

Integração de Emoção e Raciocínio em Agentes Inteligentes

Luís Filipe Graça Morgado

DI-FCUL

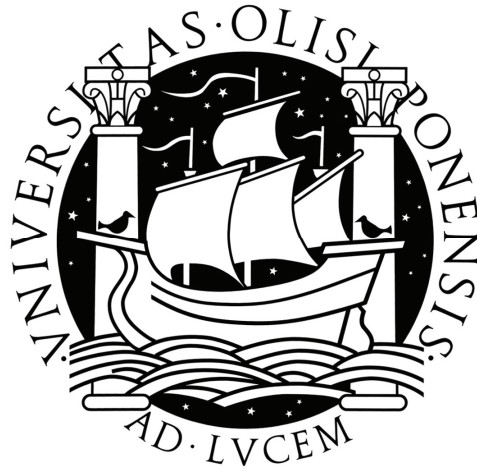
TR-06-2

Fevereiro 2006

Departamento de Informática
Faculdade de Ciências da Universidade de Lisboa
Campo Grande, 1749-016 Lisboa
Portugal

Technical reports are available at <http://www.di.fc.ul.pt/tech-reports>. The files are stored in PDF, with the report number as filename. Alternatively, reports are available by post from the above address.

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE INFORMÁTICA



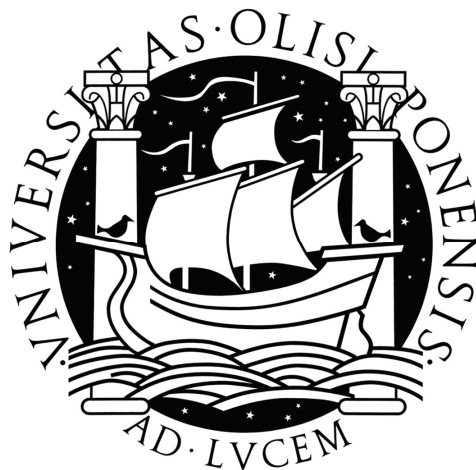
**INTEGRAÇÃO DE EMOÇÃO E RACIOCÍNIO
EM AGENTES INTELIGENTES**

Luís Filipe Graça Morgado

DOUTORAMENTO EM INFORMÁTICA

2005

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE INFORMÁTICA



INTEGRAÇÃO DE EMOÇÃO E RACIOCÍNIO EM AGENTES INTELIGENTES

Luís Filipe Graça Morgado

DOUTORAMENTO EM INFORMÁTICA

Tese orientada pela
Prof. Doutora Maria da Graça Figueiredo Rodrigues Gaspar

2005

Resumo

Esta dissertação aborda a influência que fenómenos de base emocional podem ter nos processos de raciocínio e tomada de decisão de agentes inteligentes, no sentido de tornar viável a utilização de agentes com esse tipo de capacidades cognitivas, em cenários concretos, onde é exigida uma resposta em tempo-real. São contribuições específicas da tese um modelo de emoção para suporte à modelação de fenómenos emocionais em agentes inteligentes, no qual são enfatizadas as dinâmicas subjacentes ao surgimento desses fenómenos, e um modelo de agente capaz de suportar a modelação da relação entre fenómenos emocionais e cognitivos, tendo por base o carácter dinâmico e contínuo dessa relação. Estes modelos servem de suporte à implementação de mecanismos de regulação e adaptação dos processos de raciocínio e tomada de decisão, capazes de tirar partido da relação entre fenómenos emocionais e cognitivos para focar a actividade cognitiva. Esses mecanismos permitem controlar a utilização dos recursos computacionais envolvidos na actividade cognitiva dos agentes e o tempo disponível para essa actividade cognitiva, bem como tirar partido de experiências passadas para antecipar situações futuras, através da formação de memórias emocionais autobiográficas e da exploração dessas memórias por meio de raciocínio prospectivo. É proposta uma arquitectura de agente genérica, capaz de suportar a implementação de agentes de diferentes tipos e níveis de complexidade, integrando aspectos de base emocional e cognitiva. Como suporte experimental, apresentam-se três protótipos vocacionados para contextos de aplicação distintos: contexto de dinamismo do ambiente variável; contexto com restrições sensoriais e temporais; contexto de raciocínio social.

Palavras chave: agentes inteligentes, inteligência artificial, sistemas multi-agente, modelos cognitivos, modelos de emoção, adaptação, aprendizagem, raciocínio com recursos limitados.

Abstract

This thesis addresses the influence of emotional phenomena on reasoning and decision-making processes of intelligent agents, in order to make possible the use of agents with this kind of cognitive abilities in concrete environments where real-time behavior is needed. The following are specific contributions of this thesis: an emotion model to support the modeling of emotional phenomena in intelligent agents, where the dynamics underlying those phenomena are emphasized; and an agent model able to support the modeling of the relation between emotional and cognitive phenomena based on the dynamic and continuous nature of that relation. These models support the implementation of mechanisms for regulation and adaptation of the reasoning and decision-making processes, which are able to take advantage of the relation between emotional and cognitive phenomena to focus the cognitive activity. These mechanisms enable an agent to control the computational resources used in cognitive activity and the time expended in that activity. They also enable an agent to take advantage of past experiences to anticipate future situations, through the formation of autobiographic emotional memories and through the exploration of those memories by prospective reasoning. A generic agent architecture is proposed that supports the implementation of agents of different types and levels of complexity integrating emotional and cognitive aspects. As an experimental support, three prototypes are presented, aimed at different application contexts: changing environment dynamism context; sensory and time restrictions context; social reasoning context.

Keywords: intelligent agents, artificial intelligence, multi-agent systems, cognitive models, emotion models, adaptation, learning, resource-bounded reasoning.

Agradecimentos

À Professora Doutora Graça Gaspar, por ter assumido a orientação deste trabalho, pelo estímulo, colaboração e apoio que me proporcionou, e pela disponibilidade que sempre mostrou para analisar alternativas e discutir soluções. A sua ponderação e saber muito me ajudaram a encontrar os caminhos adequados para a concretização deste trabalho.

Ao Departamento de Informática da FCUL, pelas condições proporcionadas para a realização deste trabalho. Em particular ao Professor Doutor Hélder Coelho pelo seu incentivo e apoio.

Ao Departamento de Engenharia de Electrónica e Telecomunicações e de Computadores do ISEL, pelo apoio que recebi. Agradeço, em particular, à Professora Doutora Manuela Vieira, ao Professor Doutor Hélder Pita e ao Professor Doutor Arnaldo Abrantes, pelo incentivo, ideias e sugestões que me transmitiram.

Ao Professor Doutor Walter Vieira, pela leitura crítica da tese e conselhos valiosos.

Aos colegas do DEETC do ISEL e do DI da FCUL, nomeadamente, ao Paulo Araújo, Paulo Trigo, Porfírio Filipe, Paulo Marques, Jorge Silva e Pedro Félix, pelas discussões, opiniões e sugestões que comigo partilharam.

À Doutora Conceição Libânio e aos restantes funcionários da biblioteca do ISEL, pela disponibilidade e ajuda no acesso à bibliografia necessária à investigação realizada.

À minha família, que esteve sempre e incondicionalmente a meu lado.

À Ana, pelo seu amor, sabedoria e inspiração preciosa, que tanto me ajudaram nos momentos mais difíceis do desenvolvimento deste trabalho.

Agradeço o apoio prestado a este trabalho pelo programa PRODEP III 5.3/2/2001 e pelo projecto MAGO2 (POSI/39351/SRI/2001) da Fundação para a Ciência e Tecnologia.

Índice Geral

1	INTRODUÇÃO.....	1
1.1	MOTIVAÇÃO	4
1.2	CONTRIBUIÇÕES.....	6
1.3	ORGANIZAÇÃO DA DISSERTAÇÃO	7
1.4	CONVENÇÕES DE ESCRITA	8
	PARTE I - CONTEXTO	9
2	ENQUADRAMENTO	11
2.1	AGENTES INTELIGENTES	11
2.1.1	O Conceito de Agente	12
2.1.2	Agentes Deliberativos.....	14
2.1.3	Agentes Reactivos.....	15
2.1.4	Agentes Híbridos	17
2.1.5	Inteligência e Racionalidade.....	19
2.1.6	Paradigmas de Concepção de Agentes	22
2.2	EMOÇÃO: CONCEITOS E TEORIAS	26
2.2.1	O Conceito de Emoção	27
2.2.2	Teorias da Emoção: Perspectiva Histórica.....	30
2.2.3	Teorias Cognitivas da Apreciação.....	33
2.2.4	Teorias de Controlo.....	49
2.2.5	Teoria da Interrupção.....	52
2.2.6	Emoção e Racionalidade: António Damásio.....	53
2.2.7	Emoção e Adaptação: Joseph LeDoux	55
2.3	CONCLUSÃO	57
3	TRABALHO RELACIONADO.....	59
3.1	EMOÇÃO EM AGENTES INTELIGENTES	59
3.2	MODELOS DE BASE FISIOLÓGICA	60
3.2.1	Arquitectura <i>Abbott</i>	61
3.2.2	Arquitectura <i>Cathexis</i>	64
3.2.3	Outras Abordagens de Base Fisiológica	67

3.3	MODELOS DE BASE COGNITIVA.....	68
3.3.1	Modelo <i>EMA</i>	68
3.3.2	Arquitectura <i>Salt & Pepper</i>	73
3.3.3	Arquitectura MAMID	77
3.3.4	Outras Abordagens de Base Cognitiva.....	81
3.4	MODELOS DE BASE ARQUITECTURAL.....	81
3.4.1	Arquitectura <i>AFP</i>	81
3.4.2	Modelo de Processamento Dual	84
3.4.3	Outras Abordagens de Base Arquitectural	87
3.5	CONCLUSÃO	87
	PARTE II - CONCRETIZAÇÃO	91
4	O MODELO EMOCIONAL DE FLUXO	93
4.1	INTRODUÇÃO	93
4.1.1	Sistemas Biológicos e Autopoiese	95
4.1.2	Auto-regulação, Motivação e Emoção	97
4.1.3	O Paradoxo Termodinâmico.....	99
4.2	AGENTE COMO ESTRUTURA DISSIPADORA.....	100
4.2.1	Formação de Motivações	103
4.2.2	Concretização de Motivações.....	103
4.3	A ORIGEM DA EMOÇÃO	105
4.3.1	Fases de Mudança e Padrões Emocionais	106
4.3.2	Dinâmicas Emocionais	110
4.3.3	Situação Emocional	114
4.3.4	Caracterização da Situação Emocional de um Agente	115
4.4	A EXPRESSÃO DA EMOÇÃO	120
4.4.1	Contextos Cognitivos e Expressão Emocional	121
4.4.2	Escalas Temporais e Expressão Emocional	122
4.5	CONCLUSÃO	124
5	O MODELO DE AGENTE	127
5.1	INTRODUÇÃO	127
5.2	ARQUITECTURA BASE	130
5.2.1	Estrutura Cognitiva.....	134
5.2.2	Elementos Cognitivos	134
5.2.3	Períodos de Actividade Cognitiva	139
5.2.4	Espaço Cognitivo.....	140
5.2.5	Exemplo 1: Agente Prospector	142

5.3	MECANISMOS MOTIVACIONAIS	146
5.3.1	Interacção entre Elementos Cognitivos.....	147
5.3.2	Interacção entre Motivadores e Observações.....	148
5.3.3	Exemplo 2: Agente Prospector com Motivação Explícita	150
5.4	MECANISMOS DE DISPOSIÇÃO EMOCIONAL	154
5.4.1	Descrição das Dinâmicas Emocionais no Espaço Cognitivo.....	154
5.4.2	Potenciais de Disposição Emocional e Sinais Afectivos	157
5.4.3	Mecanismo de Acumulação de Disposição Emocional.....	158
5.4.4	Exemplo 3: Agente Prospector com Mecanismos Emocionais.....	159
5.5	CONCLUSÃO	162
6	INTEGRAÇÃO DE EMOÇÃO E RACIOCÍNIO.....	165
6.1	ADAPTAÇÃO E REGULAÇÃO DA ACTIVIDADE COGNITIVA.....	166
6.2	MECANISMOS DE REGULAÇÃO COGNITIVA	168
6.2.1	Mecanismo de Integração.....	170
6.2.2	Mecanismo de Focagem de Atenção	171
6.2.3	Mecanismo de Focagem Temporal.....	173
6.2.4	Caso Experimental 1: Raciocínio Adaptativo em Ambientes Dinâmicos.....	174
6.3	MECANISMOS DE ADAPTAÇÃO DE LONGO PRAZO.....	181
6.3.1	Memórias Emocionais Autobiográficas	182
6.3.2	Mecanismos de Memória.....	183
6.3.3	Integração de Mecanismos de Memória e de Atenção.....	185
6.4	INFLUÊNCIA DE MEMÓRIAS EMOCIONAIS NO RACIOCÍNIO E TOMADA DE DECISÃO.....	186
6.4.1	Processos de Decisão.....	186
6.4.2	Nível Associativo.....	188
6.4.3	Nível Prospectivo	189
6.4.4	Caso Experimental 2: Adaptação e Decisão Guiada por Memórias Emocionais	191
6.5	CONTEXTO COGNITIVO INTER-AGENTES: EXPLORAÇÕES PRELIMINARES.....	197
6.5.1	Raciocínio Social	197
6.5.2	Caso Experimental 3: Raciocínio Social em Ambientes Incertos e Dinâmicos.....	198
6.6	CONCLUSÃO	203
7	CONCLUSÃO	205
7.1	DIRECÇÕES PARA TRABALHO FUTURO	207
7.2	CONSIDERAÇÕES FINAIS	209

APÊNDICE A - Suporte Experimental	213
APÊNDICE B - Definições Relativas à Constituição dos Elementos Cognitivos	221
BIBLIOGRAFIA.....	225

Índice de Figuras

Figura 2.1.	Arquitectura de subsunção.....	16
Figura 2.2.	Arquitecturas híbridas	18
Figura 2.3.	Representação bidimensional de termos emocionais.....	29
Figura 2.4.	Sistema límbico.....	32
Figura 2.5.	Dimensões de apreciação.....	34
Figura 2.6.	Perfis de apreciação para diferentes emoções.....	35
Figura 2.7.	Estrutura global de tipos de emoção no modelo OCC.....	38
Figura 2.8.	Processo emocional proposto por Frijda	39
Figura 2.9.	Alguns perfis de emoção propostos por Frijda	42
Figura 2.10.	O Modelo de Processo de Emoção por Componentes	44
Figura 2.11.	Uma ilustração do fenómeno de histerese na relação entre zanga e frustração	46
Figura 2.12.	Modelação do dilema fugir-atacar em cães, com base num modelo de catástrofe.....	47
Figura 2.13.	Modelo de apreciação emocional baseado num modelo de catástrofe.....	48
Figura 2.14.	Ilustração esquemática de uma malha de realimentação.....	50
Figura 2.15.	Hierarquia de malhas de realimentação	50
Figura 2.16.	Caminhos de processamento emocional envolvendo a amígdala	56
Figura 3.1.	Relações predefinidas entre estímulos e emoções.....	63
Figura 3.2.	Arquitectura <i>Cathexis</i>	64
Figura 3.3.	Arquitectura ALEC.....	67
Figura 3.4.	Arquitectura onde o modelo EMA se integra	69
Figura 3.5.	Modelo EMA	70
Figura 3.6.	Exemplo de enquadramento de apreciação	71
Figura 3.7.	Exemplo de enquadramento de “coping”	72
Figura 3.8.	Arquitectura <i>Salt & Pepper</i>	73
Figura 3.9.	Processo de geração de emoção	75
Figura 3.10.	Exemplo de regras de apreciação num nível afectivo	77
Figura 3.11.	Módulos e representações cognitivas e emocionais na arquitectura MAMID	78
Figura 3.12.	Módulo de apreciação emocional.....	79
Figura 3.13.	Influência de características individuais	80
Figura 3.14.	Arquitectura AFP	82
Figura 3.15.	Organização do modelo de processamento dual	85
Figura 3.16.	Uma implementação do modelo de processamento dual	86
Figura 4.1.	Causalidade circular numa malha de realimentação.....	96
Figura 4.2.	Agente como uma estrutura dissipadora.....	102
Figura 4.3.	Relação entre agente e ambiente.....	105
Figura 4.4.	Relação entre vários aspectos envolvidos no surgimento de fenómenos emocionais.....	110

Figura 4.5.	Vector DE como uma função de δP e de $\delta F'$	112
Figura 4.6.	Relação entre os potenciais e condutâncias que caracterizam o agente e a sua relação com o ambiente.....	116
Figura 4.7.	Variação da condutância de concretização ao longo do tempo para os dois cenários considerados	118
Figura 4.8.	Evolução da disposição emocional e da situação emocional do agente.....	118
Figura 4.9.	Evolução das componentes da situação emocional nos vários quadrantes	119
Figura 5.1.	Ilustração esquemática do desenvolvimento filogenético de diferentes componentes do comportamento	129
Figura 5.2.	Relação entre componentes do comportamento e fenómenos emocionais	129
Figura 5.3.	Representação abstracta de um agente como um conjunto de processos estruturados a diferentes níveis de organização.....	131
Figura 5.4.	Relação entre diferentes tipos de processos cognitivos.....	132
Figura 5.5.	Arquitectura genérica de agente.....	133
Figura 5.6.	Formação de elementos cognitivos como resultado de actividades de percepção.	136
Figura 5.7.	Mediadores como interface para acção concreta.	138
Figura 5.8.	Evolução dos períodos de actividade cognitiva ao longo do tempo.....	139
Figura 5.9.	Representação de um elemento cognitivo no <i>espaço cognitivo</i>	141
Figura 5.10.	Representação de uma trajectória de evolução de um elemento cognitivo ao longo do tempo.....	141
Figura 5.11.	Agente prospector simples: processos de percepção, acção e estrutura cognitiva.	143
Figura 5.12.	Detalhe do mecanismo de transdução subjacente ao processo de acção.....	144
Figura 5.13.	Comportamento do agente prospector num ambiente com três alvos.	145
Figura 5.14.	Arquitectura do agente prospector como concretização da arquitectura de agente genérica.	145
Figura 5.15.	Mecanismo motivacional como suporte à interacção entre observações e motivadores.	149
Figura 5.16.	Interacção entre uma observação e um motivador descrita no espaço cognitivo.....	149
Figura 5.17.	Motivador para minimizar os estímulos de obstáculos e maximizar os estímulos de alvos	150
Figura 5.18.	Agente prospector com motivação explícita: processos de percepção, acção e mecanismo motivacional.....	151
Figura 5.19.	Comportamento do agente para diferentes intensidades das componentes do motivador.....	152
Figura 5.20.	Decisão de acção perante resultado dos mecanismos motivacionais	153
Figura 5.21.	Arquitectura do agente prospector como concretização da arquitectura de agente genérica.	154
Figura 5.22.	Elementos participantes na concretização de um motivador num espaço cognitivo bidimensional.	155
Figura 5.23.	Movimento de uma observação em relação a um mediador no espaço cognitivo.....	156
Figura 5.24.	Caracterização afectiva no plano de disposição emocional.....	158
Figura 5.25.	Comportamento do agente num ambiente com um alvo e um obstáculo.	160
Figura 5.26.	Evolução da disposição emocional e da situação emocional do agente.....	161
Figura 5.27.	Relação entre características do modelo de agente e características dos agentes.....	163

Figura 6.1.	Mecanismos subjacentes à regulação da actividade cognitiva de um agente. ...	170
Figura 6.2.	Mecanismo de integração.	171
Figura 6.3.	Mecanismo de focagem de atenção.	172
Figura 6.4.	Ambiente <i>Tileworld</i> (protótipo implementado).....	174
Figura 6.5.	Arquitectura do agente implementado.....	176
Figura 6.6.	Eficácia dos agentes para diferentes graus de dinamismo do ambiente.....	179
Figura 6.7.	Custo total de planeamento para diferentes graus de dinamismo do ambiente.	180
Figura 6.8.	Relação entre os mecanismos que participam na formação de memórias emocionais.	182
Figura 6.9.	Integração de mecanismos de memória com mecanismos de disposição emocional, integração e de focagem de atenção.	185
Figura 6.10.	Influência da situação observada pelo agente na tomada de decisão.....	186
Figura 6.11.	Influência de recordação directa de memórias emocionais na tomada de decisão.	188
Figura 6.12.	Influência de recordação de trajectórias autobiográficas na tomada de decisão.	189
Figura 6.13.	Comportamento do agente num dos ambientes de teste.....	192
Figura 6.14.	Representação no espaço cognitivo das memórias emocionais formadas pelo agente até atingir o primeiro alvo	193
Figura 6.15.	Influência do raciocínio prospectivo nos processos de tomada de decisão.....	194
Figura 6.16.	Desempenho do agente ao longo do tempo.	196
Figura 6.17.	Exemplo de ambientes utilizados para testar o desempenho do agente	196
Figura 6.18.	Influências ambientais e cognitivas no comportamento do agente	201
Figura 6.19.	Comportamento do agente para diferentes valores do nível de atenção.....	203

Índice de Tabelas

Tabela 2.1. Relação entre variação de taxa de discrepância e afecto	51
Tabela 3.1. Variáveis fisiológicas utilizadas para definir o estado corporal de um agente	61
Tabela 3.2. Relações entre motivações para acção e impulsos (“drives”) resultantes da variação dos parâmetros fisiológicos	62
Tabela 3.3. Activação de comportamentos e respectivos efeitos fisiológicos	62
Tabela 3.4. Dimensões de classificação emocional.....	75
Tabela 3.5. Papel/função das emoções <i>vs.</i> processo mediador	76
Tabela 3.6. Dimensões de apreciação emocional e níveis de processamento	88
Tabela 4.1. Caracterização da situação do agente de acordo com o potencial de concretização e o fluxo de concretização.	106
Tabela 4.2. Comportamentos típicos de cada situação.	108
Tabela 4.3. Relação entre padrões de mudança e emoções consideradas básicas.	109
Tabela 4.4. Relação entre fenómenos emocionais e escalas temporais em que ocorrem	123
Tabela 6.1. Perfis de fiabilidade.....	199

Life is a meaningful exchange of energy

Wong Kiew Kit

Capítulo 1

Introdução

O desenvolvimento de meios auxiliares da acção humana é uma das marcas distintivas da evolução da humanidade. O desenvolvimento desses meios com características progressivamente mais sofisticadas tem estado relacionado com importantes transições qualitativas das sociedades, como é o caso do surgimento da agricultura ou da revolução industrial. Essa crescente sofisticação tem um dos seus apogeus com o surgimento dos dispositivos electrónicos, nomeadamente dos computadores digitais, no início do século XX.

Duas características principais marcam a crescente sofisticação dos meios auxiliares da acção humana ao longo do tempo: *flexibilidade* e *autonomia*. Flexibilidade traduz uma capacidade de adaptação a diferentes situações e condições de utilização. Autonomia traduz uma capacidade de acção por iniciativa própria.

A expressão mais abrangente dessas características de flexibilidade e de autonomia ocorre nos sistemas computacionais. Considerando de modo específico esse tipo de sistemas, o que se pode observar é uma evolução progressiva caracterizada por cinco tendências principais que se têm mantido ao longo do tempo [Wooldridge, 2002]:

- ubiquidade;
- interligação;
- delegação;
- humanização;
- inteligência.

Ubiquidade traduz uma crescente presença de dispositivos computacionais nos mais variados aspectos das sociedades humanas actuais. A miniaturização e a contínua redução de preços têm sido factores determinantes nessa evolução. A ubiquidade traduz-se, assim, numa distribuição massiva de dispositivos computacionais, muitos deles capazes de interagir entre si. Essa distribuição massiva é suportada por uma *interligação* alargada, incorporando diferentes infra-estruturas

de comunicação, com um carácter cada vez mais homogéneo e abrangente ao nível aplicacional, do qual a *Internet* é o exemplo mais óbvio.

Uma terceira tendência traduz-se na crescente *delegação* de actividades tipicamente humanas a sistemas computacionais. Delegação implica a transferência do controlo para o sistema computacional. Por exemplo, actualmente é comum sistemas computacionais desempenharem tarefas que envolvem um elevado risco, como é o caso da navegação aérea automática.

A quarta tendência, *humanização*, corresponde a uma evolução continuada no sentido do afastamento de perspectivas de programação e utilização centradas nos dispositivos computacionais, para uma perspectiva baseada em conceitos e metáforas próximas da nossa visão do mundo como humanos. Esta tendência é evidente na forma como a interacção com os dispositivos computacionais tem evoluído. Por exemplo, é cada vez mais comum a utilização de personagens virtuais em interfaces de utilização [Rist *et al.*, 2003].

Todas estas tendências são potenciadas por uma evolução drástica das capacidades de processamento e de memória dos dispositivos computacionais, originando uma expansão crescente dos tipos e complexidade das tarefas que é possível automatizar e delegar a sistemas computacionais. Por sua vez, este tipo de contextos computacionais e de utilização requer que os sistemas operem de modo cada vez mais independente, sem intervenção humana directa e, para além disso, que se comportem de modo a representar os nossos interesses [Wooldridge, 2002]. É neste contexto que surge a noção de *agente inteligente*, como um sistema ao qual se pode delegar, para realização autónoma, tarefas tipicamente realizadas por humanos [Russell & Norvig, 2003].

Pelo facto dos contextos de utilização de agentes inteligentes coincidirem em grande medida com os contextos da actividade humana, independentemente da sua concretização específica, quer sob a forma de sistemas robóticos, quer sob a forma de agentes virtuais, tem-se tornado claro que um aspecto central da concepção de agentes inteligentes é a capacidade de operar de modo eficaz em ambientes concretos, onde a incerteza e o dinamismo são generalizados e onde tempo e recursos são tipicamente limitados [Gigerenzer & Selten, 1999].

O projecto e implementação de agentes inteligentes capazes de comportamento eficaz neste tipo de ambientes levanta problemas importantes, relacionados quer com a capacidade adaptativa, quer com a complexidade computacional dos processos

cognitivos envolvidos, em particular de processos de raciocínio e tomada de decisão (e.g. [Picard, 1997; Russell, 1999]).

Em contrapartida, o facto dos seres humanos conseguirem lidar com este tipo de problemas de modo extremamente eficaz, resolvendo problemas que, numa perspectiva computacional, são aparentemente intratáveis, é algo que tem escapado aos modelos clássicos de raciocínio e tomada de decisão.

Até à década de 90, uma hipótese latente no âmbito da comunidade de inteligência artificial era que as capacidades humanas de resolução de problemas assentavam numa capacidade computacional muitas vezes superior à disponível nos sistemas computacionais da época. No entanto, a evolução drástica das capacidades de processamento e de memória dos dispositivos computacionais, aproximando-se do que supostamente seriam os limiares dessa capacidade computacional humana [Butazzo, 2001], não confirmou essa hipótese.

Pelo contrário, cada vez mais evidências experimentais da área das neurociências indicam que a capacidade de resolução de problemas, que caracteriza os humanos, assenta não em cálculo intensivo, mas na exploração das características e da dinâmica da relação com o ambiente, expressa, em grande medida, sob a forma de fenómenos emocionais (e.g. [Cacioppo & Gardner, 1999; Damásio, 2000; Matthews *et al.*, 2002]).

Apesar destas evidências terem ganho um carácter mais acutilante nos últimos anos, em grande medida devido ao desenvolvimento das tecnologias de imagem da actividade cerebral, já nos anos 60, Herbert Simon, ao escrever sobre as bases da cognição, enfatizou o papel que os fenómenos emocionais poderiam desempenhar no âmbito de capacidades cognitivas como raciocínio e tomada de decisão [Simon, 1967].

Em contraponto a estas evidências está a noção clássica de emoção como algo que pode perturbar o raciocínio e a tomada de decisão, com a consequente enfatização dos aspectos do comportamento humano considerados como “racionais” [Kaszniak, 2001]. O que a investigação na área das neurociências tem revelado nos últimos anos é, no entanto, um panorama diferente. Por exemplo, resultados experimentais apresentados por Damásio [2000] indicam que uma redução selectiva da emoção é, pelo menos, tão prejudicial para a racionalidade como a emoção excessiva. Por sua vez, Gray *et al.* [2002] apresentaram evidências experimentais de interacções fortes entre emoção e cognição, com perda de especialização funcional, indicando que emoção e cognição podem estar intrinsecamente ligadas. Por outro lado, a importância dos fenómenos

emocionais na aprendizagem e comportamento adaptativo está amplamente documentada (e.g. [LeDoux, 1996; LeDoux, 2000; Panksepp, 1998]).

De facto, os resultados das áreas das neurociências, ciências cognitivas e psicologia têm confirmado de modo progressivo e consistente o papel chave que os fenómenos emocionais desempenham na actividade cognitiva humana, nomeadamente no que se refere a actividades de percepção, atenção, planeamento, raciocínio, aprendizagem, memória e tomada de decisão (e.g. [Ekman & Davidson, 1994; Lane & Nadel, 2000; Rolls, 2001; Damásio, 2003]).

Esta evidência de um papel abrangente dos fenómenos emocionais na actividade cognitiva tem estimulado o desenvolvimento de modelos cognitivos que incorporam aspectos emocionais (e.g. [Botelho & Coelho, 2001; Gratch & Marsella, 2004; Hudlicka, 2004]). Mesmo no âmbito da investigação em teoria da decisão clássica, os fenómenos emocionais têm ganho relevância, tendo começado a ser incorporados em alguns modelos e teorias, como é o caso da *teoria afectiva da decisão* [Mellers, 2000], bem como de outras propostas que incorporam a emoção como um aspecto integrante dos processos de deliberação e raciocínio perante limitação de recursos (e.g. [Hanoch, 2002; Doshi & Gmytrasiewicz, 2004]).

1.1 Motivação

Do que atrás foi exposto, resulta a principal motivação desta tese: *explorar a influência que fenómenos de base emocional podem ter nos processos de raciocínio e tomada de decisão de agentes inteligentes, no sentido de tornar viável a utilização de agentes com esse tipo de capacidades cognitivas em cenários concretos, onde é exigida uma resposta em tempo-real.*

Por resposta em tempo-real entende-se não apenas a satisfação de restrições temporais *explícitas* [Laplante, 1992], mas sobretudo a satisfação das restrições temporais que resultam da relação dinâmica entre agente e ambiente, tipicamente não antecipáveis em tempo de projecto. Ou seja, os agentes não só necessitam de controlar o uso que fazem dos recursos computacionais ao seu dispor, como necessitam de adaptar esse uso às condições do ambiente, gerando eles próprios uma indicação da premência da resposta.

O trabalho aqui apresentado teve a sua origem num contexto de raciocínio social (e.g. [Conte & Castelfranchi, 1995]), tendo como objectivo principal tornar prática a

capacidade de raciocínio sobre os outros agentes, perante cenários complexos, incertos e dinâmicos.

Pelas razões anteriormente referidas, os fenómenos emocionais surgiram como um suporte promissor. Nesse sentido, procurou-se um modelo computacional de emoção que permitisse a integração com mecanismos de raciocínio social. Essa tarefa revelou-se particularmente difícil, pois os modelos de emoção encontrados apresentavam grandes obstáculos a essa integração, como é o caso do comprometimento com aspectos fisiológicos específicos, do carácter essencialmente classificativo ou do comprometimento com arquitecturas de agente específicas. Em particular, a caracterização da emoção sob a forma de etiquetas verbais (“medo”, “zanga”, etc.), mostrou-se particularmente inadequada, uma vez que não expressa o que se observou ser evidente nos fenómenos emocionais, isto é, a sua natureza intrinsecamente dinâmica e contínua (e.g. [Scherer, 2000; Carver & Scheier, 2002]).

Por outro lado, procurou-se que a integração entre aspectos emocionais e cognitivos tivesse um carácter genérico, passível de instanciação em agentes de diferentes tipos e níveis de complexidade, e não comprometido com uma arquitectura de agente específica.

A opção pela modelação do carácter dinâmico da relação entre emoção e raciocínio, simultaneamente independente do tipo e nível de complexidade dos agentes, levou a um distanciamento significativo em relação ao objectivo inicial de integração de fenómenos emocionais no contexto do raciocínio social, ganhando um âmbito mais alargado, não restrito ao raciocínio social. Esta integração de fenómenos emocionais com processos de raciocínio no contexto geral da concepção de agentes inteligentes levanta, no entanto, problemas importantes, nomeadamente: (i) a relação estreita e multifacetada entre fenómenos emocionais e cognitivos, a qual é dificilmente compatível com o adicionar de aspectos emocionais como mais um subsistema [Arzi-Gonczarowski, 2002]; (ii) a natureza dinâmica e contínua dos fenómenos emocionais, a qual é fortemente restringida pela noção clássica de um estado emocional discreto, bem como pela sua descrição sob a forma de etiquetas verbais [Scherer, 2000].

Para abordar estes problemas, procurou-se ir para além da distinção clássica entre emoção e cognição, reconhecendo a sua relação simbiótica. Isto é, a emoção é um resultado da actividade cognitiva e a actividade cognitiva é modulada pela emoção, num processo dinâmico que se desenrola ao longo do tempo de acordo com a interacção entre agente e ambiente.

A adopção desta abordagem implicou não só a definição de um modelo alternativo de emoção, como a definição de um modelo de agente capaz de suportar a implementação concreta desse modelo de emoção, bem como a sua integração com os aspectos de raciocínio e tomada de decisão. O esforço envolvido e o risco dessa opção foram consideráveis. No entanto, a decisão foi aceitar o risco e abordar o problema a um nível mais fundamental e abrangente. A tese que de seguida será apresentada é o resultado desse trabalho árduo mas compensador.

1.2 Contribuições

Da motivação atrás apresentada resulta a hipótese subjacente ao presente trabalho, segundo a qual, *os fenómenos de base emocional podem desempenhar um papel adaptativo, regulador dos processos cognitivos de um agente, fornecendo um suporte para tornar esses processos viáveis perante restrições dinâmicas de tempo e de recursos*. Da validação desta hipótese resultam as seguintes contribuições específicas:

- *Um modelo de emoção*, para suporte à modelação de fenómenos emocionais em agentes inteligentes, no qual são enfatizadas as dinâmicas subjacentes ao surgimento desses fenómenos;
- *Um modelo de agente*, capaz de suportar a modelação da relação entre fenómenos emocionais e cognitivos tendo por base o carácter dinâmico e contínuo dessa relação, bem como suportar a implementação de mecanismos de regulação e adaptação dos processos de raciocínio e tomada de decisão, capazes de tirar partido da relação entre fenómenos emocionais e cognitivos;
- *Uma arquitectura genérica de agente*, capaz de suportar a implementação de agentes de diferentes tipos e níveis de complexidade, integrando aspectos de base emocional e cognitiva;
- *Mecanismos concretos de regulação cognitiva*, capazes de tirar partido de fenómenos emocionais para focar a actividade cognitiva de um agente e controlar a utilização dos respectivos recursos computacionais;
- *Mecanismos concretos de adaptação de longo prazo*, capazes de permitir a um agente tirar partido de experiências passadas para antecipar situações futuras, através da formação de memórias emocionais autobiográficas e da exploração dessas memórias por meio de raciocínio prospectivo.

1.3 Organização da Dissertação

A dissertação está organizada em duas partes distintas. Na primeira parte é feita uma apresentação do contexto da tese, onde são descritos os principais temas envolvidos, nomeadamente a noção de agente inteligente, os conceitos e teorias da emoção, e a interligação dessas duas áreas sob a forma de modelos de agentes que integram aspectos emocionais. Na segunda parte são apresentadas as contribuições específicas da tese.

Para além deste capítulo de introdução, os temas abordados encontram-se organizados em mais seis capítulos, cujo conteúdo é o seguinte:

- **Parte I - Contexto**
 - **Capítulo 2:** capítulo de enquadramento, onde são abordados os conceitos subjacentes quer à noção de *agente inteligente*, quer à noção de *emoção*, em termos das principais teorias e modelos que lhes estão subjacentes.
 - **Capítulo 3:** capítulo onde é apresentado o trabalho relacionado, sendo discutidas diferentes abordagens de integração de aspectos de base emocional no contexto de agentes autónomos, sendo igualmente discutidos os principais modelos que estão subjacentes a essas abordagens.
- **Parte II - Concretização**
 - **Capítulo 4:** capítulo onde é apresentado o modelo de emoção proposto, nas suas diferentes vertentes.
 - **Capítulo 5:** capítulo onde são apresentados os aspectos base do modelo de agente proposto, bem como a arquitectura genérica que o suporta.
 - **Capítulo 6:** capítulo onde é abordada, de modo específico, a integração entre aspectos emocionais e de raciocínio e tomada de decisão, no contexto do modelo de agente proposto.
 - **Capítulo 7:** capítulo onde são apresentadas algumas conclusões resultantes do trabalho realizado, bem como possíveis caminhos para desenvolvimentos futuros.

No sentido de facilitar a compreensão dos conceitos apresentados, optou-se por apresentar os resultados experimentais de modo progressivo ao longo do texto, no encadeamento da apresentação dos conceitos envolvidos.

Sendo objectivo do modelo de agente proposto suportar a modelação e implementação de agentes capazes de raciocínio conceptual, cuja descrição é feita tipicamente tendo por base representações simbólicas, foi dada particular atenção à abstracção de aspectos específicos relacionados com processamento sub-simbólico, nomeadamente no que refere aos aspectos relativos a processamento de sinais. A noção de *espaço cognitivo*, apresentada no capítulo 5, é disso um exemplo.

É igualmente de realçar que não é objectivo dos modelos propostos a sua plausibilidade em termos de fenómenos biológicos ou termodinâmicos, mas tão somente tirar partido de analogias naturais no sentido de construir um suporte formal e arquitectural, capaz de suportar de modo eficaz a abordagem dos problemas anteriormente referidos.

1.4 Convenções de Escrita

No sentido de facilitar a leitura, ao longo da dissertação são utilizadas as seguintes convenções de escrita:

- as traduções de termos ou expressões originalmente em língua inglesa, na sua primeira ocorrência no texto, surgem seguidas da designação original, entre parêntesis e entre aspas;
- optou-se por não traduzir alguns termos de língua inglesa de utilização generalizada, como “hardware” ou “software”;
- em relação às siglas utilizadas, dado o seu uso generalizado na comunidade técnico-científica, são mantidas tal como aparecem no original;
- o grafismo em itálico é utilizado para destacar palavras individuais ou conjuntos de palavras no texto, como é o caso das designações de conceitos, que surgem com grafismo em itálico na primeira ocorrência que seja relevante para a sua descrição;
- surgem igualmente com grafismo em itálico as expressões latinas como *a priori*.

Parte I

Contexto

Capítulo 2

Enquadramento

Neste capítulo são apresentadas as principais áreas abordadas nesta dissertação. O capítulo encontra-se dividido em duas partes. Na primeira parte é discutido o conceito de agente inteligente, em termos de motivação, caracterização e paradigmas de concepção de agentes. Na segunda parte é discutido o conceito de emoção, em termos das principais teorias e modelos subjacentes à sua conceptualização.

2.1 Agentes Inteligentes

Até aos anos 80, a investigação em *inteligência artificial* focou-se tradicionalmente em sistemas individuais, tentando replicar nesses sistemas capacidades específicas associadas à inteligência humana, como é o caso da aprendizagem ou da resolução de problemas. No entanto, nos anos 80 começou a tornar-se claro que os aspectos da *interacção* e *distribuição* tinham de ser considerados como aspectos básicos da inteligência. Em particular, Marvin Minsky propôs que a inteligência é o resultado de inúmeros módulos interactuando entre si, cada um resolvendo tarefas primitivas específicas [Minsky, 1985].

Estas ideias ganharam uma abrangência crescente, nomeadamente quando integradas com o modelo de objectos interactivos, designados *actores*, anteriormente proposto por Carl Hewitt [1973]. Surgiram, assim, as bases para o que hoje é designado como *inteligência artificial distribuída*. Esses actores eram caracterizados por um estado interno, operando em paralelo e respondendo a mensagens de outros objectos. A evolução do conceito de actor conduziu ao que hoje é o conceito de *agente*.

Compreende-se, deste modo, que a noção de agente seja actualmente utilizada nos mais diferentes contextos. Pode ser visto numa perspectiva de um “processo de software, executado concorrentemente, que encapsula algum estado e comunica com outros agentes através de mecanismos de troca de mensagens” [Wooldridge &

Jennings, 1995], como acontece no âmbito da engenharia de software baseada em agentes, ou até como um programa com alguma forma de controlo persistente, como acontece em aplicações para a Internet. Este tipo de caracterização corresponde ao que se designa por *noção fraca de agente* [Wooldridge & Jennings, 1995].

Em contrapartida, uma *noção forte de agente* é característica de áreas como a inteligência artificial, nas quais, para além das características anteriores, conceitos como conhecimento, crenças, intenções, desejos, obrigações ou emoções, podem ser atribuídos aos agentes [Wooldridge, 2000; Schweitzer, 2003].

O termo agente é, assim, utilizado em diversas áreas com diferentes significados, não sendo por isso de estranhar que seja difícil definir a noção de agente de forma consensual [Coelho & Paiva, 1998]. Nesse sentido, existe algum consenso em que, apesar da noção de agente ser necessária, esta deve ser utilizada somente como uma ferramenta para classificar sistemas e não como uma caracterização absoluta do mundo dividido entre agentes e não-agentes [Russell & Norvig, 2003]. É, no entanto, possível identificar algumas características gerais, tipicamente subjacentes à noção de agente, tal como será abordado de seguida.

2.1.1 O Conceito de Agente

A designação *agente* resulta do facto dos agentes serem entidades capazes de acção autónoma e independente, no sentido de concretizar os objectivos para os quais foram projectados. Dito de outro modo, são capazes de decidir por eles próprios o que fazer numa determinada situação [Wooldridge, 2000; Luck *et al.*, 2002]. Das características tipicamente subjacentes à noção de agente podem-se destacar as seguintes [Jennings & Wooldridge, 1998]:

- *autonomia*: capacidade de um agente actuar sem a intervenção directa de humanos ou de outros agentes, tendo de alguma forma controlo sobre o seu estado interno e sobre as acções que realiza;
- *pró-actividade*: capacidade de um agente agir não apenas em resposta a estímulos do ambiente envolvente, mas também de forma orientada para atingir os seus objectivos, tomando a iniciativa quando apropriado;
- *reactividade*: capacidade de um agente detectar mudanças no ambiente em que está inserido e reagir atempadamente a essas mudanças;

- *sociabilidade*: capacidade de interagir com outros agentes e, eventualmente, com humanos, no sentido de atingir os seus objectivos e, se for caso disso, ajudar a atingir os objectivos de outros.

Devido à relevância que estas características têm para a caracterização do conceito de agente, vale a pena aprofundar um pouco mais o seu significado.

Por *autonomia* entende-se a capacidade do agente decidir e agir de forma independente. Esta noção de autonomia pode ser entendida num sentido ainda mais abrangente, no qual o agente tem as suas próprias crenças, motivações e intenções, as quais não são subservientes às de outros agentes [Wooldridge, 2000]. Isto não significa que estes aspectos tenham necessariamente uma representação explícita interna ao agente, apenas significa que cada agente é independente dos restantes agentes na forma como desenvolve a sua actividade.

Por *pró-actividade* entende-se a capacidade de exibir comportamento orientado por objectivos. Se um agente tem um determinado objectivo ou intenção, então o agente deverá tentar concretizar essa intenção. Para além disso, pró-actividade significa a capacidade de tirar partido do curso dos acontecimentos, em particular da ocorrência esporádica de situações favoráveis à concretização dos respectivos objectivos.

Por *reactividade* entende-se a capacidade de responder às mudanças no ambiente. Esta capacidade permite aos agentes lidar com situações que possam colocar em causa as actividades em execução, através, por exemplo, da escolha de planos de acção alternativos. A resposta às mudanças pode ser do tipo “*estímulo* → *resposta*” ou pode envolver deliberação acerca das alternativas disponíveis.

Por *sociabilidade* entende-se a capacidade de coordenar a actividade com outros agentes no sentido de atingir os objectivos pretendidos. Esta capacidade de um agente interagir com outros agentes de modo a coordenar as respectivas actividades é muito mais complexa do que a simples capacidade de trocar informação, é o caso, por exemplo, do raciocínio social.

De acordo com o tipo de características que apresentam, podem ser identificados diferentes tipos de agentes, nomeadamente, agentes *deliberativos*, *reactivos* e *híbridos* [Müller, 1997; Weiss, 2000; Wooldridge, 2002].

2.1.2 Agentes Deliberativos

Um agente deliberativo pode ser definido como um agente que contém uma representação explícita e simbólica do ambiente e (eventualmente) dos outros agentes, e no qual as decisões (por exemplo acerca de que acções realizar) são feitas através de mecanismos de inferência com base em manipulação simbólica (e.g. [Genesereth & Nilsson, 1987; Newell, 1990; Wooldridge, 2002; Russell & Norvig, 2003]). Um agente deliberativo é normalmente caracterizado recorrendo a noções cognitivas, como estados ou atitudes mentais, por analogia ao que é feito em algumas áreas da psicologia para caracterização do comportamento humano. A questão da atribuição de atitudes mentais a máquinas vem de longa data [McCarthy, 1979], estando subjacente a este tipo de descrição a ideia que, em situações onde o nível de complexidade de um agente torna difícil a sua descrição por outros processos, a utilização de noções cognitivas, como atitudes mentais, pode ser útil. Nesse sentido, diversos autores têm tentado identificar o conjunto base de atitudes que melhor permitam caracterizar um agente, como é o caso de desejos, intenções, obrigações ou crenças, as quais são normalmente agrupadas em categorias. Por exemplo em [Wooldridge & Jennings, 1994] é feito o agrupamento em duas categorias base: (i) atitudes relativas à informação que o agente possui, como crenças e conhecimento, designadas atitudes *informativas*; (ii) atitudes que, de alguma forma, guiam as acções do agente, como desejos e intenções, designadas *pró-atitudes*.

Por sua vez em [Kiss, 1996] são apresentadas três categorias de atitudes: (i) atitudes *cognitivas*, tais como crenças e conhecimento, associadas, normalmente, à informação que o agente possui; (ii) atitudes *afectivas*, tais como desejos e objectivos, associadas, normalmente, às motivações do agente; (iii) atitudes *deliberativas*, como intenções e compromissos, associadas normalmente à acção e controlo do agente. Qual a combinação de atitudes mentais mais apropriada para a caracterização de um agente, é uma questão em aberto. Por exemplo, na arquitectura ICE apresentada em [Werner, 1996], o estado mental de um agente consiste num tuplo $RA = (IA, SA, VA)$, onde *IA* designa a informação do agente acerca do mundo num determinado instante (estado *informativo*), *SA* consiste num conjunto de estratégias que guiam a acção de um agente (estado *intencional* ou estratégico), *VA* consiste numa função de avaliação sobre objectos, situações e estados (estado *valorativo*).

Outro tipo de arquitectura que apresenta três vertentes de atitudes mentais é a arquitectura BDI [Rao & Georgeff, 1991; Georgeff & Rao, 1998], uma das

arquitecturas mais divulgadas e que maior impacto teve na concepção de agentes [Haddadi & Sundermeyer, 1996; Wooldridge, 2002]. Esta arquitectura tem como base três atitudes mentais: crenças (vertente *denotativa*); desejos (vertente *volitiva*); intenções (vertente *deliberativa*). Em termos gerais, as crenças correspondem à informação que o agente tem do mundo e sobre ele próprio (a qual poderá eventualmente não ser correcta). Desejos e intenções referem-se àquilo que o agente pretende, distinguindo-se as intenções por estarem tipicamente ligadas a compromissos em relação ao comportamento futuro do agente [Bratman, 1990; Cohen & Levesque, 1990]. Assim, apesar de um agente poder ter diversos desejos, pode nunca chegar a atingi-los. Em contrapartida, se o agente tem intenção de atingir algo, esta intenção vai levar o agente a encontrar os meios apropriados para atingir o que pretende.

Diversas outras arquitecturas têm sido propostas com base neste tipo de arquitectura [Haddadi & Sundermeyer, 1996; Wooldridge, 2002]. Essas arquitecturas apresentam, normalmente, uma representação explícita de crenças, desejos e intenções, estando as intenções, em geral, relacionadas com o processo de planeamento, o qual é típico dos agentes deliberativos.

2.1.3 Agentes Reactivos

Os agentes reactivos baseiam-se num paradigma não simbólico, ou seja, não possuem explicitamente modelos simbólicos do ambiente onde actuam nem dos restantes agentes, sendo a sua construção realizada, em geral, a um nível sub-simbólico (e.g. [Arkin, 1998; Brooks, 1999; Pfeifer & Scheier, 1999; Liu, 2001]). Um agente reactivo funciona num modelo estímulo-resposta, reagindo a mudanças no ambiente ou a mensagens provenientes de outros agentes. Não tem a capacidade de manipular os seus objectivos, sendo as suas acções realizadas como resultado da activação de regras ou de planos predefinidos. Uma das principais características dos agentes reactivos tem a ver com o facto da sua estrutura ser relativamente simples e de interagirem com os restantes agentes de modo igualmente simples. Apesar disso, padrões de comportamento complexos podem emergir das interacções entre os diversos agentes, quando o comportamento dos agentes é observado numa perspectiva global. Em [Maes, 1991] são identificadas três ideias base subjacentes à realização de agentes reactivos: (i) *comportamento emergente*, ou seja, a dinâmica das interacções leva a uma complexidade emergente; (ii) *decomposição funcional*, no sentido em que

um agente é composto por um conjunto de módulos que operam de forma autónoma, onde cada um é responsável por uma tarefa específica (por exemplo sensores, controlo motor, etc.), sendo a comunicação entre os módulos mínima; (iii) *representação de baixo nível*, ou seja, tendem a operar com base em representações muito próximas dos dados sensoriais obtidos do ambiente, em contraste com as representações simbólicas de alto nível que caracterizam os agentes deliberativos.

As principais motivações para a realização de agentes reactivos têm a ver, por um lado, com a tentativa de ultrapassar os problemas relacionados com a complexidade computacional, associada a algumas vertentes dos agentes deliberativos, por outro lado, com a possibilidade de construção de agentes robustos e adaptáveis, isto é, com elevados níveis de tolerância a falhas e capazes de lidar com ambientes dinâmicos.

Das arquitecturas propostas, a arquitectura de subsunção [Brooks, 1991; Brooks, 1999] é uma das mais divulgadas. Esta arquitectura consiste num conjunto de módulos agrupados e distribuídos por níveis, sendo a sua operação assíncrona. A interacção entre os módulos é mínima, podendo os módulos de um nível mais elevado inibir os módulos de níveis inferiores. Cada módulo implementa um comportamento predefinido, como evitar obstáculos ou vaguear. A figura 2.1 apresenta a organização geral deste tipo de arquitectura.

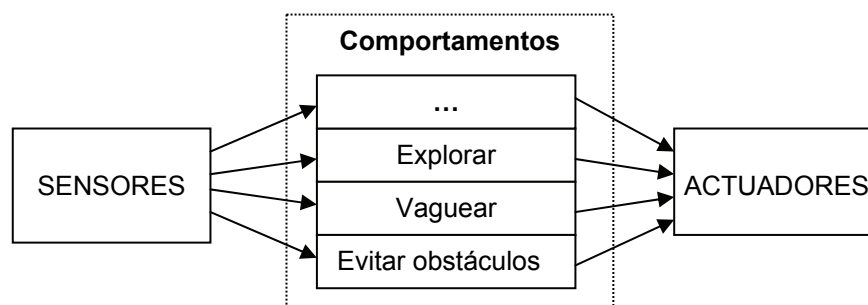


Figura 2.1. Arquitectura de subsunção.

Os comportamentos são representados como máquinas de estados finitos aumentadas (AFSM – “Augmented Finite State Machine”) [Brooks, 1991; Brooks, 1999]. Cada comportamento compete com os outros comportamentos para exercer controlo sobre o agente. Os níveis inferiores representam, normalmente, comportamentos mais primitivos, como por exemplo, evitar obstáculos.

Para além da arquitectura de subsunção, diversas outras arquitecturas relacionadas têm sido propostas, nomeadamente arquitecturas onde, em vez da

estruturação hierárquica dos comportamentos, estes são entendidos como entidades que competem para obter o controlo dos actuadores do agente (e.g. [Ferber, 1999; Arkin, 1998]), um pouco no sentido da “Society of Mind” proposta por Minsky [1985].

Apesar de terem sido obtidos alguns resultados interessantes com sistemas baseados neste tipo de arquitectura, em termos gerais apresentam algumas limitações aparentemente difíceis de ultrapassar [Jennings *et al.*, 1998; Wooldridge, 2002]:

- se um agente não possuir modelos do ambiente, deve poder ter suficiente informação disponível a partir do ambiente que lhe permita determinar as acções aceitáveis;
- uma vez que os agentes puramente reactivos tomam as suas decisões baseados em informação local (ou seja, informação acerca do seu estado actual), como pode esse mecanismo de decisão ter em conta informação não-local? Ou seja, existe inerentemente uma perspectiva de curto prazo;
- uma vez que, em sistemas reactivos, o comportamento global emerge da interacção de comportamentos individuais quando um agente é colocado num ambiente, a compreensão da relação entre os comportamentos individuais, o ambiente, e o comportamento global, pode ser difícil. Este problema pode dificultar a concepção de sistemas para realizar tarefas específicas.

2.1.4 Agentes Híbridos

Tendo cada uma destas abordagens (deliberativa e reactiva) vantagens e desvantagens, têm sido apresentadas arquitecturas alternativas que tentam conciliar estas duas abordagens de forma a tirar partido dos pontos fortes de cada uma. Essas arquitecturas são, normalmente, designadas arquitecturas híbridas. Uma arquitectura híbrida incorpora aspectos deliberativos e aspectos reactivos. As arquitecturas deste tipo são, normalmente, organizadas em camadas de forma hierárquica, lidando cada camada com informação referente a diferentes níveis de abstracção. Essas camadas podem estar organizadas horizontalmente ou verticalmente, tal como ilustrado na figura 2.2.

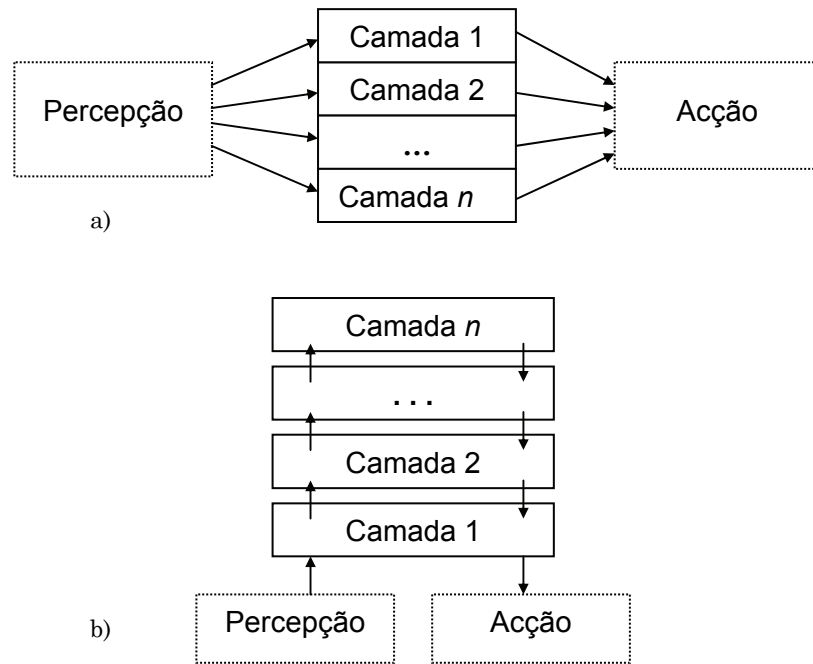


Figura 2.2. Arquitecturas híbridas: (a) organização horizontal; (b) organização vertical.

Numa arquitectura com organização vertical apenas uma camada interage com o ambiente, ao passo que numa arquitectura com organização horizontal todas as camadas interagem com o ambiente.

Diversas arquitecturas híbridas têm sido propostas (e.g. [Arkin, 1998; Pfeifer & Scheier, 1999]), como é o caso da arquitectura *Touring Machines* [Ferguson, 1992; Ferguson, 1995], caracterizada por uma organização horizontal, ou da arquitectura *InterRRap* [Müller, 1996; Müller, 1997], caracterizada por uma organização vertical.

A arquitectura *Touring Machines* consiste num conjunto de subsistemas de percepção e de acção, que realizam a interface com o ambiente, e três níveis de controlo que servem de mediadores entre os módulos de percepção e os módulos de acção. Cada nível de controlo é independente dos restantes, operando os diversos níveis de forma concorrente. O nível reactivo gera potenciais sequências de acção, em resposta a eventos para os quais o tempo de resposta deve ser bastante curto, não sendo por isso passíveis de processamento por níveis superiores. O nível de planeamento elabora planos e selecciona as acções adequadas para atingir os objectivos do agente. O nível de modelação contém representações simbólicas do estado cognitivo de outros agentes da sociedade, as quais são utilizadas para identificar possíveis conflitos na concretização dos objectivos do agente, em relação a outros agentes da sociedade. Estes três níveis podem comunicar entre si, estando organizados numa estrutura de controlo que realiza a mediação entre eles,

nomeadamente em termos de resolução de conflitos resultantes das acções propostas pelos diversos níveis.

As arquitecturas com organização vertical apresentam pelo menos três camadas: uma camada tipicamente reactiva, caracterizada por comportamentos primitivos; uma camada intermédia, que realiza a análise da informação sensorial, operando já a um nível simbólico; e uma camada puramente simbólica, normalmente associada a aspectos de coordenação e interacção social.

Como exemplo, a arquitectura *InteRRaP* é composta por um conjunto de camadas organizadas de forma vertical, onde cada camada corresponde a um nível de abstracção superior ao da camada anterior. Cada camada é, por sua vez, dividida em duas unidades distintas, uma contendo as bases de conhecimento do agente a esse nível e outra contendo os diversos elementos de controlo, os quais interagem com as bases de conhecimento do nível respectivo. O nível inferior, designado nível de interface com o ambiente, realiza a gestão da interface entre o agente e o ambiente, lidando com os aspectos da percepção, acção e comunicação. O nível seguinte, designado nível comportamental, é responsável pelos comportamentos reactivos do agente, seguido do nível de planeamento que elabora planos e selecciona as acções adequadas para atingir os objectivos do agente. Por fim, existe o nível de cooperação que contempla os aspectos sociais.

2.1.5 Inteligência e Racionalidade

A noção de inteligência é reconhecidamente difícil de definir, podendo envolver quer uma caracterização subjectiva de fenómenos mentais, quer uma classificação de comportamentos observáveis, também potencialmente subjectiva, uma vez que depende da interpretação de um observador externo [Pfeifer & Scheier, 1999; Russell & Norvig, 2003]. No entanto, para que se possa falar de sistemas inteligentes e, em particular, de agentes inteligentes, é necessário ter uma noção concreta e utilizável do que significa “inteligente”. Para além disso, essa definição deve ser directamente relacionável com o contexto de operação do sistema e com o comportamento observável por ele produzido.

No contexto da concepção de agentes capazes de comportamento autónomo, o conceito de inteligência está directamente relacionado com a capacidade de produzir um comportamento adequado [Russell, 1999; Wooldridge, 2002], ou seja, a capacidade de agir no sentido de conseguir o melhor resultado possível perante os objectivos que

se pretende atingir, a que corresponde a noção de *racionalidade* [Russell & Norvig, 2003]. Consoante o contexto de operação de um agente, quer em termos de características do ambiente, quer em termos de características internas ao agente, é possível caracterizar diferentes tipos de racionalidade (e.g. [Simon, 1967; Russell, 1999; Gigerenzer, 2000]):

- *racionalidade perfeita* (“*perfect rationality*”), ou a capacidade de gerar o melhor comportamento possível dada a informação disponível;
- *racionalidade de cálculo* (“*calculative rationality*”), ou a capacidade de, em princípio, calcular a decisão racional perfeita, dada a informação disponível no início do cálculo;
- *racionalidade de meta-cálculo* (“*meta-level rationality*”), ou a capacidade de seleccionar a combinação óptima de sequência de computação e acção, sob a restrição de a acção dever ser seleccionada pela computação;
- *optimalidade confinada* (“*bounded optimality*”), ou a capacidade de gerar o melhor comportamento possível dada a informação e os recursos computacionais disponíveis;
- *racionalidade confinada* (“*bounded rationality*”), ou a capacidade de gerar um comportamento suficientemente bom dada a informação e os recursos computacionais disponíveis.

A *racionalidade perfeita* surge como a caracterização ideal para um agente inteligente. No entanto, tal como salientado por Simon [1967; 1981], agentes perfeitamente racionais não existem na realidade. Qualquer sistema físico necessita de recursos e de tempo para realizar processamento e determinar a acção a realizar. Deste modo, o comportamento de agentes reais não pode reflectir de modo instantâneo as mudanças que ocorram no ambiente, pelo que, no caso geral, apresentará um comportamento sub-óptimo.

No caso da *racionalidade de cálculo*, a qualidade do comportamento está dependente da capacidade computacional do agente. Um agente com racionalidade de cálculo poderá (caso o cálculo termine) obter o que seria a escolha racional no início do cálculo. No entanto, a resposta certa no instante errado não tem qualquer valor em ambientes reais. No limite, uma velocidade de processamento infinita resultaria em

comportamento óptimo. Na prática, não é possível evitar o problema da complexidade computacional da generalidade dos processos inerentes ao comportamento inteligente. A resposta comum tem sido a de utilizar diferentes técnicas para acelerar o processamento ou aproximar a solução, na tentativa de obter um comportamento razoável. Apesar de alguns desses métodos poderem manter a garantia de comportamento óptimo e serem eficazes para problemas de âmbito limitado e bem estruturado, no caso geral é impossível garantir o comportamento racional dos agentes [Russell, 1999].

A *racionalidade de meta-cálculo* surge como uma possível alternativa para lidar com o problema da complexidade computacional, atrás referido. Neste tipo de racionalidade os processos computacionais do agente são, eles próprios, controlados por processos computacionais de um nível superior (nível *meta*). Na prática, o processamento num nível *meta* consome recursos computacionais, logo não pode ser realizado de modo óptimo. Desta forma, apenas se está a transferir os problemas inerentes à racionalidade de cálculo para um nível superior.

As dificuldades com a racionalidade perfeita ou a racionalidade de cálculo e meta-cálculo resultam em grande medida de restrições que são impostas sobre as condições em que o agente deve operar, as quais não são directamente controláveis durante a concepção dos agentes. Nesse sentido, Simon [1967] propõe a noção de *racionalidade confinada*, segundo a qual, perante limitações de recursos e de tempo, um agente deve tentar obter uma solução suficientemente boa, ou seja, que *satisfaça* o seu nível de aspiração em relação à qualidade da solução, em vez de tentar encontrar a solução óptima.

A noção de *optimalidade confinada*, proposta por Russell [1999], surge como uma alternativa à *racionalidade confinada*, sendo defendido que um agente deve tentar obter a melhor solução possível dados os recursos computacionais disponíveis. Nesse sentido, a *optimalidade confinada* pressupõe a capacidade de um agente ter disponível em tempo de execução uma gama variada de recursos e opções computacionais de modo a poder calcular diferentes soluções, dadas as condições correntes, e optar pela melhor solução possível. Na prática, a optimalidade confinada levanta problemas importantes, nomeadamente, a necessidade de muito conhecimento *a priori* para permitir estimativas fiáveis de custos e benefícios, e o facto dos próprios cálculos de custo/benefício terem um custo, o que leva a um problema de *regressão infinita*

[Gigerenzer & Selten, 1999], pois esses próprios cálculos necessitam de análise de custo/benefício.

A racionalidade confinada é, por isso, aquela que melhor caracteriza a actividade de um agente em cenários concretos, onde tempo e recursos são limitados (e.g. [Simon, 1981; Gigerenzer, 2000]).

No entanto, é de notar que não é especificada a forma como deve ser feita a selecção entre as opções computacionais ou de utilização de recursos disponíveis em tempo de execução, pelo que esse é um problema central em aberto. É precisamente esse um dos problemas que a presente tese aborda.

2.1.6 Paradigmas de Concepção de Agentes

O modo concreto de concepção de agentes inteligentes está directamente dependente do tipo de agentes que se pretende implementar. A distinção principal é feita entre agentes deliberativos (secção 2.1.2) e agentes reactivos (secção 2.1.3). Os agentes deliberativos são caracterizados pela existência de um modelo interno e pela capacidade de explorar esse modelo, no sentido de obter conclusões acerca de possíveis acções a realizar [Wooldridge, 2002]. Os agentes reactivos são caracterizados pela ausência desse modelo interno, pelo menos de forma explícita, sendo as suas acções determinadas pela reacção aos estímulos provenientes do ambiente, sem recurso a qualquer representação interna [Brooks, 1999].

Esta distinção tem, no entanto, implicações mais amplas. Os agentes deliberativos enquadram-se no que é designado de *paradigma cognitivo* da inteligência artificial (e.g. [Newell, 1990; Polk & Seifert, 2002; Matlin, 2003]). Em contrapartida, os agentes reactivos enquadram-se no que é designado de *paradigma comportamental* (ou *reactivo*) da inteligência artificial (e.g. [Pfeifer & Scheier, 1999; Brooks, 1999; Murphy, 2000]).

2.1.6.1 Paradigma Cognitivo

O paradigma cognitivo é caracterizado por uma perspectiva da cognição como um processo computacional que opera sobre representações simbólicas. Deste modo, a actividade interna dos agentes é caracterizada numa perspectiva de processamento de informação, sendo os agentes descritos em termos de objectivos e de conhecimento acerca dos meios para atingir os objectivos. As acções dos agentes derivam de processos de raciocínio e planeamento que operam sobre as representações internas.

Esta perspectiva de processamento de informação tem associada a noção de modularidade. Assim, os agentes são organizados em módulos funcionais bem definidos e estanques, por exemplo, percepção, aprendizagem, planeamento, memória, entre outros.

A noção de racionalidade desempenha um papel central neste paradigma. Deste modo, se um agente tem um objectivo e sabe que uma determinada acção o vai aproximar do objectivo mais que qualquer outra, o agente irá escolher essa acção.

Em termos de tipo de arquitectura de agente, o paradigma cognitivo é caracterizado por uma perspectiva de processamento central, na qual a informação proveniente das interfaces com o ambiente é integrada numa estrutura de representação central, tipicamente uma memória de curto prazo, a qual posteriormente suporta a formação de memórias de longo prazo. Subjacente a esta arquitectura, está um padrão de fluxo de controlo caracterizado por um ciclo sequencial *percepção-planeamento-acção* (“*sense-plan-act*”). Ou seja, o ambiente é percebido, resultando representações internas que são, de seguida, processadas através de processos de raciocínio e planeamento, dando origem à formação de planos que determinam as acções que o agente irá executar de seguida.

No paradigma cognitivo, a concepção de agentes é vista como um processo do geral para o específico (“top-down”). Ou seja, é especificado o conhecimento que o agente tem *a priori* sobre o que deve fazer e sobre os meios para o fazer (*nível de conhecimento* [Newell, 1982]), sendo essa descrição posteriormente convertida em aspectos concretos de implementação.

2.1.6.2 Paradigma Comportamental

O paradigma cognitivo tem-se mostrado pouco apto para lidar com a necessidade de comportamento em tempo-real, a par da necessidade de lidar com o dinamismo e incerteza dos ambientes reais.

Em grande medida motivada pela incapacidade de abordar este tipo de problemas no contexto do paradigma cognitivo, no início dos anos 90 consolidou-se o que é actualmente o paradigma comportamental. Este paradigma tem uma forte inspiração na biologia e etologia, em particular no que se refere a organismos menos complexos, como é o caso dos insectos, em contraponto à inspiração da psicologia e comportamento humano que caracteriza o paradigma cognitivo.

A característica fundamental do paradigma comportamental é que todas as acções do agente são resultado de *comportamentos* que realizam um mapeamento directo

entre as entradas sensoriais e os mecanismos de acção (ver secção 2.1.3). A noção de comportamento é o elemento central da caracterização de um agente. Esses comportamentos são implementados como unidades computacionais independentes que operam de modo concorrente. O comportamento global é emergente, não existindo um módulo central que determine que comportamentos activar ou qual a acção a realizar. Contudo, podem existir mecanismos de coordenação, apesar desse controlo ser descentralizado.

A utilização de representações explícitas de conhecimento é evitada, mesmo sendo referentes a um comportamento específico. Isto é, apenas informação sensorial local é utilizada. Deste modo, o paradigma comportamental rompe com o ciclo de *percepção-planeamento-acção* que caracteriza o paradigma cognitivo, enfatizando antes um acoplamento directo entre percepção e acção. De igual modo, a noção de modelo interno, fundamental no paradigma cognitivo, é no paradigma comportamental substituída pela noção de “o mundo é o modelo” [Brooks, 1999], resultante da ausência de representações internas explícitas.

No paradigma comportamental, a concepção de agentes é vista como um processo do específico para o geral (“bottom-up”) sendo especificados em primeiro lugar comportamentos particulares e simples, evoluindo-se progressivamente para comportamentos de carácter mais genérico. Neste tipo de concepção, a coordenação e controlo da interacção dos comportamentos é um problema central.

A capacidade de resposta em tempo-real, mesmo em cenários complexos, típicos de ambientes reais, é um dos pontos fortes do paradigma comportamental. No entanto, a dificuldade em conceber agentes capazes de comportamentos complexos, nomeadamente capazes de antecipar situações futuras e, de modo sistemático, determinarem cursos de acção para com elas lidar, limita a sua aplicabilidade. A principal resposta a este problema tem sido a concepção de agentes híbridos (ver secção 2.1.4). Essa abordagem tenta conciliar as capacidades deliberativas inerentes ao paradigma cognitivo, com a capacidade de resposta em tempo-real típica do paradigma comportamental. No entanto, esse tipo de abordagem levanta outros tipos de problemas, como é o caso de desadaptação de representações e de escalas temporais de operação distintas, que têm como resultado arquitecturas fragmentadas entre subsistemas reactivos e deliberativos.

Em particular, as abordagens híbridas não lidam com o problema da complexidade computacional inerente ao paradigma cognitivo. Em vez disso, contornam esse

problema através de mecanismos capazes de respostas rápidas em situações específicas, mas inoperantes em situações onde as capacidades prospectivas e de antecipação são necessárias.

Iremos ver que os mecanismos de base emocional podem desempenhar um papel chave quer em termos de resposta para lidar com a complexidade computacional inerente ao paradigma cognitivo, quer como elo de ligação entre estes dois paradigmas, aparentemente tão afastados conceptualmente. Para isso, de seguida será feita uma introdução aos conceitos e teorias subjacentes à modelação de fenómenos emocionais.

2.2 Emoção: Conceitos e Teorias

Desde tempos imemoriais que o ser humano se tem questionado sobre a natureza da emoção. Na tradição clássica da antiga Grécia, Platão pareceu desvalorizar a emoção, considerando-a como algo que se intromete e interfere com a razão humana. Em contrapartida, Aristóteles via as emoções como facetas mais interessantes da existência humana, considerando-as o resultado da combinação da vida cognitiva superior e da vida sensual inferior [Strongman, 1996].

Esta visão abrangente de Aristóteles antecede em muito o que as modernas ciências cognitivas reconhecem, que é a relação estreita entre emoção e cognição. Contudo, foi a visão de Platão que em grande medida prevaleceu na tradição filosófica e científica ocidental, sendo a emoção tradicionalmente encarada como impedimento à racionalidade.

Estudos teóricos e experimentais importantes têm vindo a alterar esta visão de modo consistente. Particularmente relevantes nesta mudança de atitude foram os trabalhos realizados por António Damásio [1994; 2000; 2003] e por Joseph LeDoux [1996; 2000], os quais mostraram de forma clara o papel central que os fenómenos emocionais desempenham em múltiplos aspectos da cognição, em particular, no raciocínio, tomada de decisão e na aprendizagem. É esta estreita relação entre emoção e cognição que será explorada nesta secção, no sentido de motivar e enquadrar a abordagem proposta. Para isso, começar-se-á por apresentar uma perspectiva sintética das principais vertentes actuais em teorias da emoção no âmbito das ciências cognitivas. Pela abrangência do tema, são focadas de modo específico teorias que influenciaram o trabalho realizado ou que permitam contrastar diferentes abordagens nos seus aspectos principais, não se pretendendo, por isso, uma discussão exaustiva do tema.

É com este enquadramento que será abordado no próximo capítulo o problema da emoção em agentes autónomos, onde será feita a análise e discussão de alguns dos principais modelos que têm sido propostos. Como iremos ver, existem problemas importantes em aberto que levaram à construção de um modelo alternativo, o qual será apresentado no capítulo 4.

2.2.1 O Conceito de Emoção

Apesar da experiência da emoção ser um dos fenómenos mais essenciais da existência humana, tem-se mostrado extremamente difícil de descrever em termos científicos. Um problema central é a sua natureza subjectiva. Apesar das emoções terem correspondências físicas, como é o caso de expressões comportamentais e fisiológicas ou em termos de actividade cerebral, os fenómenos emocionais são essencialmente descritos pelas etiquetas que lhes são atribuídas em termos de experiência consciente: sentimentos de alegria e de tristeza, de zanga e de medo, etc. [Matthews *et al.*, 2002]. Como resultado, as teorias propostas para explicação dos fenómenos emocionais divergem consideravelmente. Assim, a emoção tem sido relacionada com os mais variados aspectos, nomeadamente, com sistemas cerebrais específicos, com dimensões de apreciação subjectiva ou com mecanismos de processamento de informação para auto-regulação, entre muitos outros.

Também a relação entre emoção e noções psicológicas clássicas como motivação e cognição é alvo de debate. Tradicionalmente, emoção, motivação e cognição formam um suporte tripartido de classificação dos fenómenos mentais (e.g. [Matthews *et al.*, 2002]). Nessa perspectiva, a emoção representa um sistema distinto, separado da motivação e da cognição. Em contrapartida, para outros investigadores a emoção é descrita como resultante da integração de diferentes aspectos. Por exemplo, para Lazarus [1991] as emoções combinam processos motivacionais, cognitivos e fisiológicos, num estado complexo e único que envolve vários níveis de análise.

A natureza subjectiva dos fenómenos emocionais torna igualmente difícil a sua classificação e caracterização. Por exemplo, é normal fazer-se a distinção entre emoção e estado de espírito (“mood”), sendo as emoções consideradas transientes e relacionadas com eventos ou situações específicas, ao passo que os estados de espírito são mais prolongados no tempo e não possuem objectos específicos [Scherer, 2001]. No entanto, não é clara a fronteira entre estas duas noções, nem sequer se são, ou não, diferentes facetas do mesmo fenómeno. Para lidar com este tipo de problemas, alguns investigadores utilizam o termo afecto num sentido abrangente, no qual incluem as noções de emoção e de estado de espírito [Matthews *et al.*, 2002]. Em contrapartida, outros investigadores utilizam a noção de afecto para se referir, especificamente, ao carácter agradável ou desagradável da experiência subjectiva associada aos fenómenos emocionais [Scherer, 2001].

Resultam, assim, múltiplas concepções dos fenómenos emocionais, desde uma propriedade cognitiva de alto nível (e.g. [Ortony *et al.*, 1988; Lazarus, 1991]) a um processo dinâmico intrinsecamente ligado à fisiologia dos organismos (e.g. [Damásio, 1994; Panksepp, 1998]). Disto resulta, também, a existência de múltiplas definições de emoção. Existe no entanto um consenso crescente acerca de três vertentes constitutivas da expressão emocional [Scherer, 2001]:

- *resposta fisiológica*: corresponde a alterações em diversos sistemas orgânicos, produzindo sensações como aceleração do ritmo cardíaco ou tremores;
- *expressão comportamental*: corresponde a alterações na actividade motora e na expressão corporal, como alteração da expressão facial ou do tom de voz;
- *experiência subjectiva*: corresponde à percepção consciente da experiência emocional, tipicamente caracterizada através de termos emocionais como alegria, zanga ou medo.

São estas três vertentes que servem de base à caracterização emocional em geral. Uma das formas de caracterização da emoção é através da identificação de dimensões de classificação emocional. A principal dimensão de classificação emocional está relacionada com a experiência subjectiva de prazer ou de dor, agradável ou desagradável ou, como é designada normalmente no âmbito da psicologia, valência positiva ou negativa, a que corresponde a noção de afecto (positivo ou negativo), anteriormente referida [Scherer, 2001].

Outras dimensões de classificação emocional têm sido propostas, como é o caso de excitação *vs.* depressão, ou tensão *vs.* relaxamento (e.g. [Plutchik, 1980]), tendo sido realizado um grande número de estudos para identificar dimensões relevantes. Esses estudos têm identificado de modo consistente duas dimensões: valência (positiva *vs.* negativa) e excitação *vs.* depressão, mais recentemente descrita como activação *vs.* passividade. Para além de identificarem estas duas dimensões, esses estudos têm identificado, igualmente, um conjunto de termos emocionais, localizados em regiões específicas do espaço bidimensional formado pelas duas dimensões anteriores, tal como ilustrado na figura seguinte.

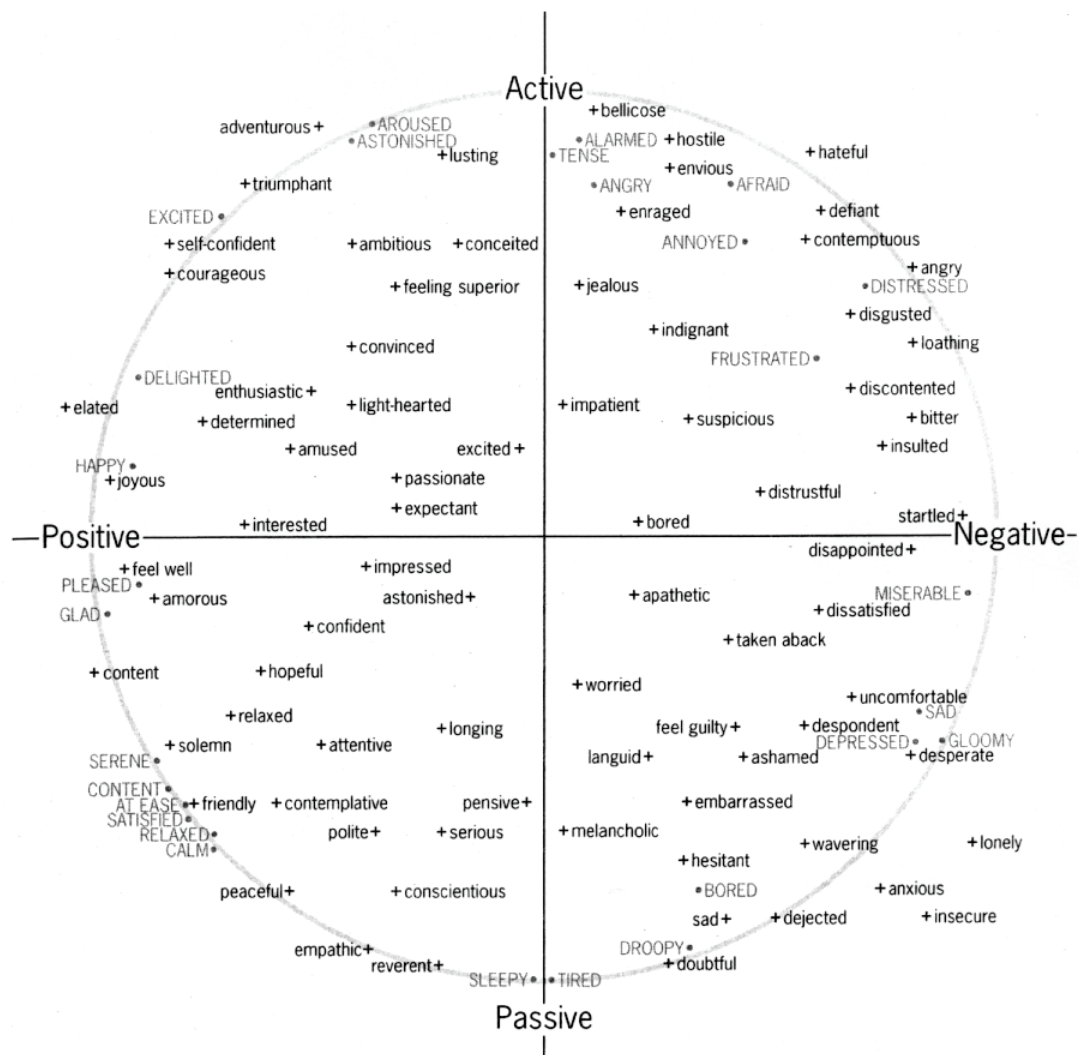


Figura 2.3. Representação bidimensional de termos emocionais - dimensão vertical: activação (activo/passivo); dimensão horizontal: valência (positiva/negativa) [Scherer, 2001].

Em termos fisiológicos e comportamentais, a dimensão activação (activo *vs.* passivo) é também relacionada com a noção de mobilização/alerta (“arousal”), a qual é normalmente utilizada para descrever o estado de mobilização do organismo, desde estados de sono profundo, estados de mobilização e elevada atenção, até estados de grande excitação. No entanto, alguns investigadores fazem a distinção entre mobilização (“arousal”) e activação (e.g. [Frijda, 1986]).

É importante notar que o facto de ser possível fazer uma projecção de termos emocionais numa representação bidimensional ou tridimensional, é sempre uma simplificação. De facto, o número enorme de termos e expressões emocionais existentes em diferentes línguas e culturas (e.g. [Ekman & Davidson, 1994]) indica claramente a existência de uma diferenciação muito mais subtil e, possivelmente, mais adequada à caracterização dos fenómenos emocionais.

2.2.2 Teorias da Emoção: Perspectiva Histórica

Na tradição moderna de estudo da emoção, um dos primeiros passos foi dado por Charles Darwin. No seu livro *“On the Expression of the Emotions in Man and Animals”* (1872), Darwin propõe que os humanos expressam as suas emoções através de mecanismos similares aos presentes noutras formas de vida ancestrais e exemplificados nos vertebrados actuais, nomeadamente nos mamíferos [Shepherd, 1994]. Darwin proporcionou, assim, uma motivação para a exploração das bases neurobiológicas da emoção. No entanto, as técnicas para a realização destes estudos só surgiriam no início do século XX.

Entretanto, na perspectiva prevalecente à altura, as emoções eram simplesmente uma classe de sentimentos, diferenciados da sensação corporal pela sua qualidade [Shepherd, 1994]. Em 1884, William James propôs uma perspectiva diferente, na qual as emoções resultam da percepção subjectiva das modificações corporais que ocorrem durante episódios emocionais (no que viria mais tarde a ser conhecida como teoria James-Lange, pois o fisiologista dinamarquês Carl Lange propôs a mesma ideia cerca de um ano mais tarde).

Segundo James, quando percebemos que estamos em perigo, um conjunto de respostas corporais ocorrem. É a percepção subjectiva dessas respostas que constitui, por exemplo, o sentimento de medo. Deste modo, a teoria de James-Lange tinha como premissa base o facto das emoções serem acompanhadas de reacções físicas, como a aceleração do ritmo cardíaco, ou o aperto no estômago. Seria essa realimentação (“feedback”) fisiológica que daria a cada emoção a sua qualidade singular. A reacção física seria, assim, crucial na génese da experiência emocional.

Um importante problema com esta teoria é que ela não fornece uma explicação adequada para a diferenciação das emoções. Este argumento foi apresentado por Walter Cannon em 1927. Enquanto para James o que distingue as emoções é o facto de elas envolverem a percepção de um conjunto único de alterações corporais, Cannon argumentou, com base em evidências experimentais, que reacções somáticas características de diferentes emoções, como é o caso do medo e da zanga, são idênticas, logo não nos permitem diferenciar esses estados emocionais.

Cannon contrapôs, ainda, o facto de as reacções do sistema nervoso autónomo serem demasiado lentas para poderem preceder o surgimento da emoção, a qual já é sentida no momento em que essas reacções têm lugar. Em contrapartida, propõe uma

perspectiva alternativa em que circuitos cerebrais específicos são considerados essenciais para a geração de emoções [Panksepp, 1998].

Apesar de Canon ter apresentado evidências experimentais convincentes para a época, os resultados experimentais actuais apontam para uma relação profunda e abrangente entre as mudanças associadas a múltiplos sistemas fisiológicos, que ocorrem durante os episódios emocionais, e a actividade do sistema nervoso central. É o caso, por exemplo, da interacção entre processos cerebrais e imunológicos, onde muitas das moléculas associadas a diferentes aspectos imunológicos têm efeitos fortes e directos sobre as regiões cerebrais associadas à emoção [Panksepp, 1998].

A proposta de Canon foi, contudo, um ponto de partida importante para o estudo e reconhecimento da neurofisiologia da emoção. Dois marcos importantes desse estudo são os trabalhos realizados por James Papez [1937] e por Paul MacLean [1952].

Segundo Papez, a relevância da emoção é tão grande que os seus mecanismos, quaisquer que sejam, devem ter uma base estrutural. Nesse sentido, propõe um conjunto de regiões cerebrais, interligadas, como suporte geral dos fenómenos emocionais, as quais viriam a ser conhecidas como *circuito de Papez*. Para Papez, a emoção não é função de nenhum centro cerebral específico, mas antes de um circuito que envolve várias estruturas interligadas por múltiplas redes nervosas. De entre as estruturas nervosas identificadas encontra-se o hipotálamo, actualmente reconhecido como um dos principais centros de regulação endócrina e homeostática, e o hipocampo, actualmente considerada uma das regiões chave na formação de memória, apesar dos processos envolvidos serem em larga medida desconhecidos. Este circuito, actuando de modo harmonizado, seria o responsável pelas funções associadas à emoção, bem como pela sua expressão somática [Rolls, 2001].

Mais tarde, Paul MacLean [1952], desenvolveu o trabalho de Papez, nomeadamente incluindo novas estruturas no circuito como é o caso do córtex prefrontal e de importantes agrupamentos subcorticais, como é o caso da amígdala. Para designar esse circuito alargado, MacLean propôs a designação de *sistema límbico*, tendo ajudado a estabelecer o conceito de sistema límbico como a principal estrutura cerebral associada aos fenómenos emocionais.

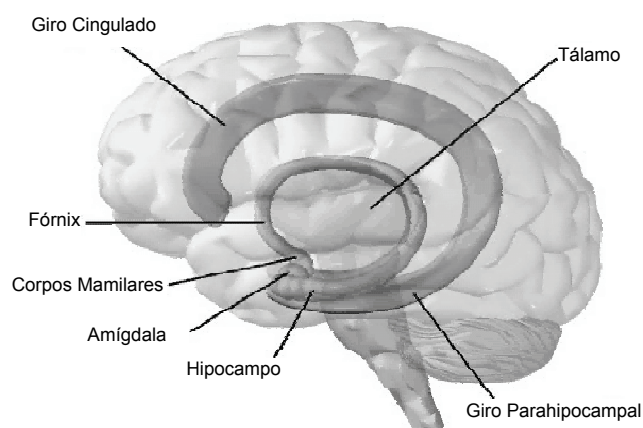


Figura 2.4. Sistema límbico (adaptado de [Stewart, 2001]).

As evidências da relevância dos processos cerebrais nos fenómenos emocionais, tal como descrito por Cannon, Papez ou MacLean, a par com a “revolução” cognitiva surgida no início dos anos 60 (e.g. [Oatley & Jenkins, 1996]), deram uma nova orientação ao estudo dos fenómenos emocionais. Um dos marcos dessa mudança foi um estudo realizado por Stanley Schachter e Jerome Singer (1962) no qual se pretendia estudar a influência de aspectos cognitivos na emoção. Nesse estudo, sujeitos injectados com epinefrina, uma substância estimulante do sistema nervoso autónomo, foram colocados em duas situações distintas. Alguns foram colocados numa sala onde um actor fez um papel de zangado, enquanto outros foram colocados numa sala onde outro actor fez um papel alegre/eufórico. Em ambos os casos o estado emocional dos sujeitos testados tendeu a seguir o estado emocional expresso pelo actor. A conclusão que Schachter e Singer retiraram desta experiência foi que, apesar de algumas formas de resposta somática serem directamente discrimináveis em termos de estados emocionais, no caso geral não é possível fazer a distinção entre estados emocionais a partir de observações fisiológicas. Ou seja, a diferenciação das emoções não é fisiológica mas sim cognitiva [Oatley & Jenkins, 1996].

O trabalho de Schachter e Singer resultou numa teoria onde se propõe que é a resposta somática em conjunto com uma “avaliação” cognitiva que está na base da experiência subjectiva da emoção. Esta abordagem cognitiva da emoção baseia-se numa perspectiva computacional dos processos mentais. Nessa perspectiva é possível distinguir entre o substrato neuronal, o suporte físico (*hardware*) e as estruturas simbólicas que determinam o processamento de informação que ocorre nesse substrato, isto é, o suporte lógico (*software*). É aceite que, em última análise, todo o processamento de informação envolve o suporte físico, mas assume-se que a descrição

da actividade mental pode ser feita num nível conceptual (e.g. [Newell, 1982; Pylyshyn, 1984]), em vez de num nível físico.

Nesse sentido, as teorias cognitivas da emoção têm por objectivo relacionar os fenómenos emocionais com uma arquitectura cognitiva de processamento de informação em termos de: (i) unidades de processamento cognitivo envolvidas; (ii) regras que regem a computação; (iii) fluxos de informação entre unidades de processamento. Este tipo de caracterização reflecte bem a influência do paradigma cognitivo, anteriormente discutido (secção 2.1.6.1), na formação das teorias cognitivas da emoção. Este tipo de caracterização permite, no entanto, a abordagem de um aspecto importante, não contemplado nas teorias da emoção de base fisiológica, que é o carácter intencional de alguns fenómenos emocionais [Elster, 1999]. A exploração desse carácter é um dos aspectos centrais das teorias cognitivas da apreciação, abordadas de seguida.

2.2.3 Teorias Cognitivas da Apreciação

O facto dos fenómenos emocionais poderem envolver uma componente intencional, a qual não é contemplada numa perspectiva meramente fisiológica, levou a uma ideia central das teorias cognitivas, segundo a qual as emoções envolvem alguma forma de avaliação, designada genericamente por *avaliação* (“*appraisal*”).

A noção de apreciação emocional foi introduzida por Magda Arnold e J. Gasson (1954), segundo os quais uma emoção relaciona o sujeito com um objecto. Nesse sentido, as emoções são essencialmente relacionais. Para Arnold e Gasson, uma emoção é sentida como uma tendência de aproximação a um objecto julgado favorável, ou de afastamento de um objecto julgado desfavorável, reforçada pelas mudanças fisiológicas de acordo com o tipo da emoção [Oatley & Jenkins, 1996].

Assim, a apreciação dá origem a atracção ou aversão e a emoção é considerada como uma tendência de atracção sentida para o que é favorável ou uma tendência de repulsão do que é desfavorável, correspondendo a uma emoção positiva ou negativa, respectivamente. Para além desta avaliação base, outras podem ser observadas, dependendo das características do objecto e da relação entre o sujeito e o objecto. A partir dessas distinções, Arnold e Gasson propõem uma tabela de emoções positivas e negativas, cada uma definida com base num conjunto de apreciações específicas. Esta ideia foi a base das diversas teorias da apreciação emocional que se desenvolveram a

partir daí. Essas teorias seguem as linhas gerais do que é proposto por Arnold e Gasson, diferindo sobretudo na ênfase dada a aspectos específicos da apreciação.

O impacto da apreciação emocional nas teorias da emoção foi grande, originando a ideia de que a apreciação emocional é o elemento central dos fenómenos emocionais. Por exemplo, Richard Lazarus [1991] defende que as apreciações são necessárias e suficientes para a emoção, vendo a identificação das várias emoções como completamente determinável a partir dos padrões de apreciação que lhe estão na origem. Por sua vez, Frijda [1986, p. 225] refere que a apreciação é a questão central na teoria da emoção.

A ideia central das teorias da apreciação é, assim, o postular que as emoções surgem e são diferenciadas na base de uma avaliação pessoal e subjectiva (designada apreciação), da relevância de uma situação, objecto ou evento de acordo com um conjunto de critérios ou dimensões de apreciação [Scherer, 1999].

Um dos aspectos que sobressai das teorias da apreciação é a existência de múltiplos critérios de apreciação de acordo com a ênfase dada a diferentes aspectos dos fenómenos emocionais. A figura seguinte dá uma ideia de alguns conjuntos de dimensões de apreciação que têm sido propostos.

Scherer	Frijda	Roseman	Smith/Ellsworth
Novelty	Change		Attentional activity
• Suddenness			
• Familiarity	Familiarity		
• Predictability			
Intrinsic pleasantness	Valence		Pleasantness
Goal significance		Appetitive/aversive motives	
• Concern Relevance	Focality	Certainty	Importance
• Outcome probability	Certainty		Certainty
• Expectation	Presence		
• Conduciveness	Open/Closed	Motive consistency	Perceived obstacle/ Anticipated effort
• Urgency	Urgency		
Coping potential			
• Cause: agent	Intent/self-other	Agency	Human agency
• Cause: motive			
• Control	Modifiability	Control potential	Situational control
• Power	Controllability		
• Adjustment			
Compatibility Standards			
• External	Value relevance		Legitimacy
• Internal			

Figura 2.5. Dimensões de apreciação [Scherer, 1999, p.639].

Note-se a diferença de critérios propostos por diferentes autores, onde alguns tentam identificar dimensões consideradas essenciais, como é caso de Roseman, enquanto outros tentam garantir uma abrangência alargada, como é o caso de

Scherer. Nesse sentido, vários autores têm formulado perfis de apreciação específicos para diferentes emoções, considerados como representando o significado central da emoção a que se referem, tal como ilustrado na figura seguinte para o modelo proposto por Scherer.

Stimulus evaluation checks	Anger/rage	Fear/Panic	Sadness
Novelty			
• Suddenness	High	High	Low
• Familiarity	Low	Open	Low
• Predictability	Low	Low	Open
Intrinsic pleasantness	Open	Open	Open
Goal significance			
• Concern Relevance	Order	Body	Open
• Outcome probability	Very high	High	Very High
• Expectation	Dissonant	Dissonant	Open
• Conduciveness	Obstruct	Obstruct	Obstruct
• Urgency	High	Very high	Low
Coping potential			
• Cause: agent	Other	Other/nature	Open
• Cause: motive	Intent	Open	Change/neg
• Control	High	Open	Very low
• Power	High	Very low	Very low
• Adjustment	High	Low	Medium
Compatibility Standards			
• External	Low	Open	Open
• Internal	Low	Open	Open

Figura 2.6. Perfis de apreciação para diferentes emoções [Scherer, 1999, p.639]

A identificação e estruturação dos múltiplos critérios de apreciação resulta em grande medida de estudos experimentais, onde a estratégia mais comum é pedir às pessoas sujeitas ao estudo para recordarem episódios emocionais e para descreverem as avaliações realizadas e os resultados delas resultantes [Scherer, 1999].

Discussão

Importantes críticas têm sido feitas às teorias da apreciação emocional. Uma primeira questão resulta da ênfase cognitiva presente nessas teorias. É argumentado que os fenómenos emocionais resultam em grande medida de mecanismos sub-conceptuais, de reacção rápida, não compatíveis com critérios de apreciação de natureza cognitiva [Plutchik, 1980; Panksepp, 1998]. Este facto tem sido confirmado por evidências experimentais (e.g. [LeDoux, 1996]). Perante essas evidências, alguns investigadores das teorias de apreciação têm proposto que os processos de apreciação podem ocorrer a diferentes níveis do sistema nervoso, envolvendo processamento sub-conceptual, nomeadamente, processamento associativo/esquemático (e.g. [Leventhal & Scherer, 1987]), naquilo que designam por teorias multi-nível [Scherer, 2000].

Uma segunda questão importante resulta dos modelos propostos se basearem em relatórios verbais de memórias pessoais, constrangidos pela própria natureza dos questionários, o que resulta em inferências tautológicas, que apenas descrevem a estrutura semântica do vocabulário emocional [Scherer, 1999]. Esta questão é reconhecida como muito mais séria, o que leva alguns investigadores a reconhecer que as teorias da apreciação não apresentam uma explicação adequada para todos os tipos de fenómenos emocionais [Scherer, 1999].

Outras questões, como o carácter dinâmico e contínuo dos fenómenos emocionais, permanecem em larga medida fora do âmbito das teorias da apreciação. No entanto, a ênfase nos aspectos cognitivos trouxe a estas teorias uma grande relevância, quer no âmbito da psicologia, quer em termos de implementação computacional, pelo que de seguida serão analisadas três teorias consideradas representativas de três principais vertentes de teorias da apreciação.

2.2.3.1 Teoria de Ortony, Clore e Collins

Uma das mais conhecidas teorias da apreciação emocional foi proposta por Ortony, Clore e Collins [Ortony *et al.*, 1988]. O modelo de apreciação proposto é conhecido como modelo OCC, tendo sido significativamente explorado na modelação de emoção em agentes inteligentes. É também um exemplo característico do que pode ser designado como vertente estrutural das teorias da apreciação. O modelo OCC ilustra essa vertente, a qual enfatiza essencialmente a estrutura semântica associada aos termos emocionais [Scherer, 2000]. Desse modo, os fenómenos emocionais são descritos com base numa discriminação linguística de possíveis estados emocionais, onde a caracterização emocional é feita com base em conjuntos de regras de apreciação.

Ortony, Clore e Collins caracterizam emoção como uma reacção afectiva a eventos, agentes ou objectos, sendo a natureza específica dessa reacção determinada pela forma como a situação em causa é apreciada. Nesse sentido, o elemento central da teoria é a definição de critérios através dos quais essa apreciação pode ser feita.

A definição dos critérios de apreciação propostos assenta na noção de tipo de emoção, o qual expressa um conjunto de formas relacionadas e reconhecíveis de emoção [Ortony *et al.*, 1988]. O objectivo é lidar com a multiplicidade de fenómenos emocionais focando tipos específicos de emoção através do agrupamento e caracterização conjunta de fenómenos emocionais que se consideram relacionados. Por

exemplo, o tipo de emoção felicidade (“joy”) é definido da seguinte forma [Ortony *et al.*, 1988, p. 86]:

JOY EMOTIONS

TYPE SPECIFICATION (appraisal): (pleased about) a desirable event

TOKENS: contented, cheerful, delighted, ecstatic, elated, euphoric,
feeling good, glad, happy, jubilant, pleasantly surprised,
pleased, etc.

VARIABLES AFFECTING INTENSITY:

(1) the degree to which extent the event is desirable

Para cada tipo de emoção identificado, é definido um conjunto de condições de discriminação (“eliciting conditions”), o qual determina as condições em que a emoção correspondente é desencadeada. Para descrever a intensidade da emoção são definidos conjuntos de variáveis, consideradas como passíveis de afectar a intensidade da emoção, quer em termos específicos de um tipo particular de emoção, quer em termos globais de todas as emoções. As variáveis globais incluem: sentido de realismo (caracterizado como o nível de crença de que a situação é real); proximidade (subjectiva) da situação; imprevisibilidade; e mobilização (“arousal”). Aumentos nestas variáveis são considerados como intensificadores da experiência emocional.

Os vários tipos de emoção estão organizados de acordo com uma hierarquia de critérios de apreciação. Durante a apreciação de um evento, o tipo de emoção resultante depende da decisão feita em cada nível da hierarquia de critérios de apreciação. A figura seguinte ilustra a estrutura base da hierarquia de critérios de apreciação proposta.

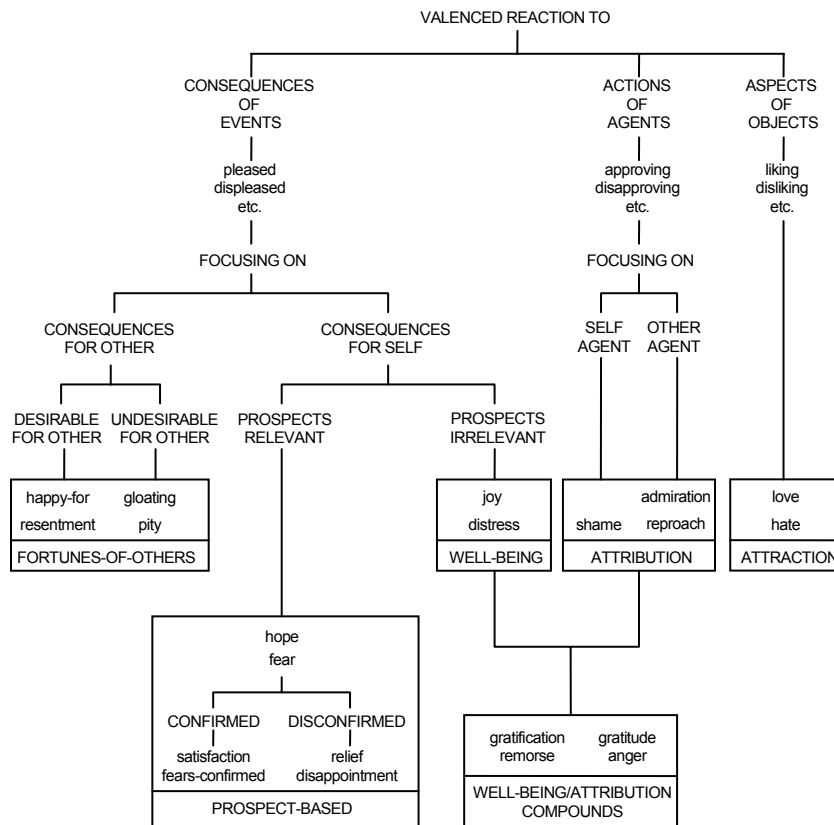


Figura 2.7. Estrutura global de tipos de emoção no modelo OCC [Ortony *et al.*, 1988; Clore & Ortony, 2000].

O facto do modelo OCC especificar um conjunto concreto de critérios de apreciação passíveis de serem expressos sob a forma de regras formais, directamente implementáveis em modelos computacionais, tornou o modelo OCC um dos principais suportes de modelação de emoções em agentes inteligentes durante os anos 90 (e.g. [Elliot, 1992]).

Discussão

Apesar do modelo OCC poder ser uma abordagem apelativa em termos de implementação computacional, este tipo de teorias da emoção é, essencialmente, uma forma de classificação de termos emocionais e não propriamente um suporte efectivo para a explicação dos fenómenos emocionais. Nomeadamente, não proporciona uma explicação para a natureza dinâmica e contínua, típica dos fenómenos emocionais, nem para os processos que estão subjacentes à emoção [Smith & Kirby, 2001].

No que se refere a uma implementação computacional, o facto de serem utilizadas caracterizações psicológicas muito gerais, torna problemática a especificação concreta dos critérios de apreciação ou, por exemplo, do cálculo da intensidade emocional,

deixando, deste modo, uma grande margem para implementações *ad hoc* (e.g. [Sloman, 2004]).

Várias alternativas a esta perspectiva estrutural têm sido propostas no âmbito das teorias da apreciação (e.g. [Oatley & Jenkins, 1996; Frijda, 1986]). Apesar de manterem a ênfase nos aspectos de apreciação emocional, dão igualmente uma grande relevância aos processos subjacentes aos fenómenos emocionais. Um dos modelos de grande relevância nesta área é o proposto por Frijda [1986; 2001], tendo igualmente sido objecto de implementação computacional por Moffat e Frijda [1995].

2.2.3.2 Teoria de Frijda

A teoria proposta por Frijda [1986; 2001] assenta em três princípios base: (i) a existência de uma base biológica para a emoção; (ii) o facto dessa base biológica ser modificada por factores sociais e cognitivos; (iii) a regulação como um aspecto integral da emoção. Um dos elementos centrais da teoria é a definição de um processo da emoção, o qual caracteriza os diferentes aspectos que relacionam um estímulo com uma resposta emocional. A figura seguinte ilustra esse processo.

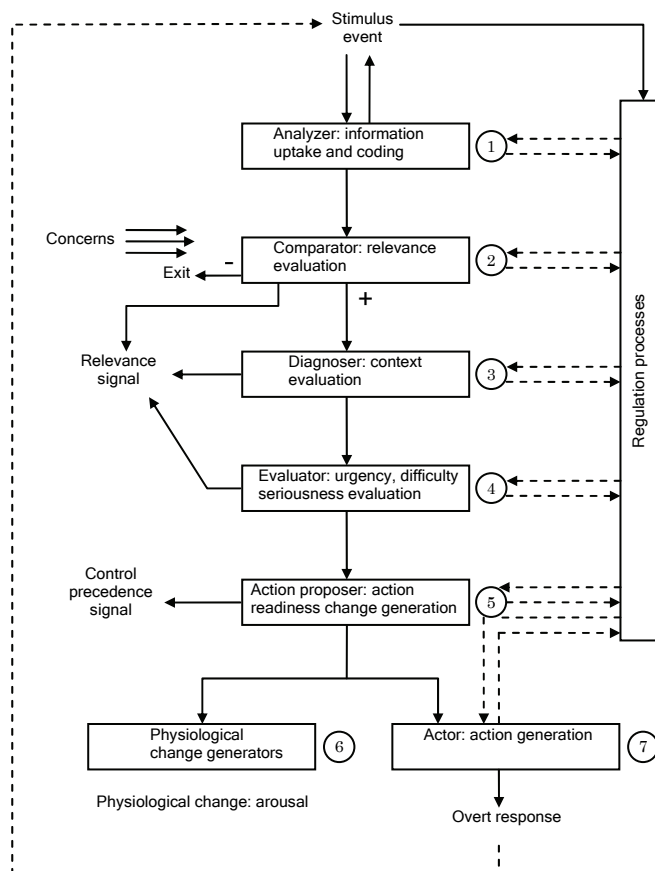


Figura 2.8. Processo emocional proposto por Frijda [1986].

Como se pode observar na figura 2.8, o processo proposto é composto por sete “elementos”:

- *Analizador*: recebe e estrutura a informação associada aos estímulos;
- *Comparador*: avalia os estímulos em termos da sua relevância, gerando sinais que traduzem prazer, dor, espanto, desejo ou irrelevância;
- *Diagnosticador*: avalia os estímulos em termos do contexto em que ocorrem e do que se pode fazer em relação a eles;
- *Avaliador*: realiza a avaliação da urgência, dificuldade e seriedade da situação, gerando um sinal de controlo respectivo;
- *Promotor de acção*: realiza a preparação para a acção através do estabelecimento de um plano de acção;
- *Gerador de modificação fisiológica*: altera as condições fisiológicas de acordo com as condições observadas;
- *Gerador de Acção*: determina a acção a realizar.

Tal como ilustrado na figura 2.8, todas estas fases são influenciadas ou reguladas por um processo de regulação, nomeadamente em termos de controlo voluntário. Segundo Frijda, este processo emocional não ocorre de forma isolada. Em vez disso, propõe a existência de uma contínua monitorização do ambiente, a par com uma contínua tendência para a acção e para mudanças fisiológicas. Este aspecto de prontidão para a acção é um aspecto central da teoria, numa relação estreita com o que designa por *preocupações* (“*concerns*”). As preocupações são definidas como disposições para desejar a ocorrência ou não ocorrência de uma situação. Por exemplo, a preocupação de obter comida.

Um aspecto que distingue claramente a teoria de Frijda das teorias estruturais do tipo proposto por Ortony, Clore e Collins, é a sua ênfase no dinamismo e na evolução temporal. Assim, os eventos associados ao processo da emoção são vistos como decorrendo ao longo do tempo e não numa perspectiva estática. Segundo Frijda, tudo está em constante mudança. À medida que a apreciação da situação é revista, fazem-se revisões de estimativas do futuro, das possibilidades de ajustamento para *lidar com a situação* (“*coping*”), de qual a susceptibilidade de controlo dos eventos, entre vários outros aspectos. Tudo isto num contexto de um mecanismo de realimentação, que relaciona os vários componentes do processo.

Um outro elemento importante da proposta de Frijda é a relação dos processos de apreciação com a concretização de objectivos. Assim, sugere que os processos de apreciação geram sinais internos, tais como sinais de relevância de prazer ou dor, mas também sinais de controlo de precedência de objectivos. Esses sinais de controlo de precedência podem causar a interrupção das actividades actuais do sistema, de modo a gerar e adoptar novos planos para lidar com a nova situação, ou gerando um novo modo de activação, como uma mobilização (“arousal”) reforçada.

Frijda propõe também uma monitorização central, de modo a integrar os múltiplos fluxos de informação envolvidos nos processos emocionais, no que designa por estruturas de significado situacional. Estas estruturas de significado situacional referem-se ao facto das emoções serem geradas por diferentes tipos de acontecimentos cujo significado determina o tipo de emoção. Estas estruturas são o suporte da discriminação de estados emocionais através da apreciação cognitiva, típica das teorias da apreciação. No entanto, Frijda refere que estas apreciações podem ocorrer a diferentes níveis, não implicando a consciência dessa apreciação. Propõe, assim, que a apreciação ocorre frequentemente num nível automático, embora possa envolver deliberações e raciocínio consciente.

As várias estruturas de significado situacional são representadas como conjuntos de componentes discriminativos, formando o que designa por perfis. Esses perfis definem várias emoções em termos do seu significado situacional e do contexto cognitivo em que ocorrem. A figura seguinte mostra alguns dos perfis propostos para várias emoções.

	Positive character	Negative character	Desire	Interest	Positive valence	Negative valence	Presence	Absence	Certainty	Uncertainty	Change	Open	Closed	Intentionality of other	Intentionality of self	Controlability	Noncontrolability	Modifiability	Finality	Object	Event	Focality	Globality	Strangeness	Familiarity	Value
Joy	X				X		X				(X)	X						X			X	X				
	X					X		X			(X)	X						X			X	X				
Distress		X			X		X										X				X	X				
		X				X	X										X				X	X				
Sorrow		X			X						(X)										X	X				
Fear		X			X			X		X			X				X	X			X	X				
		X				X	X			X			X				X	X			X	X				
Hope	X					X		X		X		X									X	X				
	X				X		X			X		X									X	X				
Anger	X	X				X	X							X		X						X	X			
	X	X			X			X						X		X						X	X			
Hate		X				X	X										X				X	X				
Frustration		X			X			X					X	X			X				X		X			
		X				X	X						X	X			X				X		X			
Guilt		X				X	X								X							X	X			X
Love	X				X		X														X		X			
Pride	X				X		X								X						X		X			

Figura 2.9. Alguns perfis de emoção propostos por Frijda (adaptado de [Frijda, 1986]).

Como podemos observar, este tipo de caracterização é, no essencial, análoga à proposta por Ortony, Clore e Collins. De facto, as estruturas de significado situacional podem ser directamente descritas como conjuntos de regras de modo semelhante ao que acontece no modelo OCC.

Discussão

Existe na teoria de Frijda a preocupação com o processo inerente à emoção, o que não acontece nas abordagens meramente estruturais. Apesar disso, o processo emocional definido é em grande medida abstracto, mantendo a noção de emoção como resultando da apreciação cognitiva de eventos e situações, e mantendo a mesma perspectiva semântica/linguística de discriminação de estados emocionais.

Apesar de vários autores realçarem o carácter altamente limitativo deste tipo de descrição dos fenómenos emocionais, só no final da década de 90 essas críticas ganharam uma abrangência significativa, impulsionadas em grande medida pelos desenvolvimentos na área das abordagens baseadas em sistemas dinâmicos.

Um dos principais proponentes de uma via alternativa para o estudo da emoção é Klaus Scherer [1984; 2000; 2001]. Scherer defende que é o processo da emoção que deve ser enfatizado, isto é, a contínua e dinâmica relação entre os organismos e o ambiente, desenvolvida ao longo do tempo através da contínua monitorização e avaliação do ambiente [Scherer, 2000]. Realça, assim, que apesar de se reconhecer teoricamente a importância da caracterização da emoção como um processo, o facto é que a maior parte da investigação está fortemente assente na noção de estado emocional e da sua caracterização em termos de etiquetas emocionais [Scherer, 2000, p. 70]. Propõe, por isso, uma mudança de paradigma para uma abordagem essencialmente dinâmica, a qual iremos apresentar de seguida.

2.2.3.3 Teoria de Scherer

Segundo Scherer [2000], o problema fundamental de qualquer teoria da emoção é explicar num único modelo: (i) a natureza dinâmica e contínua dos processos emocionais e, simultaneamente, a existência de etiquetas linguísticas discretas referentes a estados específicos; (ii) a natureza psicobiológica da emoção, incluindo a sua continuidade filogenética e, simultaneamente, o seu carácter e significado cultural; (iii) a distinção fenomenológica entre emoção e cognição e, simultaneamente, o intrincado entrelaçamento entre cognição e emoção. O modelo que propõe tem por objectivo abordar estes aspectos. Nesse modelo, emoção é definida como uma sequência de mudanças nos estados de cada um de cinco subsistemas funcionais orgânicos:

- Sistema cognitivo (*apreciação*);
- Sistema nervoso autónomo (*mobilização* - “*arousal*”);
- Sistema motor (*expressão*);
- Sistema motivacional (*tendências de acção*);
- Sistema de monitorização (*sentimento*).

Essas mudanças ocorrem de modo interdependente e interrelacionado (quando comparado com o funcionamento normal, mais independente, desses subsistemas), em resposta à avaliação de estímulos, eventos ou mudanças no próprio organismo que sejam de importância central para as suas necessidades ou objectivos [Scherer, 1984; Scherer, 2000]. O surgimento de determinado estado emocional exige um certo grau de sincronização entre as mudanças a ocorrer nos vários subsistemas.

O modelo prevê a estruturação de um conjunto de padrões de processamento nos quais:

- o organismo aprecia o evento com base num conjunto de *verificações de avaliação* (“*stimulus evaluation checks*”), entre os quais: novidade, agradabilidade intrínseca, relevância para os objectivos e compatibilidade com as normas;
- o resultado de cada verificação individual modifica o estado de cada subsistema do organismo numa direcção de adaptação ao evento;
- o padrão de reacção emocional é o resultado cumulativo de todas as modificações resultantes das apreciações efectuadas e das consequentes interacções por realimentação (“feedback”) ou alimentação directa (“feedforward”).

A estrutura geral do modelo é ilustrada na figura seguinte.

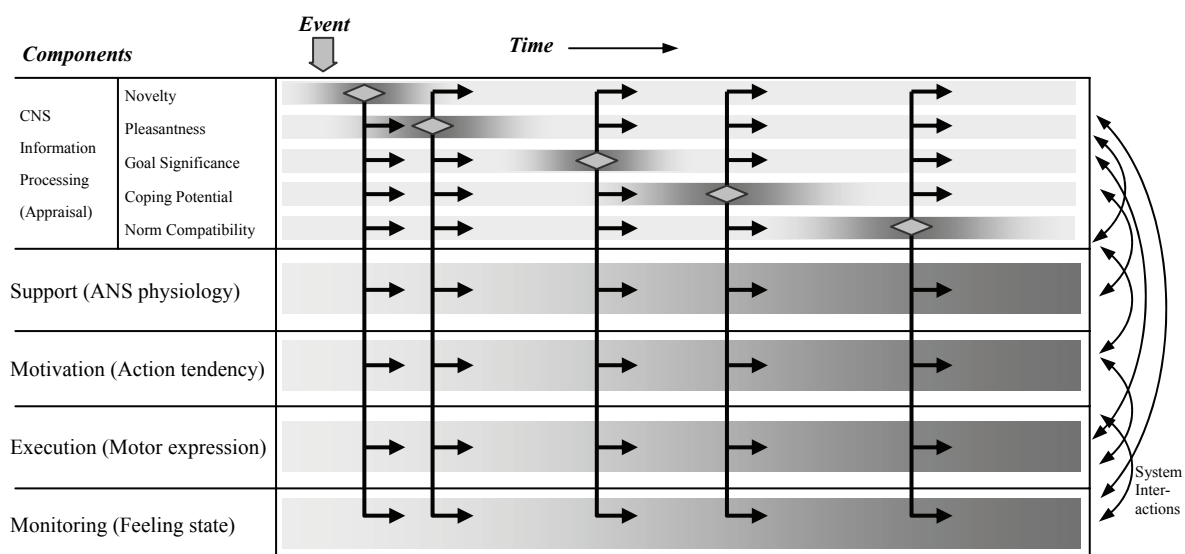


Figura 2.10. O Modelo de Processo de Emoção por Componentes [Scherer, 2000].

Na figura são mostradas as modificações em cinco subsistemas, como resultado de um evento, durante um determinado período de tempo após a ocorrência do evento. O período de sincronização desses subsistemas é considerado como um episódio emocional. Deste modo, as respostas destes cinco subsistemas são vistas como os componentes da emoção. A identificação e diferenciação dos tipos de sincronização é determinada pelas cinco principais verificações de avaliação (ou apreciação). Tal como

ilustrado pelas bandas paralelas da figura, o processamento de informação subjacente às cinco verificações de avaliação, ocorre em paralelo. No entanto, é esperado que ocorram interações entre os vários subsistemas como resultado dessas apreciações (representado pelas setas verticais no lado direito da figura). A maior ou menor coloração das bandas na figura pretende reflectir o processo de decisão, onde o losango indica o seu auge, isto é, onde a apreciação é completada e informação e comandos são enviados para outros subsistemas.

Considera-se que o tempo de decisão em cada processo pode variar de acordo com o tipo de verificação de avaliação, dependendo das características próprias da avaliação ou de restrições de processamento. Considera-se, por exemplo, que as verificações de novidade e agradabilidade intrínseca ocorram num nível sensorial-motor ou de processamento esquemático/associativo. Scherer apresenta, assim, explicitamente, a distinção entre processamento cognitivo e processamento esquemático/associativo, essencialmente ausente da perspectiva puramente cognitiva [Scherer, 2000].

O facto do atraso de decisão poder ser diferente para os vários subsistemas levanta um problema de sincronização, que Scherer resolve propondo que as várias apreciações de avaliação têm uma ordenação temporal específica, apesar da actividade dos processos envolvidos ser concorrente, o que na prática se traduz na sequencialidade da avaliação. No entanto, refere que apesar da sequencialidade da avaliação o processo ocorre em permanência, em ciclos sucessivos, pelo que o resultado das avaliações num determinado instante pode influenciar as avaliações seguintes.

A teoria proposta é claramente conceptual, não sendo indicadas formas concretas de a operacionalizar, como o próprio Scherer admite. Argumenta não existirem suportes conceptuais ou metodológicos que permitam lidar com o nível de complexidade exibido pelos processos emocionais, não vendo como seja possível, no contexto do paradigma cognitivo actual, ultrapassar essas limitações [Scherer, 2000, p. 77]. No entanto, apresenta pistas bastante interessantes para o que poderá ser uma mudança de paradigma. Propõe que é necessário “... passar de um pensamento em termos de caixas discretas, etiquetas ou mesmo de redes neuronais, para uma perspectiva não linear e dinâmica da emoção” [Scherer, 2000, p. 80].

Para suportar esta perspectiva propõe a abordagem do problema da apreciação em termos da teoria dos sistemas dinâmicos não lineares, em particular da teoria das catástrofes, com um realce particular para a noção de histerese, tal como ilustrado na figura seguinte.

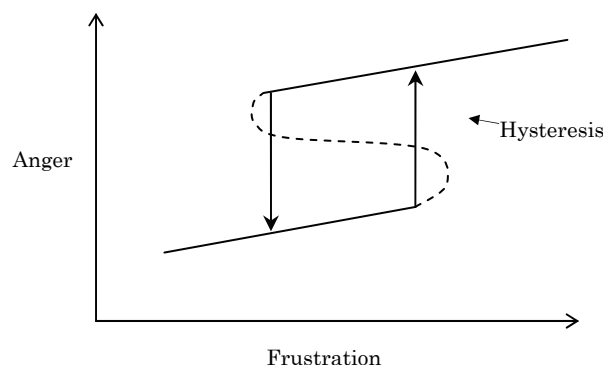


Figura 2.11. Uma ilustração do fenómeno de histerese na relação entre zanga e frustração [Scherer, 2000].

A figura ilustra a relação não linear entre o nível de frustração e o nível de zanga, típico de fenómenos emocionais (e.g. [Zeeman, 1976]). Ao contrário de um modelo linear, onde o nível de zanga iria variar proporcionalmente ao nível de frustração, num modelo com histerese existe uma região inacessível, levando a que a intensidade de zanga mude abruptamente para determinados graus de zanga.

Segundo Scherer, isto explica duas observações importantes de fenómenos emocionais. Em primeiro lugar, as transições bruscas já referidas na relação entre o nível de frustração e o nível de zanga, típicas de sistemas não lineares (relação que se observa, também, noutras emoções como é o caso do medo). Em segundo lugar, a sensibilidade aos valores iniciais, também presente nos sistemas não lineares. Por exemplo, após determinado nível de frustração ter produzido zanga, é necessário que a frustração diminua para um nível significativamente inferior ao que levou ao estado de zanga, para que um retorno ao estado normal ocorra.

Tal como Scherer realça, estas características dos fenómenos emocionais são bem conhecidas. No entanto, os modelos actuais de emoção não permitem prever esses fenómenos [Scherer, 2000].

Esta ideia de utilização de modelos não lineares na caracterização dos fenómenos emocionais não é nova. Por exemplo, Zeeman [1976] propôs esse tipo de modelação. A figura seguinte ilustra um modelo tridimensional, baseado na teoria das catástrofes, para classificação de um fenómeno comportamental clássico em etologia, que é o conflito de resposta entre atacar ou fugir, num cão que enfrenta um adversário de tamanho desconhecido [Lorenz, 1950].

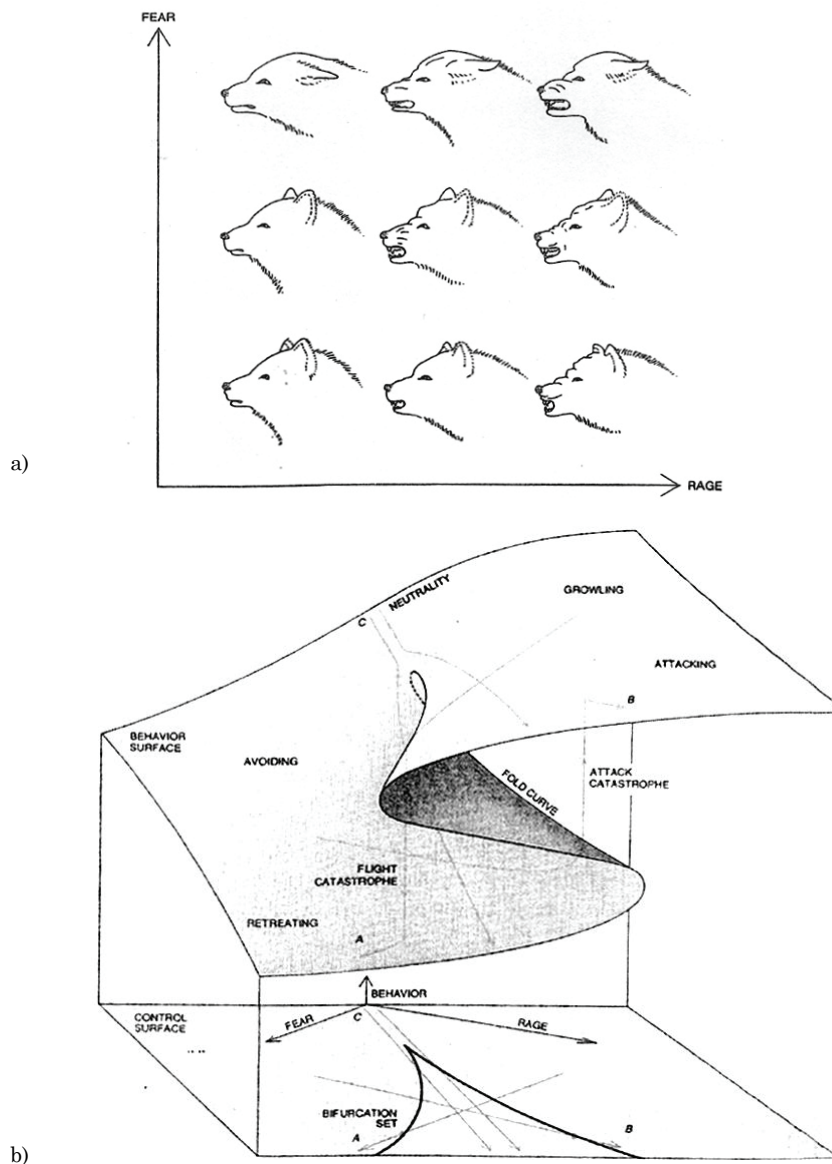


Figura 2.12. Modelação do dilema fugir-atacar em cães, com base num modelo de catástrofe (b). O modelo assume que o comportamento agressivo é controlado por dois factores que se contrapõem, a raiva e o medo, representados como eixos do plano horizontal (a superfície de controlo). Estes factores de controlo podem ser medidos a partir da expressão facial dos cães (a). A partir destes indicadores é possível descrever o estado emocional dos cães (adaptado de [Zeeman, 1976]).

A conceptualização proposta consiste na definição de um espaço de controlo (apresentado na base do gráfico da figura 2.12.b e no corte horizontal correspondente à figura 2.12.a) que expressa os factores controladores do comportamento do animal. A parte superior do gráfico mostra a superfície comportamental (ou de resposta). Observa-se que a não linearidade da superfície de resposta é consistente com características típicas bem conhecidas do comportamento emocional [Scherer, 2000]. Com base neste modelo, Scherer propõe um modelo de apreciação emocional onde

identifica regiões correspondentes a discriminações linguísticas clássicas de diferentes emoções, tal como ilustrado na figura seguinte.

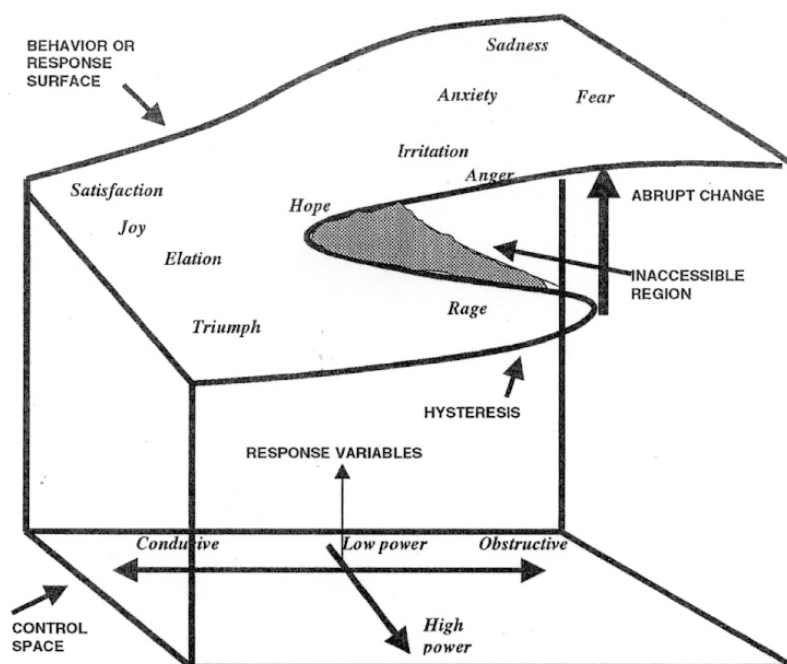


Figura 2.13. Modelo de apreciação emocional baseado num modelo de catástrofe [Scherer, 2000].

É interessante notar o possível mapeamento entre caracterização emocional (figura 2.13) e caracterização comportamental (figura 2.12), algo que será abordado mais à frente.

Quando confrontamos este modelo com os conjuntos de critérios de apreciação emocional, típicos das teorias da apreciação, sobressai o facto de haver uma grande divergência no número de dimensões de apreciação. Enquanto o modelo não linear proposto por Scherer apenas apresenta duas dimensões (condutividade e poder de acção), outras abordagens apresentam múltiplos critérios, como vimos anteriormente. Em relação a isso, Scherer refere que será necessário a definição de modelos não lineares de ordem superior, onde múltiplas dimensões de apreciação possam ser incluídas.

Iremos ver que essa poderá não ser a melhor solução. Pelo contrário, as duas dimensões base que são propostas para conciliar a perspectiva comportamental (figura 2.12) com a caracterização emocional (figura 2.13), poderão efectivamente indiciar princípios chave intrínsecos aos fenómenos emocionais.

Um número crescente de investigadores partilha a visão de Scherer, considerando que as teorias da emoção têm sido constringidas pelas abordagens cognitivas, perpetuando o debate entre cognição *vs.* emoção, dificultando assim uma modelação integrada destes dois aspectos. Em particular, grande parte da teoria da emoção deriva da psicologia cognitiva, onde o processamento sequencial de informação tem sido a metáfora dominante [Lewis & Granic, 2000], perspectiva que é claramente patente nas teorias da apreciação.

No entanto, outras teorias têm sido propostas, algumas com relevância para a abordagem apresentada nesta tese, nomeadamente as teorias que têm por base modelos de controlo e a teoria proposta por Herbert Simon, directamente relacionada com a modelação de agentes inteligentes.

2.2.4 Teorias de Controlo

O carácter adaptativo e regulador da emoção é amplamente reconhecido no estudo da emoção, incluindo no âmbito das teorias da apreciação (e.g. [Lazarus, 1991]). Contudo esse reconhecimento é tipicamente formal, forçado em grande medida pelas inequívocas evidências experimentais, não tendo, por isso, uma expressão concreta em termos dos modelos propostos, tal como vimos anteriormente. Isso não acontece com um conjunto de teorias baseadas nas noções de realimentação (“feedback”) e auto-regulação, derivadas da teoria de controlo de sistemas, como é o caso das abordagens propostas por Hsee e Abelson [1991], Coulson [2004] ou Carver e Scheier [2000; 2002]. A ideia central destas propostas é no essencial idêntica, tentando explicar os fenómenos emocionais como resultado de processos de auto-regulação. Estas teorias partem de uma noção central da teoria do controlo de sistemas que é a redução de discrepância numa malha de realimentação, seguindo assim parte da tradição da cibernética iniciada por Wiener [1948] e Ashby [1964], entre outros.

Nestas teorias, um estado actual observado é comparado com um valor de referência (por exemplo, um objectivo). Quando uma discrepância é detectada entre o estado actual e o valor de referência, um comportamento específico é realizado para reduzir a discrepância. Deste modo, a realimentação do sistema com a informação de que existe uma discrepância conduz à manutenção do estado actual próximo do valor de referência. Este processo é ilustrado na figura seguinte.

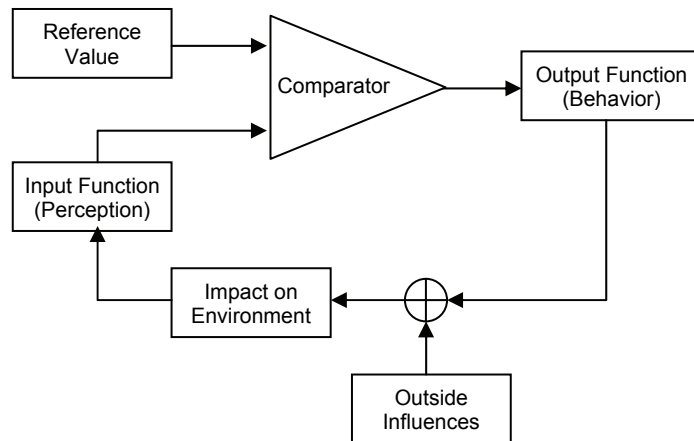


Figura 2.14. Ilustração esquemática de uma malha de realimentação, a unidade base de controlo cibernético [Carver & Scheier, 2000].

Em termos de comportamento humano, o controlo por realimentação significa que um indivíduo age de modo a minimizar qualquer discrepância discernível entre a situação actual e os padrões comportamentais de referência [Coulson, 2004].

Um outro princípio base subjacente a esta abordagem é a noção de organização hierárquica do comportamento [Carver & Scheier, 2000]. Em termos de teoria de controlo, isto significa que vários subsistemas de realimentação podem estar interligados de modo que a saída de um sistema de realimentação num determinado nível pode determinar os valores de referência noutros níveis da hierarquia. A figura seguinte ilustra esta organização.

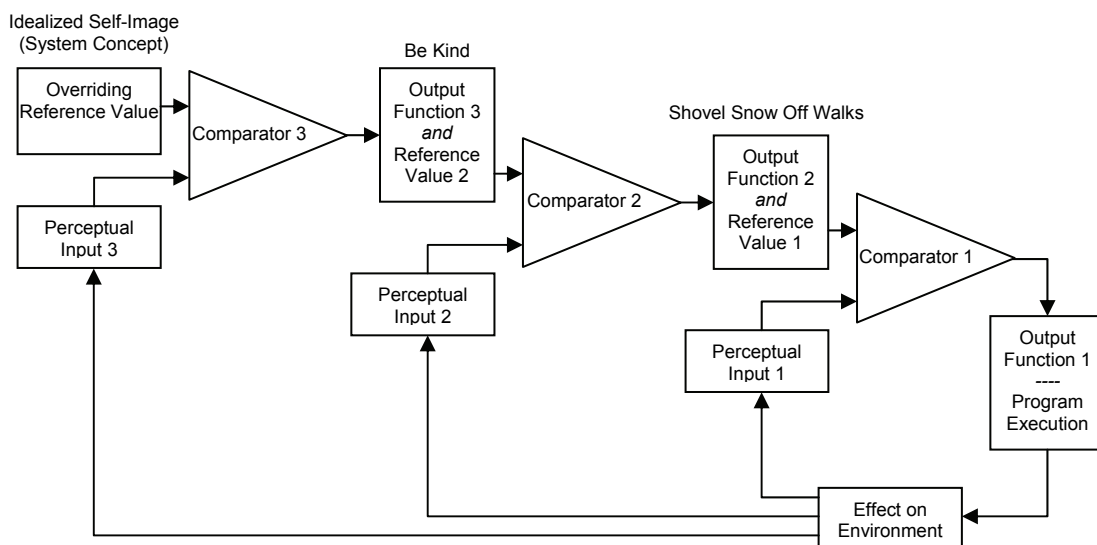


Figura 2.15. Hierarquia de malhas de realimentação [Carver & Scheier, 2000].

A organização hierárquica da figura 2.15 mostra três níveis de controlo. No nível mais elevado (designado *conceitos de sistema*) encontram-se valores de carácter abstracto, tornando-se mais concretos à medida que se desce na hierarquia.

No modelo proposto por Carver & Scheier, para além dos sistemas de realimentação, são propostos mecanismos que permitem auto-avaliar a actividade dos vários sistemas de retroacção. Com base nessa capacidade de auto-avaliação, um sistema de meta-monitorização verifica a taxa de redução de discrepância ao longo do tempo. Os autores propõem que o próprio nível de monitorização funcione como uma malha de realimentação. Neste caso, o valor de referência será a taxa de redução de discrepância desejável. Tal como nos restantes subsistemas, a existência de discrepância a este nível é motivadora de comportamento num sentido de corrigir essa discrepância.

Carver e Scheier propõem que o resultado destes mecanismos de comparação é manifestado de duas formas: (i) através de uma sensação de expectativa de resultados; e (ii) sob a forma de uma qualidade afectiva, expressa através de uma sensação de positividade ou de negatividade. A tabela seguinte ilustra esta ideia.

Comportamento	Acção (nível de acção)	Situação (nível de monitorização)	Afecto
Progresso em direcção ao objectivo, a uma taxa igual à de referência	Redução de discrepância em relação ao objectivo	Não há discrepância (da taxa de referência)	Nenhum
Progresso em direcção ao objectivo, a uma taxa inferior à de referência	Redução de discrepância em relação ao objectivo	Discrepância (da taxa de referência)	Negativo
Progresso em direcção ao objectivo, a uma taxa superior à de referência	Redução de discrepância em relação ao objectivo	Discrepância positiva (da taxa de referência)	Positivo

Tabela 2.1. Relação entre variação de taxa de discrepância e afecto (adaptado de [Carver & Scheier, 2000]).

No entanto, não é o nível de discrepância, mas sim a taxa de redução da discrepância que os autores consideram estar subjacente à experiência afectiva.

O que estes modelos apresentam como comparação para redução de discrepância pode ser considerado análogo à noção de apreciação existente nos modelos de apreciação. Para além disso, mantêm todo um enquadramento cognitivo em termos das estruturas conceptuais que utilizam, o que torna problemática a sua implementação concreta.

No entanto, evidenciam uma característica central dos fenómenos emocionais, que é o facto destes estarem intrinsecamente associados a uma relação contínua e dinâmica entre a capacidade de um organismo agir sobre o ambiente e a forma como esse ambiente evolui.

Para além disso, apresentam outro aspecto central claramente não contemplado nas abordagens cognitivas, que é o facto do carácter emocional estar essencialmente associado não à discrepância entre situações, mas sim à taxa a que essa discrepância evolui, o que é consistente com a perspectiva não linear proposta por Scherer.

2.2.5 Teoria da Interrupção

Herbert Simon [1967] propõe uma abordagem ao estudo da emoção centrada na noção de *racionalidade confinada* (“*bounded rationality*”), como reacção à noção de utilidade esperada desenvolvida no âmbito das teorias económicas de tomada de decisão. A proposta de Simon é que as representações e processos mentais são diferentes das assunções da racionalidade clássica (ver secção 2.1.5), na qual um agente conhece a informação exacta que necessita para computar, em princípio, qualquer utilidade que deseje. A sua ideia é que os organismos estão adaptados ao seu ambiente de tal modo, que a capacidade cognitiva de um organismo é suficiente para lhe permitir sobreviver nesse ambiente. Não importa que o organismo viole as normas da racionalidade desde que consiga encontrar comida e consiga evitar o perigo. Ou seja, um organismo satisfaz um determinado nível de aspiração, no sentido de encontrar uma solução suficientemente boa perante as condições em que se encontra, em vez de maximizar, no sentido de tentar encontrar a solução óptima [Simon, 1967; Simon, 1981].

O reconhecimento do facto que os organismos são limitados em termos de recursos, quer físicos (como energia e tempo), quer cognitivos (como capacidade de processamento e de memória), é um dos pontos de partida para a proposta seguinte, que consiste na possibilidade da emoção desempenhar um papel importante para a racionalidade, num contexto de limitação de recursos [Simon, 1967].

De acordo com Simon, sendo o sistema nervoso limitado em termos de capacidade de processamento e de memória, um evento no ambiente, uma memória ou uma motivação deve poder interromper o processamento actual. Por exemplo, a presença de um predador pode causar a geração de novos objectivos no sentido de lidar com a situação, entre outras possibilidades, gerando uma resposta do tipo *lutar ou fugir*

(“*fight or flight*”). Outros efeitos da interrupção através de um estímulo incluem a mobilização fisiológica (“arousal”) e a produção de sensações emocionais.

Simon estava principalmente interessado no processo de interrupção de objectivos, com a consequente mudança de comportamento, em vez de nos fenómenos fisiológicos ou subjectivos. Ou seja, a interrupção é necessária para garantir os requisitos de comportamento em tempo-real de um organismo. No entanto, Simon distingue entre interrupção adaptativa e não adaptativa. Os estímulos emocionais são considerados essencialmente interrupções adaptativas, ajudando o organismo a concretizar os seus objectivos. No entanto, em certos casos onde existe persistência ou intensidade excessiva, considera que poderá ser não adaptativa.

A ideia central da teoria de Simon é “... explicar como um processador de informação essencialmente sequencial, dotado de múltiplas necessidades, se comporta de modo adaptativo e sobrevive num ambiente que apresenta ameaças e oportunidades imprevisíveis” [Simon, 1967]. A explicação é baseada em dois mecanismos centrais: (i) um mecanismo executor de objectivos; e (ii) um mecanismo de interrupção, a emoção, a qual permite ao processador responder a necessidades urgentes em tempo real.

A teoria de Simon não contempla todos os aspectos dos fenómenos emocionais, no entanto explica a necessidade de arquitecturas cognitivas gerirem recursos que são limitados, bem como a forma como isso pode ser feito através de mecanismos emocionais, os quais muito rapidamente podem focar a atenção de um organismo para o que é importante ou urgente.

Esta relação estreita entre emoção e racionalidade, proposta por Simon há mais de 30 anos, viria a ter confirmação experimental no anos 90, em particular através do trabalho de investigação realizado por António Damásio [1994; 2000; 2003], o qual será abordado de seguida.

2.2.6 Emoção e Racionalidade: António Damásio

Existe uma longa tradição de crítica à racionalidade das emoções. Nessa tradição, que persiste em grande medida na cultura actual, a emoção é vista como um impedimento ao pensamento racional [Damásio, 1994; Ben-Ze’ev, 2000], tendo sido em grande medida ignorada no âmbito das ciências cognitivas [LeDoux, 2000]. Não admira, por isso, que tenha sido igualmente negligenciada em inteligência artificial,

em consequência da preocupação com a noção clássica de racionalidade como medida de capacidade de raciocínio.

Os desenvolvimentos nas neurociências trouxeram, entretanto, um conhecimento acrescido acerca do papel abrangente da emoção em múltiplos processos cognitivos, como formação de memórias, aprendizagem ou raciocínio e tomada de decisão [Damásio, 2000; Dolan, 2002]. De facto, evidências crescentes revelam uma intrincada relação entre emoção e cognição em geral. Por exemplo, Gray *et al.* [2002] apresentam evidências experimentais de ampla interação entre processamento emocional e cognitivo, com perda de especialização funcional, indicando uma efectiva integração entre processamento emocional e cognitivo.

Apesar de já nos anos 60 Herbert Simon ter proposto a emoção como um elemento central para a construção de sistemas inteligentes capazes de lidar com as características e restrições do “mundo real”, só nos anos 90 essa ideia veio a ganhar expressão. Dois trabalhos em particular contribuíram para essa mudança: as investigações sobre os mecanismos do medo realizadas por Joseph LeDoux [1996; 2000] e, em especial, as investigações sobre a relação entre emoção e racionalidade, realizadas por António Damásio [1994; 2000].

Um dos trabalhos de António Damásio que maior impacto teve resultou do estudo de indivíduos claramente racionais na forma como governavam as suas vidas até que, como resultado de danos neurológicos em regiões específicas do cérebro, perderam a capacidade de tomar decisões racionais e, simultaneamente, perderam a capacidade de processamento emocional normal. Apesar de terem mantido a capacidade de lidar com os aspectos pertinentes do mundo que os rodeia e, em particular, da sua capacidade de compreender e lidar com a lógica de um problema se manter intacta, muitas das suas decisões pessoais e sociais passaram a ser irracionais [Damásio, 1994].

Estas observações sugerem que uma redução da emoção é, pelo menos, tão prejudicial para a racionalidade quanto emoção excessiva. Mais, indicam que os processos de raciocínio e tomada de decisão humana são em grande medida suportados por processos emocionais [Damásio, 2000].

Damásio classifica as emoções em duas classes gerais: emoções primárias, despoletadas por estímulos externos, associadas a respostas fisiológicas como aceleração do ritmo cardíaco ou tremores; e emoções secundárias, relativas a recordações “emocionalmente carregadas” de eventos passados. Propõe que as emoções

primárias estão relacionadas com a amígdala, enquanto as emoções secundárias estão baseadas no córtex préfrontal, mas suportadas na amígdala. Existe assim uma segunda forma de activação emocional, para além dos estímulos exteriores, que consiste na activação de imagens de eventos passados.

Segundo Damásio, as mesmas áreas que estão envolvidas no raciocínio e planeamento a longo prazo, são igualmente responsáveis pela capacidade de ter uma resposta emocional a determinados estímulos, indicando que emoção e raciocínio estão fortemente interligados.

Para explicação destes fenómenos, Damásio propõe a hipótese do marcador somático, segundo a qual determinadas imagens mentais são marcadas com uma representação somática, isto é, do estado fisiológico do corpo. Os efeitos dos marcadores somáticos podem ser fisiológicos ou hipotéticos, “curto-circuitando” a expressão corporal através de um mecanismo “como-se”, mas que manteria as mesmas características. Deste modo, o corpo desempenha um papel central nos processos emocionais, o que Damásio designa como “teatro das emoções” [Damásio, 1994].

Outros aspectos centrais das conclusões de Damásio são a relação estreita entre emoção e regulação homeostática e a sua relação com a promoção da sobrevivência e manutenção da vida, bem como o facto das emoções serem inseparáveis dos estados de prazer ou dor, da ideia de bom ou de mau, de consequências favoráveis ou desfavoráveis e da recompensa ou punição por uma acção [Damásio, 2000].

2.2.7 Emoção e Adaptação: Joseph LeDoux

Um outro aspecto chave da relação entre emoção e cognição é o papel que os fenómenos emocionais desempenham na aprendizagem e capacidade de adaptação dos organismos a ambientes dinâmicos. Esse papel não é, no entanto, estanque, abrangendo igualmente os aspectos de raciocínio e tomada de decisão.

Joseph LeDoux identifica o medo como uma emoção chave desses mecanismos adaptativos. A ideia base que LeDoux propõe é a existência de dois caminhos distintos envolvidos numa resposta de medo, o que designa por *caminho superior* e por *caminho inferior*, tal como ilustrado na figura seguinte.

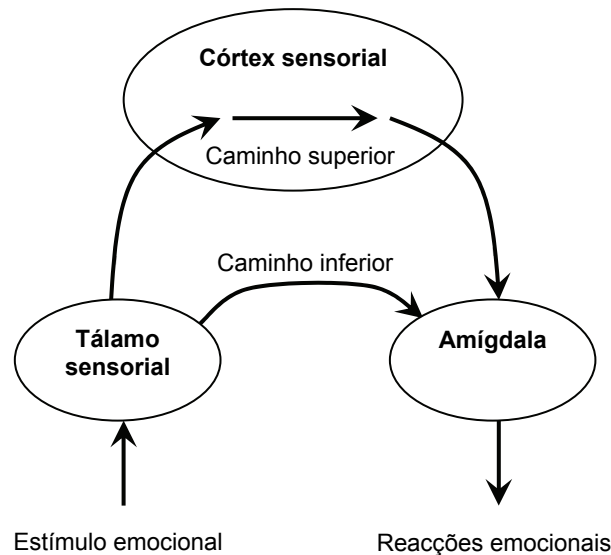


Figura 2.16. Caminhos de processamento emocional envolvendo a amígdala (adaptado de [LeDoux, 1996]).

Os estímulos emocionais são, assim, processados numa primeira etapa a nível do tálamo (ver figura 2.4). Neste ponto, surgem dois caminhos distintos. Um caminho leva à amígdala (caminho inferior), permitindo uma resposta rápida aos estímulos a partir de um processamento rudimentar. Um outro caminho leva ao córtex sensorial (caminho superior), permitindo um processamento mais elaborado dos estímulos, mas originando uma resposta mais lenta. Estes caminhos distintos são identificados por LeDoux como estando relacionados com respostas emocionais conscientes (caminho superior) e inconscientes (caminho inferior).

Para além de controlarem a resposta directa do organismo a estímulos, estes caminhos estão também relacionados com as capacidades de adaptação e aprendizagem através da modulação emocional das memórias formadas. A amígdala, em particular, apresenta múltiplas ligações a diferentes regiões cerebrais, havendo evidências de estar relacionada com a formação de memórias emocionais de longo-prazo, bem como com a experiência emocional que resulta da activação dessas memórias.

Os trabalhos de Damásio e LeDoux acerca da neurofisiologia da emoção mostraram, de modo claro, a existência de uma relação estreita entre processamento emocional e cognitivo no cérebro, a qual não se restringe a simples mecanismos de interrupção, mas antes tem uma abrangência global, nomeadamente em termos de raciocínio e tomada de decisão, e de aprendizagem e adaptação.

2.3 Conclusão

A emoção é um tema que tem sido estudado de inúmeras perspectivas, umas confluentes outras inerentemente contraditórias. Na tradição moderna, há linhas comuns e pontos de divergência, desde a perspectiva evolutiva de Darwin, à perspectiva fisiológica de James, até à perspectiva cognitiva moderna. No entanto, independentemente da perspectiva, predomina essencialmente uma visão funcional/modular da emoção, controlada por sistemas cerebrais discretos, moldados pela evolução de modo a permitir lidar com os desafios adaptativos de ambientes eventualmente hostis e em permanente mudança. É também aceite uma relação entre esses sistemas cerebrais e o processamento de informação a nível cognitivo, onde mecanismos de apreciação emocional produzem estados emocionais, independentemente desse processamento ocorrer a nível consciente ou subconsciente.

Esta perspectiva da emoção tem sido suportada nos estudos realizados para a identificação de dimensões de caracterização emocional, independentes de raça ou cultura, bem como pelas características neuro-anatómicas que indicam uma continuidade evolutiva dos sistemas cerebrais relacionados com a emoção, tal como propôs Darwin.

No entanto, à medida que mais investigação é feita sobre os fenómenos emocionais, mais esta visão computacional e modular se vai esvanecendo. De facto, os progressos significativos que têm ocorrido no estudo dos fenómenos emocionais, indicam que a emoção não é uma propriedade de sistemas cerebrais específicos (e.g. [Panksepp, 1998]). Mesmo a noção de sistema límbico é questionada por alguns investigadores (e.g. [LeDoux, 1996]), sendo considerada problemática devido ao carácter multifacetado dos fenómenos emocionais [Matthews *et al.*, 2002].

Em contrapartida, surge uma visão da emoção como um fenómeno contínuo e dinâmico (e.g. [Scherer, 2000; Lewis & Granic, 2000]), resultante da actividade conjunta de múltiplos processos, estruturados a diferentes níveis de organização e com continuidade filogenética (e.g. [Panksepp, 1998; Freeman, 2000]), responsáveis pela manutenção da viabilidade dos organismos em ambientes dinâmicos, incertos, limitados em recursos e, eventualmente, hostis (e.g. [Simon, 1967; Frijda, 1986; Oatley & Jenkins, 1996]), e que são um suporte essencial à actividade cognitiva desses organismos, nomeadamente ao raciocínio e tomada de decisão.

Compreende-se, assim, o crescente interesse em compreender e modelar fenómenos emocionais no âmbito da concepção e implementação de agentes inteligentes, cujos

modelos e arquitecturas têm sido em grande medida determinados pela visão clássica de racionalidade, discutida no início deste capítulo.

Capítulo 3

Trabalho Relacionado

Neste capítulo é feita uma panorâmica de algumas das principais abordagens de modelação de emoção no contexto de agentes inteligentes. São consideradas três perspectivas de enquadramento desses modelos: *fisiológica*, *cognitiva* e *arquitectural*. Em cada uma dessas perspectivas são apresentados, como exemplo, alguns modelos e arquitecturas que reflectem a perspectiva em causa ou que permitem contrastar diferentes abordagens.

3.1 Emoção em Agentes Inteligentes

Apesar dos trabalhos pioneiros de alguns investigadores como Herbert Simon [1967], Aaron Sloman e Monica Croucher [1981] ou Marvin Minsky [1985] que, desde há longa data, têm chamado a atenção para a necessidade de incorporar os aspectos emocionais como parte integrante dos modelos e arquitecturas de suporte à concepção de agentes inteligentes, a orientação prevalecente na comunidade de inteligência artificial distribuída foi, durante muito tempo, pouco receptiva a essas ideias.

No entanto, durante a década de 90 essa orientação mudou significativamente, em grande medida como resultado dos trabalhos de António Damásio e Joseph LeDoux, discutidos no capítulo anterior. Como resultado dessa mudança, surgiram várias linhas de investigação tendo como objectivo principal a integração de aspectos emocionais em modelos e arquitecturas de agentes inteligentes. As motivações para esses trabalhos são diversas, desde o facilitar o desenvolvimento de personagens mais realísticas para animação, entretenimento ou interacção com humanos, até àquilo que constitui o nosso objectivo central, que é a utilização de mecanismos emocionais para a implementação de agentes capazes de lidar com os problemas que resultam da acção em ambientes reais.

Reflectindo as múltiplas orientações que caracterizam o estudo dos fenómenos emocionais, também na integração de emoção em agentes inteligentes, diferentes tipos de modelos têm sido propostos. A caracterização desses modelos pode passar por diferentes vertentes, como o nível de especificidade (e.g. arquiteturas de agentes *vs.* modelação de fenómenos individuais), paradigma computacional de suporte (e.g. cognitivo *vs.* comportamental), tipos de fenómenos considerados ou tipo de aplicação (e.g. [Martínez-Miranda & Aldea, 2004; Hudlicka *et al.*, 2004; Strongman, 1996]). O critério aqui utilizado tem por base a caracterização desses modelos de acordo com a sua principal orientação em termos de teorias da emoção, sendo identificados três tipos genéricos de modelos:

- Modelos de base fisiológica;
- Modelos de base cognitiva;
- Modelos de base arquitetural.

Esta caracterização serve essencialmente como guia, não sendo de modo nenhum estanque, uma vez que muitos dos modelos propostos partilham características de mais do que um tipo de modelo. No entanto, permite-nos ter um enquadramento global da orientação dos modelos, o que é importante para a discussão que será feita de seguida.

3.2 Modelos de Base Fisiológica

Os modelos de base fisiológica têm a sua inspiração na relação entre os fenómenos emocionais e os fenómenos bioquímicos e homeostáticos que caracterizam os organismos biológicos (e.g. [Cañamero, 2003; Schulkin, 2003]). Nesse sentido, são caracterizados pela definição de uma fisiologia sintética, que tenta de alguma forma modelar os processos corporais associados aos fenómenos emocionais, bem como pela definição de mecanismos de auto-regulação inspirados nos mecanismos homeostáticos biológicos. Como exemplo, são discutidos nesta secção os modelos propostos por Cañamero (arquitetura *Abbot*) [Cañamero, 1997; Cañamero, 2003] e por Velásquez (arquitetura *Cathexis*) [Velásquez, 1998].

3.2.1 Arquitectura *Abbott*

Dolores Cañamero propõe uma abordagem baseada no conceito de Sociedade da Mente [Minsky, 1985], na qual utiliza as noções de motivação e emoção como base para a selecção e coordenação de comportamentos de um agente. Uma concretização dessa abordagem é a arquitectura *Abbott* [Cañamero, 1997]. Nessa arquitectura, os agentes implementados representam criaturas simuladas, com capacidades de percepção e acção limitadas, as quais habitam num mundo virtual bidimensional, discreto (em grelha), onde existem recursos (comida e água) e obstáculos. Nesse mundo existem duas espécies de criaturas, os *Abbotts* (que correspondem à implementação do modelo de agente) e os *Inimigos* (que possuem um comportamento simples predefinido). Os *Abbotts* são caracterizadas por uma fisiologia sintética, representada recorrendo a variáveis controladas homeostaticamente e à simulação de “hormonas”, as quais constituem o estado interno do agente, produzindo motivações controladas de forma homeostática. A tabela seguinte ilustra alguns parâmetros fisiológicos considerados.

Parameter	Initial Val.	Set Point	Range
Blood pressure	12	12	± 4
Blood sugar	30	30	± 10
Energy	120	100	± 50
Heart Rate	75	75	± 25
Pain	0	0	± 2
Respiration Rate	8	8	± 7
Temperature	37	37	± 3
Vascular Volume	25	20	± 10
Adrenaline	10	10	± 5
Dopamine	10	10	± 5
Endorphine	20	20	± 10

Tabela 3.1. Variáveis fisiológicas utilizadas para definir o estado corporal de um agente [Cañamero, 1997].

As variáveis controladas activam motivações - agressão, frio, curiosidade, fadiga, fome, auto-protecção, sede, calor - induzidas pelas necessidades internas, ou *impulsos* (“*drives*”), quando o respectivo valor sai dos limites de viabilidade. A tabela seguinte

ilustra alguns impulsos considerados, bem como os estados motivacionais a que dão origem.

Motivation	Drive
Aggression	Decrease adrenaline
Cold	Increase temperature
Warmth	Decrease temperature
Curiosity	Increase endorphine
Fatigue	Increase energy
Hunger	Increase blood sugar
Thirst	Increase vascular volume
Self-protection	Decrease pain

Tabela 3.2. Relações entre motivações para acção e impulsos (“drives”) resultantes da variação dos parâmetros fisiológicos [Cañamero, 1997].

Em cada momento, a motivação com a maior intensidade tentará activar comportamentos que possam da melhor forma satisfazer as necessidades mais urgentes. A execução desses comportamentos pode afectar quer o ambiente externo quer a fisiologia do agente. Como resultado da interacção com o ambiente, os agentes podem ter diferentes estados emocionais, os quais afectam, por sua vez, a fisiologia, atenção, percepção, motivação e comportamento geral do agente. A tabela seguinte ilustra a relação entre estes vários aspectos.

Stimulus	Motivation	Behavior	Main Effect
Living being	Aggression	Attack	Decrease adrenaline
Water	Thirst	Drink	Increase vascular volume
Food	Hunger	Eat	Increase blood sugar
Abbott, Block	Curiosity	Play	Increase endorphine
Top flat block	Fatigue	Rest	Increase energy
Free Space	Cold	Walk	Increase temperature
Pain	Self-protection	Withdraw	Decrease pain

Tabela 3.3. Activação de comportamentos e respectivos efeitos fisiológicos [Cañamero, 1997].

O modelo inclui mecanismos explícitos para simulação de emoções. Esses mecanismos são caracterizados por um conjunto de parâmetros que determinam quando determinado tipo de emoção é activada. Para isso, é definido um conjunto de

emoções “primárias” ou “básicas”: zanga, aborrecimento, medo, alegria, interesse e tristeza. Estas emoções são utilizadas para monitorar e lidar com as situações que resultam da relação entre os agentes e o ambiente. A figura seguinte mostra algumas das emoções consideradas e os eventos que as iniciam.

Emotion	Triggering Event
<i>Fear</i>	Presence of enemy
<i>Anger</i>	Accomplishment of a goal menaced or undone
<i>Happiness</i>	Achievement of a goal
<i>Sadness</i>	Inability to achieve goal
<i>Boredom</i>	Repetitive activity
<i>Interest</i>	Presence of a novel object or event

Figura 3.1. Relações predefinidas entre estímulos e emoções [Cañamero, 1997].

A emoção activada influencia o mecanismo de selecção de acção de duas formas base:

- através do aumento ou diminuição da intensidade da motivação actual, por meio da libertação de sinais químicos de controlo;
- através da modificação dos parâmetros dos sensores que monitorizam as variáveis que a emoção pode afectar, alterando deste modo a percepção do estado corporal.

Apesar de haver uma caracterização discreta dos estados emocionais, são associadas a cada emoção duas propriedades contínuas:

- *valência*: indica a positividade ou negatividade da avaliação;
- *mobilização* (“*arousal*”): indica o nível de mobilização fisiológica como preparação para a acção.

3.2.1.1 Discussão

A arquitectura *Abbott* é um exemplo interessante de como mecanismos emocionais podem ser utilizados no suporte à selecção e coordenação de comportamentos em agentes autónomos. Neste caso, o modelo de emoção proposto é baseado em mecanismos de inspiração fisiológica, como é o caso de mecanismos hormonais, sendo

a emoção relacionada quer com a regulação homeostática das variáveis correspondentes, quer com estímulos ambientais predeterminados.

Este tipo de modelação levanta alguns problemas importantes. Particularmente notório é o facto do modelo biológico ser muito primitivo [Allen, 2001]. De facto, a natureza dos fenómenos fisiológicos associados aos fenómenos emocionais é de tal modo abrangente e complexa, que as abordagens desde tipo têm um carácter inerentemente *ad hoc* [Sloman, 2001]. Por exemplo, Cañamero refere que um dos principais problemas da arquitectura que propõe é o facto de ser muito difícil de afinar, em particular no que respeita às ligações entre os vários elementos da fisiologia [Cañamero, 2003, p. 17].

É igualmente problemático o relacionamento directo de emoções específicas com estados de variáveis fisiológicas ou com estímulos ambientais específicos. Como vimos anteriormente, esse tipo de modelação não é consistente com a natureza dos fenómenos emocionais. Para além disso, gera implementações muito específicas, só possíveis em ambientes simulados simples, como é o caso do ambiente *Gridland* [Cañamero, 1997], utilizado para teste da arquitectura *Abbott*.

3.2.2 Arquitectura *Cathexis*

A arquitectura *Cathexis*, proposta por Juan Velásquez [1998], baseia-se na noção de redes comportamentais, proposta por Maes e Blumberg (e.g. [Maes, 1991; Blumberg *et al*, 1996]), à qual junta mecanismos emocionais. Os elementos base da arquitectura são unidades computacionais não lineares interligadas em redes. Essas unidades estão organizadas em cinco subsistemas distintos, tal como ilustrado na figura seguinte.

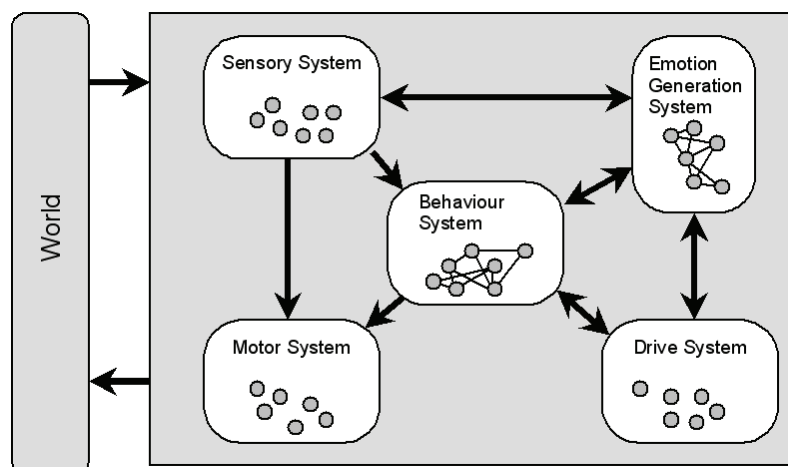


Figura 3.2. Arquitectura *Cathexis* [Velásquez, 1998]

O subsistema de percepção obtém informação do ambiente e disponibiliza essa informação para os subsistemas comportamentais e emocionais sob a forma de características dos estímulos e objectos detectados. Estes sistemas também recebem sinais de erro dos sistemas geradores de *impulsos comportamentais* (“*drives*”). Os sistemas emocionais detectam a relevância emocional dos estímulos, induzindo as respostas comportamentais futuras, bem como a percepção. O subsistema motor executa as acções indicadas pelo sistema comportamental.

Cada um dos sistemas possui um mecanismo de apreciação e um conjunto de saídas. O elemento central do mecanismo de apreciação são unidades computacionais designadas “*releasers*”, as quais realizam uma filtragem sensorial da informação que lhes chega, identificando condições específicas e produzindo reacções excitatórias ou inibitórias para os sistemas a que estão ligados, de acordo com essas condições. Todos os “*releasers*” possuem uma memória de curto prazo que permite a adaptação a estímulos com base num modelo de sensibilidade à taxa de estimulação, tal como proposto por Staddon [1983].

A resposta (nível de activação) de cada sistema base é uma função não linear das suas entradas, podendo cada entrada estar associada à saída de diferentes “*releasers*”, sendo a sua influência ponderada pelos pesos das respectivas ligações.

3.2.2.1 Sistemas emocionais

Os sistemas emocionais representam diferentes famílias de respostas afectivas, por exemplo, susto, medo, pânico, onde cada tipo de emoção corresponde a um sistema base. São consideradas seis famílias base: *zanga*, *medo*, *tristeza*, *felicidade*, *repugnância* e *surpresa*. A activação destes sistemas emocionais depende das características dos “*releasers*” que os constituem, sendo definidos quatro tipos de “*releasers*”:

- *neurais*: têm em conta o efeito de *neurotransmissores*, *hormonas* e outros elementos de natureza fisiológica;
- *sensoriais-motores*: têm em conta aspectos relacionados com a expressão emocional, como a postura corporal ou a expressão facial;
- *motivacionais*: incluem todas as motivações que levam à emoção, nomeadamente *impulsos comportamentais* (“*drives*”), por exemplo, *fome* ou *sede*; outras emoções, por exemplo, *zanga* ou *alegria*; e a regulação da dor;
- *cognitivos*: inclui todo o tipo de cognição que possa activar emoção, como é o caso da apreciação de eventos, crenças ou desejos, bem como memórias.

A activação dos sistemas emocionais difere da activação dos restantes sistemas, pelo facto de terem ligações excitatórias e inibitórias a outros sistemas e, para além disso, serem caracterizados por um decaimento temporal do seu nível de activação. Alguns dos “*releasers*” e sistemas emocionais implementados são:

- interacções com geradores de *impulsos comportamentais*: impulsos não satisfeitos produzem *aflição* e *zanga*; *aflição* é também produzida quando os impulsos são sobre-satisfeitos; a satisfação dos impulsos gera *felicidade*.
- interacções com o ambiente: os objectos rosa/vermelhos geram *felicidade*; objectos amarelos geram *repugnância*; escuridão e objectos azuis geram *medo*; ruídos altos geram *surpresa*.

No modelo proposto por Velásquez é feita a distinção entre *emoções primárias* “rápidas”, modeladas pelos sistemas emocionais base, e *emoções secundárias*, resultantes de processamento mais complexo, nomeadamente envolvendo memória. Essas emoções primárias são activadas de acordo como os respectivos “*releasers*”, como é o caso da repugnância ou do medo, correspondendo à detecção de situações que exigem atenção imediata, como é o caso de perigo, no sentido de gerar respostas defensivas que maximizem a probabilidade de sobrevivência [Velásquez, 1998].

São também consideradas misturas de emoções primárias e emoções secundárias, resultantes de memórias emocionais, inspiradas nas propostas de LeDoux [1996] e, em particular, na hipótese do marcador somático de Damásio [1994]. Essas memórias são modeladas como uma rede associativa, do tipo *K-lines* proposto por Minsky [1985], na qual estímulos relevantes (e.g. características de objectos ou outros agentes) são associados a emoções primárias. Segundo o autor, as memórias emocionais têm um duplo objectivo. Por um lado permitir o surgimento de emoções secundárias como generalizações das emoções primárias. Por outro lado, servirem como mecanismos adaptativos capazes de influenciar as decisões e comportamento futuro do agente [Velásquez, 1998].

3.2.2.2 Discussão

A arquitectura *Cathexis* explora a noção de emoção a nível de motivação comportamental e na focagem de atenção em motivações específicas. Velásquez refere, por exemplo, que “os sistemas emocionais constituem as principais motivações do agente” [Velásquez, 1998]. É também explorado o aspecto de aprendizagem através da formação de memórias emocionais.

Tal como acontece com a arquitectura *Abbott*, o modelo de emoção tem uma base fisiológica, resultando as emoções da regulação homeostática de um conjunto de variáveis como temperatura, fome ou fadiga, expressa sob a forma de *impulsos comportamentais*. Ao contrário do modelo proposto por Cañamero, não é enfatizada a existência de uma fisiologia sintética. No entanto, abstraindo os nomes das variáveis, o efeito é idêntico, sendo, por isso, idênticos os problemas. Por exemplo, ao construir um agente, é necessário especificar em detalhe os tipos emocionais considerados e a sua relação com estímulos ambientais específicos e/ou *impulsos comportamentais*.

3.2.3 Outras Abordagens de Base Fisiológica

A estreita relação entre fenómenos emocionais e fenómenos fisiológicos tem sido uma importante fonte de inspiração na construção de modelos de agentes incorporando emoção, como é o caso, para além dos trabalhos já referidos, de [Lisetti *et al.*, 2002; Gadanho, 1999; Breazeal, 2002; Aylett & Delgado, 2001; Almeida *et al.*, 2004], entre outros. Por exemplo, Sandra Gadanho [Gadanho, 1999; Gadanho, 2002] propõe uma arquitectura de agentes onde mecanismos emocionais são utilizados para guiar a adaptação dos agentes em ambientes dinâmicos.

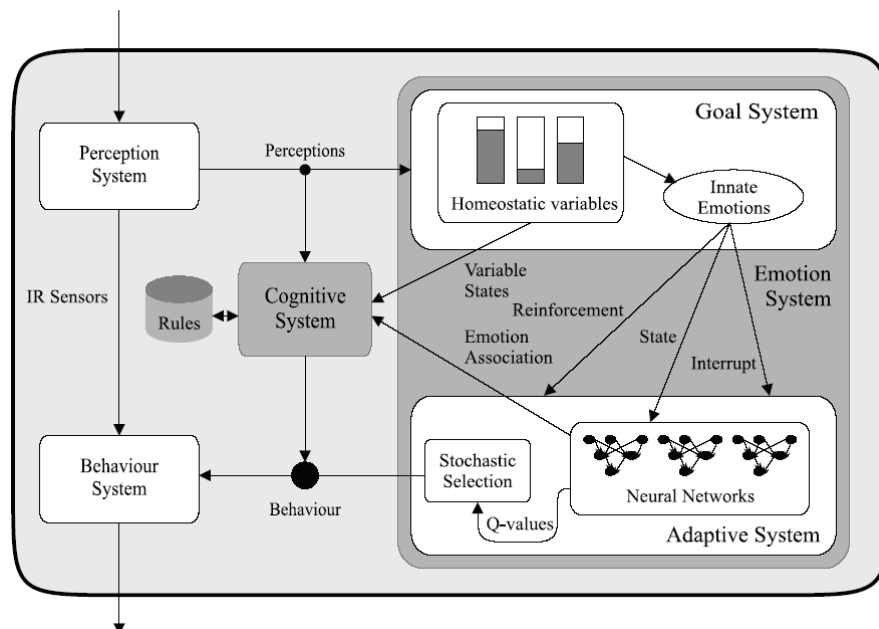


Figura 3.3. Arquitectura *ALEC* [Gadanho, 2002].

Na arquitectura *ALEC*, proposta por Gadanho, os agentes possuem emoções inatas que definem os seus objectivos, aprendendo associações entre emoções, acções e

estados do ambiente à medida que nele vão actuando. Para isso é utilizada aprendizagem por reforço com base no algoritmo *Q-learning* [Watkins, 1989]. Apesar da inspiração de base fisiológica, é igualmente definido um sistema de base cognitiva, cujo objectivo é “ajudar o sistema emocional a tomar melhores decisões” [Gadanh, 2002].

Este tipo de modelação está na linha dos modelos de apreciação multi-nível, propostos, por exemplo, por Scherer [1999], onde a apreciação emocional ocorre não apenas a nível conceptual, mas a outros níveis, nomeadamente, em termos dos mecanismos de regulação homeostática.

3.3 Modelos de Base Cognitiva

Os modelos de base cognitiva baseiam-se na perspectiva de processamento de informação inerente ao paradigma cognitivo. Nesse sentido, são caracterizados por uma grande ênfase em representações conceptuais, sendo os fenómenos emocionais resultantes de processos de apreciação emocional que operam sobre essas representações. As teorias da apreciação (ver secção 2.2.3) são, por isso, o principal suporte teórico deste tipo de modelos. Como exemplo, são discutidos nesta secção os modelos propostos por Gratch e Marsella (modelo *EMA*) [Marsella & Gratch, 2003; Gratch & Marsella, 2004], por Botelho e Coelho (arquitectura *Salt & Pepper*) [Botelho & Coelho, 2001] e por Hudlicka (arquitectura *MAMID*) [Hudlicka, 2004].

3.3.1 Modelo *EMA*

O modelo *EMA* (“*Émile Model of Emotional Appraisal*”) [Marsella & Gratch, 2003] tem por objectivo o suporte a simuladores de treino social. A arquitectura implementada é baseada na arquitectura *SOAR* [Laird *et al.*, 1987], evoluindo vários aspectos relacionados com a interacção com interlocutores humanos, nomeadamente no que se refere à geração de linguagem natural e à expressão não verbal, tal como ilustrado na figura 3.4.

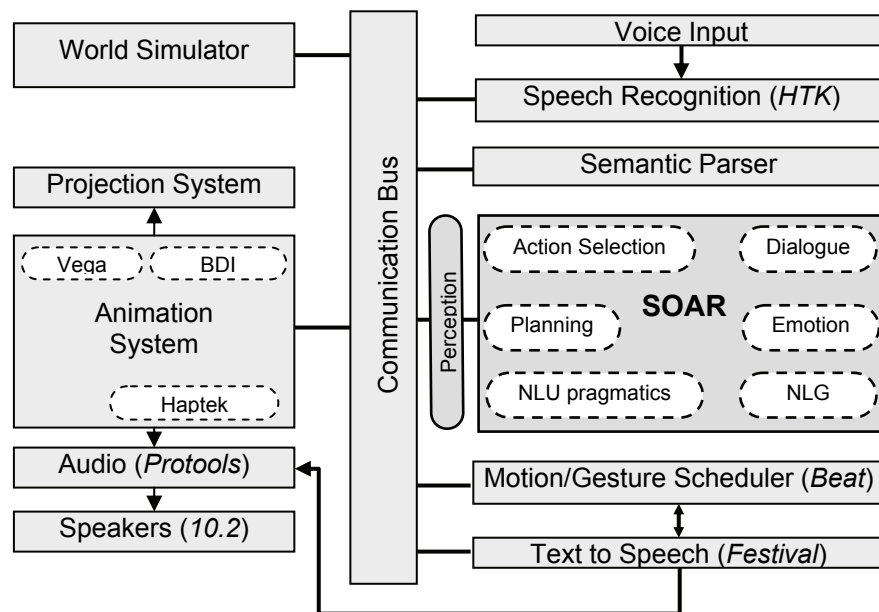


Figura 3.4. Arquitetura onde o modelo *EMA* se integra [Gratch & Marsella, 2003].

O *EMA* é um modelo computacional de emoção baseado em grande medida nas teorias cognitivas da apreciação. Um dos elementos centrais do modelo é, por isso, a noção de apreciação emocional. O modelo é inspirado pela teoria de Smith e Lazarus [2001], a qual se centra na interpretação pessoal da relação da pessoa com o ambiente. Segundo Gratch e Marsella, essa interpretação é construída por processos cognitivos, sintetizada através de variáveis de apreciação e alterada de acordo com as respostas às situações (“coping”). A ideia base que propõem é relacionar esse processo de avaliação com técnicas de planeamento e com métodos de modelação explícita de compromissos, envolvendo crenças e intenções.

Para esse efeito, os planos são utilizados como representação causal entre eventos e estados, de modo a permitir a avaliação da situação do agente de acordo com as dimensões de apreciação emocional definidas. Em particular, permitem a identificação da *relevância* dos eventos em relação aos objectivos dos agentes, a qual é uma das dimensões de apreciação emocional definidas pelos autores.

Outras dimensões de apreciação utilizadas, como o *grau de desejo* (“*desirability*”) e *verosimilhança* (“*likelihood*”), são relacionadas com as noções de utilidade e probabilidade que caracterizam os métodos de planeamento e tomada de decisão utilizados [Gratch & Marsella, 2004]. A figura seguinte ilustra a relação entre os principais aspectos do modelo proposto.

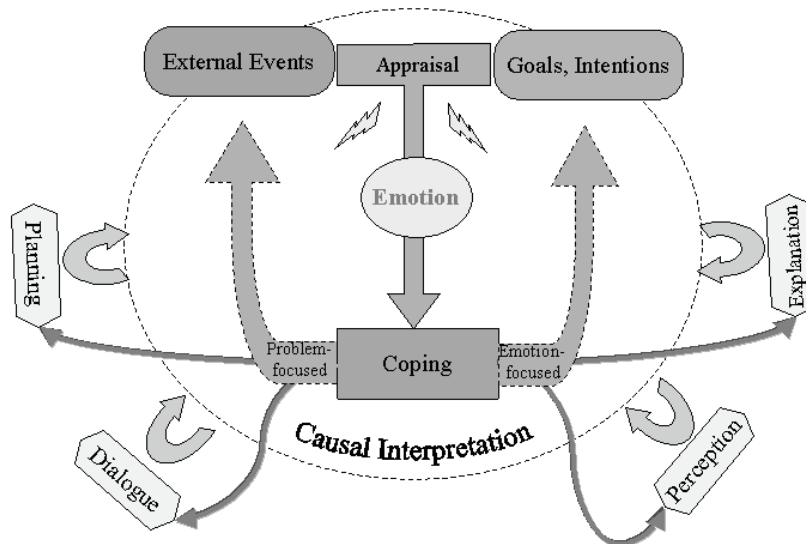


Figura 3.5. Modelo EMA [Marsella & Gratch, 2003].

Cada evento percebido é apreciado através de funções descritas como independentes do domínio, as quais examinam a estrutura sintáctica da interpretação causal envolvida de acordo com o seguinte conjunto de dimensões de apreciação:

- *perspectiva* (“*perspective*”): qual o agente sob cuja perspectiva o evento está a ser apreciado;
- *grau de interesse* (“*desirability*”): qual é a utilidade do evento na perspectiva identificada;
- *verosimilhança* (“*likelihood*”): qual a probabilidade do evento;
- *atribuição causal* (“*causal attribution*”): quem merece o crédito ou a culpa;
- *grau de controlo* (“*controllability*”): pode o resultado ser alterado pela acção do agente cuja perspectiva está a ser considerada;
- *grau de mudança* (“*changeability*”): pode o resultado ser alterável, mesmo que por outro agente.

Cada evento apreciado é mapeado numa instância de emoção de acordo com o esquema proposto por Ortony *et al.* [1988], discutido anteriormente (secção 2.2.3). Os mecanismos de apreciação propriamente ditos são modelados como conjuntos de detectores de características, os quais traduzem a interpretação causal em variáveis de apreciação estruturadas sob a forma de *enquadramentos de apreciação* (“*appraisal frames*”), tal como exemplificado na figura seguinte.

<p>Appraisal Frame: distress332</p> <p>Perspective: sergeant</p> <p>Emotion-type: Distress</p> <p>Expected utility: -59.3 (out of range -100..100)</p> <p>Intensity: 59.3</p> <p>Type: facilitator</p> <p>Annotation: progress-towards-undesired-state</p> <p>Desire-self: undesirable</p> <p>Status: confirmed</p> <p>Object: health-status (child, critical-injuries)</p> <p>Cause: collision (mom,driver)</p> <p>Evaluation: potential-blame {sergeant, mom}</p>

Figura 3.6. Exemplo de enquadramento de apreciação [Marsella & Gratch, 2002].

Outro aspecto central do modelo *EMA* é o processo de *lidar com as situações* (“*coping*”). Esse processo determina as respostas aos eventos considerados relevantes. Para isso são propostas estratégias de “*coping*” que permitem manter situações desejáveis ou alterar situações indesejáveis. As estratégias de “*coping*” operam no sentido oposto ao da apreciação emocional, identificando na interpretação causal os precursores da emoção que devem ser mantidos ou alterados, por exemplo, crenças, desejos, intenções ou expectativas. Essas estratégias incluem:

- *acção*: seleccionar uma acção para execução;
- *planeamento*: formação de intenções para realizar uma acção;
- *procura de apoio instrumental*: pedir ajuda a alguém que possa controlar os acontecimentos;
- *adiamento*: esperar por um evento externo que possa mudar as circunstâncias actuais;
- *reinterpretação positiva*: aumentar a utilidade positiva dos efeitos colaterais de uma acção com resultados negativos;
- *aceitação*: desistir de uma intenção;
- *negação*: diminuir a probabilidade de uma situação negativa latente;
- *desmobilização mental*: baixar a utilidade de um estado desejado;

- *alteração de culpa*: alterar a responsabilidade de uma acção para outro agente;
- *procurar/suprimir informação*: formar uma intenção positiva ou negativa para monitorar um estado pendente ou desconhecido.

As estratégias de “coping” geram informação que os processos cognitivos utilizam para executar as directivas indicadas, ocorrendo o processo de “coping” em várias etapas. Primeiro é percebido um determinado evento que ocorre. Esse evento dá relevância a alguma *preocupação*, *parte de plano* ou *história causal*, relacionada com o evento. A partir daí, um *enquadramento de “coping”* (“*coping frame*”) é gerado, o qual relaciona o evento, o objecto cognitivo (e.g. uma preocupação), a apreciação emocional e os elementos relevantes da situação do agente e do respectivo contexto social. A figura seguinte ilustra um desses enquadramentos.

Coping Frame:

Concern: Collision (mom.driver)

Emotion: Distress, 59.3

Focus-Event: Understand Speech

Type: Information Request

Agency: Superior

Emotion: Distress, 40.3

Responsibility:

Direct: Unspecified

Indirect: Humvee Driver, Mom

Potential: Sgt (superior of Humvee Driver)

Figura 3.7. Exemplo de enquadramento de “coping” [Marsella & Gratch, 2002].

3.3.1.1 Discussão

A perspectiva adoptada no modelo *EMA* enfatiza a componente cognitiva de apreciação emocional, numa perspectiva de processamento de informação. Gratch e Marsella salientam, no entanto, o papel central que a emoção desempenha noutros aspectos do processamento cognitivo dos agentes, nomeadamente em termos de planeamento, influenciando a determinação do curso de acção seguido, e em termos de resolução de inferências, determinando as representações consideradas, bem como no

que respeita à interação, como é o caso da geração de linguagem natural e da expressão não verbal.

A ênfase do modelo é, no entanto, a caracterização emocional, não havendo preocupações específicas com os problemas de limitação de recursos computacionais ou de restrições temporais dinâmicas, típicas de ambientes reais.

3.3.2 Arquitectura *Salt & Pepper*

Luis Botelho e Helder Coelho [Botelho & Coelho, 2001] propõem um modelo de emoção onde os fenómenos emocionais são vistos como resultando de um processo sequencial envolvendo a apreciação do estado global do agente, a geração de sinais emocionais e a geração de respostas emocionais. Um aspecto central do modelo proposto é o facto de ser feita a distinção entre apreciação afectiva e apreciação cognitiva, o que não acontece noutros modelos cognitivos, como é o caso do modelo proposto por Gratch e Marsella. O modelo proposto é implementado na arquitectura *Salt & Pepper*, cuja organização geral é apresentada na figura seguinte.

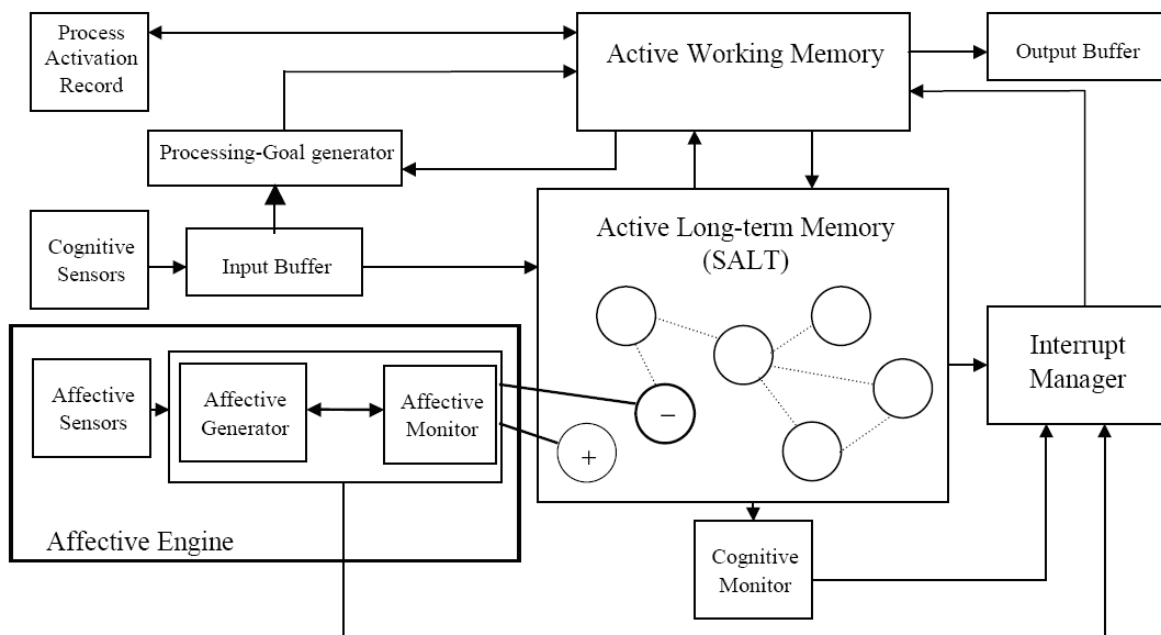


Figura 3.8. Arquitectura *Salt & Pepper* [Botelho & Coelho, 2001].

A arquitectura *Salt & Pepper* é caracterizada por três subsistemas principais: o subsistema cognitivo; o subsistema comportamental; e o subsistema afectivo e de gestão de interrupções. O subsistema afectivo inclui sensores afectivos, que detectam

aspectos relevantes para a geração de estados emocionais, um gerador afectivo e um monitor afectivo. Todos os componentes da arquitectura operam em paralelo sendo realçada a minimização do acoplamento entre subsistemas.

A arquitectura proposta inclui, também, uma memória de longo prazo, sob a forma de uma rede associativa com propagação de activação [Botelho & Coelho, 1996], o que permite associar os sinais afectivos gerados aos vários elementos envolvidos no processamento cognitivo, produzindo memórias emocionais.

No modelo de emoção proposto a emoção é caracterizada como um processo, na linha do que é sugerido por Frijda [1986] ou por Scherer [1984], tal como discutido anteriormente. Esse processo é caracterizado por diferentes etapas de apreciação, das quais resulta a geração de sinais emocionais utilizados para regular o comportamento do agente.

Na primeira etapa, um conjunto de estruturas de apreciação, ou um mecanismo alternativo específico para o efeito, é utilizado para avaliar o estado global do agente (interno e externo). Se as condições o propiciarem, é gerado um sinal emocional que informa o agente do resultado da etapa de apreciação emocional.

Os sinais emocionais gerados podem ter valência positiva ou negativa. Se o estado global do agente for de conflito com as respectivas motivações (e.g. *instintos*, *objectivos*, *intenções*, *valores*, *atitudes*), um sinal emocional negativo é gerado. Se o estado global do agente for particularmente favorável em relação às motivações do agente, um sinal positivo é gerado. Em determinadas situações poderá haver não apenas um, mas vários sinais emocionais gerados com as consequentes respostas emocionais. A experiência emocional do agente é definida como incluindo não apenas os sinais emocionais, mas também as respostas cognitivas e comportamentais que são produzidas.

Também é considerada a questão da apreciação em termos simples e de modo subconsciente, tal como proposto por esses autores. Para isso, os sinais gerados durante a etapa de apreciação emocional são dirigidos aos diferentes sistemas cognitivos e comportamentais. De acordo com a situação, uma resposta emocional pode ser uma acção reflexa ou envolver processamento cognitivo de alto nível como seleccionar um novo plano. A geração de uma resposta emocional tem como consequência a alteração do estado global do agente, num processo que é caracterizado como análogo a uma malha de realimentação, típica da teoria de controlo. A figura 3.9 ilustra as principais etapas do processo proposto.

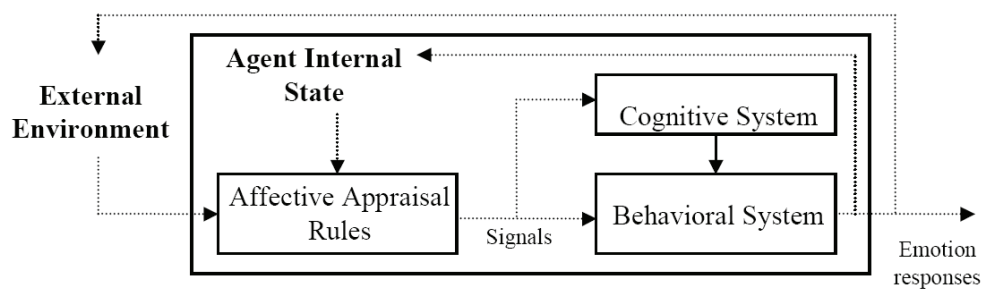


Figura 3.9. Processo de geração de emoção [Botelho & Coelho, 2001].

Em termos de classificação de aspectos emocionais são consideradas sete dimensões de classificação, tal como apresentado na tabela seguinte.

Dimensão de classificação	Exemplos / Explicação	Componente do processo
Papel/função da emoção	Aviso de mudança de atenção, avaliação de desempenho, avisos de deficiência de operação interna ou de conflito	Sinal emocional
Processo através do qual a emoção desempenha o seu papel	Ações reflexas, criação de motivações, definição de critérios de selecção de planos, intensificação motivacional	Etapa de apreciação, resposta emocional
Urgência do processo de reparação	Urgente (e.g. necessidade de atenção imediata a um aspecto do ambiente), não urgente (e.g. necessidade de melhoria a longo prazo da selecção de planos)	Sinal emocional
Intensidade	A intensidade dos sinais emocionais aumenta a possibilidade de gerarem uma resposta emocional	Sinal emocional
Valência	Negativa / Positiva	Sinal emocional
Objecto da apreciação	Ambiente externo, estado interno, eventos passados, eventos actuais, persistente, temporário	Etapa de apreciação
Tipo de apreciação	Apreciação afectiva, apreciação cognitiva	Etapa de apreciação

Tabela 3.4. Dimensões de classificação emocional [Botelho & Coelho, 2001].

Com base, por exemplo, no *papel/função* da emoção, diferentes tipos de respostas emocionais podem ser produzidas. A tabela seguinte relaciona as duas primeiras dimensões de classificação da tabela anterior, apresentando respostas possíveis que

podem ocorrer como resultado de sinais emocionais correspondentes a diferentes papéis/funções da emoção.

		Papel/Função	
Processo mediador		Aviso de mudança de atenção	Avaliadores de desempenho
	Resposta reflexa	Interromper processamento actual; iniciar novo processamento	Activar processos de aprendizagem adaptativa que evite resultados não desejados ou atinja estados desejados. Realizar uma acção que repare resultados não desejados, devido a fraco desempenho, ou mantenha estados desejados
	Criação de motivações	Interromper processamento actual; criar um objectivo de mudança de atenção diferido no tempo; reactivar o processamento actual	Gerar objectivos cujo cumprimento repare resultados não desejados, devido a fraco desempenho, ou mantenha estados desejados
	Seleção de planos	Adoptar planos com monitorização regular do ambiente externo	Adoptar planos que contenham acções que reparem resultados não desejados ou mantenha estados desejados

Tabela 3.5. Papel/função das emoções *vs.* processo mediador [Botelho & Coelho, 2001].

Os vários tipos de resposta emocional traduzem também formas de prontidão para a acção. Por exemplo, os avisos de mudança de atenção são úteis quando ocorre um evento externo, podendo representar, designadamente, uma oportunidade importante ou um risco.

Em termos de apreciação emocional propriamente dita, é dada grande importância à apreciação afectiva, no sentido de permitir uma resposta rápida, pelo que são evitadas comparações com o estado global do agente. Em vez disso, são definidas regras envolvendo as motivações impedidas ou facilitadas pelo estado global e pelos motivos do agente. A figura seguinte ilustra algumas regras apresentadas.

Explicit comparison with the survival instinct
<i>if someone risks dying, he or she will feel a lot of fear;</i> risks_dying(A, Risk) \Rightarrow activate (fear, negative, Risk) <i>if someone risks running out of food, he or she risks dying;</i> risks_running_out_of_food(A, Risk) \Rightarrow risks_dying(A, f ₁ (Risk)) <i>if someone risks running out of money, he or she risks running out of food;</i> risks_running_out_of_money(A, Risk) \Rightarrow risks_running_out_of_food(A, f ₂ (Risk)) <i>if someone loses some amount of money, he or she risks running out of money</i> loses_money(A, Loss) \Rightarrow risks_running_out_of_money(A, f ₃ (Loss))
Implicit comparison with the survival instinct
loses_money(A, Loss) \Rightarrow activate (fear, negative, f ₁ (f ₂ (f ₃ (Loss))))

Figura 3.10. Exemplo de regras de apreciação num nível afectivo [Botelho & Coelho, 2001].

Uma vez que não envolve inferências complexas, este tipo de regras é proposto como um suporte para a apreciação emocional a nível subconsciente (e.g. [Zajonc, 1980; Lazarus, 1991]).

3.3.2.1 Discussão

Apesar do modelo proposto ter uma base cognitiva, a sua definição mostra uma preocupação com um carácter mais abrangente, nomeadamente considerando a distinção entre apreciação afectiva (a qual pode não envolver representações conceptuais) e apreciação cognitiva. Ou seja, é aberta uma porta a possíveis influências resultantes não apenas de mecanismos de apreciação cognitiva, mas também a influências de carácter sub-cognitivo. Esta perspectiva permite considerar escalas de resposta temporal mais curtas, o que é um aspecto central para a operação em cenários reais.

3.3.3 Arquitectura MAMID

Um outro exemplo de modelos de base cognitiva é o proposto por Eva Hudlicka [1998; 2002; 2004]. O objectivo do modelo proposto por Hudlicka é permitir estudar o impacto da emoção na cognição, percepção e comportamento. Para isso explorou a integração de mecanismos emocionais com arquitecturas cognitivas clássicas, como a arquitectura *SOAR* [Laird *et al.*, 1987] e a arquitectura *ACT* [Anderson, 1983].

Mais recentemente propôs a arquitectura *MAMID* [Hudlicka, 2002] que mantém uma sequência de processamento clássica do tipo *percepção-planeamento-acção*

(“sense-plan-act”), mas onde foram integrados mecanismos emocionais, tal como ilustrado na figura seguinte.

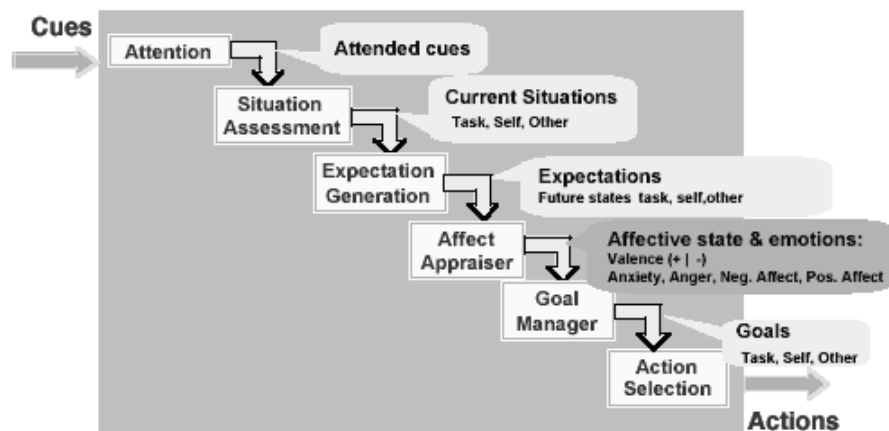


Figura 3.11. Módulos e representações cognitivas e emocionais na arquitectura *MAMID* [Hudlicka, 2004].

O ciclo percepção-planeamento-acção da arquitectura *MAMID* envolve os seguintes módulos: *pré-processamento sensorial*, que transforma os dados de entrada em representações internas; *atenção*, que filtra as representações internas, seleccionando um subconjunto para processamento; *avaliação da situação*, que integra as representações individuais numa avaliação global da situação; *geração de expectativas*, que projecta a situação corrente em possíveis estados futuros; *apreciação afectiva*, que produz um estado afectivo a partir das representações e avaliações geradas; *selecção de objectivos*, que determina quais os objectivos prioritários para execução; e *selecção de acção*, que determina as melhores acções para atingir os objectivos.

Este processo de transformação dos dados de entrada em comportamento ocorre através de uma série de estruturas intermédias de representação interna (situações, expectativas e objectivos), designadas genericamente *construções mentais* (“*mental constructs*”). Estas estruturas são suportadas por memórias de longo prazo associadas a cada módulo, implementadas sob a forma de redes de crenças ou de regras. As construções mentais são caracterizadas por atributos como *familiaridade*, *novidade*, *saliência*, *nível de ameaça* e *valência*, entre outros, os quais influenciam o seu consequente escalonamento para processamento em termos da possibilidade de virem a ser processadas.

O módulo de apreciação afectiva incorpora elementos de várias teorias da emoção, organizados em múltiplos níveis [Leventhal & Scherer, 1987; Smith & Kirby, 2001],

tendo por base diferentes tipos de estruturas de apreciação. A figura seguinte ilustra alguns dos aspectos envolvidos nessa apreciação.

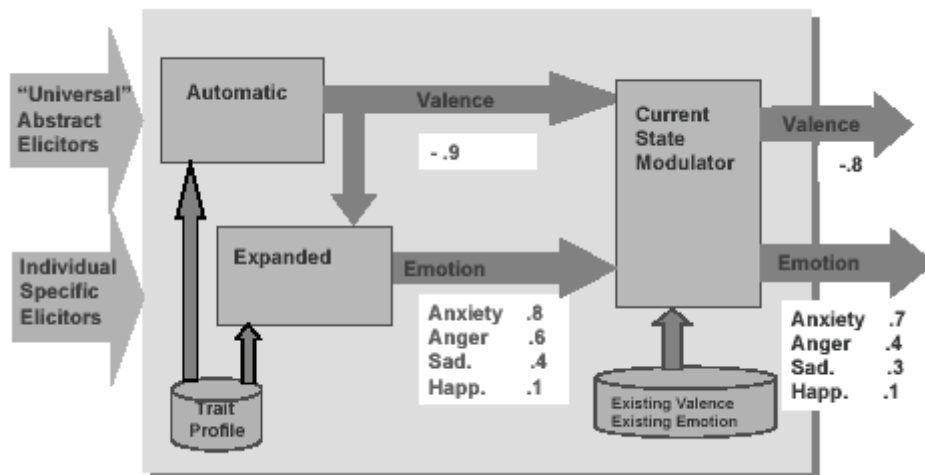


Figura 3.12. Módulo de apreciação emocional [Hudlicka, 2004].

Esta organização de apreciação multi-nível permite gerar quer apreciações de “baixa resolução”, onde os estímulos são analisados essencialmente em termos da sua valência, quer apreciações de “alta resolução” onde os estímulos são categorizados em termos de quatro emoções básicas: *ansiedade/medo*, *zanga*, *tristeza* e *felicidade*. As estruturas de apreciação respectivas utilizam dimensões de apreciação predefinidas, como *novidade*, *perigosidade* ou *agradabilidade*, para gerar apreciações afectivas em termos de valência positiva ou negativa. A nível cognitivo são definidas apreciações mais complexas, envolvendo, por exemplo, a história individual, a expectativa ou a congruência com os objectivos.

Um dos objectivos principais do modelo proposto por Hudlicka é o estudo da influência dos fenómenos emocionais em vários aspectos da actividade cognitiva, como é o caso da memória, da atenção ou dos determinantes da tomada de decisão. Essa influência é modelada através de um conjunto de parâmetros da arquitectura, tais como *velocidade* e *capacidade de processamento* dos módulos individuais. As influências emocionais nesses parâmetros são também determinadas por um análogo de *traços de personalidade* (“*trait*”), os quais determinam a influência de aspectos específicos dos estados emocionais, como taxas de decaimento ou intensidades máximas. Todas estas influências são mapeadas nos parâmetros da arquitectura que controlam o processamento em cada módulo. Os mapeamentos são feitos com base em dados experimentais sobre a relação entre fenómenos emocionais e actividade

cognitiva. Por exemplo, a redução da atenção e da capacidade de memória associadas ao medo e ansiedade, são modeladas através da redução da capacidade de atenção e da memória de trabalho dos vários módulos da arquitectura. A figura seguinte ilustra as principais relações entre estes vários aspectos.

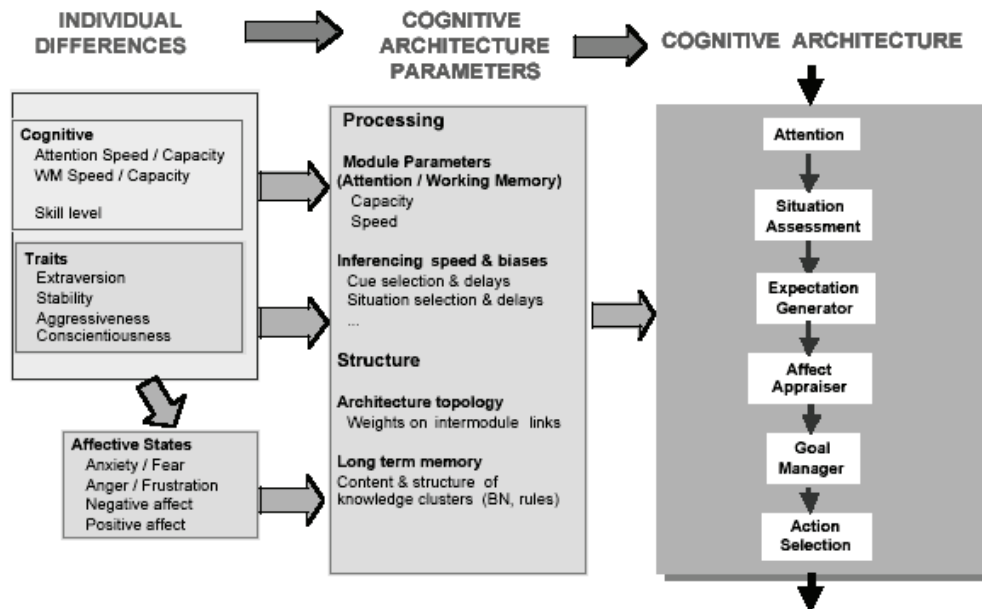


Figura 3.13. Influência de características individuais e dos fenómenos emocionais que delas resultam, sobre os parâmetros de uma arquitectura cognitiva [Hudlicka, 2004].

A partir da configuração individual dos traços de personalidade é possível definir tipos de comportamento (ansioso, agressivo, etc.), os quais são utilizados para estudar as influências emocionais nos parâmetros da arquitectura e no desempenho dos agentes.

3.3.3.1 Discussão

A arquitectura *MAMID* distingue-se das anteriores pelo facto de considerar de modo explícito o controlo da actividade cognitiva dos agentes. Como vimos anteriormente, a influência dos fenómenos emocionais em aspectos dessa actividade, como a atenção, a percepção ou a tomada de decisão, tem sido extensivamente documentada (e.g. [Williams *et al.*, 1997; Forgas, 2003]).

Uma das características salientes desta arquitectura é a multiplicidade de parâmetros que a caracterizam, bem como a heterogeneidade dos suportes representacionais. Isso origina o que Hudlicka designa como “fragilidade” do modelo, bem como uma necessidade de ajustes permanentes e muito específicos, quer dos

parâmetros da arquitectura, quer do próprio conteúdo da memória de longo prazo [Hudlicka, 2004].

3.3.4 Outras Abordagens de Base Cognitiva

Os modelos de base cognitiva têm sido uma das principais vertentes de modelação emocional em agentes inteligentes (e.g. [Elliot, 1992; Bates, 1994; Staller & Petta, 1998; Alvarado *et al.*, 2001], apesar de alguns destes modelos partilharem, também, alguma inspiração fisiológica (e.g. [Oliveira & Sarmento, 2003]).

A teoria da emoção proposta por Ortony, Clore e Collins [Ortony *et al.*, 1988], descrita no capítulo anterior, tem sido um dos principais suportes de modelos de base especificamente cognitiva. No entanto, a prática tem vindo a mostrar as limitações dessa teoria. Por exemplo, se considerarmos uma das principais áreas de aplicação deste tipo de modelos, a interacção com humanos, os resultados que se têm obtido são comportamentos estereotipados e repetitivos, tornando-se rapidamente fastidiosos, o que leva os utilizadores a desactivarem esses comportamentos. Para ultrapassar esses problemas, modelos mais recentes, como o proposto por Gratch & Marsella [2004] ou por Hudlicka [2004], apresentam desenvolvimentos no sentido das teorias de apreciação multi-nível [Scherer, 1999], anteriormente referidas.

3.4 Modelos de Base Arquitectural

A perspectiva de base arquitectural (e.g. [Sloman & Croucher, 1981; Beaudoin & Sloman, 1993; Wright, 1997; Sloman, 2001]) caracteriza-se pela definição dos fenómenos emocionais como resultantes das propriedades específicas das arquitecturas dos agentes e não como resultado de processos de apreciação cognitiva, da simulação de uma fisiologia sintética ou de processos homeostáticos. Como exemplo, são discutidos nesta secção os modelos propostos por Sloman (Arquitectura *AFP*) [Sloman, 2001] e por Ventura e Pinto-Ferreira (*Modelo de Processamento Dual*) [Ventura, 2000; Ventura & Pinto-Ferreira, 2004].

3.4.1 Arquitectura *AFP*

A arquitectura *AFP* (“*Attention Filter Penetration*”) de Sloman [2001] tem por objectivo suportar a implementação de agentes com diferentes tipos e níveis de

complexidade, em termos de controlo e de processamento de informação. Sloman explora explicitamente a analogia com a organização cognitiva humana. Nesse sentido, é postulada a existência de três camadas cognitivas com actividade concorrente, que terão evoluído em períodos diferentes nas espécies biológicas, na linha do que é proposto por MacLean com o modelo de organização cerebral triuno [MacLean, 1990]. As três camadas correspondem a diferentes tipos de processamento, não tendo nenhuma das camadas controlo total sobre a actividade cognitiva global. As várias camadas operam em concorrência, influenciando-se mutuamente. A figura seguinte dá uma ideia da arquitectura proposta.

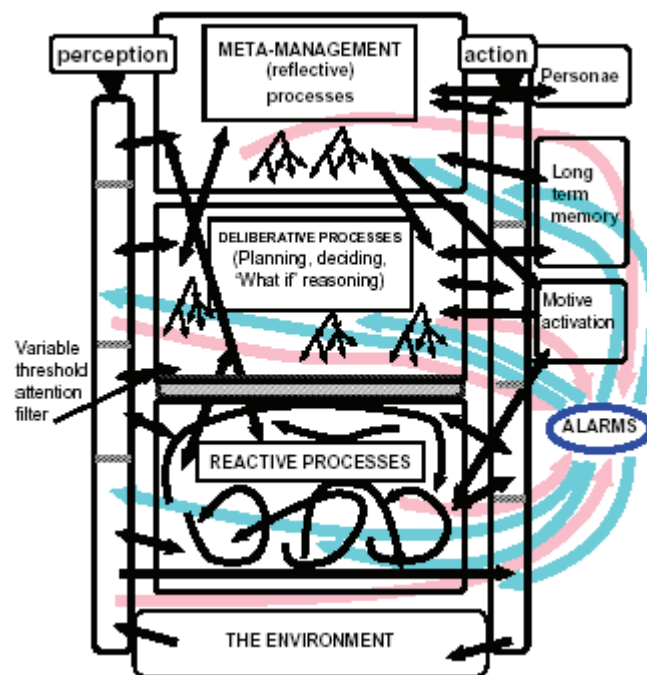


Figura 3.14. Arquitectura AFP [Sloman, 2001].

A primeira camada é constituída por mecanismos reactivos, os quais são activados de modo automático sempre que as condições apropriadas ocorram. A segunda camada é caracterizada como deliberativa, suportando as capacidades de raciocínio e planeamento necessárias à antecipação e explicação causal de acontecimentos. A terceira camada disponibiliza capacidades de meta-gestão, como a monitorização, avaliação e controlo parcial dos processos internos.

Uma vez que as várias camadas operam de modo concorrente, é possível que informação nova e relevante surja, requerendo a atenção das camadas deliberativa ou de meta-gestão, a qual deve ser tida em conta, independentemente das tarefas em

curso ou de possíveis limitações de recursos. Nesse sentido, uma das ideias centrais da arquitectura AFP é que a informação nova deve poder levar à interrupção das actividades em curso, salvaguardando, no entanto, a possibilidade de sobrecarga dos mecanismos deliberativos devido a limitações de recursos de processamento ou de memória. Para isso é proposta a noção de *filtro de atenção de limiar variável* (“*variable threshold attention filter*”) [Sloman & Croucher, 1981; Sloman, 1993; Sloman, 2004], cujo objectivo é adequar a gama de interrupções à actividade cognitiva em curso. Esta noção de interrupção está, também, na base da modelação dos aspectos emocionais.

3.4.1.1 Modelação emocional

O modelo de emoção proposto é baseado na teoria da emoção elaborada por Simon, discutida no capítulo anterior. A teoria de Simon é estendida com a noção de filtro de atenção de limiar variável, atrás referida.

A necessidade de um mecanismo de filtragem é justificada pelo facto dos recursos cognitivos serem limitados, pelo que as actividades cognitivas necessitam de ser protegidas de sobrecargas de interrupções, de modo a serem interrompidas quando for efectivamente necessário, garantido assim um comportamento adaptativo. A variação do limiar do filtro permite que essa protecção seja ajustada de acordo com o contexto.

Enquanto a teoria de Simon foca sobretudo a interrupção dos objectivos actuais, a teoria de Sloman foca as disposições para interromper o processamento cognitivo actual. Assim, no modelo proposto, os estados emocionais envolvem a capacidade de desviar a atenção do agente sem que isso implique, necessariamente, a interrupção dos objectivos ou acções actuais.

As interrupções resultam dos estados motivacionais em que o agente se encontra, sendo definidas quatro dimensões de caracterização desses estados: *insistência*, *importância*, *urgência* e *intensidade*. A intensidade corresponde à propensão para interromper a atenção, sendo considerada uma medida heurística da importância e da urgência do estado motivacional utilizada para comparação com o limiar de atenção. Por sua vez, a insistência é considerada como um estado disposicional, ou seja, traduz uma tendência para perturbar ou fazer divergir a atenção, o que pode não chegar a acontecer.

Sloman argumenta que alguns estados, ditos emocionais, possuem um potencial de perturbação ou divergência da atenção, introduzindo a noção de *perturbação*. Nesse sentido, propõe que a existência de estados perturbantes requer a existência de

mecanismos de vigilância e de auto-controlo, os quais considera típicos do nível de meta-gestão, como é o caso do filtro de atenção [Sloman, 2004]. Deste modo, os fenómenos emocionais são vistos como resultantes das propriedades de uma arquitectura de agente, nomeadamente, como resultado de mecanismos de interrupção da actividade cognitiva.

3.4.1.2 Discussão

O modelo proposto por Sloman apresenta algumas limitações importantes no que se refere à modelação de fenómenos emocionais, as quais são partilhadas pelas teorias de interrupção em geral. Nomeadamente, não apresenta uma explicação para a geração de sinais afectivos em termos de valência positiva e negativa, característicos de fenómenos emocionais, o que alguns autores designam como o problema do tom hedónico e o problema da valência (e.g. Wright [1997]).

A própria definição de emoção como um fenómeno de interrupção é muito limitativa, expressando apenas um aspecto particular e, em grande medida, extremo, da relação entre emoção e cognição.

3.4.2 Modelo de Processamento Dual

O *modelo de processamento dual* proposto por Rodrigo Ventura e Carlos Pinto-Ferreira [1998; 2001; 2004] é inspirado em grande medida pelos trabalhos de Damásio [1994] e LeDoux [1996] em termos de neurofisiologia da emoção. Nele é defendido um paradigma de representação dupla onde se propõe que os estímulos são processados sob duas perspectivas diferentes. Numa perspectiva, é extraída a partir dos estímulos uma imagem cognitiva detalhada o suficiente para permitir uma posterior reconstrução dos estímulos. Numa segunda perspectiva, é criada uma imagem perceptual que é simples e compacta, correspondendo a um conjunto reduzido das características essenciais dos estímulos, definidas como sendo as mais significativas para o agente. Por exemplo, representações vectoriais de tamanho, movimento rápido, aproximação rápida, cor dominante e outras. Estas duas perspectivas traduzem uma divisão do modelo numa *camada perceptual* e numa *camada cognitiva*, tal como apresentado na figura seguinte.

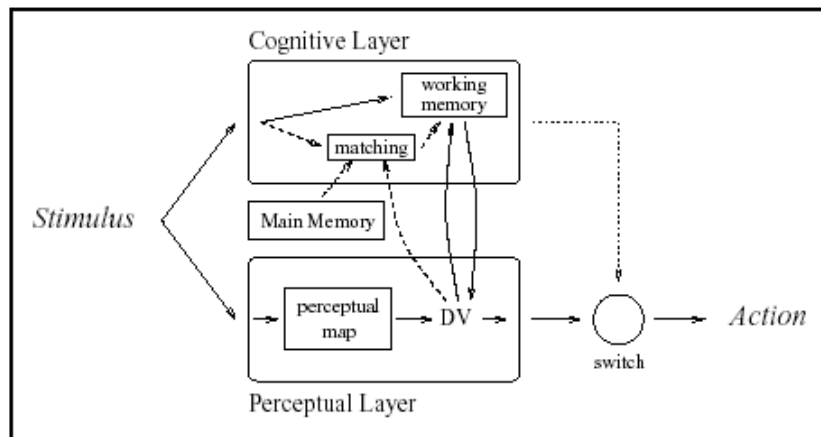


Figura 3.15. Organização do modelo de processamento dual perceptual/cognitivo [Ventura, 2000].

Na camada perceptual é igualmente formado um vector de apreciação emocional designado “*desirability vector*” (*DV*). Cada um dos componentes desse vector representa um tipo básico de avaliação de estímulos. Esses componentes podem estar activos ou neutros, podendo o seu valor variar de forma discreta ou contínua. Quando um componente é activado significa que o estímulo é avaliado em termos de ser favorável ou desfavorável. Por exemplo, um estímulo ameaçador pode activar o componente de “medo” de um vector *DV*, o que irá produzir um comportamento de medo [Ventura, 2000]. Ou seja, é este vector *DV* que expressa o carácter afectivo-emocional associado às imagens perceptuais ou cognitivas que o agente forma.

Um outro aspecto central do modelo é a existência de vários níveis de memória, que permitem a associação das imagens cognitivas e perceptuais a vectores de desejabilidade, de acordo com as circunstâncias em que ocorreram, na linha da hipótese do marcador somático proposta por Damásio [1994]. A figura seguinte mostra uma implementação da arquitectura proposta onde se ilustra a noção de marcador somático.

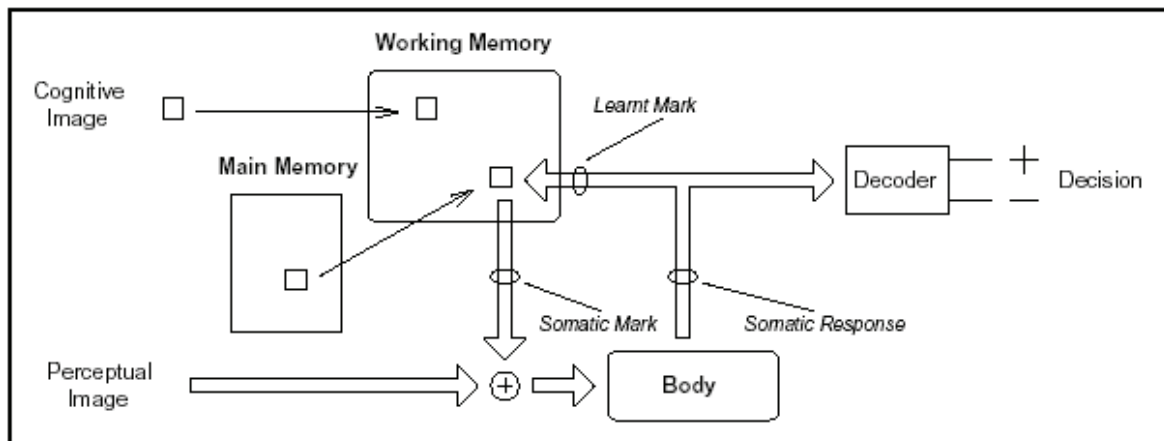


Figura 3.16. Uma implementação do modelo de processamento dual [Ventura, 2000].

Nessa implementação o agente percebe um estímulo externo através dos dois canais, um canal perceptual e um canal cognitivo. O estímulo cognitivo é recebido numa memória de trabalho de curto prazo, utilizada para recordar associações anteriores existentes na memória de longo prazo. As memórias que resultam desse processo (contendo marcadores somáticos) são combinadas com o estímulo perceptual, gerando uma resposta somática. Essa resposta somática é utilizada para gerar uma decisão de acção, mas também para actualizar, na memória de longo prazo, a associação entre estímulos e resposta somática.

Mais recentemente Ventura e Pinto-Ferreira [2004] apresentaram uma implementação da memória de longo prazo como uma memória autobiográfica, no sentido do que é proposto por Damásio com a noção “*filme-no-cérebro*” (“*movie-in-the-brain*”). Essa memória autobiográfica é um registo organizado no tempo de memórias contendo associações entre estímulos e respostas somáticas, tornando possível a um agente estabelecer relações causais entre acontecimentos [Ventura & Pinto-Ferreira, 2004].

3.4.2.1 Discussão

Apesar de explorar aspectos arquitecturais, o modelo de processamento dual tem influências quer de base fisiológica, quer de base cognitiva, expressas em grande medida pelas duas camadas propostas, tal como se pode observar numa implementação do modelo apresentada na figura 3.16.

Quando comparado com os modelos anteriores, podemos observar que partilha com alguns deles aspectos fundamentais, nomeadamente a existência de critérios de apreciação especializados, com base nos quais são gerados os componentes do vector

DV, e a possibilidade da apreciação poder envolver quer aspectos perceptuais, quer aspectos cognitivos, tal como proposto, por exemplo, por Botelho e Coelho. No entanto, no modelo de processamento dual esses dois níveis de processamento têm expressão directa sob a forma de imagens perceptuais e cognitivas distintas.

3.4.3 Outras Abordagens de Base Arquitectural

Outras abordagens à modelação de emoção em agentes inteligentes têm sido propostas que podem ser consideradas de base arquitectural, como é o caso da arquitectura *CMattie* [McCauley *et al.*, 2000] ou a abordagem proposta por Pfeifer [1993], segundo o qual a emoção é um fenómeno essencialmente emergente. Para Pfeifer, o que é importante não é uma definição de emoção, mas sim compreender quais os mecanismos subjacentes aos comportamentos caracterizados como emocionais. Nesse sentido, propõe que a emoção não é algo que esteja explicitamente embutido na arquitectura de um agente, mas sim o resultado da operação conjunta dos elementos que a constituem e cuja expressão depende da complexidade dessa arquitectura. Assim, propõe que, mesmo em agentes simples, é possível observar comportamentos de base emocional. Ou seja, a emoção está na visão do observador.

3.5 Conclusão

As várias abordagens de modelação de fenómenos emocionais no contexto de arquitecturas de agente, atrás apresentadas, reflectem linhas de orientação comuns, predominantes deste domínio de investigação. Assim, os modelos de base fisiológica assentam em arquitecturas comportamentais, caracterizadas por um nível de especialização muito elevado em relação às características do ambiente. Por seu lado, os modelos de base cognitiva assentam em arquitecturas cognitivas caracterizadas pela dependência de representações conceptuais internas, com base nas quais ocorre a apreciação emocional. Os modelos de base arquitectural apresentam aspectos de ambos os paradigmas de concepção de agentes (comportamental e cognitivo), propiciando o surgimento de arquitecturas nas quais os aspectos emocionais têm um carácter híbrido, podendo resultar quer de aspectos arquitecturais, quer de aspectos de base fisiológica ou cognitiva.

Estes modelos partilham, no entanto, outras características comuns importantes. Assim, podemos observar que, à excepção do modelo proposto por Sloman, todos os

restantes modelos assentam numa caracterização dos fenómenos emocionais em termos estruturais, discretos e tendo por base algum tipo de etiqueta afectiva ou emocional (*alegria, tristeza, medo, zanga, etc.*). Ou seja, na sua essência, estes modelos baseiam-se em alguma forma de classificação de características, consideradas relevantes para a actividade (ou objectivos) de um agente, a partir da qual são determinadas etiquetas emocionais.

De facto, esta relação entre experiência percebida e o que se pretende atingir é um elemento central da modelação de fenómenos emocionais, como vimos no capítulo anterior. No entanto, algumas questões importantes se levantam. A primeira questão tem a ver com a especificidade dos mecanismos de classificação. No caso das abordagens de base fisiológica os modelos estão claramente comprometidos com aspectos específicos da relação entre agente e ambiente. No caso das abordagens de base cognitiva existe um comprometimento com aspectos específicos das representações conceptuais que constituem o modelo do mundo, interno aos agentes. Em alguns modelos, nomeadamente os modelos propostos por Hudlicka, por Botelho e Coelho, e por Ventura e Pinto-Ferreira, a classificação emocional pode envolver diferentes contextos de actividade cognitiva, permitindo uma apreciação emocional multi-nível, no sentido do que é proposto nas evoluções mais recentes das teorias da apreciação. Por exemplo, Leventhal e Scherer [1987] fazem uma interessante síntese em termos de três níveis base de processamento de apreciação (“processing level”) e das dimensões de apreciação emocional (“appraisal dimensions”) que estão envolvidas em cada nível, tal como ilustrado na tabela seguinte.

		Appraisal dimensions				
		Novelty	Pleasantness	Goal/Need Conduciveness	Coping Potential	Norm/Self Compatibility
Processing level	Conceptual Level	Expectation: Cause/effect, probability estimates	Recalled, anticipated, or derived positive- negative evaluations	Conscious goals, plans	Problem solving ability	Self ideal, moral evaluation
	Schematic Level	Familiarity: schemata matching	Learned preferences/ aversions	Acquired needs, motives	Body schemata	Self/social schemata
	Sensory-motor Level	Sudden, intense stimulation	Innate preferences/ aversions	Basic needs	Available energy	(Empathic adaptation?)

Tabela 3.6. Dimensões de apreciação emocional e níveis de processamento [Leventhal & Scherer, 1987].

Ou seja, tem-se tornado evidente que os fenómenos emocionais não são uma propriedade de determinado tipo de processamento cognitivo específico, sendo antes um fenómeno inerente a toda a actividade cognitiva, quer simbólica quer sub-simbólica (e.g. [Scherer, 2001; Fellous, 2004]).

Uma segunda questão importante tem a ver com a natureza discreta do processo de classificação, cujo resultado são etiquetas emocionais, independentemente do tipo dessas etiquetas. É a partir dessas etiquetas emocionais discretas que são determinados padrões de actividade cognitiva e de comportamento dos agentes. A não consonância deste tipo de modelação com o que se observa das características dos fenómenos emocionais biológicos fez com que mesmo os modelos baseados em teorias cognitivas da apreciação incluíssem características como intensidade ou taxa de decaimento, tipicamente contínuas. É o caso, por exemplo, dos modelos propostos por Hudlicka ou por Gratch & Marsella. No entanto, deste tipo de abordagem resulta um problema fundamental, o qual tem em grande medida sido negligenciado: como quantificar as características contínuas associadas a etiquetas emocionais discretas. O resultado são modelos específicos, em grande medida *ad hoc* [Sloman, 2001], com as dificuldades de concepção e configuração que Cañamero [2003] e Hudlicka [2004] referem.

Não só este tipo de modelação não expressa o carácter contínuo e dinâmico dos fenómenos emocionais (ou se o expressa é de uma forma forçada), como limita e dificulta a modelação de uma relação abrangente entre fenómenos emocionais e cognitivos, reconhecida como essencial (e.g. [Scherer, 2000; Fellous, 2004; Arzi-Gonczarowski, 2002]), o que constitui uma terceira questão importante: como caracterizar de modo conciso e sistemático a relação entre fenómenos emocionais e cognitivos, sem ser de forma estereotipada e muito específica.

A dificuldade em responder adequadamente a estas questões no contexto dos actuais modelos de emoção, foi a motivação que levou ao desenvolvimento do modelo de emoção que será de seguida apresentado.

Parte II

Concretização

Capítulo 4

O Modelo Emocional de Fluxo

Neste capítulo é apresentado o modelo de emoção proposto, designado *modelo emocional de fluxo*, pela ênfase que é dada ao carácter dinâmico dos fenómenos emocionais modelados, expresso sob a forma de fluxos de diferentes tipos. O modelo proposto tem por inspiração analogias naturais, quer da biologia, quer da termodinâmica, pelo que no início será feita uma breve introdução a esses temas.

4.1 Introdução

O estudo dos fenómenos emocionais e, em particular, a concepção de modelos de emoção para a construção de agentes inteligentes, assentam, em grande medida, numa perspectiva da emoção como um fenómeno humano, eventualmente partilhado por alguns animais próximos de nós no contínuo evolutivo, como vertebrados, sobretudo mamíferos. Nos modelos de base cognitiva essa perspectiva é levada ao extremo, sendo a analogia do cérebro como um sistema de processamento de informação o elemento central das teorias. Mesmo os modelos de base fisiológica têm a sua principal inspiração na fisiologia e comportamento humanos.

A íntima ligação entre emoção, cognição e comportamento resulta em múltiplos e complexos cambiantes destes fenómenos à medida que aumenta a capacidade cognitiva dos organismos. A complexidade desses fenómenos atinge o seu auge, precisamente, nos humanos [Hebb & Thompson, 1979; Scherer, 2001]. A caracterização dos fenómenos emocionais nos humanos é por isso extremamente difícil e complexa. Compreende-se, assim, a grande diversidade de teorias e modelos de emoção que têm sido propostos e, sobretudo, o seu comprometimento com aspectos muito específicos, nomeadamente de classificação emocional, ou o facto de serem demasiado gerais para determinarem uma descrição a nível operacional, essencial para a modelação e implementação de agentes inteligentes.

A adopção de uma perspectiva antropomórfica é partilhada pelas ciências cognitivas em geral. Os modelos que caracterizam o paradigma cognitivo da inteligência artificial são disso exemplo (e.g. [Newell, 1990]). Perante as limitações desses modelos, novas abordagens têm sido propostas, como é o caso da abordagem comportamental (e.g. [Arkin, 1998; Brooks, 1999]), onde a biologia e comportamento de organismos menos complexos, nomeadamente insectos, é adoptada como analogia base, ou a abordagem dinâmica (e.g. [Port & Van Gelder, 1995; Kelso, 1999]), onde são adoptadas analogias com sistemas físicos em geral.

O surgimento destas diferentes perspectivas ilustra um problema central da modelação de sistemas inteligentes, que é a dificuldade em definir noções chave como inteligência e cognição (e.g. [Steels, 1995]). À perspectiva cognitiva clássica, onde cognição é sinónimo de estruturas cerebrais complexas, contrapõem-se evidências de comportamentos sofisticados em organismos simples (e.g. [Kühn *et al.*, 2003]), incluindo organismos unicelulares (e.g. [Staddon, 2001; Dusenbery, 1996]). A estas observações juntam-se as evidências de um contínuo filogenético no desenvolvimento das espécies que, apesar da ocorrência de transições qualitativas, não permite identificar qualquer fronteira que delimite a existência ou não existência de cognição ou de emoção.

Em alternativa à perspectiva antropomórfica, alguns investigadores têm proposto uma perspectiva significativamente diferente, na qual a cognição é inerente à vida, independentemente da complexidade organizacional dos organismos (e.g. [Maturana & Varela, 1987; Bateson, 1980]). Maturana e Varela definem cognição como “*a acção eficaz de um ser vivo no seu ambiente*”. Ou seja, cognição é caracterizada como uma propriedade global dos organismos, expressa através da capacidade de realizar a acção adequada dadas as condições do ambiente, isto é, uma acção que permite a um ser vivo continuar a sua existência num determinado ambiente.

É interessante notar a proximidade desta definição com o conceito clássico de inteligência como sinónimo de racionalidade, segundo o qual um sistema é racional se faz a “*acção certa*” dado o conhecimento que possui [Russell & Norvig, 2003]. A diferença essencial reside no forçar de um contexto cognitivo específico, que Newell [1982] refere como o *nível do conhecimento*, que a noção de racionalidade implica.

Em contrapartida, a ideia central da teoria proposta por Maturana e Varela é a identificação da cognição, o processo de conhecer, com o próprio processo da vida.

Nesta perspectiva, mesmo os organismos mais simples, como uma bactéria ou uma planta, são capazes de percepção, cognição e acção.

Este conceito de cognição é assim muito mais abrangente que a definição clássica, onde cognição implica capacidade de raciocínio e alguma forma de representação simbólica (e.g. [Newell, 1990]). Em vez disso, ao caracterizar cognição e emoção como intrínsecas ao processo da vida, torna-se possível a identificação de princípios base subjacentes a esses fenómenos, independentes de contextos cognitivos específicos, nomeadamente da cognição humana, como linguagem, raciocínio conceptual ou aspectos sociais e culturais.

Esta independência do contexto cognitivo é um aspecto central na concepção e características do modelo de emoção proposto neste capítulo. De igual modo, a generalização da noção de cognição a organismos simples permite a identificação de princípios base, doutro modo difíceis de identificar e caracterizar no complexo emaranhado que resulta da relação entre emoção e cognição em seres vivos mais complexos, nomeadamente no ser humano. É esta a perspectiva que servirá de base ao modelo de emoção proposto, que será apresentado neste capítulo.

De seguida será apresentado o enquadramento desse modelo, tendo como analogia central a noção de agente como estrutura dissipadora [Nicolis & Prigogine, 1977]. A partir desse enquadramento serão então apresentadas a estrutura e noções base do modelo, com particular ênfase na modelação do carácter dinâmico e contínuo dos fenómenos emocionais.

Como iremos ver, o modelo de emoção proposto é um modelo genérico, não comprometido com cenários ou arquitecturas de agente específicas. No capítulo 5 será apresentada a sua instanciação concreta no contexto do modelo de agente proposto, o qual também não impõe uma arquitectura rígida, permitindo a implementação de agentes de diferentes tipos e níveis de complexidade.

4.1.1 Sistemas Biológicos e Autopoiese

Uma das características chave dos seres vivos é a capacidade de continuamente se recriarem a si próprios. Cada célula incessantemente sintetiza e dissolve estruturas, numa combinação aparentemente paradoxal de mudança e estabilidade. Em organismos complexos, tecidos e órgãos substituem as suas células em ciclos contínuos mantendo, ao mesmo tempo, a consistência do organismo como um todo. Esta

capacidade de auto-criação dinâmica é designada *autopoiese* por Maturana e Varela [1987].

Num sistema autopoietico cada componente participa na criação ou transformação de outros componentes do sistema, numa rede de interdependências. Deste modo, o sistema cria-se continuamente a si próprio. Este processo começa com a própria diferenciação do organismo em relação ao meio que o rodeia através da formação de uma estrutura divisória, como é o caso da membrana celular. Essa membrana permite a existência de um meio interior diferenciado, mas também de um metabolismo que o mantém, o qual gera, por sua vez, a própria membrana celular. Ou seja, não estamos perante dois processos sequenciais, mas antes de dois aspectos distintos de um fenómeno unitário, onde não é possível separar as partes do todo. A interrupção de qualquer dos processos levaria ao fim da existência do organismo [Maturana & Varela, 1987].

Existe, assim, uma relação cíclica de *realimentação* (“*feedback*”) através da qual é mantida a estrutura interna do sistema. Cada componente do sistema afecta os outros componentes, mas esses componentes afectam, por sua vez, os primeiros componentes. Ou seja, existe uma relação de causa e efeito circular, em que qualquer mudança num componente é realimentada ao componente que lhe esteve na origem, tal como ilustrado na figura 4.1.

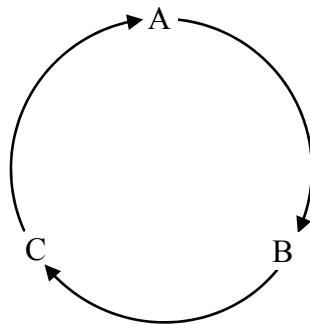


Figura 4.1. Causalidade circular numa malha de realimentação.

A realimentação pode ocorrer como uma influência recorrente que reforça ou amplifica a mudança inicial, o que se designa por *realimentação positiva*, ou no sentido contrário à mudança inicial, o que se designa por *realimentação negativa*. Deste modo, a realimentação negativa tem um carácter estabilizador, ao passo que a realimentação positiva tem um carácter auto-amplificador.

Uma característica central dos ciclos de realimentação é a capacidade de se auto-regularem. Por exemplo, quando o metabolismo de uma célula está deficitário de substâncias específicas existentes no meio exterior, a membrana celular poderá tornar-se mais permeável a essas substâncias. Em contrapartida, a presença no meio exterior de substâncias que possam danificar a membrana celular, poderá levar ao desencadear de processos metabólicos no sentido de afastar a membrana dessas substâncias (e.g. [Popa, 2004]). Os mecanismos de auto-regulação baseados em ciclos de realimentação desempenham deste modo um papel fundamental quer na manutenção de um meio interior estável dentro dos limites de viabilidade fisiológica, um processo conhecido como *homeostase* [Schulkin, 2003], quer na geração de comportamento conducente à manutenção da viabilidade global do organismo.

4.1.2 Auto-regulação, Motivação e Emoção

Os mecanismos de auto-regulação são uma característica central dos sistemas autopoieticos, desempenhando um papel de interposição entre o meio exterior e o meio interior. Ou seja, todas as variáveis que definem o estado interior são, por acção dos mecanismos de auto-regulação, tornadas independentes das variáveis que definem o meio exterior [Lorigny, 1992]. Deste modo, um sistema autopoietico é inerentemente autónomo, no sentido em que determina o que lhe é adequado [Maturana e Varela, 1987]. Essa autonomia traduz-se em comportamento pro-activo, motivado pelas dinâmicas resultantes dos seus processos auto-regulados. Os mecanismos de auto-regulação e as forças comportamentais que deles resultam são, por isso, elementos centrais dos fenómenos motivacionais (e.g. [Schulkin, 2003; Berridge, 2004]).

Este facto é patente, por exemplo, nas teorias da motivação em biologia e etologia, onde a noção de regulação fisiológica, nomeadamente o conceito de homeostase, desempenha um papel central [Barnard, 2004]. O termo *homeostase* designa a capacidade de um organismo manter o seu estado interno estável [Manuila *et al.*, 1999]. Na sua concepção clássica, isso é conseguido através de mecanismos de auto-regulação que continuamente monitorizam o ambiente, comparando os valores observados com um *valor de referência* (“*set-point*”). A diferença entre os valores observados e o valor de referência representa a motivação para produzir comportamento no sentido de reduzir essa diferença [Berridge, 2004]. Esta visão clássica de homeostase assenta na ideia de convergência dos valores observados para

os valores de referência por meio de mecanismos de realimentação negativa, sendo a realimentação positiva considerada desestabilizadora.

Nesta perspectiva, o comportamento motivado, ou pro-activo, é entendido num contexto de ser dirigido para um determinado objectivo [Tinbergen, 1951], sendo guiado por mecanismos de realimentação negativa. É esta perspectiva que tem prevalecido na modelação dos fenómenos motivacionais no âmbito da construção de agentes autónomos. Por exemplo, nos modelos cognitivos de agente a concretização de um objectivo implica uma regular monitorização da situação actual e a comparação com o objectivo pretendido (o qual serve de valor de referência), sendo gerado um comportamento no sentido da convergência para o objectivo. Esta ênfase na realimentação negativa é particularmente evidente no contexto da cibernética [Wiener, 1948; Ashby, 1964], onde o conceito de objectivo e o carácter teleológico do comportamento de agentes autónomos é modelado explicitamente em termos de ciclos de realimentação negativa [Skyttner, 2001].

De facto, enquanto a realimentação negativa tem um carácter estabilizador, trazendo os desvios do estado interno de um sistema de volta ao estado original, a realimentação positiva tem um carácter desestabilizador, induzindo a mudança de estado do sistema. Deste modo, o carácter auto-amplificador da realimentação positiva tem sido encarado como essencialmente destrutivo, em particular no que respeita aos modelos de influência cibernética.

No entanto, o que investigação em diferentes áreas tem vindo a revelar (e.g. [Nicolis & Prigogine, 1977; Heylighen, 2003]) é que a realimentação positiva desempenha um papel chave em processos de auto-organização e adaptação. Por exemplo, a noção de homeostase, que tem sido concebida em termos de realimentação negativa, tem sido questionada como sendo demasiado rígida [Schulkin, 2003]. Em alternativa é proposta a noção de *alostase* [Sterling & Eyer, 1988] para caracterizar sistemas reguladores nos quais, entre outros aspectos, os equilíbrios resultam de ciclos de realimentação negativa e positiva que interagem para a regulação do meio interno [Schulkin, 2003].

Inerente à relação entre motivação e fenómenos de realimentação em organismos biológicos está, também, a intrínseca ligação entre motivação e emoção. Por exemplo, Berridge refere que virtualmente todas as motivações de base fisiológica produzem reacções afectivas [Berridge, 2004, p.196]. O mesmo se verifica com as motivações de carácter subjectivo como desejos ou intenções (e.g. [Oatley & Jenkins, 1996; Frijda,

1986]). Assim sendo, os aspectos motivacionais, nomeadamente as dinâmicas que lhe estão subjacentes, resultantes de ciclos de realimentação negativa e positiva, deverão desempenhar um papel central num modelo de emoção, tal como acontece no modelo proposto.

4.1.3 O Paradoxo Termodinâmico

O facto dos seres vivos serem capazes de criar e manter uma estrutura organizada, longe do equilíbrio com o ambiente, está em aparente oposição com a segunda lei da termodinâmica, segundo a qual, num sistema fechado a entropia apenas pode aumentar [Nicolis & Prigogine, 1977]. Significa isto que a natureza tem uma tendência intrínseca para homogeneização, o que em termos práticos se traduz num aumento da desordem e na degradação das estruturas organizadas. Ou seja, a mudança ocorre no sentido da ordem para a desordem. Em vez disso, os seres vivos têm a capacidade de criar ordem a partir da desordem, o que vai precisamente em sentido contrário.

Uma solução para este paradoxo surgiu nos anos 60 quando Ilya Prigogine propôs a noção de *estrutura dissipadora* [Nicolis & Prigogine, 1977]. Uma estrutura dissipadora é um sistema necessariamente aberto, através do qual flui energia e matéria. Em consonância com a segunda lei da termodinâmica, uma estrutura dissipadora gera continuamente entropia, só que essa entropia é activamente dissipada, ou seja, é enviada para fora do sistema. Deste modo, uma estrutura dissipadora é capaz de gerar a sua própria organização à custa da ordem do ambiente exterior. A segunda lei da termodinâmica é assim “contornada” através da dissipação da entropia. Os exemplos mais óbvios de estruturas dissipadoras são, precisamente, os seres vivos. Plantas e animais absorvem energia e matéria de baixa entropia, sob a forma de luz ou de comida, e exportam de volta energia e matéria de alta entropia sob a forma de resíduos. Isto permite reduzir a entropia interna, contrariando assim a tendência para a degradação que a segunda lei da termodinâmica implica. No entanto, para que isso aconteça é necessário um permanente fluxo de energia e/ou matéria de baixa entropia. De facto, desde o organismo mais simples até ao homem, a manutenção da vida requer uma contínua troca de energia e matéria com o ambiente envolvente.

O fluxo constante de energia em sistemas longe do equilíbrio implica um movimento contínuo, ou fluxo, entre os componentes do sistema. Esse fluxo é

alimentado por matéria e energia de baixa entropia proveniente do ambiente que irá circular através do sistema, sendo sujeita a diferentes conversões, deixando finalmente o sistema como um fluxo de alta entropia. Por exemplo, um ecossistema desenvolve-se com base num fluxo de luz solar e de minerais que são transformados por plantas em matéria orgânica a qual, por sua vez, é convertida através de diferentes estágios, por animais, bactérias e fungos, novamente em minerais e calor. Apesar dos minerais e materiais relacionados, como a água, oxigénio ou dióxido de carbono, serem constantemente reciclados, a energia solar é dissipada e torna-se efectivamente inutilizável. De igual modo, a economia mundial utiliza materiais em bruto e energia, por exemplo derivada do petróleo, para produzir bens e os consequentes resíduos, por exemplo, dióxido de carbono e calor.

A dependência de uma fonte de energia externa torna um sistema longe do equilíbrio simultaneamente frágil e sensível a mudanças no ambiente mas, igualmente, dinâmico e capaz de reagir a essas mudanças. A fragilidade é óbvia, se a fonte de energia desaparecer a estrutura dissipadora desintegrar-se-á. Por outro lado, se a energia exterior for em grande quantidade irá amplificar os processos internos gerando reacções fortes a partir de pequenas perturbações, nomeadamente mantendo ciclos de realimentação positiva por períodos longos. Isto torna o sistema mais vigoroso no seu desenvolvimento, crescimento ou adaptação a mudanças externas. Em vez de reagir a todas as perturbações através de realimentação negativa, que traria o sistema ao mesmo estado de equilíbrio, um sistema longe do equilíbrio é, em princípio, capaz de produzir uma muito maior variedade de acções de regulação, levando a múltiplas configurações estáveis. De modo a manter uma organização particular, apesar de mudanças do ambiente, a questão que se levanta é que acção escolher e em que circunstâncias, o que corresponde ao problema da adaptabilidade.

Surgem, assim, estruturas e equilíbrios internos que necessitam de ser mantidos, caso contrário, toda a estrutura dissipadora entrará em colapso e desintegrar-se-á. Em contrapartida, se for possível manter os equilíbrios base e, eventualmente, evoluir a estrutura, esta manter-se-á.

4.2 Agente como Estrutura Dissipadora

Tendo por base a perspectiva atrás apresentada, o primeiro passo para a definição de um modelo de agente passível de descrição operacional e de implementação

concreta é a definição de um suporte que incorpore os três aspectos base característicos dos sistemas biológicos, atrás descritos, ou seja, *autopoiese*, *auto-regulação* e *motivação*. Como vimos, a noção de *estrutura dissipadora* é o suporte adequado para esse efeito.

No entanto, sendo o objectivo principal do modelo proposto o suporte à modelação de emoção em agentes inteligentes e sendo o ponto de partida do modelo a noção de estrutura dissipadora, é fundamental relacionar estas duas noções. Não existindo uma definição consensual da noção de agente, consideremos a definição proposta por Wooldridge e Jennings [1995]:

“Um agente é um sistema computacional *situado* num ambiente, capaz de *acção autónoma* nesse ambiente no sentido de atingir os objectivos para os quais foi projectado.”

(Itálico do original)

Esta definição realça dois aspectos centrais à noção de agente. A relação entre agente e ambiente, no qual se situa, e o facto de um agente ser autónomo. No contexto dos agentes inteligentes, autonomia implica pro-actividade, isto é, a capacidade de exhibir um comportamento dirigido por objectivos, tomando a iniciativa no sentido de concretizar esses mesmos objectivos [Wooldridge, 2002].

Quer a relação entre agente e ambiente, quer a autonomia assente na pro-actividade, são características base de uma estrutura dissipadora. A relação com o ambiente ocorre através de fluxos de energia de modo contínuo. A pro-actividade resulta da organização autopoietica de uma estrutura dissipadora, suportada em ciclos de realimentação. Dessa organização resultam forças motivacionais que dirigem o comportamento do agente de modo auto-determinado, no sentido da concretização das motivações que lhe estão na origem.

Deste modo, a noção de motivação abrange igualmente o conceito de objectivo, referido por Wooldridge. No entanto, vai para além disso, pois não se refere apenas à situação ou estado do mundo, que se pretende atingir e que dá direcção ao comportamento do agente, mas a todo o processo de geração do comportamento e em particular às dinâmicas motivacionais que lhe estão na origem. A noção de estrutura dissipadora propicia um suporte adequado para modelar este tipo de fenómenos de natureza essencialmente contínua, que não está disponível nos modelos de agente clássicos, quer de base cognitiva, quer de base comportamental.

Tendo por base a noção de estrutura dissipadora, um agente é caracterizado por um conjunto de potenciais internos $\{\rho_1, \rho_2, \dots, \rho_m\}$ e por um conjunto de fluxos $\{f_1, f_2, \dots, f_n\}$, tal como ilustrado na figura seguinte.

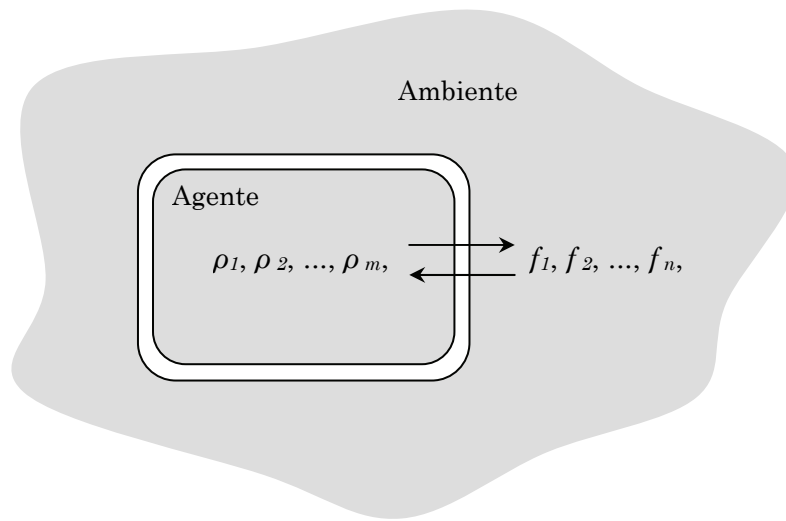


Figura 4.2. Agente como uma estrutura dissipadora.

Os potenciais internos definem a estrutura do agente, nomeadamente a sua estrutura cognitiva. Esses potenciais participam na actividade interna do agente e, simultaneamente, determinam essa actividade de acordo com a respectiva função. Ou seja, os potenciais internos variam de acordo com a actividade interna, a qual, por sua vez, é regida pela manutenção de uma determinada organização desses potenciais internos.

Deste modo, os potenciais internos são alimentados quer por fluxos energéticos provenientes do ambiente, quer por fluxos resultantes da sua regulação por meio de ciclos de realimentação. Note-se que isto não significa a maximização dos potenciais, mas sim a sua regulação no sentido de garantir a manutenção da estrutura global. Essa regulação ocorre por meio de ciclos autopoiéticos envolvendo realimentação positiva e negativa.

Como iremos ver no próximo capítulo, os potenciais internos podem formar agregados, suportando desse modo estruturas e representações complexas, as quais desempenham diferentes papéis na actividade interna. Para já, não é necessário detalhar esses aspectos. Pelo contrário, de modo a tornar o modelo simples e conciso, vamos adoptar uma perspectiva macroscópica desta noção de agente.

4.2.1 Formação de Motivações

Devido à organização autopoiética, característica das estruturas dissipadoras, onde os vários componentes da estrutura formam uma rede de interdependências cíclicas, o crescimento ou decrescimento descontrolado de um potencial pode levar ao colapso de toda a estrutura. Ou seja, a manutenção de uma estrutura interna viável, apesar de variações no ambiente, implica que os potenciais internos sejam mantidos dentro de limites de viabilidade.

É a manutenção desses limites de viabilidade por processos de auto-regulação que é considerada a fonte primária das *motivações* de um agente. Esses limites de viabilidade podem ser expressos de modo implícito, através de restrições estruturais, ou de modo explícito, sob a forma de potenciais reguladores. Estes potenciais reguladores desempenham, assim, um papel de motivadores do comportamento do agente.

Podemos distinguir dois tipos base de motivações. As *motivações predefinidas*, embutidas no agente durante a sua concepção e implementação, são conceptualmente análogas às motivações inatas, resultante de factores genéticos, que caracterizam os organismos biológicos. Para além disso, dependendo da complexidade dos agentes, podem surgir *motivações adquiridas* resultantes das motivações primárias e da interacção entre o agente e o ambiente, em particular no contexto de processos adaptativos.

Deste modo, as motivações que caracterizam um agente formam uma hierarquia causal, sendo a base, ou raiz, dessa hierarquia, as motivações predefinidas. É a satisfação dessas motivações que produz as forças directoras da actividade do agente, as quais por sua vez poderão levar à formação de novas motivações. Essas novas motivações tornar-se-ão elas próprias a base de formação de novas motivações, num processo de auto-geração típico dos sistemas autopoiéticos.

4.2.2 Concretização de Motivações

As dinâmicas motivacionais são as forças directoras do comportamento de um agente, guiando esse comportamento no sentido da concretização das motivações que lhe estão subjacentes. No entanto, para que isso possa acontecer é necessário que os agentes tenham a capacidade de produzir a mudança necessária, quer no ambiente externo, quer no meio interno. Essa capacidade de produzir mudança pode ser descrita por meio da noção de *energia*, de modo consistente com a definição

termodinâmica clássica, na qual energia representa a capacidade de produzir trabalho [Dill & Bromberg, 2003].

Desta forma, a capacidade de produzir mudança pode ser manifestada sob a forma de fluxos energéticos ou estar acumulada sob a forma de potenciais energéticos internos, os quais expressam a tendência para produção desses fluxos. Esses potenciais energéticos traduzem, assim, a capacidade potencial de um agente concretizar as suas motivações.

No entanto, um segundo factor é determinante para a concretização das motivações de um agente: as próprias características do ambiente, quer externo, quer interno. Ou seja, a actividade de um agente é determinada pela existência e características dos potenciais internos, mas também pelo favorecimento ou oposição que o ambiente (interno e externo) possa opor à expressão desses potenciais energéticos sob a forma de fluxos. Esta natureza intermutável (potencial/fluxo) da energia interna é mediada pelas estruturas subjacentes ao acoplamento entre agente e ambiente.

Em termos macroscópicos, a capacidade de produzir mudança pode ser descrita, genericamente, por um potencial P , designado *potencial de concretização*. Por sua vez, ao agir sobre o ambiente, o agente pode encontrar mais ou menos resistência ou favorecimento à mudança que está a tentar efectuar. Esse grau de favorecimento ou resistência pode ser descrito como uma condutância C , designada *condutância de concretização*.

O potencial de concretização representa o potencial de mudança que um agente é capaz de mobilizar no sentido de concretizar a situação pretendida. A condutância de concretização representa o grau de favorecimento ou resistência apresentada pelo ambiente a essa mudança. A condutância de concretização representa, igualmente, até que ponto as mudanças que estão a ocorrer no ambiente são ou não favoráveis à concretização das motivações do agente.

Numa estrutura dissipadora o potencial de concretização pode ser visto como uma força e a condutância de concretização como uma propriedade de transporte. Assim, a aplicação de um potencial de concretização P sobre uma condutância de concretização C , resulta num fluxo F , designado *fluxo de concretização*, tal como ilustrado na figura seguinte.

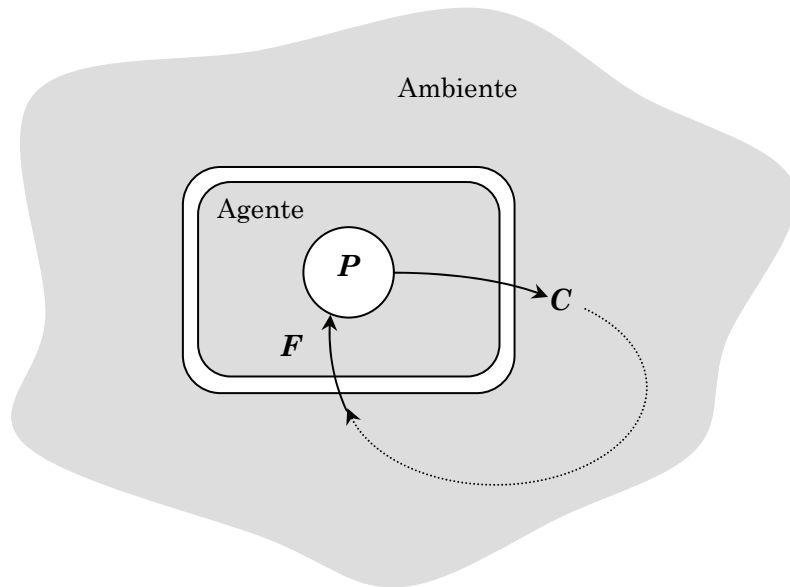


Figura 4.3. Relação entre agente e ambiente.

Deste modo, os fluxos de concretização reflectem as dinâmicas que estão subjacentes ao comportamento de um agente na sua relação com o ambiente. Esta relação entre agente e ambiente é determinante na caracterização dos fenómenos emocionais (e.g. [Lazarus, 1991]). Como iremos ver, são essas dinâmicas que constituem a raiz dos fenómenos emocionais no modelo proposto.

4.3 A Origem da Emoção

Tendo por base a noção de agente como uma estrutura dissipadora, numa relação contínua com o ambiente, determinada pelas dinâmicas motivacionais que resultam da sua actividade interna e caracterizada por potenciais e fluxos de concretização, estamos agora em condições de abordar o problema da origem dos fenómenos emocionais de acordo com a perspectiva proposta.

Os potenciais e fluxos de concretização expressam, por um lado, as dinâmicas motivacionais subjacentes ao comportamento do agente e, por outro, a relação entre agente e ambiente em termos de evolução da situação observada. Deste modo é possível identificar uma relação directa entre esses potenciais e fluxos de concretização e a expressão comportamental do agente, em cada instante, na sua relação com o ambiente. Vejamos como essa relação pode ser caracterizada.

4.3.1 Fases de Mudança e Padrões Emocionais

Como vimos anteriormente, uma das características de uma estrutura dissipadora é a permanente mudança, na qual emergem os padrões de estabilidade que caracterizam a sua estrutura e actividade interna. Podemos também identificar esses padrões no processo de concretização das motivações do agente e na sua relação com o ambiente.

Quando o potencial de concretização é elevado significa que o agente tem capacidade de produzir a mudança necessária à concretização das suas motivações. Ao contrário, se o potencial de concretização é baixo, significa que falta ao agente capacidade para produzir a mudança necessária à concretização das suas motivações.

Por sua vez, os fluxos de concretização expressam a forma como a situação do agente está a evoluir. Assim, é possível distinguir dois tipos de fluxos: *fluxos convergentes*, que expressam uma evolução da situação no sentido da concretização das motivações do agente; e *fluxos divergentes*, que expressam uma evolução da situação contrária à concretização das motivações do agente. Podemos deste modo identificar quatro padrões base de evolução da situação do agente, tal como ilustrado na tabela seguinte.

Potencial de Concretização	Fluxo de Concretização	Situação
Elevado	Convergente	Favorável
Elevado	Divergente	Adversa
Baixo	Divergente	Perigo
Baixo	Convergente	Carência

Tabela 4.1. Caracterização da situação do agente de acordo com o potencial de concretização e o fluxo de concretização.

Uma situação onde o potencial de concretização é elevado e o fluxo de concretização é convergente, é uma situação *favorável* à concretização das motivações do agente, pois o agente não só tem capacidade de produzir a mudança necessária, como a própria evolução do ambiente é conducente a essa concretização. Em contrapartida, se o potencial de concretização é elevado, mas o fluxo de concretização é divergente, significa que o ambiente está a evoluir num sentido contrário à concretização das motivações do agente, sendo, por isso, uma situação *adversa*.

No caso do potencial de concretização se tornar baixo, significa que o agente é incapaz de produzir mudança significativa. Assim, se o fluxo for divergente o agente tenderá a aproximar-se dos seus limites de viabilidade, o que corresponde a uma situação de *perigo*. No caso do fluxo de concretização ser convergente, mantendo-se o potencial de concretização baixo, significa que a evolução da situação poderá ser favorável mas o agente tem pouca capacidade de tirar partido dela, sendo uma situação de *carência* ou inércia.

Estas situações não são estados discretos mas sim diferentes fases num contínuo de mudança. Por exemplo, uma situação de carência pode resultar de situações anteriores de adversidade ou de perigo, podendo evoluir para uma situação favorável. Deste modo, uma situação de carência pode igualmente ser caracterizada como uma fase de recuperação.

O termo *situação* traduz assim uma congruência espacial e temporal passível de identificação discreta numa perspectiva macroscópica. Ou seja, as situações denotam regularidades nas dinâmicas de mudança, que poderemos designar de *padrões de mudança*.

Estes padrões de mudança envolvem não só as dinâmicas de mudança mas também o comportamento que resulta dessas dinâmicas. Esse comportamento é determinado pela natureza dos agentes, mas terá de ser adequado dadas as condições do ambiente, isto é, deverá consistir numa acção que permita convergir para a concretização das motivações do agente. É isso que caracteriza a noção de cognição tal como proposto por Maturana e Varela e que corresponde à noção de racionalidade, num sentido lato.

Nos seres vivos esses comportamentos resultam de características inatas ou são adquiridos através de processos adaptativos e de aprendizagem. É possível, no entanto, identificar padrões comportamentais típicos associados às quatro situações de concretização atrás indicadas (e.g. [Schulkin, 2003; Frijda, 1986]). A tabela seguinte mostra alguns comportamentos típicos associados a cada situação.

Situação	Comportamento
Favorável	<ul style="list-style-type: none"> • aproximar • usufruir
Adversa	<ul style="list-style-type: none"> • mobilizar • reagir
Perigo	<ul style="list-style-type: none"> • resguardar • afastar
Carência	<ul style="list-style-type: none"> • inacção • recuperação

Tabela 4.2. Comportamentos típicos de cada situação.

Ou seja, numa situação favorável o agente poderá permanecer ou aproximar-se dessa situação e usufruir do que lhe proporciona, pois permite-lhe manter ou convergir para a concretização das suas motivações. Em contrapartida, numa situação adversa o agente necessita de mobilizar os seus recursos e reagir à situação tentando alterá-la no sentido da concretização das suas motivações. No entanto, numa situação de perigo, onde a sua viabilidade pode estar em causa, o agente deverá parar de imediato e resguardar os seus recursos. Assim que a situação o propiciar, deverá afastar-se o mais rapidamente possível. Após uma situação adversa ou de perigo, o estado interno do agente poderá ter-se desviado significativamente em direcção aos limites de viabilidade, pelo que o agente necessita de minimizar actividades de usufruto no sentido de permitir a recuperação do seu estado interno, o que se traduz em inacção (inércia) e apatia.

Se compararmos as quatro situações e tipos de comportamento atrás descritos com a descrição característica de quatro emoções consideradas básicas (e.g. [Ekman & Davidson, 1994]), podemos observar as seguintes relações.

Potencial de Concretização	Fluxo de Concretização	Situação	Comportamento	Emoção
Elevado	Conducente	Favorável	<ul style="list-style-type: none"> • aproximar • usufruir 	<i>Alegria</i>
Elevado	Divergente	Adversa	<ul style="list-style-type: none"> • mobilizar • reagir 	<i>Zanga</i>
Baixo	Divergente	Perigo	<ul style="list-style-type: none"> • resguardar • afastar 	<i>Medo</i>
Baixo	Conducente	Carência	<ul style="list-style-type: none"> • inacção • recuperação 	<i>Tristeza</i>

Tabela 4.3. Relação entre padrões de mudança e emoções consideradas básicas.

Ou seja, da evolução conjunta dos potenciais e fluxos de concretização de um agente é possível observar uma relação directa e unívoca com as descrições características das quatro emoções consideradas básicas. Particularmente importante, é essa relação não ser de natureza *classificativa*, como acontece em grande medida nas teorias da apreciação ou nos modelos de base fisiológica actuais, mas antes de natureza *causal*.

Isto significa que são as próprias dinâmicas motivacionais, subjacentes à variação dos potenciais e fluxos de concretização, que produzem as condições para surgimento de determinados tipos de comportamento, regendo igualmente a expressão desses comportamentos. Deste modo, é o todo formado pela expressão das dinâmicas motivacionais, das acções a elas associadas e da percepção subjectiva que delas resulta, que pode ser caracterizado como emoção. Nesta perspectiva, os fenómenos emocionais não correspondem a qualquer tipo de representação cognitiva discreta, são antes padrões dinâmicos de evolução da estrutura e actividade de um agente, num contínuo de mudança. O que podemos designar como *padrões emocionais*.

A discriminação destes padrões emocionais como estados emocionais, típica dos modelos de emoção actuais, resulta assim da perspectiva de um observador externo ou da perspectiva do próprio agente por meio de processos auto-reflectivos. A figura seguinte ilustra alguns dos aspectos envolvidos neste processo.

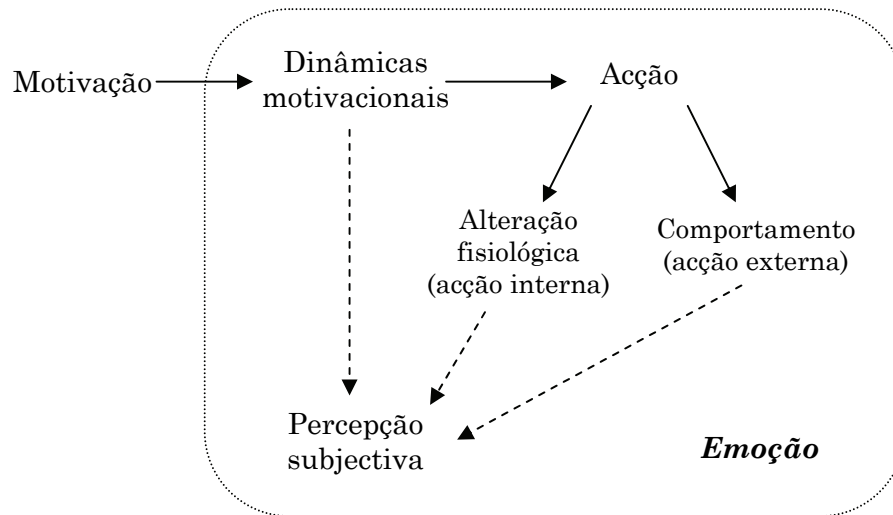


Figura 4.4. Relação entre vários aspectos envolvidos no surgimento de fenómenos emocionais.

Nesta perspectiva, os fenómenos emocionais podem ser descritos como a expressão das dinâmicas motivacionais sob a forma de acção fisiológica, comportamento observável e, eventualmente, percepção subjectiva. O termo fisiologia é aqui utilizado como analogia para referir a estrutura interna de um agente (modelada como uma estrutura dissipadora) e relacionar essa descrição com os três factores base de caracterização emocional anteriormente discutidos (ver secção 2.2.1, p. 28).

4.3.2 Dinâmicas Emocionais

Após termos caracterizado a origem dos fenómenos emocionais em termos descritivos, para tornar o modelo capaz de suportar a modelação e implementação de emoção em agentes inteligentes, é fundamental fazer a sua descrição de modo conciso e objectivo em termos operacionais.

A descrição dos fenómenos emocionais que apresentámos assenta nas dinâmicas motivacionais subjacentes ao comportamento do agente e à sua relação com o ambiente. Essas dinâmicas são expressas sob a forma de variações dos potenciais e dos fluxos de concretização, sendo a origem primária do surgimento de fenómenos emocionais. É nesse sentido que serão designadas *dinâmicas emocionais*. São essas dinâmicas emocionais que constituem o suporte para a descrição dos fenómenos emocionais no modelo proposto.

Podemos, assim, identificar duas dinâmicas base susceptíveis de explicitamente influenciar um agente. Por um lado a variação do potencial de concretização do agente, por outro lado, a variação dos fluxos de concretização resultantes das acções

de concretização das motivações do agente. Formalmente, podemos expressar essas dinâmicas em termos da variação temporal do potencial de concretização (δP) e da variação temporal do fluxo de concretização (δF). Ou seja:

$$\delta P = \frac{dP}{dt} \quad \text{e} \quad \delta F = \frac{dF}{dt} \quad (4.1)$$

Cada uma destas dinâmicas, por si só, é suficiente para determinar comportamentos específicos do agente. Por exemplo, uma diminuição do fluxo de concretização é indicadora da existência de situações adversas, implicando a necessidade de alteração do comportamento do agente para lidar com a adversidade. Em contrapartida, uma diminuição do potencial de concretização do agente, é indicadora de uma menor capacidade de acção, implicando, por exemplo, uma alteração na gestão dos recursos internos do agente ou no tipo de comportamento expresso.

Deste modo, estas duas dinâmicas emocionais são, em grande medida, complementares podendo, no entanto, influenciar-se mutuamente. Para expressar esse carácter integrado, são descritas em conjunto por uma função vectorial DE , designada *disposição emocional*, definida da seguinte forma:

$$DE \equiv (\delta P, \delta F) \quad (4.2)$$

Sendo a disposição emocional uma função que evolui no tempo (t) de acordo com o comportamento do agente e a sua relação com o ambiente, podemos observar que, para um dado instante $t = \tau$, um vector de disposição emocional DE é caracterizado por uma *qualidade*, definida pela orientação do vector (o seu argumento), e uma *intensidade*, definida pela dimensão do vector (o seu módulo). Ou seja:

$$qualidade(DE) \equiv \arg(DE) \quad (4.3)$$

$$intensidade(DE) \equiv |DE| \quad (4.4)$$

A noção de disposição emocional é assim caracterizada por dois aspectos distintos:

- *qualidade*: traduz o carácter da disposição emocional, de acordo com os padrões emocionais atrás referidos;
- *intensidade*: traduz a intensidade da “força” da disposição emocional em termos do seu carácter regulador de comportamento.

Estes dois aspectos são ilustrados na figura 4.5.a.

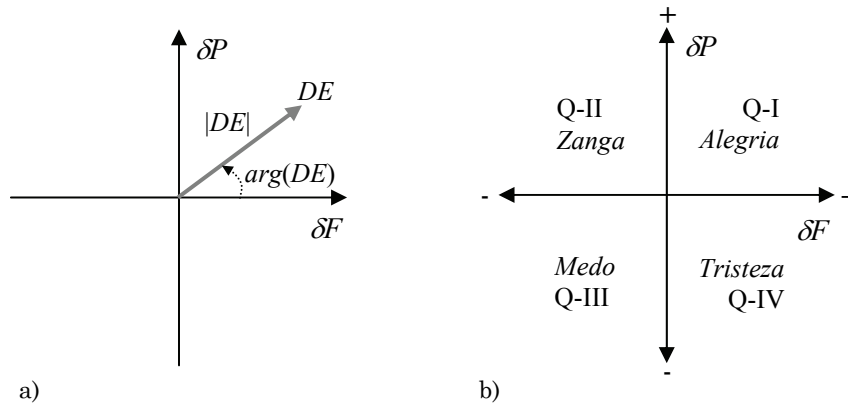


Figura 4.5. Vector DE como uma função de δP e de δF (a); relação entre os quadrantes de disposição emocional e a qualidade emocional de DE (b).

Numa perspectiva de caracterização emocional, é possível relacionar o tipo de dinâmicas emocionais predominantes em cada quadrante do espaço bidimensional $\delta P \times \delta F$ com os quatro padrões emocionais base anteriormente descritos, tal como ilustrado na figura 4.5.b.

No quadrante Q-I ($\delta P > 0$ e $\delta F > 0$) os fluxos de concretização têm um carácter convergente com as motivações do agente ($\delta F > 0$) e o potencial de concretização tem uma tendência igualmente crescente ($\delta P > 0$), o que traduz uma evolução favorável da situação, directamente relacionável com um padrão emocional de *alegria/satisfação*.

No quadrante Q-II ($\delta P > 0$ e $\delta F < 0$) os fluxos de concretização têm um carácter divergente das motivações do agente ($\delta F < 0$), o que é indicador de uma situação adversa. No entanto, o potencial de concretização tem uma tendência crescente ($\delta P > 0$), o que indica uma capacidade de mobilização de recursos para reagir à adversidade, o que é típico de um padrão emocional de *frustração/zanga/ira*.

No quadrante Q-III ($\delta P < 0$ e $\delta F < 0$) não só os fluxos de concretização têm um carácter divergente das motivações do agente ($\delta F < 0$), o que é indicador de uma situação adversa, como o potencial de concretização tem uma tendência decrescente ($\delta P < 0$), o que indica que o agente não tem capacidade de lidar com essa situação adversa, típico de situações de *ansiedade/medo/pânico*.

O quadrante Q-IV ($\delta P < 0$ e $\delta F > 0$) é particularmente interessante pois, ao contrário dos outros padrões emocionais, a *tristeza* é de difícil caracterização. Permite, por isso, verificar a adequação do modelo à descrição do fenómeno que se está a modelar. De acordo com o modelo proposto, a tristeza deverá corresponder a situações onde, apesar de não existir oposição directa à concretização, ou seja, o fluxo de

concretização é convergente ($\delta F > 0$), existe uma diminuição do potencial do agente resultante, por exemplo, de uma perda de recursos, o que se traduz numa diminuição do potencial de concretização ($\delta P < 0$). Esta caracterização é consistente com o padrão emocional de *tristeza*, caracterizado por um comportamento essencialmente apático, mesmo em situações favoráveis. Observações experimentais indicam que uma das causas importantes do surgimento de padrões de tristeza é a ocorrência de perdas súbitas (e.g. [Lazarus, 1991; Scherer, 2001]), o que pode ser descrito como uma diminuição súbita do potencial de concretização P , originando uma forte variação da dinâmica associada δP .

É importante salientar alguns aspectos importantes resultantes desta caracterização qualitativa da noção de disposição emocional. As tendências emocionais associadas a cada quadrante (*alegria*, *zanga*, *medo* e *tristeza*) são apenas indicativas da natureza essencial das disposições emocionais em cada quadrante, pois a qualidade da disposição emocional é contínua. O que é consistente com a multiplicidade de cambiantes emocionais que caracteriza os fenómenos emocionais. Deste modo é possível modelar a ocorrência de padrões emocionais básicos sem, contudo, comprometer o carácter contínuo e multifacetado da expressão emocional.

Para além disso, é importante notar que as disposições emocionais não são estados discretos no tempo, mas sim dinâmicas que evoluem ao longo do tempo. Assim, apesar de uma tendência emocional de *tristeza* poder ser associada ao quadrante Q-IV, na prática essa dinâmica poderá resultar de uma situação adversa que levou a disposição emocional do agente a passar pelos quadrantes correspondentes à *frustração/zanga* e, eventualmente, *medo*, evoluindo posteriormente para uma situação de *tristeza*.

Também este aspecto é consistente com os fenómenos emocionais biológicos, nos quais o cessar de um estímulo ou situação não produz uma mudança imediata proporcional à mudança da situação, mas antes uma evolução emocional diferida no tempo (e.g. [Scherer, 2000]).

É também importante notar que a noção de disposição emocional não constitui em si um análogo directo da noção de emoção, constituindo antes uma disposição indutora de acção, num sentido análogo a uma predisposição/prontidão para a acção, tal como salientado por Frijda [1986]. Como vimos anteriormente, emoção implica um âmbito mais alargado no qual é preciso ter em conta os contextos cognitivos em que as disposições emocionais ocorrem, tal como iremos abordar mais à frente.

4.3.3 Situação Emocional

As disposições emocionais resultam quer da interacção entre agente e ambiente, quer da actividade interna do agente, reflectindo potenciais e fluxos instantâneos, os quais produzem efeitos igualmente instantâneos sobre a actividade cognitiva do agente. No entanto, é o resultado cumulativo que esses efeitos produzem que é observável em termos macroscópicos numa escala temporal mais alargada.

Particularmente importante é o efeito cumulativo dos fluxos de concretização, uma vez que esses fluxos são indicadores da evolução da situação do agente em relação à concretização das suas motivações. Por outro lado, em termos termodinâmicos, um dos processos básicos numa estrutura dissipadora é a formação de potenciais internos a partir de fluxos de proveniências diversas. Assim, a partir de um fluxo de concretização genérico F , podemos caracterizar a formação de um potencial cumulativo P_{sit} que reflecte a *situação de concretização* do agente e que corresponde à acumulação ao longo do tempo do fluxo de concretização F . Ou seja:

$$\frac{dP_{sit}}{dt} = \alpha_{sit} \cdot F \quad (4.5)$$

onde α_{sit} é um coeficiente que determina a sensibilidade de P_{sit} ao fluxo F . O potencial P_{sit} representa, assim, o resultado cumulativo das mudanças ocorridas na situação de um agente em relação à concretização das respectivas motivações, sendo por isso designado *potencial de situação*.

Por sua vez, o fluxo F expressa o efeito cumulativo da dinâmica motivacional δF (definição 4.1), a qual pode ser vista como um impulso de mudança. Assim, o par (P_{sit}, F) expressa o efeito cumulativo das dinâmicas emocionais ao longo do tempo, o que podemos designar como *situação emocional* de um agente, descrita por uma função vectorial SE , definida da seguinte forma:

$$SE \equiv (P_{sit}, F) \quad (4.6)$$

A situação emocional pode ser vista como uma base para a emergência de um estado emocional ou para a modelação da noção de emoção como um fenómeno global à actividade cognitiva do agente. No entanto, por si só não pode ser descrita como tal, pois não abrange aspectos como a expressão comportamental ou a percepção subjectiva, os quais, como vimos anteriormente, são fundamentais para caracterizar a noção de emoção.

No entanto, o carácter cumulativo da situação emocional faz com que desempenhe um papel importante, quer a nível de mecanismos de regulação da actividade cognitiva, como será descrito no próximo capítulo, quer a nível da caracterização do estado global do agente, nomeadamente para efeitos de comunicação com interlocutores humanos ou com outros agentes.

4.3.4 Caracterização da Situação Emocional de um Agente

Tal como acontece com a noção de *disposição emocional*, também a noção de *situação emocional* é caracterizada como uma função que evolui no tempo (t) de acordo com o comportamento do agente e com a sua relação com o ambiente. Assim, num dado instante $t = \tau$, um vector de situação emocional SE é caracterizado por uma *qualidade*, definida pela orientação do vector (o seu argumento), e por uma *intensidade*, definida pela dimensão do vector (o seu módulo).

De igual modo, é possível relacionar o tipo de situações emocionais predominantes em cada quadrante do espaço bidimensional $P_{sit} \times F$ com os quatro padrões emocionais base anteriormente descritos (secção 4.3.1).

Assim, no quadrante Q-I ($P_{sit} > 0$ e $F > 0$) o fluxo de concretização é convergente com as motivações do agente ($F > 0$) e o potencial de situação indica uma situação favorável ($P_{sit} > 0$), directamente relacionável com um padrão emocional de *alegria*.

No quadrante Q-II ($P_{sit} > 0$ e $F < 0$), apesar do potencial de situação indicar uma evolução favorável até ao instante actual ($P_s > 0$), o fluxo de concretização tem um carácter divergente das motivações do agente ($F < 0$), o que é consistente com um padrão emocional de *frustração/zanga/ira*.

No quadrante Q-III ($P_s < 0$ e $F < 0$) não só o fluxo de concretização tem um carácter divergente das motivações do agente ($F < 0$), como o potencial de situação indica uma situação desfavorável ($P_{sit} < 0$), a que correspondem situações de *ansiedade/medo/pânico*.

No quadrante Q-IV ($P_{sit} < 0$ e $F > 0$) o fluxo de concretização é convergente ($F > 0$), no entanto o potencial de situação é negativo, o que indica afastamento das motivações e pouca capacidade de concretização, por exemplo perda de recursos, o que é consistente com o padrão emocional de *tristeza*.

4.3.4.1 Exemplo de Caracterização Emocional de um Agente

Para ilustrar o modo como as noções disposição emocional e situação emocional permitem caracterizar um agente na perspectiva de um observador externo, consideremos, como exemplo, um agente que é sujeito a diferentes graus de adversidade. A simulação apresentada é muito simples e tem por único objectivo tornar clara a aplicação dos conceitos atrás descritos. A aplicação do modelo em cenários concretos será feita nos capítulos seguintes.

O agente é caracterizado por um determinado potencial de concretização P e por um potencial de situação P_{sit} . No sentido de modelar a capacidade de um agente tirar partido da evolução da situação, considera-se que o potencial de concretização varia de acordo com o potencial da situação, ou seja, existe um fluxo entre P_{sit} e P , mediado por uma condutância C_P , tal como ilustrado na figura seguinte.

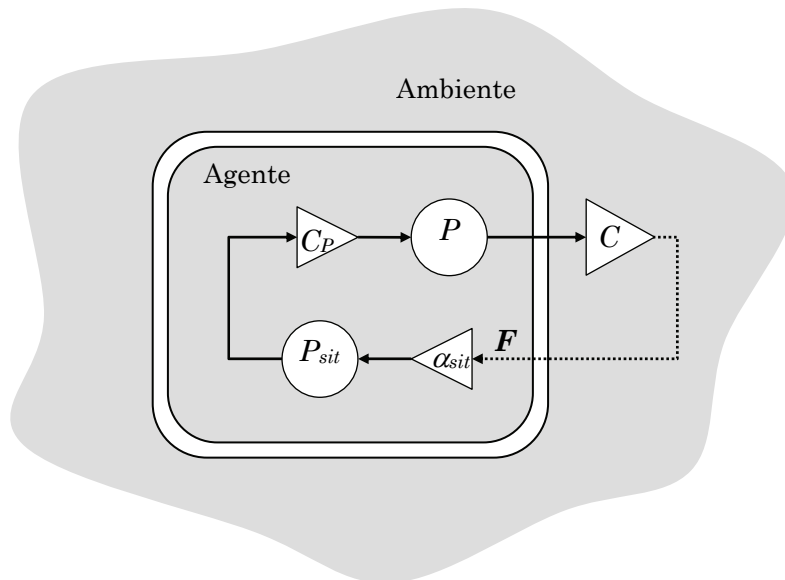


Figura 4.6. Relação entre os potenciais e condutâncias que caracterizam o agente e a sua relação com o ambiente.

Por sua vez, o potencial de concretização P é aplicado sobre o ambiente, cujas características são representadas pela condutância C . Dessa interação resulta um fluxo de concretização F . Esse fluxo de concretização dá origem à formação de um potencial de situação P_{sit} , cuja sensibilidade ao fluxo F é determinada pelo coeficiente

de sensibilidade α_{sit} . A evolução dos potenciais P , P_{sit} e do fluxo F é assim determinada da seguinte forma:

- $F = C \cdot P$ (considera-se que o fluxo gerado é proporcional ao potencial aplicado; note-se, no entanto, que em situações reais a condutância C tem um carácter tipicamente não linear)
- $\frac{dP}{dt} = C_P \cdot P_{sit}$ (evolução do potencial de concretização)
- $\frac{dP_{sit}}{dt} = \alpha_{sit} \cdot F$ (evolução do potencial de situação - definição 4.5)

São parâmetros fixos da simulação:

- o valor inicial $P(0)$ do potencial de concretização P ;
- os valores do coeficiente de sensibilidade α_{sit} e da condutância C_P , os quais foram considerados constantes.

Na simulação realizada foram utilizados os seguintes valores para os parâmetros fixos da simulação: $P(0) = 1$; $C_P = 0.1$; $\alpha_{sit} = 1$. A configuração destes parâmetros pode ser relacionada com aspectos como *personalidade* ou *temperamento* dos agentes, tal como será discutido mais à frente. A evolução do sistema foi simulada através do método de diferenças finitas (e.g. [Wylie & Barret, 1982]).

Nas simulações realizadas foram considerados dois cenários com graus de adversidade diferentes. Os graus de adversidade são expressos sob a forma de condutâncias de concretização variáveis, desde situações muito favoráveis ($C = 1$), até situações muito desfavoráveis ($C = -1$), onde C representa a condutância de concretização normalizada no intervalo $[-1,1]$. Note-se que condutâncias negativas correspondem a uma evolução contrária às motivações do agente, produzindo fluxos divergentes.

A figura 4.7 ilustra a evolução de condutância ao longo do tempo para os dois cenários considerados: um cenário de adversidade limitada (a) e um cenário de grande adversidade (b).

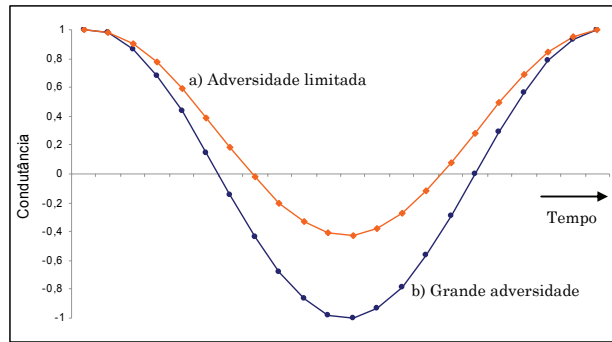


Figura 4.7. Variação da condutância de concretização ao longo do tempo para os dois cenários considerados: a) cenário de adversidade limitada; b) cenário de grande adversidade.

Em cada instante a situação emocional do agente foi avaliada. Os resultados para ambos os cenários são apresentados na figura 4.8.

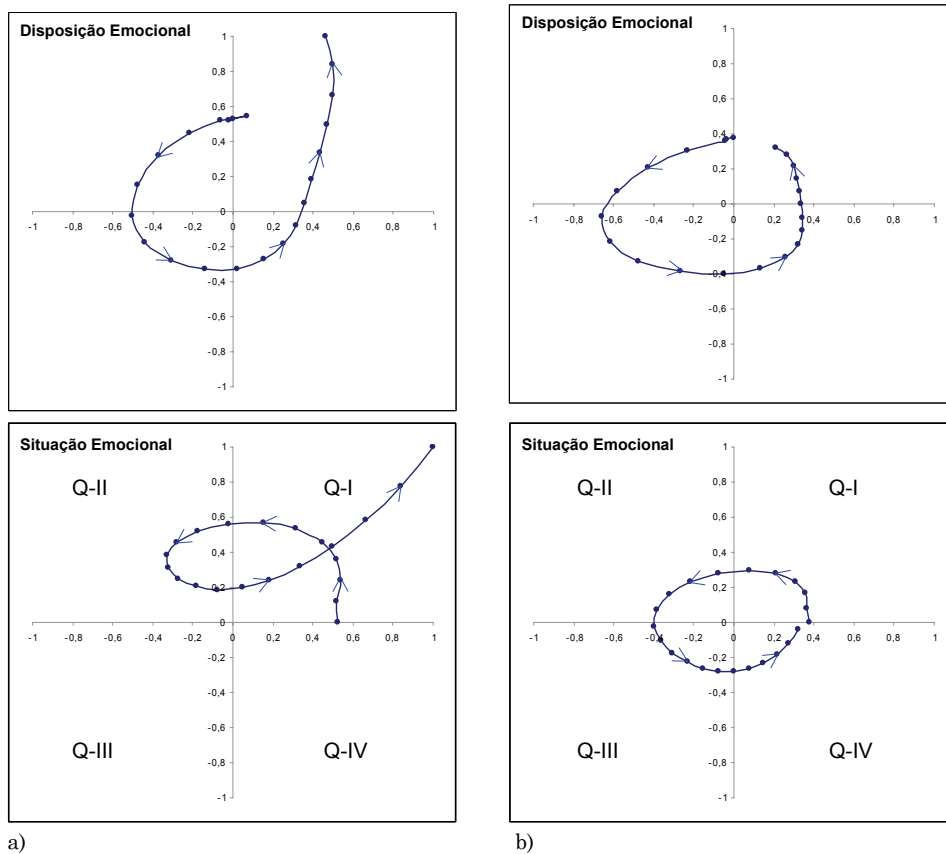


Figura 4.8. Evolução da disposição emocional e da situação emocional do agente: (a) cenário de adversidade limitada; (b) cenário de grande adversidade (valores normalizados no intervalo $[-1,1]$).

No cenário de adversidade limitada (a), à medida que a condutância varia de valores positivos para valores baixos e negativos, a *situação emocional* ganha uma tendência clara de *zanga* (Q-II). Quando a condutância torna a ser positiva, o agente retorna a uma situação emocional com um carácter de *alegria/satisfação* (Q-I). Note-se que a *disposição emocional* subjacente chega a expressar uma tendência de *medo*

(Q-III) durante o período de adversidade. No entanto, essa tendência de medo não tem a intensidade necessária nem perdura o tempo suficiente para se reflectir na situação emocional. Se a condutância de concretização for tornada mais negativa (situação mais adversa), como acontece no segundo cenário (b), a tendência de medo da disposição emocional intensifica-se, reflectindo-se na situação emocional que apresenta igualmente uma tendência de medo (Q-III) e, antes que o agente regresse novamente a uma tendência de *alegria*, o quadrante Q-IV é cruzado, a que corresponde um carácter de *tristeza/apatia*, resultante da adversidade observada.

Este exemplo, apesar de simples, permite observar claramente o carácter contínuo e dinâmico do modelo proposto. Tal como referido anteriormente, as tendências emocionais utilizadas para descrever o carácter emocional de cada quadrante são apenas indicativas da natureza essencial das disposições emocionais, bem como da situação emocional que delas resulta pois, como se pode observar na figura 4.8, a sua evolução é contínua.

É igualmente possível modelar a ocorrência de padrões emocionais básicos sem comprometer o carácter contínuo e multifacetado da expressão emocional, tal como ilustrado na figura seguinte.

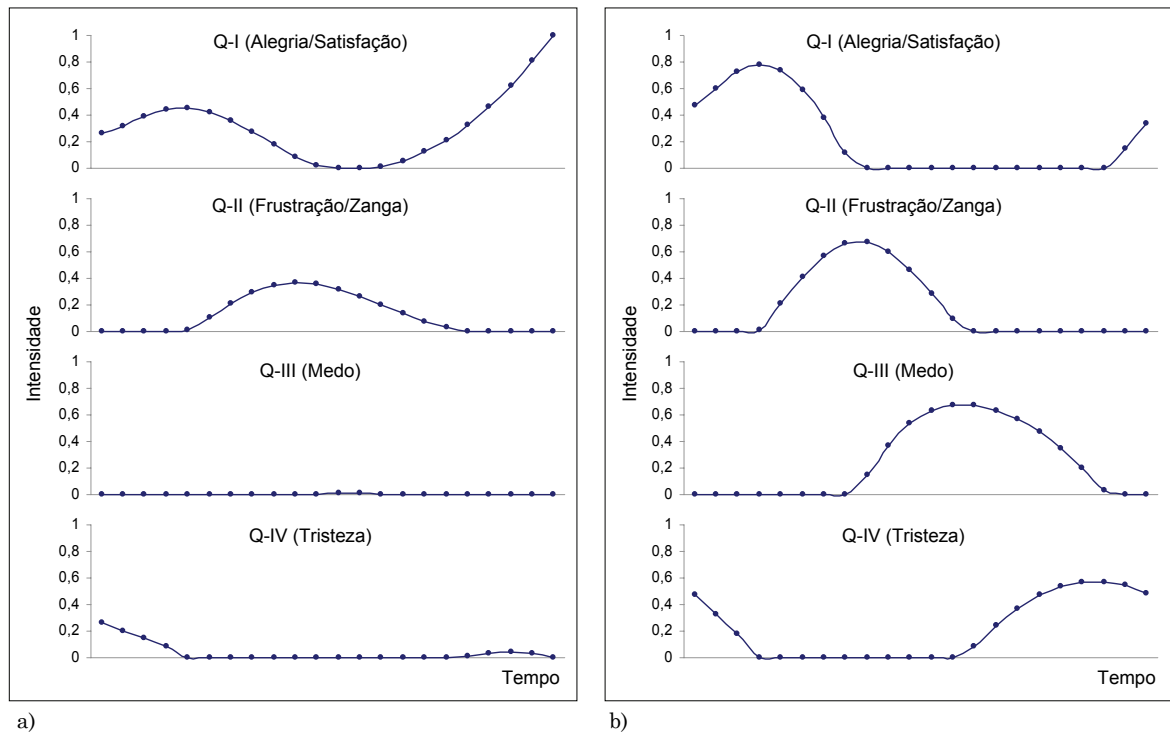


Figura 4.9. Evolução das componentes da situação emocional nos vários quadrantes (projectões do vector *SE* sobre os vectores de referência de cada quadrante): (a) cenário de adversidade limitada; (b) cenário de grande adversidade (intensidade normalizada no intervalo [0,1]).

Na figura 4.9 é apresentada a evolução das componentes da situação emocional nos vários quadrantes. Essas componentes são obtidas pela projecção do vector *SE* sobre os vectores de referência de cada quadrante (vectores unitários com a direcção da bissectriz do quadrante). Como se pode observar, é possível discriminar padrões específicos de evolução, os quais podem dar origem a uma percepção subjectiva da sua situação emocional em termos de conceitos discretos como *alegria* ou *medo*. Por exemplo, no cenário de adversidade limitada (a) podem observar-se três picos distintos: um primeiro pico de *alegria/satisfação* pela evolução favorável da situação, a que se sucede um pico de *frustração/zanga* pela adversidade encontrada e, de seguida, um novo pico de *alegria/satisfação* pelo retomar da evolução favorável. Em contrapartida, no cenário de grande adversidade (b) predominam os padrões de *medo* e de *tristeza*, resultantes da grande adversidade observada.

Pode observar-se, também, que a disposição emocional e a situação emocional ocorrem como padrões de mudança (ver secção 4.3.1) que evoluem no tempo e não como estados discretos. Como iremos ver, é o tornar explícito deste carácter dinâmico e contínuo dos fenómenos emocionais que proporciona o suporte essencial para uma integração estreita e simbiótica entre fenómenos emocionais e cognitivos. Por sua vez, é dessa integração simbiótica que resultam as capacidades adaptativas e de resposta em tempo-real que caracteriza o modelo de agente que será apresentado nos próximos capítulos.

4.4 A Expressão da Emoção

O objectivo principal do *modelo emocional de fluxo* é suportar a implementação de agentes inteligentes capazes de operar em tempo real em condições de incerteza, dinamismo e limitação de recursos, típicas de ambientes reais. Nesse sentido, foi dada ênfase à modelação de aspectos passíveis de permitirem a integração com mecanismos de raciocínio, tomada de decisão e adaptação, e não à modelação da expressão emocional.

Reconhece-se, contudo, a importância que a expressão emocional pode ter, por exemplo, no contexto da interacção com interlocutores humanos (e.g. [Picard, 1997]) ou na implementação de personagens virtuais (e.g. [Paiva *et al.*, 2004; Aylett, 2004]). Nesse sentido, e de modo a proporcionar as bases para possíveis desenvolvimentos

futuros, de seguida será feita uma breve discussão de como o modelo proposto pode suportar diferentes tipos de expressão emocional.

4.4.1 Contextos Cognitivos e Expressão Emocional

No *modelo emocional de fluxo* os fenómenos emocionais são conceptualizados como resultando de congruências espaciais e temporais das múltiplas dinâmicas emocionais que caracterizam um agente, com reflexos, em particular, na actividade cognitiva e no comportamento de um agente. Essas regularidades correspondem a fases de evolução dos processos internos do agente, podendo ter durações e intensidades variáveis, bem como predispor a diferentes tipos de processamento cognitivo ou de comportamento, dando assim origem a noções como *emoção* ou *estado de espírito* (“*mood*”). Da compartimentação no tempo desses padrões de fenómenos emocionais e da respectiva diferenciação com base em termos linguísticos, como *alegria* ou *zanga*, resulta a noção clássica de estado emocional.

Um aspecto central desses fenómenos emocionais é o contexto cognitivo em que ocorrem, sendo esse contexto que dá a característica multifacetada dos fenómenos emocionais, surgindo, por exemplo, o que se designa por emoções sociais, como *inveja* ou *culpa*, entre outras. Deste modo, quanto mais complexas forem as capacidades cognitivas de um agente, maior o repertório de fenómenos emocionais que será possível identificar. Isto é consistente com as observações experimentais que mostram que a gama e complexidade de fenómenos emocionais aumenta com o aumento da capacidade cognitiva dos organismos, sendo a sua expressão mais complexa nos humanos [Hebb & Thompson, 1979; Scherer, 2001].

É o contexto cognitivo que origina, também, os diferentes tipos de critérios de apreciação anteriormente discutidos, resultantes precisamente da discriminação linguística desses vários fenómenos e duma posterior conceptualização do tipo de avaliações e relações que poderão estar na origem desses fenómenos.

Em termos qualitativos podemos distinguir três níveis genéricos de contextos cognitivos:

- *Nível estrutural*, onde o conhecimento ocorre sob a forma de *acoplamento estrutural* entre o agente e o ambiente [Maturana & Varela, 1987], não implicando representações conceptuais ou simbólicas.

- *Nível simbólico*, onde o conhecimento ocorre sob a forma de estruturas simbólicas discretas [Newell & Simon, 1976].
- *Nível de conhecimento*, onde o conhecimento ocorre sob a forma de representações conceptuais interrelacionadas, tipicamente associadas a processos de raciocínio [Newell, 1982].

Num nível estrutural as dinâmicas emocionais têm um efeito essencialmente sobre o meio interno e em termos de comportamento, em relação directa com o ambiente. Neste caso, a expressão emocional consiste nos padrões emocionais base, sobretudo naqueles associados à existência de fluxos divergentes, ou seja, *zanga* e *medo*.

Num nível simbólico, pode ocorrer a acumulação e associação das dinâmicas emocionais a estruturas simbólicas específicas, podendo por isso surgir cambiantes de padrões emocionais mais complexos e o embrião de intencionalidade emocional. A intencionalidade emocional ganha a sua expressão efectiva num nível de conhecimento, por meio da modulação das relações entre conceitos através das dinâmicas emocionais. Por exemplo, em agentes capazes de raciocínio social, as dinâmicas emocionais podem ser direccionadas em relação ao próprio agente ou a outros agentes, dando origem a cambiantes emocionais, como por exemplo, *culpa* (*frustração/zanga* direccionada ao próprio agente) (e.g. [Ortony *et al.*, 1988]).

4.4.2 Escalas Temporais e Expressão Emocional

Um segundo aspecto da expressão emocional tem a ver com as escalas temporais em que ocorrem. Também aqui é possível identificar diferentes variações da expressão dos fenómenos temporais ao longo do tempo. O que, mais uma vez, indicia o carácter contínuo e dinâmico dos fenómenos emocionais.

Em termos de caracterização emocional, são normalmente identificados quatro tipos de fenómenos:

Tipo de Fenómeno	Escala Temporal
Episódios emocionais	Segundos
Emoções	Minutos, horas
Estados de espírito	Horas, dias, meses
Traços de personalidade	Anos

Tabela 4.4. Relação entre fenómenos emocionais e escalas temporais em que ocorrem (adaptado de [Oatley & Jenkins, 1996]).

O facto dos fenómenos emocionais se poderem dilatar no tempo, em vez de estarem restritos aos períodos em que ocorrem os estímulos ou situações que lhe dão origem, é intrínseco ao modelo proposto, uma vez que a evolução dos potenciais e fluxos, através dos quais se expressam as dinâmicas emocionais, são processos não lineares que evoluem no tempo, implicando acumulação e difusão em diferentes escalas temporais.

Para além disso, é também uma consequência do modelo proposto, que as escalas de ocorrência de fenómenos emocionais tendam a estender-se no tempo à medida que a complexidade cognitiva dos agentes aumenta. Assim, num nível estrutural, não existem estruturas cognitivas complexas nas quais possa ocorrer acumulação de potenciais emocionais, pelo que os fenómenos tenderão a ter uma escala temporal curta.

Em contrapartida, num nível de conhecimento, as múltiplas relações entre representações conceptuais produzem contextos de acumulação de potenciais cognitivos, nomeadamente sob a forma de memórias emocionais, os quais tendem a permanecer no tempo, eventualmente formando ciclos de realimentação. Esses contextos tendem, também, a tornar-se difusos à medida que as representações conceptuais que lhe estão subjacentes se diluem na memória, mantendo-se, no entanto, os potenciais emocionais, mas agora com um carácter difuso não associado a representações específicas, no que poderá ser análogo à noção de *estado de espírito* (“*mood*”).

A influência do carácter cumulativo dos fenómenos emocionais pode reflectir-se também nos próprios parâmetros da arquitectura. Este aspecto está também relacionado com os fenómenos de escala temporal mais longa, como é o caso dos *traços de personalidade* referidos por Hudlicka [2004] (ver secção 3.3.3). Esses fenómenos

podem resultar quer de padrões emocionais prevalecentes ao longo do tempo, determinando características dos próprios mecanismos emocionais, quer da definição de características desses mecanismos de modo específico em tempo de projecto. Por exemplo, o coeficiente de sensibilidade α_{sit} (definição 4.5) que determina a sensibilidade do *potencial de situação* em relação ao *fluxo de concretização*, pode ele próprio ser determinado a partir do carácter emocional prevalecente, dando origem a fenómenos não lineares como o *escalar* de tendências emocionais negativas, como zanga e medo, em contraponto ao carácter *saciante* de tendências emocionais positivas, como alegria e bem estar (e.g. [Frijda, 2001; Schwartz, 2004]). Não sendo objectivo desta tese abordar os aspectos específicos da modelação emocional, todas estas áreas representam interessantes direcções para desenvolvimentos futuros.

4.5 Conclusão

O modelo de emoção apresentado neste capítulo diverge significativamente da linha de investigação predominante nos modelos computacionais de emoção, a qual tem por base a caracterização dos fenómenos emocionais em termos classificativos sob a forma de etiquetas emocionais discretas, tal como discutido no capítulo anterior.

Em vez disso, adopta-se uma perspectiva onde os fenómenos emocionais são vistos como fenómenos inerentemente dinâmicos e contínuos. Para suportar essa perspectiva é utilizada a noção de estrutura dissipadora como analogia base para a modelação de um agente. Com base nessa analogia torna-se possível caracterizar a relação entre um agente e o ambiente, de forma contínua, sob a forma de fluxos de energia. Torna-se igualmente possível identificar e caracterizar as dinâmicas motivacionais subjacentes à actividade de um agente. Essas dinâmicas motivacionais são consideradas as forças directoras dos padrões de mudança da actividade interna (cognitiva) e externa (comportamento) de um agente e a origem primária do surgimento de fenómenos emocionais.

Torna-se assim possível abordar as questões identificadas no capítulo anterior. Em primeiro lugar, a noção estereotipada de apreciação emocional e a necessidade de mecanismos de classificação comprometidos com aspectos específicos da actividade cognitiva é considerada inadequada, pois limita a expressão do carácter contínuo e dinâmico dos fenómenos emocionais, bem como a definição de uma relação estreita e simbiótica entre cognição e emoção. Em vez disso, a qualidade e intensidade dos

fenómenos emocionais, bem como a forma como se desenvolvem no tempo, resulta numa relação complementar entre agente e ambiente, expressa sob a forma de fluxos de concretização e sintetizada, em termos formais, pela noção de disposição emocional.

Tendo a noção de disposição emocional um carácter contínuo, não são necessárias etiquetas discretas para caracterizar os fenómenos emocionais, o que é consistente com a multiplicidade de cambiantes emocionais que caracteriza esses fenómenos. Apesar disso, é possível modelar a ocorrência de padrões emocionais sem, contudo, comprometer o carácter contínuo e multifacetado da expressão emocional.

Também a necessidade de atributos de qualificação emocional, como intensidade, taxa de decaimento e outros, que caracteriza muitos dos actuais modelos de emoção (e.g. [Hudlicka, 2004; Gratch & Marsella, 2004; Oliveira & Sarmento, 2003]), deixam de ter lugar, pois o carácter contínuo e dinâmico que esses atributos têm por função reproduzir, é intrínseco às noções base do modelo de emoção proposto.

Para além da adequação em termos de caracterização emocional, o carácter dinâmico e contínuo que está subjacente ao modelo de emoção proposto pode desempenhar um papel fundamental na integração entre fenómenos emocionais e cognitivos. Para isso é necessário concretizar as noções base do modelo no contexto de um modelo de agente. É esse modelo de agente que será apresentado no capítulo seguinte.

Capítulo 5

O Modelo de Agente

No capítulo anterior adoptou-se uma perspectiva macroscópica da noção de agente como estrutura dissipadora, de modo a tornar simples e concisa a descrição do modelo de emoção proposto. Contudo, para que seja possível tirar partido desse modelo de emoção, é necessário concretizar o modelo de agente que lhe está subjacente. Essa concretização vai acontecer em duas etapas. Neste capítulo serão definidos os conceitos e arquitectura base do modelo de agente, de modo a proporcionar o suporte necessário para a modelação do carácter dinâmico e contínuo dos fenómenos emocionais. Como iremos ver, esse carácter dinâmico e contínuo é essencial para uma integração efectiva entre fenómenos emocionais e cognitivos. No próximo capítulo essa integração será concretizada, proporcionando as capacidades adaptativas e de resposta em tempo-real que caracterizam o modelo de agente proposto.

5.1 Introdução

A dicotomia entre arquitecturas de agente cognitivas (ou deliberativas) e arquitecturas comportamentais (ou reactivas), a que se tem assistido nos últimos anos, tem alimentado uma visão dos aspectos cognitivos e comportamentais como essencialmente diferentes e, por isso, requerendo diferentes suportes arquitecturais (e.g. [Van de Velde, 1995; Weiss, 2000]). Perante as limitações de ambas as abordagens, o resultado tem sido o surgimento de arquitecturas híbridas onde as distinções arquitecturais entre aspectos cognitivos e comportamentais são assumidas, dando origem a arquitecturas de agente organizadas em camadas funcionalmente estanques (ver secção 2.1.4).

Apesar da organização em camadas ser uma decomposição funcional natural, o isolamento entre camadas, onde cada camada possui suportes arquitecturais distintos, produz importantes problemas de integração (e.g. [Wooldridge, 2002]),

nomeadamente na definição das interacções entre camadas, bem como na definição das estruturas que suportam essa interacção e das escalas temporais em que as interacções ocorrem. No caso da integração de aspectos emocionais numa arquitectura de agente adoptando o modelo de emoção proposto, estes problemas tornam-se ainda mais incisivos, pois os fenómenos emocionais são vistos como resultantes de dinâmicas emocionais que ocorrem a diferentes níveis da actividade cognitiva de um agente.

Esta distinção entre aspectos cognitivos e comportamentais como essencialmente diferentes e eventualmente ortogonais, também não é consistente com o que se observa nos organismos biológicos, cujas características têm sido uma fonte de inspiração importante, sobretudo para as abordagens comportamentais. De facto, o que resulta dos dados da biologia, etologia e neurociências é uma continuidade filogenética e funcional na forma de organização dos vários aspectos subjacentes ao comportamento dos organismos, com diferentes ênfases relativas consoante o tipo de organismo.

Esta continuidade acontece num contexto de evolução funcional e estrutural, caracterizado por duas tendências principais [Barnard, 2004]:

- crescente *diferenciação/especialização*;
- crescente *centralização*.

Um exemplo deste processo de diferenciação e especialização é o desenvolvimento do sistema nervoso. Na sua forma mais simples, redes nervosas não direccionais dão origem a comportamentos estereotipados (invariantes) sob a forma de reflexos ou sequências motoras. Formas mais complexas são caracterizadas pela formação de gânglios nervosos e redes de interligação direccionais, como acontece, por exemplo, nos insectos, dando origem a comportamentos complexos, incluindo comportamentos associados à formação e manipulação de representações internas (e.g. [Giurfa & Menzel, 2003]). Nos vertebrados e, em particular, nos humanos, estas tendências de diferenciação e centralização continuam a ser claras, tornando-se no entanto muito mais complexas, dando origem a comportamentos e capacidades cognitivas sofisticadas. A figura seguinte dá uma ideia da relação entre a evolução filogenética e diferentes aspectos subjacentes ao comportamento dos organismos.

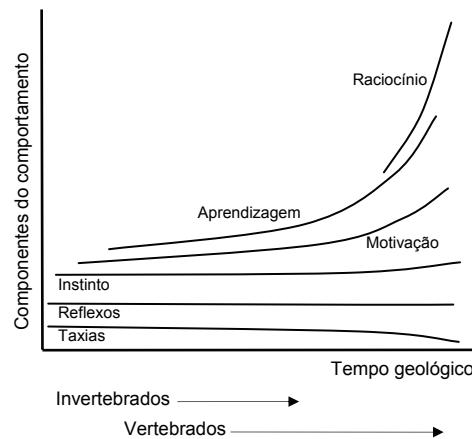


Figura 5.1. Ilustração esquemática do desenvolvimento filogenético de diferentes componentes do comportamento (adaptado de [Shepherd, 1994]).

Esta continuidade funcional, onde mecanismos mais complexos vão complementando e ampliando mecanismos cognitivos mais simples, numa crescente tendência de centralização e especialização, tem como resultado uma gama de comportamentos alargada, capaz de permitir lidar com diferentes condições ambientais, o que não é possível considerando apenas aspectos reactivos ou deliberativos.

A coexistência destes diferentes aspectos nos vertebrados e, em particular, nos humanos, tem reflexo nos diferentes tipos de fenómenos emocionais observáveis. Por exemplo, em [Fellous, 2004] é proposta uma perspectiva dos fenómenos emocionais como associados a fenómenos de neuromodulação a diferentes níveis organizacionais do sistema nervoso, tal como ilustrado na figura seguinte.

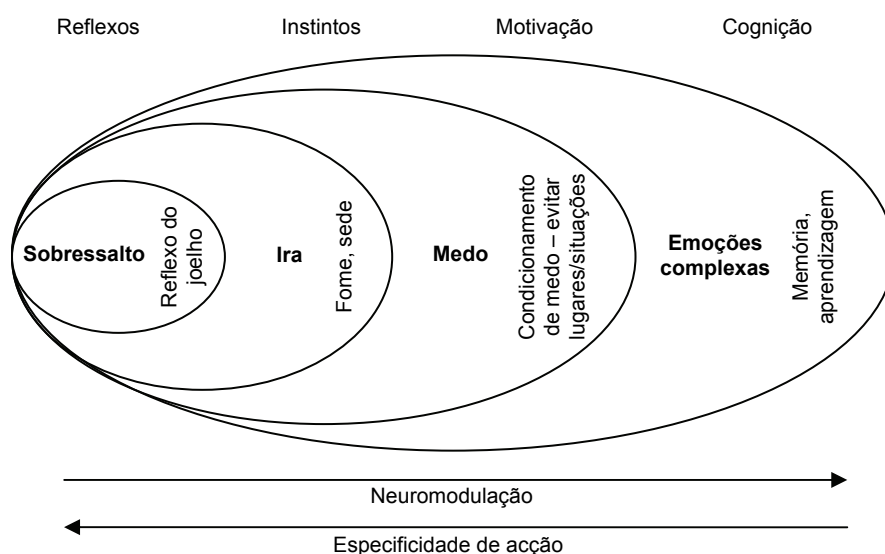


Figura 5.2. Relação entre componentes do comportamento e fenómenos emocionais (adaptado de [Fellous, 2004]).

Esta relação entre uma complexidade cognitiva crescente e diferentes tipos de fenómenos emocionais é consistente com o modelo de emoção proposto, onde a influência das dinâmicas emocionais é expressa de diferentes formas consoante os processos cognitivos envolvidos. No entanto, essa relação não é uma relação fragmentada entre camadas estanques, mas antes uma relação que resulta de suportes uniformes, quer ao nível da estrutura cognitiva, quer ao nível dos mecanismos base, dos quais se podem destacar os mecanismos emocionais.

É esta necessidade de uma integração estreita entre os vários aspectos cognitivos, e entre aspectos cognitivos e emocionais, que é um dos principais problemas na integração de aspectos emocionais em modelos de agente. Esse problema ocorre quer no contexto do paradigma cognitivo (veja-se a discussão do modelo proposto por Hudlicka no capítulo 3), quer no contexto do paradigma comportamental (veja-se a discussão do modelo proposto por Cañamero, no mesmo capítulo).

Para abordar este problema de integração entre aspectos emocionais e cognitivos, propõe-se uma perspectiva alternativa aos paradigmas clássicos de concepção de agentes, na qual a analogia entre a noção de agente e a noção estrutura dissipadora desempenha o papel central. É essa perspectiva que serve de base ao modelo de agente proposto e que será desenvolvida de seguida.

5.2 Arquitectura Base

Vimos no capítulo anterior que, tendo por base a noção de estrutura dissipadora, um agente é caracterizado por um conjunto de potenciais internos e por um conjunto de fluxos que os originam (ver figura 4.2). São os potenciais internos que definem a estrutura do agente, nomeadamente a sua *estrutura cognitiva* (noção que será concretizada mais à frente e que é um dos aspectos chave da integração entre aspectos cognitivos e emocionais). Esses potenciais internos são resultado quer de fluxos provenientes do ambiente, quer de fluxos gerados no contexto de processos internos ao agente. No caso geral, esses processos podem estar estruturados a diferentes níveis de organização, consoante o tipo e complexidade dos agentes, tal como ilustrado na figura seguinte.

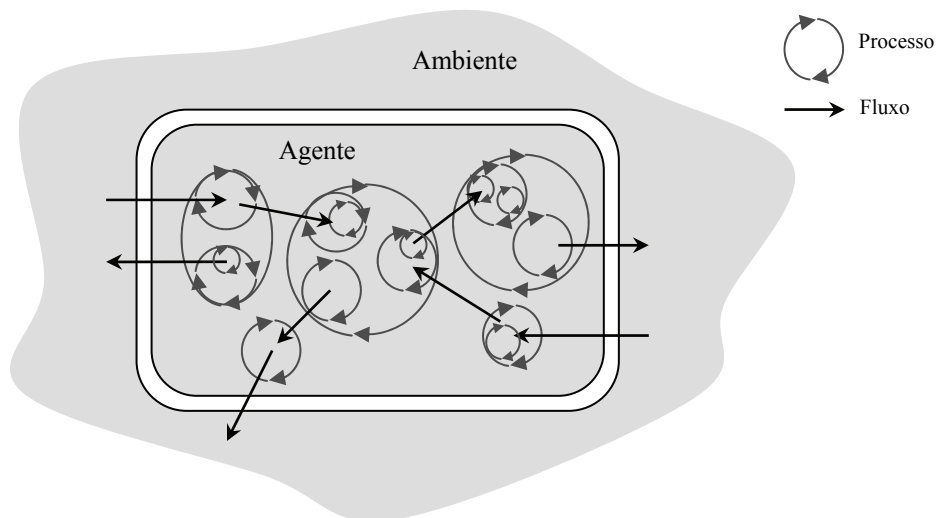


Figura 5.3. Representação abstracta de um agente como um conjunto de processos estruturados a diferentes níveis de organização, consoante o tipo e complexidade dos agentes.

Para suportar este tipo de organização, a arquitectura de agente proposta é caracterizada por múltiplos processos com actividade concorrente e fracamente acoplados. Esses processos caracterizam-se por diferentes dinâmicas e funcionalidades. Contudo, independentemente do tipo de agente, é possível identificar dois tipos de processos base, os quais se encontram directamente relacionados com o acoplamento com o ambiente:

- processos que realizam actividades de transdução e interiorização dos estímulos resultantes da interacção com o ambiente, designados genericamente *processos de percepção*;
- processos que realizam a exteriorização e transdução de sinais gerados no interior do agente, designados genericamente *processos de acção*.

Em agentes mais simples, como é o caso de agentes reactivos, a relação entre os processos de percepção e acção deriva de restrições estruturais ou funcionais, embutidas *a priori* nos agentes, e que determinam comportamentos estereotipados predefinidos. Em agentes mais complexos, como é o caso de agentes com capacidades deliberativas, é possível identificar três outros tipos de processos genéricos:

- processos que realizam a integração dos sinais gerados pelos processos de percepção em sinais compostos multimodais, bem como a incorporação desses sinais na estrutura cognitiva do agente, designados, por isso, *processos de assimilação*;

- processos que manipulam a estrutura cognitiva no sentido de derivar informação, nomeadamente no sentido de antecipar situações futuras bem como de determinar cursos de acção adequados, designados, por isso, *processos de raciocínio*;
- processos que seleccionam/determinam o curso de acção de um agente tendo por base as opções disponíveis, resultantes quer dos processos de raciocínio quer da activação de memórias de experiências passadas (recordações) geradas pelos processos de assimilação, designados *processos de decisão*.

Apesar da distinção entre tipos de processos, esses diferentes tipos de processos traduzem uma continuidade funcional, tal como ilustrado na figura seguinte.

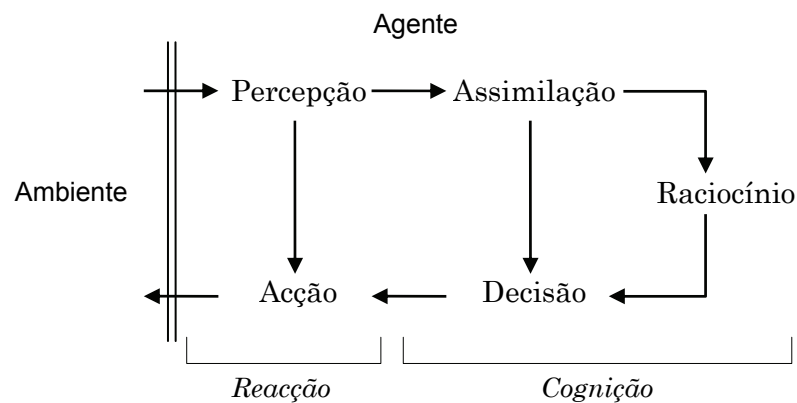


Figura 5.4. Relação entre diferentes tipos de processos cognitivos.

A separação clássica entre reacção e cognição, ilustrada na figura 5.4, corresponde a uma distinção entre agentes cujos processos internos estão directamente relacionados com o acoplamento com o ambiente e agentes com processos internos mais complexos, assentes na formação e manipulação de estruturas internas que reflectem, entre outros aspectos, a evolução da interacção entre agente e ambiente ao longo do tempo.

Subjacentes aos vários tipos de processos internos estão mecanismos base responsáveis pela sincronização e regulação da actividade cognitiva de um agente. Por *actividade cognitiva* entende-se a actividade conjunta dos diferentes tipos de processos internos que determinam o comportamento de um agente, os quais são, por isso, designados *processos cognitivos*. Note-se que esta noção de actividade cognitiva difere da noção clássica, associada ao paradigma cognitivo, onde a noção de cognição é relacionada com a capacidade de formação e manipulação de representações

Por sua vez, os *mecanismos base* possuem características específicas que, independentemente do tipo de agente, são invariantes, apesar de alguns mecanismos poderem não existir para determinados tipos de agente. Assim, o foco principal do modelo de agente proposto vai incidir na modelação da estrutura cognitiva e na modelação dos mecanismos base de suporte à actividade cognitiva.

5.2.1 Estrutura Cognitiva

O elemento central da modelação da estrutura cognitiva de um agente como estrutura dissipadora é a noção de potencial energético. Esses potenciais resultam de fluxos energéticos derivados quer da interacção com o ambiente, quer da actividade interna do agente.

Para que um potencial energético possa existir é necessário que se possa diferenciar do que o rodeia, nomeadamente de outros potenciais. Essa diferenciação pode ocorrer por separação espacial ou por diferenciação qualitativa. Diferenciação qualitativa significa que os potenciais possuem uma determinada qualidade específica que permite a sua individualização em relação a outros potenciais, mesmo quando ocupam a mesma região do espaço. Ou seja, a diferenciação qualitativa permite a superposição de potenciais energéticos sem perda das suas características específicas. Como iremos ver, esta noção de superposição de potenciais desempenha um papel chave na formação da estrutura cognitiva de um agente.

5.2.2 Elementos Cognitivos

Para que seja possível uma diferenciação qualitativa, os potenciais energéticos possuem dinâmicas próprias que os distinguem em termos da forma como evoluem no tempo e/ou no espaço. Para suportar essas características, os potenciais energéticos são modelados como sinais periódicos com uma determinada frequência angular ω , que expressa a sua qualidade e que os distingue de outros sinais com frequências diferentes. Ao adoptar esta formulação, é possível não só tirar partido da noção de estrutura dissipadora como analogia central na modelação dos agentes, como é possível, também, tirar partido de uma base sólida de conhecimentos proporcionada pela teoria dos sinais e sistemas (e.g. [Carlson, 1986; Oppenheim *et al.*, 1997]).

A discriminação qualitativa de sinais energéticos está directamente relacionada com os processos de percepção, os quais são uma das principais fontes para a formação

dos potenciais que constituem a estrutura cognitiva de um agente. Considerando as evidências da biologia, o que se observa é que os organismos biológicos possuem mecanismos específicos sintonizados para serem sensíveis a diferentes formas de energia, o que é designado por modalidades sensoriais. As modalidades sensoriais podem, por sua vez, ser subdivididas em sub-modalidades ou qualidades sensoriais.

No caso dos humanos, estes processos sensoriais estão na base da percepção de diferentes características do ambiente que nos rodeia, como sabores, cheiros, sensações de frio ou calor, ou a percepção das cores resultante da discriminação das diferentes frequências que caracterizam os sinais luminosos (e.g. [Shepherd, 1994; Levine, 2000; Coren *et al.*, 2004]).

Os receptores sensoriais que estão na origem destas percepções estão sintonizados numa determinada gama da banda espectral dos estímulos, formando um mecanismo de discriminação. Ou seja, a percepção das condições ambientais ocorre pela estimulação das interfaces sensoriais do agente, sendo os estímulos assimilados como potenciais energéticos que, em conjunto com os potenciais gerados internamente, formam a estrutura cognitiva do agente, nomeadamente sob a forma de memórias, sendo por isso designados *potenciais cognitivos*.

Os potenciais cognitivos resultam quer da interacção entre o agente e o ambiente, quer da actividade interna do próprio agente. Em qualquer dos casos, os potenciais cognitivos expressam aspectos do ambiente (externo e interno) que o agente é capaz de discriminar e perceber, o que normalmente é designado por dimensões qualitativas (e.g. [Gärdenfors, 2000]).

A discriminação dos potenciais cognitivos envolve, igualmente, uma medida quantitativa associada às qualidades discriminadas. Por exemplo, as qualidades “temperatura” ou “pressão” possuem uma gama contínua de valores. Outras qualidades, em particular qualidades resultantes de processamento conceptual, podem envolver gamas de valores discretos. Por exemplo, a caracterização contínua da qualidade “temperatura” pode, posteriormente, ser descriminada em valores discretos “quente” e “frio”. Assim, os potenciais cognitivos são modelados como uma composição de dois tipos de sinais: (i) um *senal qualitativo base* $\varphi(t)$ com uma frequência angular $\omega = \Omega$ que identifica a qualidade ou aspecto descriminados; e (ii) um *senal quantitativo* $\rho(t)$ correspondente ao valor efectivo da qualidade discriminada, expresso sob a forma de um desvio de frequência $\Delta\omega$ que modula o sinal base $\varphi(t)$. Ou seja:

$$p(t) = \rho(t)\varphi(t) \quad (5.1)$$

Através de superposição, os potenciais cognitivos formam agregados que são os elementos base da estrutura cognitiva de um agente, sendo por isso designados *elementos cognitivos*. Para garantir que a superposição ocorre sem perda de informação, considera-se que os sinais qualitativos base $\varphi_i(t)$ são ortogonais entre si e que as frequências base respectivas, Ω_i , estão suficientemente afastadas para que não ocorra interferência entre as bandas espectrais correspondentes a cada sinal qualitativo base. Assim sendo, um elemento cognitivo $\sigma(t)$ é modelado como uma composição de potenciais cognitivos:

$$\sigma(t) = \sum_{i=1}^K p_i(t) \quad (5.2)$$

onde K é o número de potenciais no agregado. A figura seguinte ilustra a formação de elementos cognitivos no contexto de um processo de percepção.

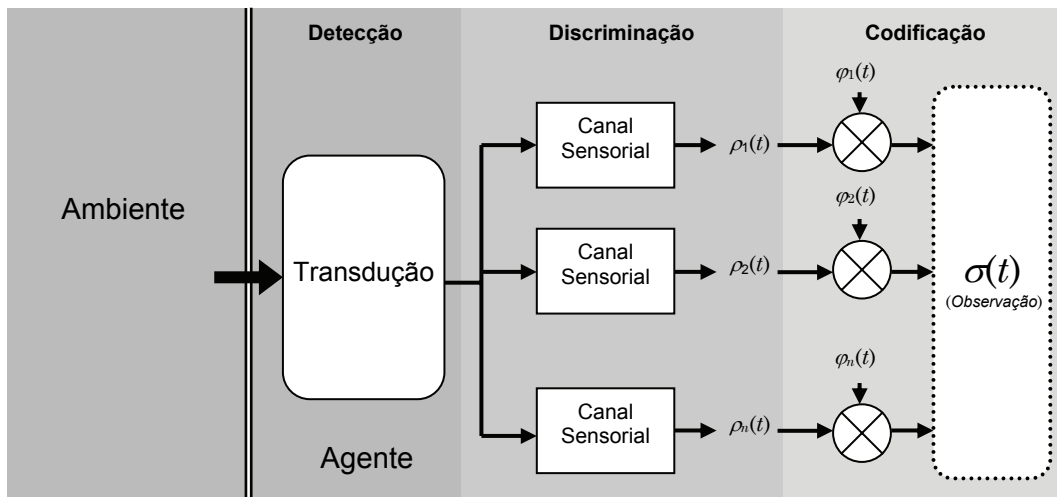


Figura 5.6. Formação de elementos cognitivos como resultado de actividades de percepção.

Como se pode observar na figura, estão envolvidas na formação de elementos cognitivos três actividades principais:

- *detecção*: na qual é feita a transdução e detecção de sinais exteriores, os quais podem surgir sob diferentes formas (electromagnéticos, químicos, mecânicos, etc.);

- *discriminação*: na qual diferentes canais sensoriais estão sintonizados para discriminar diferentes qualidades de sinais;
- *codificação*: na qual os elementos cognitivos são gerados por modulação e superposição dos potenciais cognitivos discriminados anteriormente.

As actividades de detecção são típicas de processos de percepção. Essa percepção pode ser externa (exteroceptiva) ou interna (interoceptiva), neste segundo caso os agentes são capazes de explicitamente perceber e monitorizar o seu ambiente interno. Outros processos cognitivos internos, como raciocínio ou planeamento, envolvem igualmente a geração de elementos cognitivos sem, no entanto, ocorrerem actividades de detecção no sentido das que ocorrem no âmbito da percepção.

No contexto das diferentes actividades cognitivas, os elementos cognitivos desempenham papéis específicos, sendo possível identificar três papéis base:

- *observações*: são o resultado directo de processos de percepção, representando as situações ambientais observadas;
- *motivadores*: representam as situações que o agente pretende atingir, actuando como forças impulsionadoras do comportamento dos agentes;
- *mediadores*: representam os meios que suportam a acção, formando uma interface entre o processamento cognitivo interno e a acção concreta.

Enquanto as observações são o resultado da transdução e assimilação de fluxos de energia externos, os motivadores e mediadores são produzidos internamente, como resultado da actividade cognitiva do agente, ou então explicitamente embutidos na estrutura cognitiva do agente como resultado de opções de projecto ou como restrições estruturais, resultantes da arquitectura definida. Os motivadores correspondem aos sinais de auto-regulação descritos no âmbito do *modelo emocional de fluxo*. Por sua vez, os mediadores servem de interface entre os processos cognitivos e a acção concreta. A figura seguinte ilustra este papel dos mediadores.

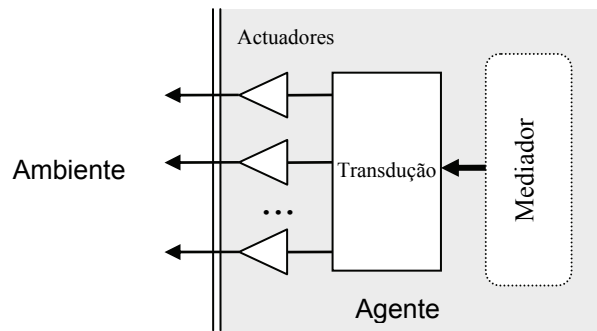


Figura 5.7. Mediadores como interface para acção concreta.

Os elementos cognitivos podem, por sua vez, formar agregados mais complexos representando aspectos integrados da experiência do agente, quer em termos de situações localizadas no tempo, quer em termos de evolução temporal dessas situações, por exemplo sob a forma de memórias autobiográficas, tal como será abordado mais à frente.

Na estrutura cognitiva de um agente os elementos cognitivos correspondem a localizações específicas. É essa localização na estrutura cognitiva que torna possível a sua diferenciação. No sentido de abstrair essa localização, serão utilizadas designações específicas como forma de diferenciar elementos cognitivos. Por exemplo, um agente com capacidade de produzir observações a partir de dois sensores, um sensor direito (DIR) e um sensor esquerdo (ESQ), será caracterizado por dois elementos cognitivos σ_{DIR} e σ_{ESQ} que representam as respectivas observações. Neste caso, DIR e ESQ representam localizações distintas na estrutura cognitiva.

Note-se, no entanto, que a existência dos elementos cognitivos é transitória, isto é, um elemento cognitivo é formado, desempenha o seu papel na actividade cognitiva durante um determinado período de tempo, e desaparece. Outro elemento cognitivo que venha a formar-se posteriormente na mesma localização do anterior, é considerado um elemento cognitivo distinto. Neste sentido, os elementos cognitivos são igualmente localizados no tempo. Por uma questão de clareza de notação, essa localização temporal será mantida, para já, implícita. A localização temporal dos elementos cognitivos será considerada de modo explícito no contexto da formação de memórias autobiográficas, tema que será abordado no capítulo 6.

5.2.3 Períodos de Actividade Cognitiva

Em ambientes reais qualquer actividade consome tempo, incluindo a actividade cognitiva. Sendo o tempo um recurso escasso, é fundamental que os processos cognitivos possam gerir a utilização desse recurso. No próximo capítulo serão discutidos os meios para fazer essa gestão. Para já, é necessário modelar explicitamente esse carácter não instantâneo da actividade cognitiva.

Nesse sentido, no modelo de agente proposto a actividade dos processos cognitivos de um agente ocorre por *períodos de actividade cognitiva*, os quais definem a granularidade temporal do processamento cognitivo. Esses períodos determinam a unidade mínima de discriminação temporal de um agente. Ou seja, é durante esses períodos que os sinais cognitivos são gerados e interagem, produzindo novos elementos cognitivos, sendo estes considerados estáveis após uma fase inicial de transição entre períodos de actividade cognitiva, tal como ilustrado na figura seguinte.

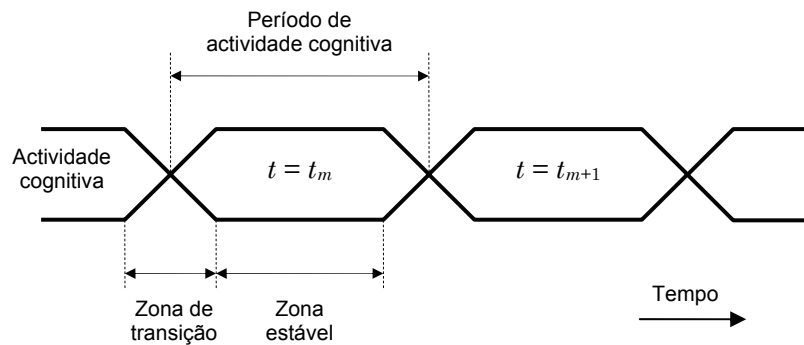


Figura 5.8. Evolução dos *períodos de actividade cognitiva* ao longo do tempo.

O facto dos períodos de actividade cognitiva terem uma duração não nula, tem como resultado um tempo subjectivo caracterizado por instantes discretos t_n com $n \in \mathbb{N}$. Note-se que a duração dos períodos de actividade cognitiva pode variar consoante os processos envolvidos e ao longo do tempo, tal como será abordado mais à frente.

Durante a zona estável de um período de actividade cognitiva, as características dos elementos cognitivos envolvidos mantêm-se inalteradas, apesar de poderem ser gerados novos elementos cognitivos. Deste modo, os elementos cognitivos são vistos como localizados no tempo, com uma existência inerentemente transiente, a não ser que tornados persistentes pela assimilação na estrutura cognitiva, por exemplo, sob a forma de memórias.

Um elemento cognitivo pode assim ser visto como o análogo da noção clássica de estado mental (e.g. [Bratman, 1987]), representando aspectos da experiência cognitiva do agente durante um determinado período de tempo. Por exemplo, uma *observação* tem um carácter *denotativo*, enquanto um *motivador* tem um carácter *volitivo*, sendo esse carácter determinante para a evolução da actividade cognitiva do agente.

Os períodos de actividade dos vários processos cognitivos podem estar ou não sincronizados, consoante o tipo e nível de complexidade dos agentes. No entanto, em agentes com capacidades de assimilação e raciocínio, essa sincronização é fundamental para garantir quer a consistência das memórias formadas, quer a consistência das representações manipuladas a nível dos processos de raciocínio e decisão.

5.2.4 Espaço Cognitivo

A estrutura cognitiva de um agente é o suporte através da qual um agente mantém um estado interno que reflecte a evolução da interacção entre agente e ambiente ao longo do tempo. Nesse sentido, e entre outros aspectos, a estrutura cognitiva de um agente suporta a formação de um *modelo interno* (ou modelo mental, e.g. [Newell, 1990]), caracterizado por uma correspondência entre representações internas e situações externas. No sentido de permitir uma descrição clara e concisa dessas representações, podemos observar que os sinais base que compõem os potenciais e elementos cognitivos podem ser modelados como um espaço de sinais. Isto torna possível uma descrição mais simples dos elementos cognitivos e da estrutura cognitiva, permitindo igualmente descrever a sua evolução, de forma concisa, como movimento nesse espaço de sinais, designado, por isso, *espaço cognitivo*.

O espaço cognitivo é caracterizado como um espaço de sinais multi-dimensional, onde cada sinal base define uma dimensão do espaço. Formalmente, um espaço cognitivo CS_K é definido por um conjunto de K vectores base ortonormais $\Phi = \{\varphi_i: i = 1, 2, \dots, K\}$ com $K \in \mathbb{N}$, onde cada vector φ_i representa um sinal base $\varphi_i(t)$ com uma qualidade específica $\omega = \Omega_i$.

No espaço cognitivo os elementos cognitivos correspondem a posições específicas. Uma vez que os elementos cognitivos são localizados no tempo, num determinado instante $t = \tau$, um elemento cognitivo $\sigma(t)$ é representado num espaço cognitivo CS_K como um vector σ , definido do seguinte modo:

$$\sigma \equiv (\rho_1, \rho_2, \dots, \rho_K) \quad (5.3)$$

Os coeficientes dimensionais $\rho_i \in \mathbb{C}$ expressam a intensidade ε_i e o desvio de frequência $\Delta\omega_i$ de cada sinal quantitativo $\rho_i(t)$ em relação ao sinal base respectivo $\varphi_i(t)$. A forma concreta desses coeficientes dimensionais depende das características específicas dos sinais quantitativos $\rho_i(t)$, em particular da sua persistência no tempo (ver apêndice B). A figura seguinte ilustra a noção de espaço cognitivo.

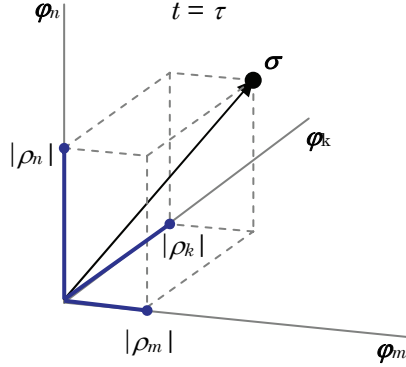


Figura 5.9. Representação de um elemento cognitivo no *espaço cognitivo* (coeficientes dimensionais representados em valor absoluto).

Na figura 5.9 cada vector cognitivo σ representa um elemento cognitivo num determinado instante $t = \tau$. Como resultado da actividade de um agente, novos elementos cognitivos vão sendo formados ao longo do tempo. O encadeamento temporal desses vários elementos pode ser representado no espaço cognitivo como uma trajectória, a qual reflecte o comportamento do agente ao longo do tempo, tal como ilustrado na figura seguinte.

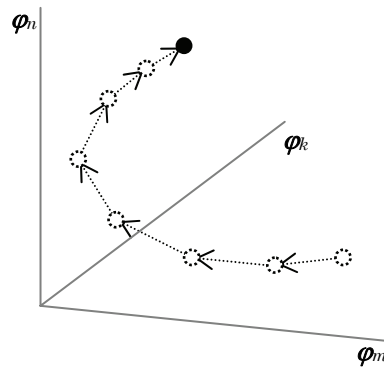


Figura 5.10. Representação no espaço cognitivo de uma trajectória formada por elementos cognitivos resultantes da actividade de um agente ao longo do tempo.

Como iremos ver, estas trajectórias desempenham um papel fundamental, quer na descrição das dinâmicas emocionais, quer como suporte à capacidade de adaptação de longo prazo de um agente, através da formação de memórias autobiográficas.

A topologia de um espaço cognitivo é caracterizada por uma métrica d que define a distância entre dois elementos cognitivos, σ_1 e σ_2 , da seguinte forma:

$$d(\sigma_1, \sigma_2) \equiv \|\sigma_1 - \sigma_2\| \quad \text{com} \quad \|\sigma\| = \sqrt{\langle \sigma, \sigma \rangle} \quad (5.4)$$

onde $\|\mathbf{x}\|$ representa a norma de um vector \mathbf{x} e $\langle \mathbf{x}, \mathbf{y} \rangle$ representa o produto escalar entre os vectores \mathbf{x} e \mathbf{y} . A noção de norma permite definir também a noção de *intensidade* de um elemento cognitivo, a qual é importante para a caracterização de alguns aspectos da actividade cognitiva que iremos ver mais à frente, nomeadamente aspectos motivacionais e de transdução. A intensidade (ε) de um elemento cognitivo é definida da seguinte forma:

$$\varepsilon(\sigma) \equiv \|\sigma\| \quad (5.5)$$

A noção de espaço cognitivo desempenha um papel central no modelo proposto, permitindo caracterizar de forma concisa, não só a estrutura cognitiva do agente, mas também as dinâmicas cognitivas, sob a forma de movimento dos elementos cognitivos, tal como será explorado mais à frente.

5.2.5 Exemplo 1: Agente Prospector

Para ilustrar a aplicação do modelo de agente proposto, bem como os conceitos envolvidos, consideremos a implementação de um agente cujo objectivo é procurar determinados objectos ou substâncias, designados alvos, evitando objectos ou substâncias nocivas, designados obstáculos.

5.2.5.1 Contexto Experimental

O contexto experimental deste exemplo consiste num ambiente composto por alvos e obstáculos. O ambiente é estático, no sentido em que não muda a não ser por acção do agente, e contínuo, no sentido que todas as suas características são definidas por valores reais. As características de cada elemento existente no ambiente (alvos e obstáculos) propagam-se através do ambiente sob a forma de campos. Esses campos são radialmente simétricos, sendo a sua intensidade local dependente da intensidade

dos elementos que os produzem e da distância a esses elementos, situação que é típica de cenários reais, quer envolvendo dispositivos sensoriais passivos (e.g. foto-sensores), quer activos (e.g. detecção por ultra-sons com pares emissor/receptor).

Consideremos um agente com dois sensores, um à esquerda e um à direita, e com dois actuadores, um que produz rotação e avanço à esquerda e outro que produz rotação e avanço à direita. Cada sensor permite discriminar os estímulos de alvos e de obstáculos.

Por uma questão de clareza, consideremos, para já, um ambiente composto só por alvos, sendo os sensores do agente também só sensíveis a alvos. O objectivo do agente é alcançar todos os alvos, ficando os alvos inactivos quando são alcançados, ou seja, deixam de gerar os campos que os caracterizam.

5.2.5.2 Arquitectura do Agente

O agente é caracterizado por um processo de percepção que realiza a transdução e geração de elementos cognitivos internos (*observações*) que descrevem os estímulos observados, e por um processo de acção que converte um elemento cognitivo indicador da acção a realizar (*mediador*) em acção concreta através dos actuadores. A figura seguinte ilustra estes processos.

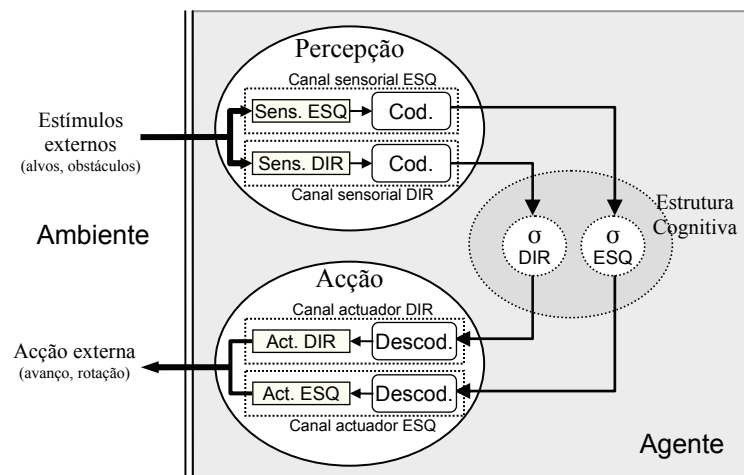


Figura 5.11. Agente prospector simples: processos de percepção, acção e estrutura cognitiva.

Neste caso são gerados dois elementos cognitivos, σ_{DIR} e σ_{ESQ} , resultantes, respectivamente, dos canais sensoriais à esquerda e à direita, e que constituem a estrutura cognitiva do agente. Uma vez que esses elementos cognitivos expressam a intensidade dos estímulos dos alvos à esquerda e à direita, podem ser utilizados para

dirigir directamente a acção do agente, tal como ilustrado na figura anterior. Esta relação directa entre percepção e acção é típica de agentes prospectores reactivos. Um exemplo clássico é o *veículo 2b* de Braitenberg [1984].

No caso de um agente com sensores de alvos e de obstáculos, o processo de percepção do agente detecta, por meio dos sensores, os campos emitidos pelos elementos do ambiente, gerando potenciais cognitivos p_{ALV} e p_{OBST} , que representam a intensidade detectada dos campos de alvos e de obstáculos. Esses sinais são formados por modulação dos sinais qualitativos base φ_{ALV} e φ_{OBST} , tal como anteriormente descrito (ver secção 5.2.2). Os potenciais cognitivos p_{ALV} e p_{OBST} são posteriormente compostos em cada canal sensorial gerando os elementos cognitivos respectivos σ_{DIR} e σ_{ESQ} . O espaço cognitivo do agente é assim constituído por duas dimensões cognitivas, φ_{ALV} e φ_{OBST} .

No caso do exemplo em estudo, o agente apenas possui sensores de alvos, pelo que os elementos cognitivos são compostos por um único potencial cognitivo p_{ALV} , sendo o espaço cognitivo igualmente caracterizado por uma única dimensão φ_{ALV} .

O processo de acção realiza a transdução dos elementos cognitivos σ_{DIR} e σ_{ESQ} em acção concreta através de um actuador específico que faz movimentar o agente. Esse actuador possui um controlo de rotação (ROT) que determina a direcção de movimentação do agente e um controlo de intensidade de avanço (AV) que determina a dimensão do passo de movimentação do agente. A figura seguinte detalha o mecanismo de transdução subjacente ao processo de acção.

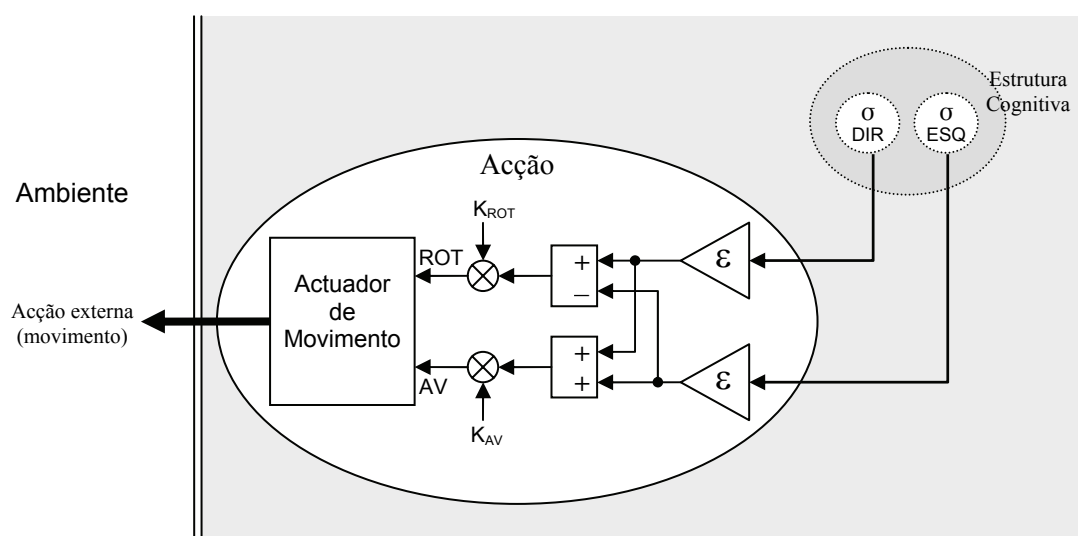


Figura 5.12. Detalhe do mecanismo de transdução subjacente ao processo de acção.

Como se pode observar na figura 5.12, a rotação é proporcional à diferença de intensidades (ε) entre os elementos cognitivos σ_{DIR} e σ_{ESQ} , enquanto o passo de avanço é proporcional à soma dessas intensidades (a intensidade do movimento do agente é tanto maior quanto maior for a intensidade dos estímulos; o agente pára quando não tem estímulos). Os coeficientes de transdução estímulo/acção são determinados pelos parâmetros κ_{ROT} e κ_{AV} . A figura seguinte mostra o comportamento do agente num ambiente com três alvos.

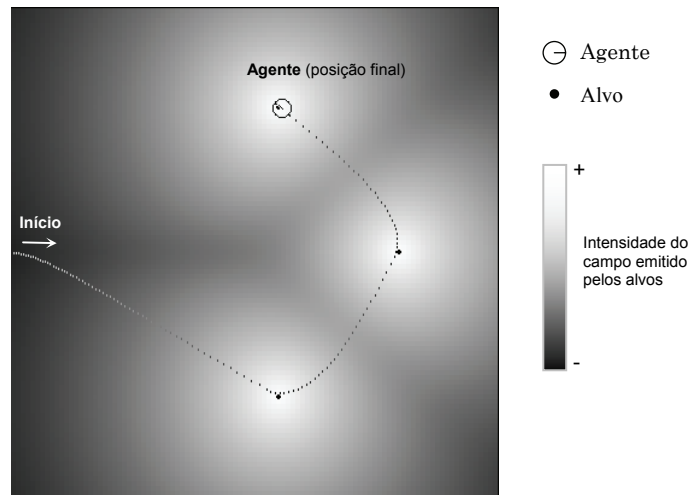


Figura 5.13. Comportamento do agente prospector num ambiente com três alvos.

Como se pode observar, o agente é guiado pelos alvos, afastando-se destes uma vez atingidos, pois deixam de emitir o respectivo campo.

A arquitectura deste agente prospector é uma concretização da arquitectura genérica proposta, na qual apenas os processos de percepção e acção estão instanciados, tal como ilustrado na figura seguinte.

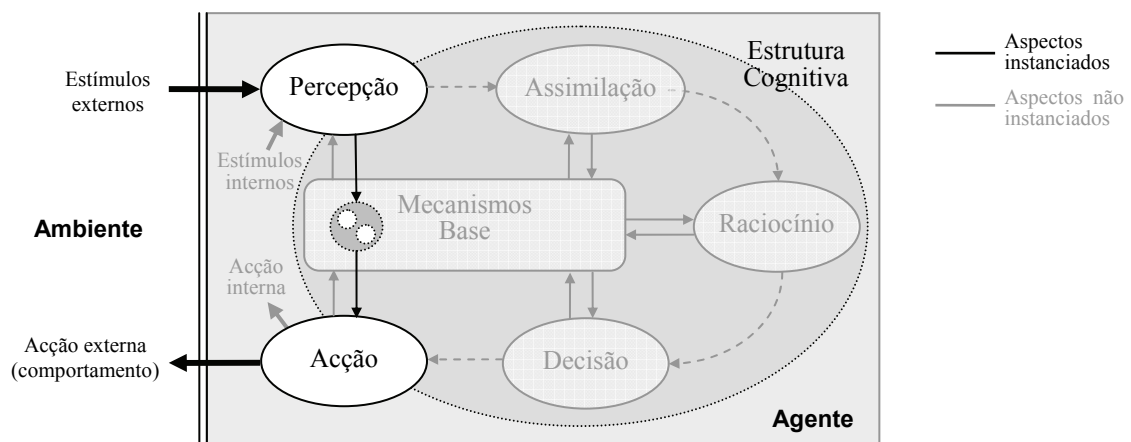


Figura 5.14. Arquitectura do agente prospector como concretização da arquitectura de agente genérica.

Neste agente não existem mecanismos base definidos. A relação entre percepção e acção é puramente reactiva, ocorrendo através da estrutura cognitiva do agente, neste caso constituída pelos dois elementos cognitivos σ_{DIR} e σ_{ESQ} , resultantes do processo de percepção.

A simplicidade da arquitectura deste agente exemplificativo permite detalhar alguns dos aspectos do modelo de agente, atrás descritos. Assim, os elementos cognitivos σ_{DIR} e σ_{ESQ} desempenham um papel de *observações*, elementos resultantes dos processos de percepção e que têm um papel de suporte da restante actividade cognitiva. No entanto, como o processamento interno se reduz à activação directa do processo de acção, estes elementos cognitivos desempenham igualmente o papel de *mediadores*, elementos caracterizados como a interface entre a actividade cognitiva do agente e a acção concreta.

Isto ilustra um aspecto importante da noção de elemento cognitivo, que é o facto do seu papel na actividade cognitiva depender do contexto cognitivo em que participa e não de características específicas desse elemento. Consequentemente, também a interpretação dos valores que os elementos cognitivos codificam é, na sua essência, subjectiva, ganhando significado no contexto cognitivo em que são manipulados. Apesar disso, é importante notar que os sinais qualitativos base podem introduzir algum pendor de significado específico, pois estão tipicamente relacionados com aspectos concretos de percepção ou representação interna. Por exemplo, em agentes mais complexos é normal surgirem dimensões cognitivas específicas para representação de mediadores. No entanto, em última análise, também o significado dessas dimensões cognitivas depende do contexto cognitivo em que estão envolvidas.

Um segundo aspecto que podemos observar tem a ver com a formação dos elementos cognitivos propriamente ditos, em termos das dimensões cognitivas que os caracterizam. Neste caso, apenas uma dimensão cognitiva é definida, a qual expressa a intensidade local dos alvos. No entanto, para compreendermos melhor este segundo aspecto é necessário evoluir para um nível seguinte de organização de agente.

5.3 Mecanismos Motivacionais

Uma das características principais do comportamento inteligente é ser orientado para a concretização de motivações. Esse carácter motivacional é essencial para que um agente possa apresentar comportamento autónomo, quer reactivo, quer pró-activo.

Como vimos anteriormente e de acordo com o modelo proposto, é também desse carácter motivacional essencial que resultam as dinâmicas cognitivas que estão subjacentes aos fenómenos emocionais. No entanto, esse carácter motivacional pode ser *implícito* ou *explícito*, consoante resulta de restrições estruturais ou funcionais embutidas na organização geral do agente, ou resulta de representações explícitas ao nível da estrutura cognitiva do agente.

A *motivação implícita* é típica de agentes reactivos, como é o caso do agente do exemplo anterior, onde as motivações se encontram embutidas nas restrições estruturais e funcionais do agente, normalmente definidas sob a forma de regras *condição* \rightarrow *acção* (ver secção 2.1.3).

A *motivação explícita* é típica de agentes deliberativos, surgindo de diferentes formas consoante a arquitectura adoptada, como é o caso de noções como *objectivo*, *desejo* ou *intenção* (ver secção 2.1.2). Em qualquer dos casos existe uma representação explícita de estados do mundo ou situações que o agente pretende atingir e que induzem e guiam a actividade cognitiva do agente.

No modelo proposto, esse papel corresponde à noção de *motivador*, isto é, elementos cognitivos que representam situações que o agente pretende atingir, actuando como forças directoras do comportamento do agente. Essas forças dão origem a dinâmicas motivacionais expressas sob a forma de disposições emocionais, tal como será discutido mais à frente.

5.3.1 Interação entre Elementos Cognitivos

Ao modelar a estrutura cognitiva de um agente a partir de agregados de sinais com dinâmicas próprias, as representações a que correspondem esses agregados deixam de ter um carácter estático e passam a ter um carácter dinâmico, podendo interagir entre si. O processo essencial subjacente a essa interação é um processo de ressonância, no qual os agregados de sinais que formam a estrutura cognitiva são activados de acordo com a semelhança com os sinais estimuladores.

A estrutura cognitiva pode assim ser vista como um sistema cuja característica global resulta da composição das características de todos os elementos cognitivos que a compõem. Quer os novos elementos cognitivos que vão sendo produzidos, quer os elementos cognitivos já existentes, podem interagir com a estrutura cognitiva, produzindo como resposta outros elementos cognitivos que expressam uma relação de semelhança entre os elementos estimuladores e os elementos da estrutura cognitiva.

Independentemente dos aspectos específicos inerentes aos fenómenos físicos de ressonância, os quais dependem de múltiplos factores, em particular da persistência temporal dos sinais envolvidos, é possível abstrair esses fenómenos em termos de um *ganho de interacção* que reflecta a semelhança entre os sinais envolvidos. Assim, a interacção entre um elemento cognitivo estimulador σ_{estim} e um elemento cognitivo base σ_{base} , é determinada por um ganho de interacção normalizado $\eta \in \mathbb{R}$, definido da seguinte forma:

$$\eta(\sigma_{estim}, \sigma_{base}) \equiv \frac{\langle \sigma_{estim}, \sigma_{base} \rangle}{\|\sigma_{base}\|^2} \quad (5.6)$$

onde $\|\sigma_{base}\|^2$ é um termo de normalização e o produto escalar $\langle \sigma_{estim}, \sigma_{base} \rangle$ expressa a semelhança entre os elementos cognitivos envolvidos. Da interacção entre um elemento cognitivo estimulador σ_{estim} e um elemento cognitivo base σ_{base} , resulta um novo elemento cognitivo σ' , definido da seguinte forma:

$$\sigma' = \eta(\sigma_{estim}, \sigma_{base}) \sigma_{base} \quad (5.7)$$

O elemento cognitivo σ' , resultante da interacção, é uma imagem modulada do elemento base original, cuja intensidade depende da semelhança (expressa pelo ganho de interacção) entre o elemento estimulador e o elemento base.

5.3.2 Interacção entre Motivadores e Observações

Podemos agora abordar a interacção concreta entre motivadores e observações, que ocorre no âmbito dos mecanismos motivacionais. Esses mecanismos produzem elementos cognitivos transientes, que são réplicas moduladas dos motivadores originais consoante a semelhança das observações com esses motivadores. Tal como acontece com a estrutura cognitiva como um todo, a interacção de uma observação σ_{OBS} e um motivador σ_{MOT} é determinada pelo ganho de interacção η (definição 5.6), onde a observação é o elemento cognitivo estimulador e o motivador é o elemento cognitivo base. Da interacção resulta um novo elemento cognitivo $\sigma' = \eta(\sigma_{OBS}, \sigma_{MOT}) \sigma_{MOT}$ (definição 5.7). A figura seguinte ilustra esquematicamente este processo.

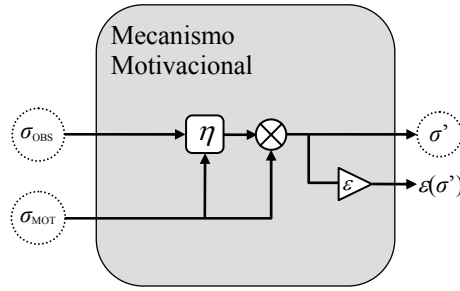


Figura 5.15. Mecanismo motivacional como suporte à interacção entre observações e motivadores.

Tal como ilustrado na figura 5.15, os mecanismos motivacionais geram igualmente um sinal $\varepsilon(\sigma')$ que corresponde à intensidade do elemento cognitivo σ' , ou seja, $\varepsilon(\sigma') = \|\sigma'\|$ (ver definição 5.5). Essa intensidade representa, na prática, uma medida de semelhança entre a observação e o motivador, sendo tanto maior quanto maior a semelhança. Note-se, no entanto, que $\sigma' = \sigma_{\text{MOT}}$ não implica que $\sigma_{\text{OBS}} = \sigma_{\text{MOT}}$, pois a intensidade do elemento cognitivo que resulta da interacção entre uma observação e um motivador é uma medida relativa da semelhança entre os dois elementos cognitivos. A figura seguinte ilustra estas relações no espaço cognitivo.

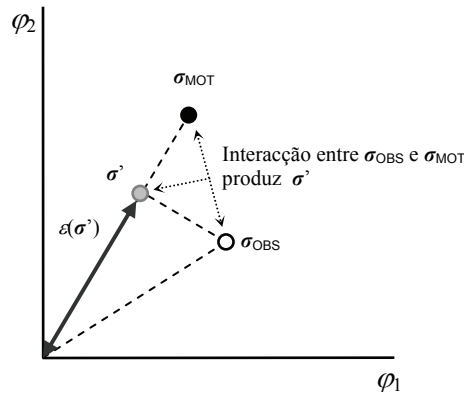


Figura 5.16. Interacção entre uma observação e um motivador descrita no espaço cognitivo.

Como se pode observar na figura 5.16, da interacção entre a observação e o motivador resulta um elemento cognitivo σ' que corresponde à projecção de σ_{OBS} sobre o vector σ_{MOT} , o que resulta directamente das definições (5.6) e (5.7). O sinal $\varepsilon(\sigma')$ corresponde à norma desse vector, indicando assim um grau de proximidade ao motivador, logo de satisfação do motivador.

5.3.3 Exemplo 2: Agente Prospector com Motivação Explícita

No exemplo anterior o agente é caracterizado por um comportamento motivado para alcançar os alvos, mas essa motivação é implícita, ou seja, é definida pela restrição estrutural de relacionar directamente o processo de percepção com o processo de acção. (ver figura 5.11). A definição de motivações de forma implícita torna-se, no entanto, inviável na prática quando a complexidade da interacção entre agente e ambiente aumenta. Mesmo quando é viável a sua definição do modo implícito, dá origem a agentes com comportamentos estereotipados e de pouca flexibilidade comportamental.

Como exemplo, consideremos o caso em que o ambiente é composto por alvos mas também por obstáculos, tendo o agente também capacidade de detectar os obstáculos. Neste caso, a interligação entre percepção e acção já não pode ser directa, implicando uma composição dos sinais resultantes da percepção dos estímulos de alvos e dos estímulos de obstáculos.

Em alternativa podemos dotar o agente de uma representação motivacional explícita. Para isso observemos que o espaço cognitivo do agente é agora composto por duas dimensões, uma associada à percepção de alvos, outra associada à percepção de obstáculos. Neste caso, o agente pode ser caracterizado pela motivação de minimizar os estímulos de obstáculos e maximizar os estímulos de alvos. A figura seguinte ilustra esta situação.

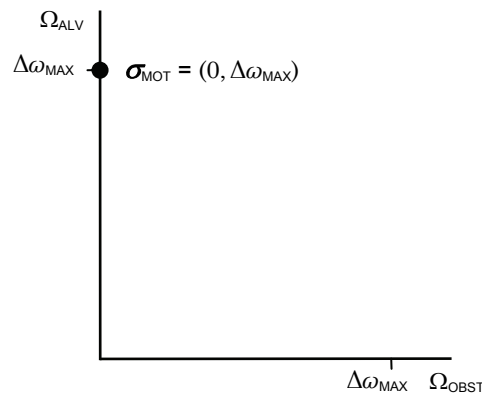


Figura 5.17. Motivador para minimizar os estímulos de obstáculos e maximizar os estímulos de alvos, num espaço cognitivo bidimensional (projectão sobre o plano de frequências; gama de variação das frequências normalizada no intervalo $[0, \Delta\omega_{MAX}]$ onde 0 representa nenhum estímulo na dimensão respectiva e $\Delta\omega_{MAX}$ representa estímulo máximo na dimensão).

Tal como anteriormente descrito, o espaço cognitivo é um espaço complexo, onde as dimensões representam quer frequência, quer intensidade. Assim, por uma questão de

clareza, na figura anterior é apresentada uma projecção no plano de frequências Ω , estando a intensidade implícita (o seu impacto será discutido mais à frente). No plano de frequências Ω , o motivador é definido pelo vector $\sigma_{\text{MOT}}^{\Omega} = (0, \Delta\omega_{\text{MAX}})$, o qual representa apenas as componentes de frequência do elemento cognitivo σ_{MOT} , tal como esquematizado na figura 5.17.

Um dos aspectos que evidenciam a diferença de flexibilidade entre um *motivador* e uma *restrição estrutural* é a facilidade de alterar as características do motivador, quer em termos qualitativos (por exemplo, para levar o agente a aproximar-se também de obstáculos bastaria alterar o motivador para $\sigma_{\text{MOT}}^{\Omega} = (\Delta\omega_{\text{MAX}}, \Delta\omega_{\text{MAX}})$), quer em termos de intensidade associada às dimensões cognitivas do motivador, o que será ilustrado mais à frente.

Para já, consideremos intensidades iguais e unitárias para ambas as dimensões cognitivas do motivador, ou seja, $\sigma_{\text{MOT}}^{\mathcal{E}} = (1, 1)$, onde $\sigma_{\text{MOT}}^{\mathcal{E}}$ representa apenas as componentes de intensidade do elemento cognitivo σ_{MOT} . Com esta configuração do motivador, a influência motivacional de ambos os tipos de estímulo é igual. A figura seguinte ilustra a arquitectura do agente prospector com motivação explícita.

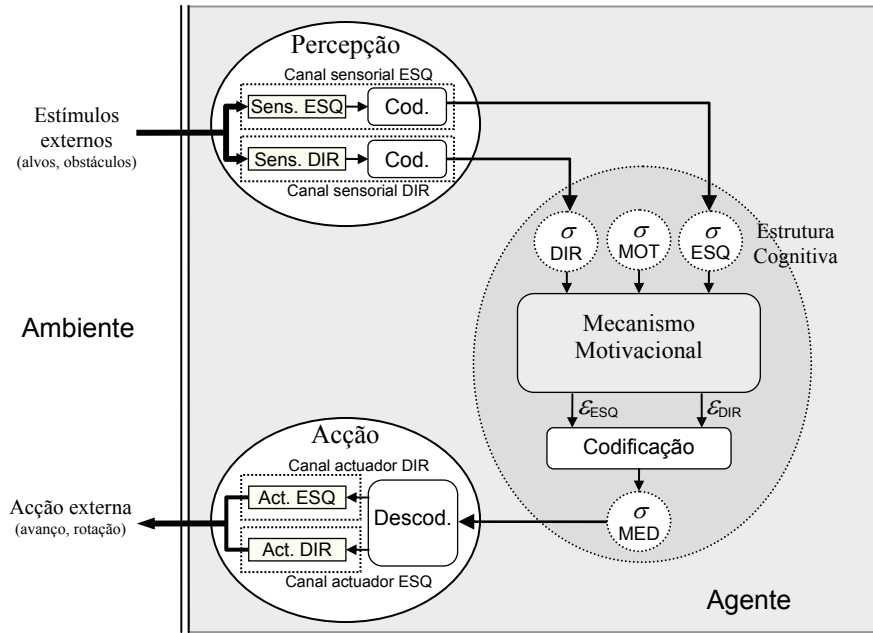


Figura 5.18. Agente prospector com motivação explícita: processos de percepção, acção e mecanismo motivacional.

Neste caso são geradas duas observações, σ_{ESQ} e σ_{DIR} , que interagem com o motivador σ_{MOT} através do mecanismo motivacional, sendo gerados, respectivamente, um sinal ϵ_{ESQ}

e um sinal ε_{DIR} , que reflectem as intensidades dos elementos cognitivos σ_{ESQ} e σ_{DIR} resultantes da interacção. Esses dois sinais são de seguida codificados (ver secção 5.2.2), formando um mediador σ_{MED} . Para esse efeito são definidas duas outras dimensões específicas do espaço cognitivo, φ_{ESQ} e φ_{DIR} , associadas à activação dos actuadores esquerdo e direito, respectivamente, pelo que o espaço cognitivo passa a ser caracterizado por quatro dimensões (φ_{ALV} , φ_{OBS} , φ_{ESQ} , φ_{DIR}). A figura seguinte mostra o comportamento do agente num ambiente com os mesmos três alvos do ambiente anterior, mas agora com obstáculos entre os alvos (representados por segmentos de recta).

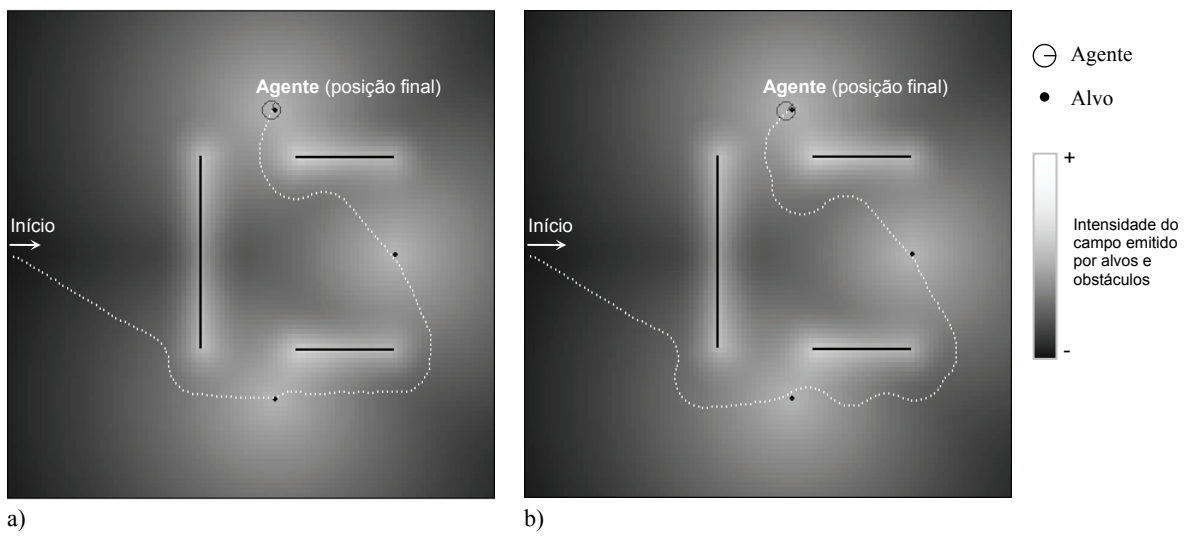


Figura 5.19. Comportamento do agente para diferentes intensidades das componentes do motivador: (a) $\sigma_{\text{MOT}}^{\varepsilon} = (1, 1, 0, 0)$; (b) $\sigma_{\text{MOT}}^{\varepsilon} = (1, 2, 0, 0)$.

Como se pode observar na figura 5.19.a, o agente afasta-se dos obstáculos e aproxima-se dos alvos, de acordo com o motivador definido. Na figura 5.19.b é possível observar o comportamento do agente para uma intensidade maior na dimensão associada aos estímulos de obstáculos ($\sigma_{\text{MOT}}^{\varepsilon} = (1, 2, 0, 0)$). Como se pode observar, o agente é mais sensível aos obstáculos, ganhando os movimentos perto dos obstáculos uma maior amplitude.

Este exemplo ilustra um aspecto importante associado à noção de motivador, anteriormente referido, que é o facto dos motivadores terem um carácter não apenas qualitativo, mas também um carácter quantitativo, expresso pela *intensidade* do motivador. Como é ilustrado no exemplo anterior, esse aspecto quantitativo tem um impacto importante no comportamento do agente, não contemplado directamente nas representações simbólicas de motivação.

Um segundo aspecto que é interessante realçar, é o facto da formação explícita do mediador σ_{MED} poder ser vista como um embrião de um processo de decisão. De facto, se representada algoritmicamente, essa operação pode ser vista como uma actividade de decisão acerca de qual a acção a realizar, tal como ilustrado na figura seguinte.

```
// Decidir acção

 $\Delta\epsilon = \epsilon_{\text{ESQ}} - \epsilon_{\text{DIR}}$  // Diferença de intensidade

IF  $\Delta\epsilon > 0$  THEN
{
     $\sigma_{\text{MED}}^{\Omega} = (0, 0, \Delta\epsilon, 0)$  // Rodar à esquerda
     $\sigma_{\text{MED}}^{\epsilon} = (0, 0, 1, 1)$ 
}
ELSE IF  $\Delta\epsilon < 0$  THEN
{
     $\sigma_{\text{MED}}^{\Omega} = (0, 0, 0, -\Delta\epsilon)$  // Rodar à direita
     $\sigma_{\text{MED}}^{\epsilon} = (0, 0, 1, 1)$ 
}
ELSE
{
     $\sigma_{\text{MED}}^{\Omega} = (0, 0, 0, 0)$  // Manter direcção
     $\sigma_{\text{MED}}^{\epsilon} = (0, 0, 1, 1)$ 
}
```

Figura 5.20. Decisão de acção perante o resultado dos mecanismos motivacionais.

É aqui que reside a principal diferença em relação à arquitectura do agente do exemplo anterior, uma vez que neste caso há uma geração explícita de um mediador, no que pode ser visto como o embrião de uma intenção. Esta caracterização é interessante uma vez que ilustra a continuidade funcional na evolução da arquitectura dos agentes, referida anteriormente. A figura seguinte mostra a concretização desta versão da arquitectura do agente prospector no contexto da arquitectura de agente genérica.

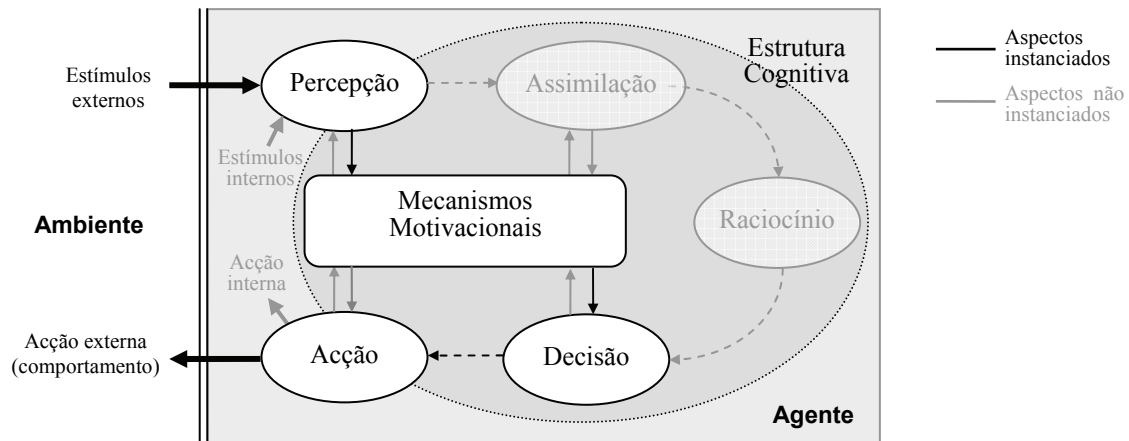


Figura 5.21. Arquitectura do agente prospector como concretização da arquitectura de agente genérica.

No entanto, a expressão das noções de decisão e de intenção só se tornam plenas de significado quando outros processos cognitivos proporcionarem múltiplas opções para deliberação. Antes de abordarmos esse contexto é necessário abordar outros mecanismos base, os mecanismos de disposição emocional.

5.4 Mecanismos de Disposição Emocional

Após termos definido uma arquitectura de agente concreta e os mecanismos motivacionais base de um agente, é agora possível definir também de forma concreta a noção de disposição emocional, uma das noções base dos modelos de emoção e de agente propostos, bem como os mecanismos que estão subjacentes à sua formação.

5.4.1 Descrição das Dinâmicas Emocionais no Espaço Cognitivo

À medida que um agente interage com o ambiente, ocorrem mudanças na sua estrutura cognitiva resultantes do comportamento realizado, as quais produzem trajetórias no espaço cognitivo. Sendo o comportamento de um agente guiado pela relação entre os motivadores que o caracterizam e a percepção da situação actual, expressa pelas observações geradas, a actividade cognitiva do agente é, consequentemente, guiada pela maximização dos fluxos de concretização que levam à redução da distância entre motivadores e observações. Este processo pode ser descrito com base no movimento que motivadores e observações descrevem no espaço cognitivo, onde os motivadores e as observações, num determinado instante, correspondem a

posições específicas, e os mediadores definem direcções de movimento, tal como ilustrado na figura seguinte.

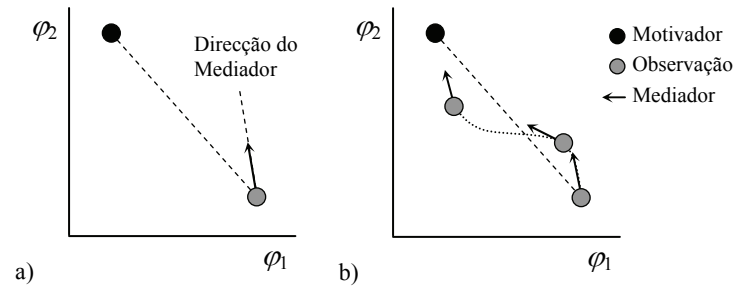


Figura 5.22. Elementos participantes na concretização de um motivador num espaço cognitivo bidimensional.

Tal como ilustrado na figura 5.22.a, a direcção dos mediadores seleccionados pode não ser a direcção exacta do motivador. Para além disso, os motivadores podem mudar e as dinâmicas do ambiente, quer externas, quer internas, podem igualmente influenciar o movimento das observações. Isto significa que o agente necessita de continuamente ajustar o seu comportamento, em especial em ambientes incertos e dinâmicos. A figura 5.22.b mostra uma possível trajectória resultante do ajuste do comportamento do agente através da selecção de mediadores diferentes.

Estes novos mediadores podem ser resultado de novas opções comportamentais ou o refinamento dos mediadores anteriores através de processos de planeamento. No entanto, independentemente dos processos específicos que geraram esses mediadores, as forças que levaram a essa mudança expressam dinâmicas motivacionais resultantes da relação entre os motivadores do agente e a percepção da forma como o ambiente evolui, representada sob a forma de observações. No modelo proposto, os fenómenos emocionais são a expressão dessas dinâmicas, caracterizadas como *disposições emocionais* (definição 4.2). No espaço cognitivo essas dinâmicas são expressas pelo movimento das observações em relação aos motivadores, tal como ilustrado na figura seguinte.

5.4.2 Potenciais de Disposição Emocional e Sinais Afectivos

Os mecanismos de disposição emocional detectam e transformam as dinâmicas constituintes das disposições emocionais, produzindo sinais concretos. Dados uma observação e um motivador, são gerados dois tipos de sinais:

- potenciais cognitivos de disposição emocional p_s e p_v ;
- sinais afectivos λ^+ e λ^- que correspondem ao carácter afectivo subjacente a esses potenciais.

Os potenciais cognitivos p_s e p_v têm dimensões qualitativas denotadas pelas frequências base Ω_s e Ω_v , e valores para essas qualidades que expressam as dinâmicas emocionais base δs e δv . Por sua vez, estas dinâmicas emocionais expressam um carácter hedónico associado ao aumento ou diminuição do “bem-estar” do agente em relação à concretização dos seus motivadores. Ou seja, valores positivos de δs ou de δv expressam uma melhoria nas condições de concretização, correspondendo, por isso, a uma valência positiva. Em contrapartida, valores negativos expressam uma deterioração das condições de concretização, correspondendo a uma valência negativa. Este aspecto de valência, agradável/desagradável ou positiva/negativa, corresponde à noção de afecto, tal como foi referido anteriormente (ver secção 2.2.1).

As situações mais favoráveis para o agente ocorrem quando δs e δv são em simultâneo positivas. No plano de disposição emocional, estas situações correspondem a vectores de *disposição emocional* localizados no quadrante Q-I. De igual modo, as situações mais desfavoráveis ocorrem quando δs e δv têm, em simultâneo, valências negativas. Essa situação corresponde a vectores de disposição emocional localizados no quadrante Q-III. Estas duas situações extremas são representadas pelos vectores de disposição emocional (DE) de referência $A^+ = (1,1)$ e $A^- = (-1,-1)$, tal como ilustrado na figura seguinte.

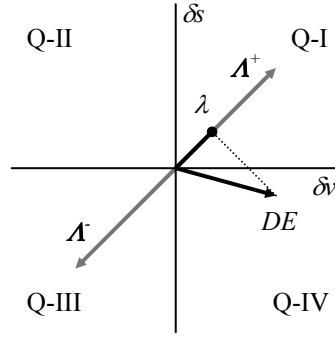


Figura 5.24. Caracterização afectiva no plano de disposição emocional.

A projecção de um vector de disposição emocional sobre um desses vectores A produz um valor afectivo $\lambda \in \mathbb{R}$ que expressa a intensidade positiva (λ^+) ou negativa (λ^-) da valência associada às dinâmicas emocionais. Ou seja:

$$\lambda^+ = \begin{cases} \text{proj}(DE, A^+) & \text{se } \text{proj}(DE, A^+) > 0 \\ 0 & \text{se } \text{proj}(DE, A^+) \leq 0 \end{cases} \quad (5.9)$$

$$\lambda^- = \begin{cases} \text{proj}(DE, A^-) & \text{se } \text{proj}(DE, A^-) > 0 \\ 0 & \text{se } \text{proj}(DE, A^-) \leq 0 \end{cases}$$

onde $\text{proj}(\mathbf{x}, \mathbf{y})$ denota a projecção ortogonal do vector \mathbf{x} sobre o vector \mathbf{y} . Em conjunto com os potenciais cognitivos p_s e p_v , os sinais λ^+ e λ^- são as principais fontes reguladoras da actividade cognitiva do agente, aspecto que será explorado no próximo capítulo.

5.4.3 Mecanismo de Acumulação de Disposição Emocional

Tal como acontece ao nível dos processos cognitivos, também os mecanismos de disposição emocional expressam uma tendência para a especialização e centralização em níveis de organização progressivamente mais complexos. No caso dos mecanismos de disposição emocional, estes caracterizam-se por um papel regulador não apenas local, mas também, e sobretudo, global. Para esse efeito um aspecto importante é a geração de sinais que expressem o efeito cumulativo das múltiplas disposições emocionais que resultam da actividade cognitiva do agente. Tendo por base o modelo de emoção proposto, isso é conseguido através de um mecanismo de acumulação quer das dinâmicas emocionais δs e δv , quer dos sinais afectivos λ^+ e λ^- , no entanto, esses

fenómenos ocorrem no âmbito de mecanismos específicos de regulação cognitiva, pelo que serão analisados mais à frente.

Tal como descrito no contexto do modelo de emoção proposto (ver secção 4.3.3), os mecanismos de acumulação são mecanismos de integração temporal. No caso das disposições emocionais são produzidos dois sinais globais δ^G_s e δ^G_v , definidos da seguinte forma:

$$\frac{d\delta^G_s}{dt} = \alpha_s \cdot \delta s \quad \text{e} \quad \frac{d\delta^G_v}{dt} = \alpha_v \cdot \delta v \quad (5.10)$$

onde α_s e α_v representam coeficientes de acumulação. Estes sinais δ^G_s e δ^G_v expressam, respectivamente, as noções de *potencial de situação* (P_{sit}) e de *fluxo de concretização* (F), apresentadas no âmbito do modelo emocional de fluxo.

Tal como descrito no capítulo anterior, os coeficientes de acumulação α_s e α_v podem ser fixos ou variar de acordo com condições específicas do agente. Por uma questão de clareza e simplicidade, e por não ser o nosso principal objectivo a modelação de aspectos específicos associados aos fenómenos emocionais, os coeficientes de acumulação serão considerados constantes. Este aspecto é, no entanto, merecedor de atenção futura, em particular no que respeita à influência deste tipo de fenómenos na regulação da actividade cognitiva do agente.

Os sinais δ^G_s e δ^G_v , para além de participarem nos processos de regulação da actividade cognitiva, descritos mais à frente, permitem igualmente caracterizar a *situação emocional* (SE) global do agente, ou seja:

$$SE \equiv (\delta^G_s, \delta^G_v) \quad (5.11)$$

A situação emocional pode ser vista como uma base para a emergência de um *estado emocional*, ou para a modelação da noção de *emoção*, como um fenómeno global à actividade cognitiva do agente. No entanto, por si só não pode ser descrita como tal, pois não abrange aspectos importantes, como é o caso da expressão comportamental ou da percepção subjectiva, os quais, como vimos anteriormente, são fundamentais para caracterizar a noção de emoção.

5.4.4 Exemplo 3: Agente Prospector com Mecanismos Emocionais

Para ilustrar o funcionamento dos mecanismos de disposição emocional, consideremos o agente prospector do exemplo 2, agora complementado com

mecanismos de disposição emocional, mas sem que esses mecanismos tenham qualquer influência na actividade cognitiva ou no comportamento do agente.

Uma vez que este agente produz duas observações, uma referente ao sensor da esquerda e outra referente ao sensor da direita, também são geradas duas disposições emocionais. Essas duas disposições emocionais são combinadas pelos mecanismos de disposição emocional, sendo gerada uma disposição emocional total que é a média das duas disposições emocionais originais. No caso geral, a composição de disposições emocionais pode ser mais elaborada envolvendo, por exemplo, a influência dos elementos σ que resultam da interacção entre observações e motivadores (ver secção 5.3.2). Este aspecto é importante se a intensidade dos motivadores variar como resultado da actividade do agente, aspecto que não foi explorado no trabalho realizado, constituindo uma área para desenvolvimento futuro.

Consideremos agora um cenário com apenas um alvo. Consideremos, também, que o agente só tem capacidade de detectar o obstáculo quando está bastante próximo deste. As figuras 5.25 e 5.26 mostram o comportamento do agente, bem como a respectiva evolução da disposição emocional e da situação emocional (resultante do efeito cumulativo das disposições emocionais).

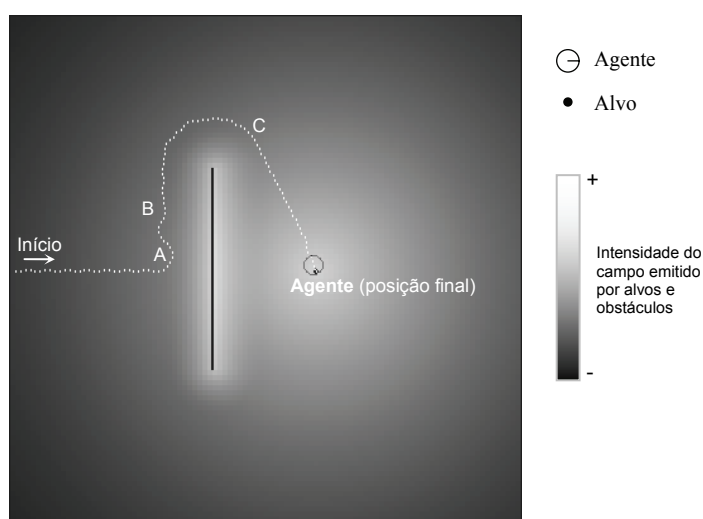


Figura 5.25. Comportamento do agente num ambiente com um alvo e um obstáculo.

Como se pode observar na figura 5.25, de início o agente dirige-se directamente para o alvo, pois não consegue detectar o obstáculo mais à frente. Durante esse período observa-se uma evolução favorável quer da disposição emocional quer da situação emocional, tal como ilustrado na figura 5.26, evoluindo ambas na transição

entre os quadrantes Q-I e Q-II, e apresentando ambas baixa intensidade, sendo por isso tendências essencialmente neutras.

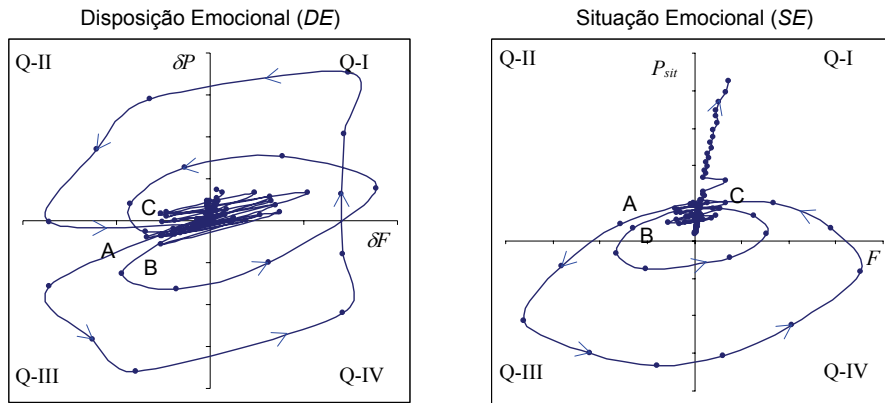


Figura 5.26. Evolução da disposição emocional e da situação emocional do agente.

Quando o agente subitamente encontra o obstáculo, são geradas dinâmicas motivacionais que levam o agente a recuar, afastando-se do obstáculo (**A**), contornando de seguida o obstáculo até encontrar um caminho livre para o alvo (**C**). A expressão dessas dinâmicas pode ser observada como um pico de disposição emocional (*DE*) no quadrante Q-III (tendência de *medo*), o que leva a situação emocional (*SE*) a evoluir também para o quadrante Q-III. Podemos, assim, observar uma consonância entre o comportamento do agente, com uma expressão que pode ser interpretada como um “sobressalto”, indiciador de uma situação típica de medo, e o carácter igualmente de medo da disposição emocional subjacente. Este padrão de evolução repete-se novamente quando o agente se aproxima do obstáculo em (**B**), só que agora numa escala mais reduzida, pois o movimento do agente já é no sentido de contornar o obstáculo.

Podemos igualmente observar que, quando o agente encontra o caminho livre para o alvo (**C**), a situação emocional ganha um carácter progressivamente mais favorável no quadrante Q-I, correspondendo a tendências emocionais do tipo *alegria/satisfação*, resultantes do fim da adversidade. Note-se, no entanto, que esta caracterização é feita na perspectiva de um observador externo.

A exploração da relação entre estes fenómenos emocionais embrionários e padrões comportamentais específicos, os quais constituem uma das vertentes de fenómenos emocionais mais complexos, é uma área merecedora de exploração em termos de trabalho futuro. O nosso foco vai ser no entanto a relação entre os fenómenos emocionais e processos deliberativos envolvendo raciocínio e adaptação.

5.5 Conclusão

O modelo de agente apresentado neste capítulo concretiza o suporte necessário para tirar partido da perspectiva dinâmica e contínua dos fenómenos emocionais, que caracteriza o modelo de emoção proposto no capítulo anterior. Para esse efeito, um dos aspectos importantes é a definição de uma estrutura cognitiva que, pelo facto de ser modelada em termos de agregados de sinais com dinâmicas próprias, permite expressar de modo adequado quer aspectos emocionais e sub-conceptuais, quer representações conceptuais estruturadas e complexas, sem perder o carácter contínuo e dinâmico dos fenómenos que lhes estão subjacentes. Isso torna possível a caracterização dos fenómenos emocionais de forma concisa, em termos da evolução relativa dos elementos que constituem a estrutura cognitiva, em particular no que se refere a *motivadores* e *observações*. Deste modo, as *disposições emocionais* surgem como resultado da actividade cognitiva de um agente, sendo tornadas explícitas pelos mecanismos de disposição emocional sob a forma de potenciais cognitivos e de sinais afectivos, os quais, por sua vez, influenciam a actividade cognitiva de um agente, como se verá no capítulo seguinte.

Um segundo aspecto importante do modelo de agente proposto, é o seu não comprometimento com contextos de utilização ou tipos de agentes específicos. Em vez disso, o que se propõe é um contínuo organizacional. Para isso é definida uma arquitectura de agente genérica, na qual o tipo e organização específica dos processos cognitivos que a caracterizam não são impostos, sendo antes definida uma estrutura cognitiva uniforme e um conjunto de mecanismos de base. Deste modo, é possível suportar a implementação de agentes de diferentes tipos e níveis de complexidade. As características concretas dos agentes implementados dependem de dois aspectos principais: o *tipo de estrutura cognitiva* e o *tipo de processos cognitivos* que caracterizam o agente. A figura seguinte ilustra a relação entre estes aspectos e algumas características principais dos agentes.

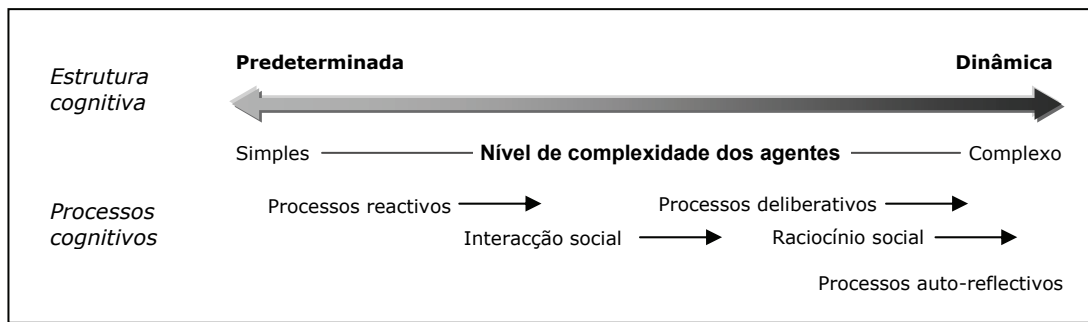


Figura 5.27. Relação entre características do modelo de agente e características dos agentes.

Os agentes mais simples são caracterizados por uma estrutura cognitiva predeterminada e por processos cognitivos de carácter reactivo. O comportamento desses agentes é essencialmente guiado pelas dinâmicas motivacionais que resultam de potenciais e fluxos instantâneos.

É interessante notar que mesmo estes processos simples podem produzir comportamentos sofisticados, pelo menos considerando o comportamento observável em organismos biológicos. Por exemplo, bactérias e outros tipos de organismos unicelulares, apesar de não terem sistema nervoso e possuírem apenas mecanismos sensoriais “rudimentares”, quando comparados com os de outros organismos, e apesar da sua aparente simplicidade, apresentam comportamentos notáveis, sendo capazes de evitar predadores e encontrar fontes de alimento de forma eficaz. Em alguns organismos, como o *Stentor roeselii* [Staddon, 2001], é possível mesmo observar comportamentos que, na perspectiva de um observador externo, podem ser caracterizados como de carácter emocional [Damásio, 2003].

Agentes mais complexos, como é o caso de agentes deliberativos, são caracterizados por uma estrutura cognitiva dinâmica, cujo conteúdo e organização é explicitamente manipulado e alterado pelos processos cognitivos no sentido de antecipar situações futuras a partir de experiências ocorridas no passado. Este carácter antecipador, baseado em fenómenos de adaptação de longo prazo, é claramente distinto do carácter reactivo, ou de adaptação de curto prazo, que caracteriza agentes mais simples.

É a este nível que a relação entre fenómenos emocionais e cognitivos se torna particularmente relevante, nomeadamente a relação entre fenómenos emocionais e os processos de raciocínio e de tomada de decisão, tal como as evidências experimentais discutidas nos capítulos 1 e 2 mostram. É essa relação que será explorada no próximo capítulo.

Capítulo 6

Integração de Emoção e Raciocínio

Sendo os recursos computacionais de um agente finitos e o tempo de resposta limitado, a actividade cognitiva do agente terá de ser confinada ao tempo e recursos disponíveis. Em ambientes reais, onde a permanente mudança é uma característica chave, o confinar da actividade cognitiva deverá reflectir, de modo dinâmico, a evolução das condições do ambiente, caso contrário haverá desperdício de recursos ou de oportunidades. Para isso, a ligação ao ambiente deverá ser permanente e não fragmentada e esporádica.

No paradigma cognitivo, a ênfase na representação interna do mundo e na sua manipulação simbólica levou a um desacoplar do ambiente, desempenhando a interacção com o ambiente um papel secundário. Em contrapartida, no paradigma comportamental, a relação com o ambiente é enfatizada à custa da eliminação das representações internas e dos processos de raciocínio que as manipulam. Essa opção tem, no entanto, um impacto importante. Sem a possibilidade de formar representações internas que reflectam a experiência acumulada de um agente, independentemente de serem simbólicas ou não, e sem processos de raciocínio que permitam explorar essa experiência acumulada de modo explícito, torna-se impossível antecipar situações futuras e explorar possíveis cursos de acção alternativos, ficando o agente limitado a um encadear de padrões de *estímulo-resposta*, independentemente do nível de organização cognitiva em que ocorram. De facto, o que se observa em organismos biológicos, mesmo nos mais simples, como os insectos, é a existência de comportamentos que claramente indiciam a formação de um modelo interno do mundo e, sobretudo, a capacidade de explorar esse modelo no sentido de antecipar situações futuras (e.g. [Giurfa & Menzel, 2003; Kühn *et al.*, 2003]), sem que isso ocorra, necessariamente, num nível simbólico.

Por sua vez, nos modelos cognitivos, as representações internas e os processos de raciocínio são um elemento central, contudo, a interacção com o ambiente resume-se a uma fonte de informação (“input”) que alimenta uma representação virtual do mundo, sobre a qual os processos cognitivos operam, gerando informação (“output”), que é transferida para o ambiente. É esta perspectiva de processamento de informação que tem caracterizado o paradigma cognitivo.

O que está omisso nesta perspectiva é que, enquanto o processamento cognitivo ocorre, o ambiente está a mudar e essa mudança necessita de se reflectir na actividade cognitiva do agente. É este facto que Simon reconhece quando propõe que a emoção pode desempenhar um papel importante na actividade cognitiva (ver secção 2.2.5). Para Simon, a emoção seria essencialmente um mecanismo de interrupção da actividade cognitiva, através do qual os processos cognitivos poderiam ser direccionados perante mudanças ocorridas no ambiente. Apesar desta perspectiva ser muito limitada, como foi anteriormente discutido, ela realça já o papel que os fenómenos emocionais podem ter no acoplamento entre actividade cognitiva e ambiente, no sentido da regulação da actividade cognitiva de acordo com a evolução do ambiente.

Actualmente, os dados experimentais mostram claramente uma relação directa entre os fenómenos emocionais e a regulação da actividade cognitiva. Essa relação inclui aspectos como atenção, criatividade, tomada de decisão, precisão da acção ou até mesmo a percepção subjectiva da passagem do tempo (e.g. [Damásio, 2003; Angrilli *et al.*, 1997; Coren *et al.*, 2004]).

Com base nos modelos de emoção e de agente propostos nos capítulos anteriores, é agora possível abordar esta relação entre fenómenos emocionais e fenómenos cognitivos, em particular no que se refere aos processos base da actividade cognitiva de alto nível, os processos de raciocínio. É essa relação que será explorada de seguida.

6.1 Adaptação e Regulação da Actividade Cognitiva

Toda a actividade concreta é caracterizada por consumo de recursos e de tempo. Os recursos podem ser computacionais, como memória ou capacidade de processamento, ou de carácter mais geral, como consumo de energia. Em ambientes reais esses recursos são limitados. Por sua vez, o dinamismo que caracteriza os ambientes reais

traduz-se na necessidade de comportamento em tempo-real, de modo a garantir uma resposta atempada perante a permanente mudança.

Em contrapartida, a noção de racionalidade subjacente ao comportamento inteligente (ver secção 2.1.5) implica que um agente seja capaz de produzir o melhor comportamento possível (ou solução para o problema que o agente está a tentar resolver) perante as condições em que se encontra, o que exige recursos e tempo. Ou seja, existe um conflito essencial entre os requisitos necessários à qualidade da solução e o tempo e recursos disponíveis para essa solução ser gerada.

Num contexto cognitivo, quando se fala de qualidade da solução gerada, está subjacente que se tem em conta os processos de raciocínio e tomada de decisão, em particular, raciocínio com limitações de recursos (“resource-bounded reasoning”) [Russell, 1999; Gigerenzer & Selten, 1999; Schut *et al.*, 2004]. No entanto, o problema da limitação de recursos e de tempo afecta todos os processos cognitivos do agente. Por exemplo, a acção pode ser mais ou menos precisa, a percepção mais limitada ou mais abrangente, a memória mais ou menos detalhada.

Neste sentido, é possível falar-se de um compromisso essencial subjacente a toda actividade cognitiva do agente (i.e., a actividade conjunta dos diferentes tipos de processos cognitivos de um agente), que consiste no balanço entre qualidade de solução e tempo e recursos consumidos.

Esse compromisso não pode ser definido de forma estática, por exemplo em tempo de projecto, pelo contrário necessita de ser dinâmico de modo a poder reflectir a evolução da relação entre agente e ambiente num contexto de permanente mudança. Ou seja, esse compromisso dinâmico tem um carácter inerentemente adaptativo e relacionado com o acoplamento entre agente e ambiente. São precisamente estes os atributos chave que caracterizam os fenómenos emocionais no contexto da abordagem proposta.

Nesse sentido, os fenómenos emocionais proporcionam o suporte necessário para a definição de mecanismos capazes de garantir um compromisso entre qualidade de solução e tempo e recursos consumidos, estabelecendo a ligação entre actividade cognitiva e evolução do ambiente, de modo dinâmico e contínuo. No modelo proposto, essa relação é caracterizada por duas vertentes:

- *regulação da actividade cognitiva* do agente como resultado das condições de concretização correntes, a que corresponde uma *adaptação de curto prazo*;

- *modulação das alterações na estrutura cognitiva* do agente como resultado das experiências acumuladas ao longo do tempo, a que corresponde uma *adaptação de longo prazo*.

A modulação das alterações na estrutura cognitiva por processos de base emocional tem como efeito a formação de memórias emocionais. Essas memórias incorporam o carácter regulador prevalecente na altura em que são formadas, sob a forma de disposições emocionais. Posteriormente, quando as memórias emocionais são activadas, essas disposições emocionais vão agir sobre os próprios mecanismos emocionais que, como iremos ver, estão na base dos mecanismos de regulação da actividade cognitiva, produzindo um efeito regulador de carácter antecipador. Deste modo, ambas as vertentes adaptativas estão directamente relacionadas com a regulação da actividade cognitiva de um agente.

6.2 Mecanismos de Regulação Cognitiva

Por *regulação cognitiva* entende-se a capacidade de adaptar a execução dos processos cognitivos de um agente de acordo com as condições de concretização observadas, quer no presente, quer no passado, de modo a obter um compromisso entre tempo e recursos consumidos e a qualidade da solução obtida.

Não sendo imposta qualquer restrição específica em relação à concepção e implementação dos processos cognitivos que caracterizam a arquitectura de agente proposta, para tornar possível esta regulação é requerido que esses processos garantam uma relação monótona e não decrescente entre qualidade da solução produzida e os recursos e tempo consumidos. Este é um requisito clássico dos métodos de raciocínio com limitação de recursos, nomeadamente no que se refere a métodos de procura e planeamento perante limitações de recursos e de tempo (e.g. [Korf, 1990; Dean & Kaelbling, 1993; Russell & Norvig, 2003; Schut *et al.*, 2004]).

Deste modo, o que se propõe não são formas específicas de implementar os processos cognitivos, nomeadamente em termos de propor novos algoritmos ou de alterar os já existentes, mas sim a capacidade de controlar de forma dinâmica o contexto de execução dos processos cognitivos, nomeadamente os parâmetros dessa execução.

No modelo proposto, esse controlo ocorre de acordo com duas perspectivas base:

- uma perspectiva *espacial*, que se refere ao espaço de elementos cognitivos sobre os quais o processamento pode ocorrer;
- uma perspectiva *temporal*, que se refere ao tempo disponível para processamento cognitivo.

Na perspectiva espacial, a acessibilidade dos processos cognitivos à estrutura cognitiva de um agente é restringida, no sentido de limitar os elementos cognitivos sobre os quais o processamento pode ocorrer (a entrada ou “input” para esses processos). Sendo a complexidade computacional desses processos uma função monótona e crescente dos dados processados, o efeito é de delimitar na prática o valor dessa função, embora sem alterar a classe de complexidade dos processos em causa, ou seja, delimitar os recursos necessários para a execução.

No entanto, o delimitar da complexidade computacional, por si só, não garante que o processamento possa ser realizado dentro dos limites de tempo e de outros recursos disponíveis e que, simultaneamente, as soluções produzidas sejam adequadas à concretização das motivações do agente perante as condições de concretização observadas. Na prática, a elevada complexidade computacional de processos como raciocínio e tomada de decisão, a par com a complexidade inerente aos ambientes reais, faz com que a restrição das entradas necessária para tornar esses processos viáveis implique uma restrição drástica do espaço de hipóteses explorado, com uma consequente degradação importante da qualidade da solução produzida.

Nesse sentido, para além da restrição das entradas sobre as quais os processos cognitivos operam, é necessário controlar outros parâmetros de execução. A perspectiva temporal desempenha esse papel através da geração de uma indicação da *premência* da resposta. Ou seja, de novo não é imposta qualquer restrição específica sobre a implementação dos processos cognitivos. Em vez disso, é disponibilizada informação a esses processos para que o respectivo processamento interno possa ser adaptado às condições de concretização observadas.

A partir da indicação da premência da resposta, a actividade dos processos cognitivos pode ser adaptada, alterada ou redireccionada no sentido de garantir uma resposta atempada, num compromisso entre qualidade da solução e premência da resposta. Note-se que a indicação de premência não é, por si, uma restrição temporal.

Como iremos ver, a indicação de premência é antes uma indicação da intensidade e do carácter da mudança que está a ocorrer.

Quer a restrição do espaço de elementos cognitivos sobre os quais os processos cognitivos podem operar, quer a indicação da premência da resposta a que devem obedecer, traduzem-se na focagem da actividade desses processos. No modelo proposto, estas duas perspectivas de focagem, *espacial* e *temporal*, correspondem a dois mecanismos base, um mecanismo de *focagem de atenção* e um mecanismo de *focagem temporal*. Ambos os mecanismos de focagem dependem dos sinais produzidos pelos mecanismos de disposição emocional, anteriormente descritos. A figura seguinte ilustra a forma como estes vários mecanismos estão interrelacionados.

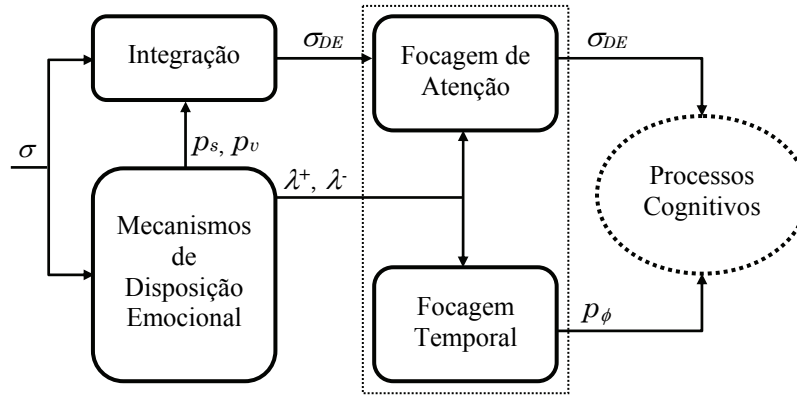


Figura 6.1. Mecanismos subjacentes à regulação da actividade cognitiva de um agente.

Os potenciais cognitivos p_s e p_v e os sinais afectivos λ^+ e λ , produzidos pelos mecanismos de disposição emocional (ver secção 5.4), são o suporte regulador de ambos os mecanismos de focagem. Os potenciais cognitivos p_s e p_v são integrados com os elementos cognitivos que lhes estão na origem, observações e motivadores. Os elementos resultantes desta integração, representados na figura 6.1 por σ_{DE} , são subsequentemente sujeitos à focagem de atenção, o que determinará a sua participação, ou não, no processamento cognitivo. Os detalhes destes mecanismos são discutidos de seguida.

6.2.1 Mecanismo de Integração

O mecanismo de integração gera agregados de elementos cognitivos que expressam a relação entre observações e motivadores, bem como a disposição emocional que resulta dessa relação. Não tendo uma influência directa na regulação da actividade

cognitiva é, porém, através deste mecanismo que as disposições emocionais são incorporadas nos elementos cognitivos, servindo posteriormente de base quer à regulação da actividade cognitiva, quer à formação de memórias emocionais. A figura seguinte ilustra a operação deste mecanismo.

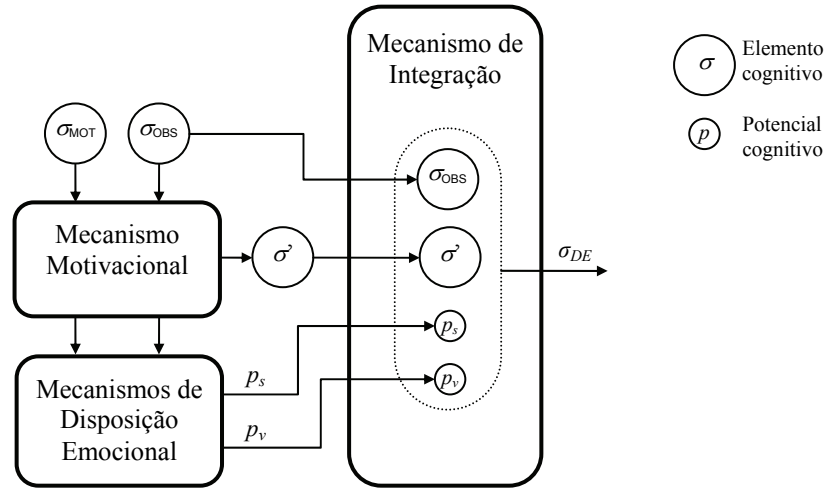


Figura 6.2. Mecanismo de integração.

Perante uma observação (σ_{OBS}) e um motivador (σ_{MOT}), o mecanismo motivacional gera um elemento cognitivo σ' que resulta da interacção entre a observação e o motivador (ver secção 5.3.2). Por sua vez, os mecanismos de disposição emocional geram os potenciais cognitivos p_s e p_v que expressam a disposição emocional resultante da relação entre observação e motivador (ver secção 5.4).

O mecanismo de integração agrega σ_{OBS} , σ' , p_s e p_v , produzindo um agregado de elementos cognitivos σ_{DE} . Estes agregados de elementos cognitivos vão posteriormente ser a base da formação de memórias emocionais (processo de assimilação) e dos processos de raciocínio e de decisão.

6.2.2 Mecanismo de Focagem de Atenção

O mecanismo de focagem de atenção restringe a atenção dos processos cognitivos de um agente a certos elementos cognitivos, de acordo com o seu conteúdo de disposição emocional, ou seja, os potenciais cognitivos p_s e p_v . Este mecanismo actua como uma barreira de depleção, produzindo um *campo de atenção* formado pelos elementos cognitivos capazes de ultrapassar a barreira. Apenas os elementos no campo de atenção são considerados pelos processos cognitivos de alto nível, como é o caso dos

processos de raciocínio e de decisão. A figura seguinte ilustra a actuação deste mecanismo.

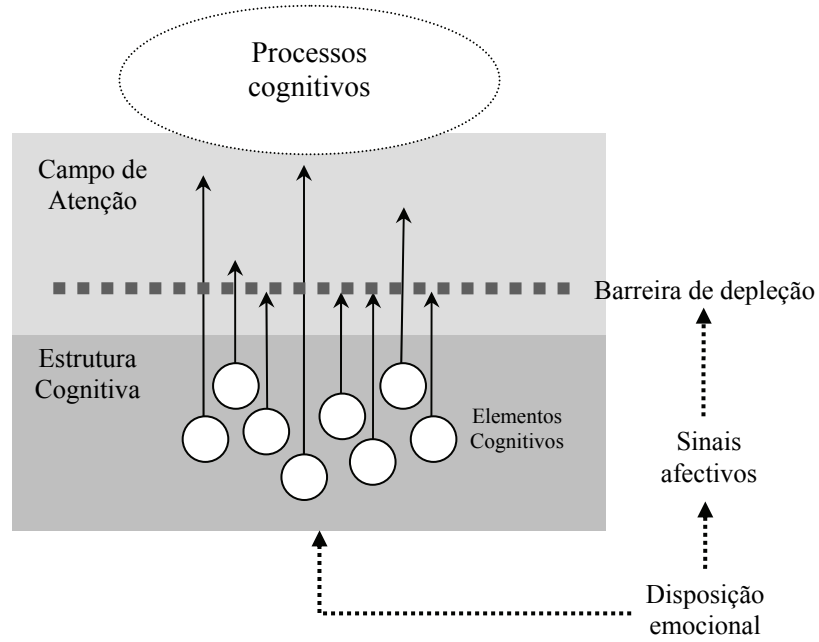


Figura 6.3. Mecanismo de focagem de atenção.

A barreira de depleção é caracterizada por uma intensidade de depleção e por uma permeabilidade. A intensidade de depleção ε , é regulada pelos sinais afectivos λ^+ e λ^- , de modo a expressar o efeito cumulativo desses sinais, ou seja:

$$\frac{d\varepsilon}{dt} = \alpha^+ \cdot \lambda^+ + \alpha^- \cdot \lambda^- \quad (6.1)$$

onde os coeficientes de sensibilidade α^+ e α^- determinam a influência dos sinais afectivos λ^+ e λ^- , respectivamente. Deste modo, a intensidade da barreira de depleção reflecte o carácter afectivo prevalecente e não os valores instantâneos, resultantes de experiências isoladas.

A permeabilidade μ determina a intensidade ε^σ da interacção entre um elemento cognitivo σ e a barreira de depleção, a qual é definida da seguinte forma:

$$\varepsilon^\sigma = \mu_s \cdot p_s^\sigma + \mu_v \cdot p_v^\sigma \quad (6.2)$$

onde μ_s e μ_v são coeficientes de permeabilidade que determinam a influência dos potenciais cognitivos de disposição emocional p_s^σ e p_v^σ do elemento cognitivo σ respectivo. Dada uma determinada intensidade de depleção ε , um elemento cognitivo σ ultrapassa a barreira de depleção e é incluído no campo de atenção se $\varepsilon^\sigma > \varepsilon$.

6.2.3 Mecanismo de Focagem Temporal

O mecanismo de focagem temporal regula o ritmo de actividade cognitiva, fornecendo uma cadência temporal para a actividade cognitiva global. Essa base temporal corresponde a um sinal p_ϕ com uma frequência ω_ϕ a qual pode ser usada para determinar o período para processamento cognitivo, isto é, o tempo disponível antes que tenha de ser produzido um comportamento.

A regulação da frequência ω_ϕ é determinada pelos sinais afectivos λ^+ e λ^- . Tal como no mecanismo de atenção, coeficientes de sensibilidade β^+ e β^- determinam a influência dos sinais afectivos λ^+ e λ^- . De igual modo, a frequência ω_ϕ expressa o efeito cumulativo dos sinais λ^+ e λ^- , de modo a reflectir o carácter afectivo prevalecente em relação às condições de concretização observadas. Ou seja:

$$\frac{d\omega_\phi}{dt} = \beta^+ \cdot \lambda^+ + \beta^- \cdot \lambda^- \quad (6.3)$$

Através da regulação do ritmo de actividade cognitiva, o mecanismo de focagem temporal tem uma influência directa no balanço entre recursos computacionais utilizados e qualidade da solução. No entanto, a influência do mecanismo de focagem temporal não se restringe aos processos de raciocínio e de decisão. Essa influência pode ser muito mais ampla, envolvendo, igualmente, os processos de percepção, acção e assimilação.

A duração variável dos períodos de actividade cognitiva, de acordo com o sinal de referência temporal p_ϕ , permite regular o tipo e âmbito do processamento realizado. A forma concreta como essa regulação ocorre depende da implementação específica de cada processo. Por exemplo, a percepção pode ser localizada ou abrangente, a acção mais lenta e precisa ou mais rápida mas menos precisa, ou a formação de memórias ser esparsa ou mais detalhada.

No trabalho desenvolvido explorou-se especificamente a regulação dos processos de raciocínio, tal como será discutido na secção seguinte. No entanto, o estudo da influência dos mecanismos de focagem sobre os restantes tipos de processos cognitivos é uma das direcções importantes para trabalho futuro.

6.2.4 Caso Experimental 1: Raciocínio Adaptativo em Ambientes Dinâmicos

É a operação combinada dos vários mecanismos anteriormente descritos que permite a um agente regular a sua actividade cognitiva de modo dinâmico, adaptando-se às condições de concretização observadas. Para ilustrar estes aspectos, vamos considerar um cenário onde o raciocínio sob condições de limitação temporal e de elevado dinamismo é necessário.

6.2.4.1 Contexto Experimental

O contexto experimental é uma implementação do domínio *Tileworld*, um ambiente virtual frequentemente utilizado como cenário de referência para o estudo do desempenho de agentes em condições de elevado dinamismo e limitação de recursos. A implementação que foi realizada segue a especificação deste cenário apresentada por Kinny e Georgeff [1991], a qual é uma referência neste domínio. Esta especificação é também a adoptada por Schut, Wooldridge e Parsons [2004], cujo trabalho irá servir como referência para avaliação de resultados.

O ambiente *Tileworld* é caracterizado por uma grelha bidimensional, na qual o agente se movimenta, tal como ilustrado na figura 6.4. O agente (representado por um círculo branco) ganha pontos por mover “peças” (representadas por quadrados cinzentos) para posições alvo, os “buracos” (representados por quadrados brancos), nos quais deve encaixar as peças. O ambiente é ainda composto por “obstáculos” (representados por segmentos pretos), dos quais o agente se deve desviar.

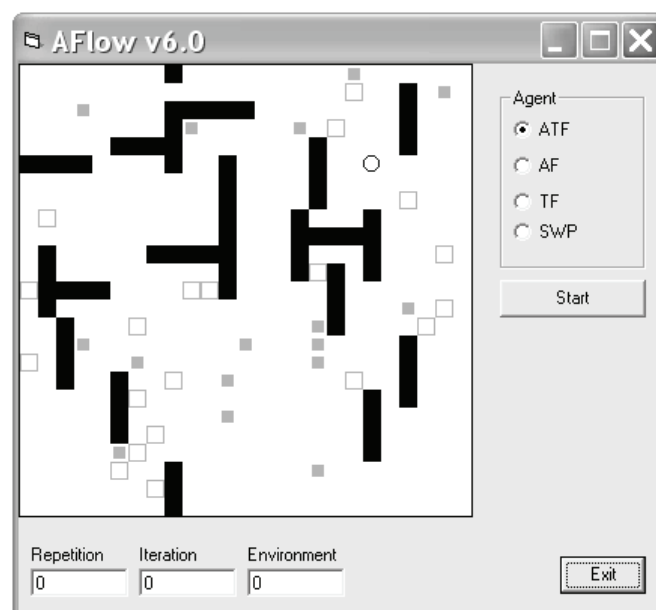


Figura 6.4. Ambiente *Tileworld* (protótipo implementado).

Na implementação proposta por Kinny e Georgeff [1991], apenas são considerados os “buracos”. Assim, quando um agente atinge um “buraco”, esse “buraco” é preenchido e desaparece. Os “buracos” aparecem e desaparecem ao longo do tempo e de modo aleatório, em posições que estejam livres. Os seus tempos de gestação e de existência são determinados por distribuições aleatórias independentes, cujas características são definidas pelos parâmetros da simulação. A tarefa do agente é visitar o maior número de buracos possível durante o tempo da simulação.

Dois parâmetros controlam as experiências:

- o *dinamismo do ambiente* (γ), tal como percebido pelo agente, denotado por um número inteiro na gama de 1 a 100 que representa o ritmo de mudança do ambiente (ver apêndice A);
- o *custo de planeamento*, que representa o custo associado à formação de um plano (o qual poderá ser medido, por exemplo, em termos do tempo necessário para a sua geração).

Kinny e Georgeff estudaram e compararam o desempenho de agentes com diferentes políticas de compromisso para custos fixos de planeamento 0, 1, 2 e 4, sem considerarem o custo efectivo de geração de um plano.

No caso do modelo proposto, de modo a ser possível avaliar o efeito dos mecanismos de regulação cognitiva, em particular da focagem temporal, é importante ter em conta o custo efectivo de planeamento. Por isso, na implementação do modelo proposto o custo de planeamento não é fixo, mas antes proporcional à dimensão da árvore de procura explorada durante a formação de um plano. O processo de planeamento recorre também a planeamento parcial, uma vez que pode ser interrompido devido à focagem temporal. O processo de planeamento é suportado num planeador do tipo A^* [Hart *et al.*, 1968; Korf, 1990].

Na implementação realizada, cada buraco detectado pelo agente é modelado como um motivador, sendo a posição corrente do agente modelada como uma observação. Tal como na implementação de Kinny e Georgeff, o agente apenas planeia para visitar um único “buraco”, e não múltiplos “buracos”.

O campo de atenção, produzido pelos mecanismos de focagem de atenção, restringe o conjunto de motivadores (“buracos”), sobre os quais o raciocínio e deliberação ocorrem. O campo de atenção também condiciona, indirectamente, a reconsideração de

intenções afectando, nomeadamente, a escolha do motivador que o agente vai tentar concretizar (o “buraco” a atingir), ou seja, a selecção da intenção corrente.

O campo de atenção controla, igualmente, a comutação entre actividades de planeamento e acção: enquanto o motivador correspondente à intenção corrente se mantiver no campo de atenção, não irá acontecer a reconsideração de intenções e será activada a acção de acordo com o plano previamente determinado; caso contrário, o motivador mais próximo da posição do agente será seleccionado como a próxima intenção corrente e o planeamento será activado.

A comutação entre planeamento e acção também é condicionada pelos mecanismos de focagem temporal: caso um período de actividade cognitiva termine durante a execução do processo de planeamento, o planeamento é interrompido e o melhor plano parcial formado até à altura é considerado, passando-se à execução da primeira acção desse plano.

Neste exemplo, a arquitectura de agente corresponde a uma implementação completa da arquitectura genérica proposta, tal como ilustrado na figura seguinte.

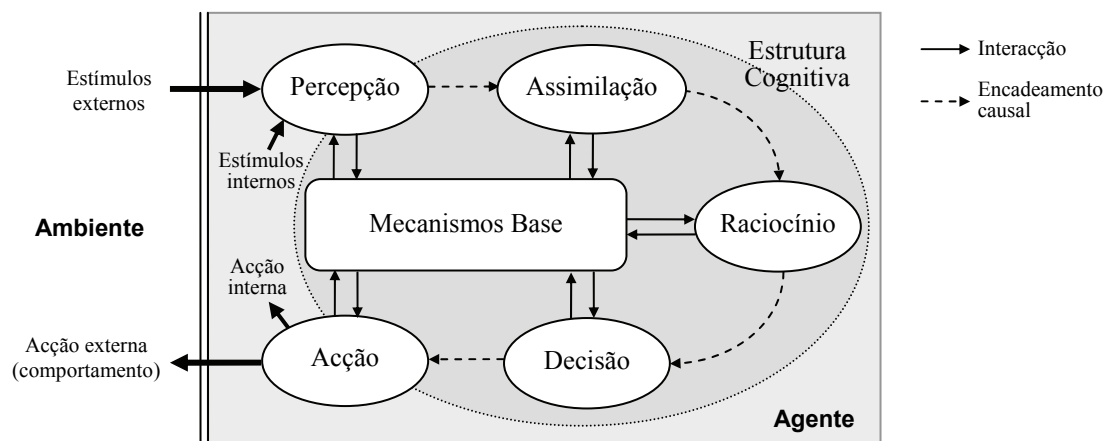


Figura 6.5. Arquitectura do agente implementado.

Nesta implementação, o processo de assimilação mantém informação acerca dos elementos que caracterizam o ambiente e acerca da posição corrente do agente, não havendo a formação de memórias de longo prazo.

No entanto, ao contrário dos agentes mais simples apresentados no capítulo anterior, cuja estrutura cognitiva pode ser vista como estática, no sentido em que a sua organização, em termos de elementos cognitivos, é predefinida, neste agente a estrutura cognitiva muda, devido à variação do número de elementos cognitivos formados e manipulados pelos processos cognitivos. A gestão dessa mudança ocorre no

âmbito do processo de assimilação. Mais à frente serão discutidos processos de assimilação onde ocorre igualmente a formação de memórias de longo prazo.

Neste caso prático, a tomada de decisão relativa a manter a intenção corrente, ou a alterá-la, segue um critério fixo, não utilizando directamente a disposição emocional associada às opções consideradas (embora dela dependente indirectamente, em resultado do campo de atenção, tal como atrás referido). A influência dos mecanismos emocionais e de regulação ocorre apenas ao nível dos processos de raciocínio. Mais à frente, iremos ver como é possível tirar partido quer da disposição emocional associada aos elementos cognitivos, quer de memórias de experiências passadas, para a tomada de decisão.

6.2.4.2 Resultados e Análise

De seguida são apresentados resultados referentes à operação individual e combinada dos mecanismos de focagem de atenção e de focagem temporal. A permeabilidade e os coeficientes de sensibilidade desses mecanismos foram considerados parâmetros fixos da arquitectura implementada, com os seguintes valores:

- $\mu_s = 1, \mu_v = 1$ (influência igual dos potenciais de disposição emocional p_s e p_v);
- $\alpha^+ = -0.8, \alpha^- = 0.3$ (maior sensibilidade a valências afectivas positivas);
- $\beta^+ = -0.8, \beta^- = 0.2$ (maior sensibilidade a valências afectivas positivas);

Os coeficientes de permeabilidade μ_s e μ_v permitem controlar a disposição emocional predominante nos elementos cognitivos do campo de atenção, neste caso uma tendência de *alegria/satisfação* representada pelo vector de disposição emocional de referência $A^+ = (1,1)$ (ver secção 5.4.2). Os coeficientes de sensibilidade $\alpha^+, \alpha^-, \beta^+$ e β^- permitem controlar o balanço entre recursos computacionais utilizados e qualidade da solução gerada. Em termos práticos, estes parâmetros condicionam o tipo de comportamento global que o agente irá apresentar.

Na implementação realizada, as disposições emocionais com valência afectiva positiva reduzem a intensidade da barreira de depleção do campo de atenção ($\alpha^+ < 0$), permitindo a mais elementos cognitivos entrar no campo de atenção, e reduzem o ritmo da actividade cognitiva ($\beta^+ < 0$), permitindo períodos de deliberação mais longos. As disposições emocionais com valência afectiva negativa têm o efeito recíproco, isto é, a intensidade da barreira de depleção do campo de atenção aumenta ($\alpha^- > 0$) e o ritmo

de actividade cognitiva aumenta ($\beta > 0$). Deste modo, uma maior sensibilidade a valências afectivas negativas traduz-se num comportamento de carácter mais reactivo, com menor pendor de planeamento e maior pendor de acção, ao passo que uma maior sensibilidade a valências afectivas positivas traduz-se num comportamento de carácter mais deliberativo, com maior pendor de planeamento. No entanto, estes comportamentos variam de acordo com o dinamismo do ambiente, tal como será discutido mais à frente.

Os valores específicos atrás indicados, adoptados para os parâmetros dos mecanismos de focagem, são os que permitiram obter o melhor comportamento do agente no contexto experimental em estudo. A configuração destes parâmetros pode ser relacionada com aspectos como *personalidade* ou *temperamento* dos agentes (ver secção 4.4), os quais não foram abordados neste trabalho, constituindo uma possível área para trabalho futuro.

Em cada experiência o dinamismo γ do ambiente foi variado. Os resultados apresentados para um determinado valor de dinamismo são valores médios referentes a 100 execuções de 20000 unidades de tempo por execução. As unidades de tempo são determinadas por um relógio interno do protótipo (ver apêndice A).

Duas variáveis distintas foram medidas: (i) a *eficácia* do agente, definida como a relação percentual entre o resultado obtido pelo agente e o resultado máximo que, em princípio, poderia ter obtido; (ii) o *custo total de planeamento*, definido como a soma dos custos de planeamento para todos os planos gerados durante uma execução.

De modo a proporcionar uma base de referência para comparação com os resultados produzidos pela abordagem proposta, também foi implementado um agente puramente deliberativo, tendo por base a melhor política de reconsideração de intenções apresentada por Schut, Wooldridge e Parsons (agente adiante designado por SWP). Essa política de reconsideração leva o agente a deliberar quando aparece um buraco mais próximo do agente que a intenção corrente (buraco a atingir), desde que não se encontre no caminho desse buraco a atingir, e quando o buraco correspondente à intenção corrente desaparece. Esta política de reconsideração é um aperfeiçoamento da melhor política de reconsideração de intenções apresentada por Kinny e Georgeff [1991]. As figuras 6.6 e 6.7 mostram os resultados obtidos para as diferentes experiências realizadas envolvendo quatro agentes: o agente SWP e três agentes correspondentes a diferentes configurações da arquitectura de agente proposta nesta tese.

Foram testadas as seguintes configurações:

- agente FTA (*Focagem Temporal e de Atenção*), tem ambos os mecanismos de focagem activos;
- agente FA (*Focagem de Atenção*), só tem o mecanismo de focagem de atenção activo;
- Agente FT (*Focagem Temporal*), só tem o mecanismo de focagem temporal activo.

A figura 6.6 mostra os resultados obtidos relativos à eficácia do agente.

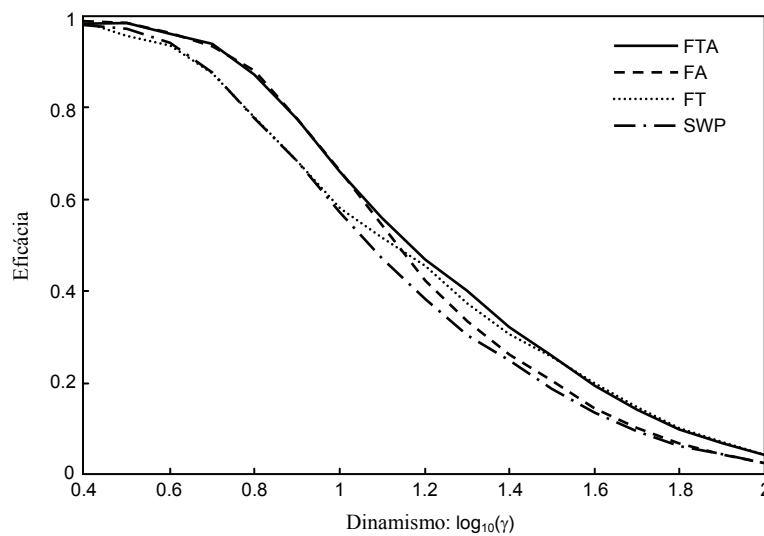


Figura 6.6. Eficácia dos agentes para diferentes graus de dinamismo do ambiente.

Tal como se pode observar, os resultados de eficácia do agente FA são melhores que os resultados do agente SWP, em particular para uma gama de valores de dinamismo baixa e média. No caso do agente FT, os resultados de eficácia são também melhores que os do agente SWP, excepto para valores baixos de dinamismo ($\log_{10}(\gamma) < 0.8$). No caso do agente FTA (mecanismo de focagem de atenção e mecanismo de focagem temporal activos), pode observar-se uma melhoria dos resultados em relação ao agente SWP, que é consistente em toda a gama de valores de dinamismo.

Estes resultados podem ser explicados pela operação combinada dos mecanismos de disposição emocional e dos mecanismos de focagem. Os mecanismos de disposição emocional proporcionam a base de suporte para a decisão de quando mudar a intenção corrente, através da modulação da relevância dos elementos cognitivos. Ou seja, através da geração de disposições emocionais que são incorporadas nos elementos

cognitivos e que posteriormente vão determinar, ou não, a presença desses elementos cognitivos no campo de atenção. Por exemplo, quando um novo buraco aparece, uma forte disposição emocional favorável é produzida, permitindo ao motivador respectivo entrar no campo de atenção. Isto acontece porque, não existindo o buraco previamente no ambiente, considera-se estar a uma distância infinita antes do seu aparecimento. O aparecimento do novo buraco produz, assim, um efeito de surpresa, que chama a atenção do agente.

Em simultâneo, são gerados sinais afectivos que regulam a focagem de atenção e a focagem temporal. Por exemplo, quando o dinamismo é baixo ou médio, o agente tem tempo para preencher todos os buracos que aparecem e, por isso, um carácter positivo prevalece, o que resulta numa baixa intensidade de depleção. Assim, a maioria dos motivadores está presente no campo de atenção, incluindo a intenção corrente. Deste modo, enquanto o agente SWP reconsidera sempre que aparece um buraco que seja mais próximo que o buraco correspondente à intenção corrente, os agentes FA e FTA raramente reconsideram. Isto reduz o dispêndio de tempo em planeamento (custo de planeamento), tal como ilustrado na figura 6.7, aumentando o tempo disponível para a acção. Como resultado o agente consegue uma melhoria da sua eficácia.

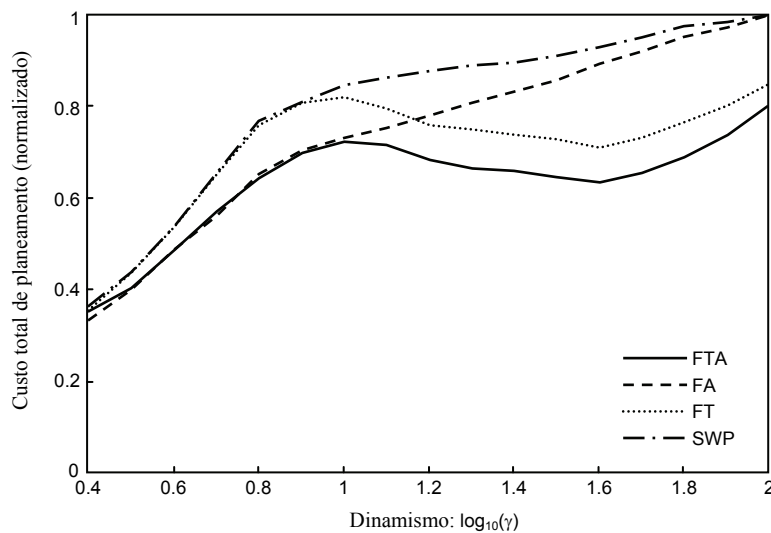


Figura 6.7. Custo total de planeamento para diferentes graus de dinamismo do ambiente.

Por outro lado, quando o dinamismo é elevado, os agentes não têm tempo para preencher a maior parte dos buracos, prevalecendo, por isso, um carácter afectivo negativo, o qual resulta numa elevada intensidade de depleção e num período de processamento cognitivo curto. Devido à elevada intensidade de depleção, apenas os

motivadores correspondentes a novos buracos são capazes de entrar no campo de atenção, o que leva os agentes FA e FTA a reconsiderar mais frequentemente.

No caso de estar activo apenas o mecanismo de focagem de atenção (agente FA), o comportamento global converge para o comportamento do agente SWP, tal como ilustrado na figura 6.6. No entanto, quando o mecanismo de focagem temporal está activo, a diminuição do período de processamento cognitivo leva a uma restrição do tempo de planeamento, mesmo quando o agente selecciona uma nova intenção, aumentando assim a rapidez de resposta do agente à mudança rápida que ocorre no ambiente, produzindo um melhor desempenho dos agentes FT e FTA. Nestas condições os efeitos da focagem temporal tornam-se predominantes, tal como a curva de eficácia do agente FT na figura 6.6 mostra. Na figura 6.7 é possível observar que ambos os agentes FT e FTA são capazes de regular o tempo dispendido em planeamento de acordo com as condições de concretização, o que é um comportamento claramente adaptativo.

Estes resultados mostram que é possível utilizar mecanismos de base emocional para controlar os processos de raciocínio, produzindo efeitos semelhantes (ou até melhores do que) aqueles obtidos por meio de técnicas de meta-raciocínio, pelo menos no domínio *Tileworld*.

6.3 Mecanismos de Adaptação de Longo Prazo

Nos mecanismos de regulação cognitiva atrás apresentados, a influência das disposições emocionais que resultam da interacção entre agente e ambiente foi explorada numa perspectiva de curto prazo, isto é, apenas o efeito das observações correntes foi considerado. Vamos agora explorar a sua influência numa perspectiva de longo prazo.

Ao considerar uma perspectiva de longo prazo, o aspecto central é a capacidade de registar a experiência do agente através de alterações persistentes da estrutura cognitiva, originando memórias das experiências do agente ao longo do tempo. São essas memórias que vão permitir ao agente antecipar situações futuras e explorar prospectivamente, por meio de processos de raciocínio, cursos de acção alternativos.

Ao mesmo tempo, se essas memórias expressarem o carácter regulador prevaLENcente na altura em que são formadas, permitem ampliar a capacidade de regulação cognitiva dos mecanismos de focagem, atrás apresentados, através da

realimentação desses mecanismos. Deste modo, os fenómenos de regulação cognitiva podem ganhar um âmbito temporal alargado. Para isso é necessário que as alterações da estrutura cognitiva possam reflectir o contexto *afectivo-emocional* em que ocorreram. São estes aspectos que serão abordados de seguida.

6.3.1 Memórias Emocionais Autobiográficas

À medida que os elementos cognitivos evoluem ao longo do tempo, descrevem trajectórias no espaço cognitivo que reflectem as experiências ocorridas. Estas trajectórias podem, ou não, ser assimiladas na estrutura cognitiva. Por exemplo, os agentes de carácter reactivo, apresentados no capítulo anterior, não realizam qualquer tipo de assimilação com efeitos permanentes.

Em contrapartida, em agentes com processos cognitivos capazes de alterar a respectiva estrutura cognitiva, as trajectórias que os elementos cognitivos descrevem ao longo do tempo podem ser assimiladas, formando memórias autobiográficas. O carácter autobiográfico destas memórias resulta de reflectirem um encadeamento temporal e não apenas um aglomerado de memórias sem qualquer relação entre si.

Subjacentes à formação destas memórias estão agregados de elementos cognitivos, que resultam dos mecanismos de integração (ver secção 6.2.1). Esses elementos cognitivos possuem uma disposição emocional integrada, dando origem à formação de *memórias de disposição emocional*, as quais podem ser relacionadas com o que outros autores designam como *memórias emocionais* (e.g. [Bower, 1994; Damásio, 2000]). A figura seguinte ilustra a relação entre os mecanismos que participam na formação de memórias emocionais.

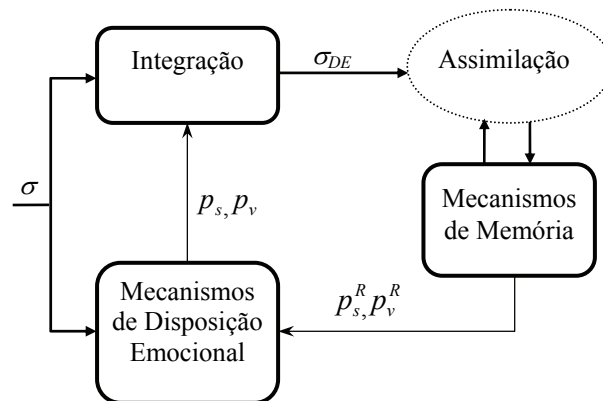


Figura 6.8. Relação entre os mecanismos que participam na formação de memórias emocionais.

Os potenciais de disposição emocional, p_s e p_v , produzidos pelos mecanismos de disposição emocional, são integrados com cada elemento cognitivo, produzindo os elementos cognitivos σ_{DE} , os quais são assimilados na estrutura cognitiva com o suporte dos mecanismos de memória.

Ao mesmo tempo, os conteúdos de disposição emocional associados aos elementos de memória recordados (os potenciais p_s^R e p_v^R apresentados na figura 6.8) vão realimentar os mecanismos de disposição emocional. Uma vez que estes mecanismos de disposição emocional são a base da regulação da actividade cognitiva, o efeito é a adaptação da actividade cognitiva, com carácter antecipador, predispondo o agente para lidar com uma situação futura similar à recordada. Este tipo de fenómenos pode ser relacionado com o que no âmbito do estudo da emoção é conhecido como *preparação para a acção* (e.g. [Frijda, 1986]). Estes mecanismos desempenham um papel importante em ambientes dinâmicos onde o tempo disponível para raciocínio possa ser variável e limitado. É o caso do cenário experimental que será apresentado mais à frente.

6.3.2 Mecanismos de Memória

Por uma questão de clareza, os sinais cognitivos têm sido modelados como variando apenas ao longo do tempo. No entanto, no caso geral esses sinais variam igualmente no espaço. Esse aspecto não foi explorado até agora, sendo a diferenciação dos elementos cognitivos feita com base na sua localização na estrutura cognitiva, representada sob a forma de índices numéricos ou textuais, como é o caso de σ ou σ_{MOT} . No entanto, a natureza oscilatória dos potenciais que formam a estrutura cognitiva de um agente propicia, igualmente, um outro tipo de modelação, no qual os elementos cognitivos assimilados (elementos de memória) não se encontram confinados a localizações espaciais específicas, mas antes se encontram distribuídos por uma determinada região da estrutura cognitiva, formando um *campo de memória*. Este tipo de organização tem um carácter paralelo e associativo, apresentando ainda características de robustez e tolerância a erros muito interessantes, mas cuja discussão está fora do âmbito deste trabalho (e.g. [Kohonen, 1980; Pribram, 1999]).

Considerando um campo de memória caracterizado por uma dimensão espacial abstracta x , as memórias são acumuladas por superposição espacial e diferenciadas por frequências espaciais específicas v . Ou seja, tal como a frequência temporal ω

permite uma discriminação qualitativa, a frequência espacial ν permite a discriminação de memórias individuais num campo de memória.

Para suportar este tipo de representação os elementos de memória necessitam de ser representados como sinais que variam não só no tempo, mas também no espaço. Para esse efeito os mecanismos de memória produzem um sinal de referência temporal $\phi_\nu(x)$ cuja frequência espacial varia contínua e monotonamente ao longo do tempo. Este sinal é utilizado para modular cada elemento cognitivo $\sigma(t)$, produzindo um elemento cognitivo $\sigma(t, x)$, o qual pode ser incorporado na memória, ou ser apenas utilizado para interagir com essa memória. Ou seja:

$$\sigma(t, x) \equiv \sigma(t) \phi_\nu(x) \quad (6.4)$$

No espaço cognitivo, esta nova representação de um elemento cognitivo está relacionada com a anterior (definição 5.3) da seguinte forma (para uma justificação mais detalhada ver apêndice B):

$$\sigma \equiv (\rho_1, \rho_2, \dots, \rho_K) \rho_\nu \quad (6.5)$$

onde o coeficiente $\rho_\nu \in \mathbb{C}$ expressa a intensidade e frequência ν do sinal $\phi_\nu(x)$ num determinado instante $t = \tau$. Esta expressão torna claro que o tempo não é representado apenas como mais uma dimensão do espaço cognitivo. Em vez disso, a passagem do tempo tem um efeito modulador na representação de cada dimensão qualitativa.

Um elemento cognitivo σ , que interactua com um campo de memória, activa múltiplos elementos de memória (recordações). Dado um elemento de memória σ_M , previamente assimilado, a sua activação produz um elemento de memória recordado σ_R , formado do seguinte modo:

$$\sigma_R = \eta(\sigma, \sigma_M) \sigma_M \quad (6.6)$$

onde η representa o ganho de interacção entre elementos cognitivos, anteriormente apresentado (definição 5.6). Os elementos de memória recordados são imagens moduladas dos elementos de memória originais, cuja intensidade depende da correlação (expressa pelo ganho de interacção) entre o elemento cognitivo estimulador e os elementos de memória. Deste modo, um campo de memória actua como uma memória associativa na qual os elementos cognitivos são activados por contacto qualitativo e temporal (e.g. [Kohonen, 1980]).

6.3.3 Integração de Mecanismos de Memória e de Atenção

Dada a natureza inerentemente associativa e paralela de um campo de memória, a interação de um elemento cognitivo com um campo de memória produz, potencialmente, um grande número de elementos de memória recordados com diferentes intensidades de activação. Perante esse grande número de recordações, um agente tem de decidir em quais se focar, ou a tomada de decisão em tempo útil não será possível.

O mecanismo de focagem de atenção, anteriormente descrito, permite lidar com este problema, restringindo a atenção dos processos cognitivos a elementos cognitivos específicos, neste caso elementos de memória, de acordo com o seu conteúdo de disposição emocional, ou seja, de acordo com os potenciais cognitivos p_s e p_v de cada elemento cognitivo. A figura seguinte mostra como os vários mecanismos envolvidos se integram.

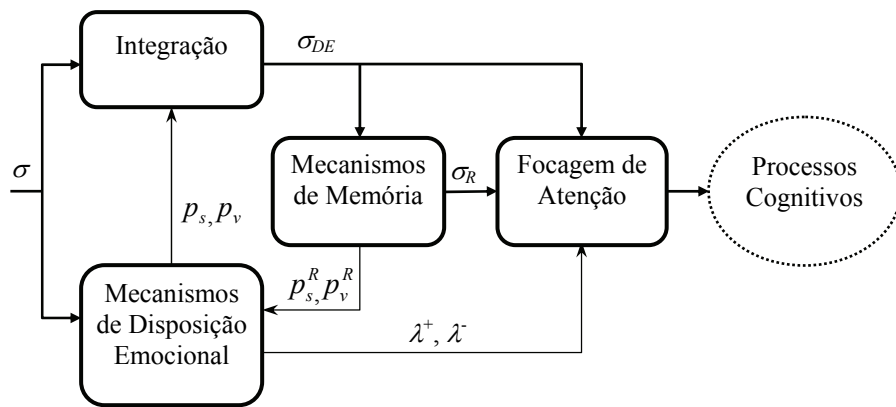


Figura 6.9. Integração de mecanismos de memória com mecanismos de disposição emocional, integração e de focagem de atenção.

Como se pode observar, os elementos de memória recordados σ_R são sujeitos à focagem de atenção antes que possam, eventualmente, participar no processamento cognitivo. Essa participação depende da disposição emocional que os caracteriza e da intensidade da barreira de depleção do campo de atenção, a qual é determinada pelos mecanismos de disposição emocional através dos sinais λ^+ e λ^- .

6.4 Influência de Memórias Emocionais no Raciocínio e Tomada de Decisão

A influência de memórias emocionais na actividade cognitiva de um agente pode ocorrer a múltiplos níveis. Por exemplo, num nível “reactivo” as disposições emocionais que resultam das memórias emocionais recordadas podem produzir directamente a activação de padrões comportamentais específicos, como é o caso de padrões do tipo *lutar ou fugir* (“*fight or flight*”) [Barnard, 2004]. No entanto, particularmente interessante é a influência de memórias emocionais no raciocínio e tomada de decisão, onde podem ter uma contribuição significativa para tornar esses processos viáveis perante restrições dinâmicas de tempo e de recursos.

Após a definição dos mecanismos base subjacentes à formação de memórias emocionais, é agora possível abordar esta relação entre memórias emocionais e processos de raciocínio e tomada de decisão. Para isso é necessário concretizar esses processos, tal como será feito de seguida.

6.4.1 Processos de Decisão

No modelo proposto, a tomada de decisão resulta da avaliação das diferentes opções disponíveis para a acção, quer produzidas pelos processos de raciocínio, quer resultantes de outros processos cognitivos, nomeadamente da recordação de memórias específicas.

As opções disponíveis são elementos cognitivos que desempenham o papel de mediadores, pois determinam a acção concreta que o agente irá realizar. Como vimos anteriormente, no espaço cognitivo os mediadores definem direcções de movimento, tal como ilustrado na figura seguinte.

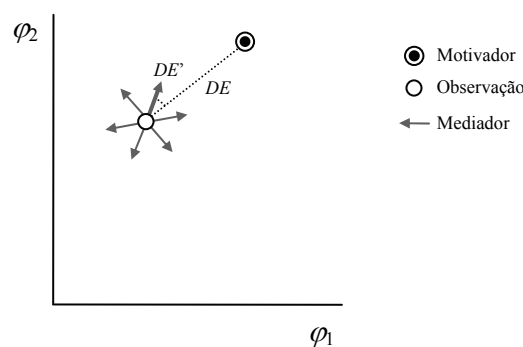


Figura 6.10. Influência da situação observada pelo agente na tomada de decisão.

Ou seja, perante os vários mediadores possíveis, o agente necessita de decidir por qual optar. A opção é feita de acordo com a disposição emocional associada a cada mediador, sendo seleccionado o mediador com a disposição emocional mais favorável (ver secção 5.4).

O primeiro passo para a tomada de decisão é, pois, a formação de uma disposição emocional associada a cada mediador. Essa disposição emocional é determinada por três tipos de elementos influentes:

- *situação observada*, a qual expressa a relação entre os motivadores do agente e as observações correntes;
- *recordação directa*, a qual resulta da activação de memórias emocionais individuais;
- *recordação de trajectória*, a qual resulta de raciocínio prospectivo, envolvendo a activação encadeada de memórias emocionais.

A disposição emocional resultante de cada uma destas influências é ponderada de acordo com a correlação entre a direcção determinada pelo mediador, e a direcção entre a observação que serve de origem ao mediador e o elemento (motivador ou recordação) com o qual essa observação se relaciona, produzindo uma disposição emocional associada ao mediador. Por exemplo, na figura 6.10, a disposição emocional DE resulta da relação entre a observação e o motivador, cuja direcção está indicada a ponteadado. Por sua vez, cada mediador define uma direcção específica de movimento. É a correlação entre essas duas direcções que é o factor de ponderação aplicado à disposição emocional original (DE) para produzir a disposição emocional DE' associada ao mediador indicado.

No caso de múltiplas influências, os coeficientes de ponderação de cada influência correspondem à correlação entre a direcção, relativa à observação corrente, de cada uma das influências, e a direcção do mediador, definida como:

$$\Gamma_{\sigma}(\sigma_1, \sigma_2) \equiv \cos(\sigma : \sigma_1 \wedge \sigma_2) \quad (6.7)$$

onde $\sigma : \sigma_1 \wedge \sigma_2$ representa o ângulo entre as direcções dos elementos cognitivos σ_1 e σ_2 tendo como referência um dado elemento cognitivo σ .

Considerando influencias resultantes de vários elementos cognitivos $\sigma_1, \sigma_2, \dots, \sigma_N$, a disposição emocional DE'_{med} associada a um mediador σ_{med} , aplicado no ponto de

origem dado pelo elemento cognitivo σ , é definida da seguinte forma:

$$DE'_{med} = \frac{1}{N} \sum_{n=1}^N \Gamma_{\sigma}(\sigma_{med}, \sigma_n) DE_{\sigma_n} \quad (6.8)$$

Em agentes sem capacidades de assimilação nem de raciocínio, a tomada de decisão ocorre tendo por base apenas a situação observada. Ou seja, depende apenas das disposições emocionais que resultam das interacções entre a observação corrente (resultante dos processos de percepção) e cada motivador.

No caso de existir um único motivador, o mediador seleccionado será aquele cuja direcção mais se aproximar da direcção desse motivador, tal como ilustrado na figura 6.10. A decisão com base apenas neste nível de influência é típica de agentes simples, tal como o agente apresentado no exemplo 2 do capítulo anterior.

Em agentes com processos de assimilação capazes de formar memórias emocionais, resultantes das experiências do agente ao longo do tempo, a tomada de decisão envolve também essas memórias .

A influência das memórias emocionais na tomada de decisão ocorre a dois níveis: (i) num *nível associativo*, onde as memórias emocionais são activadas pelas observações correntes e filtradas pelos mecanismos de focagem de atenção: (ii) num *nível prospectivo*, onde o espaço cognitivo é explorado com base nas trajectórias autobiográficas recordadas. Consideremos para já o nível associativo.

6.4.2 Nível Associativo

No nível associativo, a tomada de decisão depende não apenas das disposições emocionais resultantes da situação observada, mas também das disposições emocionais (DE_i) que resultam das memórias emocionais recordadas (m_i), tal como ilustrado na figura seguinte.

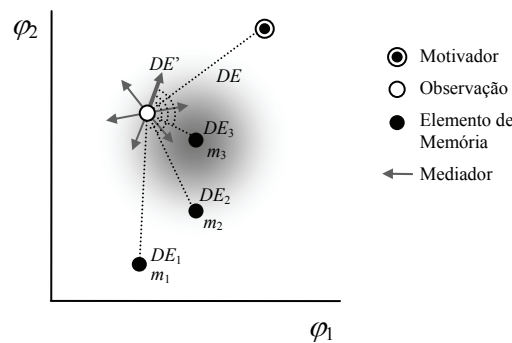


Figura 6.11. Influência de recordação directa de memórias emocionais na tomada de decisão.

Estas memórias são activadas pelas observações correntes e filtradas através do mecanismo de atenção (ver figura 6.9). Como exemplo, a figura 6.11 ilustra a influência de memórias que resultam da experiência passada do agente, eventualmente sobrepondo-se à influência da situação corrente.

A influência das memórias recordadas (ilustrada a cinzento) depende da interacção entre essas memórias e a observação em causa. Ou seja, as memórias emocionais directamente recordadas têm um âmbito de influência local.

A decisão com base apenas nestes dois níveis de influência é típica de agentes capazes de dinamicamente formar padrões estereotipados de comportamento, tendo por base a resposta a estímulos específicos.

As tendências emocionais de medo desempenham aqui um papel central, pois permitem criar padrões de comportamento no sentido de afastar o agente de situações que lhe são muito desfavoráveis. No entanto, o âmbito local, que caracteriza este nível de influência na tomada de decisão, faz com que seja muito sensível às configurações locais do espaço cognitivo.

6.4.3 Nível Prospectivo

Neste terceiro nível de influência, os processos de raciocínio desempenham o papel principal, através da exploração de trajetórias possíveis no espaço cognitivo. Desta forma proporcionam um âmbito alargado à influência das memórias emocionais na tomada de decisão.

O raciocínio prospectivo tem por base as memórias directamente recordadas, a partir das quais o espaço cognitivo é explorado. Essa exploração é guiada pelas trajetórias que partem das memórias directas, tal como ilustrado na figura 6.12.

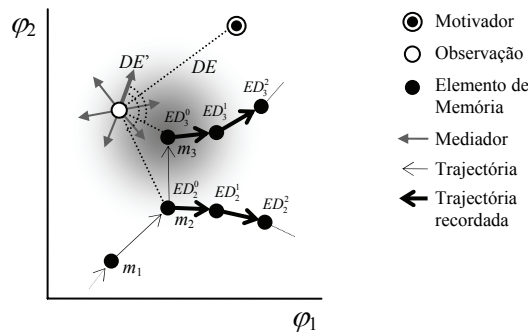


Figura 6.12. Influência de recordação de trajetórias autobiográficas na tomada de decisão.

A exploração de trajectórias corresponde a um recordar encadeado e progressivo de memórias emocionais autobiográficas. A extensão da exploração é controlada pelos mecanismos de focagem temporal descritos na secção 6.2.3.

Note-se que o mecanismo de focagem temporal é regulado pelos mecanismos emocionais através dos sinais afectivos λ^+ e λ^- , os quais, por sua vez, são realimentados pelos potenciais de disposição emocional resultantes das memórias emocionais recordadas (ver figura 6.8). Deste modo, as memórias emocionais não só proporcionam o ponto de partida para o raciocínio prospectivo, como produzem também um efeito regulador, através das disposições emocionais recordadas, o qual limita a extensão da exploração.

No raciocínio prospectivo, não são directamente utilizadas as disposições emocionais das memórias recordadas ao longo da exploração das trajectórias. Em vez disso são geradas novas disposições emocionais que expressam expectativas do que poderá ser a evolução da situação, caso as correspondentes direcções de evolução sejam seguidas. Esta geração de novas disposições emocionais é fundamental, porque o enquadramento cognitivo no qual as memórias de trajectória foram formadas pode ter mudado. Nomeadamente, podem ter mudado os motivadores.

No caso da recordação directa de memórias, esse facto não é relevante, pois a própria activação das memórias depende da semelhança do contexto cognitivo em que foram criadas com o contexto cognitivo corrente. No entanto, isso não acontece na recordação de trajectórias, uma vez que estas são activadas apenas por encadeamento temporal, independentemente da semelhança entre o contexto cognitivo em que foram formadas e o contexto cognitivo corrente.

As trajectórias autobiográficas surgem, assim, como caminhos que tornam possível a exploração de espaços cognitivos potencialmente contínuos, logo com um número de hipóteses de exploração potencialmente infinito. Tendo por base esses caminhos e as condições de concretização observadas, é possível explorar novos caminhos tirando partido das experiências do passado.

6.4.4 Caso Experimental 2: Adaptação e Decisão Guiada por Memórias Emocionais

Para testar a operação dos mecanismos de adaptação e decisão apresentados procuraram-se cenários onde os agentes tenham de operar em situações totalmente desconhecidas à priori e onde, simultaneamente, existam fortes restrições temporais em termos do tempo disponível para concretizar os objectivos que lhes são atribuídos.

Os cenários de busca e salvamento em situações de risco ou catástrofe são cenários típicos onde estas condições ocorrem. Para além disso, são cenários com requisitos muito estritos em termos de informação disponível para adaptação e tomada de decisão. Estes cenários são, por isso, particularmente interessantes para testar o modelo proposto.

6.4.4.1 Contexto Experimental

O contexto experimental que serve de base ao exemplo é a simulação de ambientes onde um agente de busca e salvamento deve mover alvos para uma posição segura. Os ambientes representam casas com diferentes quartos onde estão localizados os alvos. Os alvos emitem um sinal de socorro que o agente de salvamento é capaz de detectar. No entanto, considera-se capacidade sensorial mínima, típica de situações de catástrofe. Por exemplo, no caso de incêndios a visibilidade pode ser praticamente nula, o mesmo se passando com outras modalidades sensoriais. Deste modo, o agente de salvamento apenas é capaz de detectar o sinal de socorro e, a partir desse sinal, derivar a direcção e uma medida da distância do alvo. O agente é, igualmente, incapaz de detectar as paredes dos quartos a não ser por contacto directo (colisão).

Os ambientes são modelados como contínuos, no sentido em que todas as suas características são descritas por valores reais. Este aspecto é importante para aferir a escalabilidade do modelo. Isto é, se é um modelo demonstrativo de conceitos, mas que em termos operacionais é restrito a ambientes simulados do tipo grelha (“*gridworlds*”), ou se é capaz de suportar ambientes complexos, próximos do real, onde uma das fontes de complexidade é, precisamente, a natureza contínua desses ambientes.

O agente é caracterizado por um mecanismo auto-referencial que permite conhecer a sua posição no ambiente e por mecanismos sensoriais que permitem derivar a direcção e a distância ao alvo. Deste modo, por cada período de actividade cognitiva o agente forma uma observação, que indica a sua posição actual, e um motivador, que

indica uma estimativa da posição do alvo. Ambos os elementos cognitivos são caracterizados por duas dimensões cognitivas φ_x e φ_y , que representam as coordenadas de localização no ambiente.

Em cada operação de salvamento apenas um alvo está activo, ou seja, emite o sinal de socorro. Após encontrar o alvo, o agente selecciona como motivador a posição segura, a qual é fixa para cada experiência.

6.4.4.2 Resultados e Análise

No início de cada experiência o ambiente é totalmente desconhecido para o agente, tendo este um tempo limitado para realizar o salvamento, antes que as condições ambientais tornem o salvamento impossível. Deste modo, não é possível qualquer fase de treino. Ou seja, o agente tem de actuar directamente no ambiente e aprender em tempo de execução como se movimentar nele. A figura seguinte mostra o comportamento do agente num dos ambientes de teste utilizando dois alvos distintos (representados por círculos cinzentos). A posição inicial e final do agente são as mesmas e correspondem à posição segura para a qual deve mover os alvos (representada por uma circunferência pontilhada).

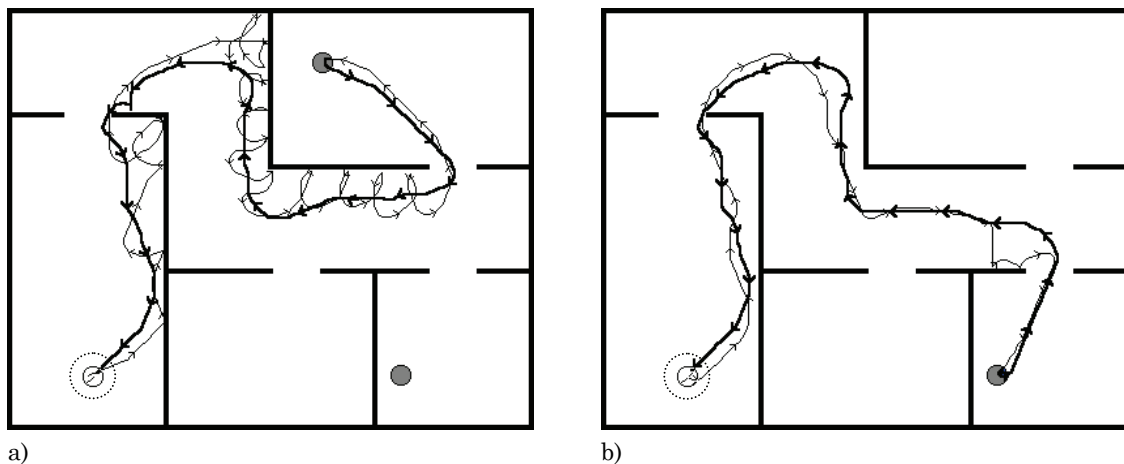


Figura 6.13. Comportamento do agente num dos ambientes de teste: (a) salvamento do primeiro alvo; (b) salvamento do segundo alvo.

Quando o agente começa a movimentar-se, tentando alcançar o primeiro alvo, ocorrem múltiplas colisões com as paredes. Sempre que uma colisão ocorre, o agente roda 180° e afasta-se da parede. Durante este processo são formadas memórias emocionais que levam a uma melhoria progressiva no desempenho do agente, à medida que mais experiência se acumula. A figura seguinte mostra uma

representação no espaço cognitivo das memórias emocionais formadas até o agente atingir o primeiro alvo.

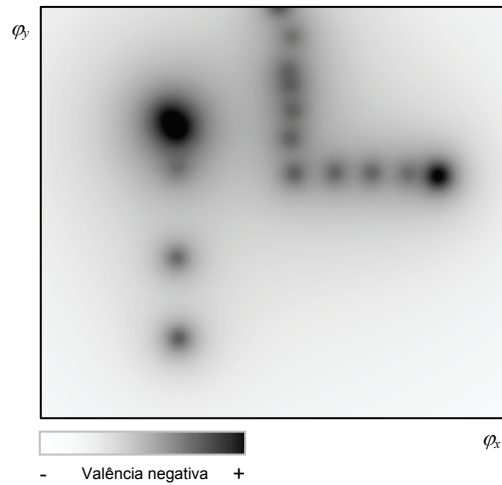


Figura 6.14. Representação no espaço cognitivo das memórias emocionais formadas pelo agente até atingir o primeiro alvo (projectão sobre as dimensões cognitivas φ_x e φ_y ; tonalidade representa a valência negativa das memórias emocionais).

Como se pode observar, nos pontos de embate com as paredes o agente forma memórias, cuja influência se estende às regiões circundantes do espaço cognitivo. A intensidade da influência conjunta das memórias aumenta à medida que se vão acumulando mais memórias numa determinada região.

Durante a progressão para o primeiro alvo, a decisão em relação à direcção do movimento a realizar só é influenciada pela situação observada (posição do agente *vs.* posição estimada do alvo) e pelas recordações directas das memórias emocionais formadas (ilustradas na figura 6.14), uma vez que ainda não há trajectórias de que o agente possa tirar partido em direcção ao alvo. Deste modo, o comportamento do agente tem um carácter mais reactivo.

Em contrapartida, no caminho de retorno (figura 6.13.a, trajectória mais escura) e na segunda operação de salvamento (figura 6.13.b), o agente já dispõe de trajectórias passíveis de ser exploradas através de raciocínio prospectivo. Isso reflecte-se numa melhoria significativa do desempenho do agente, como se pode observar no caminho de retorno e na segunda operação de salvamento.

Note-se que, quando o agente chega ao primeiro alvo e inicia o caminho de retorno, o motivador muda para a posição inicial do agente (posição segura). Logo, quando o agente realiza a exploração de trajectórias a partir de um elemento de memória recordado, não devem ser consideradas as disposições emocionais contidas nessas

memórias. Em vez disso, devem ser calculadas novas disposições emocionais em função do novo motivador. Esta situação ilustra bem o facto da influência de memórias emocionais na tomada de decisão necessitar de raciocínio prospectivo e não apenas de uma mera recordação de trajectórias. A figura seguinte ilustra alguns dos aspectos envolvidos no raciocínio e tomada de decisão do agente.

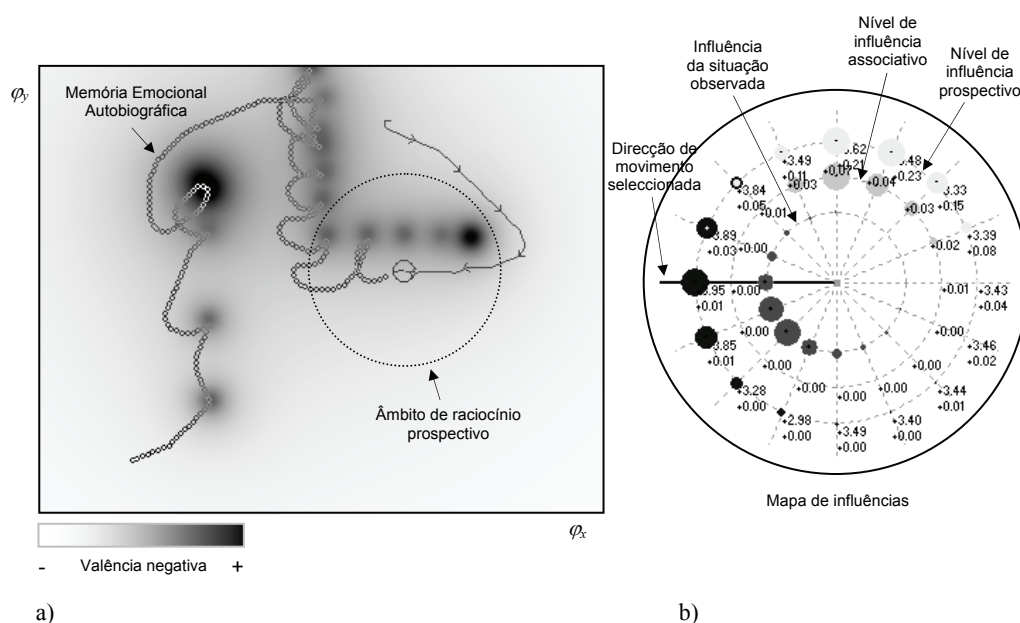


Figura 6.15. Influência do raciocínio prospectivo nos processos de tomada de decisão.

A figura 6.15.b mostra um mapa de influências que permite observar vários aspectos subjacentes ao raciocínio prospectivo e à tomada de decisão do agente. As três circunferências concêntricas representam os vários níveis de influência. O centro dessas circunferências representa a observação referente à posição do agente, representando cada direcção (segmentos a tracejado) um mediador. As influências dos vários níveis em cada mediador são representadas por círculos (a tonalidade indica o tipo de influência e a valência, a dimensão indica a intensidade).

A figura 6.15.a ilustra a exploração das trajectórias de memória autobiográfica em termos do raciocínio prospectivo do agente. O âmbito dessa exploração (indicado pela circunferência a ponteados) é regulado pelo mecanismo de focagem temporal, variando ao longo do tempo de acordo com o carácter afectivo prevaiente. Por exemplo, se o agente se aproxima de uma região com uma grande densidade de memórias emocionais de carácter afectivo negativo ($\lambda^- > \lambda^+$), o âmbito de exploração é reduzido, pelo que a tomada de decisão é mais focada na situação actual e nas recordações directas. Quando o carácter afectivo prevaiente é positivo ($\lambda^+ > \lambda^-$) o âmbito de

exploração alarga-se, pelo que a tomada de decisão é igualmente influenciada pelo raciocínio prospectivo, isto é, existe uma perspectiva de raciocínio e decisão alargada no tempo, que tem em conta possíveis evoluções futuras, em contraponto a uma perspectiva de curto prazo, necessária para lidar com possíveis contingências da situação actual.

É importante notar que o raciocínio prospectivo não consiste num mero seguir de trajectórias autobiográficas, o que levaria o agente a cometer os mesmos erros do passado. Pelo contrário, guiado pela composição dos vários níveis de influência, o raciocínio prospectivo tira partido das memórias de experiências anteriores explorando o espaço cognitivo no sentido de encontrar novos caminhos, o que resulta numa melhoria progressiva das trajectórias geradas, como se pode observar na figura 6.13.

A figura 6.15.b mostra um mapa de influências onde estão ilustrados os vários níveis de influência na tomada de decisão. É a composição dessas diferentes influências (ver definição 6.8) que determina a direcção de movimento seleccionada pelo agente. Também nesta composição de influências para tomada de decisão as disposições emocionais desempenham um papel chave. Assim, para além do seu carácter adaptativo e regulador, a noção de disposição emocional funciona também como uma *medida de avaliação*, essencial para a composição de influências heterogéneas na tomada de decisão, o que alguns autores designam por *moeda comum* (“*common currency*”) e que consideram ser um papel importante desempenhado pelos fenómenos emocionais (e.g. [Wright, 1997; Belavkin, 2001]).

Para testar o desempenho do agente em diferentes ambientes e ao longo do tempo, a tarefa de salvamento foi repetida em 10 ambientes diferentes com 10 alvos em cada ambiente. O indicador de desempenho considerado foi o tempo necessário para o salvamento de cada alvo. A figura 6.16 mostra os resultados (valores máximo, mínimo e médio para os 10 diferentes ambientes).

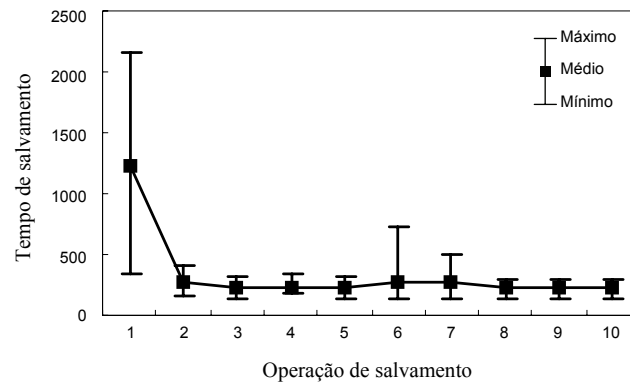


Figura 6.16. Desempenho do agente ao longo do tempo.

Tal como pode ser observado, ocorre uma melhoria significativa na tomada de decisão e na adaptação ao ambiente logo durante o primeiro salvamento. A partir do segundo salvamento o agente atinge um bom desempenho, independentemente do ambiente. A partir daí o desempenho do agente continua a melhorar mas numa escala menor. As variações no desempenho do agente da segunda operação de salvamento em diante resultam do comportamento exploratório associado ao raciocínio prospectivo, anteriormente referido.

A diferença que se pode observar entre valores máximos e mínimos na primeira operação de salvamento corresponde aos diferentes tipos de ambiente que foram escolhidos, de modo a reflectir diferentes níveis de complexidade, tal como se pode observar na figura seguinte.

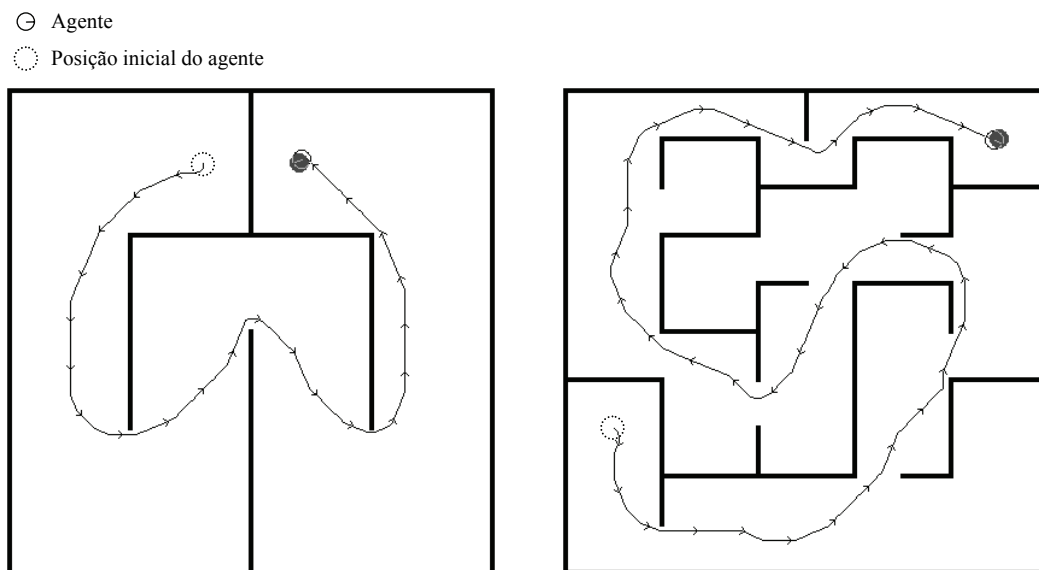


Figura 6.17. Exemplo de ambientes utilizados para testar o desempenho do agente (trajectórias referentes à última operação de salvamento).

Uma das características salientes destes resultados é a capacidade do agente muito rapidamente tirar partido da experiência anterior de forma eficaz, mesmo que não óptima. Por exemplo, após uma operação de salvamento o agente consegue percorrer a parte da casa percorrida anteriormente sem bater em qualquer obstáculo e sem fazer mudanças bruscas de direcção, como se pode observar na figura 6.13. Este compromisso entre qualidade da solução e tempo e recursos consumidos para obter essa solução é um dos resultados importantes da integração entre aspectos emocionais e de raciocínio que caracteriza o modelo proposto.

A comparação dos resultados obtidos com resultados de outras abordagens típicas deste tipo de problemas, como a aprendizagem por reforço ou métodos de planeamento em ambientes contínuos, é uma das áreas importantes para trabalho futuro.

6.5 Contexto Cognitivo Inter-Agentes: Explorações Preliminares

Uma das preocupações subjacentes à elaboração do modelo de agente proposto foi garantir a independência do tipo de processamento cognitivo envolvido. Deste modo é possível explorar a influência dos fenómenos emocionais em diferentes contextos por meio de instanciações específicas do modelo.

Nesse sentido, também a integração entre fenómenos emocionais e cognitivos, nomeadamente de raciocínio, suportada pelos mecanismos base que caracterizam o modelo, é independente da implementação concreta dos processos de raciocínio. No entanto, consoante a sofisticação dos processos de raciocínio, mais rica se pode tornar a relação entre emoção e raciocínio. De facto, o que se observa experimentalmente é uma crescente complexidade dos fenómenos emocionais à medida que a capacidade cognitiva dos organismos aumenta (e.g. [Hebb & Thompson, 1979; Scherer, 2001]).

6.5.1 Raciocínio Social

Um dos contextos onde os fenómenos emocionais ganham uma relevância particular é no contexto da interacção social [Scherer, 2001], no qual os agentes podem interagir entre si. Essa capacidade de interacção levanta problemas de comunicação e coordenação importantes, em particular no que se refere à capacidade de raciocinar sobre os outros agentes no sentido de concretizar motivações individuais ou de grupo.

Esse raciocínio, designado *raciocínio social* (e.g. [Conte & Castelfranchi, 1995]), envolve a capacidade de explicitamente manipular representações conceptuais de outros agentes e de explorar hipóteses de coordenação de diferentes grupos de agentes por diferentes meios, por exemplo, cooperação ou troca social. Daqui resulta uma das características do raciocínio social, que é a sua elevada complexidade computacional (e.g. [d'Inverno *et al.*, 1997]). Para além disso, em ambientes incertos e dinâmicos as próprias características dos agentes e das actividades de coordenação variam com o tempo. Por exemplo, pode haver quebra de compromissos levando à frustração das expectativas de um agente.

Este tipo de problemas exige da parte dos agentes a capacidade de adaptação à mudança e a capacidade de focagem do raciocínio no sentido de limitar os recursos consumidos de modo a tornar o raciocínio social viável em tempo útil. Estes são, precisamente, os aspectos centrais do modelo de agente proposto, pelo que de seguida será explorada a sua aplicação num contexto de raciocínio social.

6.5.2 Caso Experimental 3: Raciocínio Social em Ambientes Incertos e Dinâmicos

O caso experimental que será apresentado de seguida aborda o problema da coordenação do comportamento de múltiplos agentes no sentido de tirar partido das suas capacidades heterogéneas. Os modelos de raciocínio social abordam este problema dotando os agentes da capacidade de explicitamente raciocinar em relação aos outros agentes [Conte & Castelfranchi, 1995]. No entanto, em ambientes incertos e dinâmicos, onde os agentes possam ter diferentes perfis de comportamento esta capacidade de raciocínio em relação a outros agentes levanta problemas importantes. Por exemplo, o interesse de um agente como parceiro de cooperação pode mudar como resultado da formação de novas coligações, da mudança nas prioridades das tarefas e por inúmeros outros factores perturbadores. Para além disso, em cenários reais é necessária uma capacidade de resposta em tempo útil, o que é particularmente problemático no contexto do raciocínio social, dada a elevada complexidade computacional típica dos modelos de raciocínio social.

Tal como referido no capítulo de introdução, foi na tentativa de abordar estes problemas que surgiu pela primeira vez a ideia de utilização de mecanismos emocionais para lidar com os problemas da adaptabilidade e complexidade computacional em agentes inteligentes. Nesse sentido, este cenário experimental

constitui um caso prático interessante, embora muito simplificado, para ilustrar as características da abordagem proposta.

6.5.2.1 Contexto Experimental

O contexto experimental consiste numa sociedade de agentes que possuem diferentes perfis comportamentais e que possuem a capacidade de raciocinar sobre os outros agentes, de modo a coordenar as suas actividades.

Cada agente é caracterizado por um perfil de fiabilidade, definido a partir do nível de sucesso na concretização dos objectivos que se propõe atingir, representado por um valor no intervalo $[0, 1]$ (0 representa falha total, 1 representa sucesso total). Cada agente tem também características próprias em termos das suas capacidades, objectivos e planos, que se traduzem num maior ou menor interesse de cooperação por parte do agente cujo raciocínio se vai simular. Para não estarmos a entrar nos pormenores do mecanismo de raciocínio social (para mais informação ver e.g. [Sichman & Demazeau, 2001; Morgado, 1999]) são considerados três tipos de interesse de cooperação: *pouco-favorável*, *medianamente-favorável* e *muito-favorável*. Para simular um ambiente incerto e dinâmico, onde as características dos agentes mudam com o tempo, quer os perfis de fiabilidade quer os interesses de cooperação são determinados por distribuições aleatórias. Os interesses de cooperação são determinados por uma distribuição aleatória *uniforme*, e os perfis de fiabilidade por uma distribuição aleatória *normal*, tal como ilustrado na tabela seguinte.

	<i>Distribuição Normal (desvio padrão = 0.1)</i>
Perfil de Fiabilidade	Nível de Concretização médio
Baixo	0.2
Médio	0.5
Alto	0.8

Tabela 6.1. Perfis de fiabilidade.

O agente cujo raciocínio social está a ser simulado, tem a capacidade de observar os restantes agentes da sociedade. Cada observação é caracterizada por três dimensões

cognitivas: ϕ_{ag} , representa a identificação do agente; ϕ_{fiab} , representa a fiabilidade observada; ϕ_{coop} , representa o tipo de interesse de cooperação observado.

Uma vez que as características dos agentes mudam com o tempo, de acordo com as distribuições aleatórias atrás indicadas, o agente em estudo, ao decidir cooperar com um outro agente, não tem qualquer garantia que as características observadas anteriormente em relação a esse agente se vão manter. Pelo contrário, necessita de se adaptar a uma possível mudança dessas características. Nesse sentido, após cada interacção o agente observa o respectivo resultado (em termos de fiabilidade e interesse de cooperação), formando uma memória emocional dessa observação, de acordo com a evolução da situação.

Neste sentido, os perfis de fiabilidade e os interesses de cooperação expressam dois tipos distintos de condições ambientais, de acordo com as distribuições aleatórias que os determinam. Assim, no caso dos perfis de fiabilidade, apesar de as condições ambientais serem relativamente adversas devido à incerteza e mudança permanente nas características dos agentes, é possível tirar partido da experiência passada estimando os parâmetros da distribuição normal que caracteriza o perfil, no sentido de melhor lidar com situações futuras. Já no caso das características de cooperação isso não é possível, pois as condições de incerteza são extremas, isto é, não é possível extrair qualquer informação da experiência passada que possa ser utilizada no futuro, devido à mudança ser de carácter aleatório uniforme.

De seguida são apresentados resultados de experiências envolvendo diferentes aspectos do modelo proposto. Nessas experiências foi observado o comportamento de um agente específico. Esse agente é caracterizado por um motivador, σ_{fiab} , que guia o agente no sentido de seleccionar como parceiros agentes de elevada fiabilidade, e por um motivador, σ_{coop} , que guia o agente para seleccionar agentes com elevado interesse de cooperação. A intensidade ε desses motivadores mudou ao longo das experiências realizadas. A intensidade da barreira energética que determina o conteúdo do campo de atenção foi regulada por um parâmetro normalizado, designado *nível de atenção* ($\gamma \in [0,1]$). O nível de atenção determina a proporção do total de opções que estão presentes no campo de atenção (quando $\gamma = 0$ apenas uma opção está presente no campo de atenção).

Cada opção corresponde a um agente específico, com o qual o agente observado se pode coligar. Neste exemplo o ambiente é composto por 30 agentes. Os resultados

apresentados são valores médios para 10 execuções de 1000 episódios de cooperação por experiência.

Na primeira experiência o *nível de atenção* foi $\gamma = 0$ (apenas uma opção é permitida no campo de atenção). Deste modo o comportamento do agente é determinado exclusivamente pelas forças de disposição emocional, pois não existem opções alternativas para ser ponderadas pelos processos de raciocínio. Para estudar a influência específica de cada motivador, apenas um motivador esteve activo em cada experiência. A figura 6.18.a mostra os resultados para o motivador de cooperação e a figura 6.18.b mostra os resultados para o motivador de fiabilidade.

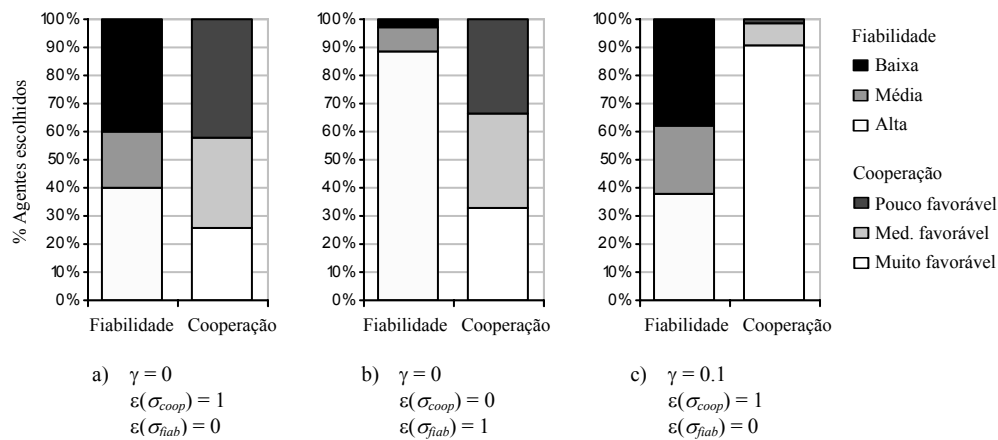


Figura 6.18. Influências ambientais e cognitivas no comportamento do agente (fiabilidade e características de cooperação dos agentes seleccionados).

Como se pode observar na figura 6.18.a, sob a influência exclusiva do motivador de cooperação, a influência emocional levou a um comportamento deficiente, no qual o agente seleccionou principalmente agentes de baixa ou média fiabilidade e de cooperação não favorável ou medianamente favorável. Este resultado deve-se à mudança aleatória uniforme das características de cooperação dos agentes, não proporcionando qualquer informação a partir de experiências passadas que pudesse ser útil para guiar o comportamento do agente.

Em contrapartida, na figura 6.18.b, que mostra os resultados sob influência exclusiva do motivador de fiabilidade, podemos observar que o agente foi muito eficiente na sua adaptação ao ambiente. Este resultado resulta da possibilidade de explorar regularidades no comportamento dos agentes ao longo do tempo, proporcionadas por uma variação aleatória mas *normal* das características dos agentes.

Numa segunda experiência, o *nível de atenção* foi colocado em $\gamma = 0.1$ para permitir mais do que uma opção no campo de atenção (10% do total de opções são permitidas no campo de atenção). A figura 6.18.c mostra os resultados obtidos sob a influência exclusiva do motivador de cooperação, tal como na figura 6.18.a. Pode agora observar-se um nível de desempenho elevado na selecção de agentes com características de cooperação favoráveis. Esta melhoria resulta da possibilidade do agente raciocinar em relação a diferentes opções, pois no campo de atenção passou a estar mais do que uma opção para o agente considerar. Neste caso o raciocínio envolve a consideração dos interesses de cooperação em relação ao motivador σ_{coop} . A tomada de decisão ocorre pela ponderação da influência de ambos os motivadores, de acordo com a respectiva intensidade, tal como discutido na secção 6.4.1.

Estes resultados mostram o interesse de combinar as influências da emoção e do raciocínio. Ou seja, os aspectos emocionais proporcionam um suporte adaptativo importante, limitado no entanto pelas características ambientais, nomeadamente em termos dos padrões de incerteza e mudança do ambiente. Nesse sentido, o raciocínio é particularmente importante para limitar e compensar os efeitos potencialmente não adaptativos das influências emocionais.

Em contrapartida, o raciocínio apresenta uma complexidade intrínseca que pode pôr em causa o comportamento eficaz dos agentes perante essa mudança constante. Neste caso, são as dinâmicas emocionais que desempenham um papel chave na regulação e focagem dos processos de raciocínio, tornando-os praticáveis.

Pode assim observar-se uma relação estreita e simbiótica entre os fenómenos emocionais e cognitivos, tal como proposto inicialmente.

No exemplo que está a ser descrito, esta capacidade de focar os processos de raciocínio para lidar com a sua complexidade computacional pode ser regulada pelo *nível de atenção* γ . A figura 6.19 mostra os resultados da variação do nível de atenção perante a influência combinada dos motivadores de cooperação e fiabilidade.

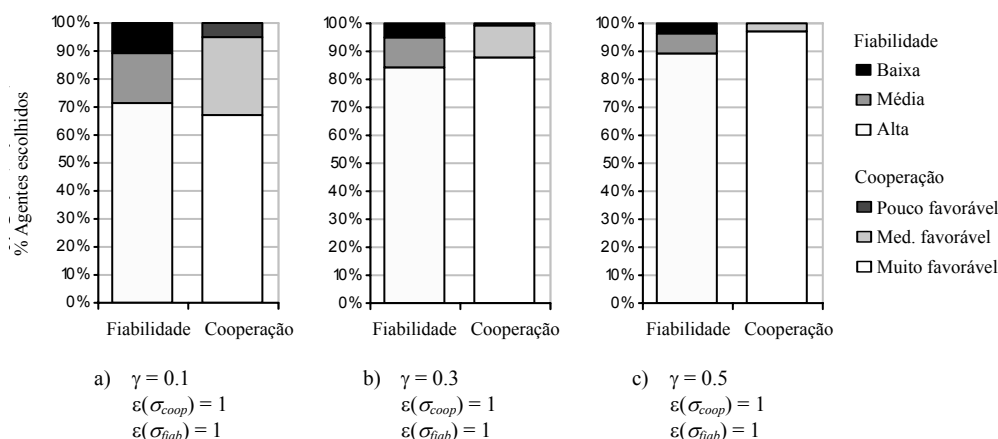


Figura 6.19. Comportamento do agente para diferentes valores do nível de atenção.

Como se pode observar, mesmo com um nível de focagem dos processos de raciocínio muito elevado, é possível obter resultados de eficácia significativa. Por exemplo, com uma redução de 50% das opções consideradas, o agente atinge um desempenho próximo do óptimo. Isto é possível devido à acção combinada dos aspectos emocionais e de raciocínio.

O efeito da variação do nível de atenção é mais notório nas características de cooperação dos agentes seleccionados devido à sua dependência directa dos processos de raciocínio. No entanto, devido à influência combinada dos dois motivadores, os resultados das características de cooperação também têm influência na formação da disposição emocional, levando a uma redução do comportamento exploratório do agente, o que melhora os resultados de fiabilidade à medida que mais experiências ocorrem. Isto é um exemplo de como a integração simbiótica entre aspectos emocionais e cognitivos levam a uma melhoria global do comportamento do agente perante condições de incerteza e dinamismo, como é o caso.

6.6 Conclusão

Neste capítulo foi discutida a forma como o modelo de agente proposto permite a integração entre os aspectos de base emocional e a actividade cognitiva de alto nível, nomeadamente com os processos de raciocínio e tomada de decisão. Essa integração assenta em duas vertentes base: a regulação da actividade cognitiva e a modulação emocional das alterações na estrutura cognitiva do agente.

A regulação da actividade cognitiva é caracterizada por duas perspectivas distintas: uma perspectiva espacial e uma perspectiva temporal. Cada uma destas

perspectivas é concretizada por mecanismos de regulação específicos que, tendo por base os mecanismos emocionais, permitem focar a atenção dos processos cognitivos e regular a sua execução. A modulação das alterações na estrutura cognitiva por disposições emocionais têm como resultado a formação de memórias emocionais, as quais desempenham um papel chave nos processos de raciocínio, permitindo delimitar e guiar esse raciocínio. Estas duas vertentes, regulação da actividade cognitiva e modulação da estrutura cognitiva, interagem mutuamente. Por um lado, os mecanismos de regulação cognitiva permitem focar os processos de raciocínio, nomeadamente, o raciocínio prospectivo, delimitando o âmbito de exploração. Por outro lado, os mecanismos de memória emocional realimentam os mecanismos de regulação cognitiva, produzindo um efeito antecipador, que predispõe o agente para responder adequadamente a possíveis situações futuras.

Subjacente a esta relação entre fenómenos emocionais e cognitivos estão dinâmicas emocionais, expressas pela noção de disposição emocional que, ao mesmo tempo, resultam da actividade cognitiva e modulam essa actividade cognitiva. Concretiza-se, assim, o proposto no início desta tese, a integração estreita e simbiótica entre fenómenos emocionais e cognitivos, na qual os fenómenos de base emocional desempenham um papel adaptativo, regulador dos processos cognitivos de um agente, fornecendo um suporte adequado para tornar esses processos viáveis perante restrições dinâmicas de tempo e de recursos.

Capítulo 7

Conclusão

A construção de sistemas inteligentes, capazes de auxiliar a actividade humana em tarefas tipicamente realizadas por humanos, as quais exigem elevados graus de autonomia e flexibilidade, tem sido um dos aspectos que tem caracterizado a civilização actual. Em paralelo, tem havido um questionar incessante acerca da natureza da inteligência, em particular se esta será replicável em sistemas artificiais. A procura de respostas para estas questões ganhou consistência com o advento da *inteligência artificial* como área de investigação científica.

A origem da investigação em inteligência artificial foi, em grande medida, inspirada pela tentativa de compreender os fenómenos naturais que poderiam estar por detrás do comportamento inteligente e a partir daí derivar analogias para a modelação e implementação de sistemas inteligentes artificiais (e.g. [Rosenblueth *et al.*, 1943; Turing, 1950; Von Neumann, 1951; Rosenblatt, 1958]). Progressivamente, a ligação aos fenómenos naturais foi sendo substituída pela ênfase nas abstracções conceptuais que deles derivaram. Em particular, prevaleceu uma perspectiva da inteligência como um processo de cálculo assente em representações simbólicas, sendo a inteligência vista como sinónimo de racionalidade [Simon, 1955]. A hipótese do *sistema físico de símbolos*, proposta por Newell e Simon [1976], é um dos marcos dessa perspectiva.

Nesta perspectiva, que se tornou preponderante na investigação em inteligência artificial, os fenómenos emocionais foram considerados como meros aspectos colaterais do comportamento inteligente, não tendo sido dado qualquer relevo ao seu estudo [Newell, 1983].

No início dos anos 90 esta perspectiva foi confrontada com evidências experimentais na área das neurociências, que indicavam uma estreita relação entre emoção e racionalidade. O trabalho realizado por António Damásio desempenhou nesse sentido um papel importante. Estas evidências levaram a um crescente

interesse pela investigação dos fenómenos emocionais, em especial no contexto da construção de agentes inteligentes.

No entanto, a perspectiva simbólica prevaleceu na maioria das abordagens, sendo os fenómenos emocionais modelados como atributos ou estados discretos etiquetados com expressões emocionais ou, em alternativa, como fenómenos de interrupção do processamento cognitivo.

Como resultado, não tem sido possível definir uma relação clara e genérica entre fenómenos emocionais e cognitivos que permita tirar partido de fenómenos emocionais na implementação de agentes de diferentes tipos e níveis de complexidade. Nomeadamente no que se refere à capacidade de implementar agentes capazes de actividades cognitivas de alto nível, como raciocínio e tomada de decisão e, simultaneamente, capazes de lidar com as restrições dinâmicas de tempo e de recursos, que caracterizam os ambientes reais.

É este o problema central abordado nesta tese. Nesse sentido foi proposta uma abordagem onde emoção e cognição são vistas como duas facetas complementares e interdependentes do comportamento inteligente. Nessa abordagem é igualmente reconhecido o carácter dinâmico e inerentemente contínuo dos fenómenos emocionais. Não sendo esta perspectiva compatível com os modelos de emoção que têm sido propostos, onde os fenómenos emocionais são vistos numa perspectiva meramente classificativa e discreta, foi proposto um modelo de emoção alternativo, onde o carácter dinâmico e contínuo dos fenómenos emocionais é uma característica base.

Ao adoptar esta perspectiva dinâmica e contínua como elemento chave na modulação dos fenómenos emocionais, a questão da integração entre fenómenos emocionais e processos cognitivos envolvendo representações conceptuais, como raciocínio e tomada de decisão, tornou-se igualmente um problema central. Nesse sentido foi proposto um modelo de agente capaz de suportar essa integração.

O modelo de agente proposto tem por base a analogia com a noção de estrutura dissipadora, a qual permite caracterizar, dum ponto de vista termodinâmico, sistemas capazes de manter uma estrutura interna estável em condições de não equilíbrio com o ambiente envolvente. Como resultado dessa analogia, a estrutura cognitiva dos agentes é modelada como agregados de sinais, cujas dinâmicas de evolução no tempo e de interacção permitem descrever, de modo adequado, as dinâmicas emocionais que caracterizam o modelo de emoção proposto.

De forma a suportar a implementação de agentes concretos de diferentes tipos e níveis de complexidade, foi definida uma arquitectura de agente genérica. Essa arquitectura é caracterizada por um conjunto de mecanismos base que concretizam e tiram partido da integração entre fenómenos emocionais e cognitivos, nomeadamente através da regulação da actividade cognitiva e através da modulação das alterações na estrutura cognitiva de um agente por fenómenos emocionais.

Os casos experimentais apresentados mostram a viabilidade da abordagem proposta, validando a hipótese central da tese, ou seja, os fenómenos de base emocional podem efectivamente desempenhar um papel adaptativo, regulador dos processos cognitivos de um agente, fornecendo um suporte para tornar esses processos viáveis perante restrições dinâmicas de tempo e de recursos.

7.1 Direcções para Trabalho Futuro

O trabalho desenvolvido caracterizou-se por uma grande abrangência dos temas em estudo, tendo sido explorados vários caminhos que, pela necessidade de limitar o âmbito do trabalho, ficaram por desenvolver de modo específico. Como resultado, diferentes direcções para trabalho futuro se perspectivam como relevantes e promissoras. De seguida é apresentada uma síntese de algumas dessas direcções para trabalho futuro.

7.1.1 Emoção e Interação Social

A relevância dos fenómenos emocionais na interação social é reconhecida no contexto da psicologia social (e.g. [Castelfranchi, 2000; Hewstone & Stroebe, 2001]). Apesar do contexto de interação social ter sido abordado em termos de cenário experimental, não se explorou a relação desse contexto com o surgimento e caracterização dos fenómenos emocionais, por exemplo em termos de emoções sociais, nem o papel que os fenómenos emocionais podem ter num contexto multi-agente, por exemplo, em termos de negociação ou de comportamento em grupo. Qualquer destes aspectos constitui interessantes caminhos para desenvolvimentos futuros.

7.1.2 Emoção, Memória e Aprendizagem

No trabalho desenvolvido, abordou-se o problema da formação de memórias emocionais numa perspectiva de suporte aos mecanismos de raciocínio. No entanto, a relação entre emoção e memória é um tema onde muitas questões se perspectivam, nomeadamente a relação entre os fenómenos emocionais e a consolidação de memórias de longo prazo, bem como a relação entre a formação e recordação de memórias e as tendências emocionais prevaletentes. Será também importante a investigação da abordagem proposta em termos de complexidade espacial e do desenvolvimento de mecanismos que permitam reduzir essa complexidade.

Por outro lado, no modelo proposto não são explorados aspectos específicos de aprendizagem. Nesse sentido, considera-se também importante o estudo da relação entre a abordagem proposta e métodos clássicos de aprendizagem, como é o caso da aprendizagem por reforço (“reinforcement learning”) ou da aprendizagem baseada em casos (“case-based learning”), no sentido de comparar resultados, clarificar condições de aplicabilidade e identificar pontos de contacto.

7.1.3 Expressão Emocional e Personalidade

Uma das áreas que têm impulsionado significativamente o estudo de fenómenos emocionais no contexto de agentes inteligentes é a expressão emocional. São exemplos de aplicações nesta área o desenvolvimento de interfaces de utilização envolvendo aspectos emocionais (e.g. [Picard, 1997; Rist *et al.*, 2003]), ou o desenvolvimento de personagens virtuais para interacção com interlocutores humanos (e.g. [Paiva *et al.*, 2004; Aylett, 2004]).

Apesar de não ser objectivo do trabalho realizado abordar este tema da expressão emocional, pelo facto de este ser um aspecto intrínseco aos fenómenos emocionais, ao longo da tese foram surgindo pontos de contacto com este tema, quer em termos da caracterização emocional dos agentes, apresentada nos capítulos 4 e 5, quer em termos da configuração dos parâmetros da arquitectura proposta, tal como referido no capítulo 6. Este facto é indicador do interesse que o desenvolvimento destes aspectos pode ter, em particular no que se refere à caracterização dos fenómenos emocionais envolvendo diferentes contextos cognitivos e escalas temporais. A abordagem desses aspectos está igualmente relacionada com a questão das emoções sociais, atrás referida, e com questões como *humor* (“*mood*”) e *personalidade* dos agentes.

7.1.4 Exploração de uma Abordagem Baseada em Sinais para Modelação Cognitiva

A abordagem proposta assenta numa perspectiva de modelação cognitiva com base em sinais, em contraponto às abordagens simbólicas e numéricas clássicas. Não sendo objectivo do trabalho realizado explorar os aspectos específicos que resultam deste tipo de modelação, optou-se por abstrair esses aspectos. Deste modo, as implicações e o potencial desta abordagem ficaram em grande medida por explorar. Este aspecto constitui assim uma importante área para desenvolvimento futuro, em particular no que se refere a uma implementação física sob a forma de *hardware* específico.

7.1.5 Ferramentas para Suporte ao Desenvolvimento de Agentes com Base nos Modelos Propostos

Ao longo do trabalho desenvolvido foram implementados vários protótipos, como consequência natural da exploração das ideias que estão na base dos modelos propostos e como forma de validar experimentalmente esses modelos. No entanto, não foi objectivo do trabalho realizado a implementação de uma ferramenta para suporte ao desenvolvimento de agentes com base nos modelos propostos, pelo que os protótipos realizados têm um carácter puramente experimental.

Para tornar possível a transição de um nível de investigação, onde a ênfase principal é a teoria de suporte, concretizada em cenários experimentais específicos, para um nível de investigação de carácter mais aplicacional, onde a ênfase possa ser em aplicações concretas, nomeadamente em cenários reais, é importante o desenvolvimento de ferramentas que permitam o desenvolvimento de modo eficiente e sistemático de agentes com base nos modelos propostos. Esta área constitui, por isso, uma direcção importante para trabalho futuro.

7.2 Considerações Finais

A implementação de agentes inteligentes capazes de resolver problemas reais em ambientes reais, é um objectivo da investigação em inteligência artificial cuja concretização se tem mostrado fugidia. Às expectativas idealizadas criadas nos anos 60 e 70 no contexto das abordagens simbólicas, juntaram-se as expectativas igualmente idealizadas, do “mundo como o modelo”, criadas nos anos 90 pelas abordagens comportamentais. No meio deste balançar entre extremos, pouco lugar

houve para considerar a hipótese dos fenómenos emocionais não serem um mero efeito colateral da actividade cognitiva, uma relíquia do nosso passado não racional, mas sim um aspecto central e basilar da própria inteligência. Os dados experimentais mostram de modo cada vez mais inequívoco essa relação chave entre emoção e inteligência.

Uma das principais ideias que resultam desta tese, é que a transposição para sistemas computacionais dessa relação entre emoção e inteligência não passa por uma mera extensão ou reformulação dos modelos cognitivos já existentes, como tem sido proposto pelas principais correntes de investigação na área, mas implica antes um reequacionar das próprias bases desses modelos. O modelo de emoção e de agente propostos e a abordagem baseada em sinais que lhes está subjacente, constituem direcções em que essa mudança pode ocorrer.

Apêndices

Apêndice A

Suporte Experimental

Ao longo do desenvolvimento da tese foram implementados vários protótipos no sentido de concretizar as ideias propostas. Cada protótipo constitui uma concretização específica da arquitectura de agente apresentada nos capítulos 5 e 6.

No desenvolvimento dos protótipos foi utilizada a plataforma *Win32* em conjunto com as linguagens de programação *C*, *C++* e *Prolog*. Como suporte para a execução concorrente dos processos cognitivos, foi utilizada uma interface de multi-programação que abstrai as primitivas de gestão e sincronização de processos disponíveis na plataforma *Win32*. As características específicas de cada protótipo são descritas de seguida.

A.1 Protótipo de Suporte aos Exemplos do Capítulo 5

O protótipo de suporte aos exemplos do capítulo 5 implementa os mecanismos motivacionais e de disposição emocional. Implementa também as diferentes configurações da arquitectura que correspondem aos agentes descritos nos exemplos.

Este protótipo não suporta processos de raciocínio, mas suporta a formação de memória emocional autobiográfica. A representação da memória autobiográfica é feita sob a forma de uma tabela indexada, onde cada posição da tabela corresponde a um elemento de memória, estando os vários elementos de memória ordenados sequencialmente de acordo com o instante de tempo a que se referem.

Foram utilizadas duas linguagens de programação na implementação deste protótipo: *C* e *C++*. A linguagem *C* foi utilizada, por uma questão de eficiência, na implementação dos mecanismos base, nomeadamente mecanismos motivacionais, mecanismos de disposição emocional e mecanismos de memória emocional autobiográfica. A linguagem *C++* foi utilizada na definição da organização do agente e na implementação da interface de utilização do protótipo, a qual é apresentada na figura seguinte.

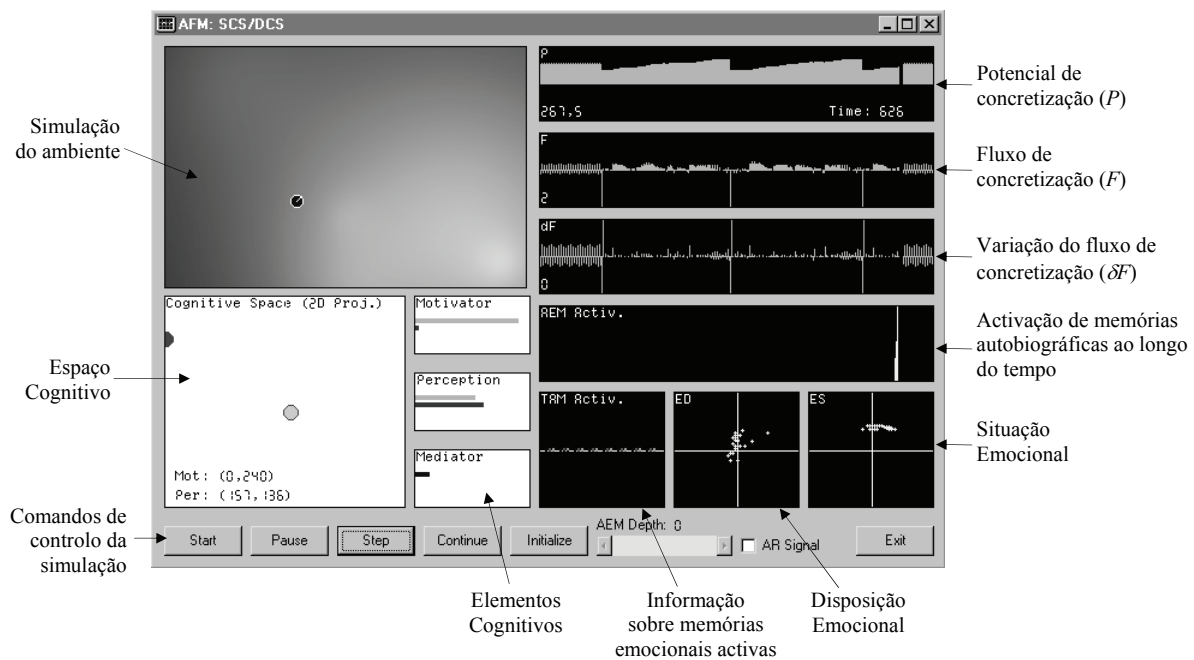


Figura A.1. Interface de utilização do protótipo de suporte aos exemplos do capítulo 5.

Este protótipo permite monitorar em tempo de execução vários aspectos que caracterizam o agente, nomeadamente a evolução dos elementos cognitivos e dos sinais subjacentes à formação da disposição e situação emocional, tal como ilustrado na figura A.1.

A.2 Protótipo de Suporte ao Cenário Experimental 1

O protótipo de suporte ao caso experimental 1, apresentado no capítulo 6, implementa, para além dos mecanismos motivacionais e de disposição emocional, os mecanismos de focagem de atenção e de focagem temporal. Implementa, também, a configuração da arquitectura que corresponde ao agente em estudo. Nessa configuração são suportados os processos de raciocínio e de assimilação. No entanto, os processos de assimilação apenas mantêm informação acerca do ambiente e da posição do agente, não formando memórias emocionais.

Este protótipo implementa ainda um agente com uma arquitectura BDI clássica (e.g. [Rao & Georgeff, 1991; Wooldridge, 2002]). Nessa implementação é utilizado o critério de reconsideração de intenções proposto por Schut, Wooldridge e Parsons

[2004], o qual serve de referência para comparação de resultados. A figura seguinte mostra a interface de utilização do protótipo.

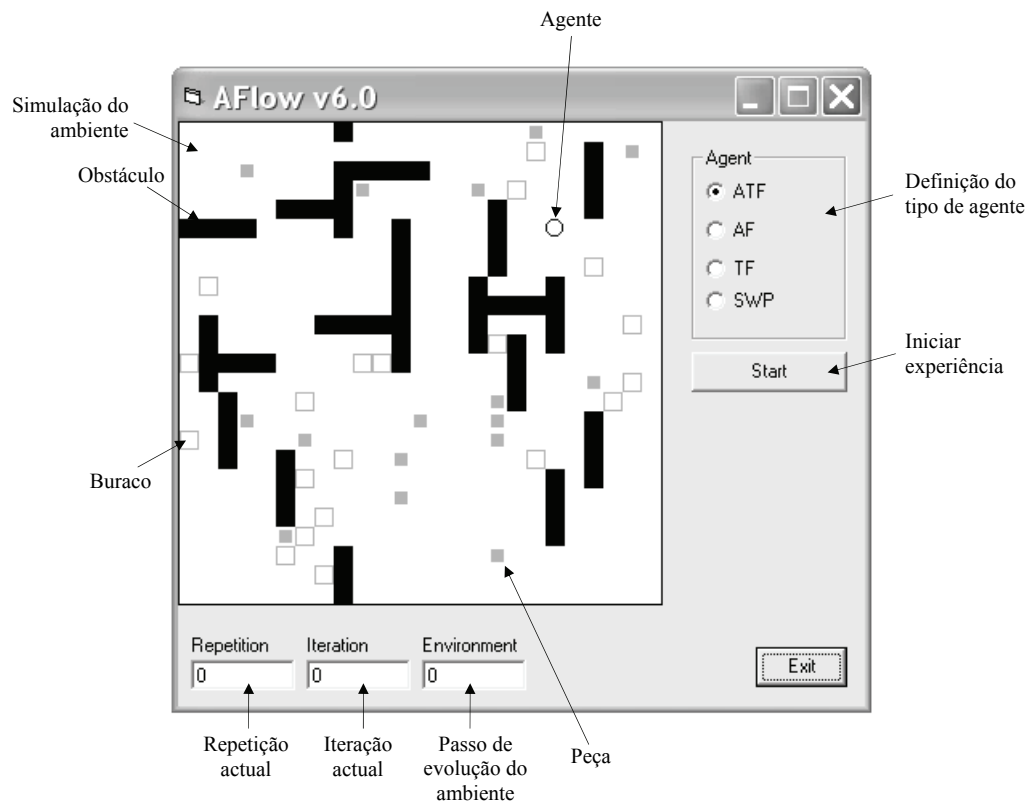


Figura A.2. Interface de utilização do protótipo de suporte ao cenário experimental 1.

De modo a garantir uma base de tempo independente da plataforma de execução, foi implementado um relógio de referência que gera impulsos de temporização periodicamente. O período desse relógio determina a *unidade de tempo* para efeitos de controlo de execução e análise de resultados. Ou seja, por cada impulso do relógio de referência, o tempo virtual da simulação avança uma unidade.

No sentido de ser possível definir de forma precisa o dinamismo do ambiente, tal como percebido pelo agente, quer a evolução do ambiente, quer a actividade cognitiva do agente são síncronas com o relógio de referência: por cada N_{Amb} impulsos do relógio de referência o ambiente evolui; por cada N_{Ag} impulsos do relógio de referência inicia-se um período de actividade cognitiva do agente. No entanto, a duração de um período de actividade cognitiva é variável. Essa duração depende do mecanismo de focagem temporal e da dimensão da árvore de procura explorada durante a formação dos planos.

Os parâmetros de divisão de frequência N_{Ag} e N_{Amb} , permitem gerar bases de tempo distintas, mas síncronas, para o agente e para o ambiente. A tabela seguinte mostra os valores destes parâmetros e os respectivos valores de dinamismo do ambiente γ , tal como definidos para o caso experimental apresentado na secção 6.2.4.

N_{Ag}	N_{Amb}	γ	$\log_{10}(\gamma)$
1000	1000	1,0	0,0
1000	794	1,3	0,1
1000	631	1,6	0,2
1000	501	2,0	0,3
1000	398	2,5	0,4
1000	316	3,2	0,5
1000	211	4,7	0,6
1000	200	5,0	0,7
1000	158	6,3	0,8
1000	126	7,9	0,9
1000	100	10,0	1,0
1000	79	12,7	1,1
1000	63	15,9	1,2
1000	50	20,0	1,3
1000	40	25,0	1,4
1000	32	31,3	1,5
1000	25	40,0	1,6
1000	20	50,0	1,7
1000	16	62,5	1,8
1000	13	76,9	1,9
1000	10	100,0	2,0

Tabela A.1. Valores dos parâmetros de divisão de frequência N_{Ag} e N_{Amb} , e respectivos valores de dinamismo do ambiente γ , para o caso experimental apresentado na secção 6.2.4.

A evolução do ambiente traduz-se no aparecimento e desaparecimento de “buracos”, de acordo com os respectivos tempos de gestação e de existência. Estes tempos são medidos em relação à base de tempo do ambiente, a qual é determinada a partir do relógio de referência, por divisão de frequência, de acordo com o valor de N_{Amb} . Deste modo, o ritmo de aparecimento e desaparecimento de “buracos” pode ser variado, originando diferentes graus de dinamismo do ambiente percebidos pelo agente, desde muito baixo dinamismo ($N_{Amb} = 1000$), até elevado dinamismo ($N_{Amb} = 1$).

O parâmetro de divisão de frequência N_{Ag} é fixo, tendo por único objectivo permitir obter as relações de divisão de frequência correspondentes aos diferentes níveis de dinamismo do ambiente definidos para estudo.

A.3 Protótipo de Suporte ao Cenário Experimental 2

O protótipo de suporte ao caso experimental 2, apresentado no capítulo 6, implementa os mecanismos motivacionais e de disposição emocional, os mecanismos de focagem de atenção e de focagem temporal e os mecanismos de memória emocional autobiográfica. Implementa, também, a configuração da arquitetura que corresponde ao agente em estudo. Nessa configuração são suportados os processos de assimilação, raciocínio e decisão, incluindo raciocínio prospectivo. A figura seguinte mostra a interface de utilização do protótipo.

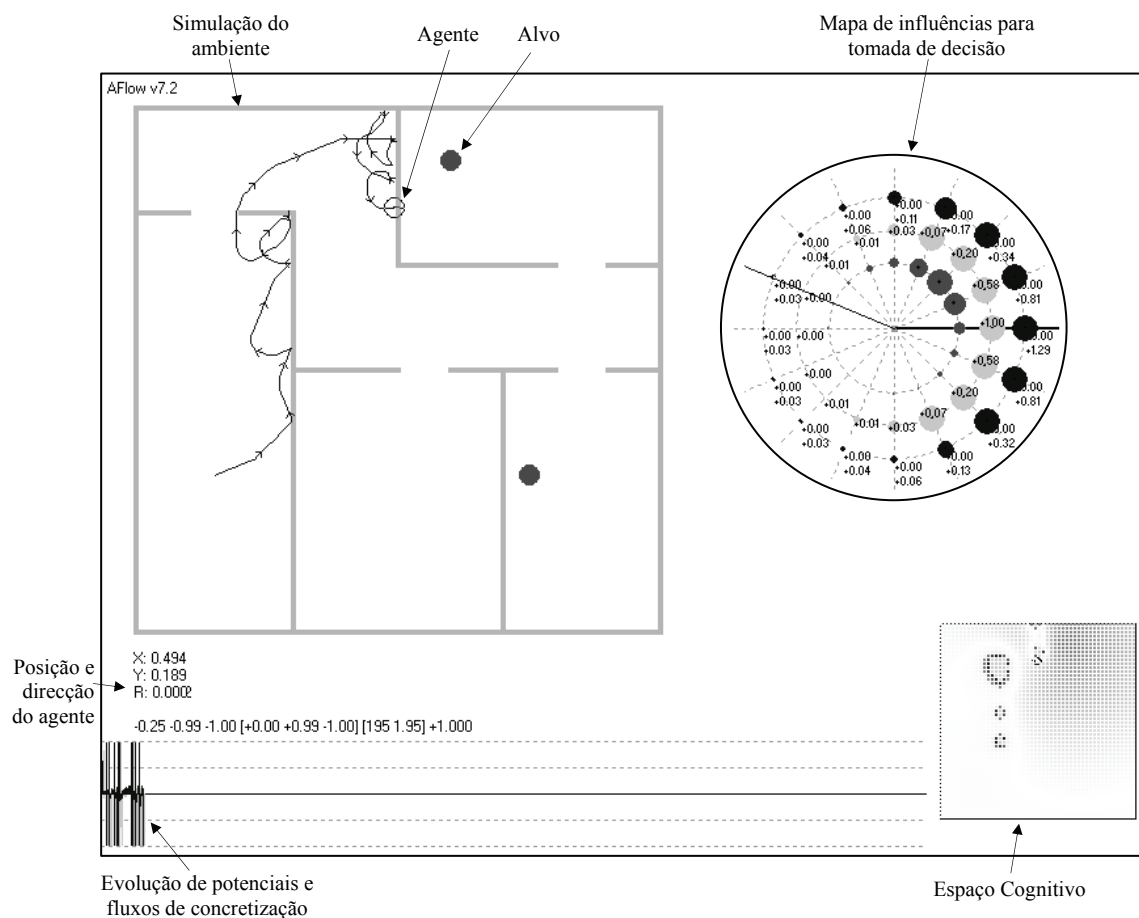


Figura A.3. Interface de utilização do protótipo de suporte ao cenário experimental 3.

A interface de utilização implementada permite observar a movimentação do agente no ambiente, bem como observar vários aspectos subjacentes à actividade interna do agente, nomeadamente a evolução no espaço cognitivo dos elementos de memória que vão sendo formados e as várias influências que determinam o raciocínio e tomada de decisão do agente. Essas influências são representadas sob a forma de um

mapa de influências, no qual três circunferências concêntricas representam os vários níveis de influência. O centro dessas circunferências representa a observação referente à posição corrente do agente, representando cada direcção (segmentos a tracejado) um mediador. As influências dos vários níveis referentes a cada mediador são representadas por círculos. A tonalidade de cada círculo indica o tipo de influência e a valência, a dimensão do círculo indica a intensidade.

A.4 Protótipo de Suporte ao Cenário Experimental 3

O protótipo de suporte ao caso experimental 3 implementa os mecanismos motivacionais e de disposição emocional, os mecanismos de focagem de atenção e de focagem temporal e os mecanismos de memória emocional autobiográfica. Implementa, também, a configuração da arquitectura que corresponde ao agente em estudo nesse caso experimental. Nessa configuração são suportados os processos de assimilação, raciocínio e decisão, incluindo raciocínio prospectivo. Este protótipo implementa, também, um mecanismo de raciocínio social baseado na *teoria da dependência* [Conte & Castelfranchi, 1995; Sichman, 1995; Sichman & Conte, 2002; Morgado, 1999].

Na teoria da dependência, as relações de dependência social existentes entre os agentes desempenham um papel central no raciocínio social. Nesse sentido, são definidos vários tipos de dependência social de acordo com as capacidades, planos e objectivos dos agentes envolvidos. Por exemplo, dependência *unilateral* ou *bilateral*, sendo bilateral se é dependência *mútua* ou *recíproca*, e outros tipos mais específicos de dependência social. O mecanismo de raciocínio social tira partido do conhecimento que um agente tem acerca dos outros agentes para identificar as relações de dependência que existem entre eles. As relações de dependência identificadas são utilizadas na tomada de decisão acerca da formação de coligações entre os agentes. Nesse sentido são definidos critérios de avaliação de situações de dependência, os quais definem relações de ordem entre os vários tipos de dependência (e.g. [Sichman, 1995; Sichman & Demazeau, 2001]).

Este protótipo implementa o mecanismo de raciocínio social baseado na teoria da dependência descrito em [Morgado, 1999], tendo sido utilizada a linguagem *Prolog* no seu desenvolvimento. A integração entre os subsistemas desenvolvidos nas linguagens

C e C++, e o subsistema de raciocínio social, foi feita com base na plataforma de desenvolvimento *Sicstus Prolog*.

No caso experimental 3, apresentado no capítulo 6, pretendeu-se abstrair os detalhes dos mecanismos de raciocínio social, pelo que foi implementado um modo de simulação do raciocínio social. Nesse modo, o resultado do raciocínio social é caracterizado em termos de interesse de cooperação. Para isso foram definidos três tipos de interesse de cooperação: *pouco-favorável*, *medianamente-favorável* e *muito-favorável*. Esses interesses de cooperação são determinados por uma distribuições aleatórias, tal como descrito no contexto desse caso prático. Este protótipo não tem interface gráfica.

Apêndice B

Definições Relativas à Constituição dos Elementos Cognitivos

Notação Utilizada

$j = \sqrt{-1}$	Unidade imaginária
z^*	Complexo conjugado
$\mathbf{v} = (v_1, v_2, \dots, v_n)$	Vector
$\langle u(t), w(t) \rangle$	Produto escalar entre sinais unidimensionais
$\langle u(t, x), w(t, x) \rangle$	Produto escalar entre sinais bidimensionais (espacio-temporais)
t	Variável tempo
x	Variável espaço
ω	Frequência temporal
Ω	Frequência temporal de referência de uma qualidade
ν	Frequência espacial
$\Delta\omega$	Desvio de frequência temporal
$\Delta\tau$	Período de tempo
Δx	Extensão de espaço
ε	Intensidade de um potencial cognitivo
\mathcal{G}	Coefficiente de amortecimento temporal
κ	Coefficiente de amortecimento espacial
ξ	Coefficiente de normalização de $\Delta\omega$

B.1 Definições gerais

$$\langle u(t), w(t) \rangle \equiv \int_{-\infty}^{\infty} u(t) w^*(t) dt$$

Produto escalar entre sinais unidimensionais

$$\langle u(t, x), w(t, x) \rangle \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(t, x) w^*(t, x) dt dx$$

Produto escalar entre sinais bidimensionais

$$\text{rect}(\zeta) \equiv \begin{cases} 0 & |\zeta| > 1/2 \\ 1 & |\zeta| \leq 1/2 \end{cases}$$

Função janela rectangular

B.2 Potenciais Cognitivos

$$\varphi_n(t) \equiv e^{-j\Omega_n t}$$

Sinal qualitativo base

$$\rho_n(t) \equiv \varepsilon_n e^{-\mathcal{G}t} e^{-j\xi\Delta\omega_n t} \text{rect}\left(\frac{t-\Delta\tau/2}{\Delta\tau}\right)$$

Sinal quantitativo de uma qualidade (sinal relativo a um período de actividade cognitiva de duração $\Delta\tau$)

$$p_n(t) \equiv \rho_n(t) \varphi_n(t)$$

Potencial cognitivo

B.3 Elementos Cognitivos

$$\sigma(t) \equiv \sum_{n=1}^K p_n(t)$$

Elemento cognitivo

$$\sigma \equiv (\rho_0, \rho_1, \dots, \rho_K)$$

Representação de um elemento cognitivo no espaço cognitivo num dado instante de tempo

$$\rho_n = \frac{\varepsilon_n}{\mathcal{G} + j\xi\Delta\omega_n}$$

Coefficiente dimensional de cada dimensão qualitativa

B.3.1 Derivação dos Coeficientes Dimensionais ρ_n

$$\rho_n = \langle p_n(t), \varphi_n(t) \rangle$$

Um coeficiente dimensional representa a correlação entre um potencial cognitivo e o sinal qualitativo base respectivo, expressa pelo produto escalar entre esses sinais.

$$= \int_{-\infty}^{\infty} p_n(t) \varphi_n^*(t) dt$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \varepsilon_n e^{-\vartheta t} e^{-j\xi\Delta\omega_n t} e^{-j\Omega_n t} \text{rect}\left(\frac{t-\Delta\tau/2}{\Delta\tau}\right) e^{j\Omega_n t} dt \\
&= \int_0^{\Delta\tau} \varepsilon_n e^{-\vartheta t} e^{-j\xi\Delta\omega_n t} e^{-j\Omega_n t} e^{j\Omega_n t} dt \\
&= \int_0^{\Delta\tau} \varepsilon_n e^{-\vartheta t} e^{-j\xi\Delta\omega_n t} dt \\
&= \left[\frac{\varepsilon_n}{-\vartheta - j\xi\Delta\omega_n} e^{-\vartheta t} e^{-j\xi\Delta\omega_n t} \right]_0^{\Delta\tau} \\
&= \frac{\varepsilon_n}{-\vartheta - j\xi\Delta\omega_n} (e^{-\vartheta\Delta\tau} e^{-j\xi\Delta\omega_n\Delta\tau} - 1) \\
&\approx 0 - \frac{\varepsilon_n}{-\vartheta - j\xi\Delta\omega_n} \\
&= \frac{\varepsilon_n}{\vartheta + j\xi\Delta\omega_n}
\end{aligned}$$

A intensidade ε_n dos potenciais cognitivos é constante durante cada período cognitivo.

No fim de um período cognitivo o termo de decaimento $e^{-\vartheta\Delta\tau}$ é aproximadamente nulo (pela definição do coeficiente de decaimento ϑ), logo $e^{-\vartheta\Delta\tau} e^{-j\xi\Delta\omega_n\Delta\tau} \approx 0$.

B.4 Elementos de Memória

$$\phi_v(x) \equiv e^{-\kappa x} e^{-j\nu x} \text{rect}\left(\frac{x-\Delta x/2}{\Delta x}\right)$$

Sinal de modulação espacial

$$\sigma(t, x) \equiv \sigma(t) \phi_v(x)$$

Elemento de memória

$$\boldsymbol{\sigma} \equiv (\rho_0, \rho_1, \dots, \rho_K) \rho_v$$

Representação de um elemento de memória no espaço cognitivo num dado instante de tempo

$$\rho_n = \frac{\varepsilon_n}{\vartheta + j\xi\Delta\omega_n}$$

Coefficiente dimensional de cada dimensão qualitativa

$$\rho_v = \frac{1}{\kappa + j\nu}$$

Coefficiente dimensional de modulação espacial

B.4.1 Derivação do Coeficiente Dimensional ρ_v

Considerando:

$$\sigma \equiv (\rho_0', \rho_1', \dots, \rho_K')$$

Obtém-se para os coeficientes dimensionais:

$$\begin{aligned} \rho_n' &= \langle p_n(t)\phi(x), \varphi_n(t) \rangle \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_n(t)\phi(x)\varphi_n^*(t)dt dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varepsilon_n e^{-\vartheta t} e^{-j\xi\Delta\omega_n t} e^{-j\Omega_n t} \text{rect}\left(\frac{t-\Delta\tau/2}{\Delta\tau}\right) e^{-\kappa x} e^{-j\nu x} \text{rect}\left(\frac{x-\Delta x/2}{\Delta x}\right) e^{j\Omega_n t} dt dx \\ &= \int_0^{\Delta x} \int_0^{\Delta\tau} \varepsilon_n e^{-\vartheta t} e^{-j\xi\Delta\omega_n t} e^{-j\Omega_n t} e^{-\kappa x} e^{-j\nu x} e^{j\Omega_n t} dt dx \\ &= \int_0^{\Delta\tau} \varepsilon_n e^{-\vartheta t} e^{-j\xi\Delta\omega_n t} dt \int_0^{\Delta x} e^{-\kappa x} e^{-j\nu x} dx \\ &= \left[\frac{\varepsilon_n}{-\vartheta - j\xi\Delta\omega_n} e^{-\vartheta t} e^{-j\xi\Delta\omega_n t} \right]_0^{\Delta\tau} \cdot \left[\frac{1}{-\kappa - j\nu} e^{-\kappa x} e^{-j\nu x} \right]_0^{\Delta x} \\ &\approx \left(0 - \frac{\varepsilon_n}{-\vartheta - j\xi\Delta\omega_n} \right) \cdot \left(0 - \frac{1}{-\kappa - j\nu} \right) \\ &= \frac{\varepsilon_n}{\vartheta + j\xi\Delta\omega_n} \cdot \frac{1}{\kappa + j\nu} \\ &= \rho_n \rho_v \quad \text{com} \quad \rho_n = \frac{\varepsilon_n}{\vartheta + j\xi\Delta\omega_n} \quad \text{e} \quad \rho_v = \frac{1}{\kappa + j\nu} \end{aligned}$$

Pelo que:

$$\begin{aligned} \sigma &\equiv (\rho_0 \rho_v, \rho_1 \rho_v, \dots, \rho_K \rho_v) \\ &= (\rho_0, \rho_1, \dots, \rho_K) \rho_v \end{aligned}$$

Bibliografia

Bibliografia

- Adolphs, R., Damásio, A. (2001). *The Interaction of Affect and Cognition: A Neurobiological Perspective*. In Forgas, J. (Ed.), *Affect and Social Cognition*, 27-45. Lawrence Erlbaum Publishers.
- Allen, S. (2001). *Concern Processing in Autonomous Agents*. PhD Thesis, School of Computer Science, University of Birmingham.
- Almeida, L., Silva, B., Bazzan, A. (2004). *Towards a Physiological Model of Emotions: First Steps*. In Hudlicka, E., Cañamero, L. (Eds.), *Architectures for Modeling Emotion: Cross-Disciplinary Foundations: AAAI Spring Symposium*, 1-4.
- Alvarado, N., Adams, S., Burbeck, S., Latta, C. (2001). *Project Joshua Blue: Design considerations for evolving an emotional mind in a simulated environment*. Proceedings of the 2001 AAAI Fall Symposium Emotional and Intelligent II: The Tangled Knot of Social Cognition.
- Alvarado, N., Adams, S., Burbeck, S. (2002). *The role of emotion in an architecture of mind*. IBM Research (<http://www.csupomona.edu/~nalvarado/PDFs/Emotion.pdf>).
- Anderson, J. (1983). *The Architecture of Cognition*. Lawrence Erlbaum Publishers.
- Angrilli, A., Cherubini, P., Pavese, A. (1997). *The Influence of Affective Factors on Time Perception*. *Perception & Psychophysics*, 59(6), 972-982.
- Antunes, L., Coelho, H. (1999). *Redesigning the Agents' Decision Machinery*. Proceedings of the International Workshop on Affective Interactions (IWAI).
- Arkin, R. (1998). *Behavior-Based Robotics*. MIT Press.
- Arzi-Gonczarowski, Z. (2002). *AI Emotions: Will One Know Them When One Sees Them?*. Proceedings of the 16th European Meeting on Cybernetics and Systems Research.
- Ashby, W. (1964). *An Introduction to Cybernetics*. Methuen.
- Ashby, F., Valentin, V., Turken, U. (2002). *The Effects of Positive Affect and Arousal on Working Memory*. In Moore, S., Oaksford, M. (Eds.), *Emotional cognition: from brain to behaviour*, 245-287. John Benjamins Publishing.
- Atkins, E., Abdelzaher, T. (2001). *Planning and Resource Allocation for Hard Real-time Fault-Tolerant Plan Execution*. *Autonomous Agents and Multi-Agent Systems*, 4, 57-78.
- Aubé, M. (1998). *A Commitment Theory of Emotions*. Proceedings of the 1998 AAAI Fall Symposium Emotional and Intelligent: The Tangled Knot of Cognition.
- Aylett, R. (2004). *Agents and Affect: Why Embodied Agents Need Affective Systems*. Proceedings of the Intelligent Virtual Environments and Virtual Agents (IVEVA 2004) Workshop.
- Aylett, R., Delgado, C. (2001). *Emotion and Agent Interaction*. Proceedings of the 2001 AAAI Fall Symposium Emotional and Intelligent II: The Tangled Knot of Social Cognition, 15-16.
- Barnard, C. (2004). *Animal Behavior: Mechanism, Development, Function and Evolution*. Prentice Hall.

- Bates, J. (1994). *The role of emotion in believable agents*. Communications of the ACM, 37(7), 122-125.
- Bateson, G. (1980). *Mind and Nature - A Necessary Unity*. Bantam Books.
- Beaudoin, L. and Sloman, A. (1994). *Goal Processing in Autonomous Agents*. PhD Thesis, School of Computer Science, University of Birmingham.
- Beaudoin, L. and Sloman, A. (1993). *A study of motive processing and attention*. In Sloman, A., Hogg, D., Humphreys, G., Partridge, D., and Ramsay, A., (Eds.)m Prospects for Artificial Intelligence, 229-238. IOS Press.
- Beer, R. (1995). *A Dynamical Systems Perspective on Agent-Environment Interaction*. Artificial Intelligence, 72(1-2), 173-215.
- Belavkin, R. (2001). *The Role of Emotion in Problem Solving*. Proceedings of the AISB'01 Symposium on Emotion, Cognition and Affective Computing, 49-57.
- Ben-Ze'ev, A. (2000). *The Subtlety of Emotions*. MIT Press.
- Berridge, K. (2004). *Motivation concepts in behavioral neuroscience*. Physiology & Behavior, 81(2), 179-209.
- Blumberg, B., Todd, P., Maes, P. (1996). *No Bad Dogs: Ethological Lessons for Learning*. Proceedings of International Conference of the Society for Adaptative Behaviour (SAB'96), 1191-1194.
- Botelho, L., Coelho, H. (1996). *Emotion-based attention shift in autonomous agents*. Proceedings of the ECAI'96 Workshop on Agent Theories, Architectures and Languages (ATAL-96), 277-291.
- Botelho, L., Coelho, H. (1998). *Adaptive agents: emotion learning*. Proceedings of the Workshop on Grounding Emotions in Adaptative Systems, Fifth International Conference of the Society for Adaptative Behaviour (SAB 98), 19-24.
- Botelho, L., Coelho, H. (2001). *Machinery for Artificial Emotions*. Cybernetics and Systems, 32(5), 465-506.
- Bower, G. (1994). *Some Relations Between Emotions and Memory*. In Ekman, P., Davidson, R. (Eds.), The Nature of Emotion: Fundamental Questions, 303-305. Oxford Univ. Press.
- Bower, G. , Forgas, J. (2001). *Mood and Social Memory*. In Forgas, J. (Ed.), Affect and Social Cognition, 95-117. Lawrence Erlbaum Publishers.
- Brafman, R., Tennenholtz, M. (1997). *Modeling agents as qualitative decision makers*. Artificial Intelligence, 94, 57-77.
- Braitenberg, V. (1984). *Vehicles: Experiments in Synthetic Psychology*. MIT Press.
- Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Harvard University Press.
- Bratman, M. (1990). *What is intention?*. In Cohen, P., Morgan, J., Pollack, M., (Eds.), Intentions in Communication, 15-32. MIT Press.
- Bratman, M., Israel, D., Pollack, M. (1988). *Plans and Resource-Bounded Pratical Reasoning*. Computational Intelligence, 4(4), 349-355.
- Breazeal, C. (2002). *Designing Sociable Machines*. MIT Press.
- Brooks, R. (1986). *A robust layered control system for a mobile robot*. IEEE Journal of Robotics and Automation, 14-23.
- Brooks, R. (1990). *Elephants Don't Play Chess*. In Maes P. (Ed.), Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back. MIT Press.

- Brooks, R. (1991). *Intelligence without reason*. Proceedings of the 12th International Joint Conference on Artificial Intelligence.
- Brooks, R. (1999). *Cambrian Intelligence: The Early History of the New AI*. MIT Press.
- Burt, A. (1998). *Emotionally Intelligent Agents: The Outline of a Resource-Oriented Approach*. Proceedings of the 1998 AAAI Fall Symposium Emotional and Intelligent: The Tangled Knot of Cognition.
- Butazzo, G. (2001). *Artificial Cosnciousness: Utopia or Real Possibility?*. IEEE Computer, 34(7).
- Cacioppo, J., Gardner, W. (1999). *Emotion*. Annual Reviews of Psychology, 50, 191-214.
- Cahill, L., McGaugh, J. (1998). *Mechanisms of Emotional Arousal and Lasting Declarative Memory*. Trends in Neuroscience, 21(7), 294-299. Elsevier.
- Cañamero, L. (1997). *A Hormonal Model of Emotions for Behavior Control*. Proceedings of the 4th European Conference on Artificial Life (ECAL'97).
- Cañamero, L. (2000). *Designing Emotions for Activity Selection in Autonomous Agents*. In R. Trappl, (Ed.), Emotions in Humans and Artifacts. MIT Press.
- Cañamero, L. (2003). *Designing Emotions for Activity Selection in Autonomous Agents*. In Trappl, R., Petta, P., Payr, S. (Eds.), Emotions in Humans and Artifacts. MIT Press.
- Cañamero, L., Corruble, V. (1999). *Situated Cognition: A Challenge to Artificial Intelligence*. In J. Bliss, P. Light, R. Saljo (Eds.), Learning Sites: Social and Technological Contexts for Learning Socially Intelligent Agents: AAAI Fall Symposium. Elsevier.
- Cañamero, L., Van de Velde, W. (1999). *Emotionally Grounded Social Interaction*. In K. Dautenhahn, (Ed.) Human Cognition and Social Agent Technology. Advances in Consciousness Research Series. John Benjamins Publishers.
- Carlson, A. (1986). *Communication Systems*. McGraw-Hill.
- Carver, C., Scheier, M. (2000). *Origins and Functions of Positive and Negative Affect: A Control-Process View*. Motivational science: social and personality perspectives, 256-272. Psychology Press.
- Carver, C., Scheier, M. (2002). *Control Processes and Self-organization as Complementary Principles Underlying Behavior*. Personality and Social Psychology Review, 6, 304-315.
- Castelfranchi, C. (2000). *Affective Appraisal vs Cognitive Evaluation in Social Emotions and Interactions*. In A. Paiva (Ed.), Affective Interactions, Lecture Notes on Artificial Intelligence, vol. 1814. Springer-Verlag.
- Castelfranchi, C. (2002). *The theory of social functions: Challenges for computational social science and multi-agent learning*. Complexity in Social Science. COSI Summer School.
- Castelfranchi, C., Rosis, F., Falcone, R. (1997). *Social Attitudes and Personalities in Agents*. Proceedings of the 1997 AAAI Fall Symposium Socially Intelligent Agents.
- Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. MIT Press.
- Clark, A., Toribio, J. (Eds.) (1998). *Machine Intelligence: Perspectives on the Computational Model*. Garland Publishing.
- Clore, G., Ortony, A. (2000). *Cognition in Emotion: Always, Sometimes, or Never?*. In Lane, R., Nadel, L. (Eds.), Cognitive Neuroscience of Emotion, 24-61. Oxford Univ. Press.
- Coelho, H., Paiva, A. (1998). *A Mente e o Mundo lá Fora: Uma Introdução à Inteligência Artificial Distribuída*. Faculdade de Ciências da Universidade de Lisboa / Instituto Superior Técnico.

- Cohen, P., Levesque, H. (1990). *Intention Is Choice with Commitment*. Artificial Intelligence, 42.
- Conte, R., Castelfranchi, C. (1995). *Cognitive and Social Action*. UCL Press.
- Coren, S., Ward, L., Enns, J. (2004). *Sensation and Perception*. John Wiley & Sons.
- Coulson, M. (2004). *An "Everything But" Framework for Modeling Emotion*. In Hudlicka, E., Cañamero, L. (Eds.), *Architectures for Modeling Emotion: Cross-Disciplinary Foundations: AAAI Spring Symposium*.
- Damásio, A. (1994). *O Erro de Descartes*. Europa-América.
- Damásio, A. (2000). *A Second Chance for Emotion*. In Lane, R., Nadel, L. (Eds.), *Cognitive Neuroscience of Emotion*. Oxford Univ. Press.
- Damásio, A. (2003). *Ao Encontro de Espinosa: As Emoções Sociais e a Neurologia do Sentir*. Europa-América.
- Davis, M., Whalen, P. (2001). *The amygdala: vigilance and emotion*. Molecular Psychiatry 6, 13-24.
- Dean, T., Kaelbling, L. (1993). *Planning with Deadlines in Stochastic Domains*. Proceedings of the 11th National Conference on Artificial Intelligence.
- Dennet, D. (1987). *The Intentional Stance*. MIT Press.
- Dill, K., Bromberg, S. (2003). *Molecular driving forces: statistical thermodynamics in chemistry and biology*. Garland Science.
- d'Inverno, M., Luck, M., Wooldridge, M. (1997). *Cooperation Structures*. Proceedings of the 15th International Joint Conference on Artificial Intelligence, 600-605.
- Dolan, R. (2002). *Emotion, cognition, and behavior*. Science, 298.
- Doshi, P., Gmytrasiewicz, P. (2004). *Towards Affect-based Approximations to Rational Planning: A Decision-Theoretic Perspective to Emotions*. In Hudlicka, E., Cañamero, L. (Eds.), *Architectures for Modeling Emotion: Cross-Disciplinary Foundations: AAAI Spring Symposium*.
- Dürr, H., Popp, F., Schommers, W. (2002). *What is life? : Scientific approaches and philosophical positions*. World Scientific.
- Dusenbery, D. (1996). *Life at small scale: the behavior of microbes*. Scientific American Library.
- Ekman, P., Davidson, R. (1994). *The nature of emotion: fundamental questions*. Oxford University Press.
- Eliasmith, C., Anderson, C. (2003). *Neural engineering: computation, representation, and dynamics in neurobiological systems*. MIT Press.
- Elliot, C. (1992). *The Affective Reasoner: A Process Model of Emotions in a Multi-agent System*. The Institute for the Learning Sciences, Technical Report No. PhD thesis, Northwestern University.
- Elster, J. (1999). *Alchemies of the mind: rationality and the emotions*. Cambridge University Press.
- Fall, C., Marland, E., Wagner, J., Tyson, J. (2002). *Computational Cell Biology*. Springer-Verlag.
- Feigenbaum, E., Buchanan, B., Lederberg, J. (1971). *On generality and problem solving: A case study using the DENDRAL program*. Machine Intelligence, 6, 165-190.

- Fellous, J. (2004). *From Human Emotions to Robot Emotions*. In Hudlicka, E., Cañamero, L. (Eds.), *Architectures for Modeling Emotion: Cross-Disciplinary Foundations: AAAI Spring Symposium*.
- Ferber, J. (1999). *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence*. Addison-Wesley.
- Ferguson, I. (1992). *Turing Machines: Autonomous Agents with Attitudes*. IEEE Computer, 25(5).
- Ferguson, I. (1995). *Integrated control and coordinated behavior: a case for agent models*. In Wooldridge, M., Jennings, R. (Eds.), *Intelligent Agents: Theories, Architectures and Languages*, 203-218.
- Fiedler, K. (2001). *Affective Influences on Social Information Processing*. In Forgas, J. (Ed.), *Affect and Social Cognition*, 163-183. Lawrence Erlbaum Publishers.
- Fliedner, D. (2001). *Six Levels of Complexity: A Typology of Processes and Systems*. Journal of Artificial Societies and Social Simulation vol. 4, no. 1, <<http://www.soc.surrey.ac.uk/JASSS/4/1/4.html>>.
- Forgas, J. (2003). *Affective Influencees on Attitudes and Judgments*. In Davidson, R., Scherer, K., Goldsmith, H. (Eds.), *Handbook of Affective Sciences*.
- Franklin, S., McCauley, L. (2004). *Feelings and Emotions as Motivators and Learning Facilitators*. In Hudlicka, E., Cañamero, L. (Eds.), *Architectures for Modeling Emotion: Cross-Disciplinary Foundations: AAAI Spring Symposium*, 48-51.
- Freeman, W. (2000). *Emotion is Essential to All Intentional Behaviors*. In Lewis, D., Granic, I. (Eds.), *Emotion, Development, and Self-Organization: Dynamic Systems Approachs to Emotional Development*, 209-235. Cambridge University Press.
- French, A. (1971). *Vibrations and Waves*. W. W. Norton & Company.
- Frijda, N. (1986). *The Emotions*. Cambridge University Press.
- Frijda, N. (2001). *The Laws of Emotion*. In Parrott, W. (Ed.), *Emotions in social psychology: essential readings*, 57-69. Psychology Press.
- Frijda, N., Moffat, D. (1993). *A Model of Emotion and Communication*. Proceedings of the IEEE International Workshop on Robot and Human Communication, 29-34.
- Gadanhó, S. (1999). *Reinforcement Learning in Autonomous Robots: An Empirical Investigation of the Role of Emotions*. PhD thesis, University of Edinburgh.
- Gadanhó, S. (2002). *Emotional and Cognitive Adaptation in Real Environments*. Proceedings of the 16th European Meeting on Cybernetics and Systems Research.
- Gärdenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought*. MIT Press.
- Genesereth, R., Nilsson, N. (1987). *Logical Foundations of Artificial Intelligence*. Morgan Kaufman.
- Georgeff, M., Rao, A. (1998). *Rational Software Agents: From Theory to Practice*. In Jennings, N., Wooldridge, M. (Eds.), *Agent Technology - Foundations, Applications, and Markets*. Springer-Verlag.
- Gigerenzer, G. (2000). *Adaptive Thinking*. Oxford University Press.
- Gigerenzer, G., Selten, R. (Eds.) (1999). *Bounded Rationality: The Adaptive Toolbox*. MIT Press.
- Giunti, M. (1997). *Computation, dynamics, and cognition*. Oxford University Press.

- Giurfa, M., Menzel, R. (2003). *Cognitive Architecture of a Mini-Brain*. In Kühn, R., Menzel, R., Menzel, W., Ratsch, U., Richter, M., Stamatescu, I. (Eds.), *Adaptivity and Learning: An Interdisciplinary Debate*, 23-48.
- Gratch, J., Marsella, S. (2003). *The Architectural Role of Emotions in Cognitive Systems*. USC Institute for Creative Technologies.
- Gratch, J., Marsella, S. (2004). *Evaluating a General Model of Emotional Appraisal and Coping*. In Hudlicka, E., Cañamero, L. (Eds.), *Architectures for Modeling Emotion: Cross-Disciplinary Foundations: AAAI Spring Symposium*.
- Gray, J., Braver, T., Raichle, M. (2002). *Integration of Emotion and Cognition in the Lateral Prefrontal Cortex*. *Proceedings of the National Academy of Sciences USA*, 99, 4115-4120.
- Habib, M. (1998). *Bases Neurologiques des Comportements*. (Edição Portuguesa - Climepsi Editores, 2000). Masson Éditeur.
- Haddadi, A. (1998). *Towards a Pragmatic Theory of Interactions*. In Huhns, M., Singh, M. (Eds.), *Readings in Agents*. Morgan Kaufman.
- Haddadi, A., Sundermeyer, K. (1996). *Belief-Desire-Intention Agent Architectures*. In O'Hare, G., Jennings, N. (Eds.), *Foundations of Distributed Artificial Intelligence*. John Wiley & Sons.
- Hanoch, Y. (2002). *Neither an Angel nor an Ant: Emotion as an Aid to Bounded Rationality*. *Journal of Economic Psychology*, 23.
- Hart, P., Nilsson, N., Raphael, B. (1968). *A Formal Basis for the Heuristic Determination of Minimum Cost Paths*. *IEEE Trans. Syst. Sci. Cybern.*, 4, 100-107.
- Hebb, D., Thompson, W. (1979). *The Social Significance of Animal Studies*. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology* (2nd ed.), 729-774. Addison-Wesley.
- Henninger, A., Jones, R., Chown, E. (2003). *Behaviors that Emerge from Emotion and Cognition: Implementation and Evaluation of a Symbolic-Connectionist Architecture*. *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multi-Agent Systems*, 321-327.
- Hewitt, C. (1973). *A Universal Modular ACTOR formalism for AI*. *Proceedings of the 3rd International Joint Conference on Artificial Intelligence*, 235-245.
- Hewstone, M., Stroebe, W. (2001). *Introduction to Social Psychology*. Blackwell Publishers.
- Heylighen, F. (2003). *The Science of Self-organization and Adaptivity*. In *Knowledge Management, Organizational Intelligence and Learning, and Complexity*. EOLSS Publishers.
- Higgins, E. (2001). *Promotion and Prevention Experiences: Relating Emotions to Nonemotional Motivational States*. In Forgas, J. (Ed.), *Affect and Social Cognition*, 186-208. Lawrence Erlbaum Publishers.
- Higgins, E., Kruglanski, A. (2000). *Motivational science: social and personality perspectives*. Psychology Press.
- Hofstadter, D. (1995). *Fluid concepts and creative analogies: computer models of the fundamental mechanisms of thought*. Penguin Books.
- Holland, J. (1995). *How Adaptation Builds Complexity*. Addison-Wesley.
- Hoppensteadt, F. (1997). *An Introduction to the Mathematics of Neurons: Modeling in the frequency Domain*. Cambridge University Press.

- Hsee, C., Abelson, R. (1991). *Velocity Relation: Satisfaction as a Function of the First Derivative of Outcome Over Time*. Journal of Personality and Social Psychology, 60, 341-347.
- Hudlicka, E. (1998). *Modeling Affect Regulation and Induction*. Proceedings of the 1998 AAAI Fall Symposium Emotional and Intelligent: The Tangled Knot of Cognition, 51-56.
- Hudlicka, E. (2002). *This time with feeling: Integrated Model of Trait and State Effects on Cognition and Behavior*. Applied Artificial Intelligence, 16 (7-8), 1-31.
- Hudlicka, E. (2004). *Two Sides of Appraisal: Implementing Appraisal and Its Consequences within a Cognitive Architecture*. In Hudlicka, E., Cañamero, L. (Eds.), Architectures for Modeling Emotion: Cross-Disciplinary Foundations: AAAI Spring Symposium, 70-78.
- Hudlicka, E., Sloman, A., den Dulk, P. (2004). *Computational Models of Affect*. <http://emotion.salk.edu/Emotion/EmoRes/CompAI/CompAI.html>.
- Ifeachor, E., Jervis, B. (2002). *Digital signal processing: a practical approach*. Prentice Hall.
- Ingrand, F., Georgeff, M., Rao, A. (1992). *An Architecture for Real Time Reasoning and System Control*. IEEE Expert, 34-44.
- Jackendoff, R., Bloom, P., Wynn, K. (2002). *Language, logic, and concepts: essays in memory of John Macnamara*. MIT Press.
- Jennings, N., Wooldridge, M. (1998). *Applications of Intelligent Agents*. In Jennings, N., Wooldridge, M. (Eds.), Agent Technology - Foundations, Applications, and Markets. Springer-Verlag.
- Jennings, N., Sycara, K., Wooldridge, M. (1998). *A Roadmap of Agent Research and Development*. Autonomous Agents and Multi-Agent Systems, 1, 7-38. Kluwer Academic Publishers.
- Johnson, M., Munakata, Y., Gilmore, R. (2002). *Brain development and cognition*. Blackwell Publishers.
- Johnston, V. (1999). *Why we feel: the science of human emotions*. Perseus Books.
- Kaszniak, A. (2001). *Emotion and Consciousness: Current Research and Controversies*. In Kaszniak, A. (Ed.), Emotions qualia and consciousness. World Scientific.
- Kelso, J. (1999). *Dynamic patterns: the self organization of brain and behavior*. MIT Press.
- Kinny, D., Georgeff, M. (1991). *Commitment and Effectiveness of Situated Agents*. Proceedings of the 12th International Joint Conference on Artificial Intelligence.
- Kiss, G. (1996). *Agent Dynamics*. In O'Hare, G., Jennings, N. (Eds.), Foundations of Distributed Artificial Intelligence. John Wiley & Sons.
- Kohonen, T. (1980). *Content-Addressable Memories*. Springer-Verlag.
- Kondepudi, D., Prigogine, I. (1998). *Modern Thermodynamics: From Heat Engines to Dissipative Structures*. J. Wiley & Sons.
- Korf, R. (1990). *Real-time heuristic search*. Artificial Intelligence, 42(2-3), 189-211.
- Kühn, R., Menzel, R., Menzel, W., Ratsch, U., Richter, M., Stamatescu, I. (2003). *Adaptivity and Learning: An Interdisciplinary Debate*. Springer-Verlag.
- Laird, J., Newell, A., Rosenbloom, P. (1987). *SOAR: An architecture for general intelligence*. Artificial Intelligence, 33(1), 1-64.
- Lane, R., Nadel, L. (2000). *Cognitive neuroscience of emotion*. Oxford University Press.
- Laplace, P. (1992). *Real-Time Systems Design and Analysis*. IEEE Press.
- Lazarus, R. (1991). *Emotion and adaptation*. Oxford University Press.

- LeDoux, J. (1996). *The Emotional Brain*. Simon and Schuster.
- LeDoux, J. (2000). *Cognitive-Emotional Interactions: Listen to the Brain*. In Lane, R., Nadel, L. (Eds.), *Cognitive Neuroscience of Emotion*, 129-155. Oxford Univ. Press.
- Leventhal, H., Scherer, K. (1987). *The relationship of emotion to cognition: A functional approach to a semantic controversy*. *Cognition and Emotion*, 1, 3-28.
- Levine, M. (2000). *Fundamentals of Sensation and Perception*. Oxford University Press.
- Lewis, M., Granic, I. (2000). *Emotion, development, and self organization: dynamic systems approaches to emotional development*. Cambridge University Press.
- Lisetti, C., Glinos, D., Murphy, R., Tardif, R. (2002). *A Multilevel Emotion-Based Architecture for Autonomous Agents*. In Working Notes of the AAAI Fall Symposium Series on Human-Robot Interaction.
- Liu, J. (2001). *Autonomous Agents and Multi-Agent Systems: Explorations in Learning, Self-Organization and Adaptive Computation*. World Scientific.
- Lorenz, K. (1950). *The Comparative Method in Studying Innate behavior Patterns*. *Mechanisms of Learning and Motivation*, 367-398. Lawrence Erlbaum Publishers.
- Lorigny, J. (1992). *Les système autonomes*. Dunod.
- Luck, M., et al. (2002). *Agent Technology Roadmap*. AgentLink II, the European Network of Excellence for Agent-Based Computing (IST-1999-29003).
- Lynch, S. (2004). *Dynamical systems with applications using MATLAB*. Birkhäuser.
- Lytton, W. (2002). *From computer to brain: foundations of computational neuroscience*. Springer.
- Maass, W., Bishop, M. (Eds.) (1999). *Pulsed Neural Networks*. MIT Press.
- MacLean, P. (1952). *Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (visceral brain)*. *Electroen. Clin. Neuro.*, 4(4), 407-418.
- MacLean, P. (1990). *The triune brain in evolution*. Plenum Press.
- Maes, P. (Ed.) (1991). *Designing Autonomous Agents: Theory and Practice From Biology to Engineering and Back*. MIT Press.
- Mainzer, K. (2004). *Thinking in complexity: the computational dynamics of matter, mind, and mankind*. Springer Verlag.
- Malec, J. (2000). *On augmenting reactivity with deliberation in a controlled manner*. In Hannebauer, M., Wendler, J., Pagello, E. (Eds.), *Balancing Reactivity and Social Deliberation in Multi-Agent Systems*. Springer-Verlag.
- Mallot, H. (1999). *Spatial Cognition: Behavioral Competences, Neural Mechanisms, and Evolutionary Scaling*. *Kognitionswissenschaft*, 8, 40-48. Springer-Verlag.
- Manuila, L., Manuila, A., Lewalle, P., Nicoulin, M. (1999). *Dictionaire Médical*. (Edição Portuguesa - Climepsi Editores, 2000). Masson Éditeur.
- Marsella, S., Gratch, J. (2002). *A step toward irrationality: using emotion to change belief*. *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-agent Systems*.
- Marsella, S., Gratch, J. (2003). *Modeling coping behaviors in virtual humans: Don't worry, be happy*. *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multi-agent Systems*.
- Martínez-Miranda, J., Aldea, A. (2004). *Emotions in Human and Artificial Intelligence*. *Computers in Human Behavior*. Elsevier.

- Matlin, M. (2003). *Cognition*. John Wiley & Sons.
- Matthews, G., Zeidner, M., Roberts, R. (2002). *Emotional Intelligence: Science & Myth*. MIT Press.
- Maturana, H., Varela, F. (1987). *The Tree of Knowledge: The Biological Roots of Human Understanding*. Shambhala Publications.
- McCall, S. (2002). *Deliberation Reasons and Explanation Reasons*. In Jackendoff, R., Bloom, P., Wynn, K. (Eds.), *Language, Logic, and Concepts*. MIT Press.
- McCarthy, J. (1979). *Ascribing Mental Qualities to Machines*. Stanford Artificial Intelligence Laboratory, Memo AIM-326.
- McCauley, L., Franklin, S., Bogner, M. (2000). *An Emotion-Based "Conscious" Software Agent Architecture*. In A. Paiva (Ed.), *Affective Interactions, Lecture Notes on Artificial Intelligence*, vol. 1814. Springer-Verlag.
- McFarland, D., Bösser, T. (1993). *Intelligent behavior in animals and robots*. MIT Press.
- Mellers, B. (2000). *Choice and the Relative Pleasure of Consequences*. *Psychological Bulletin*, 126, 910-924.
- Mellers, B., McGraw, A. (2001). *Anticipated Emotions as Guides to Choice*. *Current Directions in Psychological Science*, 10(6).
- Michaud, F. (2002). *EMIB - Computational Architecture Based on Emotion and Motivation for Intentional Selection and Configuration of Behaviour-Producing Modules*. *Cognitive Science Quarterly*, 1.
- Minsky, M. (1985). *The Society of Mind*. Simon & Schuster.
- Mitchell, M. (1997). *Theories of Structure Versus Theories of Change*. *Behavioral and Brain Sciences*.
- Moffat, D. & Frijda, N. (1995). *Where there's a Will there's an agent*. *Proceedings of the workshop on agent theories, architectures, and languages on Intelligent agents*, 245-260.
- Moore, S., Oaksford, M. (2002). *Emotional cognition: from brain to behaviour*. John Benjamins Publishing.
- Morgado, L. (1999). *Coordenação em Sistemas Multi-Agente: Formação de coligações com base em noções de dependência social*. Dissertação de Mestrado. Faculdade de Ciências da Universidade de Lisboa.
- Morgado, L., Gaspar, G. (2003). *Emotion in Intelligent Virtual Agents: The Flow Model of Emotion*. In Rist, T., Aylett, R., Ballin, D., Rickel, J. (Eds.), *Intelligent Virtual Agents*, LNAI 2792, 31-38. Springer-Verlag.
- Morgado, L., Gaspar, G. (2003). *Extending BDI with Emotional Dispositions for Adaptive Social Reasoning*. *Proceedings of the International Workshop on Regulated Agent-Based Social Systems: Theories and Applications (RASTA'03)*.
- Morgado, L., Gaspar, G. (2004). *A Generic Agent Model Allowing a Continuous Characterization of Emotion*. *Proceedings of the 17th European Meeting on Cybernetics and Systems Research (EMCSR 2004)*.
- Morgado, L., Gaspar, G. (2004). *Focusing Reasoning Through Emotional Mechanisms*. *Proceedings of the 16th European Conference on Artificial Intelligence*.
- Morgado, L., Gaspar, G. (2004). *Unifying Emotion and Cognition: The Flow Model Approach*. *Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004)*.

- Morgado, L., Gaspar, G. (2005). *Emotion Based Adaptive Reasoning for Resource Bounded Agents*. Proceedings of the 4th International Joint Conference on Autonomous Agents and Multi-Agent Systems.
- Müller, J. (1996). *The Design of Intelligent Agents: A Layered Approach*. Springer-Verlag.
- Müller, J. (1997). *A cooperation model for autonomous agents*. In Wooldridge, M., Jennings, R. (Eds.), *Intelligent Agents III*, 245-260. Springer.
- Murphy, R. (2000). *Introduction to AI Robotics*. MIT Press.
- Newell, A. (1980). *Physical Symbol Systems*. Cognitive Science, 4(2), 135-183.
- Newell, A. (1982). *The knowledge level*. Artificial Intelligence, 18 (1), 87-127. Elsevier.
- Newell, A. (1983). *Intellectual Issues in the History of Artificial Intelligence*. In Machlup, F., Mansfield, U., (Eds.) *The Study of Information - Interdisciplinary Messages*, 187-227. John Wiley & Sons.
- Newell, A. (1990). *Unified Theories of Cognition*. Harvard University Press.
- Newell, A., Simon, H. (1976). *Computer Science as Empirical Enquiry: Symbols and search*. Communications of the ACM, 19, 105-132.
- Nicolis, G., Prigogine, I. (1977). *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations*. John Wiley & Sons.
- Nilsson, N (1998). *Artificial Intelligence: A New Synthesis*. Morgan Kaufman.
- Norman, T., Long, D. (1994). *Goal Creation in Motivated Agents*. Proceedings of the ECAI Workshop on Agent Theories, Architectures, and Languages.
- Oatley, K., Jenkins, J. (1996). *Understanding Emotions*. Blackwell Publishers.
- Oatley, K., Johnson-Laird, P. (2002). *Emotion and Reasoning to Consistency*. In Moore, S., Oaksford, M. (Eds.), *Emotional cognition: from brain to behaviour*, 157-182. John Benjamins Publishing.
- Oliveira, E., Sarmiento, L. (2003). *Emotional Advantage for Adaptability and Autonomy*. Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multi-Agent Systems.
- Oppenheim, A., Willsky, A., Nawab, S. (1997). *Signals & Systems*. Prentice Hall.
- O'Reilly, R., Munakata, Y. (2000). *Computational exploration in cognitive neuroscience: understanding the mind by simulating the brain*. MIT Press.
- Ortony, A., Clore, G., Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge University Press.
- Paiva, A., Dias, J., Sobral, D., Aylett, R., Sobreperez, P., Woods, S., Zoll, C., Hall, L. (2004). *Caring for Agents and Agents that Care: Building Empathic Relations with Synthetic Agents*. Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi-Agent Systems.
- Palmer, J. (2002). *The living clock: the orchestrator of biological rhythms*. Oxford University Press.
- Panksepp, J. (1998). *Affective neuroscience: the foundations of human and animal emotions*. Oxford University Press.
- Papez, J. (1937). *A proposed mechanism of emotion*. Archives of Neurology and Psychiatry, 38, 725-743.
- Parrott, W. (2001). *Emotions in social psychology: essential readings*. Psychology Press.

- Pfeifer, R. (1993). *Emotions in Robot Design*. Proceedings of the IEEE International Workshop on Robot and Human Communication, 408-413.
- Pfeifer, R., Scheier, C. (1999). *Understanding Intelligence*. MIT Press.
- Picard, R. (1997). *Affective Computing*. MIT Press.
- Pippard, A. (1989). *The Physics of Vibration*. Cambridge University Press.
- Plutchik, P. (1980). *Emotion: A Psychobioevolutionary Synthesis*. Harper & Row.
- Polk, T., Seifert, C. (2002). *Cognitive modeling*. MIT Press.
- Popa, R. (2004). *Between necessity and probability: searching for the definition and origin of life*. Springer.
- Pöppel, E. (1978). *Time Perception*. In Held, R., Leibowitz, H., Teuber, H. (Eds.), *Handbook of Sensory Physiology*, Vol VIII: Perception, 713-729.
- Port, F., Van Gelder, T. (1995). *Mind as Motion: Explorations in the Dynamics of Cognition*. MIT Press.
- Prescott, T., Redgrave, P., Gurney, K. (1999). *Layered Control Architectures in Robots and vertebrates*. *Adaptive Behavior*, 7.
- Pribram, K. (1999). *Quantum Holography: Is it Relevant to Brain Function?*. *Information Sciences*, 115, 97-102. Elsevier.
- Pylyshyn, Z. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. MIT Press.
- Rao, A., Georgeff, M. (1991). *Modeling Rational Agents within a BDI-Architecture*. Proceedings of Knowledge Representation and Reasoning (KR&R-91). Morgan Kaufmann.
- Rao, A., Georgeff, M. (1995). *BDI agents - from theory to practice*. Proceedings of the First International Conference on Multiagent Systems.
- Rist, T., Aylett, R., Ballin, D., Rickel, J. (Eds.) (2003). *Intelligent Virtual Agents, Lecture Notes in Artificial Intelligence 2792*. Springer-Verlag.
- Robinson, R. (2004). *An introduction to dynamical systems: continuous and discrete*. Prentice Hall.
- Rolls, E. (2001). *The brain and emotion*. Oxford University Press.
- Rosa, A. (2002). *Dos sistemas centrados aos sistemas acentrados*. Vega Editora.
- Rosenblatt, F. (1958). *The perceptron: A probabilistic model for information storage and organization in the brain*. *Psychological Review*, 65, 386-408.
- Rosenblueth, A., Wiener, N., Bigelow, J. (1943). *Behavior, purpose and teleology*. *Philosophy of Science*, 10, 18-24.
- Russell, S. (1999). *Rationality and Intelligence*. In Wooldridge M., Rao, A. (Eds.), *Foundations of Rational Agency*. Kluwer Academic Publishers.
- Russell, S., Norvig, P. (2003). *Artificial Intelligence: A Modern Approach*. Prentice Hall.
- Sacerdoti, E. (1977). *A Structure for Plans and Behavior*. Elsevier.
- Sadio, R., Tavares, G., Ventura, R., Custódio, L. (1998). *An Emotion-Based Agent Architecture Application with Real Robots*. Proceedings of the 1998 AAAI Fall Symposium Emotional and Intelligent: The Tangled Knot of Cognition, 117-122.
- Schachter, S., Singer, J. (2001). *Cognitive, Social, and Psychological Determinants of Emotional State*. In Parrott, W. (Ed.), *Emotions in social psychology: essential readings*, 76-93. Psychology Press.

- Scherer, K. (1984). *On the nature and function of emotion: A component process approach*. In Scherer, K., Ekman, P. (Eds.), *Approaches to Emotion*. Lawrence Erlbaum Publishers.
- Scherer, K. (1999). *Appraisal theory*. In T. Dalgleish, & M. Power (Eds.), *Handbook of Cognition and Emotion*, 637-663. Wiley.
- Scherer, K. (2000). *Emotions as Episodes of Subsystem Synchronization Driven by Nonlinear Appraisal Processes*. In Lewis, M., Granic, I. (Eds.), *Emotion, Development, and Self-Organization*, 70-99. Cambridge University Press.
- Scherer, K. (2001). *Construction of the Social World: Emotion*. In Hewstone, M., Stroebe, W. (Eds.) *Introduction to Social Psychology*. Blackwell Publishers.
- Scheutz, M. (2004). *How to Determine the Utility of Emotions*. In Hudlicka, E., Cañamero, L. (Eds.), *Architectures for Modeling Emotion: Cross-Disciplinary Foundations: AAAI Spring Symposium*, 122-127.
- Schommers, W. (1998). *The visible and the invisible: matter and mind in physics*. World Scientific.
- Schulkin, J. (2003). *Rethinking homeostasis: allostatic regulation in physiology and pathophysiology*. Bradford Book.
- Schut, M., Wooldridge, M., Parsons, S. (2004). *The Theory and Practice of Intention Reconsideration*. *Journal of Experimental and Theoretical Artificial Intelligence*, 16:4, 261-293.
- Schwartz, B. (2004). *The tyranny of choice*. *Scientific American*, April-04.
- Schweitzer, F. (2003). *Brownian Agents and Active Particles: Collective Dynamics in the Natural and Social Sciences*. Springer.
- Searle, J. (1999). *Intentionality: an essay in the philosophy of mind*. Cambridge University Press.
- Shanahan, M. (1999). *What Sort of Computation Mediates Best between Perception and Action*. *Logical Foundations for Cognitive Agents: Contributions in Honor of Ray Reiter*, ed. H. Levesque and F. Pirri, 352-369. Springer-Verlag.
- Shepherd, G. (1994). *Neurobiology*. Oxford University Press.
- Sichman, J. (1995). *Du Raisonement Social Chez Les Agents: Une Approche Fondée sur la Théorie de la Dépendence*. Thèse de Doctorat, Institut National Polytechnique de Grenoble - France.
- Sichman, J., Conte, R. (2002). *Multi-Agent Dependence by Dependence Graphs*. *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems*, 483-490.
- Sichman, J., Demazeau, Y. (2001). *On Social Reasoning in Multi-Agent Systems*. *Revista Ibero-Americana de Inteligência Artificial*, 13, 68-84.
- Simon, H. (1955). *A Behavioral Model of Rational Choice*. *The Quarterly Journal of Economics*, 69, 99-118.
- Simon, H. (1967). *Motivational and Emotional Controls of Cognition*. *Psychological Review*, 74, 29-39.
- Simon, H. (1981). *The Sciences of the artificial*. MIT Press.
- Skyttner, L. (2001). *General systems theory: ideas & applications*. World Scientific.
- Sloman, A. (1993). *The Mind as a Control System*. In Hookway, C., Peterson, D. (Eds.), *Philosophy and the Cognitive Sciences*, 69-110. Cambridge University Press.

- Sloman, A. (2001). *Beyond Shallow Models of Emotion*. Cognitive Processing, 1.
- Sloman, A. (2004). *What Are Emotion Theories About?*. In Hudlicka, E., Cañamero, L. (Eds.), *Architectures for Modeling Emotion: Cross-Disciplinary Foundations: AAAI Spring Symposium*, 128-134.
- Sloman, A., and Croucher, M. (1981). *Why robots will have emotions*. Proceedings of the 7th International Joint Conference on AI, 197-202.
- Smith, C. (2004). *A Functional Perspective on Emotion Elicitation: Some Considerations for the Development of Emotional Architectures*. In Hudlicka, E., Cañamero, L. (Eds.), *Architectures for Modeling Emotion: Cross-Disciplinary Foundations: AAAI Spring Symposium*, 135-143.
- Smith, C., Kirby, L. (2001). *Affect and Cognitive Appraisal Processes in Affect and Social Cognition*. In Forgas, J. (Ed.), *Affect and Social Cognition*, 75-92. Lawrence Erlbaum Publishers.
- Smith, C., Lazarus, R. (2001). *Appraisal Components, Core Relational Themes, and the Emotions*. In Parrott, W. (Ed.), *Emotions in social psychology: essential readings*, 94-114. Psychology Press.
- Staddon, J. (1983). *Adaptive Behavior and Learning*. Cambridge University Press.
- Staddon, J. (2001). *Adaptive dynamics: the theoretical analysis of behavior*. MIT Press.
- Staller, A., Petta, P. (1998). *Towards a Tractable Appraisal-Based Architecture for Situated Cognizers*. In Numaoka C., Cañamero D., Petta P. (Eds.), *Grounding Emotions in Adaptive Systems, SAB'98 (5th International Conference of the Society for Adaptive Behavior) Workshop Notes*.
- Staller, A., Petta, P. (2001). *Introducing Emotions into the Computational Study of Social Norms: A First Evaluation*. Journal of Artificial Societies and Social Simulation vol. 4, no. 1, <<http://www.soc.surrey.ac.uk/JASSS/4/1/4.html>>.
- Steels, L. (Ed.) (1995). *The Biology and Technology of Intelligent Autonomous Agents*. Springer.
- Sterling, P., Eyer, J. (1988). *Allostasis: A New Paradigm to Explain Arousal Pathology*. In Fisher, S., Reason, J. (Eds.), *Handbook of Life Stress, Cognition and Health*. John Wiley & Sons.
- Stewart, P. (2001). *Functional Neuroanatomy*. Decker Electronic Publishing.
- Strongman, K. (1996). *The Psychology of Emotion 4th Ed.* (Edição Portuguesa - Climepsi Editores, 1998). John Wiley & Sons.
- Sutton, R., Barto, A. (1998). *Reinforcement Learning*. MIT Press.
- Tinbergen, N. (1951). *The Study of Instinct*. Oxford University Press.
- Tschacher, W. (1999). *Dynamics synergetics autonomous agents: nonlinear systems approaches to cognitive psychology and cognitive science*. World Scientific.
- Turing, A. (1950). *Computing machinery and intelligence*. Mind, 59, 433-460.
- Turner, J. (2003). *Origens das emoções humanas: um inquérito sociológico acerca da evolução da afectividade*. Instituto Piaget.
- Van de Velde, W. (1995). *Cognitive Architectures - From Knowledge Level to Structural Coupling*. in Steels, L. (Ed.), *The Biology and Technology of Intelligent Autonomous Agents*, 197-221. Springer-Verlag.
- Varela, F., Thompson, E., Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. (Edição Portuguesa - Instituto Piaget, 2001). MIT Press.

- Velásquez, J. (1998). *Modeling Emotion-Based Decision-Making*. Proceedings of the 1998 AAAI Fall Symposium Emotional and Intelligent: The Tangled Knot of Cognition, 164-169. AAAI Press.
- Velásquez, J., Fujita, M., Kitano, H. (1998). *An Open Architecture for Emotion and Behavior Control of Autonomous Agents*. Proceedings of the Second International Conference on Autonomous Agents (Agents '98).
- Ventura, R. (2000). *Emotion-based Agents*. MSc. Thesis, Instituto Superior Técnico.
- Ventura, R., Pinto-Ferreira, C. (1998). *Emotion-based agents*. Proceedings of the 1998 AAAI Fall Symposium Emotional and Intelligent: The Tangled Knot of Cognition.
- Ventura, R., Pinto-Ferreira, C. (2001). *Learning Courses of Action using the «movie-in-the-brain» Paradigm*. Proceedings of the 2001 AAAI Fall Symposium Emotional and Intelligent II: The Tangled Knot of Social Cognition.
- Ventura, R., Pinto-Ferreira, C. (2004). *Generating and Refining Causal Models for an Emotion-based Agent*. In Hudlicka, E., Cañamero, L. (Eds.), *Architectures for Modeling Emotion: Cross-Disciplinary Foundations: AAAI Spring Symposium*, 146-149.
- Von Neumann, J. (1951). *The General and Logical Theory of Automata*. In Jeffress, L. (Ed.), *Cerebral Mechanisms in Behavior*, 1-41. John Wiley & Sons.
- Ward, L. (2002). *Dynamical cognitive science*. MIT Press.
- Watkins, C. (1989). *Learning from delayed rewards*. PhD thesis, King's College.
- Webb, B., Consi, T. (2001). *Biorobotics: Methods and Applications*. MIT Press.
- Weiss, G. (2000). *Cognition, Sociability and Constraints*. in *Balancing Reactivity and Social Deliberation in Multi-Agent Systems Workshop* held at the 14th European Conference on Artificial Intelligence, 125-139.
- Weld, D. (1994). *An Introduction to Least Commitment Planning*. AI Magazine, winter 1994, 27-61.
- Werner, E. (1996). *Logical Foundations of Distributed Artificial Intelligence*. In O'Hare, G., Jennings, N. (Eds.), *Foundations of Distributed Artificial Intelligence*. John Wiley & Sons.
- Wiener, N. (1948). *Cybernetics: or Control and Communication in the Animal and the Machine*. MIT Press.
- Williams, J., Watts, F., MacLeod, C., Mathews, A. (1997). *Cognitive Psychology and Emotional Disorders*. John Wiley & Sons.
- Winograd, T., Flores, F. (1991). *Understanding Computers and Cognition*. Addison-Wesley.
- Wooldridge, M. (2000). *Reasoning about rational agents*. MIT Press.
- Wooldridge, M. (2001). *A modern Approach to Distributed Artificial Intelligence*. In Gerhard, W. (Ed.), *Multiagent Systems - A modern Approach to Distributed Artificial Intelligence*.
- Wooldridge, M. (2002). *An Introduction to MultiAgent Systems*. John Wiley & Sons.
- Wooldridge, M., Jennings, N. (1994). *Agent Theories, Architectures, and Languages: A Survey*. Workshop on Agent Theories, Architecture and Languages (ECAI-94), Lecture Notes in Artificial Intelligence. Springer-Verlag.
- Wooldridge, M., Jennings, N. (1995). *Intelligent Agents: theory and practice*. The Knowledge Engineering Review, 10(2), 115-152.
- Wooldridge, M., Rao, A. (1999). *Foundations of rational agency*. Kluwer.
- Wright, I. (1997). *Emotional Agents*. PhD thesis, University of Birmingham.

- Wylie, C., Barret, L. (1982). *Advanced Engineering Mathematics*. McGraw-Hill.
- Yang, Q. (1997). *Intelligent Planning: A Decomposition and Abstraction Based Approach*. Springer-Verlag.
- Zajonc, R. (1980). *Feeling and thinking: preferences need no inferences*. American Psychologist, 35, 151-175.
- Zeeman, E. (1976). *Catastrophe Theory*. Scientific American, 234, 65-83.
- Ziegler, J., Dittrich, P., Banzhaf, W. (1997). *Towards a metabolic robot control system*. Proceedings of the International Workshop on Information Processing in Cells and Tissues (IPCAT'97).
- Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking*. Cambridge University Press.