



UNIVERSIDADE ESTADUAL PAULISTA

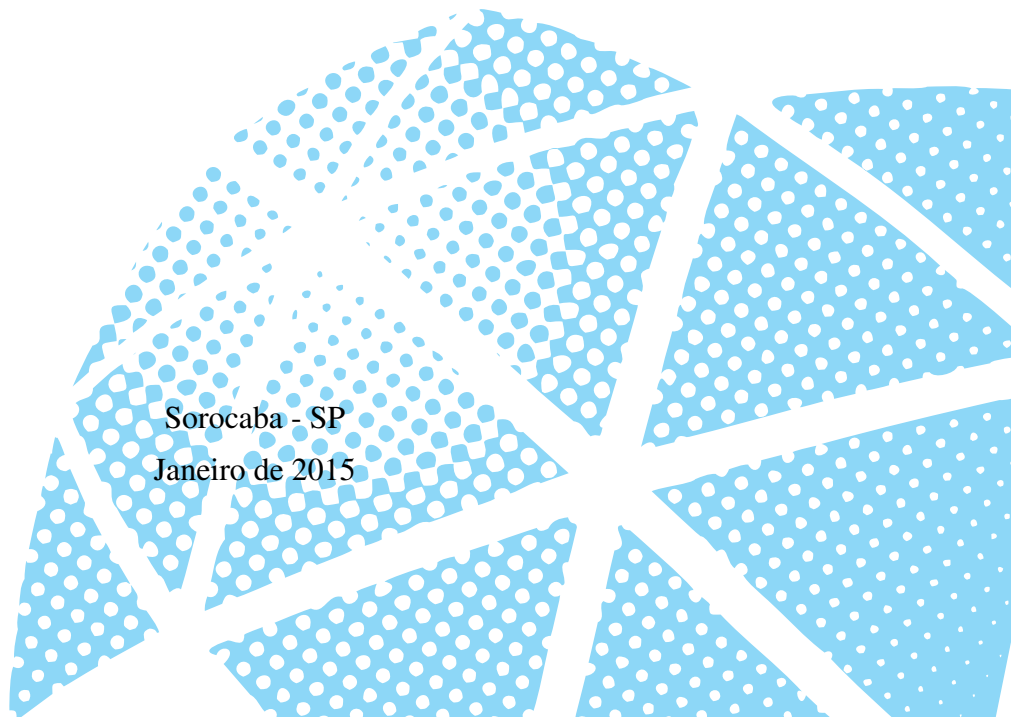
“JÚLIO DE MESQUITA FILHO”

Campus de Sorocaba

ALEXANDRE DA SILVA SIMÕES

**UM MODELO COGNITIVO BASEADO NA ATENÇÃO
PARA CONSCIÊNCIA EM ROBÔS INTELIGENTES**

Sorocaba - SP
Janeiro de 2015

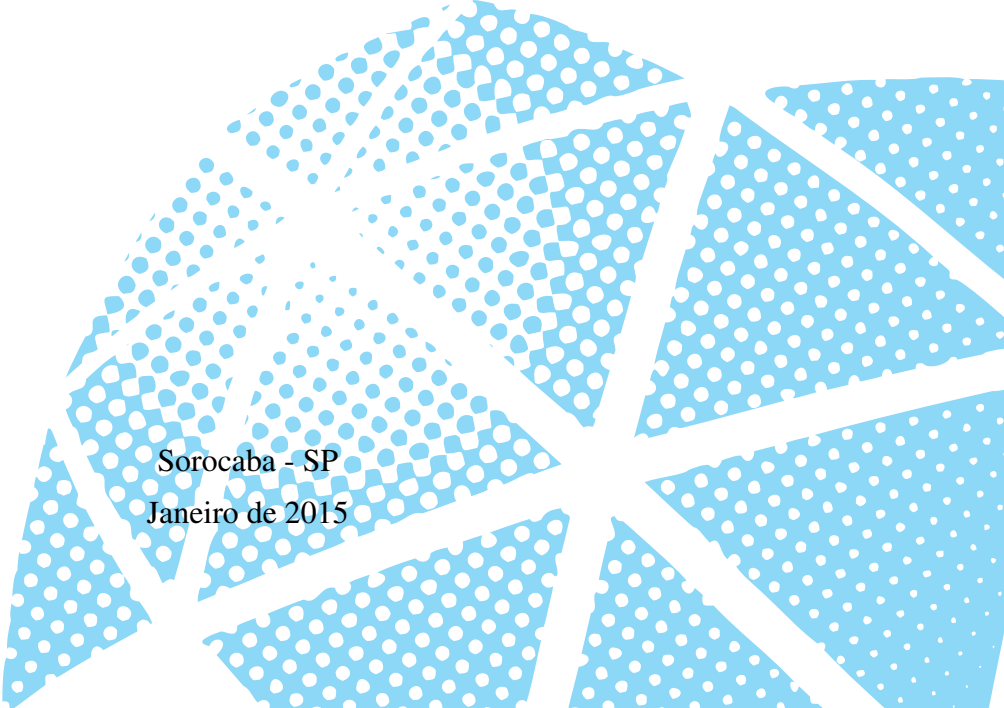


ALEXANDRE DA SILVA SIMÕES

UM MODELO COGNITIVO BASEADO NA ATENÇÃO PARA CONSCIÊNCIA EM ROBÔS INTELIGENTES

Tese apresentada ao Campus de Sorocaba da UNESP como parte dos requisitos para obtenção do título de Livre-Docente em Robótica e Inteligência Artificial.

Sorocaba - SP
Janeiro de 2015



"O infitino é realmente um dos deuses mais lindos".
Renato Russo.

Ficha catalográfica elaborada pela Biblioteca da Unesp - Câmpus de Sorocaba

S452m Simões, Alexandre da Silva.
Um modelo cognitivo baseado na atenção para consciência em robôs inteligentes / Alexandre da Silva Simões, 2015.
150 f.: il.

Tese (Livre-Docência) – Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Ciência e Tecnologia (Câmpus de Sorocaba), 2015.

1. Inteligência artificial. 2. Robótica. 3. Consciência. 4. Atenção. I. Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Ciência e Tecnologia (Câmpus de Sorocaba). II. Título.

CDD 006.3

Banca Examinadora:

Prof. Dr. Jaime Simão Sichman (POLI-USP)

Prof.^a Dr.^a Anna Helena Reali Costa (POLI-USP)

Prof.^a Dr.^a Roseli Aparecida Francelin Romero (USP - São Carlos)

Prof. Dr. Paulo Roberto de Aguiar (UNESP - Bauru)

Prof. Dr. Edvaldo Assunção (UNESP - Ilha Soleira)

Agradecimentos

A todos os companheiros da UNESP, docentes e servidores técnico-administrativos, pela amizade e pela contribuição – de muitas formas – para a realização desse trabalho ou suas partes, direta ou indiretamente, ao longo dos anos. Em particular aos amigos Marcos, Carlos, Roberta, Célia, Vanessa, Bruna, Amanda, André e especialmente aos amigos do Grupo de Automação e Sistemas Integráveis (GASI), sem os quais este momento não teria sido, literalmente, possível.

Aos meus alunos e orientados que contribuíram – cada um de sua forma única e especial – ao longo de mais de uma década, para a lapidação dos conhecimentos e para com a formulação de novas idéias, abordagens, práticas e conceitos. Obrigado por compartilharem a caminhada. Obrigado por cada dúvida, discussão ou questionamento feito ao seu professor.

A todos os meus mestres, presentes, ausentes, pelas diferentes contribuições que me trouxeram até o este dia. Por compartilharem comigo suas visões.

A meus pais, Maria José e Walsey, e minha avó Eliza, que nunca se cansaram de me incentivar, e sem os quais nem uma pequena parte dessa jornada teria sido possível.

À Prof.^a Dr.^a Sandra Luna Cirne de Azevedo pelas discussões filosóficas sobre a questão da consciência, e à Prof.^a Dr.^a Bella Luna Colombini Ishikiriama pelo carinho e injeção de ânimo. Da mesma forma ao Prof. Dr. Antonio Cesar Germano Martins pelas discussões. Ao Prof. Dr. Flavio Alessandro Serrão Gonçalves, Prof. Dr. Eduardo Paciência Godoy e Prof.^a Dr.^a Maria Glória Caño de Andrade pelo apoio das mais diversas formas.

De forma mais do que especial à Prof.^a Dr.^a Esther Luna Colombini, por todo o carinho, dedicação, inspiração e por todas as discussões sobre a natureza da inteligência, cognição e atenção, incorporados de muitas formas a este trabalho e à vida.

A Hermione, companheira fiel na redação deste documento.

Resumo

A consciência humana é um dos mais fascinantes mistérios de nosso tempo. Das antigas civilizações aos pensadores da pós-modernidade, todos já se questionaram sobre como nos tornamos conscientes de nossa existência, e como nos tornamos conscientes do mundo que nos cerca. Embora não exista uma definição precisa para o termo **consciência**, há um consenso de que este fenômeno encontra-se intimamente ligado a processos cognitivos humanos, tais como: pensamento, raciocínio, emoções, desejos. Dentre os processos-chave para o surgimento da consciência está a **atenção**, processo cognitivo capaz de promover a seleção de apenas alguns poucos estímulos dentre a imensa quantidade de informação que nos atinge a todo instante. A **consciência de máquina** é o campo da inteligência artificial que investiga a possibilidade da produção de processos conscientes utilizando dispositivos artificiais. O presente trabalho realiza uma revisão sobre a temática da consciência – natural e artificial –, abordando o tema sob os prismas filosófico e computacional, e investiga a viabilidade da utilização de um esquema atencional como base para um processamento cognitivo. Propõe-se um modelo computacional formal para agentes conscientes que integra: memórias de curto e longo prazo, raciocínio, planejamento, emoção, tomada de decisão, aprendizado, motivação e volição. Simulações computacionais no domínio da robótica móvel realizadas sobre o ambiente USARSim, proposto pela RoboCup, sugerem que o agente pode ser capaz de, utilizando esses elementos, adquirir experiências baseado em estímulos recebidos do ambiente. A adoção da arquitetura cognitiva sobre o modelo atencional tem potencial para permitir a emergência de comportamentos usualmente associados à consciência nos robôs simulados. Futuras implementações sobre esse modelo poderiam, potencialmente, permitir ao agente expressar senciência, auto-ciência, auto-consciência, consciência auto-noética, possessividade e perspectividade. Realizando computação sobre um espaço atencional, o modelo também permite ao agente aprender sobre um espaço de estados muito reduzido.

Palavras-chave: Inteligência artificial. Robótica. Consciência de máquina. Cognição. Atenção. Emoções. Memória. Qualia.

Abstract

Understanding consciousness is one of the most fascinating challenges of our time. From ancient civilizations to modern philosophers, questions have been asked on how one is conscious of his/her own existence and about the world that surrounds him/her. Although there is no precise definition for **consciousness**, there is an agreement that it is strongly related to human cognitive processes such as: thinking, reasoning, emotions, wishes. One of the key processes to the arising of the consciousness is the **attention**, a process capable of promoting a selection of a few stimuli from a huge amount of information that reaches us constantly. **Machine consciousness** is the field of the artificial intelligence that investigate the possibility of the production of conscious processes in artificial devices. This work presents a review about the theme of consciousness - in both natural and artificial aspects - , discussing this theme from the philosophical and computational perspectives, and investigates the feasibility of the adoption of an attentional schema as the base to the cognitive processing. A formal computational model is proposed for conscious agents that integrates: short and long term memories, reasoning, planning, emotion, decision making, learning, motivation and volition. Computer experiments in a mobile robotics domain under USARSim simulation environment, proposed by RoboCup, suggest that the agent can be able to use these elements to acquire experiences based on environment stimuli. The adoption of the cognitive architecture over the attentional model has potential to allow the emergence of behaviours usually associated to the consciousness in the simulated mobile robots. Further implementation under this model could potentially allow the agent to express sentience, selfawareness, self-consciousness, autothetic consciousness, mineness and perspectivalness. By performing computation over an attentional space, the model also allows the agent to learn over a much reduced state space.

Keywords: Artificial intelligence. Robotics. Machine consciousness. Cognition. Attention. Emotions. Memory. Qualia.

SUMÁRIO

Lista de Figuras	9
Lista de Tabelas	11
Lista de Equações	12
Lista de Símbolos	13
I FUNDAMENTOS TEÓRICOS	17
1 INTRODUÇÃO	18
1.1 Objetivos	20
1.2 Por que estudar a consciência artificial?	20
1.3 Organização do trabalho	21
2 FUNDAMENTOS FILOSÓFICOS	22
2.1 A relação corpo-mente nas culturas antigas	22
2.2 As doutrinas filosóficas	24
2.3 Dualismo	25
2.4 Monismo	26
3 FUNDAMENTOS DA ATENÇÃO	29
3.1 O que é atenção?	29
3.2 Taxonomia da Atenção	31
3.3 Conceitos fundamentais	32
4 FUNDAMENTOS DA CONSCIÊNCIA	35
4.1 O que é consciência?	35
4.2 Conceitos fundamentais	38
4.2.1 Memória	40
4.3 A relação entre consciência e atenção	43

4.4	Os problemas da consciência e o qualia	44
4.5	Uma consciência para cada cérebro?	45
4.6	Como avaliar a consciência?	46
4.6.1	Axiomas para consciência	47
4.6.2	ConScale: uma métrica para a avaliação da consciência	48
II MODELOS COMPUTACIONAIS		53
5 MODELOS ATENCIONAIS		54
5.1	Visão geral dos modelos	54
5.2	A arquitetura de Colombini	57
5.2.1	Controlador atencional	61
5.2.2	Modelo formal	62
6 MODELOS PARA CONSCIÊNCIA		66
6.1	Taxonomia dos modelos	66
6.2	Modelos baseados no Espaço de Trabalho Global	67
6.3	Modelos baseados na integração de informação	71
6.4	Modelos baseados no auto-conhecimento	72
6.5	Modelos baseados em representações de nível superior	73
6.6	Modelos baseados em Atenção	74
7 OUTROS MODELOS COGNITIVOS		79
7.1	O ISAC	79
7.2	O EPIROME	81
7.3	O SOAR	82
7.4	O framework de Samsonovich	83
7.5	O diagrama de relações de Stachowicz	84
III TRABALHO PROPOSTO		87
8 REVISITANDO O PROBLEMA DA CONSCIÊNCIA		88
9 O MODELO PROPOSTO		91
9.1	Conceito e hipótese adotados	91
9.2	Apresentação do modelo	92
9.3	Modelo formal	95

9.4	Operação do modelo	99
9.5	Sumário do modelo	102
9.6	Comparação com outros modelos	104
10	MATERIAIS E MÉTODOS	107
10.1	Ambiente de simulação	107
10.2	Sistemática de extração de características	109
10.2.1	Espaço de observação	109
10.2.2	Características <i>bottom-up</i>	109
10.2.3	Características <i>top-down</i>	112
10.3	EXP-01: Robô atencional <i>bottom-up</i>	113
10.4	EXP-02: Robô atencional <i>bottom-up</i> e <i>top-down</i>	114
10.5	EXP-03: Robô atencional consciente	114
11	RESULTADOS E DISCUSSÃO	122
11.1	EXP01	122
11.2	EXP02	122
11.3	EXP03	126
11.4	Uma possível abordagem para o <i>qualia</i> ?	132
12	CONCLUSÃO	134
	REFERÊNCIAS	136
	Índice Remissivo	148

Lista de Figuras

1	Organização das principais doutrinas filosóficas relacionadas à natureza da consciência	25
2	Taxonomia da atenção: estados globais, seletivos e atencionais	31
3	Representação esquemática de diferentes aspectos da atenção: influência da atenção <i>top-down</i> , <i>bottom-up</i> , atenção seletiva e atenção orientada	33
4	Relação entre consciência, inteligência, atentividade e vida	38
5	Relação entre intenção e volição: adaptação do modelo das fases da ação na terapia de reabilitação	40
6	Taxonomia da memória humana	41
7	Habilidades cognitivas do ConScale (diagrama de Hasse)	50
8	Valores possíveis para o resultado quantitativo do ConScale	52
9	Exemplos de perfis cognitivos do ConScale para algumas arquiteturas conscientes	52
10	Diagrama esquemático de funcionamento de diferentes modelos atencionais computacionais	56
11	Diagrama esquemático de funcionamento de diferentes modelos atencionais computacionais	58
12	Diagrama esquemático do modelo atencional de Colombini	59
13	Curvas responsáveis pela modelagem pelo enaltecimento ou inibição temporal e espacial no modelo de Colombini	60
14	Taxonomia dos sistemas conscientes considerando o princípio elementar para o surgimento da consciência com ênfase aos sistemas atencionais	67
15	Modelo pré-consciente IDA	68
16	Arquitetura CERA	70
17	Modelo computacional para máquina consciente de Starzyk	76
18	Taxonomia dos sistemas conscientes baseados em modelos atencionais considerando a natureza da arquitetura implementada	78
19	Modelo do robô ISAC	80
20	Arquitetura do SOAR	83

21	O <i>framework</i> de Samsonovich	85
22	Diagrama de relações de entidades em uma memória episódica segundo Stachowicz	86
23	Arquitetura do modelo proposto para um agente atencional consciente	93
24	Perfil cognitivo estimado do modelo proposto segundo o ConScale	106
25	Arquitetura modular do USARSim	108
26	Imagens da plataforma USARSim	110
27	EXP01: Configuração	115
28	EXP02: Configuração	116
29	EXP03: Configuração	120
30	Exemplo de cenário de trabalho com o modelo proposto	121
31	EXP01: Trajetória do robô no ambiente	123
32	EXP01: Dados coletados nos sonares, sensor laser e mapas de características . . .	124
33	EXP01: Dados coletados no mapa de características combinado, vencedores da competição atencional e mapa atencional	125
34	EXP02: Trajetória do robô no ambiente	126
35	EXP02: Dados coletados nos sonares, sensor laser e mapas de características . . .	127
36	EXP02: Dados coletados no mapa de características combinado, vencedores da competição atencional e mapa atencional	128
37	EXP03: Dinâmica do agente com o modelo consciente proposto	130
38	EXP03: Trajetórias do agente com o modelo proposto explorando um ambiente .	131

Lista de Tabelas

1	ConsScale - Habilidades cognitivas níveis 2 a 11 – 2: Reativo, 3: Adaptativo, 4: Atencional, 5: Executivo, 6: Emocional, 7: Auto-consciente, 8: Categórico, 9: Social, 10: Humano, 11: Supre-consciente. Como em (ARRABALES; LEDEZMA; SANCHIS, 2010b).	50
2	Diferentes relações entre consciência e atenção em vários modelos para agentes conscientes atencionais.	77

Lista de Equações

1	Modelo de Colombini: contribuição no tempo t do vencedor do processo atencional se o disparo teve causa exógena	64
2	Modelo de Colombini: contribuição no tempo t da vizinhança do vencedor do processo atencional se o disparo teve causa exógena	65
3	Modelo de Colombini: contribuição no tempo t do vencedor do processo atencional se o disparo teve causa endógena	65
4	Modelo de Colombini: contribuição no tempo t da vizinhança do vencedor do processo atencional se o disparo teve causa endógena	65
5	Modelo de Colombini: atualização dos elementos do mapa atencional	65
6	Extração de características: Nível de discrepância temporário sobre a característica de velocidade ($f_{1_{n_t}}$)	111
7	Extração de características: Nível de discrepância sobre a característica de velocidade ($f_{1_{n_t}}$)	111
8	Extração de características: Nível de discrepância temporário sobre a característica de direção ($tf_{1_{n_t}}$)	112
9	Extração de características: Nível de discrepância sobre a característica de direção ($tf_{1_{n_t}}$)	112
10	Extração de características: Disposição dos elementos em relação à característica de distância ($f_{3_{n_t}}$)	112
11	Extração de características: Ações atencionais <i>top-down</i> de distância	113
12	Extração de características: Ações atencionais <i>top-down</i> de região-objetivo	113

Lista de Símbolos

Conjuntos, vetores, coleções e funções:

α	é uma função de tomada de decisão
A	é um conjunto de ações do agente
A_A	é um conjunto de ações atencionais do agente
A_M	é um conjunto de ações motoras do agente
\mathcal{A}	é o conjunto de atuadores do agente
\mathcal{B}	é um conjunto de elementos vencedores (<i>winners</i>)
B	é o conjunto de características descritoras da estrutura física do agente
BS	é um conjunto denominado esquema corporal (<i>body schema</i>) do agente
C	é um mapa de características combinadas
C_{SA}	é um conjunto de pares estado-ação
D	é uma função de tomada de decisão (<i>decision making</i>) do agente
EM	é uma função emocional do agente
EV	é uma função avaliação (<i>evaluation</i>) do agente
\mathcal{F}_d	é um conjunto de mapas de característica
\mathcal{G}	é um conjunto de objetivos (<i>goals</i>) do modelo atencional
G	é um conjunto de objetivos (<i>goals</i>) do agente cognitivo
\mathcal{H}	é a função degrau de Heaviside
\mathcal{H}	é uma hipótese
\mathcal{L}	é o mapa de saliências
L_E	é um aprendizado episódico
L_P	é um aprendizado procedimental
L_S	é um aprendizado semântico
L	é um sistema de aprendizado
Λ	é uma função de geração de ação que decompõe uma macro-ação em ações motora e atencional
\mathcal{M}	é o mapa atencional
M	é uma memória de longo prazo
M_E	é uma memória episódica
M_P	é uma memória procedimental

M_S	é uma memória semântica
M_W	é uma memória de trabalho (<i>working memory</i>)
M_{ST}	é uma memória de curto prazo (<i>short term memory</i>)
M_{LT}	é uma memória de longo prazo (<i>long term memory</i>)
MV	é o conjunto de motivações do agente
MO	é uma função de motivação do agente
\mathcal{O}	é um conjunto de observações de sensores
P	é um processo cognitivo do agente
PL	é uma função de planejamento
PLT	é uma lista de tarefas de planejamento
Φ	é o valor atribuído à quantidade de consciência em um sistema segundo a teoria da integração da informação
Φ_{BU}	é um conjunto de funções de dimensão de característica <i>bottom-up</i>
Φ_{TD}	é um conjunto de funções de dimensão de característica <i>top-down</i>
Ψ	é a função de uma onda em mecânica quântica
Q	é a medida da qualidade da experiência para o agente
\mathfrak{R}	é o conjunto dos números reais
S	é um conjunto de estados (<i>states</i>) do agente
S_A	é um conjunto denominado estado atencional do agente
S_E	é um conjunto denominado estado emocional do agente
S_{EV}	é um conjunto denominado estado de desempenho (<i>evaluation</i>) do agente
S_P	é um conjunto denominado estado proprioceptivo do agente
\mathcal{S}	é um conjunto de sensores
Sd	é um conjunto de dimensões de sensores
T	é um conjunto de tarefas do agente
$\forall \mathcal{O}$	é uma função de volição do agente
\mathcal{W}_f	é um vetor de pesos de característica
$\zeta_{\Phi_{TD}}$	é uma função de mapeamento de dimensão <i>top-down</i>
$\zeta_{\mathcal{W}_f}$	é uma função de mapeamento de pesos

Constantes e elementos de conjuntos:

\mathcal{A}	é uma ação do agente
\mathcal{A}_M	é uma ação motora do agente
\mathcal{A}_A	é uma ação atencional do agente
b_i	é um elemento da lista de vencedores no tempo t com índice i indicando a natureza do estímulo (endógeno ou exógeno)
c_n	é um elemento do mapa de características combinado
δ_{edp}	é a constante de tempo de atraso no decaimento endógeno

δ_{ep}	é a excitação endógena pré-ativação
δ_p	é a constante de tempo de atraso pré-ativação
δ_r	é a constante de tempo de atraso pré-refração
\mathcal{D}_{ME}	é um elemento presente em um item de memória episódica
\mathcal{D}_{MS}	é um elemento presente em um item de memória semântica
f_{2d_n}	é um elemento do mapa de características
\mathcal{G}	é um objetivo do agente cognitivo
g_t	é um elemento do conjunto de objetivos (<i>goals</i>) do agente atencional
i	é o índice de posição de um disparo no conjunto de disparos
l_n	é um elemento do mapa de saliências
m_n	é um elemento do mapa atencional
m	é um elemento da memória do agente
m_E	é um elemento da memória episódica do agente
m_S	é um elemento da memória semântica do agente
m_P	é um elemento da memória procedimental do agente
m_W	é um elemento da memória de trabalho do agente
\mathcal{O}_{s_t}	é a observação da j -ésima dimensão do s -ésimo sensor no tempo t
\mathcal{P}	é um processo cognitivo do agente
ϕ_{BU_d}	é uma função de mapeamento de características <i>bottom-up</i>
ϕ_{TD_d}	é uma função de mapeamento de características <i>top-down</i>
π_G	é uma política utilizada para organização dos objetivos do agente
m_V	é uma motivação do agente
s	é um elemento do estado do agente
s_A	é um estado atencional do agente
s_E	é um estado emocional do agente
s_{EV}	é um estado de desempenho do agente
s_P	é um estado proprioceptivo do agente
s_{ns}	é um elemento do conjunto de sensores
sd_{ns}	é um elemento do conjunto de dimensões de sensores
$\sigma(\cdot)$	é o desvio padrão do argumento
\mathcal{T}	é uma tarefa do agente
τ_{aa}	é a constante de ativação exponencial ascendente
τ_{ad}	é a constante de ativação exponencial descendente
τ_{eaa}	é a constante de ativação exponencial endógena ascendente
τ_{ead}	é a constante de ativação exponencial endógena decrescente
τ_{ra}	é a constante de refração exponencial ascendente
τ_{rd}	é a constante de refração exponencial descendente
v	é uma variável do agente
v_E	é uma variável emocional do agente

-
- v_{EV} é uma variável de desempenho do agente
 v_{MV} é um valor de motivação do agente
 v_P é uma variável proprioceptiva do agente
 wf_{2d} é um elemento do vetor de pesos de características

PARTE I

FUNDAMENTOS TEÓRICOS

1 INTRODUÇÃO

“Qualquer tecnologia suficientemente avançada é indistinguível de magia.”

Arthur Charles Clarke

Ouvir os trechos mais sublimes do *Requiem* de Mozart em um quarto escuro. Apreciar o pôr do Sol em sua casa nas montanhas em um dia de verão. Saborear a água fresca corrente que brota do chão após uma longa caminhada em um dia quente. Essas são algumas das experiências mais especiais vivenciadas por nós, seres humanos, durante nossas vidas. A apreciação de experiências como essas, que chamaremos por hora de “*experiências subjetivas*” do ser, está profundamente ligada a nossos pensamentos, memórias, desejos e reflexões. Ela envolve passado, presente e futuro.

Mas, como exatamente ocorre em nós este processo? Como nos tornamos conscientes de nossa existência e do mundo que nos cerca? Como adquirimos a capacidade de apreciar estas experiências? Séculos depois do “*penso, logo existo*” de Descartes, ainda temos mais perguntas do que respostas para as questões que dizem respeito à consciência humana. Alguns negam a existência da consciência. Outros dizem que ela é, na verdade, a única coisa no universo da qual pode-se afirmar a existência, e que a reflexão sobre ela seria a única questão que faria sentido em toda a filosofia. Outros argumentam que a consciência existe, mas que, possuindo inteligência limitada, não seríamos capazes de compreender este fenômeno. Alguns dizem que a consciência é simplesmente uma manifestação dos impulsos elétricos dos neurônios presentes nos circuitos cerebrais. Outros afirmam ainda que a consciência está indissociavelmente ligada ao universo, matéria e energia.

O fato é que não conseguimos ainda, sequer, definir de forma cabal o que vem a ser a consciência. Nesse contexto de tamanha complexidade, nos inspirar nas reflexões filosóficas e nos mecanismos da cognição humana selecionados pela evolução ao longo dos milhões de anos que nos separam dos organismos unicelulares são alguns dos primeiros passos na – longa – jornada no sentido de compreender esse fenômeno da consciência.

O questionamento sobre a possibilidade da compreensão da consciência humana suscita ainda uma outra questão igualmente complexa: seria possível, ao menos em tese, construir máquinas conscientes? A resposta, evidentemente, é igualmente complexa. Vejamos. O vôo dos pássaros inspirou os homens por séculos até a construção de máquinas capazes de voar. A máquina humana para voar, contudo, é substancialmente diferente dos pássaros. Ela não utiliza músculos ou penas. Uma vez compreendidos os fundamentos envolvidos no vôo, a mente humana foi capaz de aplicar este conhecimento para produzir a sua própria versão da máquina de voar. Essa máquina supera o vôo dos pássaros em vários aspectos: é muito mais veloz do que o mais veloz dos pássaros, é capaz de transportar mais pessoas e carga do que o maior dos pássaros, dentre muitas outras vantagens. A máquina humana de voar também possui severas desvantagens em relação aos pássaros: ela é muito mais pesada do que o maior dos pássaros, utiliza muito mais energia para voar do que o maior dos pássaros e requer muito mais cuidados do que gostaríamos. Os pássaros e a máquina humana seguem dois paradigmas diferentes de vôo. Contudo, ambos são, efetivamente, capazes de voar.

Podemos esperar que a consciência se comporte de forma similar? Podemos esperar que existirá uma máquina que está para a consciência humana assim como a máquina humana de voar está para os pássaros? Em outros termos, podemos entender que o estudo dos mecanismos neurais e das funções cerebrais que constituem um indivíduo levarão à elaboração de uma teoria que, em última instância, permitirá a criação, a replicação ou a simulação da consciência – ou de um conceito similar a ela – em máquinas? Também esta teoria, evidentemente, dependeria do domínio de diversos fenômenos associados à cognição humana – tais como a formulação de pensamentos, raciocínio, memória, etc. –, da captura de seus aspectos ou fundamentos essenciais, e, então, de sua transferência, de alguma forma, para o domínio das máquinas. David Elis chamaria essa nova entidade “*machina sapiens*”.

Alan Turing afirmou que a humanidade tende a utilizar com máquinas o argumento “*as máquinas nunca poderão fazer X*”, com *X* já tendo assumido diversos valores ao longo dos séculos, tais como: voar, nadar, jogar xadrez, ir à lua, etc. Sob essa ótica, o surgimento da consciência em máquinas seria simplesmente uma questão de tempo e trabalho. Outros autores, como Ray Kurzweil, acreditam que o surgimento de máquinas conscientes é uma consequência natural do processo evolutivo, e que os limites que separam homem e máquina tenderão a desaparecer nos próximos séculos. Este processo de fusão homem-máquina já teria começado, uma vez que já utilizamos cotidianamente máquinas associadas ao nosso corpo para ampliar nossa capacidade de andar, enxergar, nos localizar, auxiliar nosso batimento cardíaco, substituir alguns de nossos membros, e, em última instância, pensar.

E, caso esse cenário venha a ser possível, o computador será o dispositivo que dará o suporte a essa máquina com consciência? Em caso afirmativo, o computador com a arquitetura de von Neumann, seguindo o paradigma de computação corrente, seria a máquina adequada? Se sim,

quais seriam as arquiteturas, modelos e formas de trabalho propícias para tratar esta máquina consciente? A consciência poderia ser implementada em *software*?

Não temos, hoje, respostas para a grande maioria destas perguntas. Nesse contexto, o presente trabalho busca reunir diferentes visões sobre a questão da consciência – natural e artificial – construídas ao longo dos séculos e elaboradas por um grande número de autores sob diferentes paradigmas, de forma que possam fornecer subsídios à proposição de caminhos para a investigação da assim denominada **consciência de máquina**, ou ainda, como preferem alguns autores, consciência artificial.

consciência
de
máquina

Se, como afirma Arthur Clarke, qualquer tecnologia suficientemente avançada é indistinguível de magia, então avançar na compreensão da tecnologia que pode levar à existência da consciência em máquinas é, em última instância, tocar um pouco da magia que existe nas experiências mais sublimes ocultas na alma humana.

1.1 Objetivos

O presente trabalho tem por objetivo promover o estudo da área denominada consciência de máquina, bem como investigar, propor e validar novos modelos cognitivos biologicamente inspirados para a consciência, especialmente aqueles relacionados com os mecanismos atencionais, foco de investigação do grupo de trabalho do autor.

1.2 Por que estudar a consciência artificial?

Há pelo menos duas óticas sobre as quais pode-se responder a esta questão. Sob uma primeira ótica, o estudo da consciência em máquinas passa pela investigação da consciência humana, um dos mistérios mais fundamentais de nossa espécie, que confunde-se com o mistério da própria vida. Avanços no entendimento deste conceito fundamental poderiam ter um profundo impacto na forma de vida de nossa sociedade, com desdobramentos em aspectos religiosos, morais, sociais, éticos, dentre muitos outros. Sob uma outra ótica, a eventual possibilidade de dotar máquinas de consciência tornaria possível a existência de uma série de novos produtos, técnicas e tecnologias hoje inexistentes, com elevado potencial para geração de novas empresas, empregos e renda.

Desta forma, o estudo da consciência artificial é justificado pelo potencial avanço que ela poderia representar no entendimento dos processos mentais humanos – notadamente através do uso de computadores na simulação de processos cognitivos –, e também pelo potencial impacto social e econômico que essas máquinas – e seus derivados – teriam em nossa sociedade.

1.3 Organização do trabalho

O presente trabalho encontra-se organizado da seguinte forma:

- **Parte I:** Reúne uma coleção de fundamentos teóricos associados à temática da consciência reunidos nas seguintes seções:
 - **Seção 1:** Introduce a consciência como tema de trabalho;
 - **Seção 2:** Apresenta fundamentos e doutrinas filosóficas relacionadas com a questão da consciência;
 - **Seção 3:** Concentra conceitos fundamentais relacionados ao mecanismo atencional, essencial para a compreensão da consciência;
 - **Seção 4:** Introduce conceitos e problemas do âmbito da consciência;
- **Parte II:** Reúne os modelos computacionais mais relevantes para o trabalho proposto. Encontra-se dividido nas seguintes seções:
 - **Seção 5:** Apresenta modelos computacionais para a atenção em agentes inteligentes;
 - **Seção 6:** Traça um panorama das arquiteturas computacionais para consciência;
 - **Seção 7:** Discorre sobre outros modelos cognitivos correlatos ao proposto no âmbito deste trabalho;
- **Parte III:** Apresenta os principais aspectos do trabalho realizado. Encontra-se dividido nas seguintes seções:
 - **Seção 8:** Revisita o problema da consciência à luz da revisão teórica realizada no âmbito deste trabalho;
 - **Seção 9:** Detalha o modelo computacional proposto para o trabalho com agentes conscientes;
 - **Seção 10:** Discorre sobre materiais e métodos utilizados nas implementações computacionais realizadas;
 - **Seção 11:** Exibe e discute os resultados obtidos nos experimentos realizados;
 - **Seção 12:** Conclui o presente trabalho.

Para facilitar a localização de conceitos e itens dispostos ao longo do texto, encontra-se disponível um índice remissivo, localizado na [página 148](#).

2 FUNDAMENTOS FILOSÓFICOS

"Os ossos e carne, e outras parte similares nossas, foram feitas como segue. O primeiro princípio de todos então foi que houve a geração da medula. Os laços da vida que unem a alma com o corpo são feitos lá, e eles são a raiz e o fundamento da raça humana".

Platão, em Tiemaus.

O presente capítulo busca apresentar um histórico do pensamento humano sobre a relação corpo-mente do ponto de vista filosófico, destacando escolas e doutrinas envolvidas. Ele encontra-se organizado da seguinte forma:

- [Seção 2.1](#): Apresenta uma visão das antigas civilizações sobre a relação corpo-mente;
- [Seção 2.2](#): Traz uma visão geral das doutrinas filosóficas que tratam deste tema;
- [Seção 2.3](#): Detalha a visão dualista do universo;
- [Seção 2.4](#): Discorre sobre a visão monista do universo.

2.1 A relação corpo-mente nas culturas antigas

Segundo [Castro e Landeira-Fernandez \(2011\)](#), desde os primórdios da humanidade, em diversas civilizações – tais como Egito, Mesopotâmia, Índia e China – observa-se a construção de diversas perspectivas acerca da relação corpo-mente, contemporâneas entre si, e que refletem uma preocupação fundamental em compreender como ocorre a relação entre nossos corpos e fenômenos mentais. Contudo, a partir da civilização grega – berço do pensamento ocidental – surgem observações mais sistemáticas sobre o tema.

É através das obras de Homero – poeta a quem são atribuídas as obras *Ilíada* e *Odisséia* e que teria vivido por volta do século VIII ou IX AC – que se tem informações sobre os hábitos e práticas dos gregos antigos. Embora os principais órgãos do corpo humano fossem conhecidos

dos gregos antigos, eles tinham pouco conhecimento funcional desses órgãos, exceto os conhecimentos mais relacionados com os ferimentos em batalha. O guerreiro do período homérico tinha a preocupação de atacar as partes adequadas do inimigo. No mais, o corpo era visto como um aglomerado de membros, e diversos sentimentos eram associados a alguns órgãos específicos. Em particular, era no coração que os sentimentos se manifestariam, não havendo distinção entre processo físico e psíquico (CASTRO; LANDEIRA-FERNANDEZ, 2011).

De acordo com Snell (2001) inexistia a separação plenamente formulada entre mente (ou alma) e corpo nas obras da época, o que, segundo o autor, refletiria a forma de pensar do grego homérico. O termo *thymos* era usado algumas vezes para definir o estado de ânimo de um sujeito, suas emoções e ímpeto, mas também poderia ser compreendido como um conceito de vida ou respiração (REALE, 2002). A *phyren*, localizada no diafragma, estaria ligada ao intelecto e ao pensamento, mas segundo alguns autores também poderia ser tomada como mente, juízo ou julgamento. Outro termo presente nas obras de Homero é *noos*, também localizado no peito, que referia-se à razão, discernimento e inteligência. Por fim, *psyché* referia-se ao *sopro da alma* e estaria associada também à respiração. No momento da morte, ela abandonaria o corpo pela boca ou através de feridas (CASTRO; LANDEIRA-FERNANDEZ, 2011).

elemento
primor-
dial

cefalocentrismo

Com o surgimento da filosofia na Grécia antiga, os primeiros filósofos buscaram uma explicação causal ligada a fenômenos da natureza para a questão do pensamento humano, postulando a existência de um **elemento primordial** como ponto de partida. Tales de Mileto (623 ou 624 AC – 546 ou 548 AC), considerado o primeiro filósofo, postulou que a *água* seria esse elemento primordial. Anaxímenes de Mileto – segunda metade do século VI AC – e Diógenes de Apolônia (499 AC – 428 AC) consideraram que a fonte do pensamento humano seria o *ar*. Empédocles de Agrigento (495 – 435 AC) considerou que a natureza era composta por quatro elementos: *água, fogo, terra e ar*, e dois princípios cosmológicos: *amor* e *ódio*. O filósofo pré-Socrático Alcmaeon de Crotona (século V AC) foi o primeiro a apontar o cérebro como sede da razão e das sensações (**cefalocentrismo**), e portanto, da mente como produto do cérebro, feito que para muitos autores é comparável com as propostas de Copérnico e Darwin em termos de revolução no conhecimento humano (DOTY, 2007). O filósofo grego Theophrastus (371 AC – 287 AC), discípulo de Aristóteles, retrataria da seguinte forma o trabalho de Alcmaeon:

"Para aqueles que não explicam a percepção por similaridade, Alcmeon primeiro determina a diferença entre homem e animal: ele diz que o homem difere de outro animal porque eles entendem sozinhos, ao passo que outros percebem, mas não entendem. Ele supõe que pensamento e percepção são distintos, não como Empédocles, sobre este assunto. Então ele discute cada um dos sentidos. Ele diz que nós ouvimos com nossos ouvidos porque existe um espaço vazio dentro deles que ecoa: a cavidade soa e o ar ecoa em retorno. Nós cheiramos com nossos narizes da mesma forma que respiramos, desenhando nossa respiração com o cérebro. Nós

discriminamos sabores com nossas línguas; Sendo quente e macias, elas dissolvem coisas com o seu calor. Os olhos vêem por meio da água que os rodeia".

Theophrastus em "Sobre os Sentidos".

Stratton (1911)

A Hipócrates (460 AC – 370 AC), considerado o pai da medicina, é atribuída a autoria da obra **Corpus hippocraticum**, uma coleção de textos médicos de grande relevância, que apontava o cérebro como centro de todas as atividades intelectuais e cognitivas, julgamento e emoções, bem como de algumas doenças como convulsões e espasmos. Contudo, os humores estariam fundamentalmente ligados a órgãos como o fígado, baço, coração e cérebro (CASTRO; LANDEIRA-FERNANDEZ, 2011).

Aristóteles (384 – 322 AC) propõe que os seres animados se diferenciariam dos seres inanimados por possuírem um princípio que lhes dá a vida, um princípio vital denominado **alma** (VALDUGA, 2007). Segundo ele, a alma possuiria três faculdades: a faculdade nutritiva, a faculdade sensitiva e a faculdade intelectual. Os vegetais teriam a *faculdade nutritiva*, o princípio mais básico e elementar da vida, responsável pelas funções biológicas como nutrição, crescimento e geração. A *faculdade sensitiva* comportaria os cinco sentidos humanos (visão, audição, olfato, tato, paladar). Os animais irracionais seriam, também, dotados da alma sensitiva, que é responsável pelas sensações e daria a seus corpos a faculdade do movimento. Já a *faculdade intelectual* (o **intelecto**) seria uma exclusividade humana. Todos os homens desejariam, por natureza, conhecer. Aristóteles então, caracteriza o intelecto como “*aquela parte da alma que permite conhecer e pensar*”.

2.2 As doutrinas filosóficas

Com a solidificação do entendimento do cérebro como centro da mente, e a natural evolução dos conceitos de matéria (ou corpo) e mente (ou alma) – e, portanto, da consciência –, surgiram ao longo dos séculos as doutrinas filosóficas que abordam essa relação. Elas podem ser agrupadas de acordo com o número de categorias, reinos ou substâncias que, em suas visões, compõem o universo. Sob essa ótica, há três doutrinas principais: *i*) o monismo, que defende a existência de uma única categoria ou substância no universo; *ii*) o dualismo, que parte do princípio que existem duas categorias distintas de substâncias no universo, e *iii*) o pluralismo, que propõe a existência de múltiplas categorias. Uma taxonomia das diferentes visões filosóficas e seus derivados é apresentada na [figura 1](#). Uma visão geral delas e sua relação com o problema da consciência pode ser encontrada em [Reggia \(2013\)](#). As duas vertentes filosóficas de maior relevância histórica (dualismo e monismo) serão discutidas nas seções a seguir.

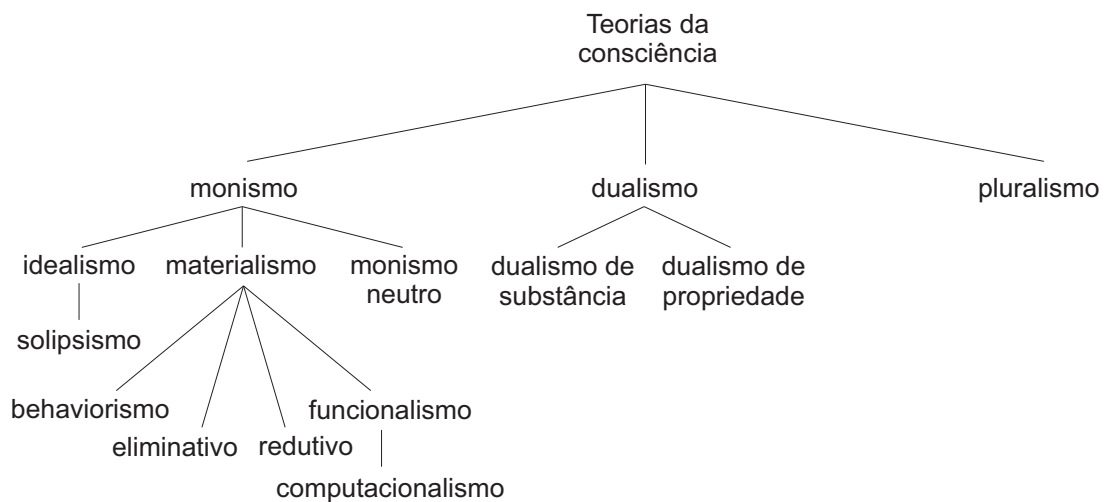


Figura 1 - Organização das principais doutrinas filosóficas relacionadas à natureza da consciência. Baseado em (REGGIA, 2013) e (SEARLE, 2004a).

2.3 Dualismo

dualismo

A doutrina **dualista** parte do princípio de que existem duas categorias ontologicamente distintas de fenômenos no universo: uma física e outra não-física. De acordo com esta corrente, nenhuma dessas duas categorias pode, de qualquer forma, ser reduzida à outra. Algumas das reflexões nesse sentido remetem ao filósofo e matemático grego Platão (429–347 AC), que teve influência do *Corpus hippocraticum*. Ele acreditava que a alma existia previamente ao corpo em uma forma incorpórea, e que pertencia a um “reino de formas”, fora do alcance do espaço e do tempo. A alma imortal, em seu estado desencarnado, antes de sua encarnação em um corpo (para a **teoria da transmigração das almas** ver (PLATÃO, 2011)), teria mantido contato com essas formas e seria então delas recordada (SILVERMAN, 2012). A parte imortal da alma seria o *logos*, que se referia ao intelecto, e estaria abrigada no cérebro. A alma teria ainda duas outras partes mortais: uma localizada no coração, responsável pelos sentimentos, e outra na região próxima ao fígado, sede dos desejos. A posição cefalocentrista de Platão foi disseminada através de sua obra *Timeu* para o mundo Árabe e para a Europa medieval e renascentista (CASTRO; LANDEIRA-FERNANDEZ, 2011). O conceito dos elementos fundamentais, contudo, permanecia presente na literatura da época.

teoria da transmigração das almas

cogito ergo sum

Em 1637, o filósofo francês René Descartes (1596–1650) enuncia o "*cogito ergo sum*" (penso, logo existo) (DESCARTES, 2001). Descartes concluiu que, ao colocar em dúvida a afirmação de sua existência, então algo ou alguém deveria estar formulando a dúvida. Portanto, o próprio fato de a dúvida existir provaria a existência de seu pensamento, e a sua própria. Segundo ele, a mente pensante imaterial e indivisível (*res cogitans*) seria capaz de interagir com um corpo mecânico e divisível do mundo material (*res extensa*). O ponto de interação entre corpo e mente, segundo Descartes, seria a glândula pineal, pequena glândula endócrina

que se localiza no centro do cérebro, entre os dois hemisférios. Corpo e mente seriam, portanto, duas substâncias essencialmente distintas. A visão defendida por Descartes ficaria conhecida como **dualismo de substância** ou dualismo cartesiano, e é aceita até os dias atuais por diversas correntes religiosas.

Uma outra variação do dualismo é o **dualismo de propriedade**, que afirma que pode não haver uma distinção de substância entre mente e corpo, mas que estas poderiam ser propriedades distintas de uma mesma substância, mas ainda categoricamente distintas e não redutíveis uma à outra. Embora nem todos os pesquisadores refutem o dualismo (ECCLES, 1994), sua aceitação junto à comunidade científica tem sido muito baixa em função de seus aspectos não-físicos, e a consequente impossibilidade de sua investigação segundo métodos científicos.

2.4 Monismo

O **monismo** é a doutrina filosófica que defende que tudo o que conhecemos decorre da existência de uma única substância no universo. Sob uma perspectiva monista, o **idealismo** (KANT, 2003) é a vertente que defende que apenas o pensamento existe, e que a matéria, tal como a conhecemos é, na verdade, uma mera ilusão, uma manifestação da consciência. Uma das correntes derivadas do idealismo é o **solipsismo**, que argumenta que a única coisa da qual se assegura a existência é da própria mente. Segundo o proposto pelo pensador grego Gorgias de Leontini (483–375 AC), em seu trabalho "*Sobre a natureza ou o não-existente*", do qual não existem cópias e cujo teor é conhecido apenas através de citações por outros autores, há três premissas fundamentais: *i*) nada existe; *ii*) Ainda que algo exista, nada pode se conhecer sobre isso, e *iii*) Ainda que algo pudesse ser conhecido sobre isso, o conhecimento sobre isso não poderia ser comunicado aos outros. O solipsismo é atribuído a Max Stirner, pseudônimo do filósofo alemão Johann Kaspar Schmidt (1806–1856). Sob um outro prisma, o **monismo neutro** James (1904) é a corrente que defende que a natureza intrínseca de tudo o que existe não é física nem mental: ambas, matéria e mente, seriam manifestações de um mesmo elemento distinto de ambos.

Em uma linha que diverge das correntes anteriores, o **materialismo**, termo cunhado em 1674 pelo filósofo irlandês Robert Boyle (1627–1691), é a corrente que sustenta que a única coisa da qual se pode afirmar a existência é a matéria. Desta forma, tudo seria composto de matéria e todos os fenômenos seriam resultado de interações materiais. Mesmo a consciência teria uma explicação científica que em última análise ainda será descoberta. O materialismo é, portanto, uma doutrina filosófica propícia ao estudo da consciência.

Um dos primeiros defensores da teoria materialista foi o filósofo inglês Thomas Hobbes (1588–1679), conhecido por seus trabalhos no campo da filosofia política. Para Hobbes, todas as experiências humanas poderiam ser explicadas através de processos biológicos presentes no

corpo:

"O mundo – não quero dizer apenas a terra (...) mas também todo o universo, isto é, toda a massa de todas as coisas que existem – é corpóreo, isto é, corpo, e tem as dimensões de grandeza, a saber, comprimento, largura e profundidade; também qualquer parte do corpo é igualmente corpo e tem as mesmas dimensões, e conseqüentemente qualquer parte do universo é corpo e aquilo que não é corpo não é parte do universo. E porque o universo é tudo, o que não é parte dele não é nada e não está em nenhum lugar".

Hobbes (2003, p. 599)

Com o passar dos anos, a filosofia materialista foi abordada sob diferentes perspectivas. O **fisicalismo** é a corrente materialista segundo a qual a linguagem da física deverá ser a linguagem universal de toda a ciência. Por sua vez, o fisicalismo pode ser abordado segundo duas óticas principais: o **fisicalismo reduutivo** afirma que todo o universo – processos orgânicos e inorgânicos – pode ser explicado pelas leis da natureza e da física. Por outro lado, o **fisicalismo não-reduutivo** tem como ideia central que mesmo derivando de estados físicos, os estados mentais teria propriedades que não seriam totalmente explicáveis pelas propriedades físicas fundamentais.

O **behaviorismo** (SKINNER, 1992) – ou *comportamentalismo* – substitui a discussão da existência da mente e dos estados mentais por uma análise baseada apenas no comportamento das entidades. O **materialismo eliminativo** argumenta que os conceitos mentais tais como os formulamos não são abordados atualmente por uma teoria adequada, e que, portanto, nossa perspectiva de senso comum sobre esse tema está fundamentalmente errada por fazer uso destes conceitos (CHURCHLAND, 2004). O **materialismo reduutivo** defende que estados mentais podem ser reduzidos a – e são efetivamente o mesmo que – estados físicos no cérebro (CHURCHLAND, 2004) (MASLIN, 2001). A mente é, portanto, equivalente ao estado eletroquímico das redes neurais biológicas (REGGIA, 2013)

Uma das teorias de maior relevância nesse cenário é o **funcionalismo**, que sustenta que a mente resulta de matéria organizada de uma forma particular. Em outros termos: os estados mentais corresponderiam a estados do cérebro devido ao funcionamento do cérebro, e não devido a sua estrutura física ou composição. Um conceito bastante próximo é o **computacionalismo**, que é a hipótese de que as relações funcionais entre entradas e saídas mentais e estados internos são computáveis. Searle (2004b) descreve o argumento computacionalismo da seguinte forma: o cérebro é um computador e a mente é um programa de computador. Do ponto de vista da possibilidade da criação de máquinas conscientes, as óticas funcionalista ou computacionalista não são impeditivas, por exemplo, à possibilidade de que uma arquitetura baseada em silício

– ou outro substrato físico qualquer – possa dar suporte ao surgimento de mentes ou consciência em entidades totalmente artificiais se esta arquitetura for capaz de apresentar as mesmas funcionalidades essenciais do cérebro:

"Se a organização é realizada em circuitos integrados, na população da China ou em latas de cerveja e bolas de pingue-pongue não importa. Contanto que a organização funcional esteja correta, a experiência consciente será determinada".

Chalmers (1996)

Esta visão também é conhecida entre os pesquisadores da área de Inteligência Artificial como **IA forte**, e defende que é possível a criação de máquinas inteligentes ou auto-conscientes baseadas em computador que sejam capazes de raciocinar e resolver problemas. Esta visão **IA fraca** contrasta com a visão da **IA fraca**, que apresenta o computador apenas como uma ferramenta útil para fazer simulações da mente sem necessariamente ser capaz de gerar um ser consciente, da mesma forma como pode ser usado para fazer simulações meteorológicas, sem, contudo, produzir frio.

3 FUNDAMENTOS DA ATENÇÃO

“Abra os olhos, ouvidos, mente e consciência para a verdade que está a sua frente. Você pode vê-la, mas só se você prestar atenção.”

Kenneth G. Ortiz

A atenção desempenha um papel de grande relevância na cognição humana. De fato, atenção e consciência são dois fenômenos próximos, com limites tênues e freqüentemente confundidos. Para o entendimento do mecanismo da consciência, faz-se necessário o entendimento do mecanismo atencional e seus principais fundamentos. O presente capítulo apresenta uma visão geral da atenção, e encontra-se dividido da seguinte forma:

- [Seção 3.1](#): Discute o conceito de atenção;
- [Seção 3.2](#): Apresenta as principais categorias da atenção;
- [Seção 3.3](#): Explicita alguns dos conceitos fundamentais da atenção.

3.1 O que é atenção?

Os seres humanos são atingidos a todo instante por uma grande quantidade de estímulos (visuais, tácteis, olfativos, etc.), não apenas de origem externa ao nosso corpo (estímulos provenientes da observação de outros seres ou do mundo), mas também de origem interna (fome, sede, cansaço, dor, etc.). Embora seja uma máquina notável, nosso cérebro é capaz de processar apenas uma fração de todos esses estímulos recebidos. O processo cognitivo que realiza a seleção desses estímulos – promovendo alguns deles em detrimento de outros – é o que chamamos **atenção**. A atenção pode ser entendida, portanto, como o ato ou estado de atender especialmente através da aplicação da mente a um objeto do sentido ou pensamento, ou ainda como uma condição de prontidão que seletivamente estreita ou foca a consciência e a receptividade dos estímulos (COLOMBINI, 2014). Uma outra definição mais próxima à neurociência é apresentada por Cohen et al. (2012):

“Atenção refere-se ao mecanismo cognitivo que permite que certas informações sejam processadas mais minuciosamente no córtex do que informações não-selecionadas. Adicionalmente, é provável que a atenção permita à informação ser mais totalmente transmitida entre as regiões corticais que a informação não-atendida.”

Cohen et al. (2012, p. 411)

É possível identificar diversas características particulares na atenção. Algumas delas são (JAMES, 1890): *i)* A possibilidade de um indivíduo exercer um controle voluntário da atenção; *ii)* A incapacidade de um indivíduo atender a diferentes estímulos simultaneamente, e *iii)* A capacidade limitada do processo atencional. Essas são apenas algumas das primeiras características identificadas dentre as várias que podem ser citadas sobre esse mecanismo. Assim como no estudo de outros processos cognitivos, o estudo da atenção também pode ser feito sob diferentes perspectivas: cognitiva, neural, sistema nervoso, celular, genético, dentre outros. Nenhuma delas é ainda capaz de, isoladamente, descrevê-la de uma forma plena (POSNER, 2004), o que torna necessário o conhecimento de diversas dessas abordagens para a compreensão desse fenômeno.

A atenção não tem uma única função e não existe um único centro da atenção no cérebro humano (GHAZANFAR; SCHROEDER, 2006). De fato, estudos com imagens cerebrais e dados neurofisiológicos corroboram a ideia de que a atenção não é nem uma propriedade de uma área do cérebro, nem do cérebro inteiro, mas um processo realizado por uma rede de áreas anatómicas (CORBETTA; SHULMAN, 2002). Os sinais da atenção gerados em diferentes regiões do cérebro influenciam fortemente o processamento motor, associativo e serial do indivíduo, e, em última instância, o comportamento humano. Esses resultados representam um substancial progresso no esforço para determinar como a atividade cerebral é regulada através da atenção (POSNER; DEHAENE, 1994).

Diversos autores entendem que a atenção, percepção e aprendizado são funções fortemente inter-relacionadas. Se atenção e percepção existem, por exemplo, então pode ocorrer aprendizado. Quando o indivíduo está atento, também pode ocorrer uma **recordação** (*recall*) voluntária. É possível estimular o desenvolvimento de habilidades para realizar algumas tarefas específicas utilizando um tipo específico de atenção: a atenção sustentada. Desta forma, a atenção é fundamental para muitas atividades cruciais executadas pelo cérebro, tais como: percepção, recordação voluntária e desenvolvimento de habilidades (PARASURAMAN, 1998).

Embora alguns autores tenham teorizado no passado que não haveria uma teoria específica sobre atenção (MORAY, 1969) ou que não existiriam mecanismos de atenção (NEISSER, 1976), a teoria mais aceita atualmente é a de que a atenção é um conjunto finito de processos cerebrais que interage com outros processos cerebrais para realizar tarefas motoras, cognitivas e da percepção (PARASURAMAN, 1998). Além disso, novos métodos e técnicas desenvolvidas

na área de neurociência cognitiva forneceram meios de detalhar e testar hipóteses atencionais (LI et al., 2007), baseando as conclusões em evidências empíricas.

3.2 Taxonomia da Atenção

Há muitas formas diferentes para classificar a atenção. A figura 2 apresenta a estrutura do sistema atencional adaptado por Colombini (2014) a partir de Gazzaniga, Ivry e Mangun (2002) e Lima (2005). Em um primeiro nível são apresentados os estados globais necessários para que a atenção exista. A atenção somente pode ser aplicada se o agente estiver em um estado *desperto*. Contudo, não é necessário apenas estar acordado, mas existe a necessidade de um estado *atentivo*, isto é, um estado geral sensitivo gerado no fluxo cerebral que prepara o sistema para receber estímulos (LENT, 2002). Uma vez que o estado de ativididade é alcançado, alguém pode atender ou ignorar um estímulo. Finalmente, de acordo com o tempo e natureza da tarefa, um dos componentes atencionais pode ser aplicado.

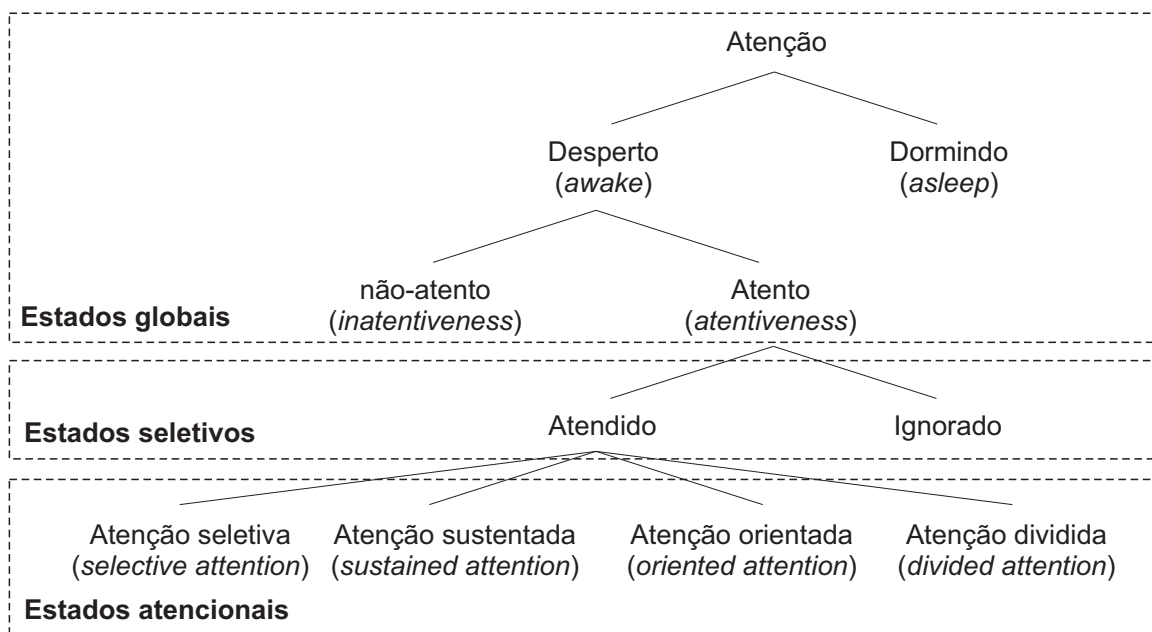


Figura 2 - Taxonomia da atenção: estados globais, seletivos e atencionais. Adaptado de (COLOMBINI, 2014).

Algumas das possíveis classificações para a atenção segundo diferentes aspectos são (COLOMBINI, 2014):

- Considerando seus componentes: seletiva (*selective*), sustentada (*sustained*), orientada (*oriented*) e dividida (*divided*);
- Considerando as funções do processo atencional: percepção seletiva e cognição seletiva;
- Considerando o *status* da atenção: evidente (*overt*) ou encoberta (*covert*);

- Considerando a natureza do processo de acordo com o estímulo: *top-down* ou *bottom-up*.

Esses conceitos serão detalhados na próxima seção.

3.3 Conceitos fundamentais

exógeno Um processo atencional pode ser disparado basicamente de duas formas distintas. A **atenção exógena**, extrínseca, automática, reflexiva ou simplesmente **bottom-up** refere-se a um processo atencional iniciado involuntariamente (POSNER, 1980). Os estímulos são apresentados de forma inesperada, e qualquer tipo de estímulo (intensidade, contraste, movimento, etc.) pode despertar a atenção desta forma. Por outro lado, o processo denominado **atenção endógena** ou simplesmente **top-down** refere-se a um processo atencional voluntário e controlado iniciado por processos cognitivos ou psicológicos (JONIDES, 1981), havendo neste caso uma mobilização consciente da atenção. Uma representação esquemática da influência da atenção *top-down* e *bottom-up* no processo atencional é apresentada na figura 3a.

atenção seletiva A **atenção seletiva** representa a capacidade do mecanismo atencional de escolher um estímulo em detrimento a outro. É o mecanismo básico que suporta a atenção agindo como o componente que permite ao cérebro dar uma resposta coerente na presença de muitos distratores e estímulos que estejam competindo. Conforme apresentado na figura 3b, a atenção seletiva recebe um conjunto de estímulos que são então pesados de acordo com sua relevância. O valor mais alto do estímulo ganha a atenção. Este é provavelmente o componente mais estudado – e compreendido – do mecanismo atencional (DUNCAN, 1984) (TIPPER, 1985) (LABERGE; BROWN, 1989) (KANWISHER; DRIVER, 1992) (POSNER; DEHAENE, 1994) (CAVE; PASHLER, 1995).

atenção orientada A **atenção orientada** ou *atenção alternada* tem a habilidade de alterar o foco atencional deixando um estímulo por outro. Está fundamentalmente ligado à habilidade de coordenar o curso de tarefas simultâneas. Usualmente esta função de controle da atenção está ligada a uma central executiva (de hierarquia mais elevada) que gerencia as atividades do processo de informação no cérebro (ALLPORT; ANTONIS; REYNOLDS, 1972). A atenção orientada é apresentada esquematicamente na figura 3c.

seleção tardia Uma outra distinção importante na atenção é entre seleção antecipada e seleção tardia (JOHNSON; PROCTOR, 2004). Na **seleção tardia** (*late selection*), toda a informação atinge o estágio de codificação semântica e é igualmente processada pelo sistema de percepção. Já na **seleção antecipada** (*early selection*), o estímulo não requer processamento completo e processamento semântico para ser aceito ou rejeitado. Neste caso a atenção atua mais proximamente a um filtro em um estágio inicial.

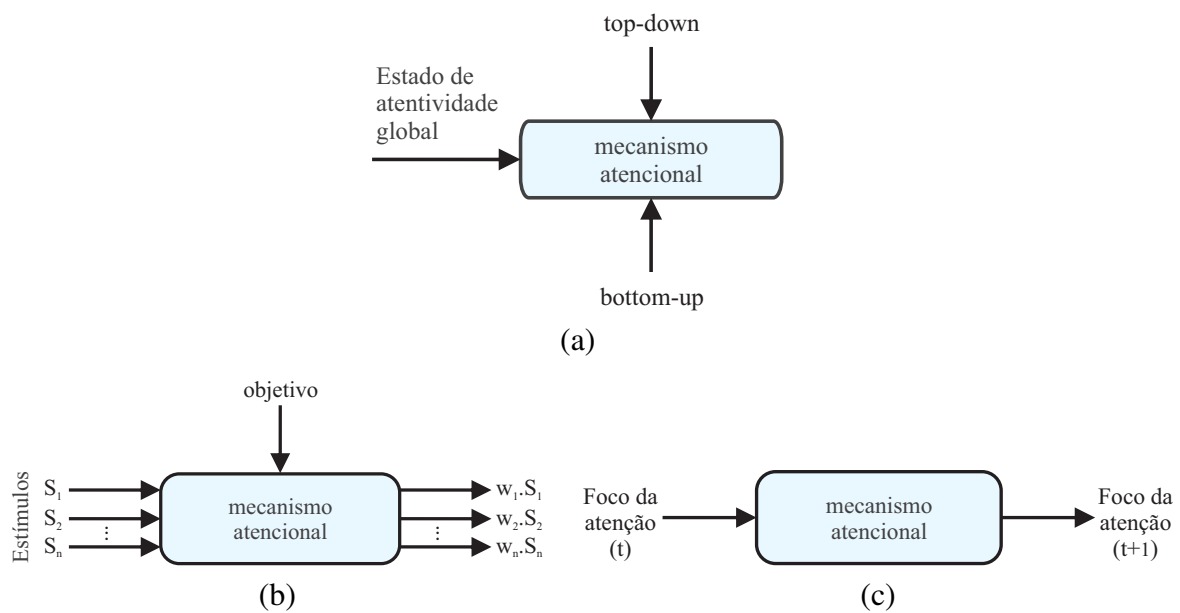


Figura 3 - Representação esquemática de diferentes aspectos da atenção: a) Influência da atenção *top-down* e *bottom-up* sobre o mecanismo atencional; b) Influência da atenção seletiva sobre o mecanismo atencional; c) Influência da atenção orientada sobre o mecanismo atencional. Adaptadas de (COLOMBINI, 2014).

atenção
susten-
tada

vigilância

A **atenção sustentada**, também conhecida como estado de **vigilância** (*vigilance*), é o sub-sistema da atenção que permite a manutenção do objetivo sobre o tempo através da concentração do foco diretamente em um estímulo específico de forma a completar uma tarefa planejada (COLOMBINI, 2014). Quando utilizada de forma eventual um de seus efeitos conhecidos é o aumento da percepção em um ponto (NARAIN, 2006). Contudo, alguns estudos apontam que a manutenção da atenção em um mesmo ponto por longos períodos de tempo pode prejudicar a atenção (LING; CARRASCO, 2006). Sua aplicação típica ocorre da seguinte forma: *i*) a percepção torna-se mais sensível ao estímulo; *ii*) a percepção é direcionada a este estímulo, e *iii*) ocorre a liberação deste ponto do foco da atenção (DEGANI; PORGES, 1990). Após certo período, para que a atenção seja mantida no mesmo ponto serão necessárias tipicamente taxas de disparo neurais muito mais elevadas.

atenção
dividida

A **atenção dividida** é o nível mais elevado da atenção. Refere-se à capacidade de responder a estímulos de múltiplas tarefas ou atender demandas de múltiplas tarefas. A atenção dividida apenas pode operar duas tarefas simultaneamente se uma delas for mediada por processos automáticos enquanto a outra é mediada por processos cognitivos (EYSENCK, 2004).

atenção
evidente

A atenção é ainda denominada **atenção evidente** (*overt attention*) tipicamente quando os olhos do indivíduo são movidos na direção do estímulo atendido: ocorre uma alteração no foco da atenção acompanhada da ação física. Contudo, é possível que essas duas ações ocorram de

atenção
enco-
berta

forma dissociada (POSNER, 1980) (WRIGHT; WARD, 2008). Na **atenção encoberta** (*covert*

attention) ocorre o ato de alguém, mentalmente, considerar a movimentação do foco dos olhos sem, contudo, movê-los. Trata-se de uma ação mental que, mesmo não sendo uma ação física, tem impactos no processo e no mecanismo atencional.

seleção
para per-
cepção

Por fim, consideremos agora a natureza dos estímulos do processo atencional. A **seleção para percepção**, ou *percepção seletiva*, refere-se aos processos que direcionam a atenção para o tratamento dos estímulos externos do ambiente. Por outro lado, a **seleção para ação**, ou algumas vezes denominada *cognição seletiva*, refere-se ao processo atencional que envolve estímulos internos, bem como processos mentais do indivíduo (motivação, memória, etc.) (COLOMBINI, 2014).

seleção
para
ação

4 FUNDAMENTOS DA CONSCIÊNCIA

"Que algo tão extraordinário como um estado de consciência surja em consequência de irradiação do tecido nervoso é tão inexplicável quanto o aparecimento do gênio quando Aladim esfregou a lâmpada".

*Thomas Huxley
citado em [Pinker \(2001, p. 144\)](#).*

O presente capítulo busca apresentar alguns dos principais conceitos, problemas e questões relacionados aos diferentes aspectos da consciência. Este capítulo encontra-se dividido da seguinte forma:

- [Seção 4.1](#): Discute as definições para consciência;
- [Seção 4.2](#): Expõe alguns dos conceitos fundamentais relacionados à consciência;
- [Seção 4.3](#): Apresenta a relação entre consciência e atenção;
- [Seção 4.4](#): Introduce os problemas fundamentais relacionados com a consciência, especificamente o conceito do *qualia*;
- [Seção 4.5](#): Discute a questão da indivisibilidade da consciência;
- [Seção 4.6](#): Apresenta os principais conceitos referentes à avaliação da consciência, destacando a metodologia denominada ConScale.

4.1 O que é consciência?

Antes de iniciar a discussão sobre a – possibilidade da – consciência em máquinas, parece relevante apresentar uma definição formal do que vem a ser consciência. Contudo, esta não é uma tarefa simples. Não há um consenso sobre a definição de consciência, e nem mesmo um consenso sobre a existência de uma definição para ela. Para Descartes, por exemplo, a

consciência seria um elemento teórico primitivo. Em outras palavras, a consciência não poderia ser explicada, possivelmente por ser aquilo que é pressuposto na explicação do que quer que seja. As definições de consciência variam conforme a linha filosófica do autor, sua área de pesquisa principal (psicologia, engenharia, neurociência, computação, etc.), dentre outros. Com o intuito de propor algumas definições introdutórias de cunho mais geral para a consciência, apresenta-se a seguir algumas das proposições feitas por autores ligados à área:

“Consciência é todo o que experienciamos.”

Tononi (2004, p. 2)

“Consciência é a habilidade humana de atribuir sentimento a capacidades mentais especializadas (tais como pensamento, emoções e fala).”

Gazzaniga, Ivry e Mangun (2002)

“Consciência é o processo que permite que informações relevantes permaneçam on-line tempo suficiente de forma que elas sejam sincronamente processadas por múltiplas redes corticais.”

Cohen et al. (2012, p. 412)

“Consciência se refere àqueles estados de sensibilidade e ciência que começam normalmente quando acordamos de um sono sem sonho e continua até que durmamos novamente, caímos em coma, morramos ou fiquemos “inconscientes”. ”

Searle (2004b)

“Nós simplesmente definimos a consciência como a experiência subjetiva que alguém tem quando acordado, seguindo sugestões anteriores de que uma definição mais precisa, dada nosso conhecimento científico inadequado da consciência neste momento, é melhor deixado para o futuro.”

Reggia (2013, p. 112)

“A consciência resulta da interação complexa da associação de áreas multimodais, ao menos, dos córtex parietotemporal e límbico no cérebro do animal, e refere-se à capacidade de experimentar a si mesmo como sendo um sujeito no passado, presente e futuro, incluindo a reflexão em si mesmo como um ser que está atento ao

ambiente que o cerca. Eu digo capacidade porque um ser consciente não precisa estar engajado em uma experiência a todo momento de sua vida. Aqui, a consciência está associada com termos como: experiência, subjetividade, reflexão sobre a consciência”

Arp (2007, p. 102)

Embora seja possível perceber que as várias definições compartilham uma filosofia comum, as diferentes interpretações trazem à discussão – algumas vezes de forma indireta – questões da maior relevância. Algumas dessas definições apresentam a consciência, por exemplo, como um atributo intrinsecamente ligado ao ser humano, excluindo desta forma a possibilidade de que outros seres vivos, tais como animais ou mesmo plantas, possam ter algum tipo de consciência, mesmo que primitiva. Outros designam consciência como um estado que poderia, em tese, ser partilhado por outros seres vivos. Segundo algumas das definições, nem mesmo todos os seres humanos seriam ditos conscientes. Alguns, por fim, acreditam que não temos – seja neste momento ou seja permanentemente – condições de compreender o que é a consciência.

De uma forma geral, contudo, podemos entender a consciência como fundamentalmente ligada às experiências subjetivas do ser, e intrinsecamente ligada a outros elementos cognitivos, tais como: pensamentos, emoções, intenção. [Greenfield \(2000\)](#) introduz ainda o conceito de **contínuo de consciência** (*continuum of consciousness*), argumentando que a consciência é um processo dinâmico, e que muda com o desenvolvimento do cérebro.

contínuo
de cons-
ciência

consciência
de
máquina

Já em se tratando da **consciência de máquina**, uma definição que parece capturar os elementos essenciais é:

“Consciência de máquina poderia ser considerada como o campo da Inteligência Artificial especificamente relacionado com a produção de processos conscientes em dispositivos de engenharia (hardware ou software). ”

Moreno, Espino e Miguel (2007)

Uma definição mais detalhada, e que via de regra será tomada como base no decorrer deste trabalho, é apresentada por [Starzyk e Prasad \(2011\)](#):

“Uma máquina é consciente se apesar dos mecanismos requeridos de percepção, ação, aprendizado, e memória associativa, ela tem uma central executiva que controla todos os processos (conscientes ou subconscientes) da máquina; A central executiva é conduzida pela motivação da máquina e seleção de objetivos, chaveamento da atenção, memória episódica e semântica e usa percepção cognitiva e entendimento cognitivo das motivações, pensamentos ou planos para controlar aprendizado, atenção, motivação e monitorar ações.”

Starzyk e Prasad (2011, p. 264)

Uma definição conjunta para os conceitos de consciência natural e artificial é proposta por Nisargadatta (1973), e encontra-se representada na figura 4. O autor propõe uma relação da consciência com três conceitos distintos: inteligência, vida e atentividade. Nessa visão, o termo consciência refere-se ao conhecimento do “eu”. A atentividade – que poderíamos definir neste momento como a capacidade de estar atento ao mundo que o cerca – seria um conceito mais abrangente, um superconjunto da consciência, e um pré-requisito para sua existência. Por outro lado, um sistema que não é inteligente não poderia ser consciente: uma planta, por exemplo, poderia estar atenta a algo, mas não possuiria as estruturas necessárias para a inteligência, e, portanto, não estaria consciente do que ocorre ao seu redor. Por fim, um sistema consciente não precisaria necessariamente estar vivo, e, portanto, a consciência artificial poderia ser implementada em uma máquina (STARZYK; PRASAD, 2011).

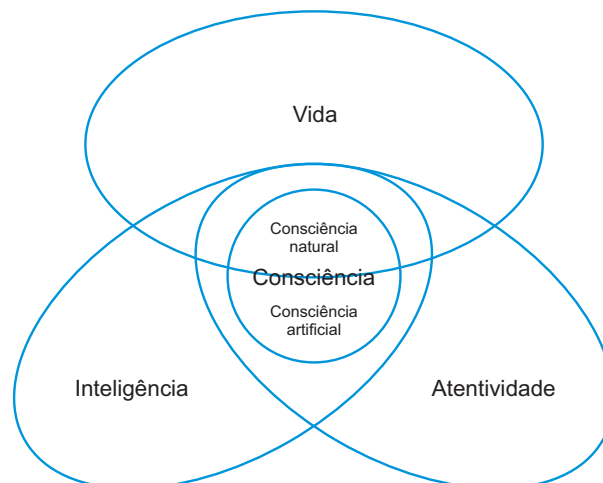


Figura 4 - Relação entre consciência, inteligência, atentividade e vida. Adaptado de (STARZYK; PRASAD, 2011).

4.2 Conceitos fundamentais

Para a discussão da consciência que será feita nos capítulos a seguir é necessário apresentarmos as definições para alguns termos. Muitos deles, contudo, possuem conceitos muito próximos, pouco precisos, ou muitas vezes conflitantes dependendo-se do autor adotado. Um problema adicional na definição de alguns termos é causado pela língua: diversos termos com significado distinto em língua inglesa (tais como *sentience*, *awareness* e *consciousness*) são muitas vezes traduzidos para a língua portuguesa como “consciência” de uma forma genérica.

Adotaremos no âmbito deste trabalho algumas definições baseadas em Arp (2007). Segundo ele, a **senciência** (*sentience*) é a consciência no senso genérico de tratar-se simplesmente

de uma criatura senciente¹, capaz de sentir e responder ao seu mundo (GULICK, 2014) (ARMSTRONG, 1981). A **Atentividade** (*awareness*) refere-se ao processamento que ocorre como um resultado da interação do sistema nervoso do animal (inclusive aparato sensorio) e seu ambiente, de forma que este processamento resulta em uma habilidade básica do animal de reagir a estímulos recebidos do ambiente. A atentividade, portanto, é associada com termos como senciência, percepção, cognição. A atentividade não precisa estar necessariamente acompanhada de consciência. Uma minhoca, por exemplo, pode ser atenta sem ser consciente. Já um ser humano pode ser atento e consciente (ARP, 2007). A **auto-ciência** (*self-awareness*) é relacionada com o ter consciência de si (SEARLE, 2004b), isto é, a capacidade para introspecção e a habilidade de reconhecer a si mesmo como um indivíduo separado do ambiente e de outros indivíduos. Já a **Auto-consciência** (*self-consciousness*) é um termo de definição mais complexa. Alguns autores entendem a auto-consciência como sinônimo de auto-ciência ou de consciência no seu sentido funcional minimalista (STARZYK; PRASAD, 2011). Contudo, é possível identificar na literatura pelo menos quatro significados distintos para auto-consciência (ARP, 2007):

- (i) Consciência de suas próprias sensações corporais e de seus próprios movimentos (presente em diversos animais);
- (ii) Habilidade de reconhecer a si mesmo em um espelho (habilidade demonstrada em chimpanzés a partir de 4,5 anos de idade, em seres humanos tipicamente entre os 15 e 24 meses, e também em golfinhos);
- (iii) Consciência de alguma coisa (suponha que você veja alguém roubando a carteira de outra pessoa. Uma coisa seria perceber o que está acontecendo e pensar sobre o ladrão, a vítima e o ato. Outra é estar consciente de que você está vendo isto e pensar sobre o que pode ser esperado de você naquela situação, o que você acredita que deveria fazer, e se você estaria em risco se outro ladrão perceber que você é uma possível testemunha do crime);
- (iv) Consciência de uma decisão (quando alguém está consciente de uma intenção de agir antes de iniciar uma ação, ele pode não somente antecipar consequências, mas confrontar diferentes possíveis cursos de ação).

O conceito de **noética** refere-se ao pensamento e à razão. A **consciência auto-noética** (*autonoetic consciousness*) é definida como a habilidade humana de mentalmente colocar-se no passado e no futuro, utilizando sua capacidade de raciocínio, anteceder situações e analisar seus próprios pensamentos. Outro conceito comum na literatura da área de consciência é a **volição** (*volition*), vontade ou intencionalidade, é o processo cognitivo pelo qual um indivíduo se decide a praticar uma ação em particular. Segundo Broonen et al. (2011), volição é a atividade mental através da qual as **intenções** são implementadas. Segundo o autor, quando alguém tem

¹Senciência: é a capacidade de sofrer ou sentir prazer ou felicidade;

a intenção de agir, mas não põe as intenções em prática, tem-se uma lacuna entre intenção e comportamento. A motivação não seria suficiente: a volição é necessária. Uma representação gráfica de um modelo de relação entre motivação e volição é apresentado na [figura 5](#).

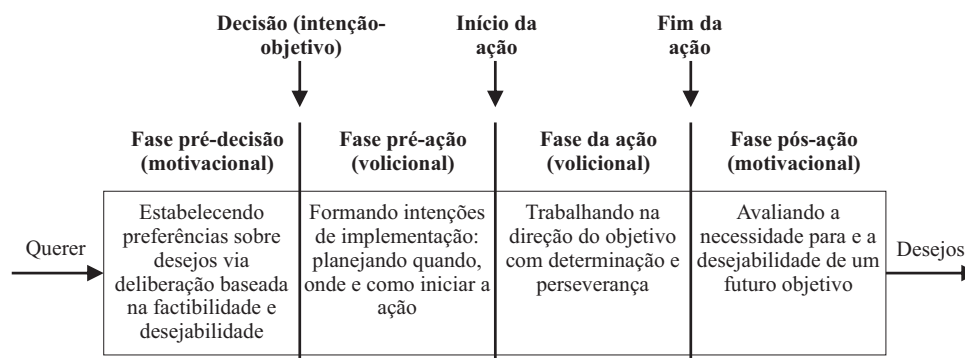


Figura 5 - Relação entre intenção e volição. Adaptação do modelo das fases da ação na terapia de reabilitação. As caixas descrevem as tarefas específicas associadas com cada uma das quatro fases. Adaptado de (BROONEN et al., 2011).

princípio da diferenciação

Por fim, dois conceitos também frequentes no universo da consciência são o **princípio da diferenciação**, que considera que a informação – na forma apropriada para o uso em sistemas cognitivos – é gerada quando existe diferenciação entre estímulos no mecanismo de percepção, e **princípio da integração**, segundo o qual experiências seriam geradas através da integração de diferentes informações.

princípio da integração

Segundo Metzinger (2000), o esboço de uma teoria que aborde a realidade sob a perspectiva da primeira pessoa seria o primeiro passo para a construção de agentes conscientes. O trabalho destaca três propriedades consideradas de relevância central para o tema:

possessividade

- **Possessividade:** (*mineness*), isto é, o senso de pertinência, a apropriação fenomenológica de partes do corpo;

individualidade

- **Individualidade:** (*selfhood*), ou seja, a experiência consciente de ser alguém;

perspectividade

- **perspectividade:** (*perspectivalness*), experiência de nossos estados mentais que requerem a adoção de um ponto de vista.

4.2.1 Memória

memória

Freqüentemente os processos cognitivos associados à consciência relacionam-se com a **memória**. O conceito que entendemos de forma genérica como memória, na verdade compreende um conjunto de tipos distintos de memória presentes em diferentes regiões cerebrais

recordar e com diferentes sistemáticas de funcionamento. Ela provê meios para armazenar experiên-
memória declarativa cias, integrando-as com conceitos mais abstratos, e é capaz de **recordar** ou lembrar (*recall*)
em situações específicas. Um primeiro conceito associado às memórias diz respeito à forma
de armazenamento de suas informações. As **memórias declarativas** podem ser descritas como
as memórias que são conscientemente acessadas. São facilmente verbalizadas e seu conteúdo é
descrito na linguagem corrente, e em geral assumem valor *verdadeiro* ou *falso* (STACHOWICZ;
memória não-declarativa KRUIJFF, 2012). Já as memórias **memórias não declarativas** não caracterizam-se pela repre-
sentação da informação como *verdadeiro* ou *falso*. Nelas existe um compromisso com a disposi-
ção da informação, e o acesso das informações é usualmente expresso por meio de desempenho
em vez de recordação (SQUIRE, 2004) (STACHOWICZ; KRUIJFF, 2012). As memórias po-
memória retrospectiva dem ainda ser classificadas como **memórias retrospectivas**, aquelas relacionadas aos eventos e
experiências passadas, e as memórias prospectivas, ou seja, aquelas ligadas a ações futuras, isto
é, aquela ligada a pensamentos ou necessidades do indivíduo no futuro (NEVID, 2009). Uma
memória prospectiva outra distinção usual na literatura é com relação ao tempo de armazenamento na memória. As
memórias de curto prazo (*short-term memories - STM*) são as que possuem a capacidade de
armazenar uma pequena quantidade de informação na mente de uma forma ativa, em um estado
prontamente disponível por um curto espaço de tempo. Já as **memórias de longo prazo** (*long-*
memória de curto prazo *term memories - LTM*) representam um estágio final de armazenamento no qual a informação
é armazenada por longos períodos de tempo (tipicamente indefinidamente). Uma taxonomia
memória de longo prazo usualmente aceita para a memória humana é apresentada na [figura 6](#).

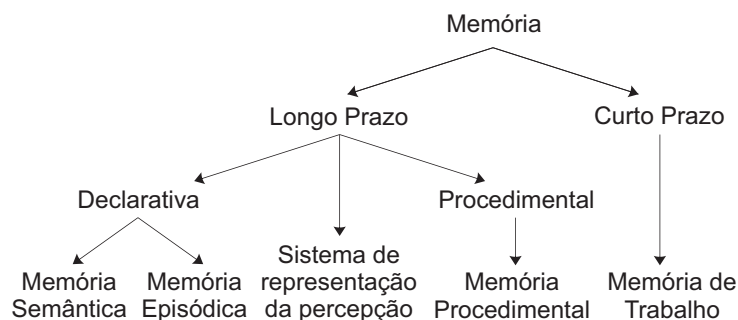


Figura 6 - Taxonomia da memória humana. Adaptado de (SCHACTER; WAGNER; BUCKNER, 2000).

Uma taxonomia que baseia-se na função realizada por cada tipo de memória foi inicialmente proposta por Tulving (1972). Segundo alguns autores (SCHACTER; WAGNER; BUCKNER, 2000) (STACHOWICZ; KRUIJFF, 2012), as memórias podem ser divididas em:

memória de trabalho

- Memória de trabalho** (*working memory*): é uma memória de curto prazo com capacidade de armazenamento limitado. Armazena informações utilizadas para o raciocínio em curso, isto é, uma representação interna/mental de situações correntes. Como uma memória de curto-prazo, especula-se que armazene tipicamente até 7 informações simultâneas

por cerca de 20 a 30 segundos;

memória
de
produção

- **Memória procedimental** ou **memória de produção** (*production memory* ou *procedural memory*): armazena conhecimento procedimental como regras de produção com pares condição-ação. As informações são armazenadas de forma não-declarativa. Em sistemas biológicos, esta memória localiza-se no corpo estriado (*striatum*) (STACHOWICZ; KRUIJFF, 2012);

memória
episódica

- **Memória episódica** (*episodic memory*): refere-se a eventos específicos (*o que, quando e onde*). É sensível ao contexto e histórico no qual a situação ocorreu, e as informações são armazenadas de forma declarativa. Mantém uma ordem cronológica dos episódios (atenvidade temporal). Contém eventos de experiência pessoal embarcados em seu contexto espaço-temporal. Em sistemas biológicos, esta memória é observada no lóbulo temporal medial, diencéfalo e córtex pré-frontal (STACHOWICZ; KRUIJFF, 2012). A memória episódica é considerada muito ligada à auto-consciência. O termo *memória como a episódica* (*episodic-like memory*) é atribuído a um sistema de memória que atende aos seguintes critérios (CLAYTON; DICKINSON, 1998) (CLAYTON; BUSSEY; DICKINSON, 2003):

critérios
memória
episódica

- **Conteúdo:** baseado na informação de uma experiência sobre o que aconteceu, *o que, quando e onde* é armazenado;
- **Estrutura:** A informação *o que, onde e quando* forma uma representação integrada, ou seja, buscando-se qualquer das características automaticamente as demais são acessadas;
- **Flexibilidade:** A informação armazenada possui natureza declarativa e pode ser depositada de forma flexível. Em particular, pode interagir com conhecimento semântico sempre, mesmo se o conhecimento foi adquirido depois que o episódio foi registrado na memória;

memória
semân-
tica

- **Memória semântica** (*semantic memory*): armazena fatos gerais e abstratos, independente de contexto específico ou da forma como as informações foram aprendidas. As informações são armazenadas de forma declarativa. Contém o conhecimento geral do indivíduo sobre o mundo. Assim como a memória episódica, em sistemas biológicos a memória semântica é localizada no lóbulo temporal medial, diencéfalo e córtex pré-frontal (STACHOWICZ; KRUIJFF, 2012).

memória
autobio-
gráfica

Alguns autores (JOCKEL; WESTHOFF; ZHANG, 2007) sugerem ainda a existência da **memória autobiográfica**, que seria composta por episódios recordados sobre a vida do próprio indivíduo. Outros autores já sugerem que a *memória autobiográfica* seria, na verdade, a própria memória episódica ou uma sub-área desta (NEVID, 2009). Alguns autores (FRANKLIN,

memória
percep-
tual

2005) admitem ainda a existência de uma **memória perceptual** (*perceptual memory*) relacionada à habilidade de interpretar estímulos recebidos através do reconhecimento de indivíduos, categorizando-os, e observando relações entre esses indivíduos e categorias. Esta memória se preocuparia, portanto, com a coleção de objetos de mesma categoria, observando, por exemplo, uma floresta ao invés das árvores individuais. Alguns argumentam que trata-se da própria memória episódica. Franklin (2005) observa que crianças que ainda não desenvolveram a memória episódica estão aptas a reconhecer e categorizar.

4.3 A relação entre consciência e atenção

A cada instante de tempo, uma pessoa somente está consciente de uma fração da informação sensorial que ela recebe (REGGIA, 2013). O mecanismo da **atenção**, desta forma, é o responsável por selecionar ativamente quais, dentre todos os estímulos recebidos pelo indivíduo, receberá o foco da atenção. Esta seleção é feita continuamente pelo sujeito baseada em parâmetros dos objetos do mundo, tais como forma, tamanho, cor, cheiro, som, dentre outros. Quando retiramos nossa atenção desse objeto, ele desaparece da consciência. Essa é a base dos principais argumentos utilizados pelos que defendem que consciência e atenção são conceitos profundamente relacionados (POSNER; DEHAENE, 1994) (VELMANS, 1996) (MERIKLE; JOORDENS, 1997) (O'REGAN; NOE, 2001) (KOCH; TSUCHIYA, 2007).

Embora trate-se de conceitos que mantêm grande proximidade, consciência e atenção não são, ao menos aparentemente, o mesmo processo (KOCH; TSUCHIYA, 2007) (TAYLOR, 2010). Alguns autores argumentam que trata-se de dois processos distintos, que têm funções e mecanismos neurais distintos (KOCH, 2004) (HARDCASTLE, 1997) (LAMME, 2003) (WOODMAN; LUCK, 2003) (TALLON-BAUDRY, 2012), e que ocorrem conjuntamente. De fato, é possível identificar situações que despertam e que não despertam a consciência em processos atencionais *top-down*. Na mesma linha de raciocínio, alguns autores colocam a questão da *percepção às cegas* (*blindsight*), uma doença onde os pacientes demonstram ser capazes de responder funcionalmente a características externas – isto é, demonstram atenção – mas não reportam ter consciência sobre elas (GRASSIA, 2001) (GRASSIA, 2004). Desta forma, corrobora-se a interpretação de que consciência e atenção seriam processos distintos (KOCH; TSUCHIYA, 2007).

Mas se atenção e consciência são processos tão próximos, porém distintos, cabe questionar: quais exatamente seriam os limites de atuação desses processos? Alguns argumentos de ordem filosófica propostos por Grassia (2001) e Grassia (2004) observam que:

- (I) O conhecimento do processamento atencional é em 3^a pessoa;
- (II) O conhecimento do estado de consciência é em 1^a pessoa;

(III) O conhecimento em 1^a e 3^a pessoas são irreduzíveis um ao outro.

Pelas afirmações realizadas, não somente corrobora-se a distinção entre esses dois processos como enfatiza-se sua clara distinção de propósito. É importante mencionar aqui que muitos autores aceitam a definição de que um agente possa estar *consciente* de um evento externo (tal como da presença de um predador, por exemplo), bem como possa estar *atento* a um evento interno (tal como a energia em uma bateria). Contudo, é fato que parte expressiva das ações associadas à consciência é realizada em 1^a pessoa, tal como o acesso ou atualização do estado de consciência, enquanto a parte mais expressiva das ações atencionais diga respeito à observação de sensores que observam a terceiros, tais como a visão, olfato e tato nos seres humanos. A consciência e a atenção manteriam ambas, portanto, suas bases na percepção, mas com objetivos e métodos distintos. Embora atenção e consciência utilizem-se, por exemplo, da percepção, a consciência pertenceria ao nível subjetivo da percepção (GRASSIA, 2001) (GRASSIA, 2004)

Além disso, parece haver por parte da consciência uma relação de dependência do processo atencional. Teoriza-se que informações que não sejam atendidas pelo sistema atencional não atinjam a consciência (COHEN et al., 2012). De fato, segundo Cohen et al. (2012), não há evidências experimentais suficientes para concluir, até o momento, que seja possível a um estímulo atingir a consciência sem que o mesmo tenha sido foco da atenção.

Segundo este entendimento, a consciência estaria posta em um nível hierárquico distinto, relacionando-se não apenas com a atenção, mas também com competição e aprendizado (GROSSBERG, 2010). Alguns autores teorizam que atenção e consciência correspondem a mecanismos distintos que alimentam um mesmo processo decisório que leva ao comportamento (TALLON-BAUDRY, 2012). Considera-se ainda a existência de um mecanismo de atenção focal (executiva) associado com a habilidade voluntária de selecionar dentre itens competitivos e regular emoções (POSNER; ROTHBART, 1998).

4.4 Os problemas da consciência e o qualia

A literatura sobre consciência frequentemente diferencia o uso desta palavra para duas situações distintas (ou dois problemas distintos). O termo consciência pode ser utilizado para se referenciar ao que é conhecido como o *problema fácil da consciência* e o *problema difícil da consciência* (CHALMERS, 1996). O **problema fácil da consciência** (*easy problem*) refere-se ao entendimento dos aspectos de processamento da informação em mentes conscientes, tais como a habilidade de integrar informação, a habilidade de focar a atenção, habilidade de acessar e reportar estados internos, etc.. Por outro lado, o **problema difícil da consciência** (*hard problem*) refere-se à experiência mais subjetiva associada com um ser consciente. Trata-se, em suma, de problemas distintos relacionados à consciência: mesmo que explique-se o problema fácil da consciência, não necessariamente ela será acompanhada da explicação para o problema

problema fácil

problema difícil

consciência fenomenológica

consciência
funcional

difícil. O termo **consciência fenomenológica** (*phenomenal consciousness*) tem sido empregado especificamente quando se trata do fenômeno da experiência subjetiva. Por outro lado, o termo **consciência funcional** ou acessível (*access consciousness*) tem sido empregado para o conceito funcionalista da consciência, representando, portanto, um conceito totalmente diferente do primeiro.

qualia

O termo **qualia** é usualmente utilizado para descrever as experiências subjetivas, tais como a sensação de ler ao por do Sol, uma dor de dente, ou o cheiro de uma rosa (REGGIA, 2013), não necessariamente relacionadas com as entradas sensoriais (HAWKINS; BLAKESLEE, 2004). Edelman (1992) destaca o qualia como a base para a teoria da consciência:

"Isso sugere uma abordagem para o problema do qualia. Como uma base para a teoria da consciência, é sensível assumir que, como em nós mesmos, qualia existe em outros seres humanos conscientes, se eles são considerados como observadores científicos ou como sujeitos. (...) Nós podemos tomar os seres humanos como as melhores referências canônicas para o estudo da consciência. Isso é justificado pelo fato de que os relatos subjetivos humanos (incluindo aqueles sobre o qualia), ações, e estruturas e funções cerebrais podem estar correlacionados. Depois de construir uma teoria baseada na premissa de que o qualia existe em seres humanos, nós podemos então olhar novamente para algumas das propriedades do qualia baseados nesta correlação. Esta é nossa habilidade de relatar e correlacionar enquanto experimentando individualmente o qualia que abre a possibilidade de uma investigação científica da consciência".

Edelman (1992, p. 115)

HAIKONEN (2009) coloca da seguinte forma a relação entre *qualia* e consciência: percepção é o processo de aquisição de informação do cérebro e o produto direto deste processo aparece no *qualia*. A mente humana é capaz de fazer uma análise introspectiva de seu conteúdo. Se eliminarmos o *qualia*, desaparecem também a percepção e a introspecção, pois o *qualia* é seu meio de manifestação. Não restará nada sobre o qual ser consciente, e, conseqüentemente, a consciência cessará. Esta observação leva à conclusão: ser consciente na forma humana é ter *qualia* (GAMEZ, 2008).

4.5 Uma consciência para cada cérebro?

Alguns relatos clínicos podem colocar em cheque a noção intuitiva de que a consciência seria indivisível e de que a mais de uma consciência não poderia coexistir em um cérebro. Duas situações relatadas em Argonov (2012) merecem destaque. Na primeira situação, relatos

demonstram que uma separação entre consciências pode ocorrer após intervenções cirúrgicas separando os hemisférios cerebrais (procedimento usualmente utilizado para o tratamento da epilepsia). Há grande número de relatos de pessoas que desenvolvem o que é relatado como duas mentes independentes: o hemisfério esquerdo controla a mão direita e a fala, enquanto o direito controla a mão esquerda (SPERRY, 1966) (GAZZINIGA, 1970). As explicações para este fenômeno são bastante variadas, mas há três hipóteses principais: *i*) a consciência realmente é dividida durante a cirurgia (NAGEL, 1971); *ii*) há duas consciências no cérebro mesmo antes da cirurgia (PUC CETTI, 1973), e *iii*) há apenas uma consciência mesmo depois da cirurgia (MARKS, 1981).

Uma segunda situação descrita trata da questão do **craniopagus**, um tipo raro de anomalia que ocorre de 4 a 6 nascimentos humanos em cada dez milhões, que faz com que o crânio de dois gêmeos fique unido. A maioria dos pacientes que sofrem desta anomalia morre durante a primeira infância, e não é capaz de relatar sua experiência subjetiva. Contudo, Argonov (2012) reporta o caso de Tatiana e Krista Hogan (ROBERTS, 2011), que atingiram a idade de 5 anos. Uma ponte conecta seus tálamos, e cada gêmea é capaz de observar os dados sensoriais da outra irmã. Contudo, cada irmã controla apenas seus próprios membros e demonstra seu próprio comportamento. Parece, portanto, biologicamente plausível que mais de uma mente compartilhe o mesmo cérebro.

4.6 Como avaliar a consciência?

A própria existência da doutrina solipsista – que argumenta que a única coisa da qual se pode ter certeza é da existência da própria mente – fornece indícios da dificuldade que existe em se caracterizar ou avaliar outras consciências, considerando que elas realmente existam. Quando se trata da possibilidade da existência de consciências artificiais, o problema é agravado, pois existem enormes diferenças de implementação entre os diferentes sistemas artificiais propostos.

Tratemos primeiramente a avaliação da consciência natural. Baseado em padrões de atividade neural e definições clínicas de entidades de medicina, usualmente aceita-se que existem alguns estados distintos de consciência em seres humanos (LEISMAN; KOCH, 2009):

estado
de coma

- Estado **inconsciente em coma**: o indivíduo não está acordado nem atento;

estado
vegeta-
tivo

- Estado **inconsciente vegetativo**: onde o indivíduo está acordado, mas não atento;

estado
minima-
mente
consci-
ente

- Estado **minimamente consciente**: onde o indivíduo está acordado e atento, mas suas funções da parte superior do cérebro não estão plenamente ativas;
- Estado **consciente**.

[Leisman e Koch \(2009\)](#) apresentam os critérios clínicos usualmente utilizados por médicos para declaração da morte de um indivíduo: *i*) não-receptividade e ausência de resposta para estímulos externos e necessidade interior; *ii*) ausência de movimentos musculares espontâneos ou respiração espontânea; *iii*) falta de reflexos no tronco cerebral que possam ser provocados. O protocolo diz que pacientes que passem nesse teste devem ser considerados mortos mesmo que seu sistema circulatório continue em funcionamento. A morte de um indivíduo, portanto, está intimamente ligada à falência de seu sistema cerebral e ausência permanente de consciência.

Quando transferimos esta questão para o universo das máquinas, alguns autores entendem que, em última instância, o problema de determinar a existência ou não de consciência – ou quantificá-la – assemelha-se fortemente ao problema de determinar se uma máquina é ou não inteligente ([KURZWEIL, 2007](#)). Para este segundo caso, o **Teste de Turing** ([TURING, 1950](#)) é o teste posto como parâmetro para a avaliação. Ou seja, na falta de qualquer outro método de avaliação objetivo de comprovada eficácia, adota-se uma avaliação subjetiva. Segundo [Kurzweil \(2007\)](#), se uma máquina puder passar num Teste de Turing válido, possivelmente vamos acreditar que ela seja consciente.

Mesmo tratando em um domínio subjetivo, diversos autores têm investigado os aspectos envolvidos na avaliação da consciência artificial. Destacam-se alguns trabalhos propondo métricas ([TONONI, 2004](#)) ([TONONI, 2008](#)) ([SETH, 2005](#)) ([ARRABALES; LEDEZMA; SANCHIS, 2009d](#)) e axiomas ([ALEKSANDER; DUNMALL, 2003](#)) para tal. Mais do que apenas avaliar os sistemas existentes, ferramentas comparativas ou de avaliação podem fornecer um direcionamento de extrema importância para o planejamento da evolução desses sistemas. Dois desses trabalhos serão destacados a seguir.

4.6.1 Axiomas para consciência

O trabalho de [Aleksander e Dunmall \(2003\)](#) descreve o que os autores denominam **axiomas para a presença de uma consciência mínima** em agentes, que devem ser considerados no projeto e avaliação da consciência em máquinas:

1. **Presença do agente no mundo:** Este axioma relaciona-se com a personificação e contextualização do agente no mundo físico;
2. **Imaginação e recordação de memórias e experiências passadas e presentes:** segundo [Starzyk e Prasad \(2011\)](#), esse axioma refere-se a: *i*) armazenamento de eventos e experiências do agente consciente na forma de memórias associativas e a habilidade de recordá-las, e *ii*) habilidade de formular cenários imaginativos e prever os desfechos dos cenários baseado nas experiências;

3. **Atenção como um determinador do contexto das experiências:** requer que o agente interaja (utilizando sua coleção de dados sensoriais, ações motoras e/ou processamento sub-cortical) com os objetos ou eventos do foco corrente da atenção;
4. **Volição:** o agente deve ter motivação e capacidade de formular objetivos, além da capacidade de tomar ações para cumprir seus objetivos;
5. **Emoção:** O agente deve ser capaz de identificar a natureza das experiências em seu próprio contexto e conjunto de emoções.

4.6.2 ConScale: uma métrica para a avaliação da consciência

Considerando que os sistemas artificiais criados até o momento inspiram-se em aspectos de organismos biológicos, parece natural que os mecanismos de avaliação possam levar em conta essa similaridade. O **ConScale** (ARRABALES; LEDEZMA; SANCHIS, 2009a) (ARRABALES; LEDEZMA; SANCHIS, 2009b) (ARRABALES; LEDEZMA; SANCHIS, 2009c) (ARRABALES; LEDEZMA; SANCHIS, 2009d) (ARRABALES; LEDEZMA; SANCHIS, 2010a) (ARRABALES; LEDEZMA; SANCHIS, 2010b) é uma proposta de *framework* que utiliza critérios arquitetural e comportamental em um nível de abstração funcional. Foca no problema de identificar as funções cognitivas mais importantes associadas com a consciência em agente artificiais e na questão de como essas funções podem ser efetivamente integradas de forma a levar à construção de agentes que ajam como humanos.

A ConScale apresenta níveis de consciência que vão do nível -1 ao nível 11. A [tabela 1](#) apresenta os 10 níveis mais comuns considerados, classificando o agente do nível reativo (ConScale 2) ao nível humano (ConScale 10) ou super-consciente (ConScale 11). Cada nível define um conjunto de **habilidades cognitivas** (*Cognitive Skills* - CS) que precisam ser satisfeitas. Um agente somente pode ser classificado em um nível *se e somente se* os níveis inferiores forem completamente satisfeitos. A dependência cognitiva das várias habilidades (CSs) é apresentada na [figura 7](#).

Nível	Habilidade Cognitiva (<i>Cognitive Skill</i> - CS)	
2	CS _{2,1} :	Resposta reativa fixa (reflexos)
3	CS _{3,1} :	Aquisição autônoma de novas respostas reativas adaptativas
	CS _{3,2} :	Uso de sensoriamento propioceptivo para respostas adaptativas embarcadas
	CS _{3,3-5} :	Seleção de informação sensorial/motor/memória relevante
	CS _{3,6} :	Avaliação (positiva/negativa) de objetos ou eventos selecionados
	CS _{3,7} :	Seleção do que precisa ser armazenado na memória
4	CS _{4,1} :	Aprendizado por tentativa e erro. Reavaliação de objetos ou eventos selecionados
	CS _{4,2} :	Comportamento direcionado para alvos especiais como seguir ou escapar
	CS _{4,3} :	Avaliação do desempenho na realização de um único objetivo

	CS _{4,4} :	Capacidade de planejamento básica: cálculo das próximas n ações sequenciais
	CS _{4,5} :	Habilidade para construir representações eloquentes (ALEKSANDER; DUNMALL, 2003) de percepções para cada modalidade sensorial disponível
5	CS _{5,1} :	Habilidade de se mover para frente e para trás em múltiplas tarefas
	CS _{5,2} :	Busca por múltiplos objetivos
	CS _{5,3} :	Avaliação do desempenho no atendimento a múltiplos objetivos
	CS _{5,4} :	Aprendizado por reforço autônomo (aprendizado emocional)
	CS _{5,5} :	Capacidade de planejamento avançada considerando todos os objetivos ativos
	CS _{5,6} :	Habilidade de gerar conteúdo mental de acordo com um significado (HAIKONEN, 2007b); Integração de diferentes modalidades em percepções explícitas diferenciadas (TONONI, 2008)
6	CS _{6,1} :	Ponderação sobre o auto-estado (emoções em background)
	CS _{6,2} :	Emoções em background causam efeito no corpo do agente
	CS _{6,3} :	Representação do efeito das emoções no organismo e planejamento (<i>feelings</i>) (DAMASIO, 1999)
	CS _{6,4} :	Habilidade de manter um mapa preciso e atualizado do esquema corporal
	CS _{6,5} :	Aprendizado abstrato (generalização de lições aprendidas)
	CS _{6,6} :	Habilidade de representar um fluxo de percepções integradas, incluindo auto-estado
7	CS _{7,1-3} :	Representação da relação entre o eu e percepções/ações/sentimentos
	CS _{7,4} :	Capacidade de auto-reconhecimento
	CS _{7,5} :	Planejamento avançado incluindo o eu como um ator dos planos
	CS _{7,6} :	Uso de estados <i>imaginacionais</i> no planejamento
	CS _{7,7} :	Aprendizado do uso de ferramentas
	CS _{7,8} :	Habilidade de representar e auto-reportar conteúdo mental (fluxo interno contínuo de percepções - imagens internas)
8	CS _{8,1} :	Habilidade de modelar outros como eus subjetivos
	CS _{8,2} :	Aprendizado por imitação de um outro
	CS _{8,3} :	Habilidade de colaborar com outros para perseguir um objetivo comum
	CS _{8,4} :	Planejamento social (planejamento com planos socialmente conscientes)
	CS _{8,5} :	Habilidade de produzir novas ferramentas
	CS _{8,6} :	Imagem interna é enriquecida com conteúdo mental relacionado com o modelo de outros e a relação entre o eu e os outros seres
9	CS _{9,1} :	Habilidade de desenvolver estratégias maquiavélicas como mentir e artimanhas
	CS _{9,2} :	Aprendizado social (aprendizado de novas estratégias maquiavélicas)
	CS _{9,3} :	Habilidades avançadas de comunicação (relatório de conteúdo mental detalhado como discurso interior básico)
	CS _{9,4} :	Grupos são capazes de desenvolver cultura
	CS _{9,5} :	Habilidade de modificar e adaptar o ambiente de acordo com as necessidades do agente
10	CS _{10,1} :	Relatório verbal preciso. Capacidades linguísticas avançadas. Discurso interior como o humano
	CS _{10,2} :	Habilidade de passar no Teste de Turing
	CS _{10,3} :	Grupos são capazes de desenvolver uma civilização e desenvolver cultura e tecnologia
11	CS _{11,1} :	Habilidade de gerenciar vários fluxos de consciência

Tabela 1 - ConScale - Habilidades cognitivas níveis 2 a 11 – 2: Reativo, 3: Adaptativo, 4: Atencional, 5: Executivo, 6: Emocional, 7: Auto-consciente, 8: Categórico, 9: Social, 10: Humano, 11: Supre-consciente. Como em (ARRABALES; LEDEZMA; SANCHIS, 2010b).

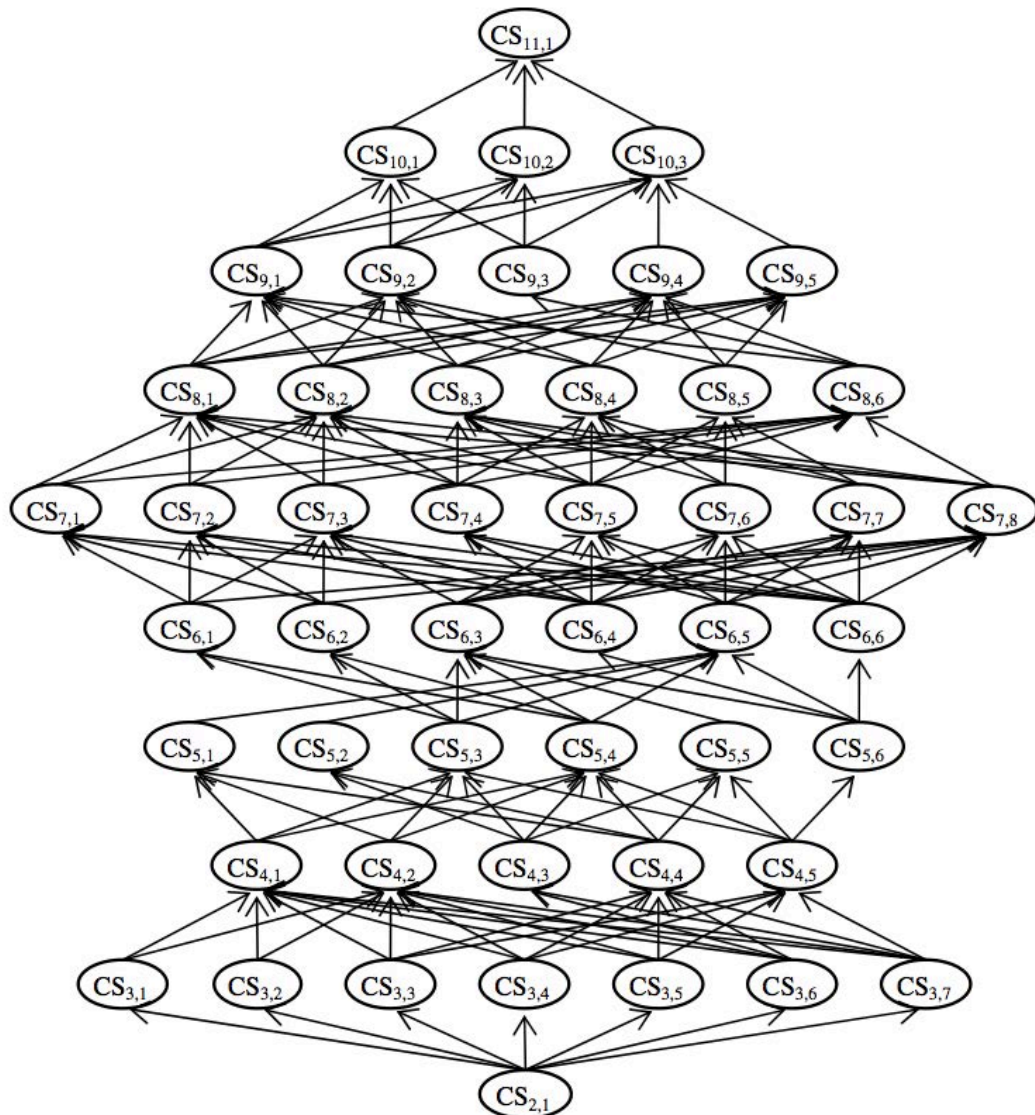


Figura 7 - Diagrama Hasse das habilidades cognitivas (CSs) do ConScale do nível hierárquico inferior (CS_{2,1}) ao nível hierárquico superior (CS_{11,1}), onde < representa uma relação de dependência cognitiva entre habilidades (ARRABALES; LEDEZMA; SANCHIS, 2010b).

A avaliação do ConScale é dada de duas formas. Primeiramente, através de uma avaliação **CQS** de **Pontuação Quantitativa do ConScale** (*Conscale Quantitative Score - CQS*). Seu objetivo é fornecer um valor numérico como indicativo do poder cognitivo da implementação sendo avaliada. É calculado em três passos (ARRABALES; LEDEZMA; SANCHIS, 2010b):

1. L_i , ou observação do nível i , fornece uma medida (entre 0.0 e 1.0) que segue uma curva exponencial como uma medida que representa a sinergia entre diferentes habilidades com o mesmo nível, isto é, quanto maior o CS atendido, maior será a contribuição de habilidades adicionais para o comportamento geral do agente;
2. **CLS** *CLS*, ou nível de pontuação cumulativa (*Cumulative Level Score* - CLS) combina todos os L_i em um único valor agregado (entre 0.0 e 1.55). Essa pontuação segue uma progressão logarítmica que impede a pontuação final de ser distorcida pelo efeito da combinação de grandes valores de L_i em níveis mais elevados e valores menores de L_i em níveis menores (isto é, boas implementações nos níveis 5 e 6, mas resultados não tão bons em níveis inferiores não deveriam fornecer pontuações elevadas);
3. *CQS* fornece um único valor (entre 0.0 e 1000) que indica a sinergia cumulativa produzida pela integração das habilidades cognitivas através de todos os níveis. *CQS* é desenvolvida como uma curva exponencial premiando as implementações que sigam o caminho de desenvolvimento implicitamente definido no nível de ordenação da ConScale.

Além da caracterização dos sistemas de consciência artificial por meio do *CQS*, os autores propõem ainda uma representação gráfica complementar denominada **perfil cognitivo** (*cognitive profile*), que apresenta graficamente as habilidades cognitivas do agente. A **perfil cognitivo** *figura 8* apresenta os valores possíveis para o *CQS*, e a *figura 9* apresenta exemplos de perfil cognitivo para algumas arquiteturas cognitivas conhecidas como o ELIZA (WEIZENBAUM, 1996), CERA UT-Bot (ARRABALES; LEDEZMA; SANCHIS, 2009d), CRONOS MAFI (MARQUES; HOLLAND, 2009), LIDA (FRANKLIN et al., 2007) e a arquitetura de Haikonen (HAIKONEN, 2007b) (arquiteturas discutidas nas próximas seções).

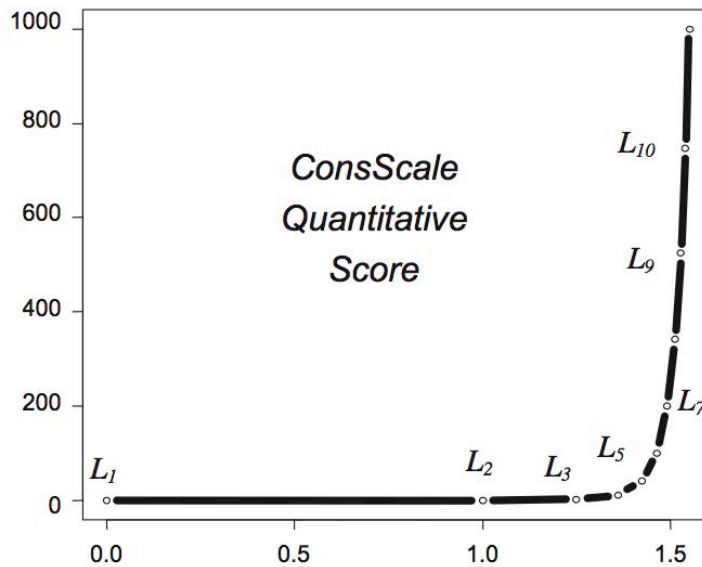


Figura 8 - Valores possíveis para o resultado quantitativo do ConScale: valores possíveis para o CQS como uma função do CLS (ARRABALES; LEDEZMA; SANCHIS, 2010b).

ELIZA	CERA UT Bot	CRONOS MAFI	LIDA	Haikonen
CS2,1; CS3,3; CS3,4; CS3,5; CS9,3.	CS2,1; CS3,1; CS3,2; CS3,3; CS3,4; CS3,5; CS3,6; CS4,1; CS4,5; CS5,2; CS5,4.	CS2,1; CS3,1; CS3,2; CS3,3; CS3,4; CS3,5; CS3,6; CS3,7; CS4,1; CS4,2; CS4,3; CS4,4; CS4,5; CS5,4; CS6,4; CS7,1; CS7,2; CS7,5; CS7,6.	CS2,1; CS3,1; CS3,2; CS3,3; CS3,4; CS3,5; CS3,6; CS3,7; CS4,1; CS4,2; CS4,3; CS4,4; CS4,5; CS5,1; CS5,2; CS5,3; CS5,4; CS5,5; CS5,6; CS6,1; CS6,2; CS6,3; CS6,4; CS6,5; CS6,6; CS7,1; CS7,2; CS7,3; CS7,6; CS8,1.	CS2,1; CS3,1; CS3,2; CS3,3; CS3,4; CS3,5; CS3,6; CS3,7; CS4,1; CS4,2; CS4,3; CS4,4; CS4,5; CS5,1; CS5,2; CS5,3; CS5,4; CS5,5; CS5,6; CS6,1; CS6,2; CS6,3; CS6,4; CS6,5; CS6,6; CS7,1; CS7,2; CS7,3; CS7,4; CS7,8; CS9,3.

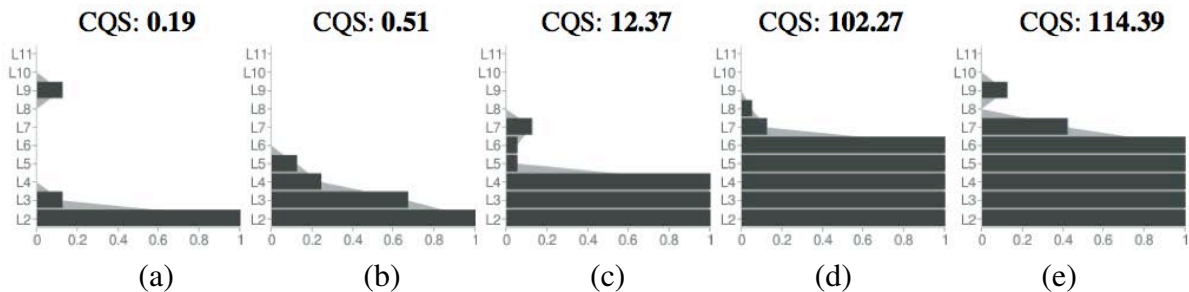


Figura 9 - Exemplos da representação gráfica de perfis cognitivos para algumas arquiteturas conscientes: a) nível 2 (Reativa); b) nível 2 (Reativa); c) nível 4 (Atencional); d) nível 6 (Emocional) e e) nível 6 (Emocional) (ARRABALES; LEDEZMA; SANCHIS, 2010b).

PARTE II

MODELOS COMPUTACIONAIS

5 MODELOS ATENCIONAIS

“Um dos meus filhos foi acordado no meio da noite por um ruído em circunstâncias que, naturalmente, requeriam cuidado, e depois de alguns minutos ele percebeu que sua boca estava escancarada. Ele, então, tornou-se consciente de que ele a tinha aberto por causa da respiração o mais silenciosamente possível (...). Quando a atenção está concentrada por um período de tempo com afimco em qualquer objeto ou assunto, todos os órgãos do corpo são esquecidos e negligenciados.”

Darwin (1899, cap. 12)

O presente capítulo discorre sobre os principais modelos computacionais para a modelagem da atenção. O capítulo encontra-se organizado da seguinte forma:

- [Seção 5.1](#): Apresenta os principais modelos atencionais encontrados na literatura;
- [Seção 5.2](#): Descreve em maior profundidade o modelo de [Colombini \(2014\)](#), adotado no contexto do presente trabalho.

5.1 Visão geral dos modelos

Diversas arquiteturas foram propostas ao longo dos anos para a modelagem de processos atencionais. Um primeiro modelo de atenção foi proposto por [Broadbent \(1958\)](#). Nesse modelo, os dados sensoriais são enviados a uma memória de curto prazo, e então passam por um filtro atencional capaz de realizar a seleção baseada nas propriedades físicas dos estímulos. Trata-se de um modelo típico de *seleção antecipada*. Não existe realimentação no sistema. O modelo é apresentado esquematicamente na [figura 10a](#). Em um modelo atencional proposto por [Norman \(1968\)](#), tenta-se agrupar elementos relacionados à memória e processos referentes à atenção. Este é um modelo típico de *atenção tardia*, onde o sistema realiza o processamento de todos os estímulos antes da decisão sobre o foco da atenção, neste caso considerando também o peso de cada estímulo. O modelo é apresentado na [figura 10b](#). Já o modelo proposto por

[Kahneman \(1973\)](#) representa um processador com capacidade limitada de alocar recursos. Uma vez recebido o estímulo, o sistema observa a capacidade disponível (recursos), e, construindo sua política de alocação de recursos baseado nos objetivos/intenções correntes e permanentes, para distribuir a atenção entre os estímulos. Muitas tarefas podem, portanto, ser executadas enquanto os recursos não excedem os disponíveis. Um diagrama esquemático do sistema é apresentado na [figura 10c](#). Um outro modelo, cuja versão final é denominada *Guided Search 2*, ou GS2, foi proposto por [Lee, Buxton e Feng \(2005\)](#). O modelo considera processamento paralelo dos estímulos e utiliza informações sobre características simples para guiar a atenção em busca de conjunções, isto é, combinações de duas ou mais características, combinando estímulos *bottom-up* e *top-down* em um mapa integrado. Uma visão geral da abordagem é apresentada na [figura 10d](#).

A teoria da integração de características (*Feature-Integration Theory - FIT*) ([TREISMAN; GELADE, 1980](#)) é um modelo para tarefas de busca visual que envolve atenção focada em dois estágios: *i*) um estágio pré-atencional paralelo, *bottom-up*, capaz de processar diferentes características dos estímulos (tamanho, orientação, cor, movimento, etc.) e *ii*) um estágio atencional serial, que direciona a atenção para locais específicos do campo visual, realizando uma busca visual até que o alvo seja localizado. O modelo é apresentado esquematicamente na [figura 10e](#).

CLAM O modelo denominado CLAM (*Closed-loop Attention Model*) ([van der VELDE; DEKAMPS; VANDERVOORTVANDERKLEIJ, 2004](#)) propõe dois estágios do modelo FIT unificados em um modelo único. Seja o modelo apresentado na [figura 11a](#). Dados dois padrões de estímulos de entrada (*displays*), o processamento será realizado por quatro módulos: *i*) Lousa visual (*visual blackboard*), composto por neurônios com campos receptivos que codificam a informação de entrada; *ii*) mapa de características (*feature maps*), que realiza a identificação dos estímulos através de seus atributos; *iii*) Mapas seriais, que processa a informação local dos objetos no campo visual, transformando-os em coordenadas espaciais da localização dos objetos, e *iv*) Memória de trabalho, que armazena as características de um objeto desejado. Desta forma, a seleção pode resultar tanto do processamento das características básicas dos estímulos (forma, cor, etc.) quanto de características complexas (formas completas, por exemplo).

SAIM O modelo SAIM (*Selective Attention for Identification Model*) ([HEINKE; HUMPHREYS, 2005](#)) tem uma arquitetura composta por três elementos básicos: *i*) A rede de conteúdo, que permite uma representação dos dados com invariância à translação; *ii*) A rede de seleção, que informa que unidades da retina serão mapeadas para o foco da atenção (FOA), e *iii*) A rede de conhecimento, que identifica o conteúdo do foco da atenção através de um processo de comparação de padrões (*matching*), realizando uma atenuação na rede de seleção via processo *top-down*. Os objetos são selecionados no campo visual utilizando apenas o critério do tamanho. Uma representação do modelo é apresentada na [figura 11b](#).

CODAM Um outro modelo de particular relevância no cenário dos modelos atencionais é o CODAM

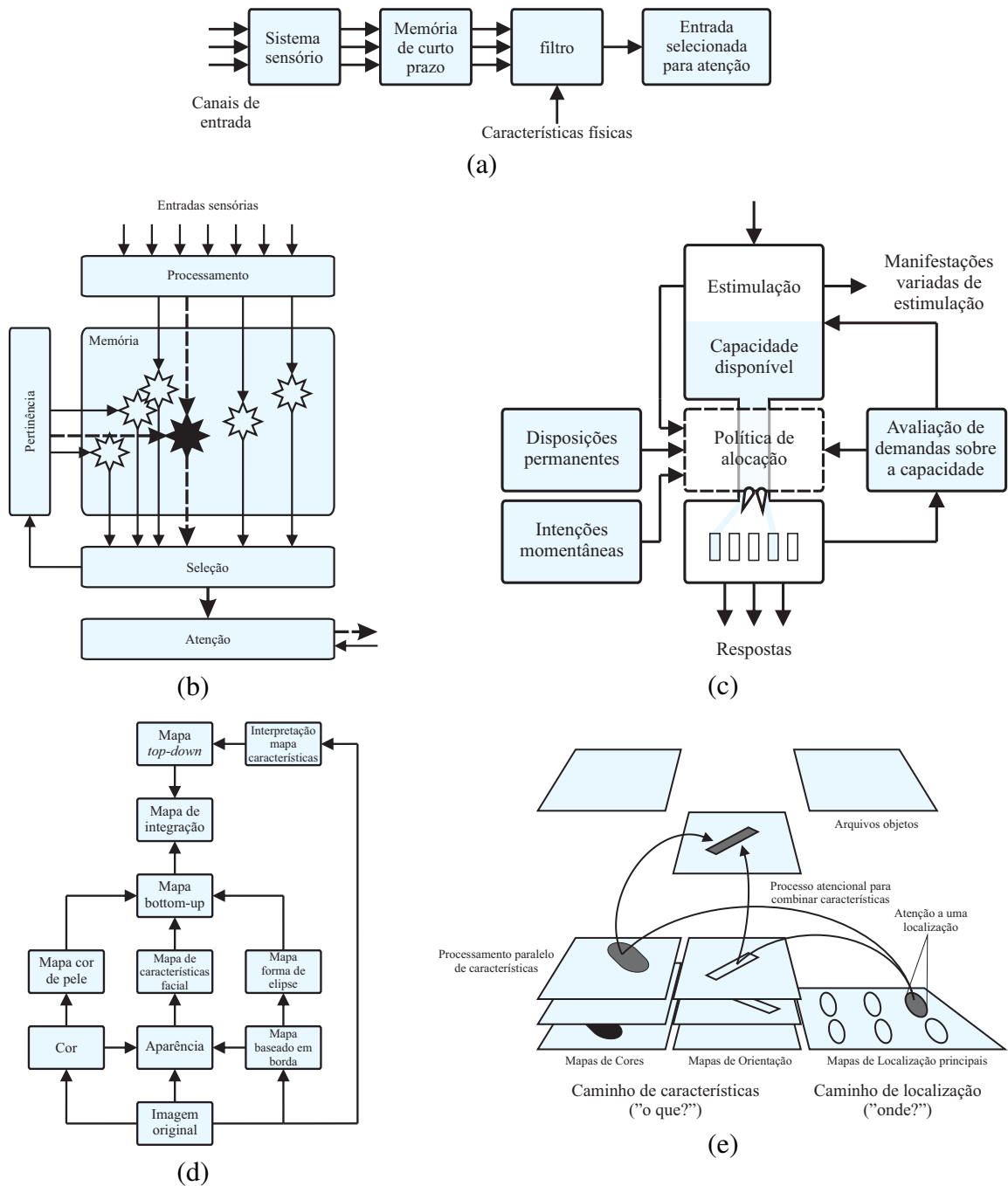


Figura 10 - Diagrama esquemático de funcionamento dos modelos atencionais: a) Modelo atencional de (BROADBENT, 1958); b) Modelo atencional de (NORMAN, 1968); c) Modelo atencional de (KAHNEMAN, 1973); d) Modelo atencional de (LEE; BUXTON; FENG, 2005); e) Modelo atencional de (TREISMAN; GELADE, 1980). Figuras adaptadas de (COLOMBINI, 2014).

(*Corollary Discharge of Attention Movement*) (KASDERIDIS; TAYLOR, 2004). O sistema é apresentado esquematicamente na [figura 11c](#). Um módulo monitor recebe observações dos sensores e os objetivos do agente e baseado num modelo interno do mundo dispara os eventos atencionais a serem executados através de um gerador de ações. Trata-se, portanto, de um sistema baseado em um *framework* de engenharia, estendendo a noção da atenção como um processador de nível sensorial para um modelo mais amplo de coordenação das funções cerebrais. Neste modelo a atenção é o gerador do conhecimento, uma vez que ela mesma controla a troca de modos automáticos para modos de aprendizado. O sistema tem comportamento orientado a objetivo, e a atenção inicia um aprendizado quando necessário, atuando sobre a resolução de conflitos de tomada de decisão. A proposta do modelo é de ampliar a noção de atenção utilizando duas frentes principais: *i*) sistema de controle de atenção interna, e *ii*) sistema de controle de atenção geral.

Por fim, um último modelo de referência é o proposto por Navalpakkam e Itti (2005). Este modelo considera apenas o elemento *bottom-up* da atenção aplicado sobre uma imagem. Uma vez percebida, a imagem é representada em dimensões de características distintas (cor, intensidade, orientação, etc.). As diferentes características são representadas em escalas distintas, de acordo com a especialização dos neurônios no caminho visual. Os mapas de característica são então calculados baseados nas diferenças entre o centro e o entorno das regiões visuais representadas. Uma vez normalizados, os mapas de contraste são construídos, permitindo a identificação das regiões de maior saliência através da combinação linear dos mapas de contraste. A região que apresentar o maior contraste tipicamente receberá o foco da atenção e sofrerá o curso do processo atencional (inibição de retorno). Um diagrama esquemático da arquitetura é apresentado na [figura 11d](#).

5.2 A arquitetura de Colombini

Uma arquitetura atencional propícia ao trabalho no domínio de sistemas robóticos é apresentada em Colombini (2014), Colombini, Simões e Ribeiro (2013) e Colombini e Ribeiro (2012). Esta arquitetura incorpora diversos aspectos de outros trabalhos correlatos e é capaz de lidar com múltiplos sistemas sensoriais, múltiplos processos para extração de características, processo decisório e suporte a aprendizado. O modelo é apresentado na [figura 12](#).

De uma perspectiva *bottom-up*, o modelo oferece suporte a múltiplos **sensores** de diferentes dimensões e naturezas que têm seus dados armazenados em uma **memória sensorial** por uma janela de tempo. Esses dados sensoriais são combinados para criar **mapas de características**, que são integrados em um único **mapa de características combinado**, que representa uma soma ponderada das diversas características. A junção do mapa de características combinado com o mapa atencional corrente (gerado anteriormente pelo agente) gera um **mapa de**

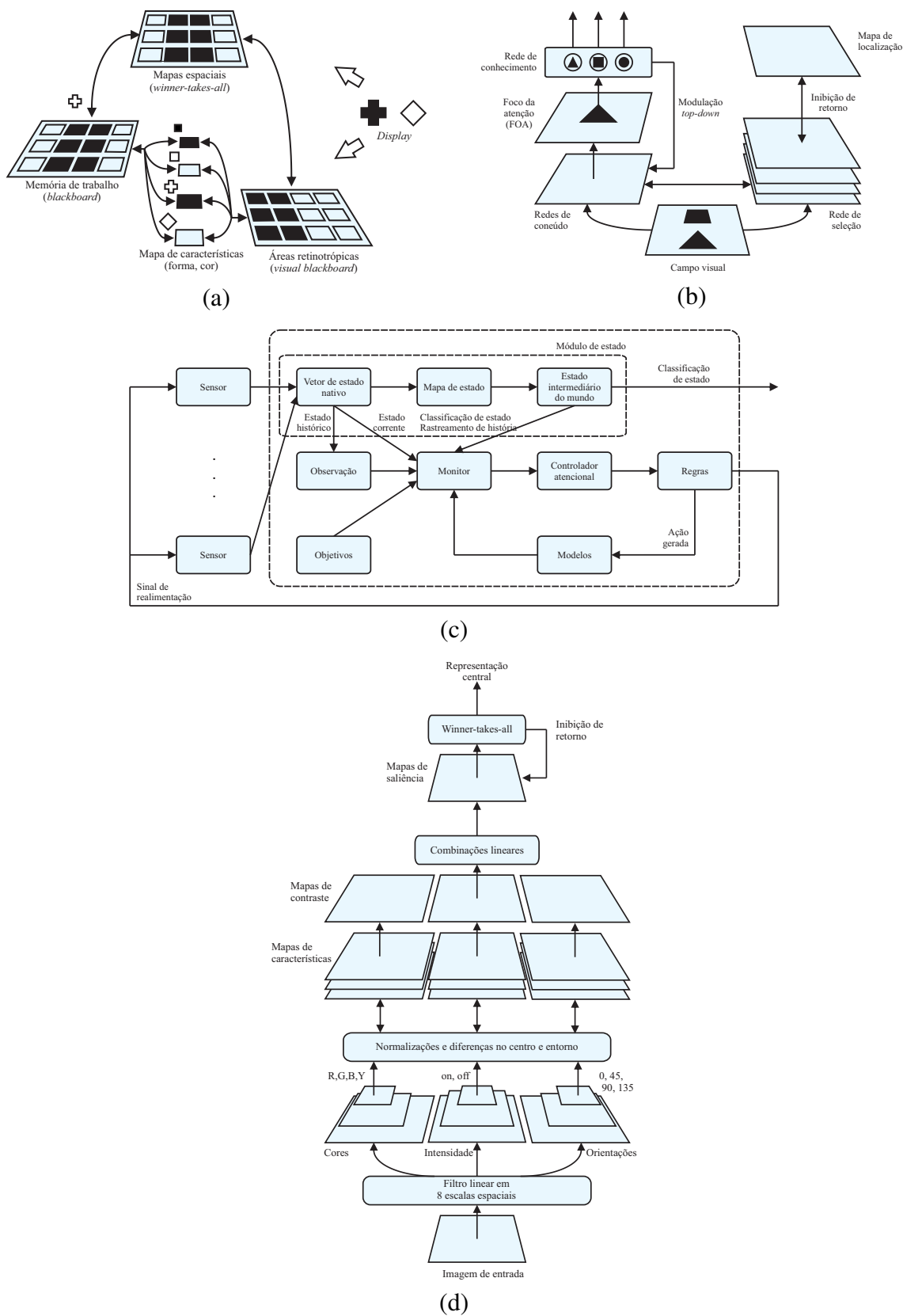


Figura 11 - Diagrama esquemático de funcionamento dos modelos atencionais: a) Modelo atencional de (van der VELDE; DEKAMPS; VANDERVOORTVANDERKLEIJ, 2004); b) Modelo atencional de (HEINKE; HUMPHREYS, 2005); c) Modelo atencional de (KASDERIDIS; TAYLOR, 2004); d) Modelo atencional de (NAVALPAKKAM; ITTI, 2005) Figuras a-c adaptadas de (COLOMBINI, 2014).

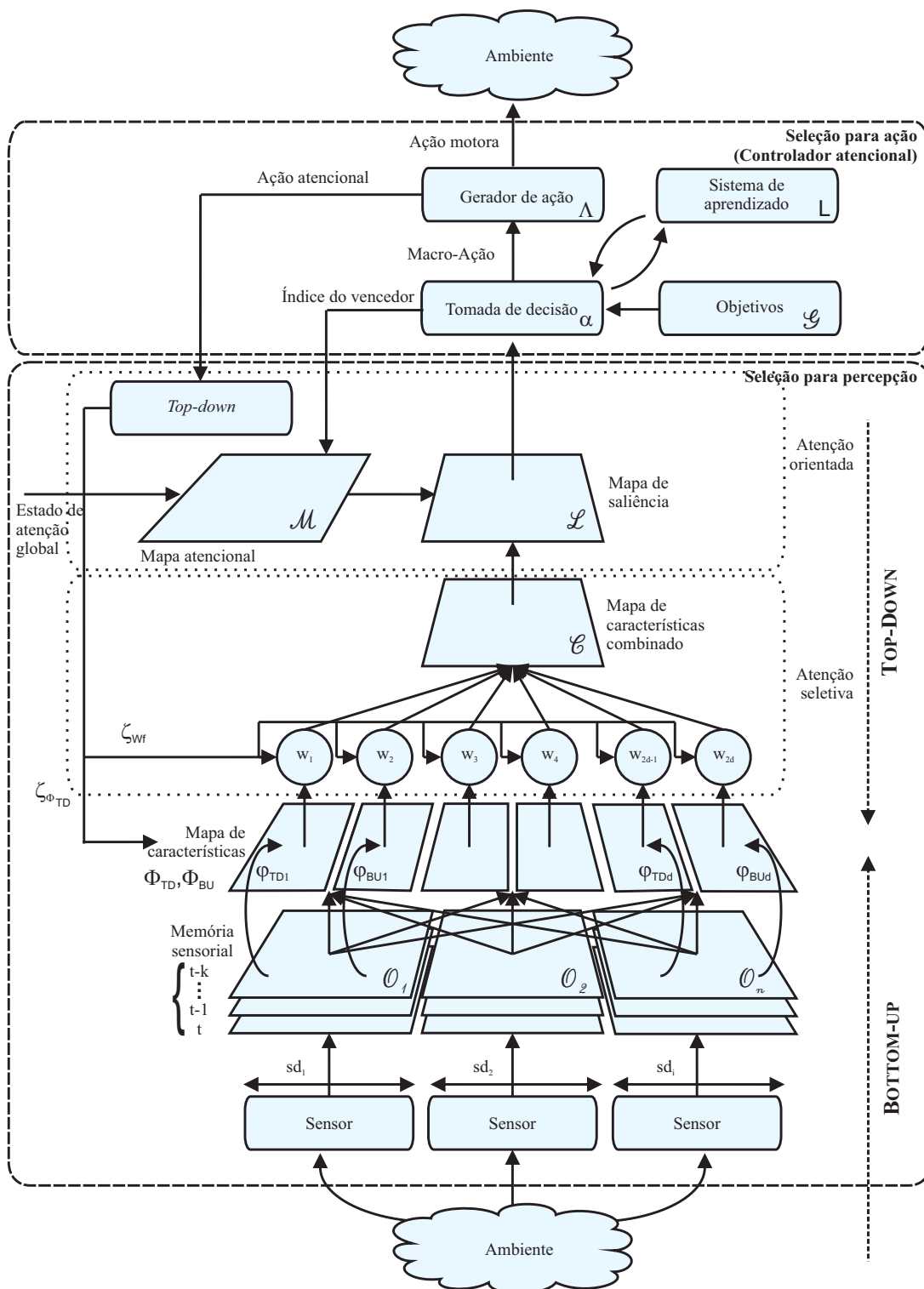


Figura 12 - Diagrama esquemático do modelo atencional proposto por (COLOMBINI, 2014) Adaptado.

saliências, que é, em essência, uma representação do ambiente como o percebido pelo agente, modulado pelo processo atencional. De acordo com as saliências uma sistemática de **seleção de vencedores** – um algoritmo *winner-takes-all*, por exemplo, – pode decidir – tipicamente observando o estímulo mais saliente no mapa de saliências – pela orientação da atenção para alguma região específica do **mapa atencional** por certo período de tempo. O mapa atencional, que representa o *status* atencional corrente do sistema, é utilizado no cômputo do mapa de saliência nos próximos instantes de tempo. O modelo prevê também uma sistemática de inibição (*inhibition of return - IOR*) (POSNER; COHEN, 1984) (KLEIN, 2000): a manutenção do foco atencional sobre uma região usualmente leva a uma adaptação que causa inibição no atendimento da mesma, ou, em outros termos, a sensibilidade desta região cai quando existe exposição prolongada de um estímulo, estimulando o chaveamento da atenção para outras regiões. Este efeito, bem como a excitação/inibição lateral nas vizinhanças do vencedor, são modelados conforme as curvas apresentadas na figura 13 e sua expressão matemática (MAASS; BISHOP, 1999), apresentada na equação 2.

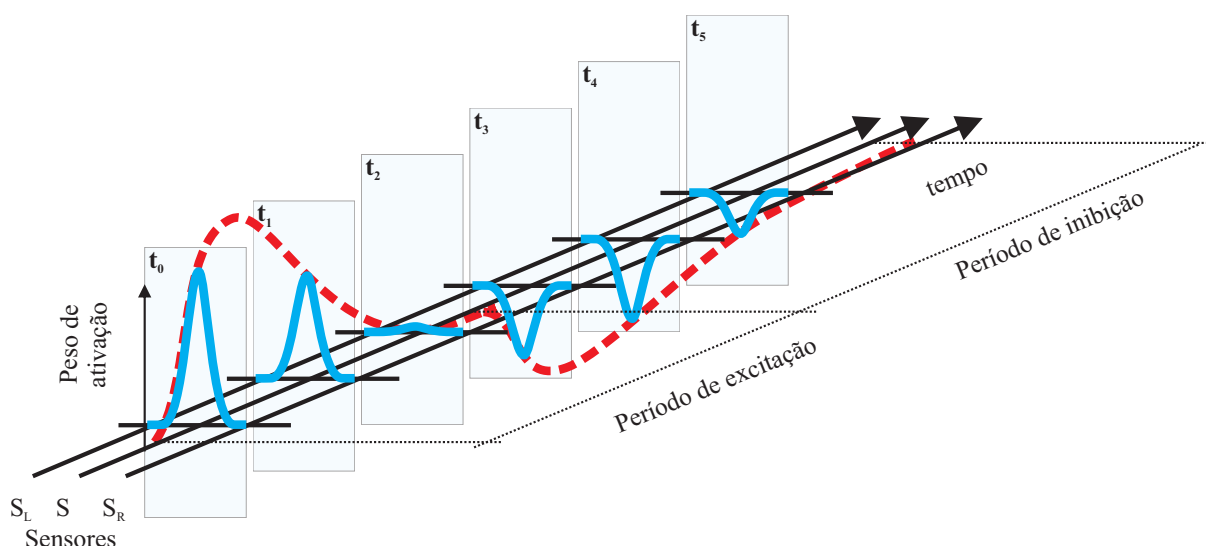


Figura 13 - Curvas responsáveis pela modelagem pelo enaltecimento ou inibição temporal e espacial: contribuição temporal e lateral quando a região S é escolhida como foco da atenção no tempo $t = t_0$. A observação do sensor S terá um peso de ativação variável no mapa atencional de acordo com o tempo, cujo valor é dado pela intersecção entre os planos t_0, t_1, \dots, t_n e a curva vermelha tracejada. A curva do peso de ativação é modelada com um período excitatório e um período inibitório. A vizinhança do sensor S , representada pelos sensores S_L (*left*) e S_R (*right*) receberá também uma contribuição lateral variável ao longo do tempo, representada pelas curvas azuis contínuas. Adaptado de (COLOMBINI, 2014).

De uma perspectiva *top-down*, dependendo do objetivo do sistema e da dinâmica atencional do estado corrente (orientada, seletiva ou sustentada), a atenção voluntária – tipicamente conforme deliberação de processos de hierarquia mais elevada – pode ser direcionada para uma região do espaço ou objeto. Esse direcionamento pode ocorrer de duas maneiras: *i*) delibe-

radamente reforçando a região no mapa atencional ou *ii*) ajustando os pesos que definem a contribuição de cada característica. Segundo Colombini (2014), em sistemas biológicos o componente endógeno da atenção usualmente inicia-se 200ms depois que o estímulo é atendido e ele permanece enquanto a tarefa o fizer necessário. De acordo com Ling e Carrasco (2006), se um estado de vigilância é mantido por mais de 6s, o atendimento ao estímulo na localização fica prejudicado, o que irá tornar mais difícil que esta região permaneça como foco da atenção. Este processo é modelado pela equação 3. Desta forma, trata-se de um modelo mais completo que os anteriormente apresentados, capaz de modelar:

- atenção sustentada
 - A *atenção sustentada*, guiada pelo componente de objetivo;
 - A *atenção seletiva*, que pode enfatizar características específicas;
- atenção seletiva
 - A *atenção orientada*, resultado do processo atencional para escolha do foco da atenção;
- atenção orientada
 - A *atenção evidente*, dado que o módulo de tomada de decisão pode gerar uma ação motora para ajustar as estruturas sensoriais;
- atenção evidente
 - A *atenção encoberta*, dado que o módulo de tomada de decisão pode não gerar nenhuma ação motora;
- atenção encoberta
 - A *atenção dividida*, dado que o módulo de tomada de decisão pode ter a habilidade de gerenciar múltiplos objetivos no tempo t ;
- atenção dividida
 - Estímulos *bottom-up* e *top-down*.

5.2.1 Controlador atencional

A arquitetura proposta por Colombini (2014) é composta por duas partes:

- seleção para percepção
 - *Seleção para percepção*: é a parte do modelo atencional que representa a parte inconsciente do agente, responsável por processar os dados sensoriais e decidir sobre os aspectos atencionais. Em última instância, o processo de seleção para percepção proposto pode ser visto como capaz de promover uma redução no espaço de estados do sistema, já que nesse paradigma o espaço do mapa atencional passa a substituir o espaço de estados formado por todas as dimensões de entrada;
- seleção para ação
 - *Seleção para ação*: é a parte que contempla o controlador atencional, que representa a parte cognitiva e/ou consciente do sistema, que considera os objetivos e intenção do agente. É um componente hierarquicamente superior à seleção para percepção.

O controlador proposto (figura 12) é composto por: *i*) Objetivos do agente; *ii*) Um sistema de tomada de decisão, que, observando os objetivos do agente, pode definir macro-ações deste agente sobre o espaço atencional; *iii*) Um gerador de ações, que recebe uma macro-ação e a desmembra em ações motoras (ações sobre os atuadores do sistema) e ações atencionais (que realimentam o sistema de seleção para percepção de acordo com os propósitos correntes do agente), e *iv*) Um sistema básico de aprendizado que pode substituir o processo de tomada de decisão para que o agente aprenda sobre seu ambiente. Quando não há objetivos correntes, o trabalho mostrou que as saliências são capazes de guiar o processo atencional, e, em última instância, todo o comportamento do agente. Já quando existe uma intenção explícita do agente, estratégias mais elaboradas podem ser necessárias, dependendo da natureza do objetivo.

5.2.2 Modelo formal

O modelo proposto na figura 12 é também o primeiro modelo com uma definição computacional formal. Ele pode ser formalmente descrito como (COLOMBINI, 2014):

- Um conjunto de ns sensores $\{s_1, s_2, \dots, s_{ns}\} \in \mathcal{S}$;
- Um conjunto de ns dimensões de sensores $\{sd_1, sd_2, \dots, sd_{ns}\} \in \mathcal{S}d$, com cada $sd_i \in \mathcal{S}d$ associada a um sensor $s_i \in \mathcal{S}$;
- Um conjunto de observações de sensores, denotado $\{\mathcal{O}_1, \mathcal{O}_2, \dots, \mathcal{O}_i\} \in \mathcal{O}$, onde $\mathcal{O}_{i,jk}$ denota a observação da j -ésima dimensão do i -ésimo sensor no tempo k , com $j \in \{1, \dots, sd_i\}$, $k \in [t-k, t]$ e t é o tempo da observação corrente;
- Um número de dimensões de características $d \in \mathbb{N}^*$ no qual as características *bottom-up* podem emergir ou características *top-down* podem atuar;
- Um conjunto de d funções de dimensão de característica *bottom-up* $\{\phi_{BU_1}, \phi_{BU_2}, \dots, \phi_{BU_d}\} \in \Phi_{BU}$;
- Um conjunto de d funções de dimensão de característica *top-down* $\{\phi_{TD_1}, \phi_{TD_2}, \dots, \phi_{TD_d}\} \in \Phi_{TD}$;
- Um conjunto de $2d$ Mapas de características n dimensionais, denotado $\mathcal{F} \in \mathfrak{R}^n$, com elementos denotados por $\{\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_{2d}\}$, com cada mapa de características \mathcal{F}_i representado por $\{f_{i_1}, f_{i_2}, \dots, f_{i_n}\}$;
- Um Vetor de pesos de características de $2d$ dimensões, denotado $\mathcal{W}_f \in \mathfrak{R}^d$, com elementos denotados por $\{wf_1, wf_2, \dots, wf_{2d}\}$;

- Um *Mapa de características combinadas* de n dimensões, denotado $\mathcal{C} \in \mathfrak{R}^n$, com elementos denotados por $\{c_1, c_2, \dots, c_n\}$;
- Um *Mapa de saliências* de n dimensões, denotado $\mathcal{L} \in \mathfrak{R}^n$, com elementos denotados por $\{l_1, l_2, \dots, l_n\}$;
- Um *Mapa atencional* de n dimensões, denotado $\mathcal{M} \in \mathfrak{R}^n$, com elementos denotados por $\{m_1, m_2, \dots, m_n\}$, com todos os elementos inicializados com valor 1;
- Um conjunto de *vencedores*, denotado \mathcal{B} , com elementos denotados por $\{b_1, b_2, \dots, b_z\}$. Cada elemento $b_i \in \mathcal{B}$ é uma 3-upla $\{\xi, t, (0, 1)\}$ onde os elementos são: um índice ξ que refere-se ao vencedor $l_i \in \mathcal{L}$, um tempo de disparo t e um elemento binário $\in \{0, 1\}$ que indica se o disparo foi causado por um processo endógeno ou exógeno;
- Um conjunto de objetivos $\{g_1, g_2, \dots, g_t\} \in \mathcal{G}$;
- Uma *função decisória* α que mapeia o mapa de saliências corrente (\mathcal{L}) e o objetivo corrente g_t para o índice de um vencedor;
- Uma *função de mapeamento de pesos top-down* ζ_{W_f}
- Uma *função de mapeamento de dimensões de características top-down* $\zeta_{\Phi_{TD}}$

A **operação do modelo** é dada por (COLOMBINI, 2014):

operação
do
modelo
de Co-
lombini

1. **Atualização Top-down.** De acordo com as funções de mapeamento ζ_{W_f} e $\zeta_{\Phi_{TD}}$, baseado no objetivo g no tempo t , atualizar o *vetor de pesos de características* \mathcal{W}_f e as *funções de dimensão de características top-down* $\{\phi_{TD_1}, \phi_{TD_2}, \dots, \phi_{TD_d}\} \in \Phi_{TD}$;
2. **Extração de características.** Para todas as dimensões d , aplicar uma *função de dimensão de característica bottom-up* ϕ_{BU_d} e uma *função de dimensão de característica top-down* capaz de extrair um mapa de características *bottom-up* de n dimensões, denotado \mathcal{F}_{BU_d} , e um mapa de características *top-down*, denotado \mathcal{F}_{TD_d} , no tempo t , estabelecendo os seguintes mapeamentos:

$$\phi_{BU_d} \\ \mathcal{O} \longrightarrow \mathcal{F}_{BU_{d_t}} : \mathfrak{R}^n$$

$$\phi_{TD_d} \\ \mathcal{O} \longrightarrow \mathcal{F}_{TD_{d_t}} : \mathfrak{R}^n$$

3. **Computação do mapa de características combinado.** Para todo $i \in \{1, 2, \dots, n\}$ computar $c_{i_t} \in \mathcal{C}_t$ fazendo:

$$c_{i_t} = \sum_{j=1}^d w f_{j_t} \cdot \mathcal{F}_{i_{j_t}}$$

4. **Computação do mapa de saliências.** Para todo $i \in \{1, 2, \dots, n\}$, computar \mathcal{L}_{d_t} , fazendo:

$$l_{i_t} = c_{i_t} \cdot m_{i_{t-1}},$$

com $m_{i_{t-1}} \in \mathcal{M}_{t-1}$ e $c_{i_t} \in \mathcal{C}_t$;

5. **Tomada de decisão.** Executar o mapeamento do Mapa de saliências \mathcal{L} e objetivo g_t para um índice de vencedor ξ . Inserir o vencedor $b_{i(\xi, t, (0, 1))}$ em \mathcal{B} . Se um processo de decisão *winner-takes-all* está em curso, o índice do vencedor ξ de b_i será computado sobre o Mapa de saliências como:

$$\xi = j, \text{ com } l_{j_t} = \max(\mathcal{L}), \text{ para } j \in \{1, 2, \dots, n\}$$

6. Atualização do Mapa atencional.

- (a) Para cada $b_i \in \mathcal{B}$ que disparou no tempo q , com $q \leq t$, faça:

- i. Se b_i disparou devido a um estímulo exógeno (*bottom-up*), computar a contribuição $\Delta b_{i(\xi, q, 0)}$ no tempo corrente t com respeito ao seu índice de vencedor ξ fazendo:

$$\Delta b_{i(\xi, q, 0)} = \left(\left[e^{\left(-\frac{t-q-\delta_p}{\tau_{aa}} \right)} - e^{\left(-\frac{t-q-\delta_p}{\tau_{ad}} \right)} \right] \cdot \mathcal{H}(t-q-\delta_p) \right) - \quad (1)$$

$$\left(\left[e^{\left(-\frac{t-q-\delta_r}{\tau_{ra}} \right)} - e^{\left(-\frac{t-q-\delta_r}{\tau_{rd}} \right)} \right] \cdot \mathcal{H}(t-q-\delta_r) \right),$$

onde:

$$\left\{ \begin{array}{l} \delta_p \text{ é a constante de tempo de atraso na pré-ativação;} \\ \delta_r \text{ é a constante de tempo de atraso na pré-refração;} \\ \tau_{aa} \text{ é a constante de ativação exponencial ascendente;} \\ \tau_{ad} \text{ é a constante de ativação exponencial descendente,} \\ \text{com } 0 < \tau_{ad} < \tau_{aa}; \\ \tau_{ra} \text{ é a constante de refração exponencial ascendente;} \\ \tau_{rd} \text{ é a constante de refração exponencial descendente,} \\ \text{com } 0 < \tau_{rd} < \tau_{ra}; \\ \xi \text{ é o componente do índice do vencedor do elemento } b_i; \\ \mathcal{H} \text{ é a função degrau de Heaviside,} \\ \text{não definida para } t < 0, \text{ definida como } 1/2 \text{ para} \\ t = 0 \text{ e } 1 \text{ para } t > 0; \end{array} \right.$$

- ii. Se b_i disparou devido a um estímulo exógeno (*bottom-up*), computar a contribuição $\Delta b_{i(j,q,0)}$ no tempo corrente t com respeito ao índice do vencedor ξ e sua vizinhança. Para cada $j \in \{1, 2, \dots, n\}$ com $j \neq \xi$, fazer:

$$\Delta b_{i(j,q,0)} = \Delta b_{i(\xi,q,0)} \cdot e^{-\frac{(j-\xi)^2}{2\tau_n^2}}, \quad (2)$$

onde τ_n é uma constante relacionada com a janela de vizinhança.

- iii. Se b_i disparou devido a um estímulo endógeno (*top-down*), computar a contribuição $\Delta b_{i(\xi,q,1)}$ no tempo corrente t com respeito ao índice do vencedor ξ fazendo:

$$\Delta b_{i(\xi,q,1)} = \left[e^{\left(-\frac{t-q-\delta_{ep}}{\tau_{eaa}} \right)} \right] \cdot \mathcal{H}(t-q-\delta_{ep}) - \quad (3)$$

$$e^{\left(-\frac{t-q-\delta_{ep}}{\tau_{ead}} \right)} \right] \cdot \mathcal{H}(t-q-\delta_{ep}),$$

onde:

$$\left\{ \begin{array}{l} \delta_{ep} \text{ é a constante de atraso na excitação endógena pré-ativação} \\ \delta_{edp} \text{ é a constante de tempo de atraso no decaimento endógeno;} \\ \tau_{eaa} \text{ é constante de ativação endógena exponencial ascendente} \\ \tau_{ead} \text{ é a constante de ativação exponencial endógena descendente, com} \\ 0 < \tau_{ead} < \tau_{eaa}; \\ \mathcal{H} \text{ é a função degrau de Heaviside, não definida para} \\ t < 0, \text{ definida como } 1/2 \text{ para} \\ t = 0 \text{ e } 1 \text{ para } t > 0; \end{array} \right.$$

- iv. Se b_i disparou devido a um estímulo endógeno (*top-down*), computar a contribuição $\Delta b_{i(j,q,1)}$ no tempo corrente t com respeito ao índice do vencedor ξ e sua vizinhança. Para todo $j \in \{1, 2, \dots, n\}$ com $j \neq \xi$, faça:

$$\Delta b_{i(j,q,1)} = \Delta b_{i(\xi,q,1)} \cdot e^{-\frac{(j-\xi)^2}{2\tau_{en}^2}}, \quad (4)$$

onde τ_{en} é a constante endógena relacionada com a janela de vizinhança.

- (b) Para todo $m_j \in \mathcal{M}$, atualize m_j com $j \in \{1, 2, \dots, n\}$ fazendo:

$$m_{j_t} = m_{j_{t-1}} + \Delta b_{i(j,q,0)} + \Delta b_{i(j,q,1)} \quad (5)$$

6 MODELOS PARA CONSCIÊNCIA

“Em cada um de nós há um outro que nós não conhecemos.”

Carl Jung
Eagleman (2012)

O presente capítulo busca apresentar uma visão geral dos modelos e arquiteturas utilizadas no trabalho computacional com agentes conscientes. É importante observar que foram considerados neste capítulo apenas modelos onde os autores, explicitamente, declararam que seu trabalho mantinha relação com aspectos da consciência. Ele encontra-se organizado da seguinte forma:

- [Seção 6.1](#): Propõe uma taxonomia para os modelos de consciência;
- [Seção 6.2](#): Apresenta os principais modelos para consciência que se apoiam no conceito do Espaço de Trabalho Global;
- [Seção 6.3](#): Apresenta os principais modelos para consciência que se apoiam no conceito da integração de informação;
- [Seção 6.4](#): Apresenta os principais modelos para consciência que se apoiam no conceito do auto-conhecimento;
- [Seção 6.5](#): Apresenta os principais modelos para consciência que se apoiam no conceito da representação superior;
- [Seção 6.6](#): Apresenta os principais modelos para consciência que se apoiam no conceito da atenção.

6.1 Taxonomia dos modelos

Os modelos para consciência em máquinas são usualmente divididos em cinco classes principais ([REGGIA, 2013](#)): *i*) modelos baseados no Espaço de Trabalho Global (GWT); *ii*) mo-

delos baseados na integração de informação; *iii*) modelos baseados no auto-conhecimento; *iv*) modelos baseados em representações de nível superior, e *v*) modelos baseados em atenção. Esses modelos serão detalhados nas seções a seguir. A [figura 14](#) apresenta uma taxonomia desses modelos tomando como base o princípio ou condição fundamental através do qual os autores consideram a emergência da consciência. Uma visão geral sobre os modelos e arquiteturas pode ser obtida em [Gamez \(2008\)](#) e [Reggia \(2013\)](#).

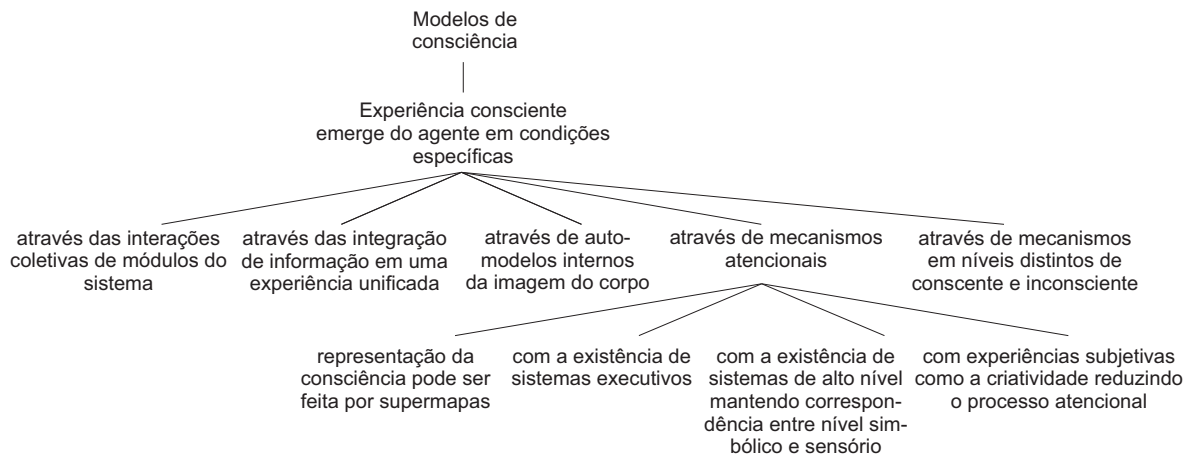


Figura 14 - Taxonomia dos sistemas conscientes considerando o princípio elementar para o surgimento da consciência com ênfase aos sistemas atencionais.

6.2 Modelos baseados no Espaço de Trabalho Global

espaço
de
trabalho
global

GWT

No modelo de trabalho da Teoria do **Espaço de Trabalho Global** (*Global Workspace Theory - GWT*), múltiplos processos especialistas competem e cooperam, paralelamente, pelo acesso ao espaço de trabalho global. A coalisão de processos vencedores pode transmitir informações a todos os outros processos em um sistema de *broadcast*. Esta proposta tem inspiração biológica no que acontece em regiões do córtex cerebral durante esforços mentais conscientes ([BAARS, 1988](#)) ([BAARS, 2002](#)) ([BAARS; RAMSEY; LAUREYS, 2003](#)) ([BAARS; FRANKLIN, 2007](#)) ([BAARS; FRANKLIN, 2009](#)). Alguns autores apresentam, por exemplo, modelos GWT inspirados no córtex sensorio e motor, gânglio basal e tálamo ([SHANAHAN, 2005](#)).

Em síntese, neste modelo de trabalho, a hipótese principal investigada pelos autores é a de que a experiência consciente emerge através de interações coletivas entre os processos especializados no espaço de trabalho ([REGGIA, 2013](#)). Via de regra, a dimensão fenomenológica, representada pela questão “*O robô está consciente da exploração que está fazendo?*” é deliberadamente negligenciada ([MORENO; ESPINO; MIGUEL, 2007](#)).

IDA

Um dos primeiros modelos computacionais para consciência é o IDA (*Intelligent Distributed Agent*) ([FRANKLIN; KELEMEN; MCCAULEY, 1998](#)) ([FRANKLIN, 2003](#)) ([FRANKLIN,](#)

2005) (BAARS; FRANKLIN, 2007), uma arquitetura cognitiva que contempla diversos tipos de memória (memória perceptual, memória de trabalho, memória episódica, memória procedimental), consciência, seleção de ações, satisfação de restrições, deliberação e volição (FRANKLIN, 2005). A arquitetura, apresentada esquematicamente na figura 15, baseia-se em um conceito de sistema multiagentes, e consiste de diversos programas em java (agentes denominados **codelets**) rodando de forma paralela. Neste modelo, o conteúdo do foco é escrito em memórias associativa e intermediária apenas depois de ter vindo à consciência. De acordo com os autores, o fato dos agentes se comunicarem com os demais via espaço de trabalho global no contexto apresentado implica que a IDA é funcionalmente consciente (FRANKLIN, 2003).

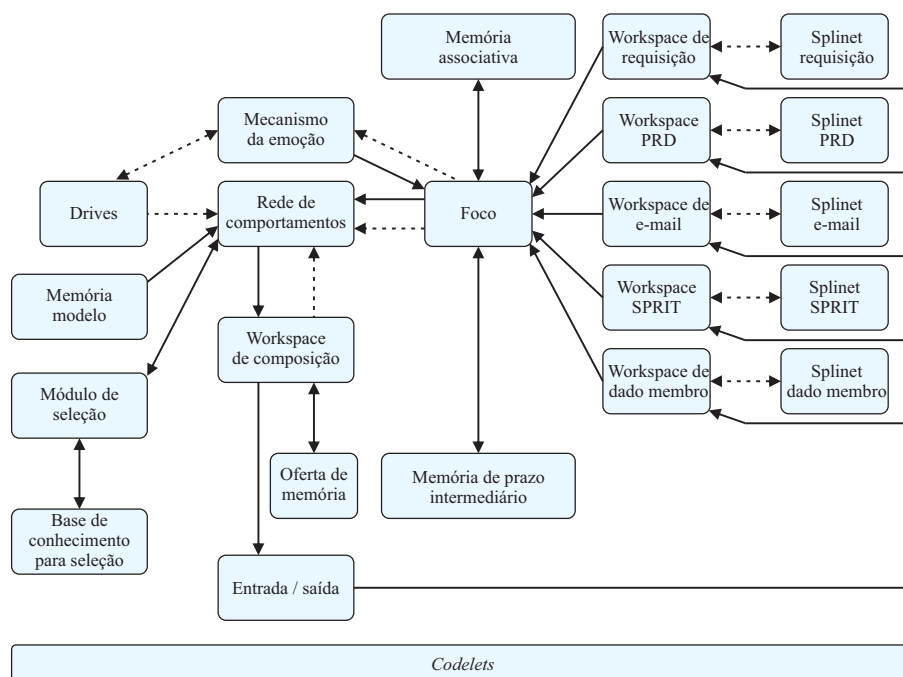


Figura 15 - Modelo pré-consciente IDA. As setas sólidas significam transferência de dados. As tracejadas indicam potencial ativação do alvo, e podem ocorrer com transferência de dados. Adaptado de (FRANKLIN; KELEMEN; MCCAULEY, 1998).

ciclo
cognitivo

Uma das contribuições importantes do IDA foi o estabelecimento de um **ciclo cognitivo**, uma combinação de papéis especializados para produzir motivação e facilitar o aprendizado (FRANKLIN, 2005):

1. **Percepção:** Estímulos sensoriais (internos ou externos) são recebidos e interpretados pela percepção produzindo significado. O modelo prevê *seleção antecipada* realizada através de *codelets perceptuais*. Esta sistema de memória perceptual identifica sentimentos/emoções relacionados com objetos, categorias e suas relações;
2. **Percepção para buffer pré-consciente:** A percepção, incluindo alguns dos dados e seus

significados, são armazenados em *buffers pré-conscientes* da memória de trabalho. Os sentimentos e emoções são parte da percepção pré-consciente;

3. **Associação local:** Utilizando as percepções recebidas e o conteúdo residual dos *buffers pré-conscientes* que incluem o conteúdo emocional, associações locais são automaticamente lembradas da memória episódica e da memória semântica;
4. **Competição para a consciência:** *codelets atencionais* observam a memória e selecionam eventos relevantes para a consciência. Os diversos *codelets* competem pelo acesso à consciência;
5. **Broadcast consciente:** Uma coalizão de *codelets* ganha acesso ao *espaço de trabalho global* e tem seu conteúdo enviado por *broadcast*;
6. **Recrutamento de recursos:** *codelets comportamentais*, com variáveis que podem ser ativadas pela informação no *broadcast consciente* respondem ao chamado;
7. **Configuração da hierarquia do contexto do objetivo:** os processadores recrutados usam os conteúdos da consciência, incluindo sentimentos e emoções, para instanciar novas hierarquias de contexto de objetivo, vinculando suas variáveis e ampliando sua ativação. Os sentimentos e emoções, portanto, geram motivações;
8. **Seleção da ação:** A rede de comportamentos seleciona um único comportamento dentre os disponíveis;
9. **Execução da ação:** A execução de um comportamento resulta nos *codelets comportamentais* realizarem sua ação especializada, que pode ter conseqüências internas, externas ou ambas.

LIDA Com a evolução desta arquitetura, os autores propuseram o **LIDA** (*Learning IDA*), buscando inserir aspectos relacionados ao aprendizado no sistema (BAARS; FRANKLIN, 2009).

CERA-CRANIUM Em outra linha de trabalho, alguns autores apresentam modelos computacionais utilizando redes neurais pulsadas (*spiking neural networks*) (SHANAHAN, 2008). Um modelo modular – denominado CERA, e, posteriormente CERA-CRANIUM – com capacidades cognitivas, tais como atenção, conhecimento do ambiente e aprendizado emocional, aplicado ao controle de um agente robótico é apresentado em Moreno, Espino e Miguel (2007) e Arrabales, Ledezma e Sanchis (2009d) baseado no competição e colaboração de processos especializados. O CERA-CRANIUM é composto por dois componentes principais: *i*) uma arquitetura de controle estruturada em camadas, denominada CERA (*Conscious and Emotional Reasoning Architecture*), e *ii*) uma ferramenta para a criação e gerenciamento de processos paralelos em espaços de

trabalho compartilhados, denominada CRANIUM (*Cognitive Robotics Architecture Neurologically Inspired Underlying Manager*). Uma visão geral desta arquitetura é apresentada na [figura 16a](#).

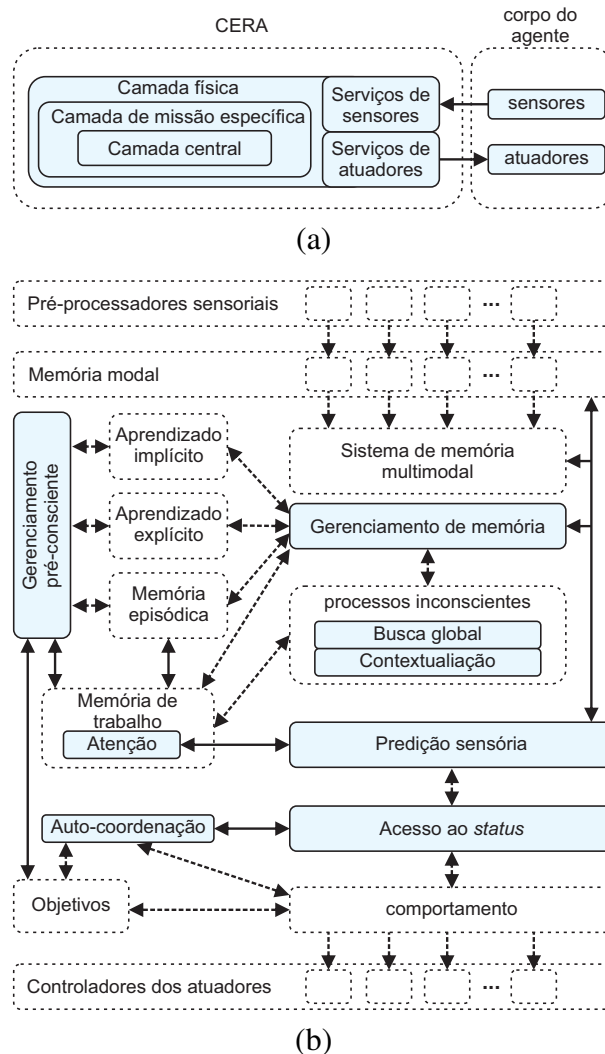


Figura 16 - Arquitetura do CERA. a) Camadas do CERA; b) Módulos da camada central do CERA. Diagramas em caixas sólidas azuis são módulos centrais do CERA. Linhas tracejadas representam camadas instanciadas, que dependentes do domínio. Linhas pontilhadas representam camada física. Adaptado de (MORENO; ESPINO; MIGUEL, 2007) e (ARRABALES; LEDEZMA; SANCHIS, 2009d).

CERA

O CERA é estruturado em quatro camadas:

- **Camada de serviços sensório-motor:** conjunto de serviços que interagem e se comunicam e que implementam um acesso requerido à leitura de sensores e/ou comandos dos atuadores. Todo sensor e atuador, portanto, possui um serviço associado;
- **Camada física:** compreende as representações de baixo nível dos sensores e atuadores. Realiza a preparação e pré-processamento de dados e tarefas análogas para os atuadores;

- **Camada específica de missão:** produz e gerencia conteúdo sensório-motor elaborado relacionado aos comportamentos vitais do agente e missões¹. Conteúdos adquiridos e pré-processados são combinados em porções mais complexas de conteúdo que tenham algum significado específico relacionado com os objetivos do agente;
- **Camada central:** compreende um conjunto de módulos que realizam funções cognitivas superiores. Segundo os autores, a definição e interação desses módulos pode ser ajustada de forma a implementar um modelo cognitivo de máquina consciente qualquer. Contudo, identificam os seguintes módulos: atenção, acesso a *status* do agente, gerenciamento pré-consciente, gerenciamento de memória e auto-coordenação.

CRANIUM provê uma biblioteca de *software* no qual o CERA pode executar milhares de processos assíncronos concorrentes, porém coordenados. Segundo os autores, o **CRANIUM** é inspirado na forma de trabalho do cérebro, onde a informação de processos de regiões especializadas é combinada dos sentidos ou de regiões especializadas. Ainda segundo os autores, de acordo com a **hipótese do acesso global** (BAARS, 2002), conexões neurais entre áreas especializadas tornam possível a emergência da coordenação global. A **figura 16b** apresenta um exemplo de estruturação da camada interna do CERA.

Um modelo que também apresenta similaridades com o GWT é apresentado por **CHELLA e GAGLIO (2012)**, investigando uma arquitetura cognitiva que parte da hipótese de que um *buffer* de grande dimensão na base do sistema e percepção gera a síntese da fonomenologia do robô.

6.3 Modelos baseados na integração de informação

A teoria de integração da informação – também denominada teoria de integração da consciência – foi formulada por **Tononi (2004)**. Esta família de arquiteturas tem como hipótese principal que a consciência corresponde à capacidade do sistema de integrar informações, um análogo às regiões individuais do cérebro que interagiriam de uma forma construtiva, levando à consciência. A troca de informação na arquitetura seria, portanto, o ponto fundamental para a consciência. O conceito de informação, no contexto desta abordagem, refere-se à teoria clássica de informação (SHANNON, 1948).

Tononi (2008) teria se inspirado nos princípios da *diferenciação e integração*. A abordagem sugere ainda que a quantidade de consciência disponível a um sistema poderia ser medida através de um valor Φ , cujo cálculo pode ser feito através de uma metodologia discutida pelos autores. Na abordagem, cada experiência consciente particular é especificada pelo valor efetivo

¹uma missão tipicamente envolve alguns objetivos particulares (ARRABALES; LEDEZMA; SANCHIS, 2009d)

Φ , em um dado tempo, das variáveis mediando interações informacionais entre os elementos de um complexo. A qualidade da consciência seria determinada, portanto, por esta relação informacional entre os elementos do complexo.

Alguns autores têm proposto arquiteturas baseadas neste conceito para, por exemplo, realizar predições sobre a distribuição de estados fenomenológicos utilizando redes neurais pulsadas com 18.000 neurônios modelando emoções e imaginação para controlar os movimentos dos olhos de um robô virtual (GAMEZ, 2010).

espaço
de qualia

A consciência é, em suma, vista como a capacidade de um sistema de integrar informação em uma experiência unificada e de diferenciar dentre uma larga gama de estímulos (REGGIA, 2013). A teoria faz ainda uma outra suposição: que os elementos de um complexo que realiza troca de informações definem um espaço relacional abstrato chamado **espaço de qualia** (TONONI, 2004).

6.4 Modelos baseados no auto-conhecimento

auto-
ciência

Alguns autores argumentam que a auto-ciência (*self-awareness*) é um aspecto crucial da mente consciente (SEARLE, 2004b) (SAMSONOVICH; NADEL, 2005). Um esboço de uma teoria que aborde a realidade sob a perspectiva da primeira pessoa é apresentada em (METZINGER, 2000).

imagem
do corpo

Os auto-modelos são biologicamente inspirados nos padrões de ativação neurais presentes no cérebro. Baseiam-se na representação interna das propriedades espaciais do corpo – a **imagem do corpo** (*body image*). Segundo Reggia (2013), os pesquisadores têm elaborado diversas suposições sobre como a experiência da consciência subjetiva poderia emergir de agentes com auto-modelos. As principais hipóteses são: *i*) agentes que raciocinam sobre si mesmos podem raciocinar sobre seu próprio raciocínio sobre si mesmos (PERLIS, 1997); *ii*) a consciência fenomenológica emergiria dos mecanismos de raciocínio introspectivo sobre a percepção (McDERMOTT, 2007), e *iii*) algumas das propriedades-chave do auto-conceito (existência, continuidade sobre o tempo, sobrevivência em um substrato físico, etc.) poderiam formar as bases para a auto-modelagem de agentes inteligentes (SAMSONOVICH; ASCOLI, 2005). Argumenta-se ainda que a introdução de auto-conceito durante o aprendizado poderia ser o evento-chave para criar *consciência computacional* em arquiteturas cognitivas apropriadas (SAMSONOVICH; DEJONG, 2004). Outros autores sugerem que os *qualia* seriam um efeito-colateral da existência de um componente executivo ou meta-gerencial como parte de uma máquina virtual que permita à mesma examinar o funcionamento de seu próprio estado interno e representações (SLOMAN; CHRISLEY, 2003).

CRONOS

É possível encontrar na literatura exemplos da implementação desta arquitetura em robôs. Um deles é o robô CRONOS (HOLLAND; GOODMAN, 2003) (HOLLAND, 2007), que possui

auto-modelo e modelo do mundo externo. Argumenta-se que este processo reflete de forma qualitativa o conteúdo cognitivo da consciência humana de uma perspectiva funcional, e que a raiz da consciência fenomenológica humana está em um modelo equivalente a um auto-modelo interno da imagem do corpo no cérebro (HOLLAND, 2007) (REGGIA, 2013). Um outro modelo que foca na auto-consciência controlado por redes neurais recorrentes executando aprendizado supervisionado foi proposto por Katayama e Takeno (2011), Takeno () e Takeno (2013). O robô foi aplicado na predição de eventos futuros baseado na seqüência de percepções. Os autores argumentam que a consistência da cognição e comportamento geram consciência (TAKENO,). Em outro projeto, o autor desenvolveu um robô com a capacidade de auto-reconhecimento em um espelho (TAKENO, 2008).

6.5 Modelos baseados em representações de nível superior

Nesse tipo de modelo, geralmente postula-se que a atividade mental consciente usa um nível mais elevado de representação que a atividade inconsciente. Como observado por alguns autores (REGGIA, 2013), a distinção entre nível de representação alto e baixo pode assumir diferentes formas. Muitas dessas arquiteturas tem sido empregadas em sistemas robóticos.

CLARION O sistema CLARION (SUN, 1997) (SUN, 1999) (SUN,) (SUN; FRANKLIN, 2007) é uma arquitetura cognitiva com conhecimento declarativo e procedimental, mecanismo de controle cognitivo e objetivos. Ela é organizada em dois níveis: um nível superior consciente – um sistema de representação simbólica de alto nível – e outro inconsciente – conceitos são representados por padrões de atividade sub-simbólicos em uma rede neural.

CERA-CRANIUM Uma arquitetura baseada em consciência com cinco níveis de representação foi proposto por (KITAMURA; TAHARA; ASAMI, 2000). O nível interior é puramente reativo, enquanto o mais elevado é um módulo simbólico de regras que analisa movimentos do robô. Os autores argumentam que a consciência emerge no nível hierárquico superior do robô quando as tarefas requerem esforço cognitivo. Uma outra arquitetura dessa natureza é a denominada CERA-CRANIUM (já abordado na seção 6.2). Embora baseie-se no GWT, esta arquitetura também possui níveis cognitivos distintos, cabendo sua classificação sob ambas as óticas. Os autores sugerem que o *qualia* e a experiência consciente emergem nos níveis mais elevados da arquitetura cognitiva. Outra arquitetura de múltiplos níveis é apresentada em Chella (2007) e Chella, Frixione e Gaglio (2008). Nesta arquitetura, com três níveis, uma “área sub-conceitual” que realiza o processamento dos dados sensoriais, uma “área conceitual” de nível médio que organiza os dados sensoriais em categorias conceituais, e uma “área linguística” baseada no cálculo de predicados de primeira ordem (REGGIA, 2013)

6.6 Modelos baseados em Atenção

Diversos autores têm enfatizado nos últimos anos a proximidade entre os mecanismos da consciência e da atenção (POSNER; DEHAENE, 1994):

"Eu proponho discutir a questão da consciência à luz das recentes descobertas sobre as redes atencionais do cérebro humano que levam à seleção da informação sensorial, ativam ideias armazenadas na memória, e mantém o estado de alerta. Eu não acredito que qualquer um desses mecanismos seja consciente por si, assim como o DNA não é vida, mas eu acredito que um entendimento da consciência deve assentar na apreciação das redes cerebrais que subservem a atenção, da mesma forma que uma análise científica da vida sem considerar a estrutura do DNA pareceria vazia".

Posner e Dehaene (1994)

Face à inegável proximidade entre esses processos, algumas arquiteturas de modelos conscientes têm sido propostas na literatura nos últimos anos tomando como base o mecanismo atencional. Uma visão geral das abordagens atencionais para esta finalidade pode ser encontrada em Reggia (2013). Muitas das abordagens utilizam a estratégia da aplicação da engenharia reversa em sistemas inteligentes e conscientes, gerando modelos bio-inspirados (GRAY; BUHUSI; SCHMAJUK, 1997). Alguns autores têm demonstrado grande preocupação com a arquitetura destes sistemas, sugerindo que quando a razão entre comportamentos aprendidos e recursos disponíveis – tais como memória, por exemplo – aumenta, as arquiteturas tendem a se tornar mais e mais limitadas com evidentes implicações para a consciência (COWARD; GEDEON, 2009).

Quando observamos os diferentes modelos e arquiteturas propostas para a consciência em agentes, é imediato observar que o conceito de consciência não é homogêneo entre os autores. Um modelo computacional – voltado à percepção visual – constituído por camadas de neurônios pulsados é apresentado em Tinsley (2008). Esses neurônios, observando algumas características distintas das imagens de entrada, seriam capazes de construir **supermapas** através da alteração de suas atividades. Os padrões de saída desses mapas poderiam, segundo o autor, ser usados como uma representação da consciência. Nesta abordagem, consciência e atenção se misturam.

Em outra abordagem (HAIKONEN, 2003) (HAIKONEN, 2007b) (HAIKONEN, 2007a), vários módulos distintos interconectados sem uma hierarquia definida recrutam outros módulos através de um processo competitivo para que atendam ao mesmo tópico que estão processando. A máquina se tornaria consciente sobre o tópico quando os vários módulos cooperarem - em uníssono - e focarem sua atenção nas percepções relativas a este tópico. Dentre os modelos

com processos decisórios centralizados, uma abordagem comum é que para que a consciência ocorra é necessário que o agente tenha um sistema executivo que, direcionando sua atenção, seja capaz de monitorar todos os aspectos do comportamento da máquina, incluindo aspectos inconscientes (STARZYK; PRASAD, 2010) (STARZYK; PRASAD, 2011). Nesses modelos, que mantém uma forte relação com o *Global Workspace Theory*, o controle do chaveamento da atenção é um mecanismo central para a seleção da consciência.

modelo
de
Starzyk

O modelo proposto por (STARZYK; PRASAD, 2010) (STARZYK; PRASAD, 2011) é apresentado na figura 17. Trata-se de uma arquitetura com três blocos principais:

- **Bloco sensório-motor:** é composto de três partes: *i*) processamento sensório integrado com memória semântica; *ii*) processamento motor integrado com habilidades motoras, e *iii*) processador sub-cortical integrado com emoções e recompensas;
- **Memória episódica e bloco de aprendizado:** é composto por duas partes: *i*) unidade de memória episódica, uma coleção de blocos funcionais menores, cada um dedicado a capturar uma seqüência espaço-temporal de relações semânticas como relações entre objetos observados em uma experiência episódica com sua significância derivada do contexto emocional. Para isso, a unidade de organização dos episódios coleciona dados de todas as unidades, incluindo percepções, motivações e interpretações de experiência cognitiva e sua significação da central executiva, e *ii*) unidade de aprendizado, que aprende sobre eventos/padrões específicos se direcionada pela central executiva;
- **Bloco da central executiva:** responsável pela coordenação e controle seletivo de todas as outras unidades. Para realizar este propósito, a central executiva precisa da capacidade de dinamicamente selecionar e direcionar a execução de programas que governam a atenção, motivação, memória episódica e monitoramento da ação. Pode ativar a memória semântica e controlar emoções. Embora direcione os aspectos cognitivos da máquina, suas operações são influenciadas por sinais – competitivos – das motivações, desejos e chaveamento da atenção. O vencedor da competição interna dirige o processo de pensamento cognitivo até que seja substituído por outro vencedor através do chaveamento da atenção e alteração na motivação para agir.

O chaveamento da atenção no modelo é realizado como um resultado da competição entre vários tipos de sinais (sinais de pensamentos internos, novas percepções devido a alterações no mundo exterior, alterações motivacionais).

tracker Considerando agentes que utilizem métodos de inferência e armazenamento de memória simbólico, ponteiros (ou *trackers*) poderiam relacionar símbolos e porções do fluxo de dados sensórios, mantendo uma correspondência temporal entre o nível simbólico e o sensório. A consciência é interpretada como a parte do mecanismo atencional que relaciona a representação

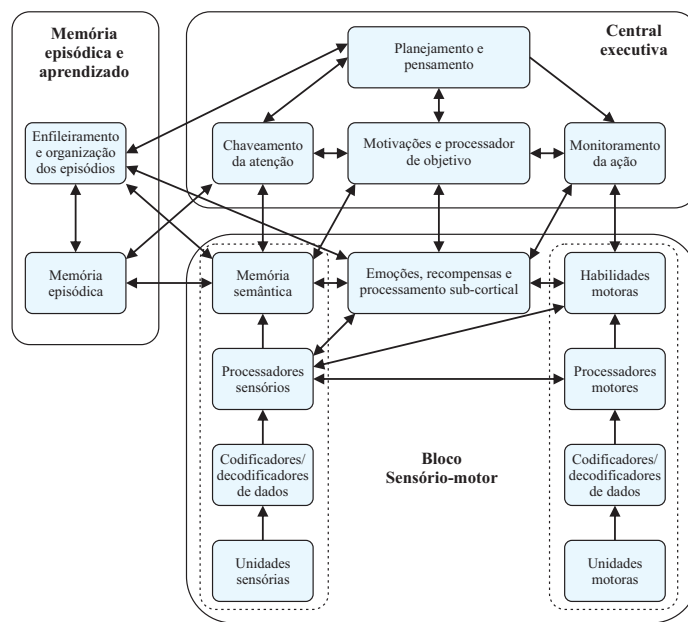


Figura 17 - Modelo computacional para uma máquina consciente. Adaptado de (STARZYK; PRASAD, 2011).

simbólica com o fluxo de dados (KUIPERS, 2005) (KUIPERS, 2008). O autor apresenta a hipótese de que qualquer sistema organizado desta forma, com componentes atencionais *top-down* e *bottom-up* que criem esses *trackers*, seria genuinamente uma consciência instanciada, com experiências subjetivas correspondentes ao qualia e auto-consciência (KUIPERS, 2005).

Um modelo que integra consciência, atenção e **criatividade** é apresentado em Taylor (2010). Segundo o autor, a criatividade poderia ser usada pelos sujeitos para ampliar suas experiências subjetivas. A manifestação da criatividade no modelo se dá através de uma redução da intensidade do processo atencional, permitindo assim a geração de pensamentos inconscientes até levar a um estado mental que tenha reconhecido valor na solução do problema. O modelo busca manter a consciência do agente sobre o estímulo sensorial sem que a atenção do modelo seja direcionada para esse estímulo.

Em uma análise geral das arquiteturas propostas para agentes conscientes baseados em mecanismos atencionais, é possível identificar duas famílias distintas: *i*) arquiteturas com modelos estruturados e *ii*) arquiteturas com modelos não estruturados. Dentre os modelos estruturados, é possível separar os modelos em grupos que possuem processamento central hierarquizado e os que não possuem. A figura 18 apresenta a **taxonomia dos modelos de sistemas conscientes baseados em mecanismos atencionais** considerando a natureza da arquitetura implementada em cada caso. A tabela 2 apresenta um sumário das diferentes visões sobre a consciência presentes nos modelos atencionais propostos por diferentes autores.

Modelo	Relação entre consciência e atenção
(TINSLEY, 2008)	Consciência se confunde com atenção
(HAIKONEN, 2007b)	Consciência é o foco da atenção sobre os módulos do sistema
(STARZYK; PRASAD, 2010)	Consciência é o foco da atenção sobre o monitoramento de aspectos do comportamento do sistema
(KUIPERS, 2005)	Consciência é a parte do mecanismo atencional que relaciona representação simbólica com dados sensoriais da percepção
(TAYLOR, 2010)	Subconsciente permite manifestação de pensamentos criativos quando processo atencional é reduzido

Tabela 2 - Diferentes relações entre consciência e atenção em vários modelos para agentes conscientes atencionais.

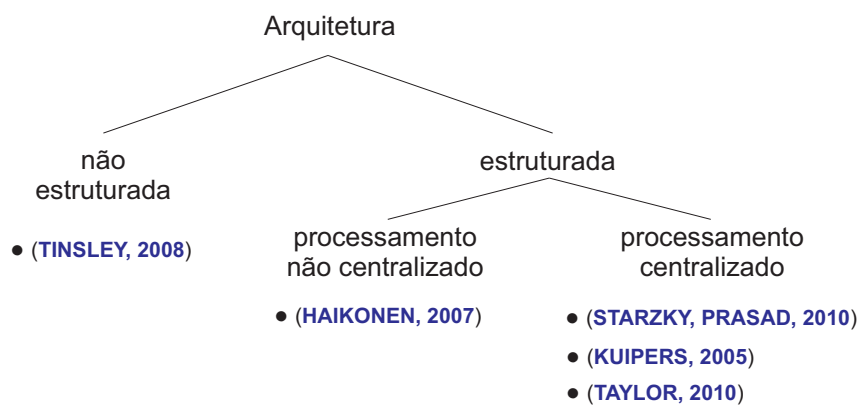


Figura 18 - Taxonomia dos sistemas conscientes baseada em modelos atencionais considerando a natureza da arquitetura implementada.

7 OUTROS MODELOS COGNITIVOS

“Não é a consciência do homem que lhe determina o ser, mas, ao contrário, o seu ser social que lhe determina a consciência”.

Karl Marx

Este capítulo concentra outros modelos cognitivos que mantêm relação com aspectos ou conceitos deste trabalho, e/ou que servem como inspiração para a proposição de arquiteturas conscientes. O capítulo encontra-se dividido da seguinte forma:

- [Seção 7.1](#): Apresenta o modelo ISAC;
- [Seção 7.2](#): Discorre sobre o modelo EPIROME;
- [Seção 7.3](#): Destaca o modelo SOAR;
- [Seção 7.4](#): Introduce o *framework* de Samsonovich;
- [Seção 7.5](#): Discorre sobre o diagrama de relações de Stachowicz.

7.1 O ISAC

ISAC O **ISAC** (*Intelligent Soft Arm Control*) (DODD; GUTIERREZ, 2005) é um robô humanoide cujo propósito é trabalhar de forma segura com humanos. O foco de estudo dos autores é a memória, mais especificamente a memória episódica e sua capacidade de recordar conceitos usando estatísticas e emoções. Contudo, o sistema apresenta uma arquitetura cognitiva com diversos elementos, tais como: memória, planejador e controle baseado em emoções. Uma representação esquemática desta arquitetura pode ser vista na [figura 19](#).

noção de
si mesmo

A **noção de si mesmo** (*sense of self*), segundo os autores, é representada em um agente composto – *agente em si* (*self agent*) – que consiste de alguns agentes simples. Esse módulo compreende os sistemas de planejamento, controle executivo, emoções, auto-monitoramento e

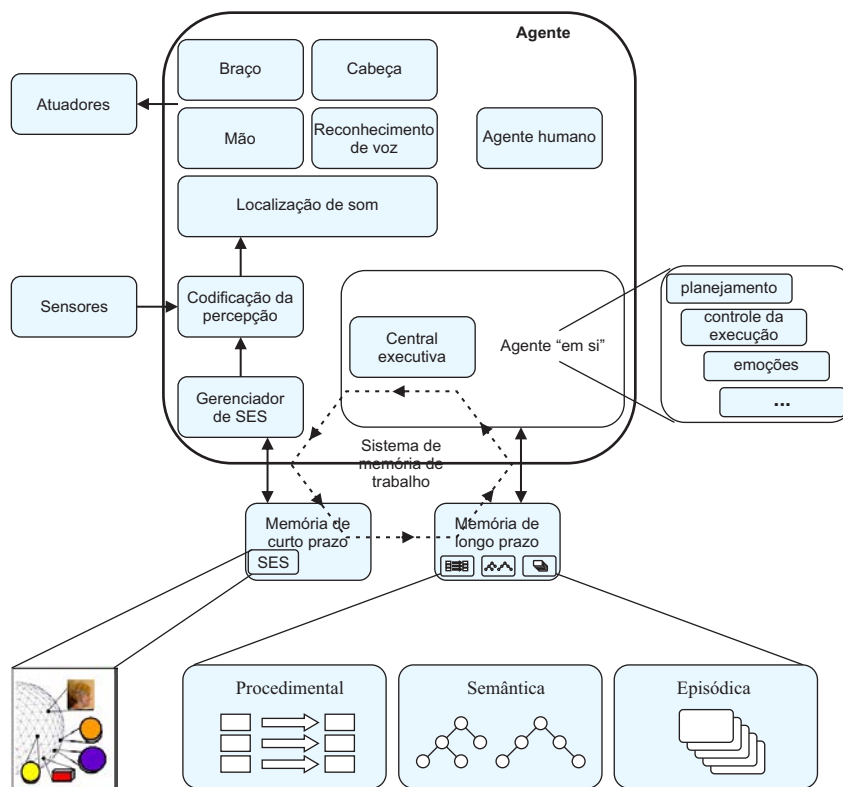


Figura 19 - Modelo do robô ISAC. Adaptado de (DODD; GUTIERREZ, 2005).

outras funções cognitivas. Um componente do *agente em si* é chamado *central executiva*, que retira informações da memória através de um programa ativo de armazenamento de memória (a memória de trabalho). A memória de trabalho monitora as necessidades da *central executiva* do agente, os conteúdos das memórias, e a saída do sistema de percepção para preencher a si mesma com as partes mais relevantes de dados, ou **chunks**. Cada memória tem seu próprio método para recordar seus *chunks*. Uma estrutura chamada **egosfera sensória** (*Sensory EgoSphere - SES*) armazena os dados da memória de curto prazo. A estrutura de dados do SES serve como uma memória de curto prazo espaço-temporal para o agente. A memória de longo prazo é dividida em procedural, episódica e semântica.

No modelo proposto, cada episódio contido na memória episódica contém um *link* para os conteúdos inteiros da memória de trabalho e para o estado emocional do agente durante a tarefa. Todos os *links* apontam para unidades da memória semântica. Desta forma, estatísticas podem ser calculadas sobre a frequência de uso de cada elemento da memória semântica. A tarefa da memória episódica é preencher automaticamente a memória episódica / memória de trabalho com os episódios corretos para uma dada situação. Segundo os autores, 5 (cinco) fatores são preponderantes na medida de relevância para os itens de memória (DODD; GUTIERREZ, 2005):

1. A memória recordada deve conter elementos da memória semântica que se assemelhem

à situação corrente;

2. Episódios recentes e/ou comumente acessados são mais úteis do que os antigos ou menos acessados;
3. Similaridades originais devem pontuar mais do que similaridades comuns com episódios;
4. O sistema executivo deve ser capaz de orientar sua atenção para certos elementos sugeridos;
5. Memórias emocionalmente salientes devem pontuar mais do que as que produziram pouca emoção.

Um modelo para o cálculo da probabilidade de que um item de memória seja relevante em um dado contexto é apresentado pelos autores. Memórias associadas a **saliências emocionais** mais fortes têm sua importância decaindo mais lentamente do que outras. O sistema emocional do agente baseia-se no seguinte conceito: emoções poderiam ser consideradas como combinações interativas de reações básicas perceptíveis do sistema. As sensações elementares do ISAC são: bom, mau, dor e prazer. O sistema apresenta um modelo comportamental com ações básicas associados a cada estado emocional do agente.

7.2 O EPIROME

O robô **TASER** (JOCKEL; WESTHOFF; ZHANG, 2007) (JOCKEL et al., 2008), uma plataforma robótica móvel multimodal para serviços, tem por objetivo colecionar memórias autobiográficas para melhorar suas ações de planejamento baseado em experiências passadas. Para tanto, os autores propõem o *framework* EPIROME, utilizado para representar memória episódica de alto nível, implementado através de eventos classificados segundo:

1. **Eventos de comando:** são especificações com relação à habilidade de produzir movimentos pela interação entre a unidade de controle e os atuadores. Esses comandos são normalmente: rotacionar, transladar, abrir, etc.;
2. **Eventos de percepção:** esses eventos focam no reconhecimento e interpretação de estímulos sensoriais. Alguns exemplos são: percepção de localização espacial, percepção do estado de travamento (*stuck*), etc.;
3. **Eventos executivos:** esta categoria de eventos contém um tipo de habilidade de alto nível que pode ser entendida como procedimento, sendo denominadas, portanto, funções executivas. São necessariamente guiadas por um objetivo. Exemplos são: manipulação, planejamento, executar um objetivo, etc.

Conforme observado por [Stachowicz e Kruijff \(2012\)](#), o EPIROME não contempla todos os requisitos para um sistema que atua como uma memória episódica (discutidos na [seção 4.2.1](#)).

7.3 O SOAR

A arquitetura SOAR (*State Operator and Result*) ([LAIRD; NEWELL; ROSENBLOOM, 1987](#)) ([NUXOLL; LAIRD, 2004](#)) ([NUXOLL; LAIRD, 2007](#)) ([BANSAL; RAJAN; SRINIVASAN, 2014](#)) foi proposta com o intuito de servir como um sistema genérico de comportamento inteligente. Com base no uso de estruturas de memória semântica como o proposto pela arquitetura cognitiva **ACT-R** ([ANDERSON, 2004](#)), alguns autores ([WANG et al., 2006](#)) ([WANG; LAIRD, 2007](#)) propuseram a integração do SOAR a uma estrutura de memória semântica, ampliando seu poder de atuação. A arquitetura resultante passou a contar, então, com memórias de curto prazo, longo prazo, memórias com armazenamento de informação na forma declarativa, procedimental, dentre outras variantes. Dentre as habilidades do SOAR estão ([STACHOWICZ; KRUIJFF, 2012](#)): observar novas situações, detectar repetição, sensação virtual (recordar entradas sensoriais passadas para auxiliar novas situações), predição de efeitos de suas próprias ações sobre outros eventos percebidos, gerenciamento de objetivos de longo termo, aprendizado retroativo (reanálise em um momento de baixa carga de trabalho das experiências adquiridas em um momento de pressão temporal) e reanálise de conhecimento.

A arquitetura base do sistema é apresentada esquematicamente na [figura 20](#). A arquitetura se caracteriza pela presença de memória com as seguintes características:

- **Memória de trabalho:** As representações são na forma estática declarativa. Cada elemento é usualmente uma tripla: identificador, atributo e valor (o valor, por sua vez, pode ser um outro identificador ou uma constante);
- **Memória procedimental:** Múltiplas regras de produção podem disparar em paralelo em uma mesma fase (de elaboração ou aplicação de operador);
- **Memória episódica:** Armazena um *snapshot* das imagens na memória de trabalho que referem-se à experiência do agente;
- **Memória semântica:** Armazena grupos de pares valores-atributo descrevendo um conceito do objeto.

Segundo [Stachowicz e Kruijff \(2012\)](#), cada vez que o SOAR toma uma ação um episódio é armazenado contendo dados da memória de trabalho, o que resulta em episódios muito menores do que, por exemplo os do ISAC (discutido na [seção 7.1](#)). O mecanismo de recordação

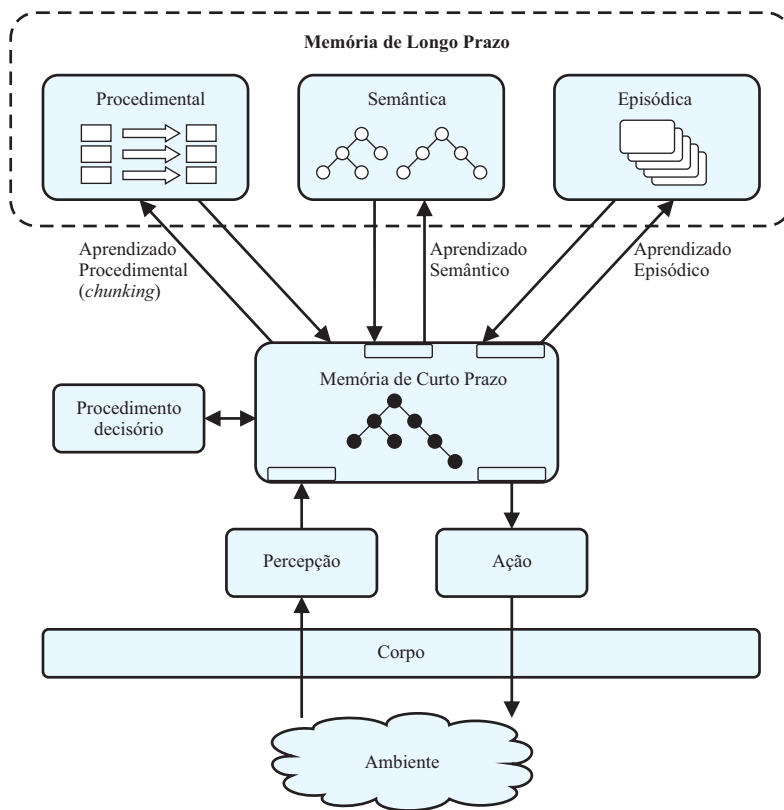


Figura 20 - Versão esquemática da arquitetura SOAR com memória semântica, episódica e mecanismos de aprendizado correspondentes. Adaptado de (WANG; LAIRD, 2007).

do SOAR também é considerado mais flexível do que o do ISAC, uma vez que o SOAR não recorda exclusivamente e automaticamente memórias similares à situação corrente. Ao contrário, ele permite recordação deliberativa deixando um *palpite de recordação* (*retrieval cue*) para a memória de trabalho.

7.4 O framework de Samsonovich

Samsonovich e Nadel (2005) propõem um *framework* com uma forma computável e objetiva para representar uma experiência. Segundo eles, ao “ver uma árvore”, por exemplo, o conteúdo do estado experimentado a ser armazenado poderia ser simplesmente “*existe uma árvore*”, se não houvesse o componente da auto-ciência. Ao considerar a auto-ciência, seria necessário adicionar ao “eu” sua experiência como um rótulo, atribuindo o significado que esta experiência conscientemente ou automaticamente atribuiu ao “eu”. Ao considerarmos alguém mais observando esta árvore, o rótulo seria diferente: deveria haver um “ele”. Desta forma, o complexo inteiro envolve dois componentes: *i*) um representando o conteúdo da experiência, e *ii*) outro representando o sujeito que está experimentando este conteúdo. Este conjunto, segundo o autor, seria um **estado mental**.

A [figura 21a](#) apresenta um *snapshot* de um estado corrente de atenção contendo experiências como: “*eu ando*”, “*estou entrando em um jardim*”. Este estado rapidamente se alteraria para outro estado com experiências como: “*eu ouço pássaros*”, “*eu estou em um jardim*”. Esta condição pode evoluir nesse instante de tempo ocasionando minha re-experimentação do episódio recordado a partir de minha perspectiva passada (o “*eu antes*”), que pode forçar o “*eu agora*” a pensar sobre os meus planos futuros e então focar neles no próximo momento de tempo (o “*eu depois*”), e assim sucessivamente.

Este modelo é completado, segundo os autores, ligando-se todas as experiências do sujeito e instâncias dele relacionadas rotuladas de uma forma auto-explicativa como segue: “*eu antes*”: unificando tudo sobre o que eu estava atento até um momento atrás; “*eu depois*”: unificando meu estado esperado de atenção no próximo instante de tempo; “*eu passado*”: apontando para um conjunto de estados mentais que aconteceram juntos para constituir uma experiência em primeira pessoa sobre mim, e assim sucessivamente. Segundo os autores, embora esses estados refiram-se a vários momentos no tempo, eles são considerados como ativos simultaneamente no cérebro. O tempo atribuído a eles é um tempo subjetivo, e não o tempo físico de sua manifestação no cérebro.

O *framework* sugere ainda esquemas para a recordação de informações na memória de trabalho e a relação desta com a memória semântica, apresentadas na [figura 21b](#). Os autores propõe a adoção de relações hierárquicas entre perspectivas (*charts*). Os conteúdos desses *charts* seriam, por sua vez, estados mentais. Em grande escala haveria uma organização entre os *charts*: o “*eu passado*”, o “*eu agora*” e o “*eu anterior*” seriam consistentes entre si. Existiria uma linha através da seqüência de *charts* levando a algum tipo de objetivo mental específico (o “*eu objetivo*”). A linha mais grossa indicaria um **cenário de trabalho**, e, adicionalmente a ele, haveriam outros *charts* (não apresentados na figura) presentes na memória. Esse esquema seria uma generalização do trabalho de [Nichols e Stich \(2000\)](#) e [Nichols e Stich \(2003\)](#).

cenário
de
trabalho

7.5 O diagrama de relações de Stachowicz

[Stachowicz e Kruijff \(2012\)](#) propuseram um diagrama de relações de entidades destinado ao desenvolvimento de uma memória como a episódica para robôs cognitivos, apresentado na [figura 22](#). O estudo para elaboração do diagrama levou em consideração os análogos biológicos das estruturas, bem como os requisitos técnicos de um robô cognitivo, resultando em um *framework* formal que originou o diagrama. A estrutura central do diagrama é o *evento*, retratado com seus atributos *tempo* (*início e fim* do evento), *localização* e *dados binários específicos* do evento. Um identificar único *IDdoEvento* é necessário, segundo os autores, principalmente para a construção de relações com o evento, tais como, por exemplo, quais eventos fariam parte de um evento complexo. De um modo geral, todo evento é de um *tipo*, e tipos de evento

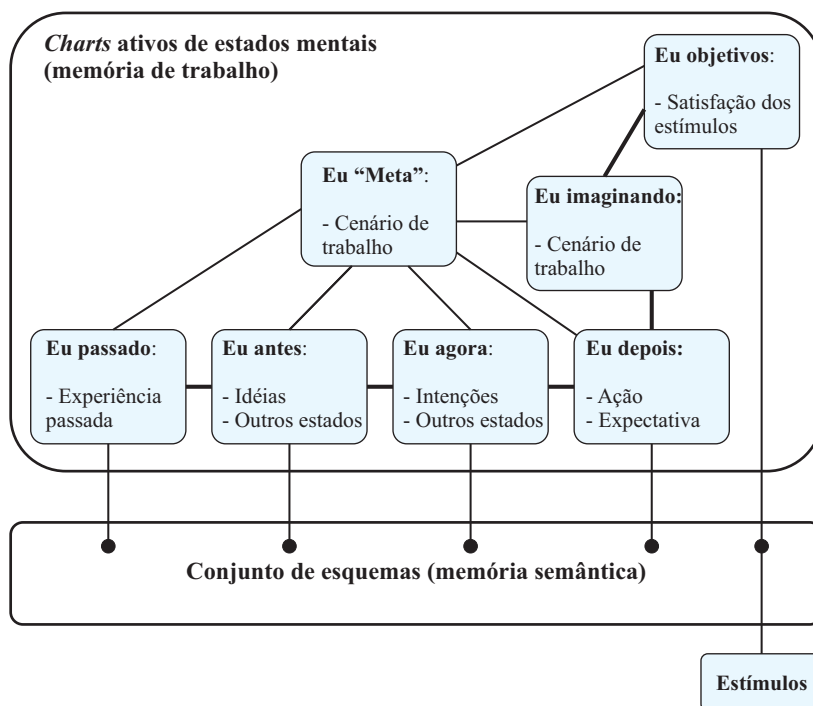
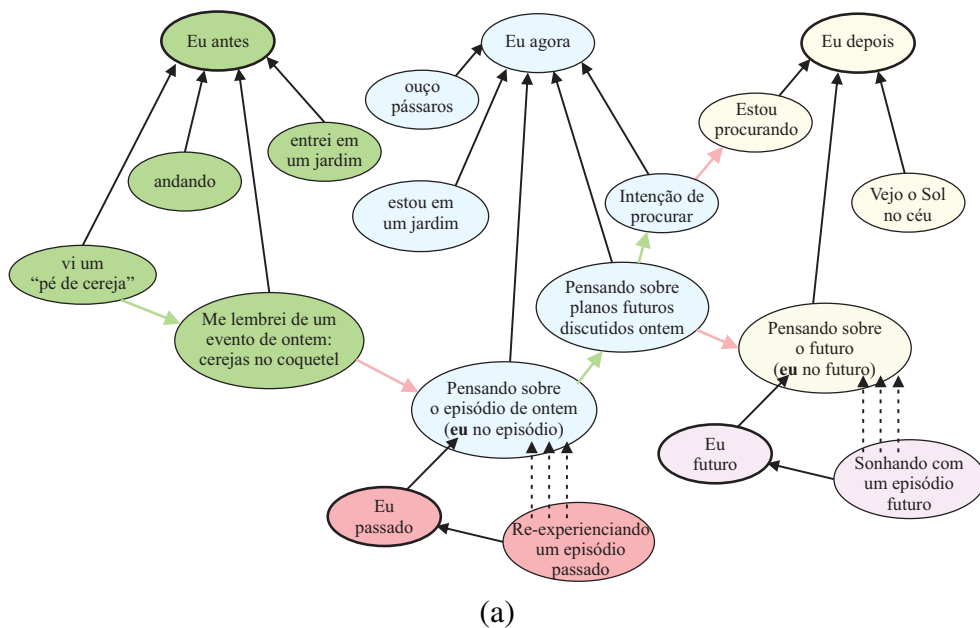


Figura 21 - O *framework* de Samsonovich. a) Um possível fragmento de um *snapshot* de um estado da memória de trabalho, vendo-os como constituindo estados mentais. A linha mais grossa constitui um *cenário de trabalho*. b) Sistema de *charts* ativos que constituem o cenário de trabalho (conectados pela linha mais grossa) e o conjunto de esquemas. Adaptado de (SAMSONOVICH; NADEL, 2005).

formam uma hierarquia. Finalmente, eventos podem ter *nomes* e *características específicas* sobre os quais o cruzamento de dados (*matching*) também é possível. Com essas características, segundo os autores, a memória episódica poderia obedecer ao princípio de “*recordando qualquer característica automaticamente recorda-se de outras características*”, e robôs utilizando a arquitetura proposta foram capazes de acumular memórias por várias horas (STACHOWICZ; KRUIJFF, 2012).

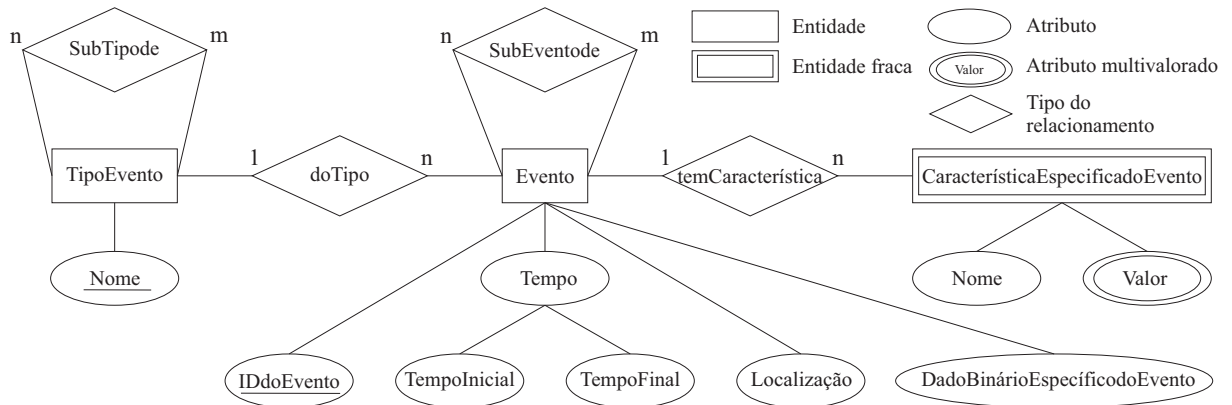


Figura 22 - Diagrama de relações de entidades em uma memória episódica básica. Adaptado de (STACHOWICZ; KRUIJFF, 2012).

PARTE III

TRABALHO PROPOSTO

8 REVISITANDO O PROBLEMA DA CONSCIÊNCIA

"Minha própria visão é que todas essas escolas [teorias sobre a consciência] estão corretas quando vistas em conjunto, mas [são] insuficientes quando vistas uma de cada vez".

Kurzweil (2007, p. 94)

Esta seção busca revisitar as questões da consciência e consciência de máquina sob as perspectivas levantadas nas seções anteriores, buscando desta forma estabelecer diretrizes que possam nortear o tratamento da questão.

As primeiras informações disponíveis – do século VIII ou IX AC – demonstram a ausência de uma separação inata no homem entre os conceitos de corpo e mente. Foi somente com o desenvolvimento da filosofia que a separação desses conceitos ganhou impulso. O cefalocentrismo somente foi considerado no século V AC. Platão é um dos pioneiros na visão dualista do universo, mas tarde sustentada por nomes como Descartes.

Ao confrontarmos essa visão dualista com os conceitos atualmente aceitos no campo da física, essa visão não se sustentaria, uma vez que todo o universo seria constituído por uma única substância, que em última análise é integralmente constituída de partículas indivisíveis como os *quarks* e os *leptrons* (ARGONOV, 2012). De forma análoga, não há na teoria física até o momento argumentos que referendem um modelo pluralista do universo. À luz dos conhecimentos atuais, portanto, a física parece fortalecer a abordagem monista.

Mais do que isso, o **materialismo** é colocado em destaque: a matéria, com seus elementos subatômicos, seria a única substância do universo, e através dela seria possível explicar relação entre matéria e consciência. Uma vez que não existe nenhuma prova até o presente momento de que o pensamento esteja ligado essencialmente à matéria, um dos conceitos que tem recebido maior destaque nos últimos anos é o **funcionalismo**, que sustenta que os estados mentais corresponderiam aos estados do cérebro devido ao funcionamento deste, e não devido

a sua estrutura ou composição. O **computacionalismo** parte do princípio que as relações mentais seriam computáveis, fato controverso na literatura. Sucita-se, portanto, investigações que possam demonstrar ou refutar a possibilidade de que a consciência possa emergir devido ao funcionamento do aparato cerebral, bem como a possibilidade de que as relações mentais sejam computáveis.

A revisão sobre os fundamentos da consciência ([capítulo 4](#)) revela que não existe um entendimento único sobre esse conceito, bem como sobre diversos outros conceitos correlatos. A maior crítica deste autor às diversas definições para consciência apresentadas nos trabalhos prévios (discutidas na [seção 4.1](#)) é que estas, embora capturem a essência da problemática da consciência são demasiadamente fragmentadas, apresentando separadamente diferentes aspectos da consciência, e, portanto, tornando-se pouco propícias para a proposição de sistemas artificiais. Já quando tratamos os conceitos de consciência de máquina, muitos dos elementos fundamentais que estariam presentes em sistemas artificiais conscientes são apresentados na definição proposta por [Starzyk e Prasad \(2011\)](#). A maior crítica, nesse caso, pode ser feita observando-se que a sistemática de funcionamento desses elementos – que sob uma ótica funcionalista-computacionalista seriam fundamentais para a discussão da consciência de máquina – apresentam ainda baixo nível de formalização. As implementações computacionais e os experimentos formais detalhados são praticamente inexistentes na literatura da área. Nenhum modelo formal para uma máquina consciente foi encontrado. Os vários modelos existentes divergem ainda na filosofia fundamental que sustenta a manifestação da consciência. Além disso, embora alguns modelos apresentem alternativas para o **problema fácil** da consciência, não há menções na literatura de tratamento para a questão do **problema difícil**. Nesse cenário, estimular o desenvolvimento de arquiteturas que possam capturar aspectos fundamentais presentes nessas diferentes famílias de modelos computacionais parece ser um caminho a considerar. Em particular, a proposição de modelos que possam fornecer alternativas para fazer frente à questão das experiências subjetivas desponta como uma demanda da área.

atenção A **atenção**, discutida no [capítulo 3](#), é, inegavelmente, um paradigma próximo à consciência. Dentre os modelos atencionais avaliados no âmbito do presente trabalho (discutidos no [capítulo 5](#)), um modelo que se destaca é o de [Colombini \(2014\)](#) devido principalmente a sua capacidade de representar de forma unificada diversos dos conceitos existentes na literatura e por tratar-se de um modelo formal bem definido. Mais do que isso, a atenção naturalmente age como filtro, reduzindo drasticamente o espaço de estados do agente ([COLOMBINI, 2014](#)). Contudo, o controlador proposto para esse modelo – utilizado para efeito de demonstração do modelo atencional – mostra-se limitado no que diz respeito ao poder cognitivo gerado, uma vez que este não contempla aspectos essenciais da cognição, tais como: raciocínio, planejamento, comportamentos complexos, modelagem de estados emocionais, dentre diversos outros que poderiam ser incorporados ao agente.

modelo de Colombini

Nesse contexto, teoriza-se que seria possível a proposição de uma arquitetura cognitiva baseada nesse modelo atencional como plataforma para a consciência de máquina, e, por conseguinte, explorando a forte redução do número de dimensões de entrada do sistema, com diversos possíveis ganhos computacionais para o agente.

9 O MODELO PROPOSTO

“A formulação de um problema freqüentemente é mais essencial do que sua solução, que pode ser meramente uma questão de habilidade matemática ou experimental. Para que surjam novas questões, novas possibilidades, para considerar velhos problemas sobre um novo ângulo requer-se imaginação criativa que marca reais avanços na ciência. ”

Albert Einstein
citado em [Einstein e Infeld \(1966\)](#)

O presente capítulo apresenta um modelo para consciência de máquina e suas principais características, bem como algumas de suas primeiras implicações. O capítulo encontra-se dividido da seguinte forma:

- [Seção 9.1](#): Apresenta o conceito fundamental de consciência de máquina utilizado no âmbito deste trabalho e a hipótese que fundamenta o mesmo;
- [Seção 9.2](#): Apresenta o modelo proposto;
- [Seção 9.3](#): Realiza uma descrição formal do o modelo proposto;
- [Seção 9.4](#): Detalha a operação do modelo formal proposto;
- [Seção 9.5](#): Traz um sumário das características do modelo proposto;
- [Seção 9.6](#): Realiza um breve comparativo entre o sistema proposto e outros presentes na literatura.

9.1 Conceito e hipótese adotados

Considerando os diferentes axiomas e definições para consciência e consciência de máquina disponíveis na literatura, no âmbito deste trabalho, será adotada a seguinte definição:

“Consciência de máquina é o campo da Inteligência Artificial que, através da proposição de um conjunto de elementos organizados segundo uma arquitetura e de operações sobre esses elementos, busca dotar um agente construído com dispositivos artificiais da capacidade de expressar um conjunto específico de habilidades que inclui: atentividade, sciência, auto-ciência, auto-consciência, consciência auto-noética, possessividade, individualidade e perspectividade, construídas com base nas experiências do próprio agente ao longo do tempo”.

Esta definição reúne um conjunto de habilidades específicas que, segundo a literatura aqui apresentada, caracterizaria os seres conscientes segundo a ótica atual. O atendimento a esta definição, hoje, passa invariavelmente pela proposição de modelos que contenham elementos de tomada de decisão, memórias, elementos de avaliação de desempenho, funções e processos cognitivos conscientes e inconscientes. A forma como os elementos do sistema artificial se relacionam é enfatizada. Ainda, o foco na aquisição e construção de experiências é trazido à tona: como as experiências do agente poderiam ser adquiridas, construídas e manipuladas utilizando esses mecanismos? Considerando a necessidade da proposição de métodos para, dinamicamente, lidar com essas experiências, e considerando a relação inata desta definição com a atenção, a seguinte hipótese foi levantada:

Hipótese \mathcal{H}_1 : Existe um modelo – composto por uma arquitetura em um conjunto de operações sobre os elementos desta arquitetura – que atende à definição de consciência de máquina aqui apresentada de forma a permitir a aquisição e manipulação de experiências pelo agente, capaz de operar sobre a arquitetura atencional proposta por Colombini – em substituição a seu controlador atencional – computando sobre o mapa de saliências de forma a reduzir significativamente o espaço de dimensões de entrada do modelo.

As seções a seguir apresentam o modelo proposto para atender à hipótese \mathcal{H}_1 .

9.2 Apresentação do modelo

O modelo é composto por um conjunto de elementos dispostos em uma arquitetura e de uma descrição formal da operação dessas estruturas. A arquitetura proposta é apresentada na [figura 23](#).

A arquitetura compreende dois sistemas principais:

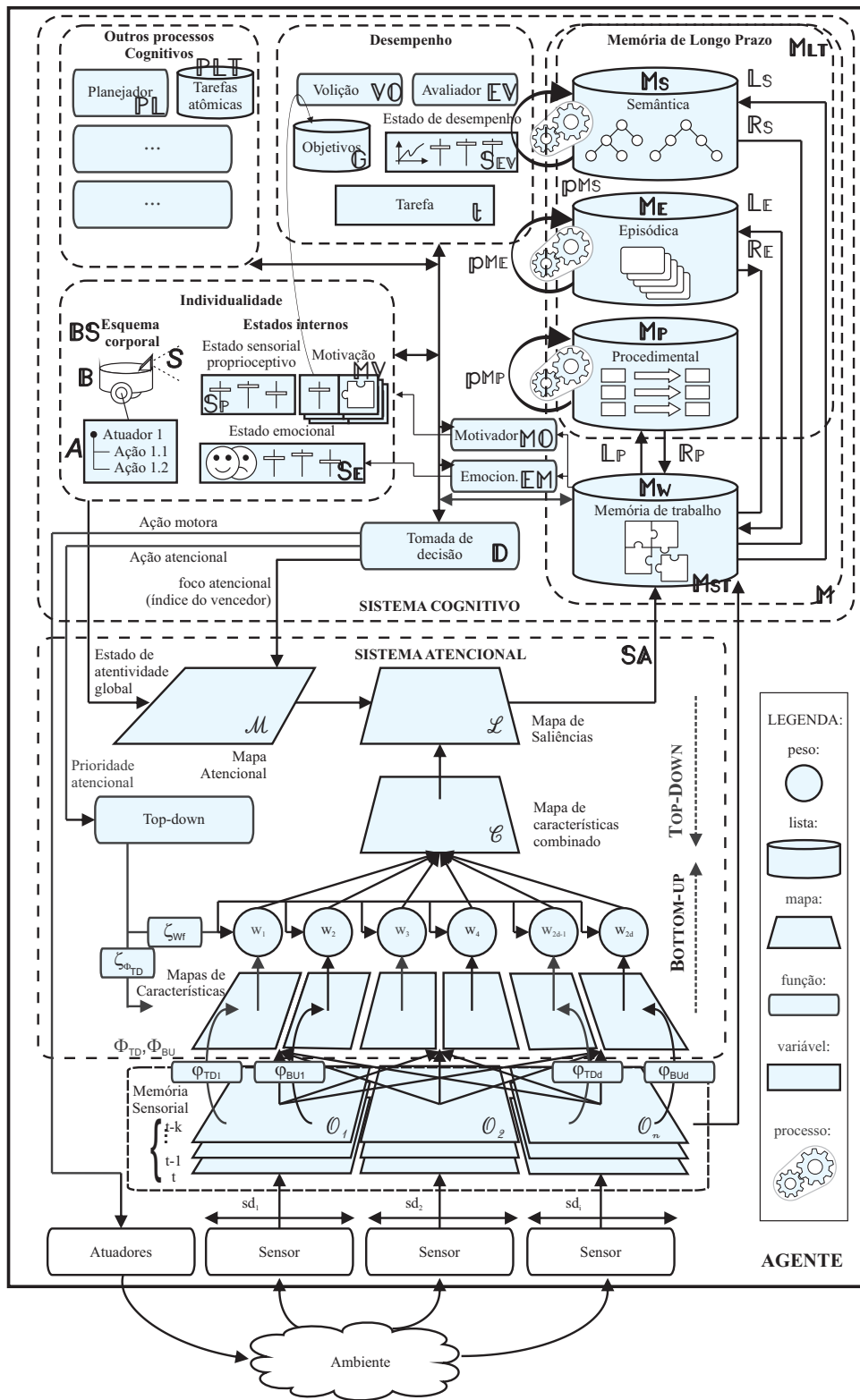


Figura 23 - Arquitetura do modelo proposto para um agente atencional consciente.

- sistema atencional

 - Um **sistema atencional**, conforme o sistema de **seleção para percepção** proposto por Colombini (2014) e discutido na [seção 5.2](#), compreendendo memória sensorial, mapas de características, pesos associados a mapas de características, mapa de características combinado, mapa de saliências e mapa atencional, e o curso da dinâmica atencional apresentado no modelo formal ([seção 5.2.2](#));

- sistema cognitivo

 - Um **sistema cognitivo** – que substitui o sistema de **seleção para ação** proposto por Colombini (2014) – composto por:
 - Bloco de **tomada de decisão**: bloco que delibera sobre as ações a serem realizadas pelo agente;
 - Bloco de **memória de curto prazo** (memória de trabalho): memória que armazena uma pequena quantidade de informação – tipicamente recebida do mapa de saliência e da memória sensorial – que poderá ser utilizada por diversos módulos do sistema cognitivo do agente, em particular pelas memórias de longo prazo;
 - Bloco de **memória de longo prazo**: memória que armazena grandes quantidades de informação por tempo indeterminado. É constituída de:
 - * Bloco de **memória episódica**: armazena dados sobre os eventos recebidos, especificamente *o que, quando e onde* na forma declarativa, utilizando uma estrutura de dados adequada;
 - * Bloco de **memória semântica**: armazena fatos e conhecimento sobre o mundo na forma declarativa, segundo uma estrutura de dados adequada;
 - * Bloco de **memória procedimental**: armazena o conhecimento sobre os procedimentos do agente, tipicamente na forma de pares estado-ação;
 - Bloco de **desempenho**: agrupa variáveis, listas e funções relacionadas com o desempenho do agente. Composto de:
 - * **Objetivos**: armazena o(s) objetivo(s) corrente(s) do agente;
 - * **Avaliações**: conjunto de métricas que visa avaliar o desempenho do agente sob diferentes perspectivas e tipicamente atua sobre seus estados internos;
 - * **Tarefa**: tarefa corrente executada pelo agente;
 - Bloco de **processos cognitivos**: reúne um número indefinido de funções de ordem cognitivas do agente de múltiplos propósitos, tais como: planejador, estimador de estados futuros, etc.;
 - Bloco de **individualidade**: composto por um conjunto de estados internos, tais como: estado proprioceptivo, motivação, estado emocional, e por um esquema corporal do agente.

processos

- Um conjunto de **processos** rodando em *background* com propósitos diversos que realizam, de forma não necessariamente sincronizada com os demais módulos, ações sobre elementos do sistema cognitivo.

O sistema opera sobre o paradigma principal de que qualquer função ou processo do sistema cognitivo pode requisitar informações de qualquer outro módulo e/ou alterar estados internos do agente em qualquer tempo, não sendo obrigatória, portanto, a passagem do fluxo de informação pelo módulo de tomada de decisão. Esta estrutura encontra inspiração no caso reportado de *craniopagus*, onde o sistema sensorio de uma das gêmeas pode responder às solicitações recebidas do sistema cognitivo da outra gêmea;

9.3 Modelo formal

Esta seção apresenta uma descrição formal para o modelo proposto. Este modelo faz uso das variáveis e funções do **sistema atencional** formalizadas por [Colombini \(2014\)](#), e descritas na [seção 5.2.2](#), razão pela qual estas não são novamente apresentadas na presente seção. O **sistema cognitivo** pode ser descrito como:

- ação** • Um conjunto de ações \mathbb{A} do agente, com um número arbitrário de elementos \mathfrak{a}_i denominados **ações**, com $\{\mathfrak{a}_1, \mathfrak{a}_2, \dots, \mathfrak{a}_n\} \in \mathbb{A}$, e com $\mathbb{A} = \mathbb{A}_\Delta \cup \mathbb{A}_M$;
- ação atencional** • Um conjunto de ações atencionais \mathbb{A}_Δ do agente, com um número arbitrário de elementos $\mathfrak{a}_{\Delta i}$ denominados **ações atencionais**, com $\mathbb{A}_\Delta = \{\mathfrak{a}_{\Delta 1}, \mathfrak{a}_{\Delta 2}, \dots, \mathfrak{a}_{\Delta n}\}$;
- ação motora** • Um conjunto de ações motoras \mathbb{A}_M do agente, com um número arbitrário de elementos $\mathfrak{a}_{M i}$ denominados **ações motoras**, com $\mathbb{A}_M = \{\mathfrak{a}_{M 1}, \mathfrak{a}_{M 2}, \dots, \mathfrak{a}_{M n}\}$;
- objetivo** • Uma lista simples de **objetivos** \mathbb{G} (*goals*) do agente, que contém um número arbitrário de elementos \mathfrak{g}_i , com $\mathbb{G} = \{\mathfrak{g}_1, \mathfrak{g}_2, \dots, \mathfrak{g}_n\}$;
- tarefas** • Um conjunto de tarefas \mathbb{T} do agente, com um número arbitrário de elementos \mathfrak{t}_i denominados **tarefas**, com $\mathbb{T} = \{\mathfrak{t}_1, \mathfrak{t}_2, \dots, \mathfrak{t}_n\}$, e com uma tarefa específica \mathfrak{t} denominada **tarefa corrente** do agente;
- estados** • Um conjunto de estados \mathbb{S} do agente, com um número arbitrário de elementos \mathfrak{s}_i denominados **estados**, onde $\mathbb{S} = \{\mathfrak{s}_1, \mathfrak{s}_2, \dots, \mathfrak{s}_n\}$, e com $\mathbb{S} = \mathbb{S}_\Delta \cup \mathbb{S}_E \cup \mathbb{S}_{EV} \cup \mathbb{S}_P$;
- estado atencional** • Um conjunto de estados atencionais \mathbb{S}_Δ do agente, com um número arbitrário de elementos de forma que $\mathbb{S}_\Delta = \{\mathfrak{s}_{\Delta 1}, \mathfrak{s}_{\Delta 2}, \dots, \mathfrak{s}_{\Delta n}\}$, onde cada elemento $\mathfrak{s}_{\Delta i}$, denotado **estado atencional**, é constituído de um *mapa de saliências* de m dimensões, denotado $\mathcal{L} \in \mathfrak{R}^m$, com elementos denotados por $\{l_1, l_2, \dots, l_m\}$, conforme o proposto no modelo de [Colombini \(2014\)](#) ([seção 5.2.2](#));

- estado emocional
 - Um conjunto de estados emocionais S_E (*emotional states*) do agente, com um número arbitrário de elementos, onde cada elemento s_{Ei} , denotado **estado emocional**, é constituído de um número arbitrário de **variáveis emocionais** v_E do agente, ou seja, $S_E = \{v_{E1}, v_{E2}, \dots, v_{En}\}$, onde $v_{Ei} \in \mathfrak{R}$;
- estado de desempenho
 - Um conjunto de estados de desempenho S_{EV} (*evaluation states*) do agente, com um número arbitrário de elementos, onde cada elemento s_{EVi} , denotado **estado de desempenho**, é constituído de um número arbitrário de **variáveis de desempenho** v_{EV} do agente, ou seja, $S_{EV} = \{v_{EV1}, v_{EV2}, \dots, v_{EVn}\}$, onde $v_{EVi} \in \mathfrak{R}$;
- estado proprioceptivo
 - Um conjunto de estados proprioceptivos S_P (*proprioceptive states*) do agente, com um número arbitrário de elementos, onde cada elemento s_{Pi} , denotado **estado proprioceptivo**, é constituído de um número arbitrário de **variáveis proprioceptivas** v_P do agente, ou seja, $S_P = \{v_{P1}, v_{P2}, \dots, v_{Pn}\}$, onde $v_{Pi} \in \mathfrak{R}$;
- motivação
 - Um conjunto de motivações MV do agente, com um número arbitrário de elementos m_{Vi} denominados **motivação**, com cada elemento $m_{Vi} \in MV$ composto por uma dupla com um valor $v_{MV} \in \mathfrak{R}$ denominado **valor da motivação** e um elemento de memória $m \in M$, ou seja, $m_{Vi} = \{v_{MV_i}, m_i\}$;
- memória
 - Um superconjunto denominado **memória** M , com $M = M_{LT} \cup M_{ST}$;
- memória de curto prazo
 - Um conjunto denominado **memória de curto prazo** M_{ST} , com $M_{ST} = M_W$;
 - Um conjunto denominado **memória de longo prazo** M_{LT} , com $M_{LT} = M_E \cup M_P \cup M_S$;
- memória de longo prazo
 - Uma lista duplamente encadeada de elementos denominada **memória de trabalho** M_W (*working memory*), capaz de comportar um número arbitrário de elementos m_W , ou seja $M_W = \{m_{W1}, m_{W2}, \dots, m_{Wn}\}$, e com entrada e saída de elementos da lista dada pelo princípio FIFO (*First-In-First-Out*). Recomenda-se a modelagem dos elementos m_W no formato proposto por [Samsonovich e Nadel \(2005\)](#), discutido na [seção 7.4](#);
- memória de trabalho
 - Uma lista duplamente encadeada de elementos denominada **memória semântica** M_S , capaz de comportar um número arbitrário de elementos m_S , com $M_S = \{m_{S1}, m_{S2}, \dots, m_{Sn}\}$, onde cada elemento de memória semântica m_{Si} contém um número arbitrário de n -uplas d_{MS} , tal que $m_{Si} = \{d_{MS1}, d_{MS2}, \dots, d_{MSn}\}$. A ordenação dos elementos m_S na lista M_S dá-se pelo valor crescente de um elemento d_{MSi} denominado **elemento indexador da memória semântica**;
- memória semântica
 - Uma lista duplamente encadeada de elementos denominada **memória episódica** M_E , capaz de comportar um número arbitrário de elementos m_E , com $M_E = \{m_{E1}, m_{E2}, \dots, m_{En}\}$, onde cada elemento de memória episódica m_{Ei} contém um número arbitrário de n -uplas d_{ME} , tal que $\{d_{ME1}, d_{ME2}, \dots, d_{ME_n}\} \in m_{Ei}$. A ordenação dos elementos m_E na lista M_E
- memória episódica
 - Uma lista duplamente encadeada de elementos denominada **memória episódica** M_E , capaz de comportar um número arbitrário de elementos m_E , com $M_E = \{m_{E1}, m_{E2}, \dots, m_{En}\}$, onde cada elemento de memória episódica m_{Ei} contém um número arbitrário de n -uplas d_{ME} , tal que $\{d_{ME1}, d_{ME2}, \dots, d_{ME_n}\} \in m_{Ei}$. A ordenação dos elementos m_E na lista M_E

dá-se pelo valor crescente de um elemento d_{MEi} denominado elemento **indexador da memória episódica**. Recomenda-se a modelagem dos elementos m_E no formato proposto por [Stachowicz e Kruijff \(2012\)](#), discutido na [seção 7.5](#);

memória
procedi-
mental

- Uma lista duplamente encadeada de elementos denominada **memória procedimental** M_P , capaz de comportar um número arbitrário de elementos m_P , com $M_P = \{m_{P1}, m_{P2}, \dots, m_{Pn}\}$, onde cada elemento de memória procedimental m_{Pi} é uma tripla formada por uma **tarefa** $t \in T$, um conjunto de **qualificadores** d_{MP} e um **conjunto estado-ação** C_{SA} , composto por um número também arbitrário de elementos $c_{SAi} \in C_{SA}$, com cada elemento C_{SA} formado por pares estado $s \in S$ e ação $a \in A$, com $c_{SAi} = \{s_i, a_i\}$, tal que $m_{Pi} = \{t_i, d_{MP}, C_{SAi}\}$;

planejamento

- Uma **função de planejamento** PL , que mapeia um objetivo $g \in G$ para um conjunto com um número arbitrário de tarefas $t \in PLT$, com PLT representando uma lista de tarefas interna ao planejador. Ou seja:

$$PL \\ g \longrightarrow \{t_1, t_2, \dots, t_n\} \in PLT$$

volição

- Uma **função de volição** VO que mapeia motivações $mv \in MV$ para objetivos $g \in G$, dada por:

$$VO \\ \{MV\} \longrightarrow \{G\}$$

avaliação

- Uma **função de avaliação** EV do agente, que mapeia o estado corrente do agente S , seus objetivos G e tarefas t e PLT para o estado de avaliação S_{EV} do agente, isto é:

$$EV \\ \{S_A, S_E, S_{EV}, S_P, G, t, PLT\} \longrightarrow \{S_{EV}\}$$

motivação

- Uma **função de motivação** MO do agente, que mapeia o conteúdo da memória de trabalho M_W para a motivação MV do agente, isto é: $MO: M_W \rightarrow MV$

emoção

- Uma **função de emocionamento** EM do agente, que mapeia o conteúdo da memória de trabalho M_W para o estado emocional S_E do agente, isto é: $EM: M_W \rightarrow S_E$

tomada
de
decisão

- Uma **função de tomada de decisão** \mathbb{D} (*decision making*) do agente, que tipicamente coordena todos os sistemas cognitivos, listas e elementos escolhendo uma ação $a \in A$ – que pode ser uma ação motora $a_M \in A_M$ ou uma ação atencional $a_A \in A_A$ – e um índice i denotado **índice do vencedor do processo atencional** ou **foco da atenção**. Em outros termos:

$$\mathbb{D} \\ \{S_A, S_E, S_{EV}, S_P, G, T, M_W, MV, PLT, BS\} \longrightarrow \{a_M, a_A, i, G, T, M_W, MV, S_E\}$$

esquema
corporal

- Um **esquema corporal** (*body schema*) BS do agente, uma n -upla formada pelo conjunto de sensores do agente \mathcal{S} , o conjunto de atuadores do agente \mathcal{A} , e uma n -upla \mathcal{B} que contém um conjunto de características descritoras da estrutura física do agente (peso, volume, centro de gravidade, formato externo, etc.), tal que $BS = \{\mathcal{S}, \mathcal{A}, \mathcal{B}\}$.

processo

- Um conjunto de **processos** assíncronos inconscientes \mathbb{P} , com cada $\{p_1, p_2, \dots, p_n\} \in \mathbb{P}$ representando um processo responsável por realizar um mapeamento de um elemento específico da arquitetura para atualização de estados e/ou listas do sistema cognitivo.

$$\mathbb{P} \\ \{S_A, S_E, S_{EV}, S_P, MV, M, M_W, G, T\} \longrightarrow \{S_A, S_E, S_{EV}, S_P, MV, M, M_W, G, T\}$$

aprendizado
semân-
tico

- Um **aprendizado semântico** \mathbb{L}_S , capaz de efetuar um mapeamento tipicamente dado por:

$$\mathbb{L}_S \\ \{m_W, M_W, M_S\} \longrightarrow \{m_S\}$$

aprendizado
episódico

- Um **aprendizado episódico** \mathbb{L}_E , capaz de efetuar um mapeamento tipicamente dado por:

$$\mathbb{L}_E \\ \{m_W, M_W, M_E\} \longrightarrow \{m_E\}$$

aprendizado
procedi-
mental

- Um **aprendizado procedimental** \mathbb{L}_P , capaz de efetuar um mapeamento tipicamente dado por:

$$\mathbb{L}_P \\ \{m_W, M_W, M_P\} \longrightarrow \{m_P\}$$

recordação
semân-
tica

- Uma **recordação semântica** \mathbb{R}_S , capaz de efetuar um mapeamento tipicamente dado por:

$$\mathbb{R}_S \\ \{m_W, M_W, M_S\} \longrightarrow \{m_W\}$$

recordação
episódica

- Uma **recordação episódica** \mathbb{R}_E , capaz de efetuar um mapeamento tipicamente dado por:

$$\mathbb{R}_E \\ \{m_W, M_W, M_E\} \longrightarrow \{m_W\}$$

recordação
procedi-
mental

- Uma **recordação procedimental** \mathbb{R}_P , capaz de efetuar um mapeamento tipicamente dado por:

$$\mathbb{R}_P \\ \{m_W, M_W, M_P\} \longrightarrow \{m_W\}$$

9.4 Operação do modelo

A operação do modelo apresentado na [seção 9.3](#) dá-se por:

1. Iniciar novo ciclo atencional:

mapa de
saliên-
cias

- 1.1. Computar o modelo de [Colombini \(2014\)](#) conforme [seção 5.2.2](#) até a geração do **mapa de saliências** \mathcal{L} ;

competição

- 1.2. Computar o **vencedor do processo competitivo** (*winner*), isto é, o índice do sensor com a máxima saliência. Para isso, fazer:

$$winner = arg(max(\mathcal{L}))$$

2. Iniciar novo ciclo cognitivo:

- 2.1. Criar um novo elemento m_W na memória de trabalho M_W , e inserir em m_W o índice do sensor vencedor *winner*, suas observações \mathcal{O}_{winner} e o mapa de saliências \mathcal{L} , isto é, computar:

$$\{winner, \mathcal{O}_{winner}, \mathcal{L}\} \longrightarrow m_W \in M_W$$

- aprendizado semântico
- aprendizado episódico
- aprendizado procedimental
- recordação semântica
- recordação episódica
- recordação procedimental
- função motivação
- motivação
- 2.2. Computar o **aprendizado semântico** L_S sobre m_W visando a inserção ou ampliação de conceitos em M_S , ou seja, realizar: $L_S : \{m_W, M_W, M_S\} \rightarrow m_S$;
 - 2.3. Computar o **aprendizado episódico** L_E sobre m_W visando a inserção ou ampliação de episódios em M_E , ou seja, realizar: $L_E : \{m_W, M_W, M_E\} \rightarrow m_E$;
 - 2.4. Computar o **aprendizado procedimental** L_P sobre m_W visando a inserção ou ampliação de procedimentos em M_P , ou seja, realizar: $L_P : \{m_W, M_W, M_P\} \rightarrow m_P$;
 - 2.5. Computar a **recordação semântica** R_S sobre m_W , inserindo em m_W o conteúdo recordado m_S ou *null*, ou seja, $R_S : \{M_W, M_S, \emptyset\} \rightarrow \{m_W\}$. R_S pode ser realizado baseado em muitos aspectos distintos d_{M_S} de elementos m_S (forma, cor, cheiro, propósito, número de *recalls*, etc.). Para cada d_{M_S} , um algoritmo distinto de correspondência (*matching*) deve realizar uma busca completa nos elementos da memória;
 - 2.6. Computar a **recordação episódica** R_E sobre m_W , inserindo em m_W os conteúdos recordados m_E ou *null*, ou seja, $R_E : \{M_W, M_E, \emptyset\} \rightarrow m_W$. R_E pode ser realizado com base em muitos aspectos distintos d_{M_E} de m_E (tempo do evento, condições ambientais, objetos, ações realizadas, etc.). Para cada d_{M_E} , um algoritmo distinto de correspondência (*matching*) deve realizar uma busca completa nos elementos da memória;
 - 2.7. Computar a **recordação procedimental** R_P sobre m_W , inserindo em m_W os conteúdos recordados m_P ou *null*, ou seja, $R_P : \{M_W, M_P, \emptyset\} \rightarrow m_W$. R_P pode ser realizado com base em qualificadores distintos d_{M_P} de uma tarefa (propósitos, restrições, condições, etc.). Algum raciocínio pode ser necessário para checar a aplicabilidade de um procedimento em uma dada situação para atingir um resultado esperado;
 - 2.8. Computar a **função motivação** M_O . Funções de motivação podem ser realizadas com base em um grande número de teorias psicológicas (BROONEN et al., 2011). Tipicamente, a motivação é reduzida quando certa atividade é realizada por muito tempo, ou se ela é entendida como chata, difícil ou se o agente tem atividades mais agradáveis a serem realizadas. Elas também dependem do agente como um indivíduo, de seu estado emocional e memórias ou de sua curiosidade com relação a um novo estímulo. Se o item em m_W indica que:
 - a. M_S , M_E ou M_P possuem um conceito, episódio ou procedimento em significativa evolução que pode ser investigado pelo agente, ou
 - b. Elementos recordados têm emoções relevantes associadas, ou
 - c. Elementos são frequentemente recordados pelas memórias;
 então **motivar** o agente a se relacionar com o conceito, episódio ou procedimento fazendo:
 - 2.8.1. Se $m_W \subset M_V$, então atualizar o valor da motivação v_{M_V} associada;

2.8.2. Caso contrário, criar um novo item $m_{v_i} \in MV$, inserir um valor alto de motivação em $v_{MV_i} \in mv_i$ e inserir m_w em $m_i \in mv_i$; realizar ainda a atualização dos valores associados a motivações $m_w \subset MV$ de acordo com uma política de decaimento no tempo, eventualmente removendo itens da lista;

estado
emocional

2.9. Computar o **gerador de emoções** EM do agente com base em m_w (dados observados e memórias recordadas) de forma a atualizar o **estado emocional** do agente SE , composto por um número de emoções v_E . Embora um largo número de emoções básicas possam ser investigadas, alguns candidatos são: raiva, desgosto, medo, felicidade, tristeza, surpresa (EKMAN, 1999) (PANKSEPP, 2005). Funções de emoção devem afetar a intensidade das emoções tentando imitar a complexa regulação biológica que ocorre no córtex, hipocampo, tálamo e regiões correlatas do cérebro humano. Para uma visão geral, veja (ZIEMKE; LOWE, 2009) (STARZYK; PRASAD, 2011) (DAMASIO, 2003);

volição

2.10. Computar a **função volição** VO do agente, avaliando MV e G de forma a decidir se deve definir um novo objetivo $g \in G$ para o agente. VO é assumida como um processo distinto de MO , dado que VO é mais relacionado com um processo deliberativo. Ela pode ser dividida em fases pré-ação e de ação. A fase pré-ação preocupa-se com a formação de interpretações e planejamento de quando, onde e como iniciar a ação. VO deve escolher considerando a real possibilidade de completar o objetivo, os reais benefícios de completá-lo, condições ambientais, força da motivação, etc. A fase de ação está relacionada com a avaliação da execução da ação e com as razões para mantê-la. Alguns fatores devem ser considerados: habilidade de resistir a distratores, capacidade do ambiente de capturar a atenção do agente, obstáculos não previstos, desencorajamento para com a realização da ação, estado emocional (BROONEN et al., 2011);

planejamento

2.11. Se necessário, computar a **função de planejamento** PL para determinar as tarefas atômicas $\{t_1, t_2, \dots, t_n\} \in PLT$ que precisam ser realizadas para cumprir os objetivos G , ou seja, computar:

$$G \xrightarrow{PL} \{t_1, t_2, \dots, t_n\} \in PLT$$

estado de
desempenho

2.12. Avaliando os objetivos G , a tarefa corrente t do agente, o estado geral do agente S e as tarefas que precisariam ser executadas para o atendimento das tarefas PLT , computar o **estado de desempenho** S_D do agente, computando:

$$\{S_A, S_{EV}, S_E, S_P, G, t, PLT\} \xrightarrow{EV} S_{EV};$$

tomada
de
decisão

2.13. Computar a **tomada de decisão** \mathbb{D} , fazendo:

2.13.1. Considerando a lista de objetivos \mathbb{G} , a tarefa corrente \mathbb{t} do agente, o estado geral do agente \mathbb{S} e as tarefas que precisariam ser executadas para o atendimento das tarefas \mathbb{PLT} , deliberar sobre a manutenção ou **alteração da tarefa corrente** \mathbb{t} . Em outros termos, computar:

$$\mathbb{D} \\ \{\mathbb{S}_A, \mathbb{S}_E, \mathbb{S}_{EV}, \mathbb{S}_P, \mathbb{G}, \mathbb{t}, \mathbb{PLT}\} \longrightarrow \mathbb{t};$$

2.13.2. Inserir \mathbb{t} como item na memória de trabalho m_W e aguardar a recordação R_P da tarefa de M_P , fazendo:

2.13.2.1. Se o procedimento para resolução da tarefa for conhecido, isto é, se $R_P: \{m_P, MP\} \rightarrow m_W$, resulta em um $m_W \neq \emptyset$ então executar a tarefa \mathbb{t} utilizando o procedimento descrito no conjunto de pares estados-ação $C_{SA} \in m_W$;

2.13.2.2. Caso contrário, se a recordação retornar *falha*, isto é, $m_W = \emptyset$, então avaliar se o agente deve realizar um procedimento de aprendizado da tarefa \mathbb{t} . Em caso afirmativo, criar um novo elemento $m_{P_i} \in M_P$ com $m_{P_i} = \{\mathbb{t}_i, C_{SA_i}\}$ e iniciar o aprendizado.

3. Iniciar novo ciclo atencional (passo 1) para $tempo = tempo + 1$;

Além da execução do modelo, prevê-se ao menos 3 (três) processos assíncronos inconscientes em *background*:

- Um processo P_{M_S} , que varre permanentemente a memória semântica M_S buscando derivar ou unificar conceitos, executando um mapeamento tipicamente dado por: $P_{M_S}: M_S \rightarrow M_S$;
- Um processo P_{M_E} , que varre permanentemente a memória episódica M_E buscando derivar ou unificar episódios, executando um mapeamento tipicamente dado por: $P_{M_E}: M_E \rightarrow M_E$;
- Um processo P_{M_P} , que varre permanentemente a memória procedimental M_P buscando derivar ou unificar procedimentos, executando um mapeamento tipicamente dado por: $P_{M_P}: M_P \rightarrow M_P$.

9.5 Sumário do modelo

De modo geral, o modelo atencional consciente proposto neste trabalho considera:

- atentividade

 - A possibilidade de trabalho com **múltiplos sensores** \mathcal{S} de diferentes categorias e dimensões, que são automaticamente fundidos pelo sistema sensorial-atencional;
 - A **atentividade**, uma vez que o agente é capaz de reagir a estímulos do ambiente recebidos através dos sensores e modulados de acordo com os sistemas atencional e cognitivo;
- senciência

 - A **senciência**, uma vez que o agente pode ter suas variáveis de estado emocional alteradas pelas experiências conscientes;
- auto-ciência

 - A **auto-ciência**, uma vez que o agente possui um modelo de individualidade que contém estados internos e esquema corporal, assim como mecanismos de planejamento, tomada de decisão e aprendizado, que dão suporte à introspecção, bem como a possíveis comparações com outros indivíduos;
- auto-consciência

 - A **auto-consciência**, uma vez que o agente possui mecanismos para ser consciente de seu corpo, para reconhecer a semelhantes e/ou a si mesmo, para ser consciente de algo que aconteça em seu entorno (formulando conceitos e assimilando experiências), bem como para ser consciente das conseqüências de suas decisões;
- consciência autooética

 - A **consciência autooética**, uma vez que o agente possui elementos que permitem o raciocínio sobre presente, passado e futuro, especialmente considerando a modelagem da memória de trabalho de acordo com o proposto por Starchowicz ([seção 7.4](#));
- possessividade

 - A **possessividade**, uma vez que objetos utilizados com frequência pelo agente podem incrementar um de seus indicadores $d_{ms} \in \mathbb{M}_s$, de forma a representar uma relação de posse;
- volição

 - A **volição** $\forall \mathcal{O}$, aqui modelada como uma função capaz de transformar motivações $m_v \in \mathbb{M}_v$ em objetivos $o \in \mathcal{O}$;
- perspectividade

 - A **perspectividade**, uma vez que o mecanismo mental contido na memória de trabalho é flexível de forma a permitir o raciocínio sobre a perspectiva do *eu* em comparação ao *ele*. A adoção de um ponto de vista dependerá da capacidade de construção de conceitos na memória semântica, e da capacidade de raciocínio do agente sobre eles;
 - O sistema cognitivo permite **inclusão de outros módulos** especializados (tais como: processamento de linguagem, estimação de estados futuros, etc.) no bloco denominado *outros processos cognitivos*, bem como permite a inclusão de outros processos \mathbb{P} ;
 - O suporte a estímulos **exógenos** (*bottom-up*) e **endógenos** (*top-down*) e seus cursos de seleção pelo processo atencional;
- atenção sustentada

 - A **atenção sustentada**, ou seja, a manutenção do foco da atenção, guiada agora pelo

bloco de tomada de decisão (\mathbb{D}), baseado em todo o esquema cognitivo consciente do agente;

atenção
seletiva

- A **atenção seletiva**, que eleva à memória apenas um pequeno subconjunto das informações sensoriais localizado no entorno do foco da atenção, reduzindo drasticamente o espaço de entrada sensorial;

atenção
orientada

- A **atenção orientada**, resultado do processo atencional para troca do foco da atenção;

atenção
evidente

- A **atenção evidente**, dado que o módulo de tomada de decisão \mathbb{D} pode optar por uma ação motora $\mathbb{C}_M \in \mathbb{A}_M$ para ajustar as estruturas sensórias;

atenção
enco-
berta

- A **atenção encoberta**, dado que o módulo de tomada de decisão \mathbb{D} pode optar, se necessário, apenas por uma ação atencional $\mathbb{C}_A \in \mathbb{A}_A$, sem que uma ação motora $\mathbb{C}_M \in \mathbb{A}_M$ esteja associada a esta.

9.6 Comparação com outros modelos

Avaliando-se a inserção do modelo proposto junto aos demais modelos conscientes existentes na literatura, abordados na [seção 6.1](#) e sintetizados pela [figura 14](#), temos:

- **Modelos baseados em espaço de trabalho global:** no modelo proposto, os estímulos sensoriais competem entre si para ganhar o foco da atenção. A competição existe, então, no nível atencional. No nível cognitivo, não existe uma competição direta pelo domínio exclusivo da máquina de processamento, mas existe uma forte cooperação entre os módulos para o processamento da informação na medida em que vários deles operarem paralelamente de forma assíncrona sobre os dados da memória de trabalho, cooperando entre si. Esta sistemática mantém forte similaridade com o GWT, promovendo a interação coletiva entre os diversos processos especializados. Mais do que isso, o conceito do *broadcast* surge naturalmente na trilha compartilhada por praticamente todos os módulos no sistema cognitivo. Abre-se, portanto, uma possibilidade para que a experiência consciente possa emergir através das interações coletivas dos vários módulos;
- **Modelos baseados na integração de informações:** dois dos princípios básicos relacionados com esses modelos são atendidos. Existe *diferenciação* entre estímulos no mecanismo de percepção, e também integração entre os diversos sistemas;
- **Modelos baseados no auto-conhecimento:** no modelo proposto, o agente possui imagem do corpo e estados emocionais, favorecendo seu auto-conhecimento. O agente possui capacidade de reconhecer seus semelhantes por similaridade com sua imagem corporal. O modelo adotado para a memória episódica também favorece que o agente raciocine

sobre si mesmo no tempo e no espaço. Abre-se, portanto, uma possibilidade para que a experiência consciente possa se manifestar através da integração da informação em uma experiência unificada;

- **Modelos baseados em representações de nível superior:** o conhecimento assume forma declarativa e não-declarativa. Mistura-se o conhecimento sensorial e simbólico, bem como processos conscientes e inconscientes;
- **Modelos baseados em atenção:** Por fazer uso de um mecanismo atencional bastante complexo e que integra diferentes abordagens, o modelo proposto contempla praticamente a totalidade dos esquemas atencionais encontrados na literatura. O esquema de mapas de características combinado assemelha-se ao sistema de supermapas. O sistema também apresenta uma central executiva, responsável pelo fluxo principal de informação. No sistema de hierarquia mais elevada, o uso da linguagem declarativa conclui o processo de mapeamento que vem do nível perceptual até o armazenamento do conhecimento nas memórias. Desta forma, ficam assegurados os mecanismos elementares para que a consciência possa emergir através dos mecanismos atencionais.

Sob um outro prisma, o modelo atende a aspectos básicos dos axiomas mínimos para a consciência propostos por Aleksander e Dunmall (2003) (seção 4.6.1), uma vez que: *i*) o agente possui níveis de **personificação**; *ii*) O agente possui capacidade para armazenamento de eventos e experiências na forma de memórias e de recordá-las; *iii*) O agente faz uso do mecanismo atencional; *iv*) O agente apresenta **volição**; *v*) O agente possui meios de relacionar emoções e seu contexto.

perfil
cognitivo

Tomando como guia o ConScale, discutido na seção 4.6.2, pode-se dizer que o modelo proposto tem potencial para atingir um **perfil cognitivo** compatível com o apresentado na figura 24. Desta forma, trata-se de uma arquitetura com potencial próximo à arquitetura de Haikonen (2007b)

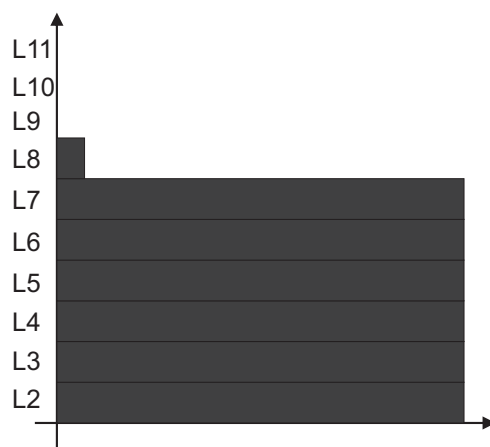


Figura 24 - Perfil cognitivo estimado do modelo proposto segundo o Conscale.

10 MATERIAIS E MÉTODOS

“Você faz experimentos e eu faço teorias. Sabe qual a diferença? Uma teoria é algo em que ninguém acredita, exceto a pessoa que a fez. Um experimento é algo em que todos acreditam, exceto quem o fez.”

*Albert Einstein
citado em [Holton \(1986\)](#)*

O presente capítulo apresenta um conjunto de experimentos computacionais destinados a investigar o modelo proposto. Ele encontra-se dividido da seguinte forma:

- [Seção 10.1](#): Descreve o ambiente utilizado para simulação e os materiais e métodos gerais utilizados na experimentação;
- [Seção 10.2](#): Explicita a sistemática de extração de características utilizada nos experimentos;
- [Seção 10.3](#): Apresenta um experimento com robô atencional *bottom-up* não consciente;
- [Seção 10.4](#): Apresenta um experimento com robô atencional *bottom-up* e *top-down* não consciente;
- [Seção 10.5](#): Apresenta um experimento com robô atencional consciente.

10.1 Ambiente de simulação

[usarsim](#)

Para a realização dos experimentos adotou-se a plataforma **USARSim** (*Urban Search and Rescue Simulator*) ([WANG; BALAKIRSKY, 2006](#)). Trata-se de um simulador proposto para competições na modalidade *Virtual Robots* da RoboCup e propício para estudos de sistemas robóticos e multiagentes. O simulador é construído sobre a *engine* de um jogo, o que permite uma excelente acuidade visual a um baixo custo computacional. O simulador fornece diversos mapas de situações simuladas de desastres, que contém diversos objetos (caixas, destroços,

rampas, vítimas humanas, etc.). Encontram-se disponíveis para uso no USARSim diversos robôs comercialmente disponíveis (terrestres, aéreos e aquáticos), bem como diversos sensores (sonar, odômetro, IR, RFID, encoders, laser, bateria e câmera). O simulador automaticamente adiciona aos sensores um ruído, que é modelado utilizando um gerador de número pseudo-aleatório com distribuição gaussiana sobre o sinal original. A [figura 26](#) fornece uma visão geral do ambiente e alguns de seus principais elementos.

Em termos computacionais, a plataforma USARSim ([figura 25](#)) provê um servidor TCP/IP, ao qual podem se conectar controladores clientes. Um cliente básico foi desenvolvido em trabalhos prévios pelo autor e seu grupo de trabalho ([COLOMBINI et al., 2008](#)) ([ARAUJO et al., 2008](#)) ([SIMÕES et al., 2009](#)) ([COLOMBINI, 2014](#)). O controlador foi desenvolvido em C++ sobre interface gráfica *cross*-plataforma *wxWidgets*. As diferentes classes foram implementadas como *threads* distintas. Este controlador foi adaptado para acomodar as classes pertinentes ao modelo de agente apresentado no [capítulo 9](#). As classes que implementam algoritmos de [PROSE](#) busca e memória em C++ foram implementadas utilizando o ambiente **PROSE** (*Problem Search Environment*), previamente desenvolvido pelo autor ([SIMÕES; COLOMBINI, 2014](#)). Os experimentos foram conduzidos em um *notebook* Intel core i7 2.6GHz com 16GBytes de memória RAM e placa de vídeo AMD Radeon R7 M265 com 2GBytes de memória de vídeo, rodando sob sistema operacional Windows 8.

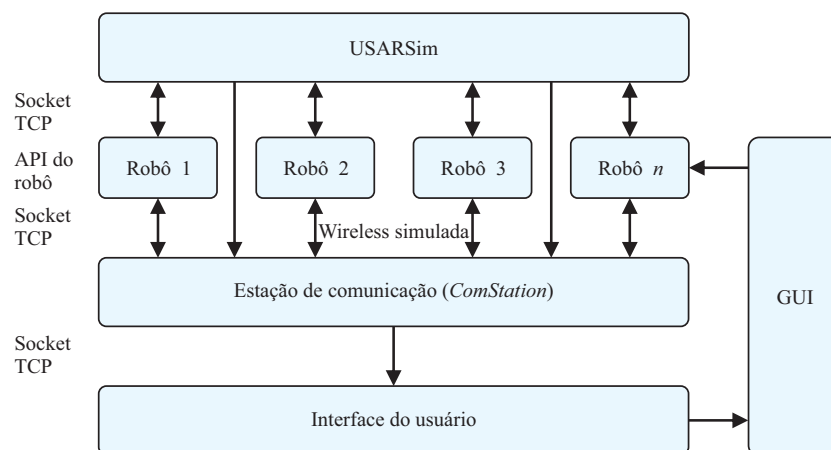


Figura 25 - Arquitetura modular do USARSim. Adaptado de ([COLOMBINI, 2014](#)).

Os experimentos descritos nas próximas seções foram conduzidos utilizando robôs Pioneer P2AT utilizando a seguinte **configuração de sensores**:

configuração
de
sensores

- **Laser:** 01 (um) *range scanner* com alcance de 20m, com abertura angular de 180° instalado no *chassis* do robô e apontado para a frente. A disposição do *range scanner* é apresentada na [figura 26e](#);
- **Sonares:** 08 (oito) sonares, cada um deles com um alcance de 5cm-5m e ângulo de

abertura $19,54^\circ$, instalados no *chassis* do robô de forma igualmente espaçada entre si e cobrindo os 180° da porte frontal do robô. A disposição dos sonares é apresentada na [figura 26f](#);

Algumas observações importantes:

- Conforme pode ser observado nas [figuras 26e-f](#), existem regiões de sombra ao redor do robô onde nenhum dos sensores é capaz de identificar estímulos;
- ruído • Todos os sensores possuem incorporado um gerador de ruído¹. Foram utilizados os valores padrão do ambiente: amplitude de ruído aleatório relativa de 5% e desvio padrão $\sigma = 0,1$;
- cenário • O **cenário** selecionado foi o *DM-Testroom-250*, apresentado na [figura 26a](#), que caracteriza-se pelo amplo espaço de circulação para os robôs, com a presença de diversas caixas de tamanhos e formatos distintos, bem como a presença de vítimas, rampas, vidro, além de ondulações no piso.

10.2 Sistemática de extração de características

10.2.1 Espaço de observação

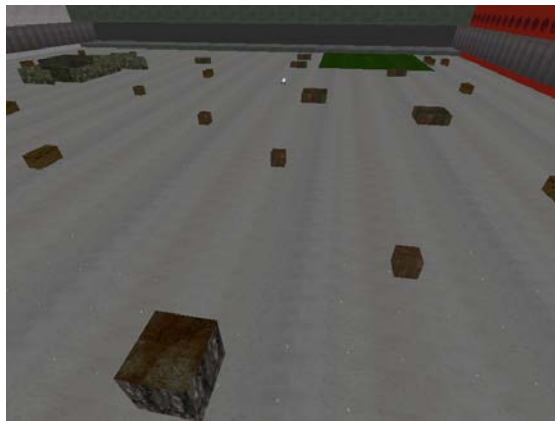
O espaço de observações foi definido como:

- O_1 representa o espaço de observação para as leituras de sonar, definido por: $o_{1_{n_t}} = \text{sonar}_{n_t}$ with $n \in [1, 8]$
- O_2 representa o espaço de observação para as leituras do scanner laser, definido por: $o_{2_{n_t}} = \text{range}_{n_t}$ with $n \in [1, 180]$

10.2.2 Características *bottom-up*

O modelo é capaz de combinar diferentes dimensões de características fornecidas por múltiplos sensores. Nos experimentos realizados, as seguintes características figuram como mais relevantes:

¹Antes que o dado seja enviado ao robô, um número aleatório é adicionado para simular o ruído aleatório. Então uma curva de distorção é usada para interpolar o valor da medida para simular o sensor real. Com o ruído, o dado terá, portanto, valor dado por: $data = data + \text{random}(\text{noise}) * data$ (WANG; BALAKIRSKY, 2006)



(a)



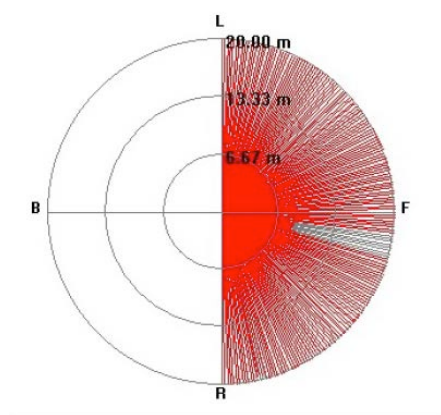
(b)



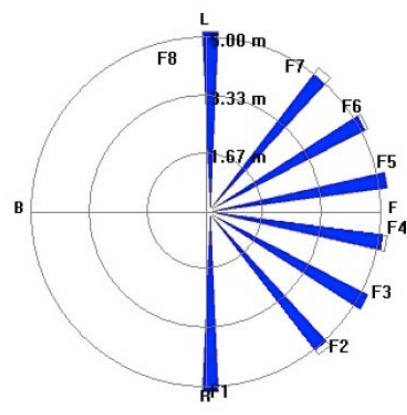
(c)



(d)



(e)



(f)

Figura 26 - Imagens da plataforma USARSim: a) visão geral do cenário DM-TestRoom250, constituído de caixas de tamanhos distintos, paredes e vítimas, e com amplo espaço de circulação; b) robô móvel virtual Pioneer P2-AT; c) Caixa de madeira; d) Vítima; e) Exemplo de leituras do *range scanner* do robô disposto na parte frontal (F) do robô; f) Exemplo de leituras dos 8 sonares (F1-F8) dispostos na parte frontal (F) do robô.

- **Movimento:** representa a intensidade de movimento detectada no ambiente;
- **Direção:** representa a direção de objetos em movimento no ambiente relativo ao agente atencional;
- **Distância:** representa a distância entre obstáculos (estáticos ou em movimento) e o agente atencional;
- **Espaço:** representa a distribuição do interesse no espaço.

Para obter essas características, foram utilizadas 3 (três) funções de mapeamento de características *bottom-up* desenvolvidas baseadas nas observações descritas acima. Essas funções foram:

\mathcal{F}_1 : Movimento

Esta característica extrai regiões com forte contraste de velocidade estimando a velocidade com que os alvos/distratores se movem na cena. Assumindo que cada obstáculo em movimento no ambiente se move com velocidade constante, computa-se $\phi_{BU_1} = z(O_1)$, onde $z(O_1)$ é dado por:

$$f_{1_{n_t}} = \left| \frac{\Delta s_{n_t}}{\Delta t} \right|, \quad (6)$$

,onde Δt é a variação de tempo retornada pelo simulador entre duas leituras do sonar e $\Delta s = o_{1_{n_t}} - o_{1_{n_t-1}}$. Para estabelecer um nível de contraste entre o conjunto de elementos, cada elemento da característica de movimento foi atualizado fazendo:

$$f_{1_{n_t}} = \left| f_{1_{n_t}} - \frac{\sum_{m=1}^n f_{1_{m_t}}}{n} \right| \quad (7)$$

Finalmente, cada elemento de característica foi normalizado dividindo-se os elementos pela máxima velocidade permitida para objetos se movendo no ambiente testado (MAX-SPEED).

\mathcal{F}_2 : Direção

Esta característica extrai informação relacionada à direção de objetos se movendo (alvos ou destratores) no ambiente. O robô atencional é usado como referência e três direções são consideradas: a) em direção ao robô atencional; b) afastando-se do robô atencional; c) não alterando a direção. A função $\phi_{BU_2} = z(O_1) \circ z(O_2)$, onde $z(O_1)$ é dada por:

$$tf_{1_{n_t}} = \begin{cases} 1 & \text{if } \frac{\Delta s_n}{\Delta t} > 0 \\ -1 & \text{if } \frac{\Delta s_n}{\Delta t} < 0 \\ 0 & \text{caso contrário} \end{cases} \quad (8)$$

onde Δt e Δs são dadas pela [equação 7](#) e $tf_{1_{n_t}}$ é um mapa de características temporário com dimensão n , com $n \in [1, 8]$. Então, para medir o nível de discrepância, cada $tf_{1_{n_t}}$ é atualizado da seguinte forma:

$$tf_{1_{n_t}} = \frac{n}{count(tf_{1_{n_t}})} \quad (9)$$

, onde $count(tf_{1_{n_t}})$ é uma função que determina o número de ocorrências deste valor de características na dimensão temporária da característica.

Para estabelecer quando há benefícios na utilização de sensores com diferentes capacidades (alcance, sensibilidade de material, etc.) para gerar o mesmo mapa de características, $z(O_2)$ é representado da mesma forma que a [equação 9](#), com Δt sendo a variação de tempo informada pelo simulador entre duas leituras de ranger scanner e $\Delta s = o_{2_{n_t}} - o_{2_{n_t-1}}$, $n \in [1, 180]$, como em O_2 . Então, $tf_{2_{n_t}} = n/count(tf_{2_{n_t}})$. A função composta $\phi_{BU_2} = z(O_1) \circ z(O_2)$ usa o espaço de características temporário para determinar cada $f_{2_{n_t}}$ do mapa de características \mathcal{F}_2 . A composição é realizada reduzindo-se a dimensão $z(O_2)$ para adequar-se a $z(O_1)$.

\mathcal{F}_3 : Distância

ϕ_{BU_3} é responsável por extrair informação relacionada à disposição dos elementos em relação ao agente atencional. $\phi_{BU_3} = z(O_2)$, onde $z(O_2)$ é dado por:

$$f_{3_{n_t}} = |o_{2_{n_t}} - \frac{\sum_{m=1}^n o_{2_{m_t}}}{n}|/mr, \quad (10)$$

, onde mr é o valor de saturação do *range scanner*.

10.2.3 Características *top-down*

O processo exógeno da atenção é extremamente relevante para detectar as cenas do mundo onde as características presentes são discrepantes. Entretanto, a maior parte da nossa atenção é dirigida sobre influência exógena. Esta seção apresenta as funções de mapeamento *top-down* formuladas para realizar a influência exógena sobre as observações da memória sensorial e do

mapa atencional. As características *top-down* modeladas são:

\mathcal{F}_4 : Distância-Objetivo

Considere o valor $f_{3_{n_t}}$ definido pela equação 10 como a disposição dos elementos em torno do agente atencional no tempo t . Dessa forma, cada $f_{4_{n_t}}$ do mapa de características *top-down* será computada como:

$$\begin{cases} \text{Se } f_{3_{n_t}} == \text{DISTANCIA_DESEJADA} & \implies f_{4_{n_t}} = 1 \\ \text{Se } f_{3_{n_t}} \neq \text{DISTANCIA_DESEJADA} & \implies f_{4_{n_t}} = f_{3_{n_t}}/\text{MAXDISTANCE} \end{cases} \quad (11)$$

, onde DISTANCIA_DESEJADA assume os valores 1m, 2m e 3m respectivamente, para as ações atencionais $\mathbb{A}_{\mathbb{A}} = \{\text{DISTANCIA_MUITO_PERTO}, \text{DISTANCIA_PERTO}, \text{LONGE}\}$.

\mathcal{F}_5 : Região-Objetivo

Esta característica representa regiões do espaço para os quais a atenção do agente pode ser direcionada. Ela permite que se preste atenção a uma região independente do que esteja acontecendo na mesma. Cada elemento $f_{5_{n_t}}$ é computado de acordo com:

$$\begin{cases} \text{Se EXTREMA_DIREITA} & \implies f_{5_{n_t}} = 1 & \text{para } 1 \leq n \leq 3 \\ \text{Se DIREITA} & \implies f_{5_{n_t}} = 1 & \text{para } 1 \leq n \leq 2 \\ \text{Se FRENTE} & \implies f_{5_{n_t}} = 1 & \text{para } 4 \leq n \leq 5 \\ \text{Se ESQUERDA} & \implies f_{5_{n_t}} = 1 & \text{para } 6 \leq n \leq 8 \\ \text{Se EXTREMA_ESQUERDA} & \implies f_{5_{n_t}} = 1 & \text{para } 7 \leq n \leq 8 \end{cases} \quad (12)$$

As ações descritas na equação 12 são ações atencionais $\mathbb{O}_{\mathbb{A}} \in \mathbb{A}_{\mathbb{A}}$.

10.3 EXP-01: Robô atencional *bottom-up*

Nesse experimento, adotou-se um robô puramente atencional, conforme o proposto por Colombini (2014). Os estímulos recebidos são tratados pelo agente como estímulos *bottom-up*. As diversas características são combinadas em um mapa de características combinadas. A seleção do vencedor no mapa de saliências – isto é, do ponto que receberá o foco da atenção – é implementada por um algoritmo *winner-takes-all*. O sistema decisório do robô (*seleção para ação*) tem influência apenas de um módulo de tomada de decisão, que executa a seguinte sistemática: *i*) rotacionar o robô sobre o próprio eixo em direção ao estímulo mais saliente, e *ii*)

depois de alinhado o robô, andar para frente de forma proporcional à saliência do estímulo. O robô foi disposto no ambiente em frente a uma caixa. O estado inicial do ambiente e o modelo de controle do agente são apresentados na [figura 27](#).

10.4 EXP-02: Robô atencional *bottom-up* e *top-down*

Neste experimento um robô atencional não consciente foi inserido no ambiente. O robô implementa os três mapas de características *bottom-up* do experimento anterior, e também uma característica *top-down*: distância-objetivo, que enaltece regiões que estejam a menos de 2,5 metros do robô. Busca-se, com isso, investigar o comportamento do agente quando sua dinâmica é guiada no ambiente pelas ações *bottom-up* e *top-down*. A seleção do vencedor no mapa de saliências – isto é, do ponto que receberá o foco da atenção – é implementada por um algoritmo *winner-takes-all*. O sistema decisório do robô (*seleção para ação*) é idêntico ao do experimento anterior. Tipicamente, espera-se que estímulos endógenos e exógenos compitam pelo foco da atenção, mas nos casos em que ambos despertem interesse por uma mesma característica, seus efeitos devem ser combinados, potencializando a chance de focar a atenção naquele elemento.

10.5 EXP-03: Robô atencional consciente

No terceiro experimento adotou-se o modelo cognitivo proposto com alguns de seus elementos principais, como o apresentado na [figura 29](#). Os elementos utilizados foram:

- Sensores: sonar e laser;
- Features *bottom-up*: $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$;
- Features *top-down*: $\mathcal{F}_4, \mathcal{F}_5$;
- Ações atencionais:
 - . \mathbb{Q}_{M1} : Rotacionar em torno do foco atencional;
 - . \mathbb{Q}_{M2} : Mover-se para a frente;
 - . \mathbb{Q}_{A1} : Direcionar o foco atencional considerando a característica top-down \mathcal{F}_4 para elementos muito próximos do agente;
 - . \mathbb{Q}_{A2} : Direcionar o foco atencional considerando a característica top-down \mathcal{F}_4 para elementos próximos do agente;
 - . \mathbb{Q}_{A3} : Direcionar o foco atencional considerando a característica top-down \mathcal{F}_4 para elementos longe do agente;

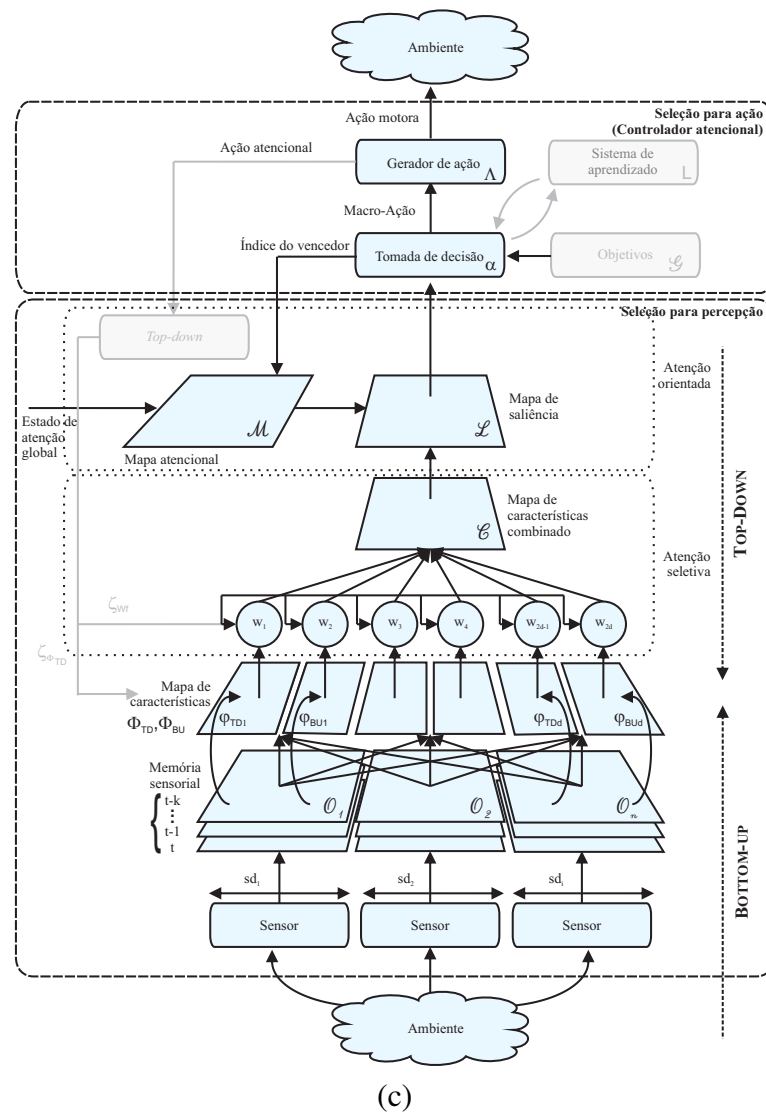


Figura 27 - Configuração do EXP01: a) Vista superior do estado inicial do ambiente: um robô atencional não consciente é disposto em frente a uma caixa; b) Vista do robô no início do experimento; c) Modelo interno do agente: utiliza-se o modelo atencional básico proposto por (COLOMBINI, 2014).

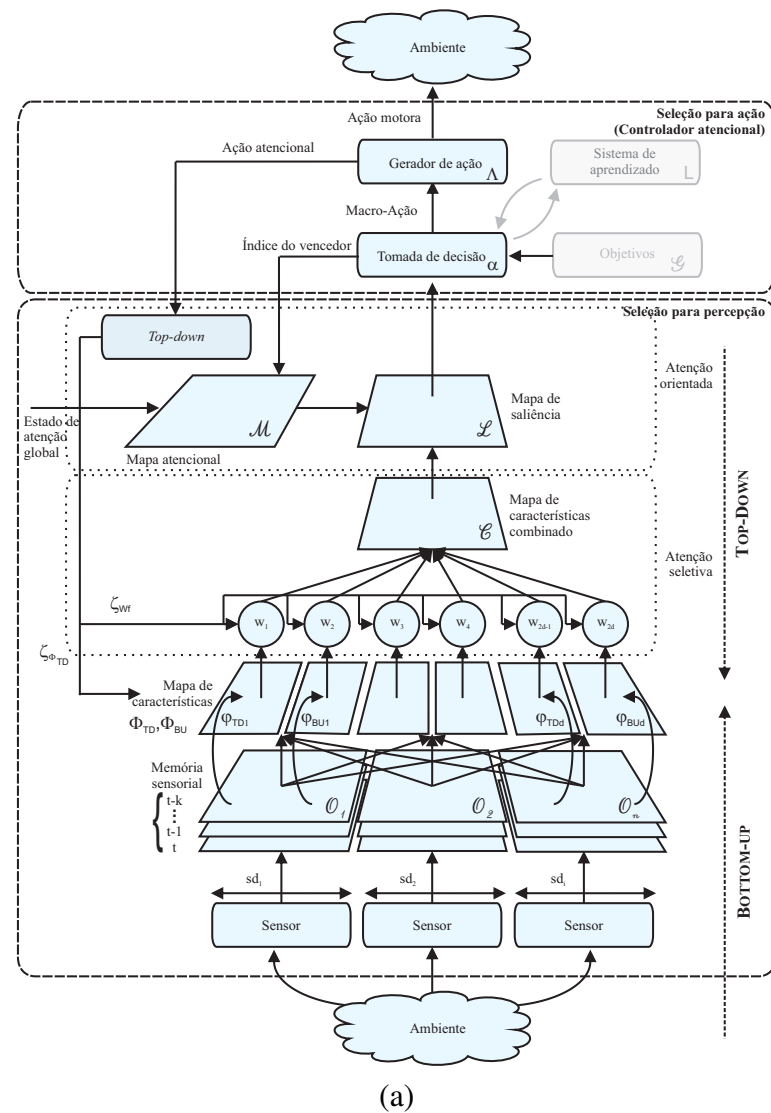


Figura 28 - Configuração do EXP02: Modelo interno do agente: utiliza-se o modelo atencional com influências *bottom-up* e *top-down* proposto por (COLOMBINI, 2014).

- . \mathbb{Q}_{A4} : Direcionar o foco atencional considerando a característica top-down \mathcal{F}_5 para a região da extrema direita do agente;
 - . \mathbb{Q}_{A5} : Direcionar o foco atencional considerando a característica top-down \mathcal{F}_5 para a região direita do agente;
 - . \mathbb{Q}_{A6} : Direcionar o foco atencional considerando a característica top-down \mathcal{F}_5 para a região à frente ao agente;
 - . \mathbb{Q}_{A7} : Direcionar o foco atencional considerando a característica top-down \mathcal{F}_5 para a região esquerda do agente;
 - . \mathbb{Q}_{A8} : Direcionar o foco atencional considerando a característica top-down \mathcal{F}_5 para a região da extrema esquerda do agente;
- Avaliação: lista de objetivos \mathbb{G} e tarefas \mathbb{T} inicialmente vazias;
 - Motivação \mathbb{MV} inicialmente vazia;
 - Função de motivação \mathbb{MO} : todo novo conceito em \mathbb{M}_W que não puder ser localizada na memória semântica do agente adiciona um elemento de motivação em \mathbb{MV} com um certo nível de intensidade, que se altera com o tempo;
 - Memória semântica \mathbb{M}_S inicialmente vazia. Forma foi adotada com o elemento chave $d_{MS} \in m_S$. A forma do objeto foi inicialmente representada usando uma nuvem de pontos. Esses pontos podem tipicamente envolver um polígono usando um algoritmo *convex hull* realizado pelo processo \mathbb{PMS} em *background* na memória semântica assim que os pontos próximos o bastante estejam disponíveis;
 - Recordação semântica \mathbb{RS} : uma busca completa na memória semântica é realizada buscando elementos que tenham pontos próximos àqueles observados no momento pelo agente de acordo com um critério de distância Euclidiana;
 - Aprendizado semântico \mathbb{LS} : as seguintes regras foram aplicadas: *i)* SE recordação $\mathbb{RS} = \emptyset$, ENTÃO criar uma nova entrada m_S em \mathbb{MS} e armazenar o dado m_W em $d_{MS} \in m_S$; *ii)* SE o objeto já existe em m_{S_i} , então anexar o dado m_W em m_{S_i} ;
 - Memória procedimental \mathbb{MP} inicialmente vazia;
 - Recordação procedimental \mathbb{RP} : uma busca completa na memória procedimental é executada buscando por tarefas \mathbb{T} que possam ser aplicadas sobre objetos m_S ;
 - Aprendizado procedimental \mathbb{LP} : implementado através de um algoritmo de aprendizado por reforço com a seguinte estrutura:
 - . Algoritmo: Q-learning;

- . Estado do agente: \mathcal{L} ;
- . Ações: \mathcal{A} conforme descrito acima;
- . Discretização do espaço de estados: 5 níveis para cada estado (total de 5^8 estados);
- . Taxa de exploração: 0,95 inicial, decaindo linearmente para 0 no última execução;
- . Taxa de aprendizado: 0,9;
- . Taxa de desconto temporal: 0,99;
- . Número máximo de tentativas de execução: 100;
- . Número máximo de passos em uma tentativa de execução: 500;
- . Estrutura de reforço:
 - $r = r + 10$ para cada novo dado inserido ao conceito sob investigação na memória semântica;
 - $r = r + 3/d$, onde d é a distância do agente ao primeiro elemento que gerou o conceito;
 - $r = r - 2$ se o robô estiver próximo de colidir;
 - $r = r - 10$ se o robô colidiu;
 - $r = r - 50$ se o robô ficou permanentemente inoperante.
- Tomada de decisão \mathbb{D} : as seguintes regras foram adotadas: *i)* SE tarefa \mathbb{t} e lista de objetivos \mathbb{G} estão ambos vazios, ENTÃO direcionar o robô para o estímulo mais saliente; *ii)* SE tarefa \mathbb{t} está vazia e lista de objetivos \mathbb{G} não está vazia, ENTÃO remover o objetivo \mathbb{g} de maior prioridade de \mathbb{G} e fazer dele a nova tarefa \mathbb{t} ; *iii)* SE tarefa \mathbb{t} requer um novo procedimento desconhecido, ENTÃO verificar o estado \mathbb{s} e executar a ação \mathbb{a} correspondente; *iv)* SE tarefa \mathbb{t} requer um procedimento desconhecido, ENTÃO aplicar o algoritmo de aprendizado e executar a ação \mathbb{a} .

Os cursos atencional e consciente são executados conforme o apresentado graficamente pela [figura 30](#). O mapa de saliência \mathcal{L} resultante, o vencedor (*winner*) do processo competitivo e suas observações \mathcal{O}_{winner} são armazenados na memória de trabalho \mathbb{M}_W . A chegada de novos dados em \mathbb{M}_W estimula recordações e aprendizado nas memórias procedimental e semântica. Como as memórias estão inicialmente vazias, nenhuma recordação é esperada nesse momento. Nesse caso, a memória semântica deve criar um novo conceito baseado nos dados recebidos do ambiente através da função de aprendizado \mathbb{L}_S . Um dos mecanismos disponíveis que pode estimular a agente a completar o conceito é a motivação. Dessa forma, tão logo o primeiro novo dado atinja \mathbb{M}_S , espera-se que uma motivação seja gerada em \mathbb{M}_V para fomentar a investigação desse conceito. A função de volição \mathbb{V} tipicamente transforma essa motivação em um objetivo, adicionando-a à lista \mathbb{G} de objetivos do agente, que estava inicialmente vazia. A função de

tomada de decisão \mathbb{D} é responsável por transformar esse objetivo em uma tarefa, inserindo-a na memória de trabalho \mathbb{M}_W de forma a recordar o procedimento para esta tarefa. Como a memória procedimental \mathbb{M}_P não tem inicialmente procedimentos armazenados, \mathbb{D} pode decidir iniciar um novo aprendizado procedimental \mathbb{L}_P para esta tarefa. Este processo é completado escolhendo-se ações adicionais $\mathbb{O}_A \in \mathbb{A}_A$ que possam enfatizar a atenção na área próxima aos pontos recentes através de ações *top-down*. O aprendizado pode continuar tipicamente até que a nuvem de pontos permita a formulação de um conceito geométrico do objeto. A cada passo do aprendizado, os pares estado-ação são armazenados na memória procedimental como uma receita para a tarefa de investigação de um novo conceito.

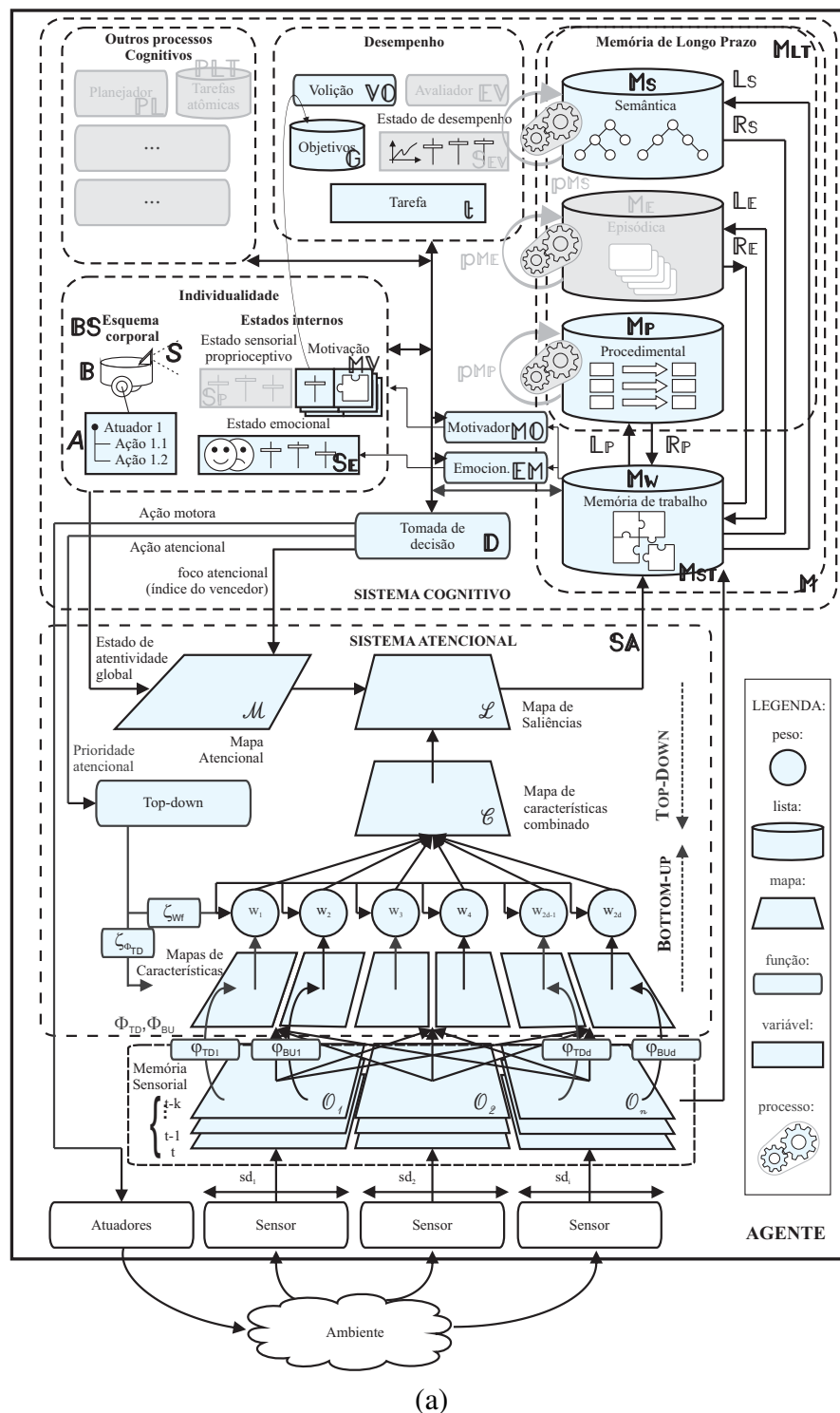


Figura 29 - Configuração do EXP03: Modelo interno do agente: utiliza-se o modelo cognitivo proposto com algumas de suas funcionalidades principais.

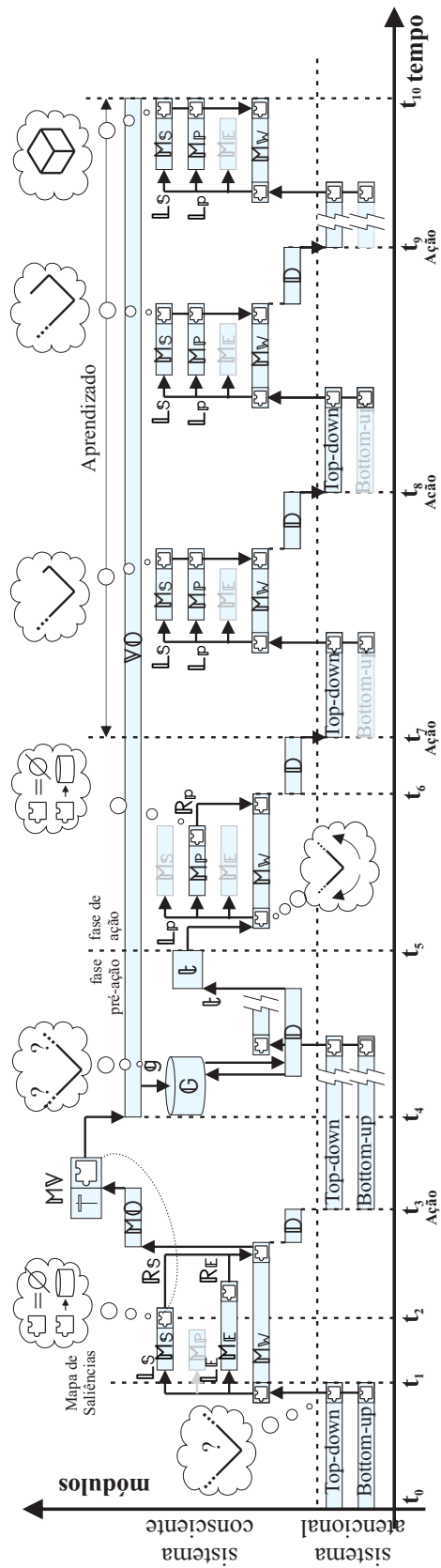


Figura 30 - Exemplo de cenário de trabalho com o modelo proposto.

11 RESULTADOS E DISCUSSÃO

“É paradoxal, mas verdadeiro, dizer que quanto mais sabemos, mais ignorantes que nos tornamos no sentido absoluto, pois é somente através da iluminação que nos tornamos conscientes de nossas limitações. Precisamente um dos resultados mais gratificantes da evolução intelectual é a abertura contínua de novas e maiores perspectivas.”

Nikola Tesla

11.1 EXP01

A [figura 31](#) apresenta a trajetória percorrida pelo robô. A [figura 32](#) e a [figura 33](#) apresentam os diversos mapas resultantes do processo atencional. O experimento demonstrou o curso atencional da seleção para percepção para um agente executando apenas o curso *bottom-up* da atenção. Apesar de possuir um comportamento simples de tomada de decisão, o modelo atencional naturalmente promove a exploração do ambiente. Entretanto, embora o agente seja capaz de responder a estímulos do ambiente, para a composição de uma arquitetura consciente seria necessária a incorporação de mecanismos que pudessem inserir os módulos cognitivos cerebrais sobre os estímulos sensoriais filtrados pela atenção.

11.2 EXP02

A [figura 34](#) apresenta a trajetória do agente neste experimento. As [figuras 35](#) e [36](#) apresentam os mapas diversos do processo atencional. Apesar da maior parte da nossa atenção ser dirigida por elementos endógenos, a pura aplicação de um processo *top-down* sem um nível cognitivo superior que possa coordenar as ações do agente parece impossibilitar a execução de tarefas mais complexas. Entretanto, o mapa de saliência utilizado como base para a tomada de decisão nesse experimento sugere que esse dado é propício para utilização por um processo

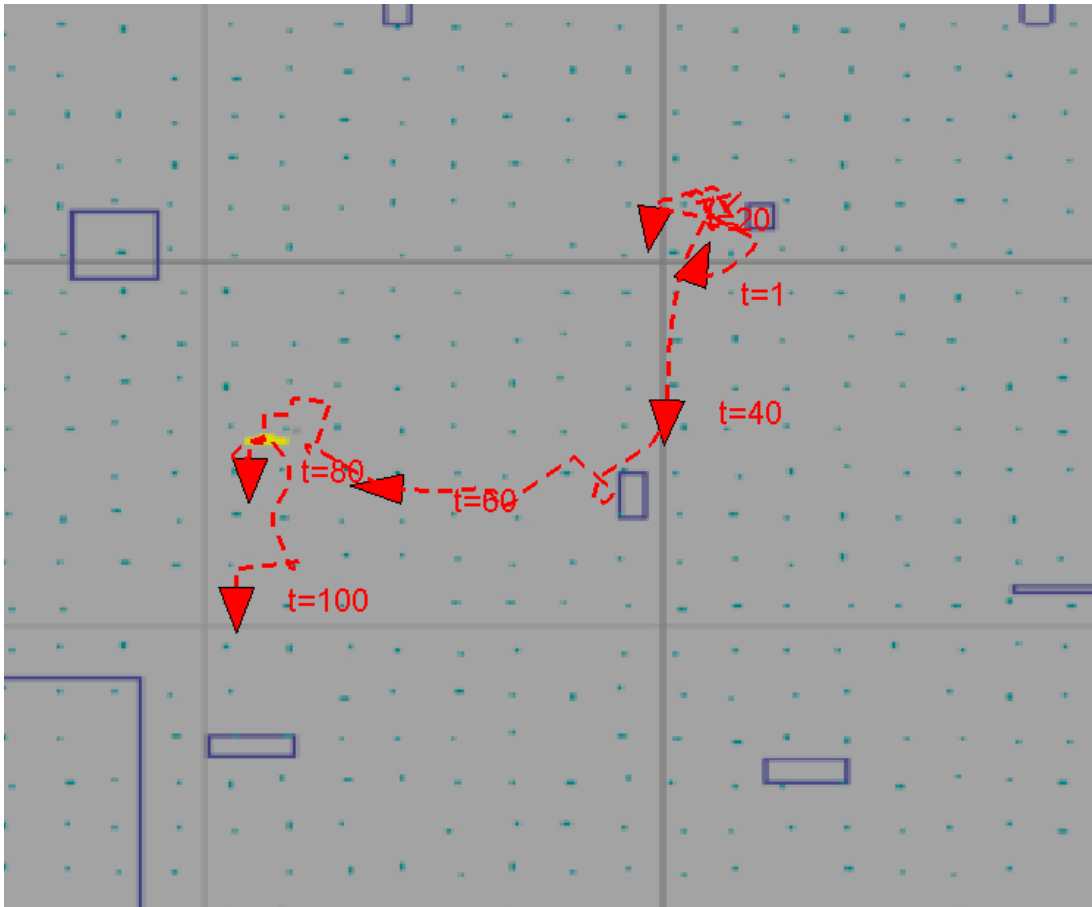


Figura 31 - **EXP01:** Trajetória (linha vermelha tracejada) percorrida pelo robô durante 100 iterações atencionais (de $t=1$ a $t=100$). As setas indicam a orientação do robô em momentos específicos. As linhas azuis contínuas indicam objetos (caixas) no ambiente. O ponto amarelo representa uma vítima (pessoa) no ambiente. O robô atencional partiu do canto superior direito da figura (equivalente à posição inicial mostrada na [figura 27a](#)). Tipicamente o robô teve sua atenção sucessivamente chaveada entre os diversos objetos presentes no ambiente, graças à discrepância entre as características levantadas sobre os estímulos e ao mecanismo de refração. O robô foi, portanto, guiado pela atenção despertada no ambiente.

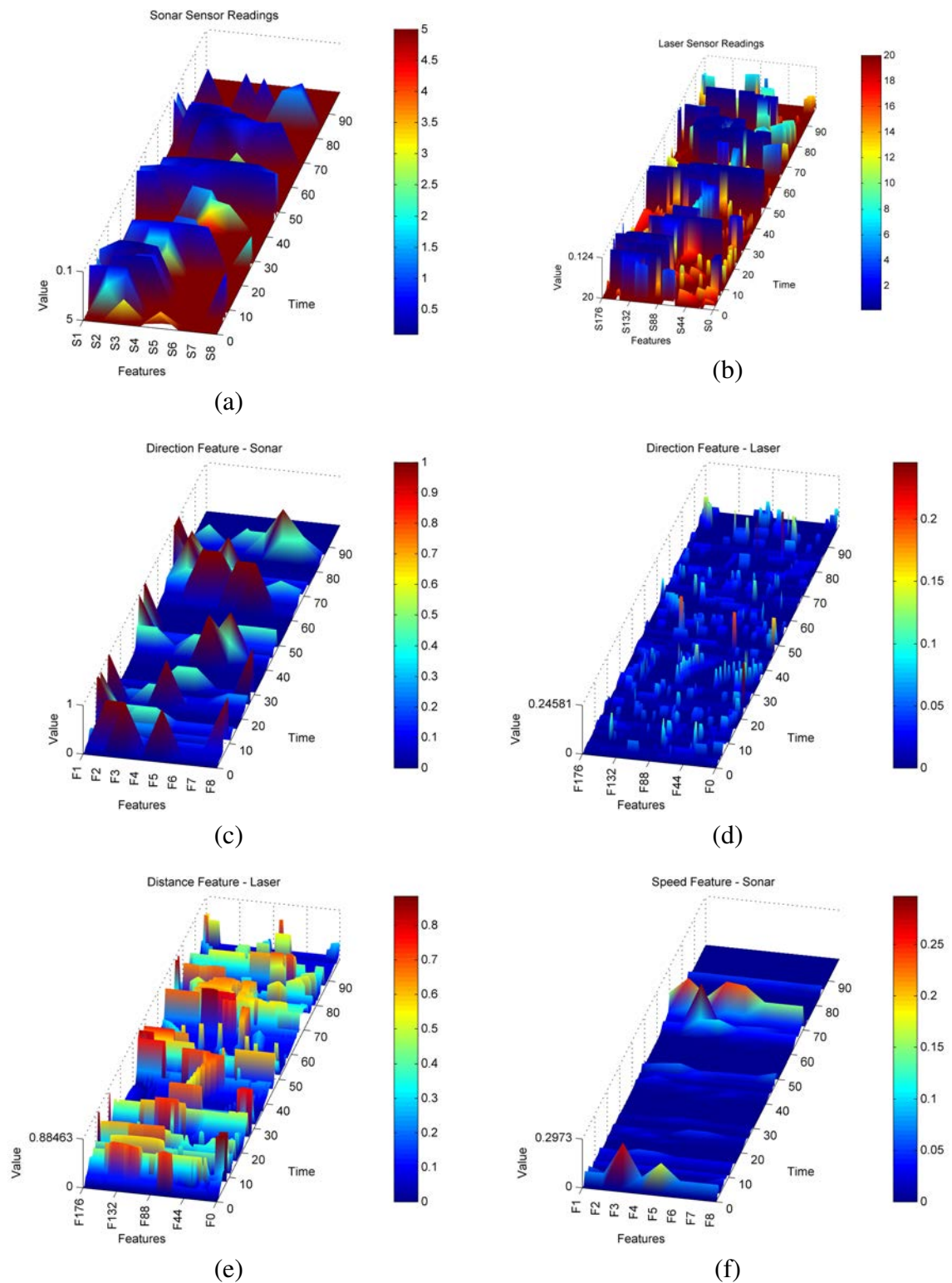


Figura 32 - **EXP01**: a) Leituras de distância dos sonares (em metros); b) Leituras de distância do sensor laser (em metros); c) Mapa de características *bottom-up* direção calculada sobre as observações dos sonares; d) Mapa de características *bottom-up* direção calculada sobre as observações do laser; e) Mapa de características *bottom-up* distância calculada sobre as observações do laser; f) Mapa de características *bottom-up* velocidade calculada sobre as observações do sonar.

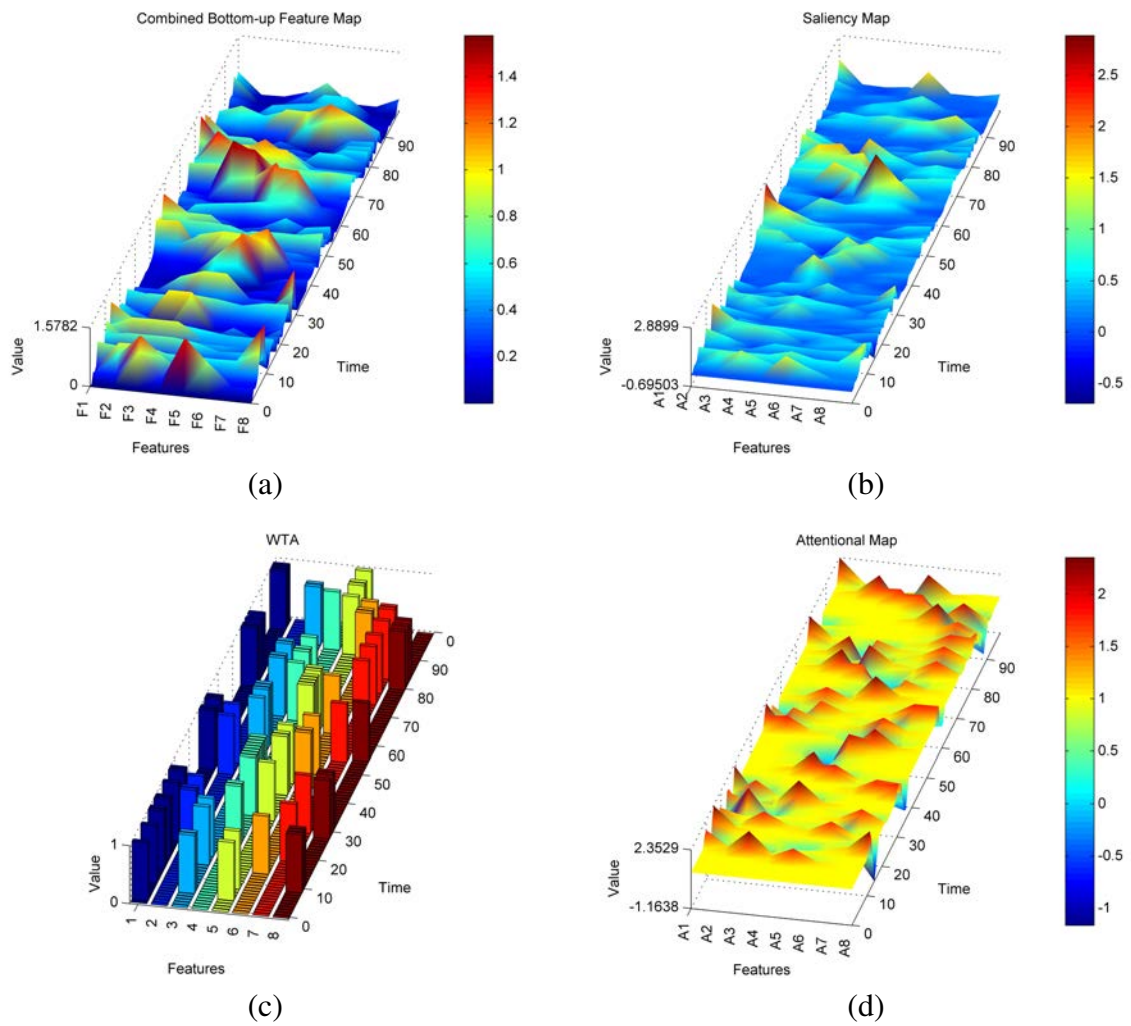


Figura 33 - **EXP01**: a) Mapa de características combinado; b) Mapa de saliências; c) Estímulos vencedores dos processos competitivos por iteração; d) Mapa atencional.

decisório de mais alto nível, visto que nos experimentos aqui realizados o sistema foi capaz de cumprir sua ação atencional *top-down* de dar importância a elementos próximos do agente e explorar as características discrepantes do ambiente.

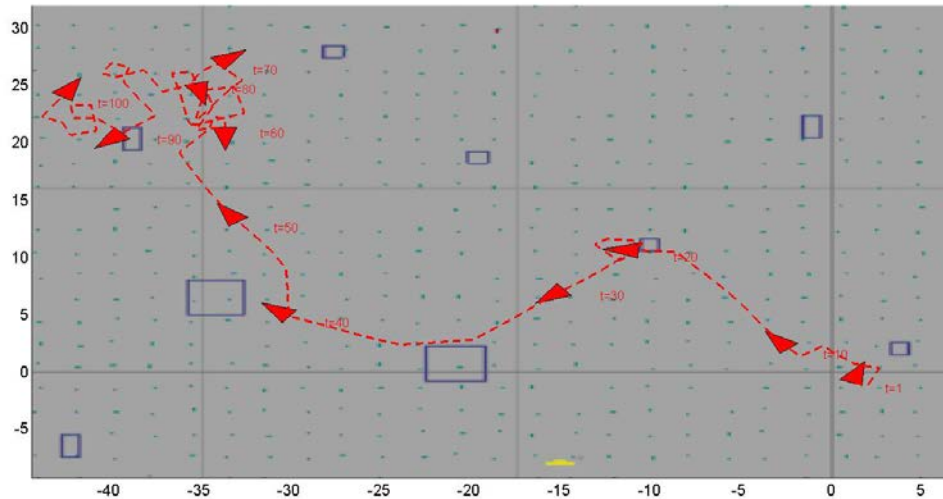


Figura 34 - **EXP02**: Trajetória percorrida pelo robô durante 100 iterações atencionais (de $t=1$ a $t=100$). Tipicamente o robô teve sua atenção sucessivamente chaveada entre os diversos objetos presentes no ambiente, graças à discrepância entre as características levantadas sobre os estímulos e ao mecanismo de refração. O robô foi, portanto, guiado pela atenção despertada no ambiente. Embora tenha havido influência *top-down*, esta não é suficiente para coordenar as ações do agente

11.3 EXP03

A dinâmica do sistema atencional guiado pelo modelo consciente proposto depois de aprendido procedimental é apresentado na [figura 37](#). A [figura 37a](#) apresenta a trajetória do agente em torno de uma baixa de $t=0$ to $t=200$ passos (dimensões em metros). A [figura 37b](#) apresenta a disposição e nomes para os 8 sonares (F1 a F8) (azul) e 180 medidas de laser (F0...F179) (vermelho). A [figura 37c](#) apresenta a representação da arquitetura adotada. A [figura 37d](#) apresenta o índice do vencedor (*winner*) do foco atencional. A [figura 37e](#) apresenta o mapa atencional \mathcal{M} . A [figura 37f](#) apresenta o mapa de saliência \mathcal{L} . A [figura 37g](#) apresenta o mapa de características combinado \mathcal{C} . A [figura 37h](#) apresenta o mapa de características *bottom-up* (discrepância de movimento). A [figura 37i](#) apresenta o mapa de características *bottom-up* do componente de sonar \mathcal{F}_2 (discrepância nas direções de movimento). A [figura 37j](#) apresenta o mapa de características *bottom-up* do componente de laser \mathcal{F}_2 (discrepância nas direções de movimento). A [figura 37k](#) apresenta o mapa de características *bottom-up* \mathcal{F}_3 (discrepância de distância). A [figura 37l](#) apresenta o mapa de características *top-down* \mathcal{F}_4 (interesse em medidas com distância específica). A

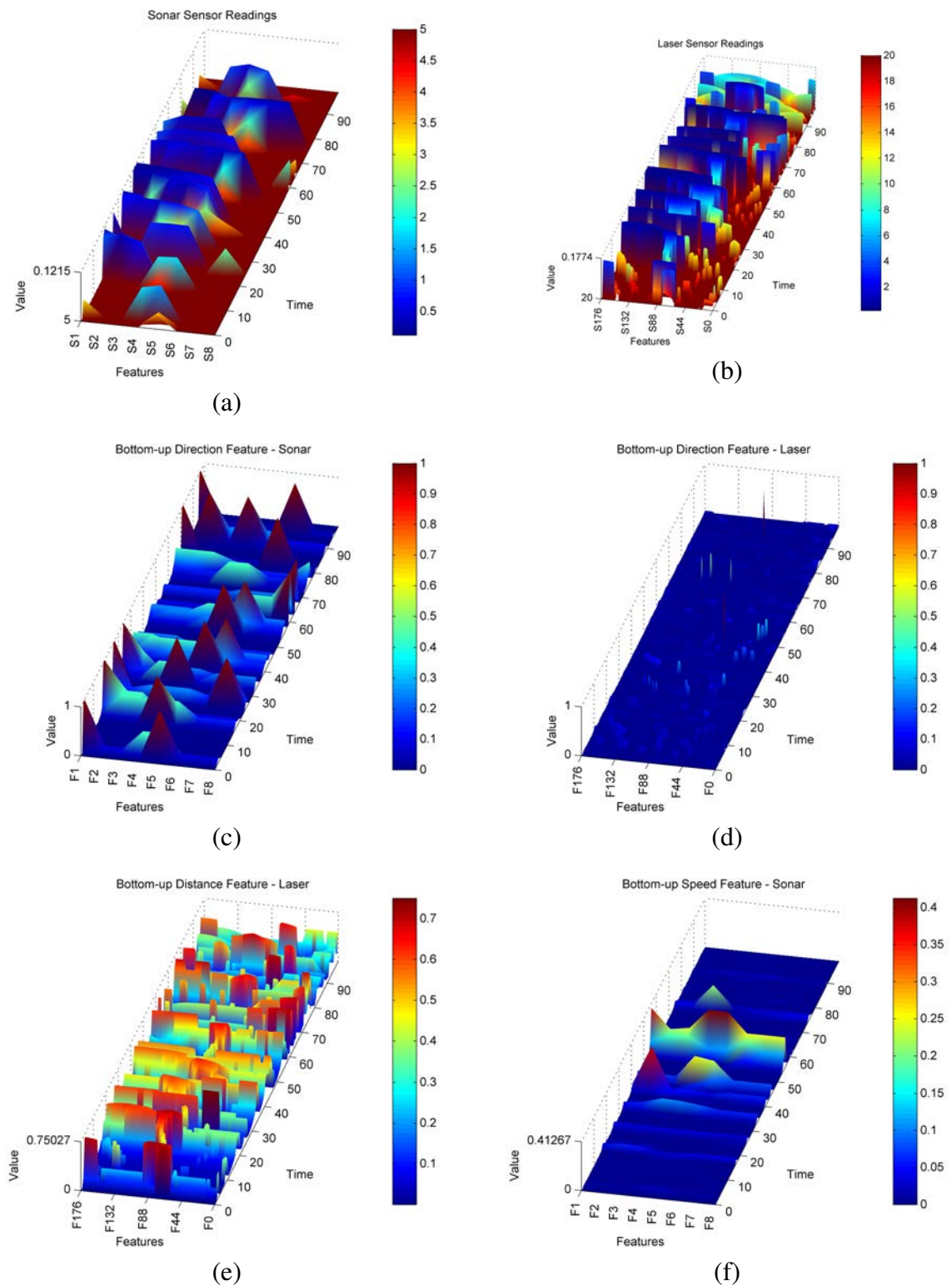


Figura 35 - **EXP02**: a) Leituras de distância dos sonares (em metros); b) Leituras de distância do sensor laser (em metros); c) Mapa de características *bottom-up* direção calculada sobre as observações dos sonares; d) Mapa de características *bottom-up* direção calculada sobre as observações do laser; e) Mapa de características *bottom-up* distância calculada sobre as observações do laser; f) Mapa de características *bottom-up* velocidade calculada sobre as observações do sonar.

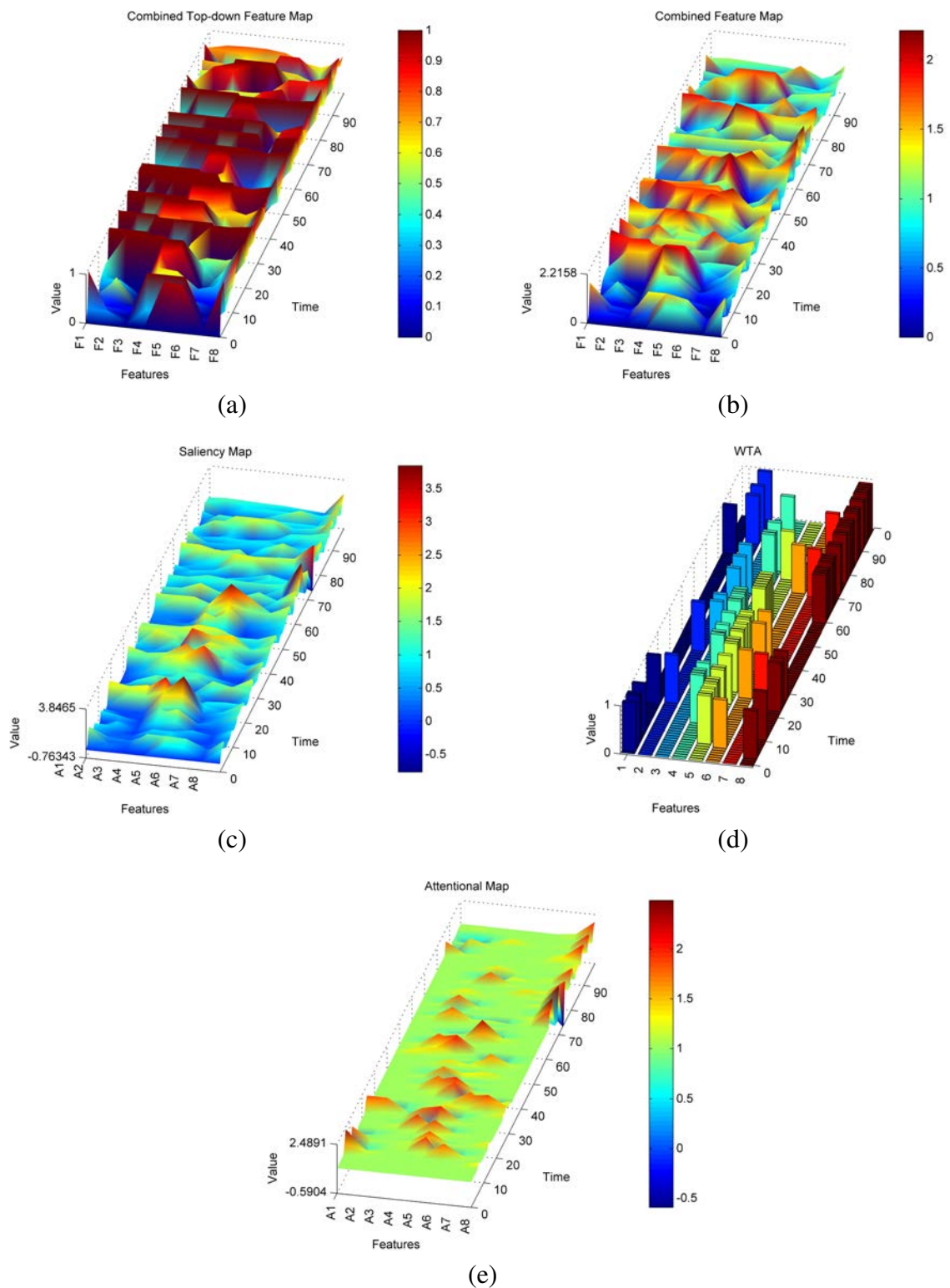


Figura 36 - **EXP02**: a) Mapa de características *top-down*; b) Mapa de características combinado; c) Mapa de saliências; d) Estímulos vencedores dos processos competitivos por iteração; e) Mapa atencional.

figura 37m apresenta o mapa de características *top-down* \mathcal{F}_5 (interesse em regiões específicas). A figura 37n apresenta as leituras dos sonares \mathcal{O}_1 (metros). Finalmente, a figura 37o apresenta as leituras dos lasers \mathcal{O}_2 (metros),

Observa-se que a figura 37m apresenta a aplicação sucessiva das ações *top-down* \mathcal{O}_{05} e \mathcal{O}_{07} (ênfase em estímulos dos lados esquerdo e direito do robô), como o decidido pelo sistema cognitivo de forma a guiar o agente na investigação da caixa.

A figura 38 compara o agente guiado por estímulo e o agente consciente. A figura 38a apresenta a trajetória de 100 tentativas a partir de 8 diferentes poses iniciais considerando que o agente era sempre guiado pelo estímulo mais saliente. A figura 38b apresenta as trajetórias do agente guiado pelo modelo consciente durante 100 ciclos de aprendizado a partir de 8 diferentes poses iniciais. O agente aprendeu a andar em torno do objeto de interesse, coletando novas informações sobre ele até que o conceito geométrico da caixa fosse formulado em sua memória semântica. As figuras 38c-d apresentam o número de pontos coletados e acumulados da caixa em ambas as abordagens. A figura 38e apresenta os valores Q médios e desvio padrão do agente ao longo das iterações de treinamento para os experimentos. A figura 38 apresenta um exemplo da aplicação de algoritmo *convex hull* na memória semântica do agente em "b" sobre a nuvem de pontos para um experimento.

Ao final desse experimento a tabela $Q(s, a)$ construída pelo *Q-learning* – um $\mathbb{C}_{SA} \in \mathbb{M}_{\mathbb{P}}$ – representa a receita de execução da tarefa aprendida armazenada. Este mesmo procedimento poderá ser recordado em futuras ocasiões, bem como poderá ser, em qualquer momento, melhorado por novas ações do agente. Em suma, através do mecanismo proposto, o agente imerso em um meio adquiriu experiência sobre esse meio ao longo do tempo.

Este experimento demonstra a viabilidade da utilização do modelo proposto para um agente consciente computando sobre o paradigma atencional. Além das características aqui apresentadas, diversas outras podem ser incorporadas ou aprimoradas. Um exemplo é o estado emocional, que poderia ser utilizado nesse cenário, por exemplo, como referência para construir a base de reforço empregada no aprendizado. Uma importante característica desta abordagem, contudo, é que nesse modelo o sistema atencional provê aos níveis cognitivos de mais alta hierarquia espaços de estados densos, que no caso do experimento apresentado, foi utilizado como base para o sistema de aprendizado, ao invés do sistema de percepção tradicional. Prontamente, nesse experimento, a utilização desse espaço atencional promoveu uma redução de 188 para 8 estados.

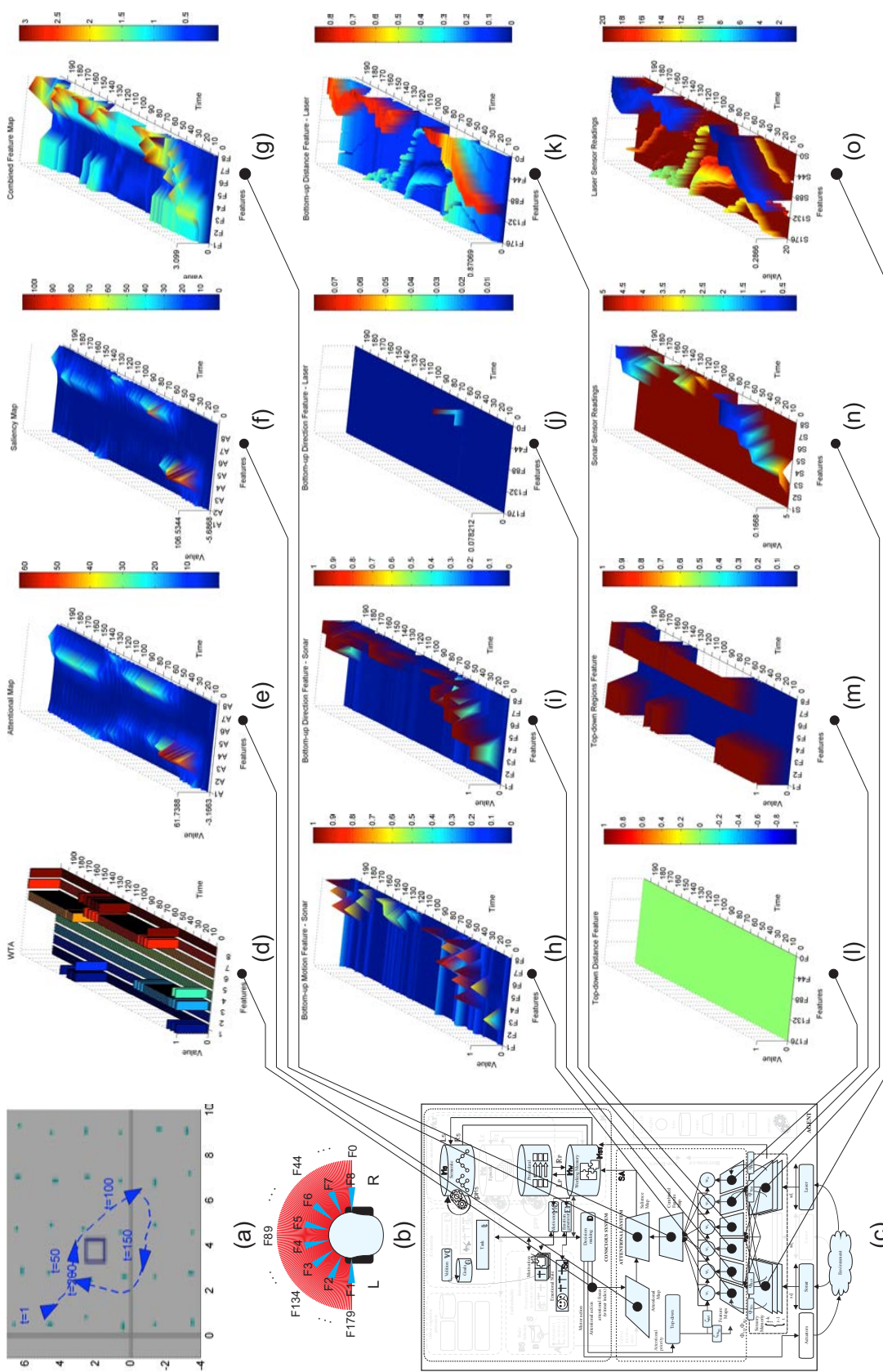


Figura 37 - EXP03: Dinâmica do agente com o modelo consciente proposto.

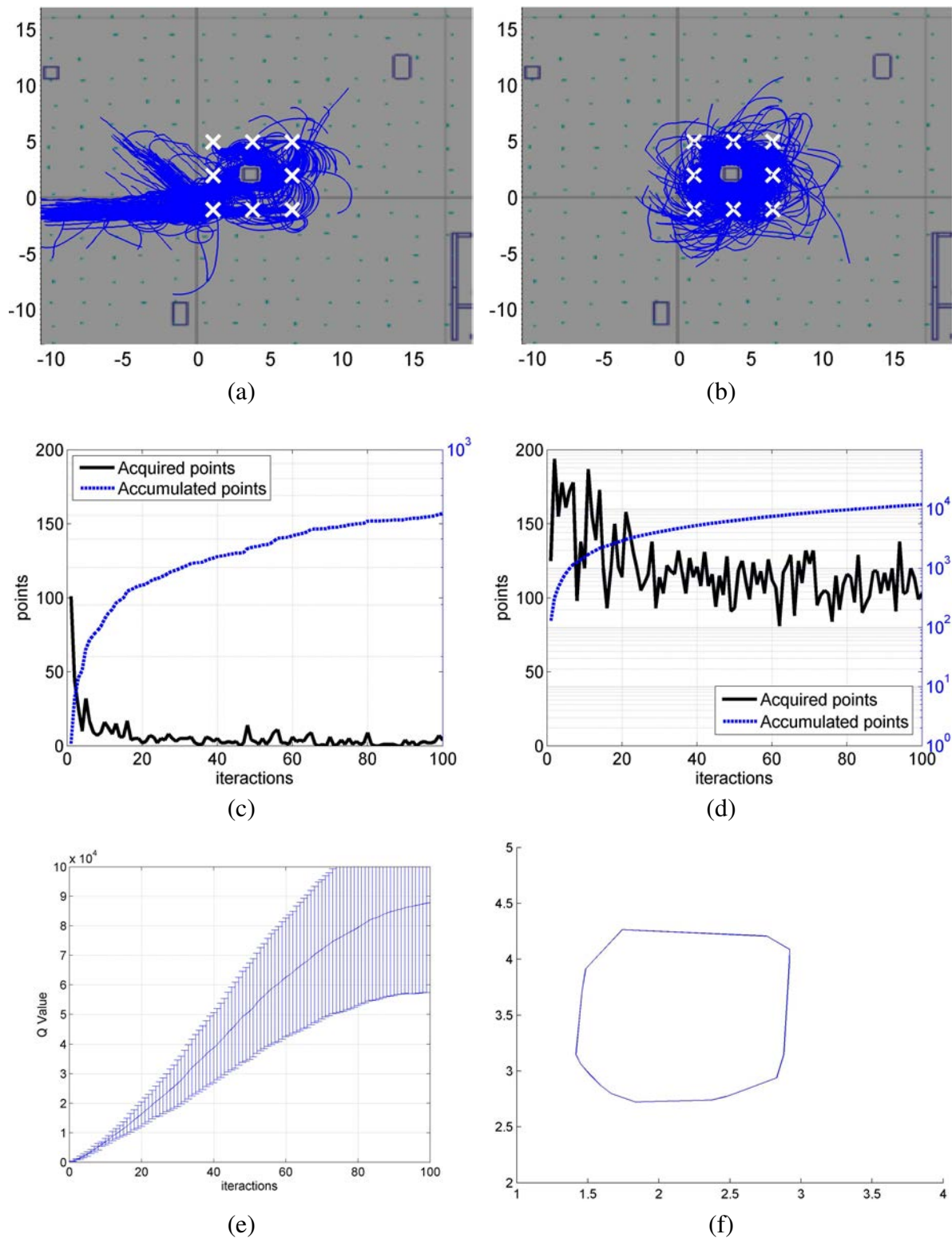


Figura 38 - **EXP03**: Trajetórias do agente com o modelo proposto quando explorando um ambiente com algumas caixas a partir de 8 diferentes poses ao redor de uma caixa.

11.4 Uma possível abordagem para o *qualia*?

qualia

Em um retorno ao problema do *qualia*, discute-se nesta seção as possíveis implicações para a interpretação das experiências subjetivas sob a ótica do modelo Colombini-Simões. Para tanto, retornemos por um momento às experiências subjetivas humanas. Ao observar um pôr do Sol no dia de hoje, por exemplo, uma pessoa pode recordar-se de várias tardes anteriores, eventos e emoções que ocorreram, trazendo esses elementos à memória. Essa pessoa também poderia realizar projeções para o seu futuro baseadas nessa observação e em suas perspectivas e objetivos pessoais, experimentar novas ou antigas emoções a partir desses pensamentos e/ou estabelecer novos objetivos para si. Trata-se, em suma, de uma experiência que envolve uma quantidade de raciocínio, experiências pessoais prévias, emoções e projeções que são, de alguma forma, maiores ou mais intensas do que as experiências cotidianas.

Sob a ótica do modelo apresentado, primeiramente vale observar que pode-se entender que os conteúdos das memórias de longo prazo $m_{LT} \in M_{LT}$ mais usualmente acessados pelo agente sejam conteúdos de grande relevância para este em seu universo. Desta forma, é factível traçarmos inicialmente um perfil da **importância dos elementos armazenados na memória** para o agente tomando como base o número de acessos durante a vida do mesmo. O modelo proposto permite, por exemplo, que essas informações de número de acessos sejam incorporadas a cada elemento de memória, facilitando assim futuros raciocínios sobre esse valor.

Da mesma forma consideremos agora que um novo elemento m_W – oriundo de uma nova observação feita pelo agente – foi inserido na memória de trabalho M_W , estimulando recordações das memórias de longo termo M_{LT} . Em função da **quantidade de recordações** realizadas e também da **importância das recordações**, é possível quantificar a importância do novo elemento m_W para o agente. É possível ainda avaliar um elemento m_W sob a ótica da **qualidade das emoções** que possam estar contidas nas memórias recordadas. Também é possível avaliar a importância de um novo elemento m_W sob a perspectiva da **importância do aprendizado para o agente**, isto é, a quantidade de conceitos, episódios ou procedimentos aprendidos nas memórias de longo termo M_{LT} utilizando suas funções de aprendizado L_E , L_P e L_S computadas sobre m_W . Em uma outra ótica, também é possível quantificar um elemento m_W baseado na **quantidade de alterações de objetivo** que este trouxe para o agente, ou simplesmente sobre a **quantidade de raciocínio** que demandou. Teoriza-se, portanto, que seria possível desenvolver uma função que leve em consideração os vários aspectos da importância de uma observação para o agente no âmbito da consciência de máquina, quantificando-a em um valor Q , que poderia ser interpretado como uma **medida da qualidade da experiência** para o agente.

Neste contexto, estabelecendo um paralelo com o cenário do pôr do Sol mencionado, pode-se teorizar que as experiências subjetivas – ou *qualias* – em agentes artificiais operando sobre o modelo proposto poderiam estar relacionados com as experiências capazes de despertar no

agente os valores mais elevados de Q , integrando-se elementos como: quantidade e importância das recordações provocadas, a importância para o aprendizado do agente, natureza das emoções provocadas no agente. Em todos os casos, especula-se que essa experiência subjetiva poderia ser simulada através do modelo aqui apresentado.

12 CONCLUSÃO

“A verdadeira viagem de descoberta não consiste em procurar novas paisagens, mas em ter novos olhos.”

Marcel Proust

Sob um prisma funcionalista, o presente trabalho apresentou uma proposta de modelo cognitivo baseado em um paradigma atencional para a **consciência de máquina** em sintonia com várias definições apresentadas para o conceito. O modelo considera elementos cognitivos das áreas de atenção, memória, emoção, aprendizado e consciência, fornecendo uma metodologia para realizar a computação através de um conjunto de itens atômicos bem definidos. Diferentemente de outros trabalhos encontrados na literatura, onde na maioria das vezes os autores apresentam apenas um breve conceito dos módulos que compõem o sistema, o modelo aqui apresentado expõe claramente todas as conexões entre seus módulos e sua forma de operação. À luz da revisão bibliográfica realizada no âmbito neste trabalho, apresenta-se aqui a primeira formalização conhecida de um modelo consciente atencional permitindo, desta forma, sua implementação e/ou replicação por outros autores.

Conforme discutido na [seção 9.5](#), o modelo proposto tem potencial para acomodar conjuntamente os conceitos de: atentividade, senciência, auto-ciência, auto-consciência, consciência autooética, possessividade, volição, perspectividade, estímulos endógenos e exógenos, atenção sustentada, seletiva, orientada, evidente e encoberta, bem como de permitir a inclusão de outros módulos (tais como processamento de linguagem, estimação de estados futuros, dentre outros). O modelo é capaz também de processar informação com a anuência do agente – deliberando através de regras explícitas concentradas, por exemplo, no módulo de tomada de decisão do agente – e também de forma inconsciente – realizando computações através de várias funções internas sem que exista necessariamente a anuência do agente. O modelo incorpora ainda algumas das visões e princípios presentes em diferentes famílias de arquiteturas conscientes disponíveis na literatura (discutidos na [seção 9.6](#)).

O modelo proposto foi submetido a testes em ambiente simulado que demonstraram a viabilidade de sua utilização para tarefas cognitivas simples, tais como: formulação de conceitos, aprendizado baseado na experimentação e agregação de diferentes experiências em uma experiência unificada. Especula-se que a arquitetura proposta seria válida também em ambientes mais complexos, com todos os módulos funcionais.

Face ao exposto, estabelece-se um modelo que tem potencial para acomodar elementos que poderiam vir a atender requisitos da **consciência funcional**, e, portanto ao **problema fácil** da consciência, relacionado aos aspectos de processamento da informação em mentes conscientes. O modelo pode também estabelecer alguma direção no que diz respeito ao **problema difícil** da consciência, relacionado às experiências subjetivas. Considerando que alguns dos modelos da área de consciência – notadamente os modelos conscientes baseados em integração de informação discutidos na [seção 6.3](#) – partem da hipótese de que a consciência corresponde à capacidade do sistema de integrar informações (TONONI, 2004), teoriza-se, conforme discutido na [seção 11.4](#), que as experiências subjetivas poderiam estar relacionadas nesta arquitetura não somente à capacidade de agregar dados, mas sim à relevância para o agente dos dados envolvidos. Desta forma, estabelece-se uma linha para trabalho futuro visando a investigação do *qualia* sob a ótica do modelo proposto.

Dentre as principais **contribuições** do presente trabalho destacam-se:

- i. A sistematização dos conceitos – muitas vezes conflitantes e dispersos – da área de consciência e consciência de máquina organizados nas seções introdutórias deste trabalho;
- ii. A proposição de um modelo para consciência de máquina que agrega vários módulos distintos em um único sistema cognitivo, pavimentando o estudo futuro de várias funções e processos que possam melhorar o desempenho do modelo em suas várias atividades;
- iii. A proposição de um modelo de sistema cognitivo integralmente compatível com o sistema atencional mais abrangente disponível na literatura;
- iv. A formalização de um modelo para implementação computacional de um agente consciente baseado em mecanismo atencional;
- v. A formalização de uma sistemática de operação para este modelo;
- vi. A investigação inicial de um novo paradigma de aprendizado: o aprendizado por reforço atencional.

REFERÊNCIAS

- ALEKSANDER, I.; DUNMALL, B. Axioms and tests for the presence of minimal consciousness in agents. *Journal of Consciousness Studies*, v. 10, p. 7–18, 2003.
- ALLPORT, A. D.; ANTONIS, B.; REYNOLDS, P. On the division of attention: a disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology*, v. 24, p. 225–235, 1972.
- ANDERSON, J. R. An integrated theory of the mind. *Psychological Review*, v. 4, n. 111, p. 1036–1060, 2004.
- ARAUJO, T. F. O.; NETO, A. T.; LIMA, A. M. N.; GUERRA, W.; COLOMBINI, E. L.; SIMÕES, A. S. *UFCG/ITA/UNESP Rescue team description paper*. São Bernardo do Campo, SP: Competição Brasileira de Robótica (CBR), 2008. Disponível em: <http://fei.edu.br/~flaviot/LARC2008/TDP/50825.pdf>. Acesso em dez/2014.
- ARGONOV, V. Y. Neural correlate of consciousness in a single electron: Radical answer to "quantum theories of consciousness". *Neuroquantology*, v. 10, n. 2, p. 276–285, 2012.
- ARMSTRONG, D. *What is consciousness? In the Nature of Mind*. Ithaca, NY: Cornell University Press, 1981.
- ARP, R. Consciousness and awareness: Switched-on rheostats: Response to de Quincey. *Journal of Consciousness Studies*, v. 14, n. 3, p. 101–106, 2007.
- ARRABALES, R.; LEDEZMA, A.; SANCHIS, A. Assessing and characterizing the cognitive power of machine consciousness implementations. In: *AAAI Fall Symposium Series*. Arlington, Virginia, USA: AAAI, 2009.
- ARRABALES, R.; LEDEZMA, A.; SANCHIS, A. Establishing a roadmap and metrics for conscious machines development. In: *Proceedings of the 8th IEEE International Conference on Cognitive Informatics*. Hong Kong: IEEE, 2009. p. 94–101.
- ARRABALES, R.; LEDEZMA, A.; SANCHIS, A. Strategies for measuring machine consciousness. *International Journal of Machine Consciousness*, n. 1, p. 193–201, 2009.
- ARRABALES, R.; LEDEZMA, A.; SANCHIS, A. Towards conscious-like behavior in computer game characters. In: *IEEE Symposium on Computational Intelligence and Games*. Milano, Italy: IEEE, 2009.
- ARRABALES, R.; LEDEZMA, A.; SANCHIS, A. ConsScale: A pragmatic scale for measuring the level of consciousness in artificial agents. *Journal of Consciousness Studies*, n. 17, p. 131–164, 2010.
- ARRABALES, R.; LEDEZMA, A.; SANCHIS, A. The cognitive development of machine

- consciousness implementations. *International Journal of Machine Consciousness*, v. 02, n. 02, p. 213–225, dez. 2010.
- BAARS, B. The conscious aspects hypothesis. *Trends in cognitive sciences*, n. 6, p. 47–52, 2002.
- BAARS, B.; FRANKLIN, S. An architectural model of conscious and unconscious brain function. *Neural Networks*, n. 20, p. 955–961, 2007.
- BAARS, B.; FRANKLIN, S. Conscious is computational: the LIDA model. *International Journal of Machine Consciousness*, n. 1, p. 23–32, 2009.
- BAARS, B.; RAMSEY, T.; LAUREYS, S. Brain, conscious experience, and the observing self. *Trends in neuroscience*, n. 26, p. 671–675, 2003.
- BAARS, B. J. *A Cognitive theory of consciousness*. Cambridge: Cambridge University Press, 1988.
- BANSAL, N.; RAJAN, N.; SRINIVASAN. Learning capability: A soar agent. In: *International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*. Ghaziabad, India: IEEE, 2014. p. 119–122.
- BROADBENT, D. E. *Perception and Communication*. Oxford: Pergamon Press, 1958.
- BROONEN, J.-P.; MARTY, M.; LEGOUT, V.; CEDRASCHI, C.; HENROTIN, Y.; SFR, S. R.; SOC, B. B. Is volition the missing link in the management of low back pain? *Joint Bone Spine*, Elsevier, v. 78, n. 4, p. 364–367, jul. 2011.
- BRYSON, J. J. A role for consciousness in action selection. *International Journal of Machine Consciousness*, 2012.
- CAGNAC, P.; NOIA, N. D.; HUANG, C.-H.; RACOCEANU, D.; CHAUDRON, L. Consciousness-driven model for visual attention. In: *International Joint Conference on Neural Networks (IJCNN)*. USA: IEEE, 2011. p. 1061–1066.
- CASTRO, F. S.; LANDEIRA-FERNANDEZ, J. Alma, corpo e a antiga civilização grega: as primeiras observações do funcionamento cerebral e das atividades mentais. *Psicologia: Reflexão Crítica*, scielo, v. 24, p. 798 – 809, 2011. ISSN 0102-7972. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttextpid=S0102-79722011000400021nrm=iso>.
- CAVE, K. R.; PASHLER, H. Visual selection mediated by location: Selecting successive visual objects. *Percept Psychophys*, v. 57, p. 421–432, 1995.
- CHALMERS, D. J. *The conscious mind*. Oxford, UK: Oxford University Press, 1996.
- CHELLA, A. Towards robot conscious perception. In: CHELLA, A.; MANZOTTI, R. (Ed.). *Artificial Consciousness*. UK: Imprint Academic, 2007. p. 124–140.
- CHELLA, A.; FRIXIONE, M.; GAGLIO, S. A cognitive architecture for robot self-consciousness. *Artificial intelligence in medicine*, n. 44, p. 147 – 154, 2008.
- CHELLA, A.; GAGLIO, S. Synthetic phenomenology and high-dimensional buffer hypothesis.

- International Journal of Machine Consciousness*, v. 04, n. 02, p. 353–365, dez. 2012.
- CHURCHLAND, P. *Matter and consciousness*. Massachusetts: MIT Press, 2004.
- CLAYTON, N. S.; BUSSEY, T. J.; DICKINSON, A. Can animals recall the past and plan for the future? *Nature*, v. 4, n. 8, p. 685–691, ago. 2003.
- CLAYTON, N. S.; DICKINSON, A. Episodic-like memory during cache recovered by scrub jays. *Nature*, v. 395, n. 6699, p. 272–274, set. 1998.
- COHEN, M. A.; CAVANAGH, P.; CHUN, M. M.; NAKAYAMA, K. The attentional requirements of consciousness. *Trends in Cognitive Sciences*, v. 16, n. 8, p. 411–417, ago. 2012.
- COLOMBINI, E. L. *An attentional model for intelligent robotics agent*. Tese (Doutorado) — Instituto Tecnológico de Aeronáutica (ITA), São José dos Campos, SP, 2014.
- COLOMBINI, E. L.; RIBEIRO, C. C. An attentive multi-sensor based system for mobile robotics. In: *IEEE International Conference on Systems, Man and Cybernetics (SMC)*. Seul: IEEE, 2012.
- COLOMBINI, E. L.; SIMÕES, A. da S.; MARTINS, A. C. G.; MATSUURA, J. P. A framework for learning in simulated humanoid robots. In: _____. *RoboCup 2007: robot soccer world cup XI*. Berlin: Springer-Verlag Berlin Heidelberg, 2008. (Lecture Notes in Computer Science, v. 5001), p. 345–352.
- COLOMBINI, E. L.; SIMÕES, A. S.; RIBEIRO, C. C. Top-down and bottom-up features combination for multi-sensor attentive robots. In: *Proceedings of the International Symposium on Attention in Cognitive Systems (ISACS). 23rd International Joint Conference on Artificial Intelligence (IJCAI)*. Beijing, China: IJCAI, 2013.
- CORBETTA, M.; SHULMAN, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, v. 3, n. 3, p. 201–215, 2002.
- COWARD, L. A.; GEDEON, T. O. Implications of resource limitations for a conscious machine. *Neurocomputing*, n. 72, p. 767–788, 2009.
- DAMASIO, A. R. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. London: Heinemann, 1999.
- DAMASIO, A. R. *Looking for spinoza: joy, sorrow and the feeling brain*. Orlando, Florida: Harcourt, 2003.
- DARWIN, C. *The expression of the emotions in man and animals*. New York: D. Appleton and Company, 1899.
- DEGANI, G.; PORGES, S. *Neuroscience Foundations of Human Performance*. United States: American Occupational Therapy Association Inc, 1990.
- DESCARTES, R. *Discurso do método*. São Paulo: Martins Fontes, 2001.
- DODD, W.; GUTIERREZ, R. The role of episodic memory and emotion in a cognitive robot.

Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on, p. 692–697, 2005.

DOTY, R. W. Alkmaion's discovery that brain creates mind: A revolution in human knowledge comparable to that of Copernicus and of Darwin. *Neuroscience*, n. 147, p. 561 – 568, 2007.

DUNCAN, J. Selective attention and the organisation of visual information. *Journal Exp. Psychol. Gen.*, v. 113, p. 501–517, 1984.

EAGLEMAN, D. *Incognito: The secret lives of the brain*. USA: Pantheon Books, 2012.

ECCLES, J. C. *How the self controls its brain*. Berlin, Heidelberg: Springer Verlag, 1994.

EDELMAN, G. *Bright Air, Brilliant Fire*. New York: Basic Books, 1992.

EINSTEIN, A.; INFELD, L. *The evolution of physics: the growth of ideas from early concepts to relativity and quanta*. Cambridge: Cambridge University Press, 1966.

EKMAN, P. Basic emotions. In: DALGLEISH, T.; POWER, M. (Ed.). *Handbook of cognition and emotion*. Sussex, UK: Wiley, 1999.

ELLIS, D. *Deus ex machina sapiens*. USA: Elysian Seller, 2011.

EYSENCK, M. W. *Psychology: An International Perspective*. London: Psychology Press, 2004.

FRANKLIN, S. IDA - a conscious artifact? *Journal of consciousness studies*, n. 10, p. 47–66, 2003.

FRANKLIN, S. Perceptual memory and learning: Recognizing, categorizing, and relating. In: *Symposium on developmental robotics*. Stanford University, Palo Alto, CA, USA: American Association for Artificial Intelligence (AAAI), 2005.

FRANKLIN, S.; KELEMEN, A.; MCCAULEY, L. IDA: a cognitive agent architecture. In: *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. California, USA: IEEE, 1998. p. 2646–2651.

FRANKLIN, S.; RAMAMURTHY, U.; DIMELLO, S. K.; MCCAULEY, L.; NEGATU, A.; SILVA, R.; DATLA, V. Lida: A computational model of global workspace theory and developmental learning. In: *AAAI Fall Symposium on AI and Consciousness: Theoretical Foundations and Current Approaches*. Arlington, Virginia, USA: AAAI, 2007.

GAMEZ, D. Progress in machine consciousness. *Consciousness and Cognition*, v. 17, n. 3, p. 887–910, set. 2008.

GAMEZ, D. Information integration based predictions about the conscious states of a spiking neural network. *Consciousness and Cognition*, v. 19, n. 1, p. 294–310, mar. 2010.

GAZZANIGA, M. S.; IVRY, R. B.; MANGUN, G. R. (Ed.). *Cognitive Neuroscience: the biology of mind*. New York: Norton, 2002.

GAZZINIGA, M. *The bisected brain*. New York: Appleton-Century-Crofts, 1970.

- GHAZANFAR, A. A.; SCHROEDER, C. E. Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, v. 10, p. 278–85, 2006.
- GRASSIA, M. Perceptual attention as the basis of consciousness. In: *PHICS - A graduate student conference on philosophy of mind, philosophy of language and cognitive science*. Ottawa, Canada: Carleton University, 2001. p. 20–39. Disponível em: <http://www.ibrarian.net/navon/page.jsp?paperid=564288&searchTerm=attention+and+perceptual>. Acesso em dez/2014.
- GRASSIA, M. Consciousness and perceptual attention: A methodological argument. *Essays in Philosophy*, v. 5, 2004. Article 18.
- GRAY, J. A.; BUHUSI, C. V.; SCHMAJUK, N. The transition from automatic to controlled processing. *Neural Networks*, 1997.
- GREENFIELD, S. *The private life of the brain*. New York: John Wiley and Sons, Inc., 2000.
- GROSSBERG, S. Linking attention to learning, expectation, competition, and consciousness. *CAS/CNS Technical Report Series*, n. 007, 2010.
- GROSSBERG, S. John Taylor, learning, attention, and consciousness. In: *International Joint Conference on Neural Networks (IJCNN)*. Dallas, Texas, USA: IEEE, 2013. p. 1–3.
- GULICK, R. V. Consciousness. In: ZALTA, E. N. (Ed.). *The Stanford Encyclopedia of Philosophy*. Spring 2014. Stanford, California, USA: Stanford University, 2014.
- HAIKONEN, P. *The cognitive approach to conscious machines*. UK: Imprint Academic, 2003.
- HAIKONEN, P. Essential issues of conscious machines. *Journal of Consciousness Studies*, n. 14, p. 72–84, 2007.
- HAIKONEN, P. O. A. *Robot Brains. Circuits and Systems for Conscious Machines*. United Kingdom: John Wiley & Sons, 2007.
- HAIKONEN, P. O. A. Qualia and conscious machines. *International Journal of Machine Consciousness*, v. 01, n. 02, p. 225–234, dez. 2009.
- HARDCASTLE, V. G. Attention versus consciousness: a distinction with a difference. *Cognitive Studies. Bulletin of the Japanese Cognitive Science Society*, n. 4, p. 56–66, 1997.
- HAWKINS, J.; BLAKESLEE, S. *On intelligence*. New York: Henry Holt and Company, 2004.
- HEINKE, D.; HUMPHREYS, G. Selective attention for identification model: simulating visual neglect. *Computer Vision and Image Understanding*, v. 100, n. 1-2, p. 172–197, out. 2005. ISSN 10773142.
- HOBBS, T. *Leviatão: ou matéria, forma e poder de uma república eclesiástica e civil*. São Paulo: Martins Fontes, 2003.
- HOLLAND, O. A strong embodied approach to machine consciousness. *Journal of consciousness studies*, n. 14, p. 97–110, 2007.

- HOLLAND, O.; GOODMAN, R. Robots with internal models. *Journal of consciousness studies*, n. 10, p. 77–109, 2003.
- HOLTON, G. *The advancement of science and its burdens*. Cambridge, USA: Harvard University Press, 1986.
- JAMES, W. *The principles of psychology*. New York: Dover, 1890.
- JAMES, W. Does consciousness exist? *Journal of Philosophy, Psychology, and Scientific Methods*, v. 1, p. 477–491, 1904.
- JOCKEL, S.; WESER, M.; WESTHOFF, D.; ZHANG, J. Towards an episodic memory for cognitive robots. p. 68–74, 2008.
- JOCKEL, S.; WESTHOFF, D.; ZHANG, J. EPIROME - a novel framework to investigate high-level episodic robot memory. In: *IEEE International Conference on Robotics and Biomimetics (ROBIO)*. Shenzhen, China: IEEE, 2007. p. 1075–1080.
- JOHNSON, A.; PROCTOR, R. W. *Attention: Theory and Practice*. United States: Sage Publications, 2004.
- JONIDES, J. Voluntary vs. automatic control over the mind's eye's movement. *Attention and Performance IX*, p. 361–394, 1981.
- KAHNEMAN, D. *Attention and Effort*. United States: Prentice Hall, 1973.
- KANT, I. *Critique of pure reason*. The Project Gutenberg EBook, 2003. Acesso em dez/2014. Versão original de 1781. Disponível em: <<http://www.gutenberg.org/ebooks/4280>>.
- KANWISHER, N. G.; DRIVER, J. Objects, attributes and visual attention: which, what and where. *Current Directions in Psychological Science*, v. 1, p. 26–31, 1992.
- KASDERIDIS, S.; TAYLOR, J. G. Attention-based learning. In: *IEEE International Joint Conference on Neural Networks 2004*. Budapest: IEEE, 2004. v. 1, p. 525–530. ISSN 1098-7576.
- KATAYAMA, K.; TAKENO, J. Conscious expectation system. In: SAMSONOVICH, A.; JÓHANNSDÓTTIR, K. (Ed.). *Proceedings Biologically Inspired Cognitive Architectures*. Arlington, Virginia, USA: IOS Press, 2011. p. 180–187.
- KITAMURA, T.; TAHARA, T.; ASAMI, K. How can a robot have consciousness? *Advanced Robotics*, n. 14, p. 263 – 275, 2000.
- KLEIN, R. M. Inhibition of return: review. *Trends in Cognitive Science*, v. 72, n. 1, p. 76–85, 2000.
- KOCH, C. *The quest for consciousness: a neurobiological approach*. USA: Roberts and Publishers, 2004.
- KOCH, C.; TSUCHIYA, N. Attention and consciousness: two distinct brain processes. *Trends in Cognitive Sciences*, v. 11, n. 1, p. 16–22, jan. 2007.

- KUIPERS, B. Consciousness: Drinking from the firehose of experience. In: *Proceedings of the 20th national conference on artificial intelligence (IAAA)*. Pittsburgh, Pennsylvania, USA: AI Press, 2005. p. 1298–1305.
- KUIPERS, B. Drinking from the firehose of experience. *Artificial Intelligence in Medicine*, n. 44, p. 155–170, 2008.
- KURZWEIL, R. *A era das máquinas espirituais*. São Paulo: Editora Aleph, 2007.
- LABERGE, D.; BROWN, V. Theories of attentional operations in shape identification. *Psychologica Review*, v. 96, p. 101–124, 1989.
- LAIRD, J. E.; NEWELL, A.; ROSENBLOOM, P. S. Soar - an architecture for general intelligence. *Artificial Intelligence*, v. 1, n. 33, p. 1–64, 1987.
- LAMME, V. A. Why visual attention and awareness are different. *Trends Cogn. Sci.*, n. 7, p. 12–18, 2003.
- LEE, K.; BUXTON, H.; FENG, J. Cue-Guided Search: A Computational Model of Selective Attention. *IEEE Transactions on Neural Networks*, v. 16, n. 4, p. 910–924, 2005.
- LEISMAN, G.; KOCH, P. Networks of conscious experience: computational neuroscience in understanding life, death, and consciousness. *Reviews in the Neurosciences*, v. 20, n. 3-4, p. 151–176, jan. 2009.
- LENT, R. *Cem bilhões de neurônios*. São Paulo: Atheneu Editora, 2002.
- LI, C.; KOCHIYAMA, T.; WU, J.; CHUI, D.; TSUGE, T.; OSAKA, K. Attention systems and neural responses to visual and auditory stimuli : an fmri study. In: *IEEE/ICME International Conference on Complex Medical Engineering*. Beijing: IEEE, 2007.
- LIMA, R. F. D. Compreendendo os Mecanismos Atencionais. *Ciências e Cognição*, v. 6, n. 113-122, p. 113–122, 2005.
- LING, S.; CARRASCO, M. When sustained attention impairs perception. *Nature Neuroscience*, v. 9, p. 1243–1245, 2006.
- MAASS, W.; BISHOP, C. M. (Ed.). *Pulsed Neural Networks*. Cambridge: MIT Press, 1999.
- MARKS, C. E. *Commissurotomy, consciousness, and utility of mind*. Massachusetts: MIT Press, 1981.
- MARQUES, H. G.; HOLLAND, O. *Architectures for functional imagination*, n. 72, p. 743–759, 2009.
- MASLIN, K. *An Introduction to the Philosophy of Mind*. United Kingdom: Polity Press, 2001.
- McDERMOTT, D. Artificial intelligence and consciousness. In: _____. Cambridge: Cambridge University Press, 2007. (Cambridge handbook of consciousness), p. 117–150.
- MERIKLE, P. M.; JOORDENS, S. Parallels between perception without attention and perception without awareness. *Conscious Cogn.*, n. 6, p. 219–236, 1997.

- METZINGER, T. The subjectivity of subjective experience: A representationalist analysis of the first-person perspective. *Neural correlates of consciousness: Empirical and conceptual questions*, p. 285–306, 2000.
- MINSKY, M. *The society of mind*. New York: Simon & Schuster, Inc., 1985.
- MORAY, N. *Attention: Selective Process in Vision and Hearing*. London: Hutchinson Educational, 1969.
- MORENO, R. A.; ESPINO, A. L.; MIGUEL, A. S. de. Modeling consciousness for autonomous robot exploration. In: *International Work-Conference on the Interplay between Natural and Artificial Computation (IWINAC)*. Murcia, Spain: Lecture Notes in Computer Science, 2007. v. 4527, p. 51–60.
- NAGEL, T. Brain bisection and the utility of consciousness. n. 22, p. 396–413, 1971.
- NAGEL, T. What is it like to be a bat? In: *Mortal Questions*. Cambridge: Cambridge University Press, 1979. p. 166.
- NARAIN, C. Pay attention now. *Nature Neuroscience*, v. 7, p. 832–833, 2006.
- NAVALPAKKAM, V.; ITTI, L. Modeling the influence of task on attention. *Vision Research*, v. 45, n. 2, p. 205–231, jan. 2005.
- NEISSER, U. *Cognition and Reality*. San Francisco: W. H. Freeman, 1976.
- NEVID, J. S. *Psychology concepts and applications*. 3rd. ed. USA: Houghton Mifflin Company, 2009.
- NICHOLS, S.; STICH, S. A cognitive theory of pretense. *Cognition*, n. 74, p. 115–147, 2000.
- NICHOLS, S.; STICH, S. *Mindreading*. Oxford, UK: Oxford University Press, 2003.
- NISARGADATTA, M. *I am that*. Bombay: Chetana Publishing, 1973.
- NORMAN, D. Toward a theory of memory and attention. *Psychological Review*, p. 522–536, 1968.
- NUXOLL, A.; LAIRD, J. E. A cognitive model of episodic memory integrated with a general cognitive architecture. In: *Proceedings of the 6th International Conference on Cognitive Modeling (ICCM)*. Pittsburgh, PA, USA: Lawrence Erlbaum Associates, 2004. p. 220–225.
- NUXOLL, A.; LAIRD, J. E. Extending cognitive architecture with episodic memory. In: *Proceedings of the 21st National Conference on Artificial Intelligence*. Vancouver, Canada: American Association for Artificial Intelligence (AAAI), 2007. p. 1560–1564.
- O'REGAN, J. K.; NOE, A. A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.*, n. 24, p. 939–973, 2001.
- PANKSEPP, J. Affective consciousness: core emotional feelings in animals and humans. *Conscious cognition*, v. 14, p. 30–80, 2005.

- PARASURAMAN, R. (Ed.). *The Attentive Brain*. Cambridge, Massachusetts, USA: The Mit Press, 1998.
- PERLIS, D. Consciousness as self-function. *Journal of consciousness studies*, n. 4, p. 509–525, 1997.
- PINKER, S. *Como a mente funciona*. São Paulo: Companhia das Letras, 2001.
- PLATO. *Platho's Phaedo*. USA: Hackett Publishing Company Inc., 1977.
- PLATÃO. *A República - texto integral*. São Paulo: Saraiva, 2011.
- POSNER, M. I. Orienting of attention. *The Quarterly Journal of Experimental Psychology*, v. 32, p. 2–25, 1980.
- POSNER, M. I. *Cognitive Neuroscience of Attention*. New York: Taylor and Francis Inc, 2004.
- POSNER, M. I.; COHEN, Y. Components of visual orienting. *Attention and performance X Control of language processes*, Erlbaum, v. 32, n. 4, p. 531–556, 1984.
- POSNER, M. I.; DEHAENE, S. Attentional networks. *Trends Neuroscience*, v. 17, p. 75–79, 1994.
- POSNER, M. I.; ROTHBART, M. K. Attention, self-regulation and consciousness. *Philosophical Transactions of the Royal Society B-Biological Sciences*, v. 353, n. 1377, p. 1915–1927, 1998.
- PUC CETTI, R. Brain bisection and personal identity. *British J for the philosophy of science*, n. 24, p. 339–355, 1973.
- REALE, G. *Corpo, alma e saúde: O conceito de homem de Homero a Platão*. São Paulo: Paulus, 2002.
- REGGIA, J. A. The rise of machine consciousness: studying consciousness with computational models. Elsevier, v. 44, p. 112–131, 2013.
- ROBERTS, R. This is beyond empathy: do these amazing conjoined twin girls share one mind? *Daily Mail*, maio 2011. Disponível em: <http://www.dailymail.co.uk/news/article-1390948/Do-amazing-conjoined-twin-girls-share-ONE-mind.html>. Acesso em dez/2014.
- SAMSONOVICH, A.; ASCOLI, G. The conscious self: ontology, epistemology and the mirror quest. *Cortex*, n. 41, p. 621–636, 2005.
- SAMSONOVICH, A.; DEJONG, K. A general-purpose computational model of the conscious mind. In: LOVETT, M.; SCHUNN, C.; LEBIERE, C.; MUNRO, P. (Ed.). *Proceedings of the sixth International Conference on Cognitive Modeling (ICCM)*. Pittsburgh, Pennsylvania, USA: Lawrence Erlbaum Associates, 2004. p. 382–383.
- SAMSONOVICH, A.; NADEL, L. Fundamental principles of mechanisms of the conscious self. *Cortex*, n. 41, p. 669–689, 2005.

- SCHACTER, D. L.; WAGNER, A. D.; BUCKNER, R. L. The Oxford handbook of memory 2000. In: TULVING, E.; CRAIK, F. I. M. (Ed.). New York: Oxford University Press, 2000. cap. Memory Systems of 1999, p. 627–643.
- SCHRÖDINGER, E. *What is life?* Cambridge: Cambridge University Press, 1944. Disponível em: http://whatislife.stanford.edu/LoCo_files/What-is-Life.pdf. Acesso em dez/2014.
- SEARLE, J. Minds, brains and programs. *Behavioral and brain sciences*, v. 3, n. 3, p. 417–457, 1980.
- SEARLE, J. R. *Mind: A brief introduction*. Oxford: Oxford University Press, 2004.
- SEARLE, J. R. *O mistério da consciência*. São Paulo: Paz e Terra, 2004.
- SETH, A. K. Causal connectivity of evolved neural networks during behavior. *Network: computation in neural systems*, n. 16, p. 35, 2005.
- SHANAHAN, M. A spiking neuron model of cortical broadcast and competition. *Consciousness and Cognition*, v. 17, n. 1, p. 288–303, jan. 2008.
- SHANAHAN, M. P. Cognition, action selection, and inner rehearsal. In: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), Workshop on Modelling Natural Action Selection*. Edinburgh, Scotland, UK: IJCAI, 2005. p. 92–99.
- SHANNON, C. A mathematical theory of communication. *Bell System Technical Journal*, n. 27, p. 379–423, 1948.
- SILVERMAN, A. Plato's middle period metaphysics and epistemology. In: ZALTA, E. N. (Ed.). *The Stanford Encyclopedia of Philosophy*. Stanford, California, USA: Stanford University, 2012.
- SIMÕES, A. S.; COLOMBINI, E. L. *PROSE - Problem Search Environment*. 2014. Disponível em: <http://code.google.com/p/prose/>.
- SIMÕES, A. S.; COLOMBINI, E. L.; PRADO, G. I.; MARTINS, A. C. G. Enhancing exploration in rescue environments through a neurofuzzy system. In: *Simpósio Brasileiro de Automação Inteligente (SBAI)*. Brasília, DF: Sociedade Brasileira de Automática (SBA), 2009.
- SKINNER, B. F. *Verbal behavior*. USA: Copley Publishing Group, 1992.
- SLOMAN, A. Cross-disciplinary reflections: Philosophical robotics. *Cognitive Systems*, v. 8, p. 441–484, 2010.
- SLOMAN, A.; CHRISLEY, R. Virtual machines and consciousness. *Journal of consciousness studies*, n. 10, p. 133–172, 2003.
- SNELL, B. *A cultura grega e as origens do pensamento europeu*. São Paulo: Perspectiva, 2001.
- SPERRY, R. Brain bisection and the neurology of consciousness. Springer-Verlag, New York, p. 298–313, 1966.

- SQUIRE, L. R. Memory systems of the brain: a brief history and current perspective. *Neurobiol. learn. Memory*, v. 82, n. 3, p. 171–177, 2004.
- STACHOWICZ, D.; KRUIJFF, G. M. Episodic-like memory for cognitive robots. *Autonomous Mental Development, IEEE Transactions on*, v. 4, n. 1, p. 1–16, 2012.
- STARZYK, J. A.; PRASAD, D. K. Machine consciousness: A computational model. *Brain-inspired Cognitive Systems*, 2010.
- STARZYK, J. A.; PRASAD, D. K. A computational model of machine consciousness. *International Journal of Machine Consciousness*, n. 3, p. 255–281, 2011.
- STRATTON. *Semicentennial publications of the University of California 1868–1918*. Califórnia: Universidade da Califórnia, 1911. Contém a tradução de “*On the Senses*”. Disponível em: <https://ia601407.us.archive.org/5/items/theophrastusgree00stra/theophrastusgree00stra.pdf>. Acesso em dez/2014.
- SUN, R. *Duality of the mind*. California, USA: Erlbaum.
- SUN, R. Learning, action and consciousness. *Neural Networks*, n. 10, p. 1317 – 1331, 1997.
- SUN, R. Accounting for the computational basis of consciousness. *Consciousness and cognition*, n. 8, p. 529 – 565, 1999.
- SUN, R.; FRANKLIN, S. Computational models of consciousness. In: ZELAZO, P.; MOSCOVITCH, M. (Ed.). *Cambridge handbook of consciousness*. Cambridge: Cambridge University Press, 2007. p. 151–174.
- TAKENO, J. MoNAD structure and self-awareness. In: SAMSONOVICH, A.; JÓHANNSDÓTTIR, K. (Ed.). *Proceedings biologically inspired cognitive architectures*. Arlington, Virginia, USA: IOS Press.
- TAKENO, J. A robot succeeds in 100% mirror image cognition. *International journal on smart sensing and intelligent systems*, n. 1, p. 891–911, 2008.
- TAKENO, J. *Creation of a conscious robot*. Singapura: Pan Stanford, 2013.
- TALLON-BAUDRY, C. On the neural mechanisms subserving consciousness and attention. *Frontiers in Psychology*, v. 2, n. Article 397, p. 397–397, jan. 2012.
- TAYLOR, J. G. The creativity effect: Consciousness versus attention. In: *International Joint Conference on Neural Networks (IJCNN)*. Barcelona, Spain: IEEE, 2010. p. 1–8.
- TINSLEY, C. J. Using topographic networks to build a representation of consciousness. *Biosystems*, v. 92, n. 1, p. 29–41, jan. 2008.
- TIPPER, S. P. The negative priming effect: Inhibitory priming by ignored objects. *Experimental Psychology*, v. 37A, p. 571–590, 1985.
- TONONI, G. An information integration theory of consciousness. *BMC Neuroscience*, v. 5, n. 42, 2004.

- TONONI, G. Consciousness as integrated information: A provisional manifesto. *Biological Bulletin*, v. 215, p. 216–242, 2008.
- TREISMAN, A. M.; GELADE, G. A Feature-Integration Theory of Attention. *Cognitive Psychology*, p. 97–136, 1980.
- TULVING, E. Organization of memory. In: TULVING, E.; DONALDSON, W. (Ed.). New York: Academic, 1972. cap. Episodic and semantic memory, p. 381–403.
- TURING, A. M. Computing machinery and intelligence. *Mind*, n. 59, p. 433–460, 1950.
- VALDUGA, A. Como o intelecto conhece segundo Aristóteles. *Revista Eletrônica Frontistés*, Faculdade Palotina (FAPAS), v. 1, n. 1, 2007.
- van der VELDE, F.; DEKAMPS, M.; VANDERVOORTVANDERKLEIJ, G. CLAM: closed-Loop attention model for visual search. *Neurocomputing*, v. 58-60, p. 607–612, June 2004. ISSN 09252312.
- VELMANS, M. *The Science of Consciousness*. UK: Routledge, 1996.
- WANG, J.; BALAKIRSKY, S. *USARSim V3.1.3 Manual: A Game-based Simulation of the NIST Reference Arenas*. United States, 2006.
- WANG, Y.; LAIRD, J. *Integrating semantic memory into a cognitive architecture*. Michigan, 2007. Technical Report - CCA-TR-2006-02.
- WANG, Y.; WANG, Y.; PATEL, S.; PATEL, D. A layered reference model of the brain (lrmb). *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, v. 36, n. 2, p. 124–133, 2006.
- WEIZENBAUM, J. Eliza: a computer program for the study of natural language communication between man and machine. *Communication of ACM*, n. 9, p. 36–45, 1996.
- WOODMAN, G. F.; LUCK, S. J. Dissociations among attention, perception, and awareness during object-substitution masking. *Psychol. Sci.*, n. 14, p. 605–611, 2003.
- WRIGHT, R. D.; WARD, L. *Orienting of attention*. New York: Oxford University Press, 2008.
- ZIEMKE, T.; LOWE, R. On the role of emotion in embodied cognitive architectures: From organisms to robots. *Cogn Comput*, Elsevier, v. 1, p. 104–117, 2009.
- ZLATEV, J. The mimetic origins of self-consciousness in phylo-, onto- and robotogenesis. *26th Annual Conference of the IEEE Industrial Electronics Society (IECON)*, v. 4, p. 2921–2928, 2000.

Índice Remissivo

- ação, 95
 ação atencional, 95
 ação motora, 95
 ACT-R, 82
 alma, 24
 aprendizado episódico, 98, 100
 aprendizado procedimental, 98, 100
 aprendizado semântico, 98, 100
 arquitetura de Colombini, 57
 atenção, 29, 48, 89
 atenção bottom-up, 32
 atenção conceitos, 32
 atenção dividida, 33, 61
 atenção encoberta, 33, 61, 104
 atenção endógena, 32
 atenção evidente, 33, 61, 104
 atenção exógena, 32
 atenção modelos, 74
 atenção orientada, 32, 61, 81, 104
 atenção seletiva, 32, 61, 104
 atenção sustentada, 33, 61, 103
 atenção top-down, 32
 atenção x consciência, 43
 atenção, taxonomia, 31
 atentividade, 38, 39, 103
 auto-ciência, 39, 72, 103
 auto-consciência, 39, 103
 avaliação, 97
 axiomas para consciência mínima, 47

 behaviorismo, 27
 bottom-up, 32, 76, 109

 cefalocentrismo, 23
 cenário, 109
 cenário de trabalho, 84
 CERA, 70
 CERA-CRANIUM, 69, 73
 chunk, 80
 ciclo cognitivo, 68
 CLAM, 55

 CLARION, 73
 CLS, 51
 CODAM, 55
 codelet, 68
 cogito ergo sum, 25
 como avaliar a consciência?, 46
 competição, 99
 computa-cionalismo, 27
 conceito do modelo de Simões, 91
 conceitos da atenção, 32
 conceitos da consciência, 38
 conclusões, 134
 configuração de sensores, 108
 conscale, 48
 consciência acessível, 45
 consciência auto-noética, 39, 103
 consciência avaliação, 46
 consciência axiomas, 47
 consciência de máquina, 20, 37
 consciência fenomenológica, 44
 consciência funcional, 45, 135
 consciência indivisibilidade, 45
 consciência natural e artificial, 38
 consciência taxonomia, 66
 consciência x atenção, 43
 contínuo de consciência, 37
 contribuições, 135
 controlador atencional de Colombini, 61
 corpo-mente em culturas antigas, 22
 corpus hippocraticum, 24
 CQS, 50
 craniopagus, 46
 CRANIUM, 71
 criatividade, 76
 critérios memória episódica, 42
 CRONOS, 72
 CS, 48
 culturas antigas, 22

 desempenho, 94

- diferenciação, 40
doutrinas filosóficas, 24
dualismo, 25
dualismo de propriedade, 26
dualismo de substância, 26
- egosfera sensória, 80
elemento primordial, 23
emoção, 48, 79, 97
endógeno, 32
Espaço de observação, 109
espaço de qualia, 72
espaço de trabalho global, 67
esquema corporal, 98
estado atencional, 95
estado de coma, 46
estado de desempenho, 96, 101
estado emocional, 96, 101
estado mental, 83
estado minimamente consciente, 46
estado proprioceptivo, 96
estado vegetativo, 46
estados, 95
exógeno, 32
- fisicalismo, 27
fisicalismo não-redutivo, 27
fisicalismo redutivo, 27
framework, 48, 84
função de avaliação, 97
função de planejamento, 97
função de volição, 97
função motivação, 100
funcionalismo, 27
fundamentos da atenção, 29
fundamentos da consciência, 35
fundamentos filosóficos, 22
- GWT, 67, 104
- habilidades cognitivas, 48
hipótese do acesso global, 71
- IA forte, 28
IA fraca, 28
IDA, 67
idealismo, 26
imagem do corpo, 72
imaginação, 47
- inconsciência, 46
individualidade, 40, 94
indivisibilidade da consciência, 45
integração, 40
integração da informação, 71
intelecto, 24
intenção, 39
introdução, 18
ISAC, 79
- justificativa, 20
- LIDA, 69
LTM, 41
- mapa de saliências, 99
materiais e métodos, 107
materialismo, 26
materialismo eliminativo, 27
materialismo redutivo, 27
memória, 40, 96
memória autobiográfica, 42
memória de curto prazo, 41, 80, 82, 96
memória de longo prazo, 41, 80, 82, 94, 96
memória de produção, 42
memória de trabalho, 41, 68, 80, 82, 83, 94, 96
memória declarativa, 41
memória episódica, 42, 68, 79, 82, 84, 94, 96
memória episódica critérios, 42
memória não-declarativa, 41
memória perceptual, 43, 68
memória procedimental, 68, 82, 94, 97
memória prospectiva, 41
memória retrospectiva, 41
memória semântica, 42, 82, 94, 96
modelo de Colombini, 89
modelo de ruído USARSim, 109
modelo de Samsonovich, 83
modelo de Simões, 92
modelo de Starzyk, 75
modelo formal de Colombini, 62
modelo proposto, 91
modelos atencionais, 54
modelos baseados em atenção, 74
modelos cognitivos, 79
modelos para consciência, 66

- monismo, 26
monismo neutro, 26
motivação, 94, 96, 97, 100
noética, 39
noção de si mesmo, 79
o que é atenção, 29
o que é consciência?, 35
objetivo, 95
objetivos, 20
operação do modelo, 99
operação do modelo de Colombini, 63
organização do trabalho, 21
perfil cognitivo, 51, 105
perspectividade, 40, 103
planejamento, 97, 101
por que estudar a consciência artificial?, 20
possessividade, 40, 103
princípio da diferenciação, 40
princípio da integração, 40
problema difícil, 44, 89, 135
problema fácil, 44, 89, 135
processo, 98
processos, 95
processos cognitivos, 94
propriocepção, 94
PROSE, 108
qualia, 45, 72, 132, 135
recordação, 47, 82, 84
recordação episódica, 99, 100
recordação procedimental, 99, 100
recordação semântica, 99, 100
recordar, 30, 41, 79, 82
relação corpo-mente em culturas antigas, 22
representações de nível superior, 73
resultados, 122
revisitando o problema da consciência, 88
ruído, 109
SAIM, 55
saliência emocional, 81
seleção antecipada, 32
seleção para ação, 34, 61, 94
seleção para percepção, 34, 61, 94
seleção tardia, 32
senciência, 38, 103
sistema atencional, 94
sistema cognitivo, 94
SOAR, 82
solipsismo, 26
STM, 41
supermapas, 74
tarefas, 95
TASER, 81
taxonomia da atenção, 31
taxonomia das arquiteturas para consciência, 66
taxonomia sistemas conscientes baseados em atenção, 76
teoria da transmigração das almas, 25
teste de Turing, 47
tomada de decisão, 94, 98, 102
top-down, 32, 76, 112
tracker, 75
usarsim, 107
usarsim modelo de ruído, 109
vigilância, 33
volição, 39, 48, 68, 97, 101, 103