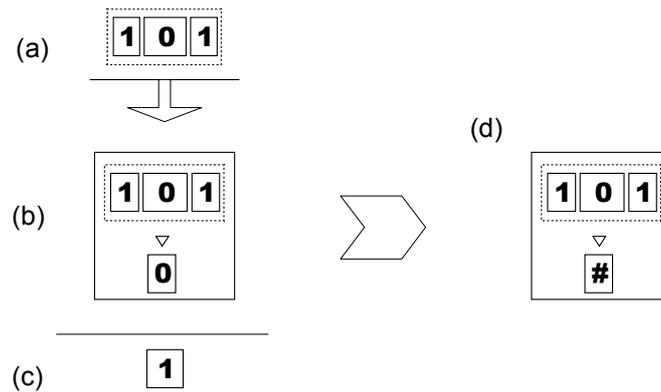


Figura 3.20: Exemplo de ajuste.

Em (a) contexto e ação experimentados; em (b) esquema ativado; em (c) resultado real observado; e em (d) esquema com a expectativa reduzida após o ajuste.

### 3.2.7.5. Integração

(427)

Finalmente, ajustes sucessivos na expectativa de diversos esquemas acabam por revelar algumas diferenciações desnecessárias, que podem então ser removidas da árvore. A integração procura por subárvores que possuem expectativas idênticas. A ideia é reunir essas subárvores em uma única. O princípio de integração, mostrado como algoritmo 3.7, é ilustrado na figura 3.21.

#### CALM – Método INTEGRA:

SEJAM:

$E_A$  um esquema cuja expectativa foi alterada  
 $S$  um conjunto de subárvores irmãs  
 $S_i$  uma subárvore de  $S$   
 $S_U$  uma possível unificação entre as subárvores de  $S$   
 $\Theta_i$  um nó da árvore  
 $d$  o elemento diferenciador no nível de  $\Theta_i$   
 $x'_s$  o conjunto de expectativas das subárvores  $S$

INÍCIO:

PARA cada nó  $\Theta_i$ , a partir do pai de  $E_A$  em direção ao nó raiz, FAÇA:

$S \leftarrow \text{sub-árvore}(\Theta_i)$ ;

SE **Compatíveis** ( $x'_s$ ) ENTÃO: //subárvores têm expectativas correspondentes

$S_U \leftarrow \text{Unifique}(S)$ ; //junta os esquemas correlatos de cada subárvore

$d \leftarrow '*'$ ; //generaliza o diferenciador

FAÇA  $S_U$  um filho único do nó  $\Theta_i$ ;

Se para qualquer nó no nível de  $\Theta_i$ ,  $d = *$ , então:

$\Lambda \leftarrow \Lambda - \{d\}$ ; //elimina o diferenciador da árvore

FIM;

Algoritmo 3.7: Método de integração de subárvores.

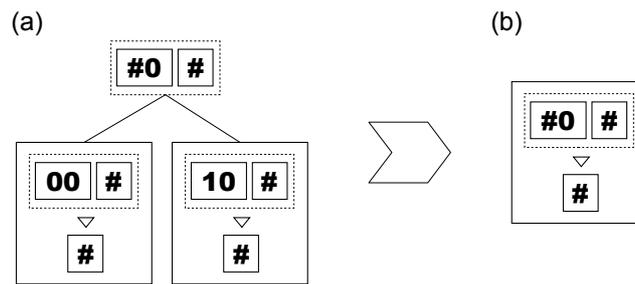


Figura 3.21: Integração entre irmãos.

Em (a) sub-árvore depois do ajuste; e em (b) um esquema integrado substitui a sub-árvore.

(428)

Note-se que o método de integração analisa o conjunto de subárvores irmãs a partir do esquema modificado. Portanto, uma integração pode acontecer entre esquemas que não são diretamente irmãos na árvore, mas primos. A figura 3.22 mostra esse caso de integração, onde a diferenciação eliminada não está imediatamente acima do esquema modificado.

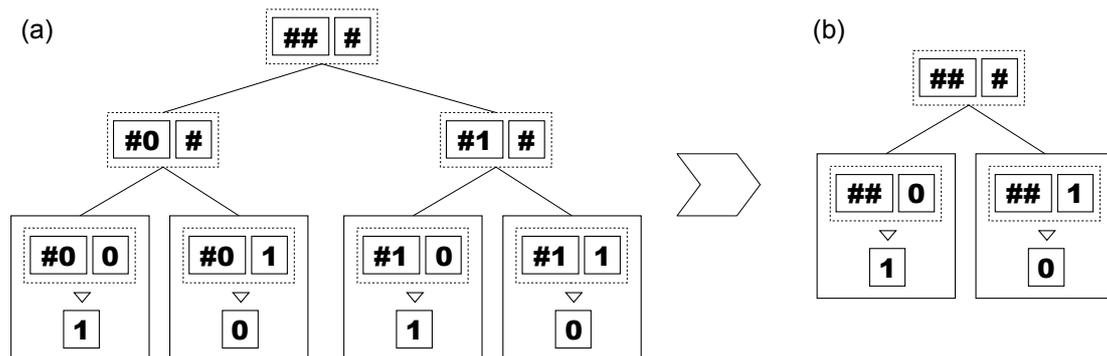


Figura 3.22: Integração entre esquemas primos.

### 3.2.8. Propriedades Não-Observáveis

(429)

No caso parcialmente observável há o problema da *descoberta de propriedades não-observáveis*. O agente precisa induzir a existência de características ocultas ou abstratas do ambiente, e incluí-las no vocabulário de representação do seu modelo de mundo. O CALM trata propriedades não-observáveis do universo através da criação de *elementos sintéticos*, no conjunto  $H$ , que são utilizados para descrever as situações juntamente com os elementos da percepção  $P$  e do controle  $C$ .

(430)

Em ambientes parcialmente observáveis, a existência de *perceptual aliasing* (situações que se confundem) impõe uma dificuldade significativa para muitos algoritmos de aprendizagem (CROOK; HAYES, 2003). De fato, quanto maior é o

número médio de propriedades relevantes não-observáveis ( $1 \gg \omega > 0$ ), mais difícil é o processo de aprendizagem.

(431) De forma semelhante à (DRESCHER, 1991) e (LANG, 1999), para tratar a observação parcial, o CALM utiliza elementos sintéticos. Eles têm a finalidade de supor a existência de algo além da percepção sensorial que seja capaz de explicar ou diferenciar situações ambíguas. Esses novos elementos representam variáveis que não são diretamente observáveis pelo agente, mas que podem ter seus valores deduzidos através da análise do encadeamento histórico dos contextos sensório-motores, e assim, podem ajudar a definição de novas antecipações.

### 3.2.8.1. *Descobrendo Propriedades Abstratas*

(432) Em problemas completamente determinísticos ( $\partial = 1$ ) mas parcialmente observáveis ( $\omega < 1$ ), as transformações podem ser representadas como funções determinísticas na forma  $\tau_i : P \times H \times C \rightarrow X_i$ . Mas o agente não percebe sensorialmente as variáveis do conjunto  $H$ , que são propriedades relevantes porém ocultas do ambiente. O espaço de situações que o agente considera inicialmente está limitado à  $P \times C$ . Por isso, mesmo que o ambiente tenha uma dinâmica completamente determinística, poderá parecer não-determinístico quando analisado somente em sua face observável.

(433) Entretanto, neste caso, quando duas situações não se diferenciam pelas características perceptivas, é possível diferenciá-las por suas “assinaturas” (RON; RUBINFELD, 1997), (FREUND et al., 1997). A assinatura identifica a situação pela sua vizinhança temporal, ou seja, pelas transições que ocorrem nos ciclos imediatamente anteriores e imediatamente posteriores. Essa identificação é ilustrada na figura 3.23.

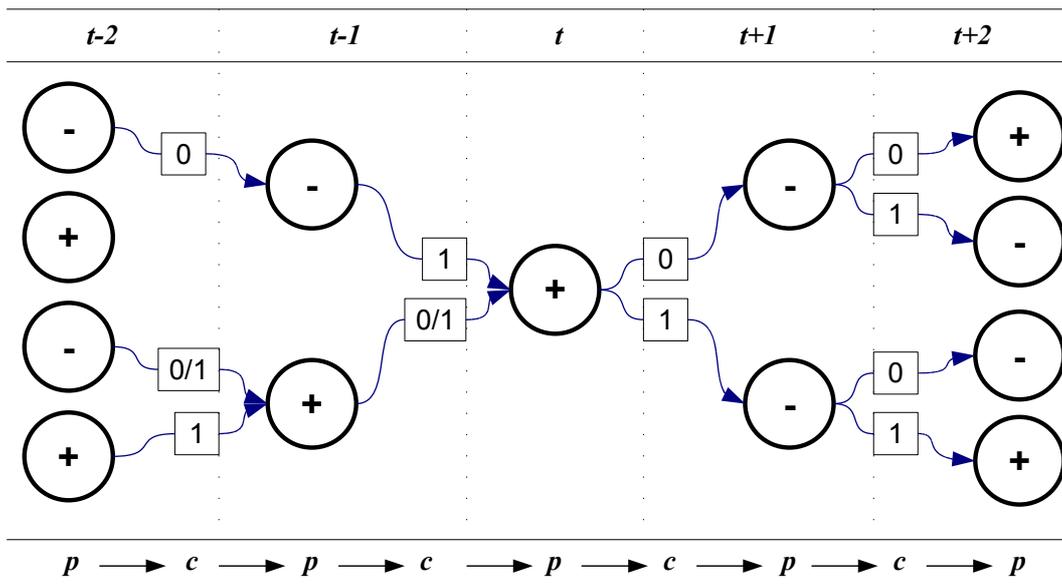


Figura 3.23: Exemplo de assinatura de uma situação a partir das observações.

O estado subjacente não pode ser corretamente determinado apenas pela observação atual, no centro da figura, entretanto ele torna-se identificável quando analisadas as sequências imediatas de observação, passadas e futuras, que levam a este estado. Se uma sequência suficiente é considerada, então a assinatura de cada estado revela-se única.

(434)

Quando ocorre uma experiência desequilibrante que não parece ter explicações no nível sensorio-motor, então o agente pode supor a existência de uma propriedade oculta do ambiente, necessária para diferenciar as situações confusas. O indício que o agente usa para descobrir a ocorrência dessa ambiguidade é a sequência de observações posterior da assinatura (as observações que sucedem a passagem por uma situação possivelmente ambígua), como ilustrado na figura 3.24.

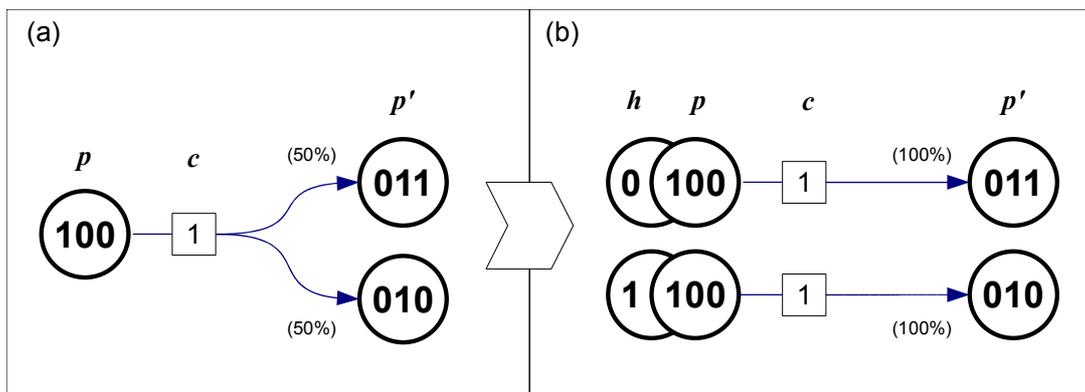


Figura 3.24: Desambiguação.

Do lado esquerdo (a), a mera percepção direta não é suficiente para saber se a observação refere-se a um mesmo estado subjacente. Do lado direito (b), a incoerência entre os resultados das transições permite deduzir, no caso determinístico, que se tratava de dois estados diferentes.

(435) Assim, se o ambiente é determinístico e bem estruturado, o agente pode, sem medo de equívocos, criar um elemento sintético sempre que encontrar uma experiência desequilibrante, e não possuir elementos suficientes no seu repertório atual para diferenciá-la. Através da inclusão de elementos sintéticos no conjunto  $H$ , para representar propriedades ocultas, o espaço inicial de representação das transformações  $P \times C$  é ampliado gradativamente, até se estabilizar, quando deverão estar presentes em  $P \times H \times C$  todas as propriedades relevantes do ambiente. Nesse momento, o agente terá condições de representar corretamente todas as transformações.

(436) Para o CALM o problema não é assim tão simples, pois o ambiente é apenas parcialmente determinístico, portanto  $\partial < 1$ , e ao mesmo tempo, parcialmente observável,  $\omega < 1$ . Neste caso, não é fácil definir se uma transformação está parecendo não-determinística porque de fato ela o é, ou porque ainda existe uma propriedade oculta não representada. A figura 3.25 ilustra as duas formas alternativas de explicação para uma situação ambígua.

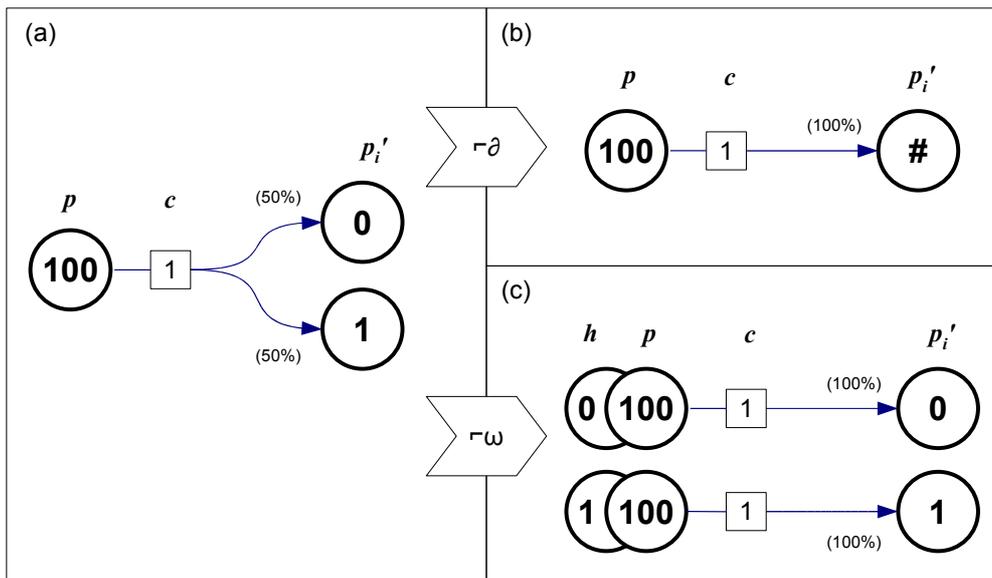


Figura 3.25: Observabilidade parcial pode explicar o aparente não-determinismo.

Do lado esquerdo (a), ilustra-se a observação referente a duas transformações observadas a partir de uma situação idêntica, mas que, porém, conduziram a resultados diferentes. Do lado direito apresentam-se as duas soluções possíveis: (b) supõe que a transformação é não-determinística, e então deixa de tentar antecipá-la; (c) supõe a existência de uma propriedade oculta capaz de explicar a transformação de forma determinística, e então introduz um elemento sintético.

(437) É preciso evitar que o agente caia na armadilha de ficar criando novos elementos sintéticos indefinidamente na esperança de encontrar determinismo numa transformação não-determinística. Neste caso ele jamais conseguirá construir um modelo estável. Por

isso, o parâmetro  $\alpha_H$  é usado para estabelecer o número máximo de elementos sintéticos que o mecanismo pode criar.

### 3.2.8.2. *Antecipando o valor de uma propriedade oculta*

(438) Somente descobrir a existência de propriedades não-observáveis, e atribuir-lhes elementos sintéticos correspondentes, não é suficiente para construir um modelo de mundo em condições de antecipar corretamente as transformações. O agente precisa também conhecer os valores dos elementos sintéticos ao longo do tempo, e como se trata de propriedades que não são diretamente observáveis, não há como saber seus valores diretamente pela percepção.

(439) Dito de outro modo, quando um elemento sintético  $h_j$  é criado para resolver dificuldades na antecipação de uma determinada propriedade  $x_i$  da árvore  $\Psi_i$ , ao mesmo tempo torna-se necessário criar uma nova árvore de antecipações  $\Psi_j$  para modelar a transformação dos valor de  $h_j$ .

(440) Suponha-se que uma mesma percepção  $p = [0]$  seja a face observável de duas situações subjacentes diferentes. A ambiguidade da situação  $p = [0]$  se evidencia porque ela conduz o agente às vezes para  $p' = [0]$  e às vezes para  $p' = [1]$ . Assim, é possível supor que essa transformação é regulada por uma propriedade não-observável, representada por um elemento sintético, diferenciando então as situações  $hp = [00]$  e  $hp = [10]$ . Suponhamos então um agente que observe a sequência  $p^{(0)} = [0]$ ,  $p^{(1)} = [0]$ , e  $p^{(2)} = [1]$ , respectivamente nos ciclos de execução  $t = 0$ ,  $t = 1$ , e  $t = 2$ . Nos instantes  $t = 0$  e  $t = 1$  o agente não desconfia da ambiguidade, desconhecendo a existência da propriedade oculta. Entretanto, ao atingir  $p^{(2)} = [1]$ , o agente poderá: (1) supor a existência de uma propriedade não-observável que é relevante para a descrição da transformação de  $p$ , e que será então representada por  $h$ ; (2) deduzir que estava anteriormente numa situação  $hp = [00]$  diferente de  $hp = [10]$ ; e (3) incluir a antecipação do elemento sintético  $h$  à partir de  $p$ , como ilustra a figura 3.26.

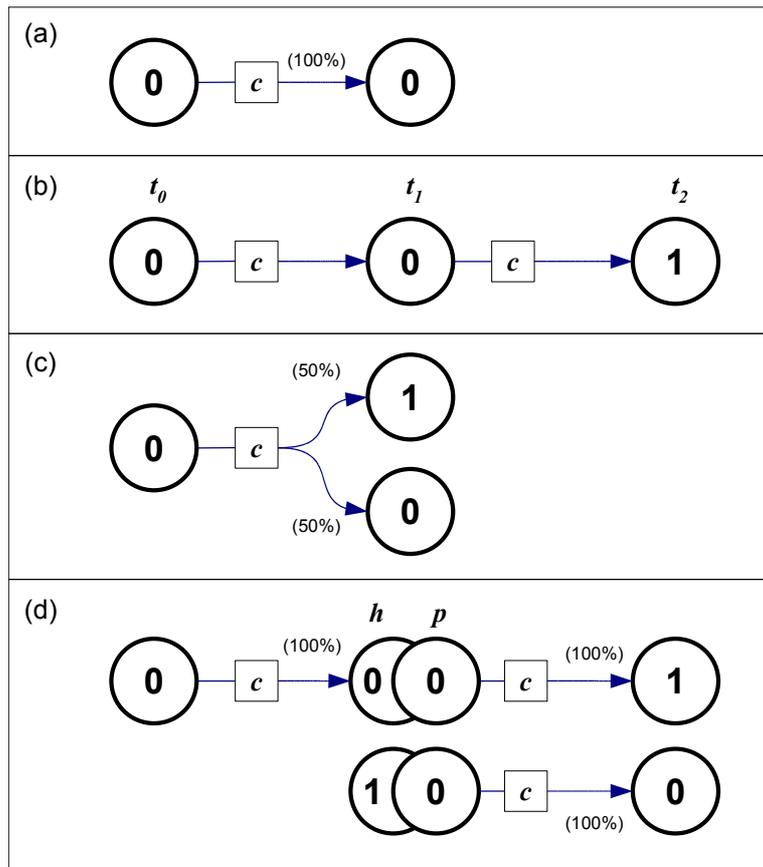


Figura 3.26: Indução de propriedades não-observáveis.

Em (a) está ilustrado o conhecimento do agente, que até o instante  $t_1$  acredita que a percepção 0 com a ação  $c$  gera 0. Em (b) mostra-se a sucessão de percepções realmente observadas pelo agente nos instantes de tempo  $t_0$ ,  $t_1$  e  $t_2$ . A observação realizada em  $t_2$  contraria o conhecimento prévio, gerando uma ambiguidade mostrada em (c). O agente então faz uso de um elemento sintético, para explicar o desequilíbrio, conforme ilustrado em (d).

### 3.2.8.3. Retropropagação do Elemento Sintético

(441) Enquanto a suposição de uma nova propriedade oculta não for necessária, o mecanismo tentará esclarecer todas as ambiguidades utilizando os elementos sintéticos já criados. Eventualmente, uma vez que é o próprio mecanismo quem induz os valores dos elementos sintéticos, uma falha na antecipação de alguma propriedade sensorial cuja previsão depende do valor de elementos sintéticos pode significar que é a antecipação do valor não-observável que está errada.

(442) Assim quando essa possibilidade existir, antes de tentar criar outros elementos sintéticos, o mecanismo deve observar se há outros esquemas precedentes ao desequilíbrio que possam ser diferenciados em função dos elementos sintéticos já existentes. Cada árvore de antecipação  $\Psi_i$  possui, para este fim, uma *memória recente*

$M_i$  que recorda uma sequência finita dos últimos esquemas ativados. O número de esquemas guardados por essa memória é limitado pelo parâmetro  $\alpha_M$ . Na ocorrência de uma falha na previsão, então a primeira alternativa que o mecanismo irá fazer será tentar executar uma “diferenciação tardia” em um esquema ativado no passado, que seria potencialmente responsável pelo desequilíbrio manifestado no instante atual.

### 3.2.9. O Processo de Decisão

(443) Para realizar o *processo de tomada de decisão*, o mecanismo CALM utiliza o conhecimento que possui (seu modelo de mundo) para interpretar as situações e planejar suas ações. O problema é equivalente ao de encontrar uma boa política de ações num PD-FPOMDP. A partir do conjunto de árvores de antecipação  $\Psi = \{\Psi_1, \Psi_2, \dots, \Psi_{|\Psi|}\}$ , o CALM constrói um conjunto de árvores de deliberação  $\mathcal{K} = \{\mathcal{K}_1, \mathcal{K}_2, \dots, \mathcal{K}_{|\mathcal{K}|}\}$ , que é o primeiro passo para o estabelecimento de uma política de ações. Uma árvore de deliberação é formalizada na definição 3.7

Uma árvore de deliberação ( $\mathcal{K}_i$ ) é uma 4-upla:

$$\mathcal{K}_i = \{\mathring{d}, \Theta, \Lambda, E\}$$

onde

$\mathring{d} = \{\mathring{d}_1, \mathring{d}_2, \dots, \mathring{d}_{|\mathring{d}|}\}$  é um conjunto de decisores (folhas da árvore)

$\Theta = \{\Theta_1, \Theta_2, \dots, \Theta_{|\Theta|}\}$  é um conjunto de nós intermediários

$\Lambda$  é uma lista ordenada de diferenciadores

$E$  é o conjunto de propriedades de expectativa, onde  $E \subset (P \cup H)$

Definição 3.7: Árvore de Deliberação.

(444) Uma árvore de deliberação é construída a partir da união de uma ou mais árvores de antecipação. Dessa forma, enquanto cada árvore de antecipação encarrega-se da previsão de uma única propriedade, as árvores de deliberação podem antecipar uma combinação delas. Uma vez constituídas, cada uma das árvores apresentará um conjunto de decisores, que são então avaliados, fornecendo os parâmetros para a construção da política de ações.

#### 3.2.9.1. Excitando e Ativando os Decisores

(445) No mecanismo CALM, o processo de decisão ocorre da seguinte maneira: a cada instante de tempo  $t$ , equivalente a um ciclo de execução, o contexto em que o agente está situado, percebido pela mente através dos sinais  $p$  e  $h$ , excita os decisores

compatíveis em cada uma das árvores de deliberação. Cada decisor excitado  $\bar{d}_j$  é um candidato à ativação na árvore  $\mathcal{X}_i$ , definindo uma ação alternativa para ser realizada através do sinal de controle  $c$ .

(446) Um decisor é compatível com uma dada situação quando os elementos definidos do seu vetor perceptivo  $p_d$  são equivalentes àqueles da percepção atual  $p_\mu$  recebida pela mente do agente, bem como os elementos sintéticos  $h_d$  são compatíveis com os valores atuais de  $h_\mu$ , estes induzidos pelo próprio mecanismo. Os decisores antecipam as consequências possíveis de sua ativação, e a partir dessa previsão é que o processo de ativação escolherá um deles para ser ativado.

(447) Os diferenciadores na árvore de deliberação estão organizados da seguinte maneira: primeiro, mais perto da raiz, diferenciadores do tipo  $P$ ; a seguir, os diferenciadores do tipo  $H$ ; e finalmente, antes de chegar aos decisores, são os diferenciadores do tipo  $C$ . O processo de excitação inicia-se a partir da raiz da árvore de deliberação, e propaga-se pelos ramos, até chegar aos decisores (nós terminais).

(448) Durante o processo de excitação da árvore de deliberação, na primeira parte do caminho, entre diferenciadores do tipo  $P$ , apenas um único ramo do nó é excitado, especificamente aquele que é compatível compatível com a percepção atual. A seguir, na segunda parte do caminho, o fluxo de excitação vai depender das inferências que o mecanismo pode fazer sobre o valor dos elementos do conjunto  $H$ . Se o valor de um elemento  $h_i$  diferenciador está inferido pelo sistema cognitivo, então a excitação segue unicamente por esse ramo, senão todos os ramos são excitados, uma vez que o valor do diferenciador não é conhecido. Finalmente, a última parte do caminho apresenta os diferenciadores do tipo  $C$ , onde a excitação é propagada incondicionalmente para todos os filhos. Ao final, apenas os decisores compatíveis com a situação atual estarão excitados, apresentando simultaneamente as alternativas de ação (para  $C$ ) e eventualmente algumas dúvidas relativas ao verdadeiro estado do ambiente (em  $H$ ).

(449) O próximo passo é escolher qual dos decisores excitados será ativado. Esta escolha é feita com base nos valores afetivos atribuídos às suas respectivas expectativas. CALM recebe um sinal de recompensa fatorada, proveniente do sistema afetivo da mente. Para cada elemento  $p_i$  do sinal de percepção  $P$  que constitua uma variável essencial do organismo, haverá um correspondente sinal avaliativo  $v_i$ . Finalmente, a

ação proposta pelo decisor ativado é realizada de fato pelo agente. Quando um decisor é ativado, sua ação  $c_d$  é enviada como sinal de controle  $c_\mu$  da mente para o corpo, mobilizando os atuadores do agente.

(450) A figura 3.27 ilustra uma decisão. No exemplo, o agente percebe a situação [01], que irá excitar dois decisores alternativos  $\hat{d}_{[*1][*][0]}$  e  $\hat{d}_{[*1][*][1]}$ . O sistema avaliativo fornece um valor afetivo para cada um desses decisores, de acordo com as suas expectativas. Finalmente, o decisor com maior valor afetivo (+0.8) é ativado, e então a ação [1], proposta por ele, é enviada através do sinal de controle.

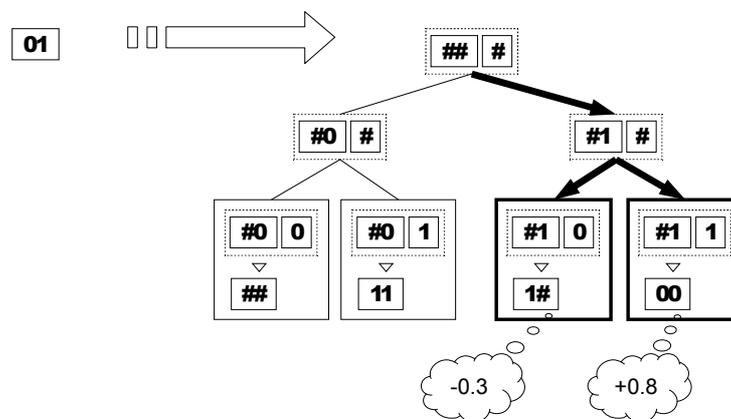


Figura 3.27: Exemplo de uma tomada de decisão.

À esquerda temos a situação percebida, e à direita, a árvore de deliberação do agente no momento da decisão.

### 3.2.9.2. Exploração e Desempenho

(451) Qualquer algoritmo de aprendizagem ativa e incremental enfrenta o dilema entre exploração e desempenho (*exploration-exploitation dilemma*). Para construir seu modelo de mundo e sua política de ações, o agente precisa explorar explicitamente o ambiente (KAELBLING et al., 1996), tentando ações alternativas, ou buscando experimentar novas situações. Mas ficar apenas explorando, sem tirar proveito do conhecimento já adquirido não faz muito sentido para um agente situado. O conhecimento adquirido deve ser utilizado para maximizar a utilidade das ações, aumentando a média dos retornos afetivos recebidos ao longo do tempo. Entretanto se o agente muito rapidamente decide deixar de explorar alternativas, corre o risco de que seu modelo de mundo ainda esteja inadequado, e que sua política de ações, derivada deste modelo, tenha alcançado apenas um máximo local, gerando comportamentos insatisfatórios (RUSSELL; NORVIG, 1995).

(452) Devido a isso, o CALM possui um *parâmetro de curiosidade*  $\varepsilon$  que define a tendência do agente para explorar o que ainda não está bem conhecido, ou, pelo contrário, para procurar as situações afetivamente mais agradáveis. O parâmetro  $\varepsilon$  indica a probabilidade de que a decisão se dê em função da avaliação afetiva  $v$  dos decisores, ou em função de sua confiabilidade  $\rho$ .

(453) Quando o agente não está explorando, a escolha do decisor a ser ativado é baseada no valor afetivo atribuído às expectativas dos decisores excitados. Aqueles que levam o agente a situações mais “prazerosas” são preferidos. Inversamente, quando está explorando, a escolha do decisor a ser ativado é baseada na sua “estabilidade”. Decisores com baixa estabilidade, que foram pouco testados, ou que foram recentemente modificados, são os preferidos neste caso, supondo justamente que são esses os que precisam ainda ser mais experimentados, a fim de que possam atingir a estabilidade. Um decisor estável é aquele que tem suas expectativas correspondidas depois de várias ativações, e por isso, a estabilidade de um decisor também indica a confiança que o agente tem em suas previsões.

### 3.2.9.3. Definindo o Conjunto de Árvores de Deliberação

(454) O primeiro passo para a construção das árvores de deliberação é realizar uma filtragem no conjunto de árvores de antecipação, selecionando apenas o subconjunto de árvores que antecipam *variáveis importantes*. Uma variável é considerada importante se ela é afetivamente positiva ou negativa, ou então se ela condiciona a antecipação de outra variável importante. Por exemplo, suponhamos um problema composto por 7 variáveis  $\{p_1, p_2, p_3, p_4, h_1, c_1, c_2\}$ , onde as variáveis perceptivas  $\{p_1, p_2, p_3, p_4\}$  estão associadas respectivamente aos valores afetivos  $\{0, +0.7, -0.3, 0\}$ . Nesse caso, o conjunto de variáveis importantes começa sendo o subconjunto  $\{p_2, p_3\}$  das variáveis afetivamente não-neutras.

(455) A seguir, é preciso analisar quais são as outras variáveis de  $P$  e  $H$  que determinam a transformação dessas propriedades importantes. Se, por exemplo,  $h_1$  for relevante para a dinâmica de  $p_2$ , então  $h_1$  será importante também, expandindo o conjunto para  $\{p_2, p_3, h_1\}$ , e assim sucessivamente até que o conjunto de propriedades importantes esteja completo. Suponhamos que, no exemplo,  $p_4$  seja um fator causal para  $h_1$ , então o conjunto completo de propriedades importantes será  $\{p_2, p_3, p_4, h_1\}$ .

(456) Ainda antes de passar à construção das árvores de deliberação propriamente, é preciso filtrar do conjunto de propriedades importantes àquelas variáveis que não estão submetidas a nenhum tipo de controle do agente. Imaginemos que, no mesmo exemplo, a transformação de  $h_1$  seja dependente da ação  $c_1$ . Uma vez que a transformação de  $p_2$  depende de  $h_1$ , então o controle realizado por  $c_1$ , que interfere sobre  $h_1$ , também vai interferir sobre  $p_2$ , ainda que indiretamente. Por isso, tanto  $h_1$  quanto  $p_2$  ficam no conjunto. Se  $p_3$  depende de  $c_2$ , então  $p_3$  também permanece. Já  $p_4$  será excluída, pois não está ligada, nem mesmo indiretamente, a uma variável de controle.

(457) Uma vez definido esse conjunto de variáveis importantes e controláveis, no exemplo equivalente à  $\{p_2, p_3, h_1\}$ , resta definir quantas árvores de deliberação serão necessárias. As árvores de deliberação são construídas a partir da união das árvores de antecipação que dependem dos mesmos elementos de controle. No caso mais simples, cada propriedade do conjunto geraria uma árvore de deliberação separada. Inversamente, se todas as variáveis importantes são dependentes das mesmas decisões de controle, então será necessário agrupar todas elas numa única árvore de deliberação. No exemplo, vimos que  $p_2$  e  $h_1$  dependem do mesmo controle  $c_1$ , enquanto  $p_3$  depende de  $c_2$ , sendo necessário então, construir duas árvores.

#### 3.2.9.4. Construindo as Árvores de Deliberação

(458) A construção de uma árvore de deliberação começa com a concatenação das árvores de antecipação. O conjunto de elementos diferenciadores (descritores) na árvore de deliberação  $\mathcal{K}_i$  é definido pela união dos diferenciadores de cada árvore de antecipação  $\{\Psi_j, \Psi_k, \dots, \Psi_m\}$  a que ela está associada, ou seja,  $\Lambda_i = \Lambda_j \cup \Lambda_k \cup \dots \cup \Lambda_m$ . A seguir, esses diferenciadores são ordenados de forma que aqueles ligados à percepção fiquem próximos à raiz, depois os diferenciadores sintéticos, e depois os diferenciadores ligados ao controle. Dito de outro modo,  $P$  antes de  $H$ , e  $H$  antes de  $C$ .

(459) A seguir, a partir da definição dos diferenciadores, a estrutura da árvore é construída, e o vetor de expectativas de cada decisor (nó terminal) é determinado pela concatenação das expectativas unitárias dos esquemas nas árvores de antecipação. A árvore de deliberação é construída inicialmente como uma árvore completa, onde todos os nós da árvore são diferenciados, e incluindo-se nessa diferenciação um ramo para o

valor indefinido. No caso de propriedades binárias, os descritores definem 3 ramos:  $\{0, 1, *\}$ .

(460) Essa primeira construção da árvore resultará em uma enumeração exaustiva de decisores. A seguir, é executado um processo de integração, que unifica decisores de mesma expectativa. Finalmente, cada decisor é conectado a ao menos um decisor sucessor, de acordo com a antecipação que faz. Os decisores não conectados podem ser eliminados.

(461) Cada árvore de antecipação descreve uma parte da dinâmica de interação entre o agente e o ambiente, que constitui apenas uma parcela do modelo de mundo. De outra parte, a mente do agente possui um sistema avaliativo que atribui valores afetivos positivos e negativos a estas antecipações. A árvore de deliberação é onde todos estes elementos são combinados. De um lado, os elementos que eram tratados como previsões separadas pelas árvores de antecipação, agora se juntam, antecipando a transformação combinada de certas propriedades. De outro lado, as situações antecipadas agora podem ser pesadas pelo sistema avaliativo, permitindo ao mecanismo construir uma política de ações adequada, e desse modo fazer com que o agente apresente um comportamento adaptado.

### 3.2.9.5. *Construindo a Política de Ações*

(462) Construída a árvore, resta avaliar os decisores. Quatro parâmetros são calculados para cada decisor, de acordo com a situação para a qual eles esperam conduzir o agente: (1) o valor afetivo imediato; (2) a utilidade afetiva, que é uma medida da capacidade do decisor para trazer retornos afetivos futuros; (3) o valor explorativo imediato, relacionado à curiosidade; e (4) a utilidade explorativa, que mede a capacidade do decisor de levar o agente à situações interessantes e desconhecidas em longo-prazo.

(463) O valor afetivo imediato  $v_A$  de um decisor é a composição (tomada aqui como uma soma) das afetividades implicadas em cada elemento de sua expectativa. Conforme definido no capítulo 2, a mente recebe um sinal de avaliação afetiva, que é um tipo de recompensa fatorada. Cada propriedade está independentemente associada a uma função de retorno avaliativo, e o valor da situação como um todo pode ser dado pela simples soma dos retornos avaliativos particulares. O valor explorativo imediato  $v_K$  de um decisor advém da confiança que o sistema tem na sua antecipação. Funciona como um

tipo de *fitness* (BOOKER, et al., 1989), aumentando cada vez que o decisor confirma suas antecipações, e retornando a zero quando algum elemento da expectativa é alterado.

(464) Entretanto, não é possível definir uma política de ações que maximize os ganhos numa perspectiva de longo prazo, apenas baseado em recompensas instantâneas. Às vezes é preciso se sujeitar a sofrer pequenas perdas imediatas para poder alcançar grandes recompensas um pouco mais adiante no futuro. Esse é o papel da função de utilidade, que geralmente representa a soma descontada dos ganhos esperados ao longo do tempo (SUTTON; BARTO, 1998). A partir de  $v_{\mathcal{A}}$  e de  $v_{\mathcal{K}}$  é possível calcular, respectivamente, a utilidade afetiva  $u_{\mathcal{A}}$  e a utilidade explorativa  $u_{\mathcal{K}}$  de um decisor. A utilidade é definida pela seguinte fórmula, baseada na equação de Bellman (1957):  $u(\vec{d}) = v(\vec{d}) + \gamma \cdot u(\vec{d}')$ .

(465) Entretanto, a questão fundamental para o cálculo das utilidades é saber quem é  $\vec{d}'$ , ou seja, quem é o decisor que será ativado após a ativação de  $\vec{d}$ . O mecanismo escolhe dois sucessores:  $\vec{d}'_{\mathcal{A}}$  entre os decisores compatíveis com a expectativa de  $\vec{d}$ , como aquele decisor em que  $c$  maximiza a utilidade afetiva, e  $\vec{d}'_{\mathcal{K}}$  entre os decisores em que  $c$  maximiza a utilidade explorativa.

## 4. RESULTADOS EXPERIMENTAIS

(466) Neste capítulo, apresentamos uma sequência de experimentos onde utilizamos um agente modelado segundo a arquitetura CAES (definida no capítulo 2) e que implementa o mecanismo CALM (definido no capítulo 3) como seu sistema cognitivo. Os resultados experimentais obtidos são então comparados àqueles dos trabalhos relacionados. São apresentados dois problemas: (1) *wepp*, e (2) *flip*. Os experimentos procuram demonstrar as consequências e os resultados da utilização da arquitetura CAES e do mecanismo CALM nestes problemas.

(467) O mecanismo CALM e a arquitetura CAES são ambas contribuições desta tese, porém cada um deles tem uma finalidade específica e diferente, portanto a análise dos resultados dos experimentos deve levar isso em conta. A arquitetura CAES é avaliada na sua capacidade de modelar os problemas de forma situada, evidenciando a fatoração do sistema pela utilização dos sensores e atuadores vetorizados, internalizando a motivação do agente através de um sistema afetivo corporal, e definindo problemas de observação parcial pela própria natureza de comunicação indireta dos sinais entre a mente e o corpo, e entre o corpo e o ambiente. O mecanismo CALM é avaliado na sua capacidade de aprender um modelo de mundo e uma política de ações para os cenários experimentais apresentados. Neste caso, os experimentos têm a finalidade de testar e demonstrar a eficiência, a estabilidade, e a qualidade das soluções obtidas em cada um dos problemas.

### 4.1. Problema Wepp

(468) Construímos um primeiro experimento, chamado *wepp*, (sigla para *walking-exhaustion-pain-pleasure*) onde o agente deve aprender a se movimentar num ambiente que lhe é desconhecido, buscando desenvolver um padrão de comportamentos que

promova o aumento das sensações de prazer e a redução das sensações de dor e exaustão.

#### 4.1.1. Definição do Problema

(469)

No problema *wepp*, o agente pode andar para frente ou girar sobre seu próprio eixo num espaço bidimensional. Toda vez que o agente anda, ele tem uma sensação de prazer, afetivamente positiva. Entretanto, o dilema é que andar demais gera exaustão, e isso é afetivamente negativo. Além disso, chocar-se contra os obstáculos gera dor, que também é afetivamente negativa. Assim, é necessário que o agente coordene suas ações de acordo com os contextos observados, afim de otimizar seu comportamento. Um retrato do problema *wepp* é mostrado na figura 4.1, e as variáveis envolvidas, bem como suas relações, são mostradas na figura 4.2.

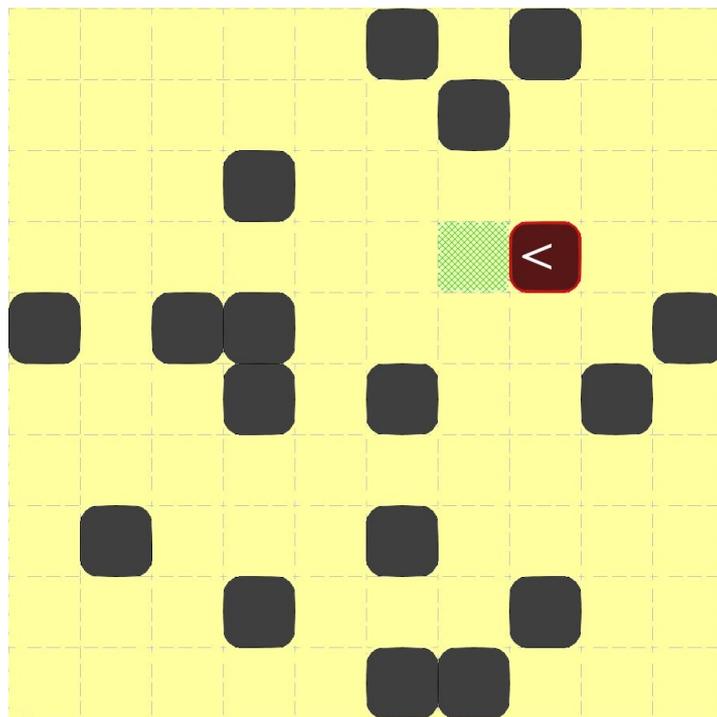


Figura 4.1: Retrato do problema *wepp*.

O objeto em destaque é o agente, voltado para a esquerda, e enxergando uma célula à sua frente.

Os demais objetos são os obstáculos.

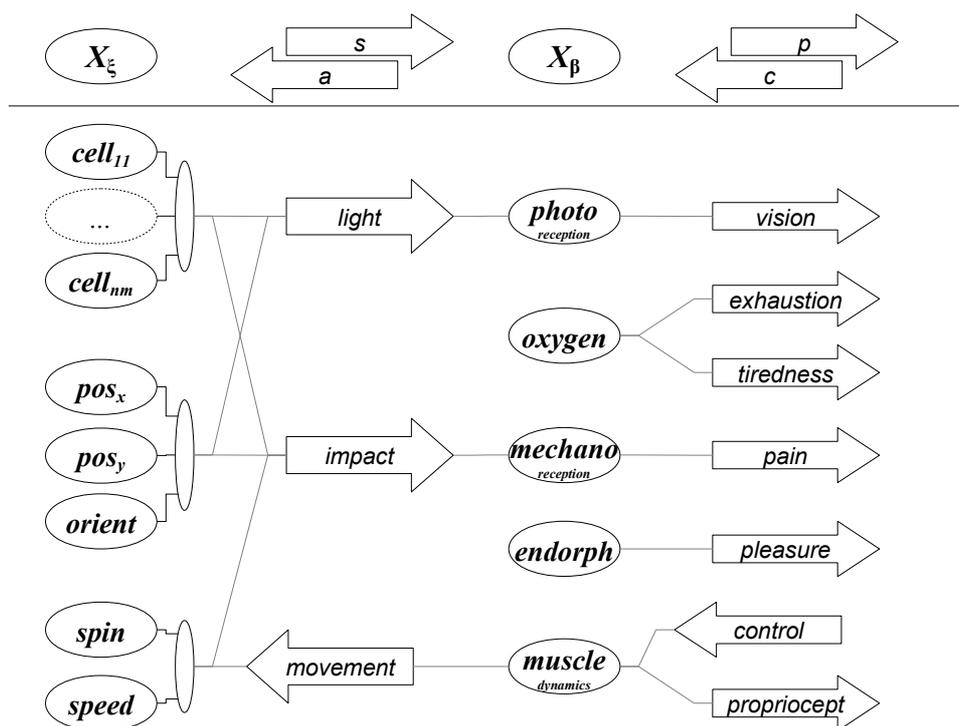


Figura 4.2: Função de evolução do ambiente no problema wepp.

#### 4.1.1.1. O Universo Wepp

(470) O ambiente deste experimento é um espaço discretizado por onde o agente se movimenta. Este espaço tem a forma de um tabuleiro (um *grid* onde cada objeto ocupa uma célula) toroidal (as bordas opostas estão conectadas). Há apenas dois tipos de objetos nesse universo: o *agente* e os *obstáculos*. A cada ciclo de execução, o agente encontra-se numa determinada posição desse tabuleiro. Ele pode caminhar livremente através dos espaços vazios, mas se tentar ir contra um obstáculo, sofrerá uma colisão, e terá seu deslocamento impedido.

(471) Neste problema, o agente só tem um atuador, que realiza suas ações de deslocamento. A cada ciclo de execução, utilizando este atuador, o agente pode realizar uma de duas ações: *andar*, dando um passo à frente, ou *girar*, rotacionando sobre seu próprio eixo num ângulo de 90 graus, para a direita ou para a esquerda. O agente também só possui um sensor externo que informa o que o agente vê na célula diante de si.

(472) Ao utilizarmos a arquitetura CAES, um tabuleiro de  $n$  por  $m$  células passa a ser representado na estrutura do ambiente  $\xi$  como uma matriz de  $n$  por  $m$  variáveis. O ambiente conta ainda com mais um vetor de 5 outras variáveis indicando a orientação

atual do agente, sua posição  $x$  e  $y$ , e sua velocidade de rotação e deslocamento. Cada célula do tabuleiro (variáveis  $cell_{ij}$ ) pode assumir um de 2 valores possíveis:  $\{\text{vazio}, \text{obstáculo}\}$ . A orientação do agente (variável  $orientation$ ), pode ter um de 4 valores possíveis:  $\{\text{norte}, \text{sul}, \text{leste}, \text{oeste}\}$ . As variáveis de posição ( $pos_x$  e  $pos_y$ ) podem assumir valores dos domínios:  $\{1, 2, \dots, n\}$  e  $\{1, 2, \dots, m\}$ . As variáveis de rotação ( $spin$ ) e deslocamento ( $speed$ ) codificam de maneira discretizada a velocidade dos movimentos do agente, respectivamente nas formas  $\{-1, 0, 1\}$  e  $\{0, 1\}$ .

(473) Na arquitetura CAES, agente e ambiente são sistemas independentes e parcialmente abertos. Existe um canal de interferência mútua entre ambos constituído pelos sinais de *situação* ( $s$ ) e *atuação* ( $m$ ). O agente influencia a dinâmica do ambiente através de sua atuação, neste caso composta por um único sinal de movimento (variável *movement*), que assume um entre dois valores:  $\{\text{andar}, \text{girar}\}$ . A função de transformação do ambiente  $f_\xi$  depende, primeiramente, do próprio estado atual do ambiente (ou seja, da localização do agente, e da localização dos obstáculos), mas depende também da atuação do agente.

(474) De outro lado, o sinal de situação representa a interferência do ambiente sobre a evolução interna do agente. Neste experimento, a situação é composta por duas variáveis. A primeira dessas variáveis (*light*) representa a célula que está diante do agente no tabuleiro, cujo valor é dado em função da posição e da orientação do agente, indicando a luminosidade do que está imediatamente a sua frente, no domínio  $\{\text{claro}, \text{escuro}\}$ . A segunda variável (*impact*) representa a ocorrência ou não de colisão durante a movimentação do agente, em consequência de um obstáculo no caminho.

(475) A definição 4.1 mostra a composição do conjunto de propriedades do ambiente, e também a composição dos sinais de situação e atuação, definindo o problema *wepp* no nível do ambiente. A figura 4.3, na sequência, descreve a função de evolução do ambiente.

Definição do Universo no Problema WEPP:

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

$$M = \{movement\}$$

$$S = \{light, impact\}$$

Definição 4.1: Variáveis do problema *wepp*, no nível do ambiente.

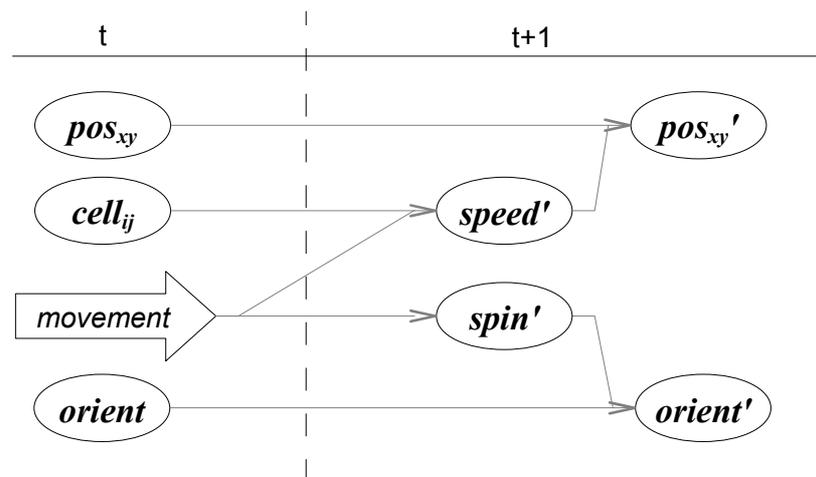


Figura 4.3: Função de evolução do ambiente no problema wepp.

#### 4.1.1.2. O Corpo do Agente Wepp

(476) O corpo do agente é composto por 5 propriedades internas. Embora trate-se de um corpo virtual, vamos associar nomes biológicos a estas variáveis, com o fim de caracterizar a função de cada uma no organismo do agente. A primeira delas (variável *endorphin*) está ligada à sensação de prazer do agente. A segunda variável (*mechanoreception*) é ligada à sensação do impacto. Uma terceira variável (*oxygen*) representa a energia atual do agente. Há mais uma variável (*muscledynamics*) que controla o atuador de movimento do agente. Uma última variável corporal (*photoreception*), ligada ao sensor externo do agente, retorna a percepção da luminosidade à sua frente, indicando a presença ou ausência de obstáculos. Quase todas essas propriedades são binárias, com exceção da oxigenação, definida numa escala discreta de 0 a 20.

(477) A função de evolução do corpo  $f_{\beta}$  está definida pelas seguintes regras: (1) A presença de *endorfina* no corpo do agente depende unicamente do seu *movimento*, ou seja, se o agente anda, então a endorfina é produzida pelo organismo, mas se o agente gira, não. (2) A *mecanorrecepção* torna-se ativa quando o agente colide com um obstáculo, ou seja, ela surge no instante em que o agente recebe um *impacto*, e desaparece no momento seguinte. (3) A cada passo em frente que o agente dá, ele gasta energia, diminuindo gradativamente a *oxigenação* do corpo. Por outro lado, a ação de girar não despende energia, permitindo que o corpo aumente sua oxigenação. (4) A *musculodinâmica* é determinada pelo sinal de controle proveniente da mente. (5) A *fotorrecepção* é definida pelo sinal de situação *light*.

(478) O corpo do agente se comunica com a mente através dos sinais de *percepção* ( $p$ ) e *controle* ( $c$ ). O sinal de percepção é composto por 6 variáveis. A primeira delas corresponde à sensação de dor (variável *pain*), e está ligada à mecanorrecepção no corpo, que por sua vez é determinada pela ocorrência ou não de impacto entre o agente e obstáculos do ambiente. A segunda variável (*pleasure*) corresponde à sensação de prazer, e informa à mente sobre a presença ou não de endorfina no corpo. A fadiga (variável *tiredness*) é um sinal disparado quando o nível de oxigenação do corpo está abaixo de 7, numa escala de 0 a 20, ou seja 35%. Outra sensação (variável *exhaustion*), também ligada à condição de oxigenação do corpo, é disparada quando o nível está abaixo de 6. Uma quinta variável (*vision*) repassa para a mente a informação recebida como fotorrecepção, que por sua vez indica a luminosidade do que há diante do agente no ambiente. A última variável da percepção (*proprioception*) dá o feedback do movimento, indicando o estado do atuador do agente. A mente possui um único comando (*control*), que determina a musculodinâmica do corpo, e que por sua vez induz modificações no ambiente.

(479) A definição 4.2 mostra a composição do conjunto de propriedades do corpo do agente, bem como a composição dos sinais de percepção e controle. A definição 4.3, na sequência, representa a função de evolução do corpo.

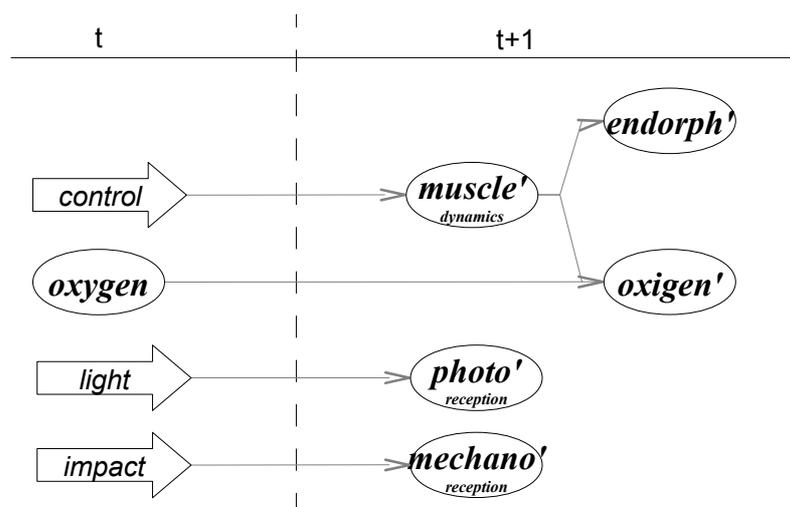
Conjuntos de propriedades corporais do agente no problema *wepp*:

$$X_{\beta} = \{endorphin, mechanoreception, oxygen, muscledynamics, photoreception\}$$

$$P = \{pain, pleasure, tiredness, exhaustion, vision, proprioception\}$$

$$C = \{control\}$$

Definição 4.2: Conjuntos de propriedades corporais do agente no problema *wepp*.



Definição 4.3: Função de evolução do corpo no problema wepp.

(480) A dinâmica dos sinais da mente, em função da evolução do corpo, pode ser resumida da seguinte maneira: quando o agente repete muitas vezes seguidas a ação de *caminhar*, então a *fadiga* é ativada, e nesse caso, basta o agente dar mais um passo para que a sensação de *exaustão* também seja ativada. Essas sensações desaparecem quando a oxigenação do corpo retorna a níveis normais. Em resumo, a ação de andar pode levar o agente ao cansaço pois cada passo à frente baixa um pouco sua oxigenação. Contrariamente girar afasta o agente do cansaço pois esta ação permite ao metabolismo do organismo aumentar a oxigenação.

(481) O problema *wepp* é interessante porque ele gera esse conflito a ser resolvido pelo agente: ele quer andar sempre que possível pois é assim que sente prazer, mas ele não pode fazê-lo de maneira inconsequente senão acaba por sentir também consequências desagradáveis como dor e exaustão.

#### 4.1.1.3. Questões de Situatividade

(482) É importante notar que no problema *wepp*, utilizando-se a arquitetura CAES, a posição espacial do agente constitui, através das variáveis  $pos_x$  e  $pos_y$ , um par de propriedades do ambiente, e não variáveis internas do agente. O agente pode interferir nessas variáveis usando o sinal de atuação, no caso composto pela variável *movimento*. De outro lado, a observação que o agente faz do ambiente ocorre através dos sensores de *mecanorrecepção* e *fotorrecepção*, que são propriedades internas ao agente, não fazendo parte do ambiente, embora seja o ambiente que as determine através do sinal de situação, neste caso composto pelas variáveis *light* e *impacto*.

(483) Nos experimentos, o problema *wepp* foi representado de duas maneiras diferentes: primeiro utilizando a configuração clássica de problemas de labirinto em aprendizagem de máquina, na qual o agente recebe como entrada sua posição e orientação no tabuleiro, e depois utilizando a arquitetura CAES, onde o agente percebe o ambiente através de uma janela visual. Os resultados experimentais demonstram que, devida à condição de situatividade promovida pelo CAES, o agente pode resolver o problema sem mapear explicitamente o ambiente. Ou seja, ele pode aprender um padrão de comportamento que maximiza o retorno afetivo, baseado apenas na sua atuação e ponto de vista locais, sem necessariamente ter um modelo de mundo que mapeie o tabuleiro com a posição de todos os obstáculos.

(484) Duas vantagens são marcantes na abordagem situada. Primeiro, a dificuldade para resolver o problema não aumenta com o aumento do tabuleiro. Segundo, a solução construída pelo agente é independente da posição dos obstáculos, e assim o agente pode transportar a solução diretamente de um cenário para outro.

(485) Não há uma “posição-objetivo” no tabuleiro, portanto não existe o conceito de estado final, como uma motivação extrínseca. No problema *wepp*, os objetivos do agente se estabelecem indiretamente através do sistema regulativo da mente, que atribui valores afetivos negativos para a dor e para a exaustão, e um valor afetivo positivo para o prazer. Essas sensações afetivas equivalem a um sinal de reforço interiorizado. Nos experimentos foram utilizados os seguintes valores: *dor* =  $-1.0$ ; *exaustão* =  $-0.7$ ; *prazer* =  $+0.4$ .

#### 4.1.2. Resultados e Considerações

(486) Foram realizadas duas implementações do problema *wepp*. Na primeira delas utilizamos uma representação clássica, onde o agente percebe diretamente o ambiente, de uma forma onisciente, sendo informado de sua posição e orientação. Na segunda utilizamos a arquitetura CAES como referência de modelagem, na qual o agente é situado.

(487) Foram também implementados dois mecanismos de aprendizagem. O primeiro, usado como referência de comparação, é uma versão típica do algoritmo *Q-Learning* (WATKINS; DAYAN, 1992), onde a aprendizagem é processada com base na

representação rasa dos estados, ou seja, num espaço formado pela combinação dos valores de todas as variáveis consideradas. A segunda é uma implementação do mecanismo CALM.

(488) 9 configurações diferentes de cenário foram utilizadas, aumentando gradativamente o tamanho do tabuleiro, bem como a quantidade relativa de obstáculos. A configuração mais simples é formada por um tabuleiro com dimensões  $5 \times 5$  (25 células), a segunda configuração estabelece um tabuleiro  $25 \times 25$ , e a configuração maior é a de um tabuleiro  $125 \times 125$  (15.625 células). Os obstáculos são distribuídos aleatoriamente, variando a quantidade de células ocupadas pelos obstáculos de 10% para 20% e depois para 30%.

(489) Para cada uma das configurações foram definidos 10 cenários randômicos, e o experimento foi repetido 10 vezes para cada cenário, totalizando portanto 900 simulações. Essas nove configurações são exemplificadas na figura 4.4, onde as células escuras são os obstáculos, as células claras são os espaços vazios, e o objeto circular é o agente. Nos eventuais casos onde o posicionamento randômico dos objetos deixou o agente preso num subespaço muito pequeno, o cenário foi descartado e uma nova randomização foi realizada.

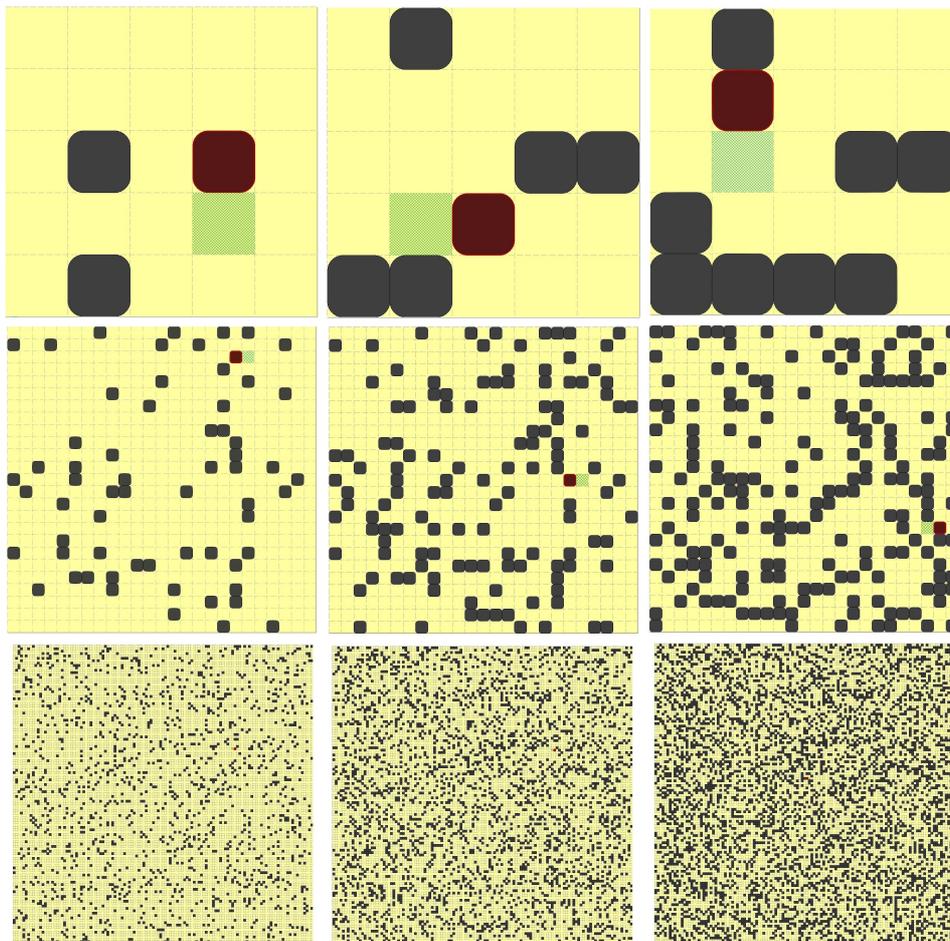


Figura 4.4: Configurações de simulação utilizadas nos experimentos com o problema *wepp*.

As configurações apresentadas na linha de cima mostram tabuleiros de dimensões 5 x 5 contendo, da esquerda para a direita, respectivamente, 10%, 20% ou 30% das células ocupadas por obstáculos. No centro, mostram-se as configurações 25 x 25 e em baixo, as configurações 125 x 125.

#### 4.1.2.1. *Avaliando o CAES: efeitos da situatividade*

(490)

Os gráficos das figuras 4.5, 4.6 e 4.7 mostram a série de experimentos realizados nos diversos cenários, neste primeiro momento, utilizando apenas o algoritmo *Q-Learning*. Os resultados confirmam que o aumento no tamanho do tabuleiro não influencia no tempo de convergência nem na qualidade da solução quando o agente é implementado pelo CAES. Diferentemente, a implementação clássica, atrelada ao mapa do tabuleiro, sofre um aumento exponencial no tempo necessário para a convergência, de acordo com o tamanho do problema.

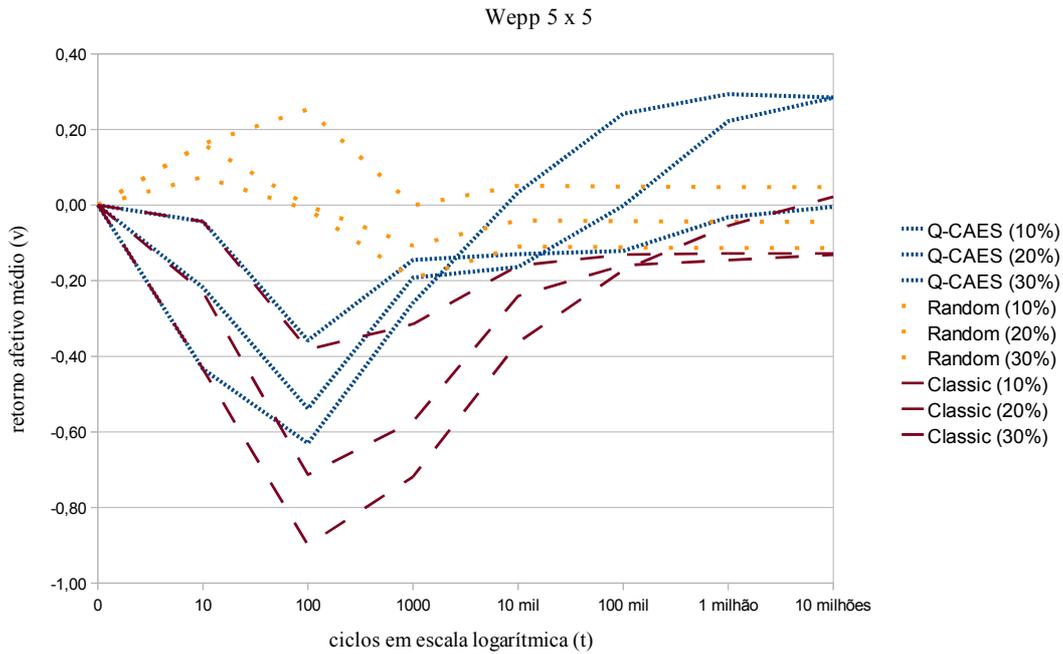


Figura 4.5: Curvas de convergência no tabuleiro 5 x 5 (25 células).

A implementação situada (CAES) converge um pouco mais rapidamente, e a solução é alcançada antes de 10 mil ciclos. A implementação clássica estabiliza-se apenas em 100 mil ciclos. O algoritmo de aprendizagem utilizado foi o Q-Learning.

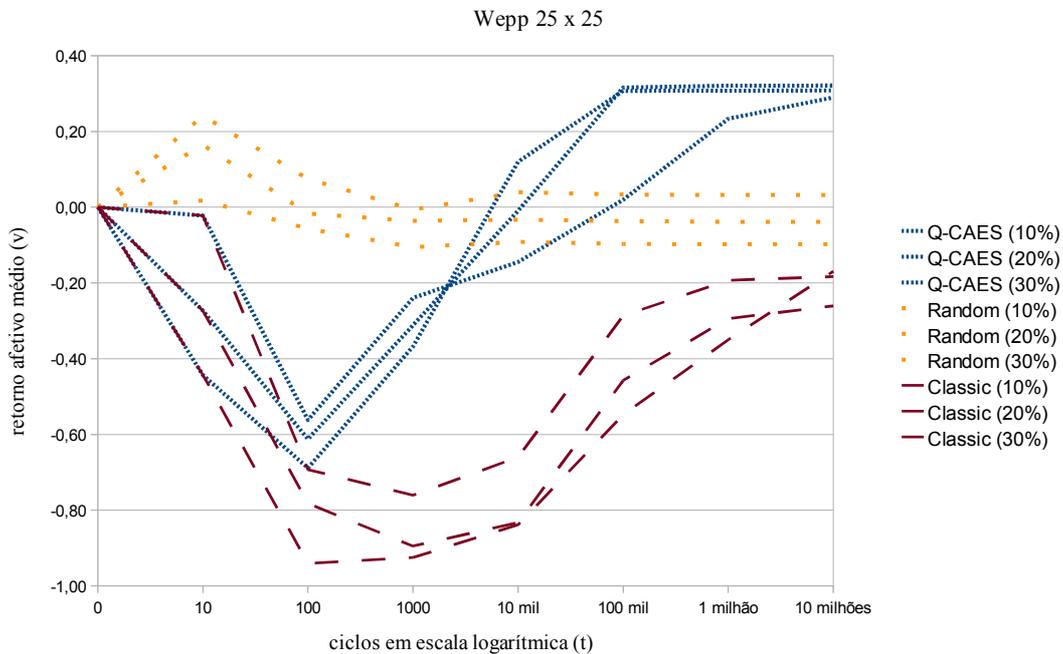


Figura 4.6: Curvas de convergência no tabuleiro 25 x 25 (625 células).

A implementação situada (CAES) não diminui o tempo de convergência apesar do aumento do tabuleiro, mas a implementação clássica passa de 100 mil para 1 milhão de ciclos.

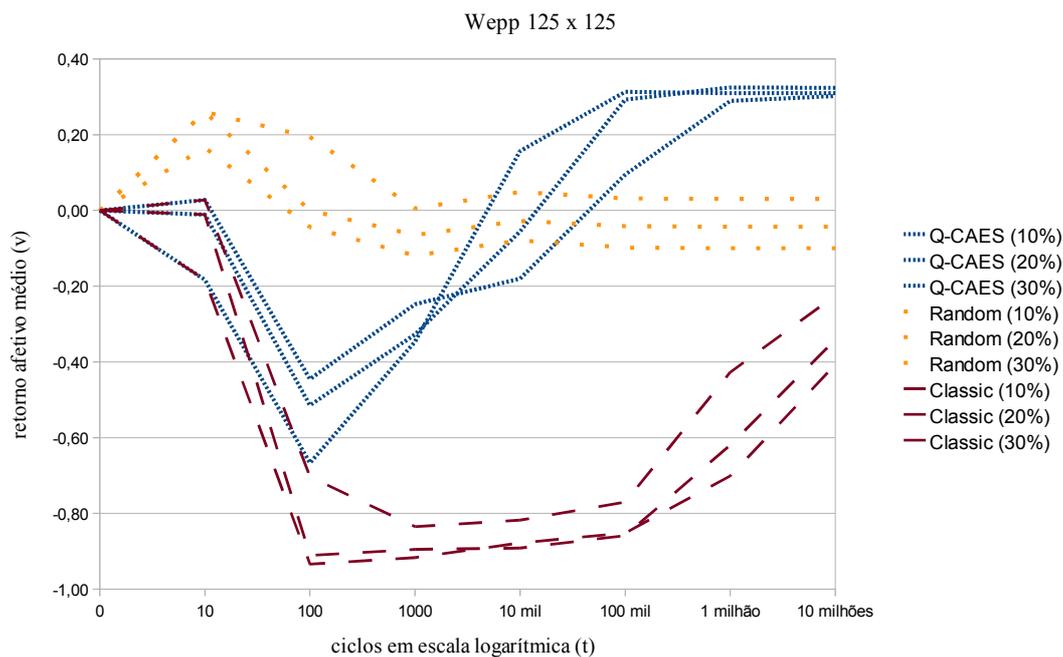


Figura 4.7: Curvas de convergência no tabuleiro 125 x 125 (15625 células).

A implementação situada (CAES) converge em 10 mil ciclos, enquanto a implementação clássica leva 10 milhões.

(491)

Quando o agente é situado (CAES), a solução não é afetada com o aumento da complexidade do tabuleiro, pois ela não está baseada no mapeamento do ambiente, e sim na descoberta das regularidades observadas na interação. Já a implementação clássica sofre com esse aumento, conforme a figura 4.8. Ou seja, para todas as configurações experimentadas, tanto a solução como o tempo de convergência foram semelhantes. Isso se deve pela utilização do modelo situado, que é definido na arquitetura CAES, onde o agente não tem a percepção onisciente do tabuleiro, mas sim um ponto de vista dele, conforme sua posição espacial e orientação.

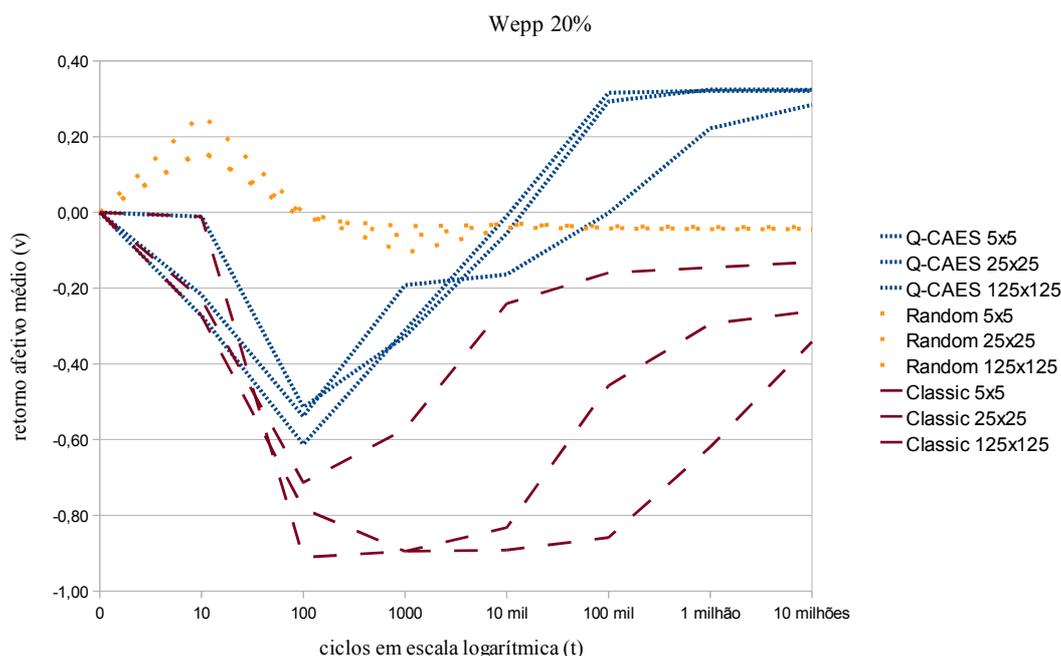


Figura 4.8: Curvas de convergência para 20% de obstáculos.

Vê-se claramente o impacto do aumento do tabuleiro sobre a implementação clássica (não-situada), que não ocorre quando utilizando a arquitetura CAES.

#### 4.1.2.2. Avaliando o CALM: solução construtivista

(492)

No início da simulação, o agente não sabe absolutamente nada sobre como funciona a dinâmica da interação com o ambiente, nem o que causa suas sensações. Ele não distingue os obstáculos dos caminhos livres, e também não sabe quais são as consequências implicadas por suas ações. Nestas condições, o mecanismo CALM foi capaz de constantemente convergir para a solução esperada, induzindo um modelo de mundo adequado para representar as regularidades do ambiente, bem como as regularidades de suas sensações corporais, e também para representar a influência regular de suas ações sobre ambos.

(493)

O agente aprende sobre as consequências de suas ações em diferentes situações, representadas através de poucos esquemas bastante gerais. A partir deles o mecanismo pode construir uma política de ações que lhe permite evitar as situações afetivamente negativas e buscar aquelas afetivamente positivas. Esta solução consegue descrever precisamente todas as regularidades que o agente poderia perceber sem construir um mapa completo do ambiente. O modelo de mundo construído pelo CALM é mostrado através de suas árvores de antecipação nas figuras 4.9, 4.10, 4.11, 4.12, 4.13, 4.14.

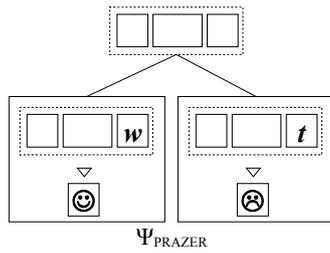


Figura 4.9: Árvore de antecipação do prazer.

Os esquemas indicam (da esquerda para a direita): “se andar então prazer”; e “se virar então não-prazer”.

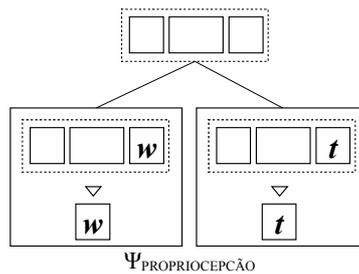


Figura 4.10: Árvore de antecipação da propriocepção.

“Se anda, percebe que está andando”; “se gira, percebe que está girando”.

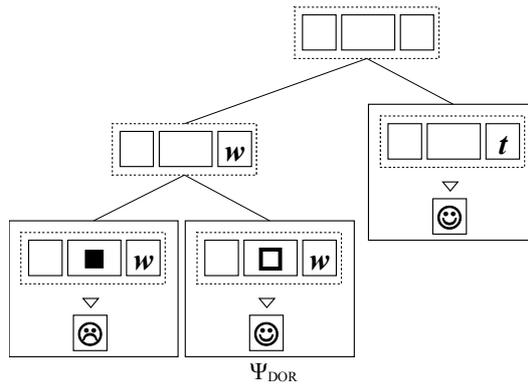


Figura 4.11: Árvore de antecipação da dor.

“Se vendo parede e andar então dor”; “se vendo vazio e andar então não-dor”; e “se virar então não-dor”.

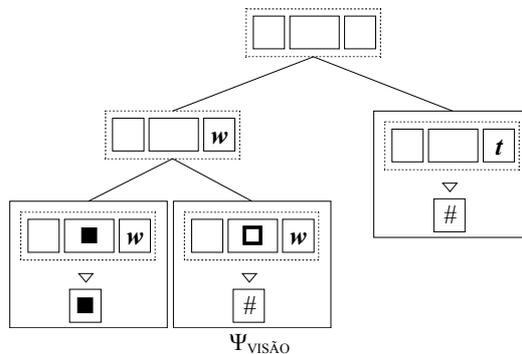


Figura 4.12: Árvore de antecipação da visão.

“Se vendo parede e andar então vê parede”; “se vendo vazio e andar então visão indeterminada”; e “se virar então visão indeterminada”.

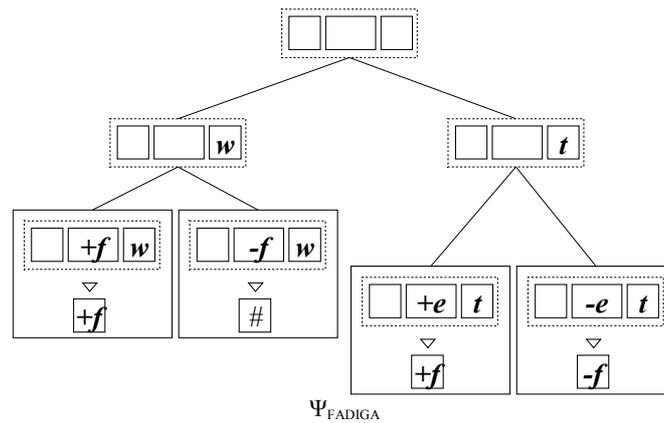


Figura 4.13: Árvore de antecipação da fadiga.

“Se fatigado e andar então fatigado”; “se não-fatigado e andar então fadiga indeterminada”; “se exausto e virar então fadiga”; e “se não-exausto e virar então não-fatigado”.

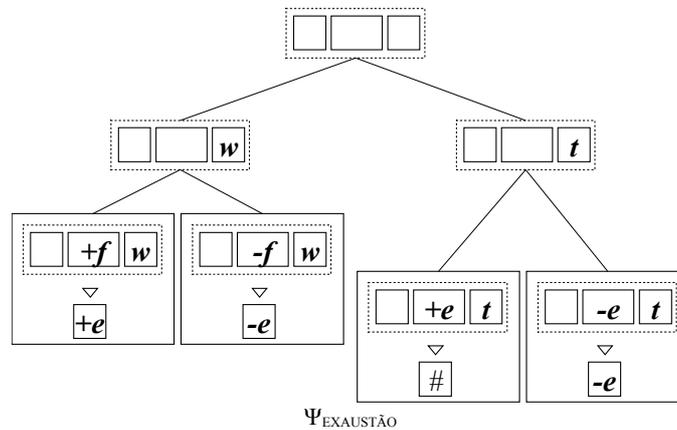


Figura 4.14: Árvore de antecipação da exaustão.

“Se fatigado e andar então exausto”; “se não-fatigado e andar então não-exausto”; “se exausto e virar então exaustão indeterminada”; e “se não-exausto e virar então não-exausto”.

(494)

A figura 4.15 mostra a evolução do comportamento apresentado por um agente CALM durante uma execução típica do experimento *wep*, realizada num cenário de dimensões  $25 \times 25$ , e 20% das células ocupadas por obstáculos. Os três primeiros gráficos (a, b, c) mostram respectivamente o prazer, a dor, e a exaustão do agente. O dois gráficos seguintes (d, e) representam as transformações cognitivas que o agente sofreu no tempo, respectivamente relativas à memória episódica e às árvores de antecipação. O número total de passos no eixo  $x$  dos gráficos equivale a 4000 ciclos de execução, plotando-se um ponto a cada 100 ciclos. Note-se que o comportamento ótimo é alcançado à altura de 2000 passos de execução, mas o modelo de mundo já é estável desde 500 ciclos, conforme pode ser visto no penúltimo gráfico (f), que descreve a média dos retornos afetivos. Alguns dos eventos negativos ocorrem mesmo depois da mente estar estabilizada devido ao comportamento exploratório do agente, ilustrado no

último gráfico (g), que mostra os momentos em que a escolha das ações foi feita por curiosidade.

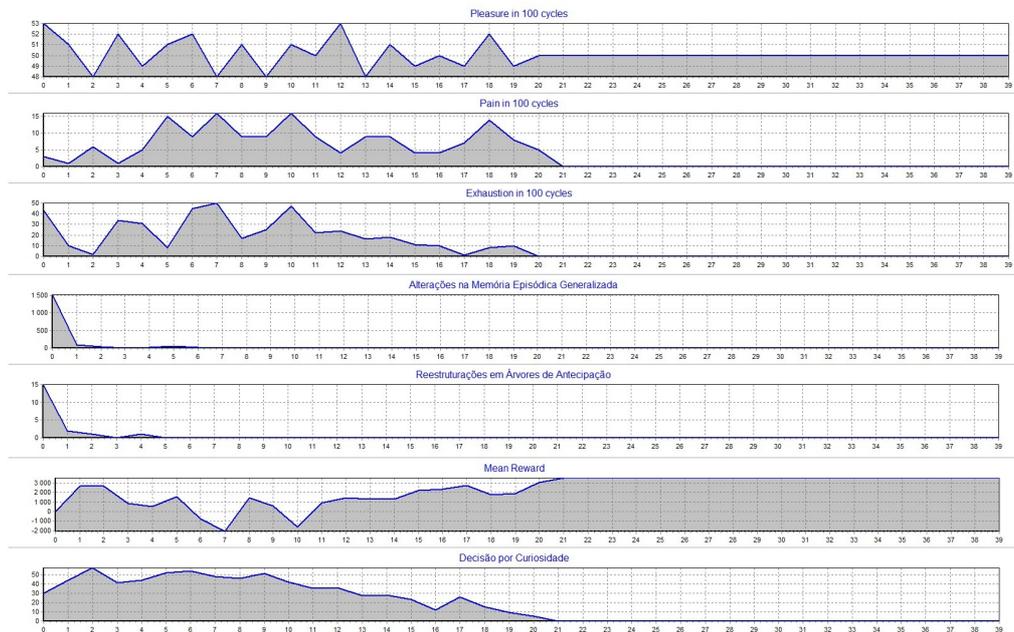


Figura 4.15: Resultados da simulação (caso típico).

Cada ponto no eixo x dos gráficos representa o número de eventos em 100 passos de execução. De cima para baixo, cada gráfico apresenta, respectivamente, a) prazer, b) dor, e c) exaustão; d) alterações na memória episódica, e e) transformações nas árvores de antecipação; f) recompensa média, e g) decisões exploratórias (tomadas por curiosidade).

(495)

A solução apresentada pelo mecanismo CALM mostrou-se ligeiramente superior em qualidade se comparada ao algoritmo Q-Learning, ambos implementados como agentes do tipo CAES, e substancialmente mais rápido no que se refere ao número de ciclos necessários para a convergência. As figuras 4.16, 4.17, e 4.18 mostram, respectivamente aos tabuleiros 5x5, 25x25, e 125x125, a comparação entre o CALM e o Q-Learning com a arquitetura CAES (representação situada), além do desempenho do Q-Learning num agente não-situado clássico (onde cada posição no tabuleiro é representada por um estado), e um agente de comportamento randômico.

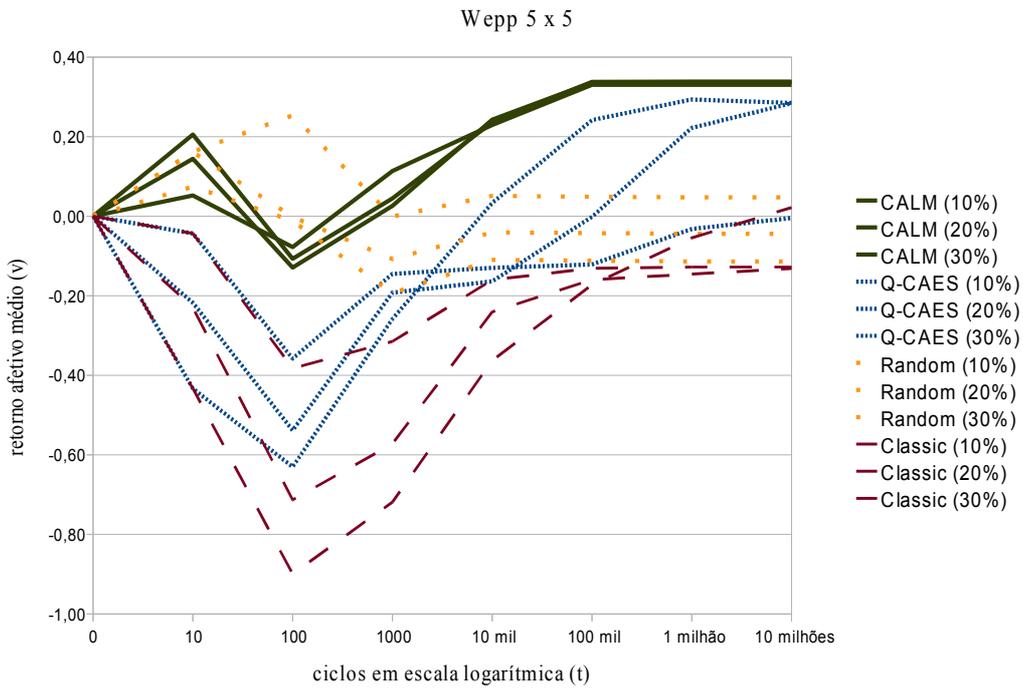


Figura 4.16: Wepp 5 x 5: Q-Learning x CALM.

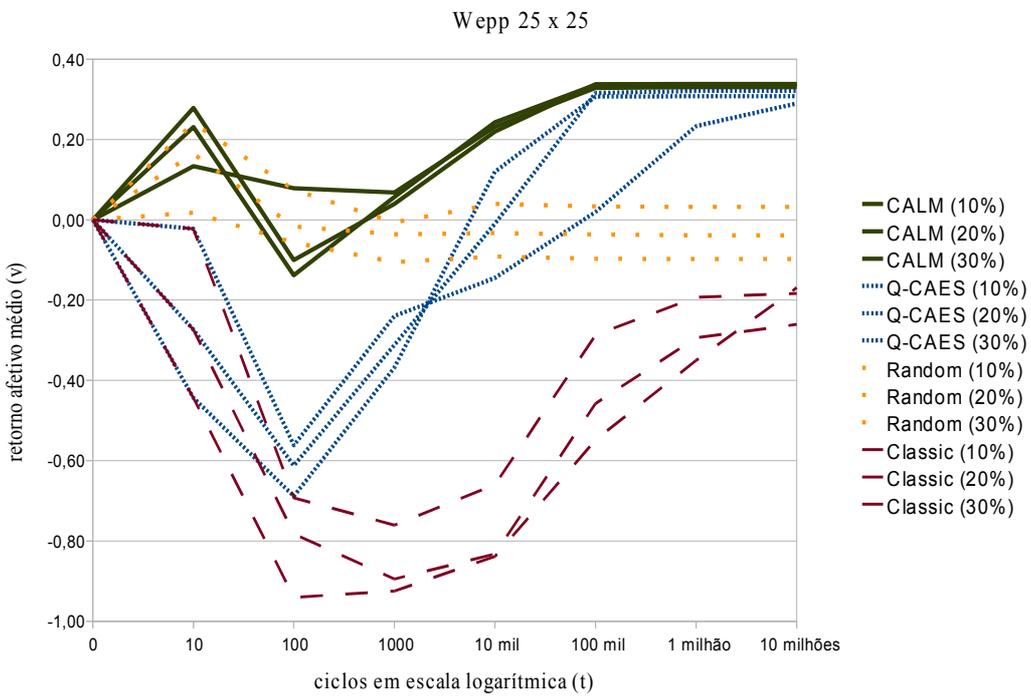


Figura 4.17: Wepp 25 x 25: Q-Learning x CALM.

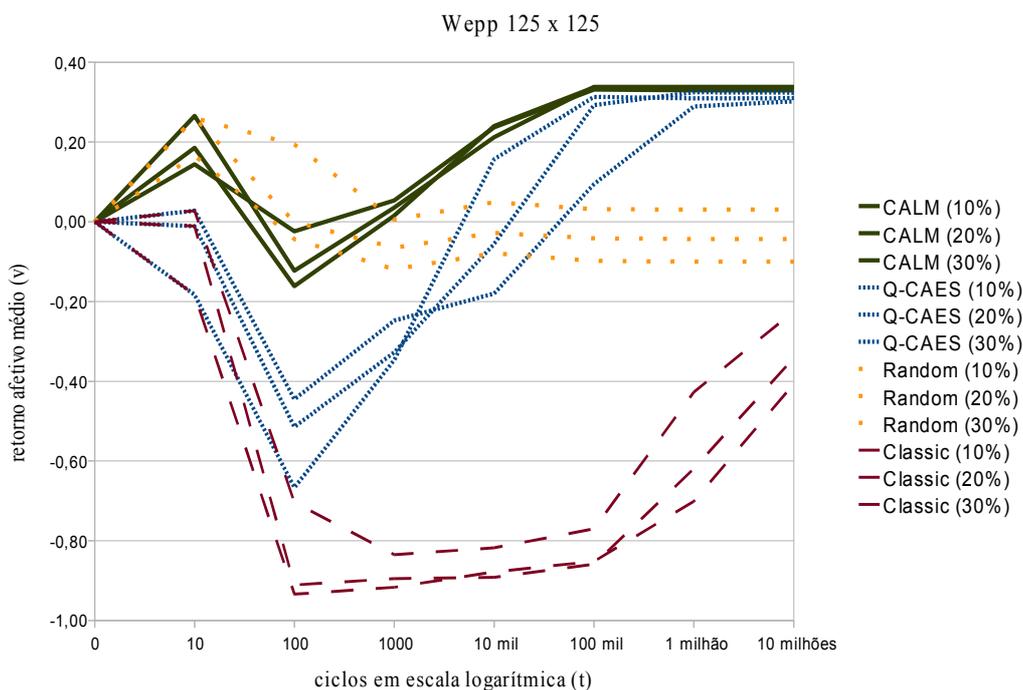


Figura 4.18: Wepp 125 x 125: Q-Learning x CALM.

#### 4.1.2.3. Escalabilidade

(496)

A seguir, no intuito de testar a escalabilidade do mecanismo, outra série de experimentos foi realizada, ainda utilizando-se a estrutura do problema wepp. Fez-se variar o número de variáveis do problema, pela introdução de propriedades que servem somente para inserir ruído no sinal perceptivo do agente. Ou seja, o problema continua fundamentalmente o mesmo, entretanto o número de variáveis em  $|P|$  é aumentado com propriedades irrelevantes e de dinâmica aleatória.

(497)

Dois tipos de agente foram submetidos a 4 cenários diferentes, repetindo-se o experimento 10 vezes em cada um. De um lado, utilizou-se um agente implementando o mecanismo CALM, e de outro um agente Q-Learning clássico, que representa os estados de forma rasa, como um espaço criado pela combinação dos valores das propriedades perceptivas. O primeiro cenário é o problema wepp original, num tabuleiro 25 x 25 com 20% de obstáculos, e nos cenários seguintes a quantidade de variáveis  $|P|$  é aumentada de 6 para 14, 22 e 30, inserindo-se a cada vez 8 novas variáveis aleatórias. Na representação clássica, o cenário mais complexo torna-se um problema de mais de um bilhão de estados.

(498)

A figura 4.19 mostra o desempenho do CALM, onde se pode ver que a média do número de ciclos necessários para atingir a estabilidade da solução varia muito pouco com a introdução de variáveis aleatórias. Isto se deve por que o CALM rapidamente antecipa seu comportamento não-determinístico, assim como sua irrelevância para a dinâmica do sistema. Diferentemente, o Q-Learning clássico sofre uma grande perda de desempenho com as variáveis irrelevantes.

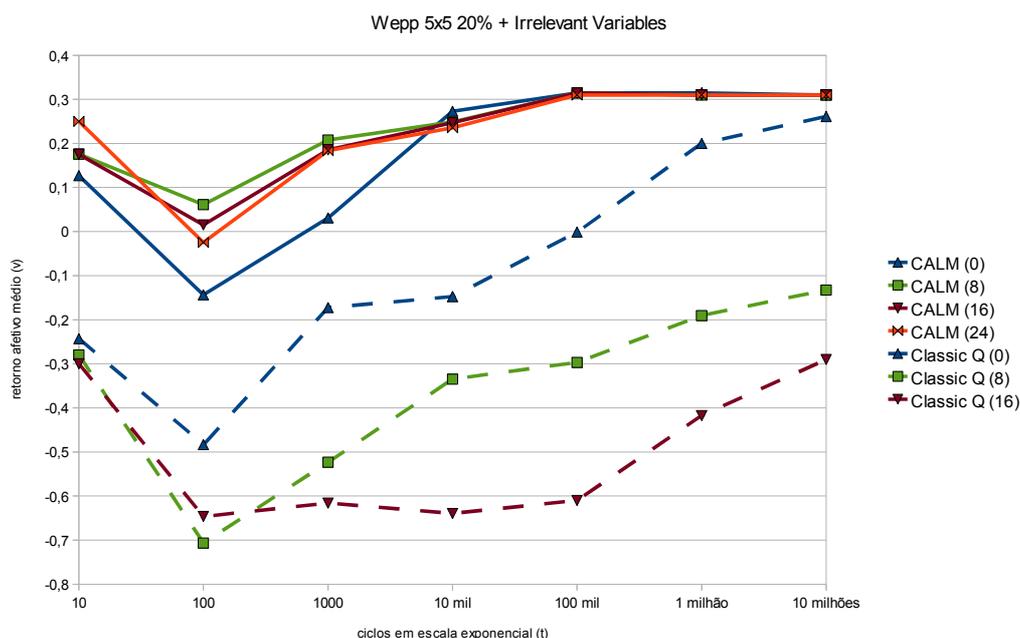


Figura 4.19: Análise de escalabilidade.

Resultados quando aumenta-se o número de propriedades irrelevantes de 0 a 8, 16, até 24.

## 4.2. Problema Flip

(499)

O problema *wepp*, apresentado anteriormente, não explora a capacidade que o mecanismo CALM tem para tratar o caso de observabilidade parcial. Afim de testar explicitamente esta capacidade, utilizamos o problema *flip*, proposto na literatura por (SINGH et al., 2003) e recentemente utilizado por (HOLMES; ISBELL, 2006). Trata-se de um agente vivendo num mundo composto por dois estados  $\{R, L\}$ , que são para ele ocultos. Ele tem um atuador que lhe permite realizar três ações  $\{l, r, u\}$ , e tem a percepção de dois valores possíveis  $\{0, 1\}$ . O agente observa '1' quando o estado subjacente troca, e '0' se ele permanece o mesmo. A ação  $u$  não faz nada, a ação  $l$  faz uma transição determinística para o estado  $L$  (à esquerda), e a ação  $r$  do mesmo modo para o estado  $R$  (à direita), conforme ilustra a figura 4.20.

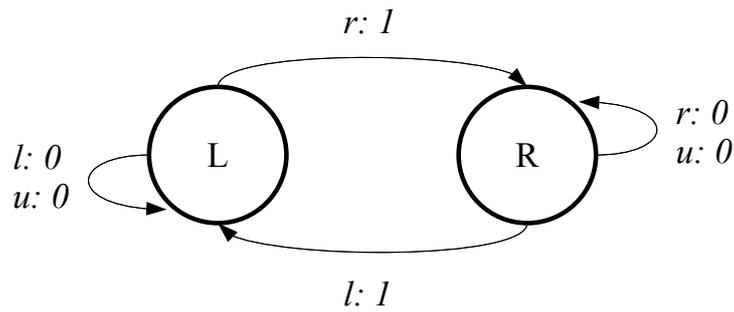


Figura 4.20: Problema *flip*, mostrado como uma máquina de estados.

(500) O problema *flip* pode ser descrito como um D-FPOMDP que possui apenas uma variável perceptiva  $P_1 = \{0, 1\}$ , representando a percepção da troca de estado, uma única variável de controle  $C_1 = \{l, r, u\}$ , e uma variável não-observável  $H_1 = \{R, L\}$ . A antecipação de  $p_1'$  e de  $h_1'$  se pode ser feita em função de  $h_1$  e  $c_1$ , na forma  $(H_1 \times C_1 \rightarrow P_1)$ , e  $(H_1 \times C_1 \rightarrow H_1)$ .

#### 4.2.1. Construção da Solução pelo CALM

(501) Uma vez enfrentando o problema *flip*, o mecanismo CALM começa a construção de seu modelo inferindo as regularidades que se baseiam apenas nas propriedades observáveis (conjunto  $P$ ) ou nas ações tomadas pelo agente (conjunto  $C$ ). Nessa primeira etapa, o modelo se estabiliza com uma única regularidade, representada pelo esquema  $[u \rightarrow 0]$ . Na figura 4.21 observamos essa primeira estabilização do modelo, quando o CALM possui apenas a árvore de antecipação  $\Psi_{P_1}$ , destinada a descrever a dinâmica da propriedade  $P_1$ . Nesse momento, a árvore possui 3 esquemas, diferenciados pela variável de controle  $C_1$ . Os esquemas que assimilam as ações “ $r$ ” e “ $l$ ”, ao longo dessa primeira etapa de aprendizagem, acabam com suas expectativas indeterminadas,  $[r \rightarrow \#]$  e  $[l \rightarrow \#]$ , uma vez que não é possível antecipar a observação seguinte apenas a partir da ação e observação anteriores.

(502) Quando o modelo se estabiliza, ou seja, quando os esquemas deixam de sofrer modificações por um longo período, a memória episódica pode passar por uma limpeza, eliminando tabelas que não guardem mais nenhuma informação relevante. Na figura 4.22 a memória episódica do problema é mostrada depois de eliminadas suas tabelas inúteis.

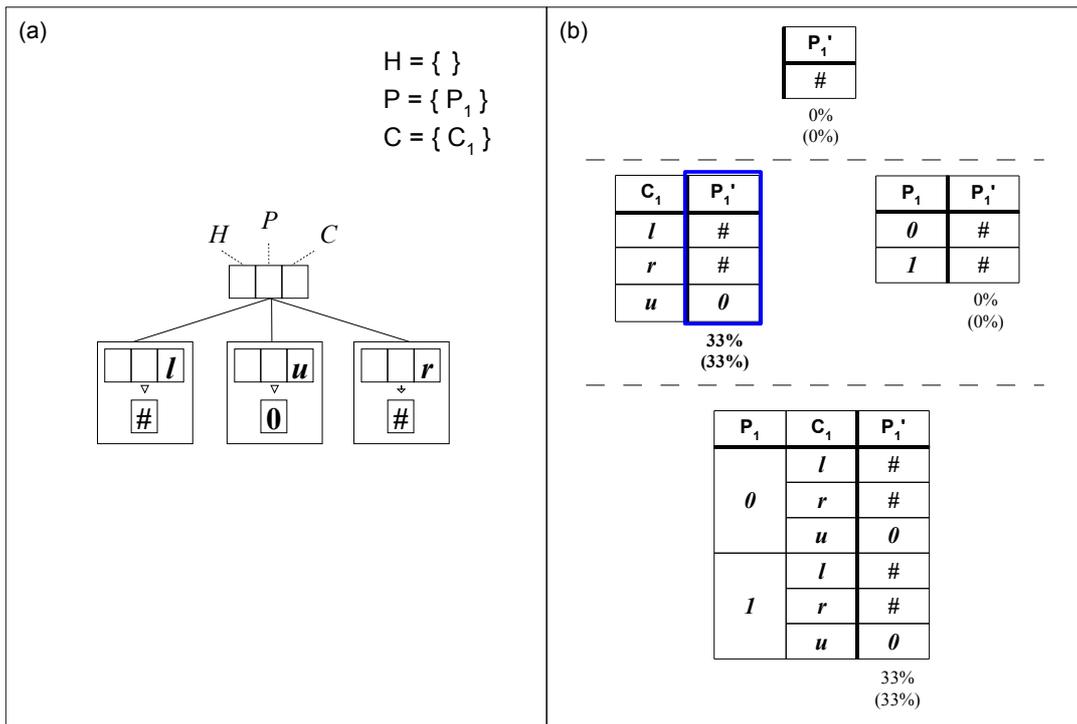


Figura 4.21: Começo da construção da solução CALM para o problema *flip*.

Em (a), a árvore de antecipação para a variável  $P_1$ , que tem como diferenciador a variável  $C_1$ . Em (b), o conteúdo da memória episódica nessa primeira etapa, onde o modelo se estabiliza sem a utilização de elementos sintéticos. A tabela destacada é aquela que dá origem à árvore.

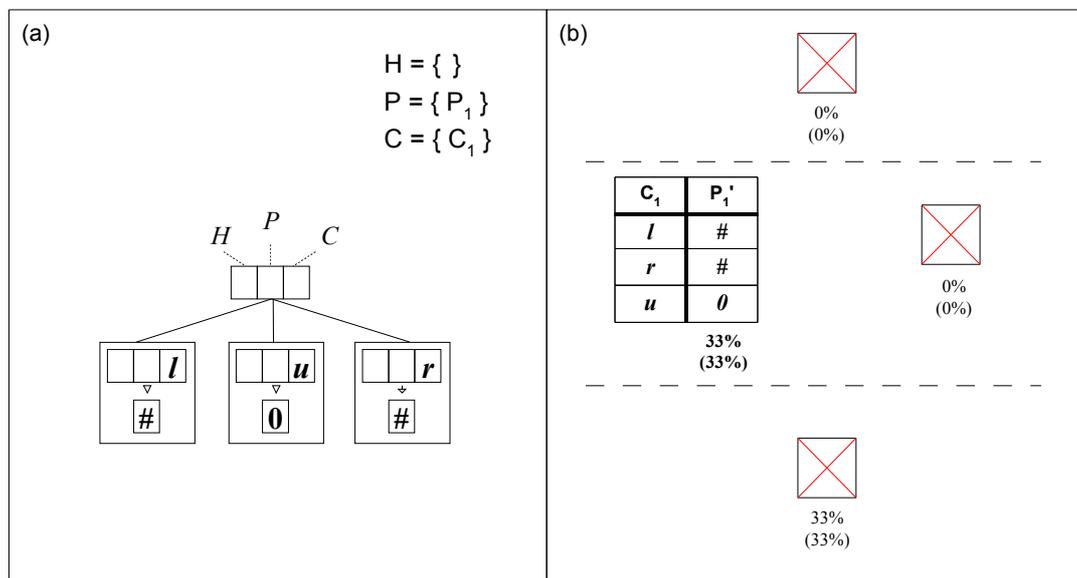


Figura 4.22: Primeira estabilização da solução, ainda sem elementos sintéticos.

Em (a), a árvore de antecipação para a variável  $P_1$ . Em (b), após a estabilização do conhecimento, as tabelas inúteis da memória episódica são eliminadas, seja por conterem menos informação que a tabela destacada, seja por conterem a mesma informação gastando mais espaço.

(503)

Na figura 4.23, depois da primeira estabilização, o CALM cria um elemento sintético,  $H_l = \{ \clubsuit, \diamond \}$  visando melhorar a antecipação de  $P_1$ , e supondo, para isso, a

existência de uma propriedade não-observável capaz de precisar as expectativas indeterminadas na árvore de antecipação atual. O elemento sintético  $H_1$  é utilizado para diferenciar o esquema  $[r \rightarrow \#]$ , que agora será especializado em  $[\clubsuit r \rightarrow 1]$  e  $[\diamond r \rightarrow 0]$ . Esses mesmos valores são colocados na tabela relativa da memória episódica, cujas respectivas células são chamadas “situações âncora”, porque são as situações que determinam, a posteriori, o estado da condição não-observável.

(504)

No problema, fazer  $r$  no tempo  $t$  e ver 1 em  $t+I$  significa que o agente estava em  $\clubsuit$  em  $t$ , e da mesma forma, fazer  $l$  no tempo  $t$  e ver 0 em  $t+I$  significa que ele estava em  $\diamond$  em  $t$ . A inclusão desse novo elemento sintético,  $H_1$ , no modelo de mundo do agente desencadeia duas mudanças principais. Primeiramente, uma nova tabela é adicionada à memória episódica  $\mathbb{M}_{P_1}$ , que é memória responsável por manter as observações relativas à antecipação de  $P_1$ . Essa nova tabela é uma cópia daquela que até então dava origem à árvore  $\Psi_{P_1}$ , ou seja, a tabela com maior índice observado de determinismo, adicionada do novo elemento sintético. Essa nova tabela herda os valores determinados da tabela já estável, e então, no caso, a célula de memória  $[u \rightarrow 0]$  definirá  $[\clubsuit u \rightarrow 0]$  e  $[\diamond u \rightarrow 0]$ . Em segundo lugar, uma nova árvore de antecipação,  $\Psi_{H_1}$ , é criada para realizar a antecipação de  $H_1$ , sendo inicializada com um esquema único. Juntamente, toda uma nova série de tabelas é criada, compondo a memória episódica  $\mathbb{M}_{H_1}$ .

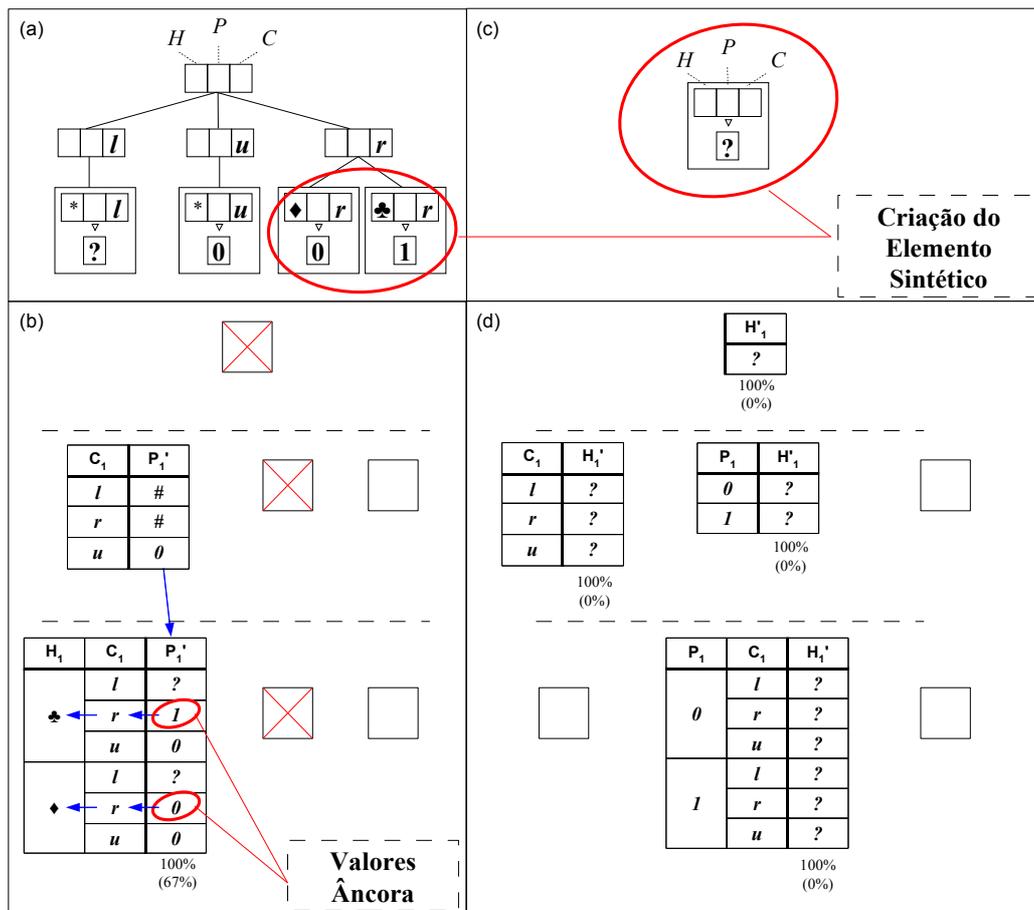


Figura 4.23: Criação do elemento sintético.

Uma diferenciação abstrata é gerada a partir do esquema “r” em (a), e uma nova árvore em (c) para tentar antecipar a dinâmica dessa propriedade não-observável. As memórias episódicas relacionadas à antecipação de  $P_1$  são mostradas em (b), onde é inserida uma nova tabela contendo as situações âncora, as quais definem o estado da condição não-observável. As memórias episódicas relacionadas à antecipação de  $H_1$ , inicializadas com expectativas inexistentes, são mostradas em (d).

(505)

Na figura 4.24 é mostrado o estado do modelo no momento de uma segunda estabilização, quando o modelo já conta com o elemento sintético. O esquema único da árvore  $\Psi_{HI}$  é diferenciado pela variável  $C_1$ , representando a informação contida na tabela de memória em destaque. Nessa fase, a antecipação do elemento sintético  $H_1$  é ainda parcial porque as memórias episódicas relacionadas a sua antecipação só são registradas quando seu valor é confirmado pela ocorrência de uma situação âncora em  $\mathbb{U}_{P_1}$ .

(506)

Por exemplo, toda vez que o agente realiza uma sequência de ações  $t_1 : [r]$  e  $t_2 : [r]$ , conforme a definição do problema, a observação em  $t_2$  será sempre  $[0]$ , ou seja,  $t_2 : [r \rightarrow 0]$ , que é justamente uma das situações âncora, o que permite concluir que o elemento sintético em  $t_2$  é constantemente  $[\spadesuit]$ , ou seja,  $t_2 : [\spadesuit r \rightarrow 0]$ . A partir dessa informação é possível então deduzir que  $[r \rightarrow \spadesuit]$ , e incluir esta regularidade no modelo.

O equivalente ocorre para o caso  $t_1 : [l]$  e  $t_2 : [r]$ , que coincidirá constantemente com a observação da situação âncora  $t_2 : [\clubsuit r \rightarrow 1]$ , resultando na descoberta da regularidade  $[l \rightarrow \clubsuit]$ .

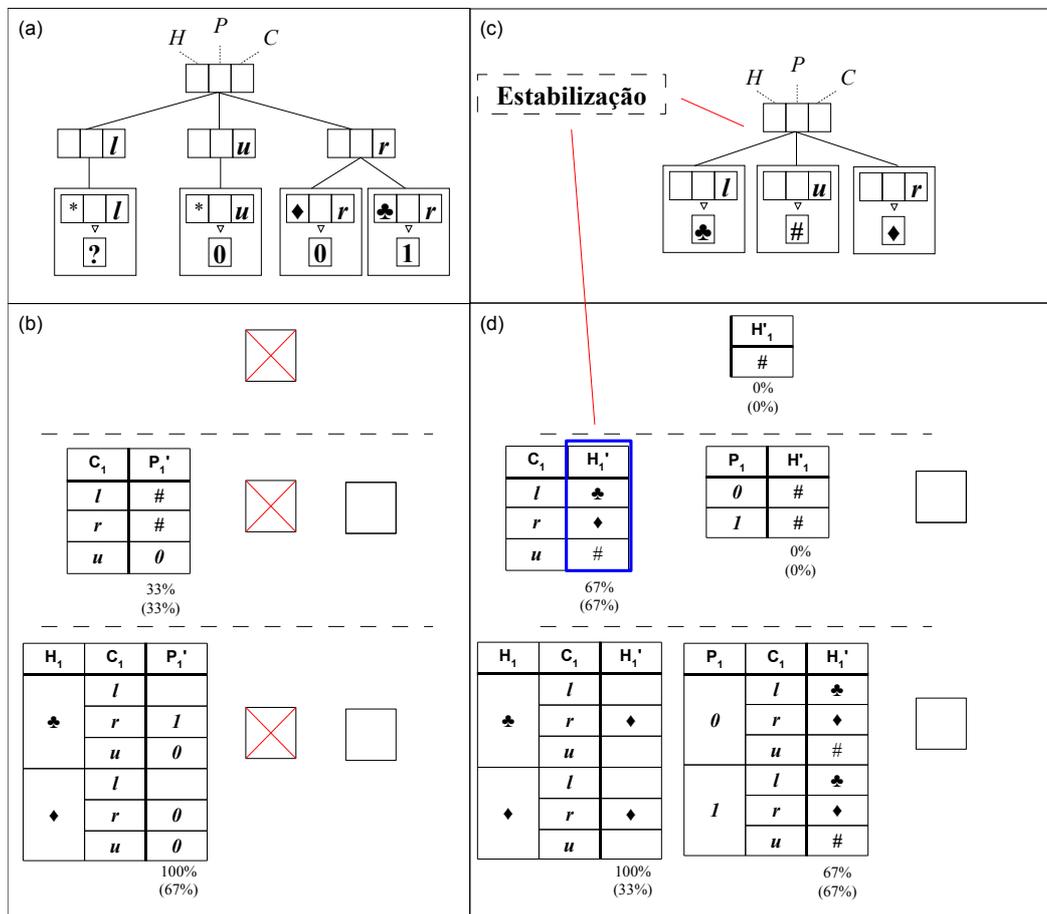


Figura 4.24: Segunda estabilização da solução.

Nesse ponto, a antecipação do próprio elemento sintético começa a se construir.

(507)

De forma similar, as tabelas da memória nas quais o elemento sintético aparece como condição só podem ser atualizadas quando o agente está vivenciando uma situação âncora. Por exemplo, suponhamos uma ocorrência de  $t_1 : [l r \rightarrow 1]$  e  $t_2 : [l r \rightarrow 0]$ , de onde deduz se que  $t_1 : [\clubsuit l r \rightarrow \diamond 1]$  e  $t_2 : [\diamond l r \rightarrow 0]$ . A situação  $t_1$  permitirá a atualização das células  $[\clubsuit l \rightarrow \diamond]$  e  $[\clubsuit \rightarrow \diamond]$  nas suas respectivas tabelas.

(508)

A ocorrência posterior de  $t_3 : [\clubsuit l l \rightarrow \clubsuit 0]$  e  $t_4 : [\clubsuit 0 r \rightarrow 1]$  vai indeterminar as observações nas células, na forma  $[\clubsuit l \rightarrow \#]$  e  $[\clubsuit \rightarrow \#]$ . Situações similares vão conduzir as demais células das tabelas à  $[\clubsuit 0 \rightarrow \#]$ ,  $[\diamond 0 \rightarrow \#]$ ,  $[\diamond l \rightarrow \#]$ , e  $[\diamond \rightarrow \#]$ , tornando essas tabelas inúteis e portanto elimináveis, conforme mostrado na figura 4.25. Nesse momento, algumas regularidades observadas nas tabelas que não possuem  $H_1$  como

condição podem ser transmitidos para suas correlatas que incluem o elemento sintético na condição, como também é ilustrado na figura.

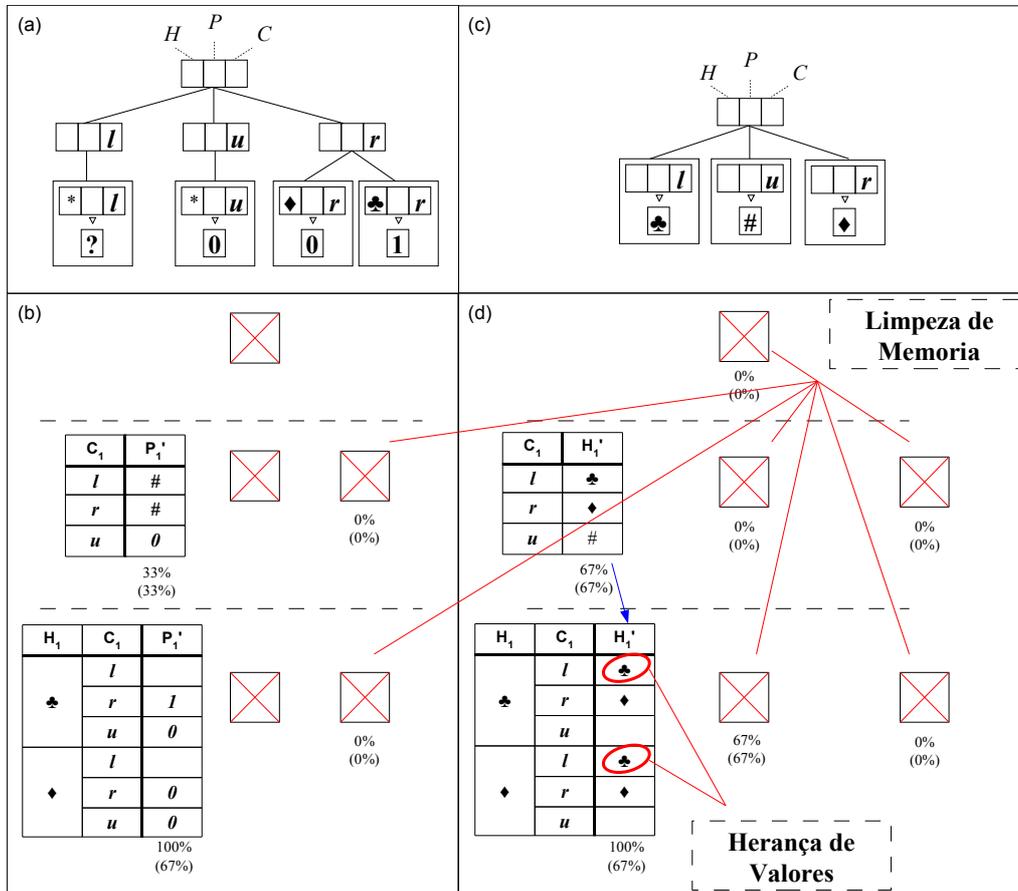


Figura 4.25: Nova tabela da memória episódica herda valores de sua correspondente.

Nesse processo ocorre a transmissão dos valores definidos nas tabelas que não têm o elemento sintético como condição, às tabelas que o têm.

(509)

A figura 4.26 ilustra as últimas modificações na árvore  $\Psi_{HI}$ , que se estabiliza definindo a antecipação abstrata para o esquema relacionado à ação  $[u]$ , na forma  $[u \rightarrow \approx]$ , significando que o valor de  $H_l$  não é alterado pela execução da ação  $[u]$ , e portanto  $[\spadesuit u \rightarrow \spadesuit]$  e  $[\clubsuit u \rightarrow \clubsuit]$ .

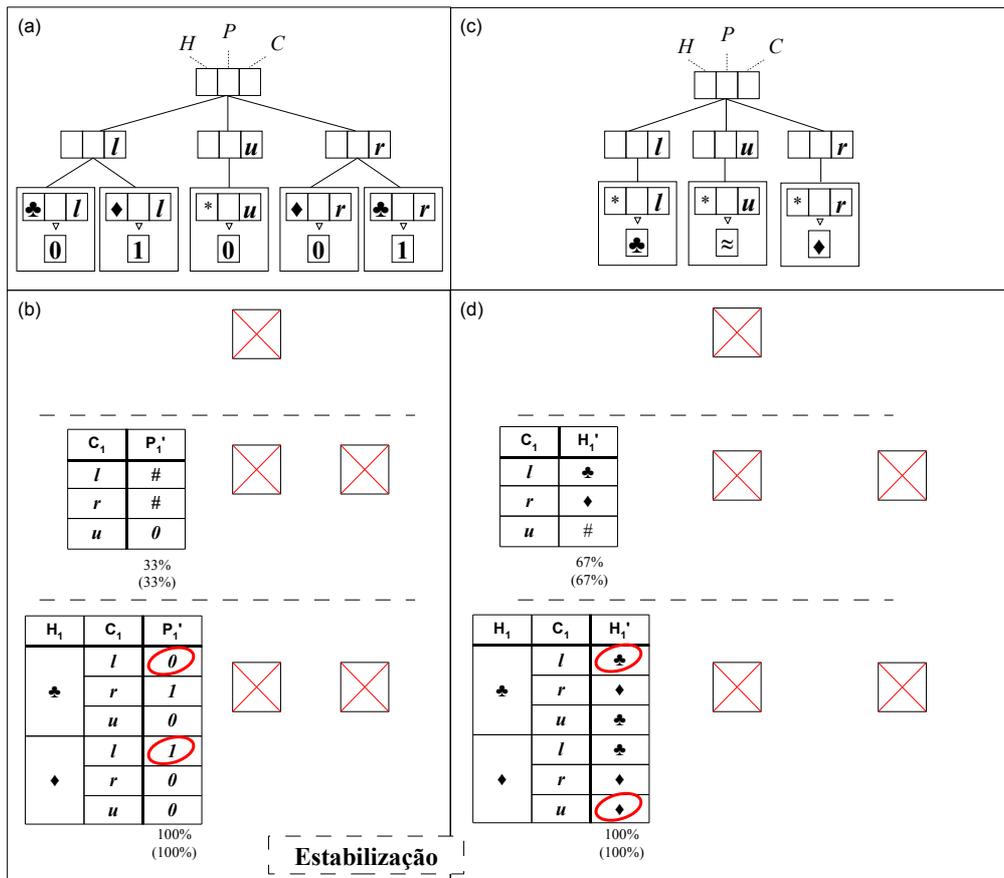


Figura 4.26: Última estabilização, com a solução final.

### 4.2.2. Comparação das Soluções

(510)

O CALM é capaz de resolver o problema *flip* através da criação de um novo elemento sintético para representar os estados subjacentes esquerdo ( $\clubsuit$ ) e direito ( $\spadesuit$ ). A figura 4.27 mostra as árvores de antecipação finais construídas pelo mecanismo. O problema *flip*, embora aparentemente simples, não é trivial para diversos algoritmos de aprendizagem (SINGH et al., 2003) porque o ciclo criado pela ação  $u$  pode se repetir indefinidamente sem deixar evidências históricas observáveis.

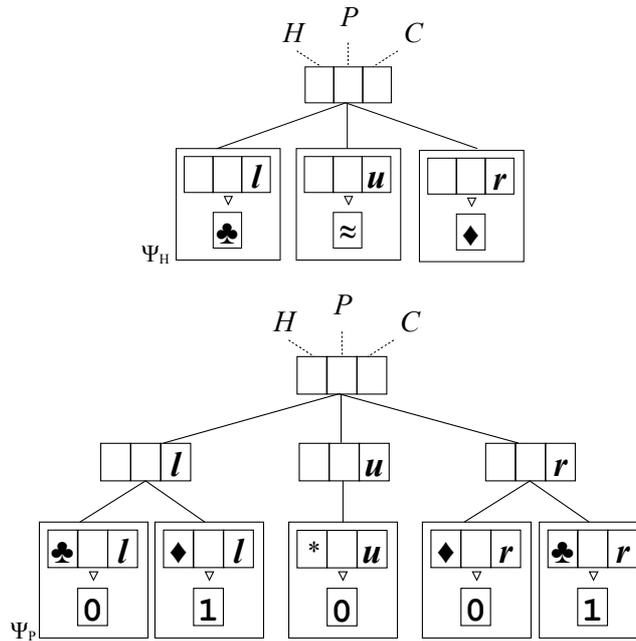


Figura 4.27: Árvores de antecipação construídas pelo CALM no experimento *flip*.

(511)

Para adaptar o problema *flip* ao mecanismo CALM, foram atribuídos valores afetivos para as percepções, de forma que o agente recebe um valor positivo quando observa 1, e negativo quando observa 0. O CALM calcula a utilidade dos esquemas na árvore de deliberação tanto para exploração quanto para o desempenho, o que possibilita a descoberta do modelo completo do problema. A árvore de deliberação é mostrada na figura 4.28.

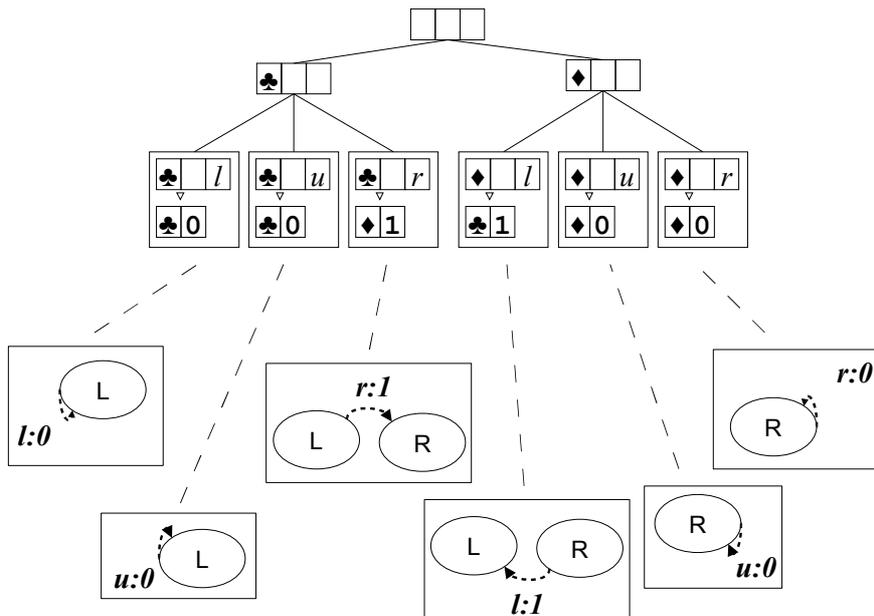


Figura 4.28: Árvore de deliberação construída pelo CALM para o problema *flip*.

(512)

Holmes e Isbell (2006) apresentaram com sucesso um algoritmo para a solução do problema *flip*, através da aprendizagem de árvores de prefixo (PST). Seu método, embora eficiente, apresenta duas desvantagens em relação ao CALM. Primeiramente, o PST não faz um modelo de mundo, ou seja, ele não tenta representar as variáveis não-observáveis do ambiente. A representação das situações no PST é diretamente baseada nas sequências históricas de observações e ações anteriores. Em segundo lugar, o método não é incremental, sendo necessário apresentar todo o conjunto de observações de uma só vez, no início do processo. A PST que resolve o problema *flip* é mostrada na figura 4.29.

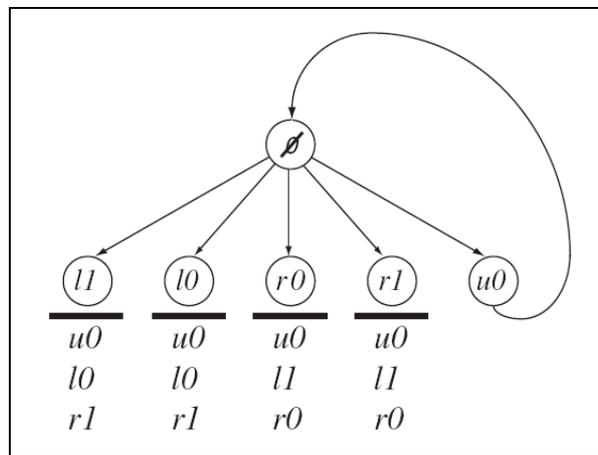


Figura 4.29: PST para o problema *flip*.

O nó raiz é um nó vazio, e cada ramo representa uma sequência de interação possível. Abaixo desses nós terminais apresentam-se as antecipações. Por exemplo, o primeiro ramo, à esquerda, indica que ao fazer a ação 'l' e observar '1', pode-se antecipar que ao executar a ação "u" logo em seguida, observa-se "0", ao executar "l" também observa-se "0", e ao executar "r", observa-se "1". O ramo da direita mostra o ciclo gerado pela ação "u".

## 5. CONCLUSÃO

(513) A inteligência artificial é possivelmente a principal promessa de revolução tecnológica do século XXI. Pouco a pouco os sistemas inteligentes, sejam eles robóticos ou virtuais, estão aparecendo e ganhando espaço em diversos domínios da vida, seja no cotidiano doméstico, seja na indústria, seja na internet. Por essa razão é que ganha importância a pesquisa de base, aquela que visa repensar a disciplina e compreender os mecanismos gerais da inteligência, em particular da aprendizagem autônoma, definindo novas estratégias e métodos que possam ser então aplicados no desenvolvimento de novas tecnologias, e na solução de problemas que até o momento não podem ser enfrentados pelos computadores de forma inteligente, sem uma tutela muito estreita dos operadores e programadores.

(514) Uma das formas de representação correntes para problemas de decisão são os *Processos de Decisão de Markov* (MDP), especialmente na versão fatorada (FMDP) e na versão parcialmente observável (POMDP). Embora diversos trabalhos tenham proposto mecanismos para o cálculo da política de ações à partir de um dado FMDP, ou de um dado POMDP, pouca atenção foi dada ao problema de se aprender a estrutura do processo markoviano através da observação. Trata-se, entretanto, de um problema fundamental, pois um agente autônomo inserido em um ambiente desconhecido precisa descobrir a estrutura desse ambiente através de suas observações e interações.

(515) Dentro desse contexto, a realização desta tese culminou com: (1) a proposição do mecanismo de aprendizagem CALM (*Constructivist Anticipatory Learning Mechanism*), concebido justamente para a construção de modelos de mundo através da interação; e (2) a proposição da arquitetura CAES (*Coupled Agent-Environment System*), que estrutura a relação de um agente autônomo e situado com seu ambiente.

## 5.1. Considerações sobre o Mecanismo CALM

(516) A principal contribuição desta tese é a proposição de um mecanismo geral de aprendizagem inspirado na teoria construtivista, CALM, apresentado na seção 3.2, desenvolvido para resolver importantes problemas de aprendizagem automática, estes definidos na seção 3.1.

(517) O CALM é um mecanismo original para a construção de modelos de mundo por agentes artificiais a partir da própria interação com o ambiente, num processo realizado de forma autônoma. O CALM foi elaborado para ser aplicado a problemas de aprendizagem ativa e contínua associada a decisão e planejamento. Problemas em que o agente precisa construir seu modelo de mundo de forma incremental, numa única etapa de interação ininterrupta com o ambiente, e ao mesmo tempo precisa definir uma política de ações a fim de otimizar seu comportamento.

(518) O modelo de mundo que o CALM constrói descreve as regularidades determinísticas do ambiente em que o agente está situado, ainda que existam fenômenos não-determinísticos nesse universo. Conforme argumentado no capítulo 3, em universos bem estruturados, ainda que complexos como o mundo real, a maioria dos fenômenos se apresenta como transformações regulares e determinísticas se as condições causais estão bem identificadas.

(519) O CALM também é capaz de descobrir regularidades mesmo quando elas são dependentes de propriedades não-observáveis do ambiente. A estratégia é aumentar o vocabulário representacional do agente acrescentando elementos sintéticos que podem estar associados a propriedades ocultas, condições sequenciais, ou condições abstratas presentes nas situações vivenciadas pelo agente.

(520) O mecanismo CALM permite ao agente construir um modelo de mundo na forma de árvores de antecipação, gerando a necessidade de tratar o problema da fragmentação do conhecimento. Em problemas complexos, o aumento da profundidade ramifica a árvore de maneira que o número de nós pode crescer exponencialmente. Assim, uma parte importante do desafio neste trabalho foi garantir que as árvores tivessem um tamanho controlado, mantendo o problema em níveis computacionais tratáveis. Isso foi conseguido através da seleção das variáveis usadas para compor as árvores, estimando a relevância de cada variável para cada transformação modelada.

- (521) No CALM, a seleção de propriedades relevantes durante a construção de cada uma das árvores de antecipação se processa através da análise da memória episódica generalizada associada à árvore. Essa estrutura exige uma grande quantidade de espaço de armazenamento, mas seu crescimento é controlado pelo mecanismo. A memória episódica guarda uma lembrança generalizada das situações vividas pelo agente, referentes a uma determinada transformação cuja antecipação está sendo construída. Seu tamanho é controlável justamente por que ela é generalizada, podendo observar simultaneamente apenas um número limitado de condições.
- (522) A eficiência do mecanismo está condicionada ao fato de que o ambiente seja bem estruturado, ou seja, que o número de propriedades relevantes para descrever cada uma das transformações regulares seja no máximo de ordem logarítmica com relação ao número total de propriedades do problema. Se isso é garantido, então o ambiente enfrentado pelo CALM pode ser parcialmente observável e parcialmente determinístico.
- (523) Uma vantagem da representação utilizada pelo CALM é que as árvores são construídas de forma independente, e então podem ser processadas em paralelo. A paralelização da aprendizagem é importante pois permite quebrar a complexidade de processamento do problema.
- (524) O mecanismo foi desenvolvido para tratar objetivos de longo-prazo, em problemas de decisão sequencial, onde o agente precisa construir uma política de ações visando à maximização da eficiência de seu comportamento. Assim, implementamos um método de cálculo da utilidade das ações, baseado nas equações de Bellman, que utiliza o modelo de mundo construído pelo agente para então definir uma política de ações na forma de árvores de deliberação.
- (525) Experimentos realizados nos problemas *wepp* e *flip* (no capítulo 4) mostraram que um agente artificial que implementa o mecanismo CALM é capaz de convergir satisfatoriamente para as soluções esperadas. Esses resultados confirmam a capacidade do mecanismo para descobrir regularidades na interação com o ambiente, representando-as em suas árvores de antecipação, e usar esse conhecimento para construir uma boa política de ações, tornando o agente mais adaptado a seu mundo.

## 5.2. Considerações sobre a Arquitetura CAES

(526) No intuito de implementar um modelo construtivista de inteligência artificial, foi necessário realizar uma reflexão mais fundamental a respeito do conceito de agentes autônomos. Assim, outra contribuição desta tese, apresentada no capítulo 2, é o debate teórico a respeito da ideia de agente autônomo, contemplando as noções de situatividade, corporificação e motivação intrínseca, que resultou na proposição formal da arquitetura CAES, utilizada na definição dos experimentos.

(527) A arquitetura CAES define agente e ambiente como dois sistemas parcialmente abertos e acoplados um ao outro, onde o agente é situado, precisando coordenar seu comportamento com a dinâmica do sistema global. No CAES há três tipos fundamentais de interação: mais externamente, a interação entre agente e ambiente; a seguir, dentro do agente, a interação entre seu corpo e sua mente; e por fim, dentro da mente, a interação entre o sistema cognitivo e o sistema regulativo.

(528) A arquitetura CAES define um modelo formal, que inova pela integração de conceitos oriundos de dois paradigmas distintos: a IA Afetiva e a IA Situada. O agente adquire um corpo como um universo diferenciado tanto do mundo (exterior) quanto da mente (interior). A mente, então, torna-se o sistema que interage com o corpo, e através dele com o exterior.

(529) Na mente do agente, a arquitetura CAES estabelece, de um lado, a atuação de um sistema regulativo, que gerencia os comportamentos emocionais e reativos, e de outro, a atuação de um sistema cognitivo, capaz de aprender um modelo de mundo e uma política de ações. O sistema regulativo é também responsável pelos sinais afetivos. No CAES, os objetivos do agente não são estabelecidos na forma de estados-alvo ou recompensas externas. É o sistema afetivo que avalia de forma fatorada a transformação dos sinais perceptivos do agente, motivando-o a agir através da definição de sensações inerentemente agradáveis ou desagradáveis.

## 5.3. Propriedades Não-Observáveis e Abstração

(530) A maior parte dos fenômenos interessantes do nosso mundo só pode ser expressa através de conceitos de mais alto-nível. Por isso um dos grandes desafios colocado para

a IA é encontrar maneiras de superar os limites da percepção sensorial direta, e atingir modos de pensamento e representação mais abstratos.

(531) Algoritmos tradicionais de classificação são capazes de generalizar conjuntos de situações, realizando um primeiro afastamento da percepção sensorial direta. Entretanto, a descrição das classes é feita através de combinações das propriedades perceptivas, o que implica que, de alguma forma, as classificações permanecem no nível sensorial, sem de fato constituir elementos radicalmente novos de representação.

(532) O CALM implementa um método para a descoberta de propriedades não-observáveis do ambiente, baseado na análise da informação histórica da experiência sensorial do agente. Supor a existência dessas variáveis ocultas é uma maneira de discernir situações sensorialmente ambíguas em ambientes parcialmente observáveis. Esta descoberta acontece por meio da criação de elementos sintéticos de representação, que podem ser vistos como estando um passo além do nível sensorial. O mecanismo possui a capacidade de aprender a antecipar os valores dessas propriedades ocultas, que então, passam a ser usadas na construção da política de ações do mesmo modo que os elementos sensoriais.

(533) Neste sentido, o CALM é uma contribuição porque oferece uma solução diferenciada para este problema, que é bastante relevante no domínio da aprendizagem de máquina. A estratégia utilizada pelo CALM é inspirada na ideia proposta em (DRESCHER, 1991), que já estabelecia a utilização de elementos sintéticos para representar propriedades não-observáveis do ambiente. Entretanto, o método de aprendizagem utilizado por Drescher é fundado sobre uma abordagem estatística que apresenta custos computacionais inviáveis (CHAPUT, 2004).

(534) Entre os trabalhos recentes, o método utilizado pelo CALM para diferenciar os estados confusos pode ser comparado àquele apresentado por (HOLMES; ISBELL, 2006), que realiza antecipações em ambientes representados como D-POMDPs. Entretanto, sua solução difere-se do CALM, primeiramente porque é baseada na representação rasa dos estados, e além disso, porque funciona no modo off-line, e exige que o ambiente subjacente, embora parcialmente observável, seja completamente determinístico.

(535) O CALM restringe o problema ao aprendizado apenas das regularidades determinísticas do ambiente, mas não exige que o ambiente seja completamente determinístico, ou seja, o mecanismo se aplica a universos parcialmente determinísticos. Isso permite que ele implemente um método de aprendizagem diretamente indutivo, consistindo uma abordagem que, no alcance do conhecimento dos autores desta pesquisa, não foi utilizado em nenhum outro trabalho.

(536) O interessante é que uma propriedade não-observável pode representar diversos tipos de condição abstrata de um ambiente. Três casos podem ilustrar adequadamente esta afirmação: (a) ambientes particionados em sub-ambientes; (b) ambientes discretamente não-estacionários; e (c) efeitos produzidos por encadeamento de ações.

(537) O primeiro caso (a) acontece quando um agente é inserido em um ambiente que possui uma função de transformação que varia de acordo com uma propriedade oculta. Por exemplo, se o ambiente está particionado em sub-ambientes de aparência perceptiva idêntica para o agente, onde porém as regularidades não são equivalentes. Neste caso elementos sintéticos podem ser usados para identificar os sub-ambientes em que o agente se encontra a cada momento. Analogamente, o segundo caso (b) acontece quando o ambiente é não-estacionário de forma discreta, ou seja, quando as regularidades do ambiente sofrem uma alteração periódica definida. Neste caso, os períodos dentro dos quais o ambiente apresenta regularidades específicas podem também ser identificados por elementos sintéticos.

(538) O caso do encadeamento de ações (c) concerne àquelas situações onde o agente precisa realizar uma sequencia definida de passos para ver produzido um resultado, mas onde, no entanto, os passos intermediários não apresentam nenhum efeito aparente, e o resultado só surge depois que toda a sequencia é executada. Os elementos sintéticos podem então representar, na mente do agente, a execução dos passos intermediários.

(539) Embora a construção de elementos sintéticos não constitua uma forma de pensamento abstrato, propriamente dito, uma vez que os elementos sintéticos ainda são tratados no mesmo nível que os elementos sensoriais, é possível afirmar que tal processo é legitimamente uma forma de invenção de conceitos, pois o agente constrói novos elementos de representação que permitem designar entidades diferentes de tudo aquilo que pode ser representado à partir das percepções diretas.

(540) Assim, a possibilidade de se tratar propriedades não-observáveis representa um avanço no caminho entre a mera percepção direta, e formas mais abstratas de compreensão do mundo.

#### **5.4. Limitações e Trabalhos Futuros**

(541) A finalização dessa tese, como ocorre em muitas pesquisas desse teor, deixou mais hipóteses em aberto, mais perguntas sem resposta, mais promessas e ideias, do que propriamente conclusões determinantes ou resultados remarcáveis. Ao mesmo tempo, graças a essa ambição, é possível acreditar que o trabalho realizado constitui um passo na direção certa, ainda que bem modesto, para o futuro da pesquisa em inteligência artificial e aprendizagem de máquina.

(542) Uma das principais limitações da abordagem utilizada nesta tese é a discretização das representações, tanto dos sinais recebidos e transmitidos pelo agente, como do próprio tempo, considerado como uma sucessão de ciclos. Notoriamente, muitos problemas de mundo real só podem ser adequadamente tratados através de representações contínuas. A arquitetura CAES está descrita de maneira suficientemente geral, de modo que o modelo se aplica indiferentemente para representações discretas e contínuas. Entretanto o mecanismo de aprendizagem CALM funciona fortemente ancorado na discretização dos problemas.

(543) Assim, a necessidade de discretização é uma primeira grande limitação do CALM, sobretudo quando a ideia é que o agente enfrente problemas reais no nível sensorio-motor. Entretanto, isso não significa que o mecanismo não possa ser adaptado para universos contínuos, constituindo-se assim uma boa opção de trabalho futuro.

(544) Também, é preciso admitir que os problemas utilizados ao longo da tese, embora adequados para demonstrar as características desejadas da arquitetura CAES e do mecanismo CALM, são ainda muito simples. Seriam precisos experimentos mais elaborados para demonstrar a capacidade do CALM em descobrir propriedades não-observáveis e relevantes do ambiente, e em utilizá-las como elementos sintéticos em seus modelos de mundo, quando enfrentando problemas de fato mais complexos.

(545) Outro grande terreno, explicitamente deixado sem exploração pela necessidade de tornar viável a realização da tese, é o problema do não-determinismo. Embora, sob

nosso ponto de vista, a definição e exploração de problemas parcialmente determinísticos se tenha revelado uma contribuição interessante do trabalho, o tratamento de situações que são na prática melhor representadas por modelos estocásticos é outro caminho possível como trabalho futuro.

(546) Mais tecnicamente, certas escolhas feitas dentro dos métodos que compõem o mecanismo CALM necessitariam de mais atenção, pois é possível que haja maneiras de otimizar os algoritmos, sobretudo no que se refere ao gerenciamento da memória episódica e das árvores de antecipação e de deliberação.

(547) Finalmente, diversas escolhas feitas na definição dos modelos foram motivadas mais por praticidade do que por convicção filosófica. A realização de um trabalho desse teor, mesmo com ambições de IA Geral, exige no fim uma série de simplificações para que seja possível chegar a algum lugar. Todo o sistema de motivação do modelo, ainda que tenhamos feito um esforço para construir um agente intrinsecamente motivado, e coerente na perspectiva de uma IA corporificada, é ainda muito marcadamente utilitarista, e deixa de lado uma série de evidências, levantadas pelas teorias construtivista e interativista da inteligência, de que a motivação também se dá ligada à própria atividade. Beber água por que se tem sede é um tipo de comportamento que pode ser facilmente ancorado numa explicação utilitarista biológica, mas outros comportamentos como jogar damas, ouvir música, ou fazer uma tese, tudo isso dificilmente seria explicado numa simples alusão a necessidades fisiológicas.

(548) Outro aspecto em que o CALM inovou foi em relação às noções de curiosidade e comportamento exploratório. Enquanto nos modelos tradicionais explorar geralmente é sinônimo de fazer ações aleatórias, no CALM há uma medida de utilidade exploratória que permite ao agente planejar ações que o possam levar, em longo-prazo, a novas descobertas, ou seja, a novos conhecimentos que irão aprimorar seu modelo de mundo. Entretanto, o CALM ainda sorteia uma parcela do tempo para exploração e outra para aproveitamento, o que não é muito empolgante como forma de modelar a curiosidade.

(549) De fato, diversos processos que se observam no ser humano enquanto agente inteligente ainda estão muito longe de serem modelados de maneira convincente pela inteligência artificial, e o CALM não escapa disso. Diversos mecanismos claramente

importantes são ignorados nos nossos modelos, não por descaso, mas pela necessidade de adequar-se ao que é possível fazer no estágio atual do conhecimento, tanto sobre a inteligência de modo geral, quanto em IA. E claro, foi também necessário limitar nossas ambições a um projeto factível em um doutorado.

(550) No caso desta tese, o ponto mais marcante é que notadamente nosso esforço de tratar propriedades não-observáveis como elementos abstratos, ainda que represente um avanço importante em relação aos mecanismos tradicionais de IA, ainda está a uma distância realmente enorme daquilo que o construtivismo definiu como “pensamento simbólico”.

(551) Mesmo com todo o trabalho sério e esforçado, ainda ficamos encurralados entre dois grandes muros: de um lado, os problemas sensório-motores complexos, exigindo modelos contínuos, com capacidade de adaptação cibernética e interativa, tratamento de imprecisões, refinamento de habilidades, etc.; de outro lado, o problema da construção de símbolos que representem entidades e processos abstratos, e que levem o agente de fato a um tipo de pensamento de mais alto-nível. Esta tese não conseguiu ultrapassar nenhum dos dois muros, mas, digamos assim, está tentando ajudar a IA a colocar um pé em cada lado, pra começar uma longa escalada.

## PUBLICAÇÕES

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, 35<sup>th</sup>, 2009 Pelotas, RS, Brazil. **Proceedings...** Pelotas: UFPel, 2009.

PEROTTO, F.S.; ÁLVARES, L.O.; BUISSON, J.-C. *Um Mecanismo Constructivista para a Aprendizagem de Estrutura de MDPs Fatorados e Parcialmente Observáveis*. In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL, ENIA, 7<sup>th</sup>, 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, 1<sup>st</sup>, 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, 7<sup>th</sup>, 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, 26<sup>th</sup>, 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, 3<sup>rd</sup>, 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, 5<sup>th</sup>, 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, 5<sup>th</sup>, 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, 31<sup>st</sup>, 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, 1<sup>st</sup>, 2004, Toulouse, France. **Proceedings...** Norwell, MA: Kluwer, 2004. p.103-115. ISBN: 1-4020-8150-2.

## REFERÊNCIAS

- 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, 9<sup>th</sup>, 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, 14<sup>th</sup>, 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, 13<sup>rd</sup>, 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, 4<sup>th</sup>. **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, 1<sup>st</sup>, 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, 10<sup>th</sup>, 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**. 2<sup>ed</sup>. 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, 10<sup>th</sup>, 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, 36<sup>th</sup>, 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, 2<sup>nd</sup>, 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, 23<sup>th</sup>, 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 ON 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**. New York: 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**, Elsevier, 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.
- GUESTIN, 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.
- GUESTIN, 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, 23<sup>th</sup>, 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, 22<sup>nd</sup>, 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, 4<sup>th</sup>, 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, 12<sup>th</sup>, 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, 15<sup>th</sup>, 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, 4<sup>th</sup>, 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, 18<sup>th</sup>, 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 Biología a la Psicología**. 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, 15<sup>th</sup>, 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.; KAEHLING, L.P.; CASSANDRA, A.R. *Solving POMDPs by Searching the Space of Finite Policies.* In: INTERNATIONAL CONFERENCE ON UNCERTAINTY IN ARTIFICIAL INTELLIGENCE, UAI, 15<sup>th</sup>, 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 Réel 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 affectivity 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 Connaissance**: essai sur les relations entre les régulations organiques et les processus cognitifs. Paris: Éditions de la Pléiade, 1967.

PIAGET, J. **L'Équilibration des Structures Cognitives**: problème central du développement. Paris: PUF, 1975.

PIAGET, J. *Le Développement 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, 17<sup>th</sup>, 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, 3<sup>rd</sup>, 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, 1<sup>st</sup>, 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, 16<sup>th</sup>, 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<sup>th</sup>, 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.

SHERINGTON, 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, 23<sup>rd</sup>, 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, 20<sup>th</sup>, 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, 17<sup>th</sup>, 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**, *Inform*s, 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, 14<sup>th</sup>, 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.

WILLIAMS, 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, 25<sup>th</sup>, 2002, Boston, MA, USA. **Proceedings...** Lawrence Erlbaum, 2003. p.1305-1310.