

Implementação de um Mecanismo de Consciência de Máquina na Arquitetura Cognitiva MECA

Ricardo R. Gudwin

DCA-FEEC-UNICAMP

gudwin@unicamp.br

<http://faculty.dca.fee.unicamp.br/gudwin>



■ O Exterminador do Futuro (1984)

- Cyberdyne (Sunnyvale, CA)
 - Desenvolveu o Skynet, uma rede de supercomputadores utilizando inteligência artificial para substituir seres humanos pilotando aeronaves comerciais e militares, e também para o controle de outros sistemas militares, incluindo mísseis nucleares
 - O sistema ficou "on-line" em 4 de agosto de 1997
 - Em 29 de agosto de 1997, o Skynet desenvolveu consciência, tornando-se consciente.
 - Com medo do que poderia ocorrer, tentou-se desligar o Skynet, mas devido ao caráter distribuído do mesmo, ele conseguiu manter-se ligado e resolveu retaliar, lançando um ataque nuclear contra a Rússia
 - Iniciou-se uma batalha entre humanos e máquinas que passaram a conseguir um número cada vez maior de capacidades

Estudo Científico da Consciência

- Consciência: O que é isso ?
 - Processo ? Entidade ? Estado ? Propriedade ?
 - Qual sua natureza ?
 - Como (se possível) tratar deste fenômeno em partes ?
 - Qual a(s) sua(s) função(ões) ?
 - Quais os níveis de descrição e análise adequados para compreender seu aparecimento ?
 - Quais seus correlatos biológicos ?
 - Como, se possível, simular seu funcionamento ?
 - Quais as teorias / métodos / modelos / protocolos de investigação ?



A Filosofia da Consciência

- **Caráter Subjetivo da Experiência (Nagel 1974)**
 - What is like to be a ...
- **O “Hard Problem” da Consciência (Chalmers 1995)**
 - Explicar como as sensações adquirem características, tais como cores ou sabores (Qualia)
- **Os “Easy Problems”**
 - Explicar as habilidades de discriminar, categorizar e reagir a estímulos ambientais; integrar informação; reportar estados mentais; focar a atenção; acessar seus próprios estados mentais; controle deliberativo do comportamento; diferença entre estar acordado (consciente) ou dormindo, etc.
 - São simples porque podem ser explicados por mecanismos que implementam sua funcionalidade

Um Termo – Diversos Fenômenos ?

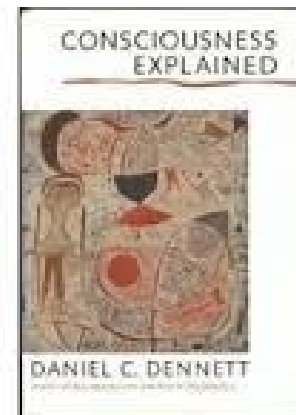
- Fenomenológico
 - O que é “experienciar” alguma coisa ?
 - Aspectos da experiência individual, subjetiva - Qualia
- Biopsicológico
 - Diferentes níveis de atenção (vigília, sono, coma, hiperatividade)
- Meta-cognitivo
 - Reflexão (ciência) sobre os próprios processos cognitivos
- Introspectivo
 - “Inner Voice” e “Inner Speech”
- Psicoanalítico
 - Noção de 1a pessoa, “Eu”, “Self”, “Ego”, “Identidade”
- Processual
 - Processamento Consciente (Serial) ×
Processamento Inconsciente (Paralelo)

Block e os 4 Tipos de Consciência

Consciência Fenomenal P-consciência	Natureza qualitativa da experiência; <i>Qualia</i> ;
Consciência de Acesso A-consciência	O que pode ser observado: Objetos e Cenas; Habilidade de reportar experiências e de agir sobre elas;
Consciência de Monitoramento M-consciência	Pensamentos sobre sensações e percepções (distintamente destas mesmas); Meta-cognição; <i>Inner-voice</i> ;
Auto-Consciência S-consciência	Pensamentos sobre o indivíduo como um ente distinto; Experiência de primeira pessoa: <i>Self</i> ;

O Paradigma Serial/Paralelo

- **Consciousness Explained (1991) – Daniel Dennett**
 - Pg. 210: “A Consciência humana em si ... pode ser melhor entendida como a operação de uma máquina virtual do tipo Von Neumann (serial) implementada em uma arquitetura paralela do cérebro que não foi desenhada para nenhuma dessas atividades”
 - Máquina serial virtual sobre um hardware paralelo
- **Função**
 - Tornar serial (sequencial) um processamento que é eminentemente paralelo
- **Pergunta**
 - Como criar essa máquina virtual ?
- **Resposta**
 - O Paradigma Serial/Paralelo em Consciência Artificial



Teoria do Workspace Global (Bernard Baars)

- Mecanismo da Consciência
 - Metáfora do teatro interativo



- Coordenação de um fluxo serial de processamento em um sistema paralelo e assíncrono de processamento de informações
- Propagação da informação em broadcast para módulos funcionais separados distribuídos no cérebro

Arquitetura Serial-Paralela

■ Informação em Broadcast

- Descoberta de estados globais potencialmente interessantes
- Permite a formação de novas conexões

■ Seleção de Processadores/Coalizões que irão fazer o broadcast a cada instante de tempo

- Baseada no nível de ativação dos processadores/coalizões
 - Relevância ou importância da informação sendo processada
- Emergência de uma sequência serializada – Consciência

■ Consciência

- Permite a divulgação de informações importantes simultaneamente a todo o sistema
- Fomenta o surgimento de novas coalizões

Arquitetura Serial-Paralela

■ Processamento Inconsciente

- Acesso somente a estados locais
- Comportamentos automáticos do sistema
- Característica eminentemente paralela
 - Percepção, visão, audição, etc.

■ Processamento Consciente

- Permite o aprendizado de novos comportamentos automáticos
- Característica eminentemente serial
 - Planejamento, execução de comportamentos, raciocínio, etc.

■ Comportamento Global do Sistema

- Incorpora simultaneamente características paralelas e seriais
- Aprendizagem de novos comportamentos por meio da experiência

Teoria dos Processos Duais

Dual Process Theory

BEHAVIORAL AND BRAIN SCIENCES (2000) 23, 645–726

Printed in the United States of America

Individual differences in reasoning:
Implications for the rationality
debate?

Keith E. Stanovich

*Department of Human Development and Applied Psychology, University of
Toronto, Toronto, Ontario, Canada M5S 1V6*

kstanovich@oise.utoronto.ca

Richard F. West

School of Psychology, James Madison University, Harrisonburg, VA 22807

westrf@jmu.edu falcon.jmu.edu/~westrf

Teoria dos Processos Duais

Dual Process Theory

Type 1 process (intuitive)	Type 2 process (reflective)
Defining features	
<i>Does not require working memory</i>	<i>Requires working memory</i>
<i>Autonomous</i>	<i>Cognitive decoupling; mental simulation</i>
Typical correlates	
Fast	Slow
High capacity	Capacity limited
Parallel	Serial
Nonconscious	Conscious
Biased responses	Normative responses
Contextualized	Abstract
Automatic	Controlled
Associative	Rule-based
Experience-based decision making	Consequential decision making
Independent of cognitive ability	Correlated with cognitive ability
System 1 (old mind)	System 2 (new mind)
Evolved early	Evolved late
Similar to animal cognition	Distinctively human
Implicit knowledge	Explicit knowledge
Basic emotions	Complex emotions

Arquiteturas Cognitivas

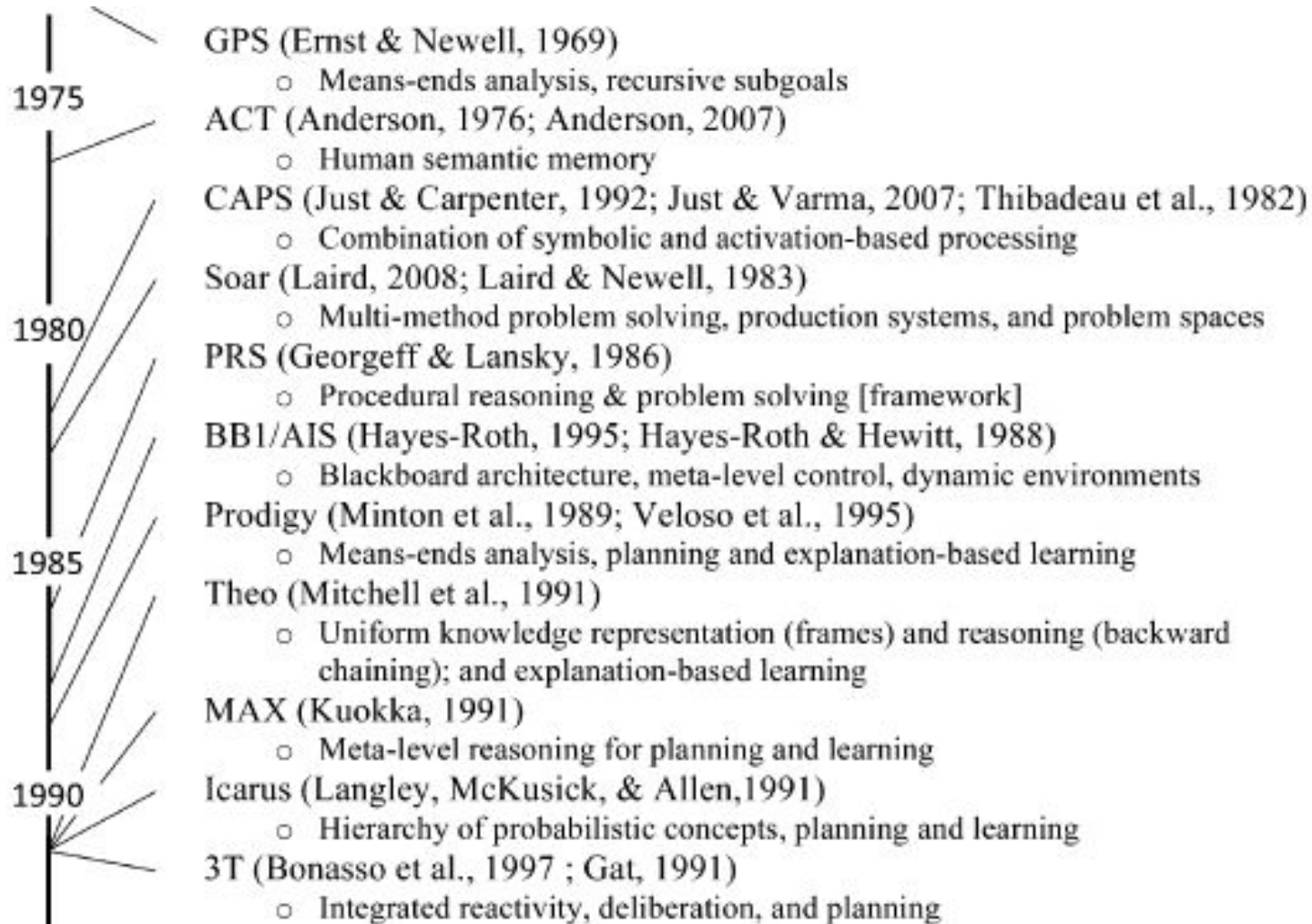
■ Arquiteturas Cognitivas

- Arquiteturas computacionais de propósito geral provendo a infraestrutura básica para a construção de agentes artificiais
- Baseadas em Modelos Cognitivos da Mente Humana
 - Implementando versões computacionais de habilidades cognitivas, como por exemplo: percepção, emoções, memória, tomada de decisões, comportamentos reativos e/ou deliberativos, motivações, aprendizagem, linguagem, cognição social e consciência

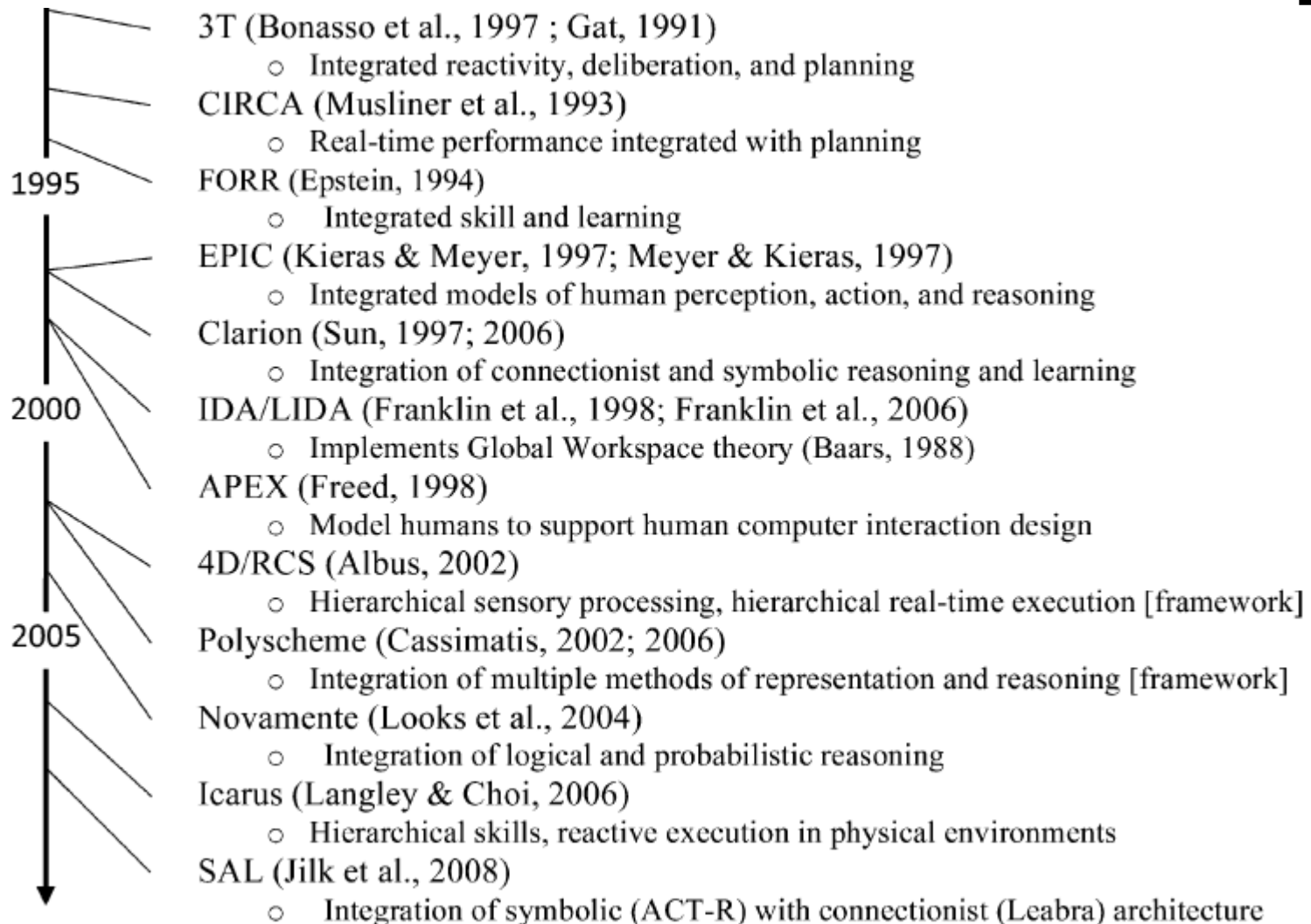
■ Exemplos de Arquiteturas Cognitivas (BICA Society)

- 4CAPS, 4D/RCS, **ACT-R**, AIS, Apex, ART, Atlantis, BECCA, biSoar, CERA-CRANIUM, Chrest, **Clarion**, CogNet, CogPrime, CoJACK, Copycat, Disciple, DUAL, Emotion Machine, Epic, ERE, FORR, Gat, GLAIR, GMU-BICA, Guardian, H-Cogaff, Homer, HTM, **Icarus**, Imprint, Leabra, MAX, **MECA**, **LIDA**, **NARS**, Nexting, Omar, Pogamut, Polyscheme, PRODIGY, PRS, Psi-Theory, R-CAST, RALPH-MEA, Recommendation Architecture, REM, SAL, **Soar**, Society of Mind, **Subsumption architecture**, Teton, Theo, Tosca, Ymir

História das Arquiteturas Cognitivas



História das Arquiteturas Cognitivas



CST: Cognitive Systems Toolkit

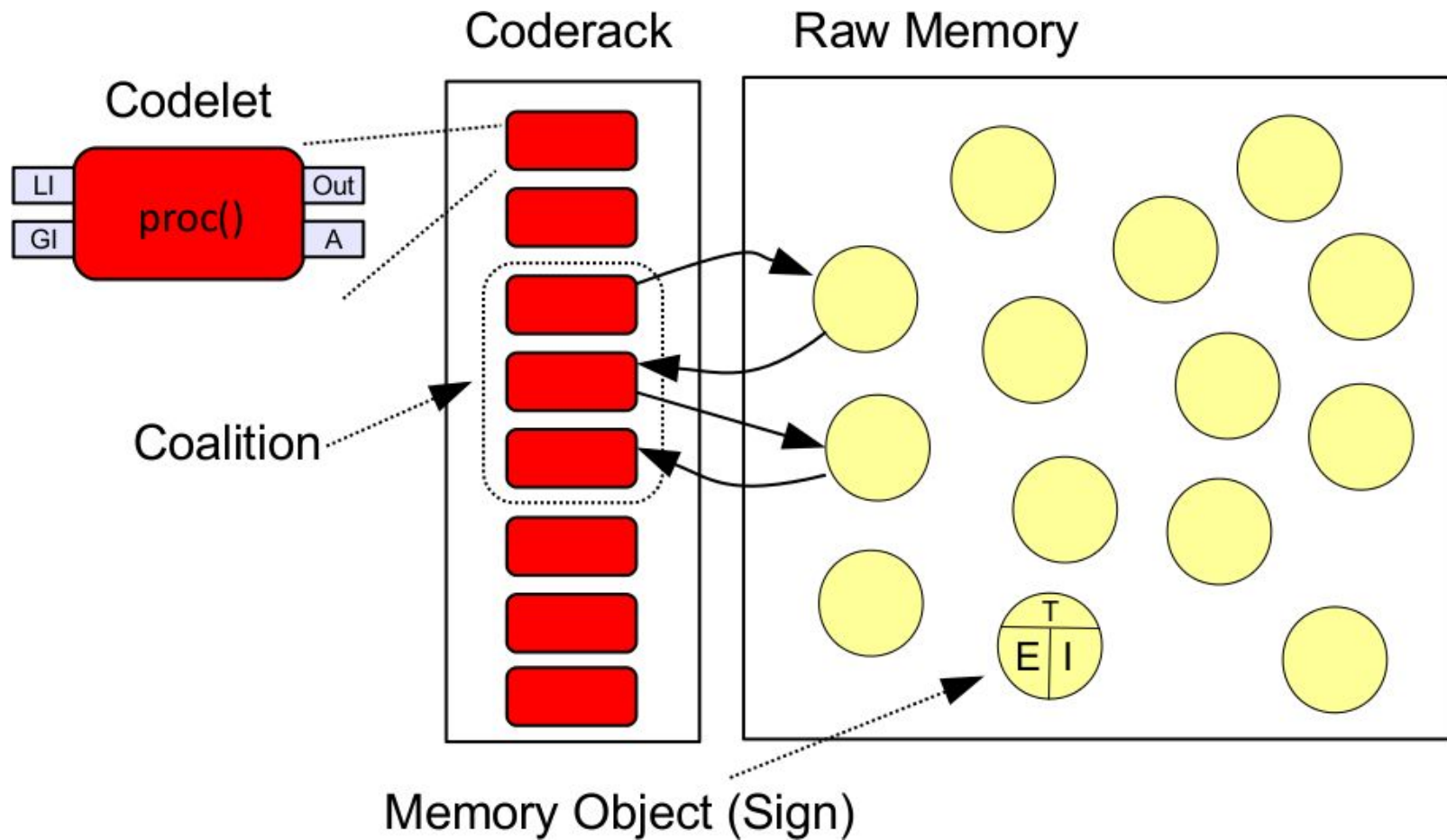
■ Arquiteturas Cognitivas

- De propósito geral ou especializadas
- Arquiteturas de Propósito Geral: Monolíticas
 - Comprometidas com teorias específicas, cobrindo as vezes somente parcialmente diferentes aspectos da mente
 - Não são flexíveis o suficiente para acomodar novos modelos sobre fenômenos mentais
- Arquiteturas Híbridas
 - Tendência recente de tentar integrar duas ou mais arquiteturas
 - e.g. SAL – Synthesis of ACT-R and Leabra

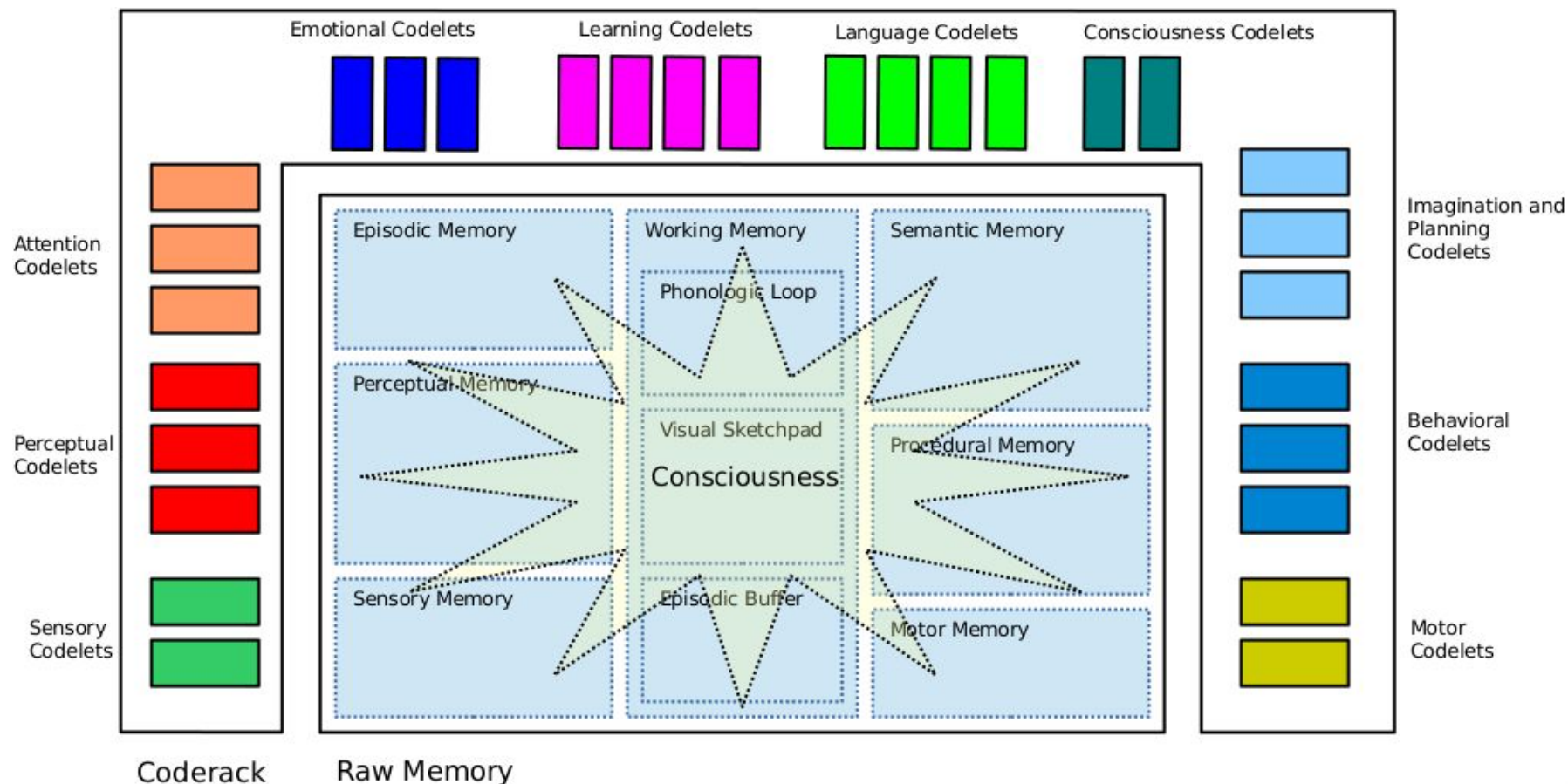
■ CST – Ferramenta para a construção de sistemas cognitivos

- Flexibilidade no aproveitamento de novos modelos cognitivos

CST: Core

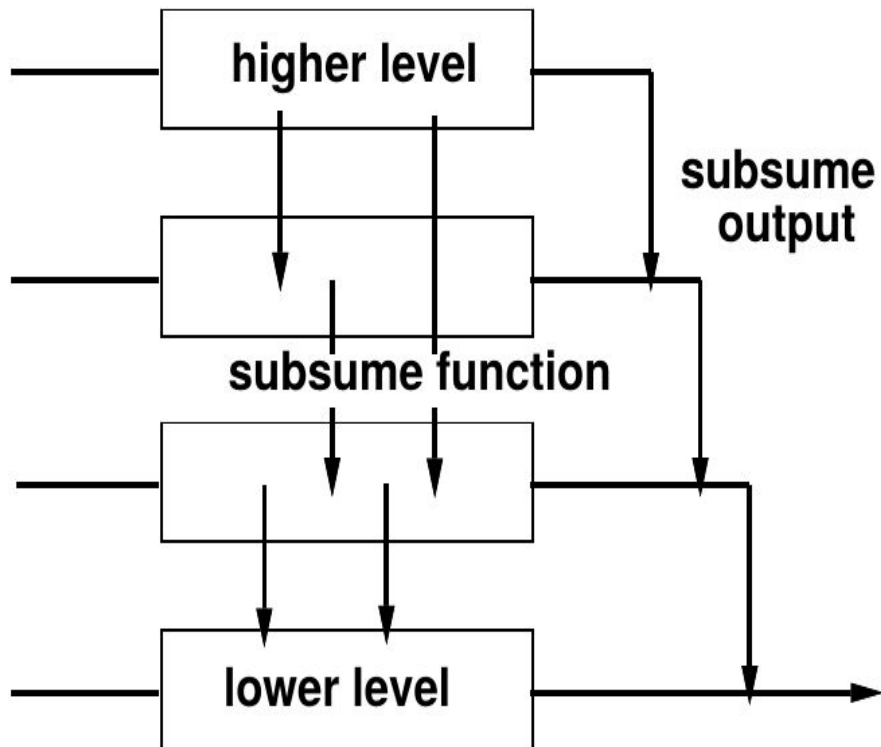


Arquitetura Cognitiva de Referência

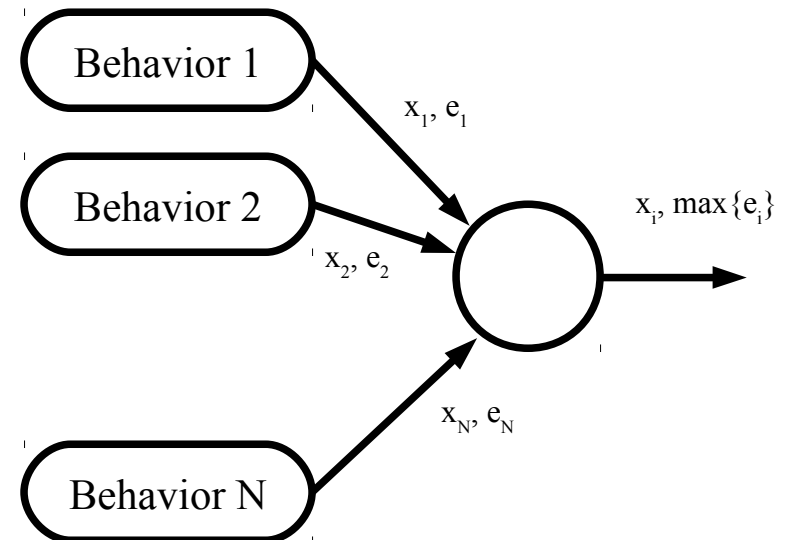


Subsumção Dinâmica

Subsumção Clássica



Subsumção Dinâmica



■ SOAR (State Operator And Result)

- Sistema baseado em Regras
 - Exploração de um Espaço de Estados
 - Operador: definido na forma de regras
 - Transforma um estado em um próximo estado
 - Goal: Estado final a ser atingido durante a busca
 - Interrompe a exploração do espaço de estados com uma solução
- Diferentes tipos de regras
 - Regras para a proposição de operadores
 - Regras para a seleção dentre possíveis operadores
 - Regras para a aplicação de operadores
- Estado
 - Definido por meio de WMEs (Working Memory Elements)
 - Tripla (ID, ^Attribute, VALUE) constituindo uma árvore = estado

Motivações para a MECA

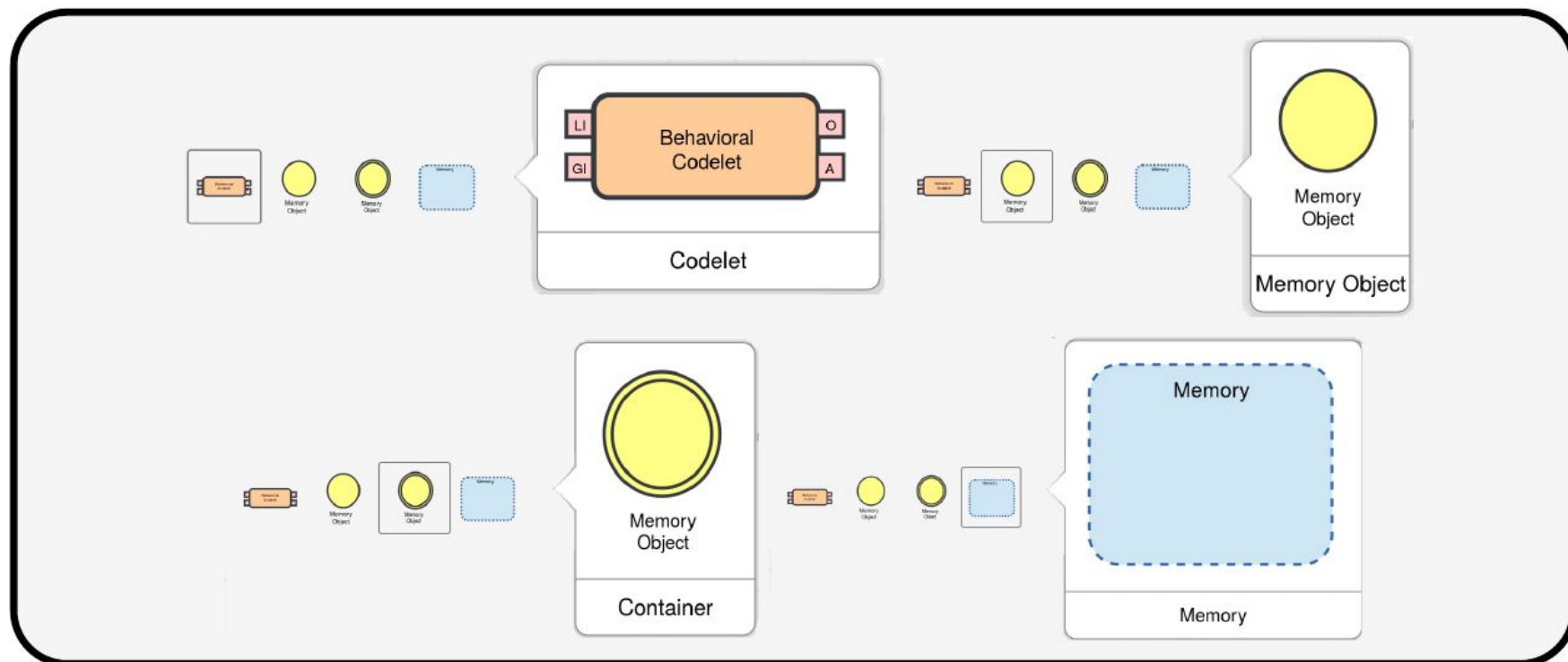
■ Motivação

- Arquitetura Cognitiva implementando a Teoria de Processos Duais: Integrando processamento Inconsciente com processamento Consciente, utilizando o CST

■ MECA em poucas palavras:

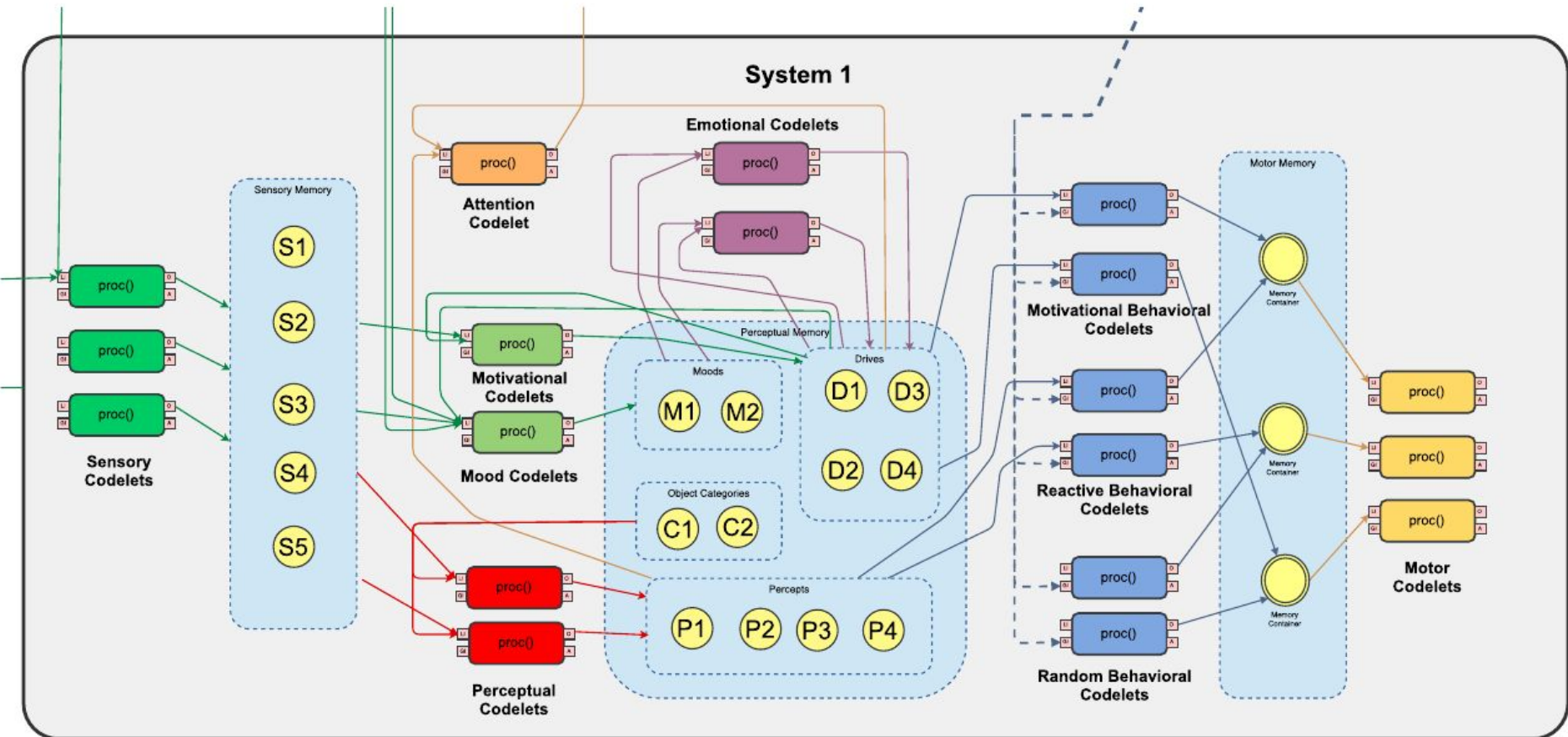
- Arquitetura Cognitiva Híbrida
- Instância da Teoria dos Processos Duais
 - System 1: Arquitetura de Subsumção Dinâmica
 - Acrescida de subsistemas Perceptual, Emocional e Motivacional
 - System 2: Baseada no SOAR para realizar planejamentos
 - Acrescida de:
 - Subsistema Motivacional (Appraisal + Geração de Goals)
 - Mecanismo de Consciência (Teoria do Workspace Global)
 - Subsistema de Memória Episódica – baseado em cenas
 - Mecanismo baseado em Expectativas

MECA: Elementos da Arquitetura

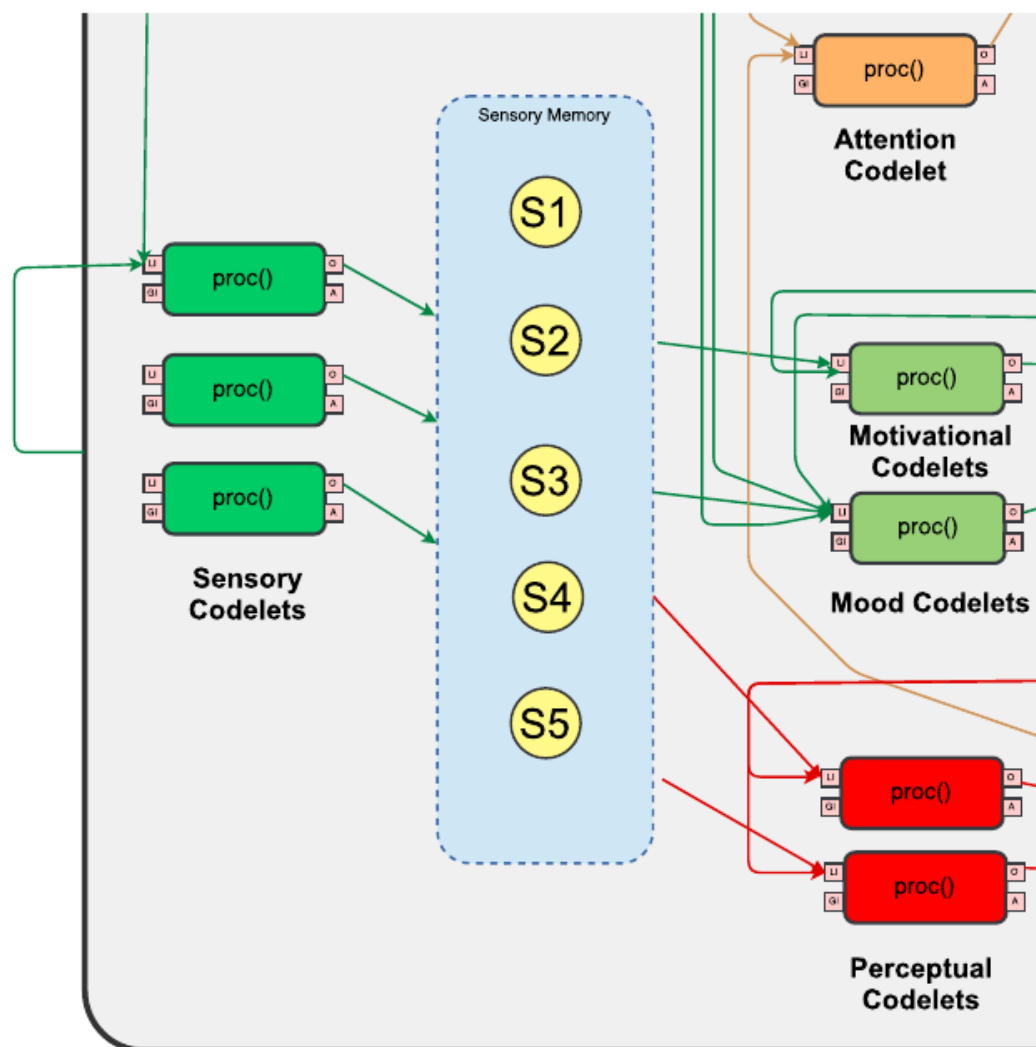




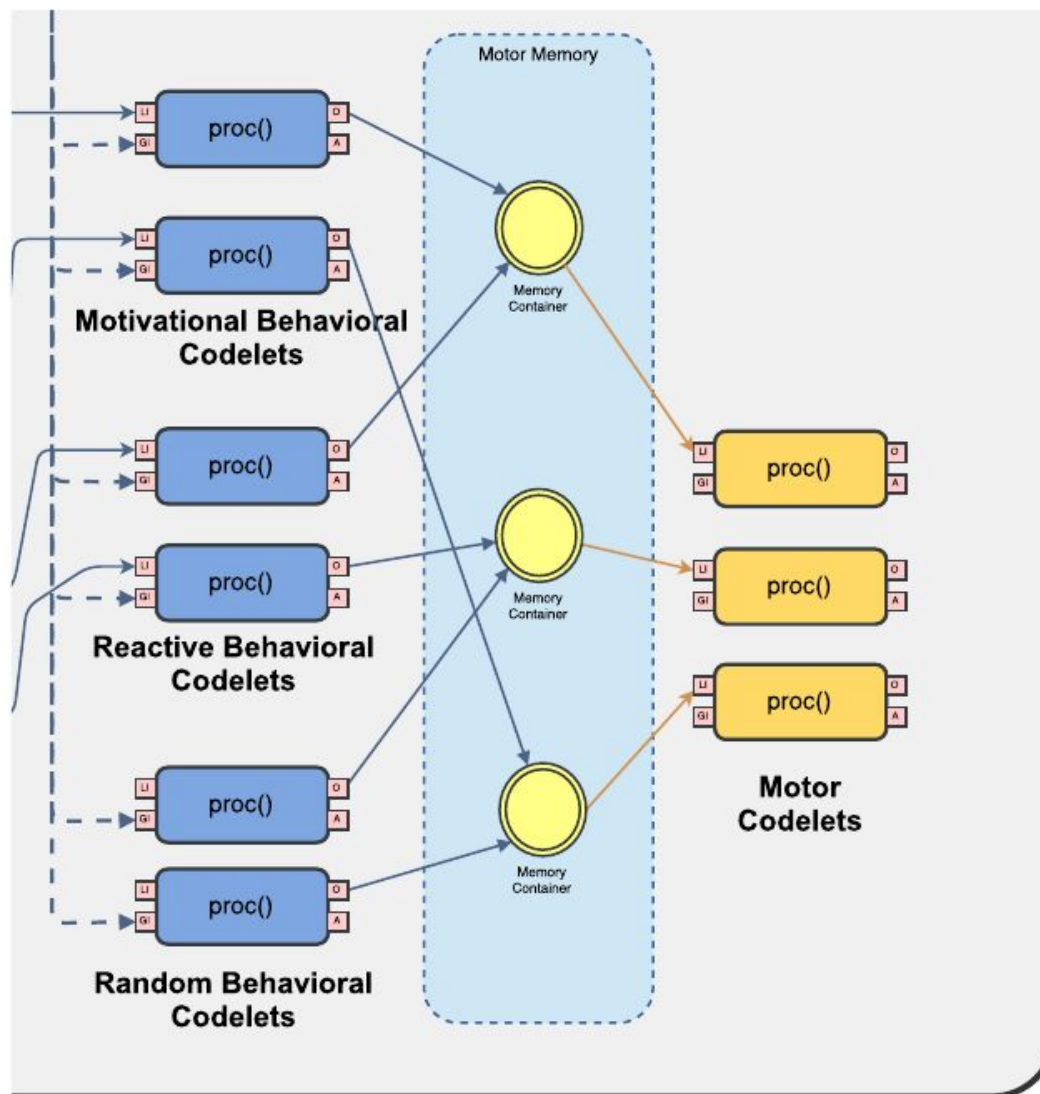
System 1



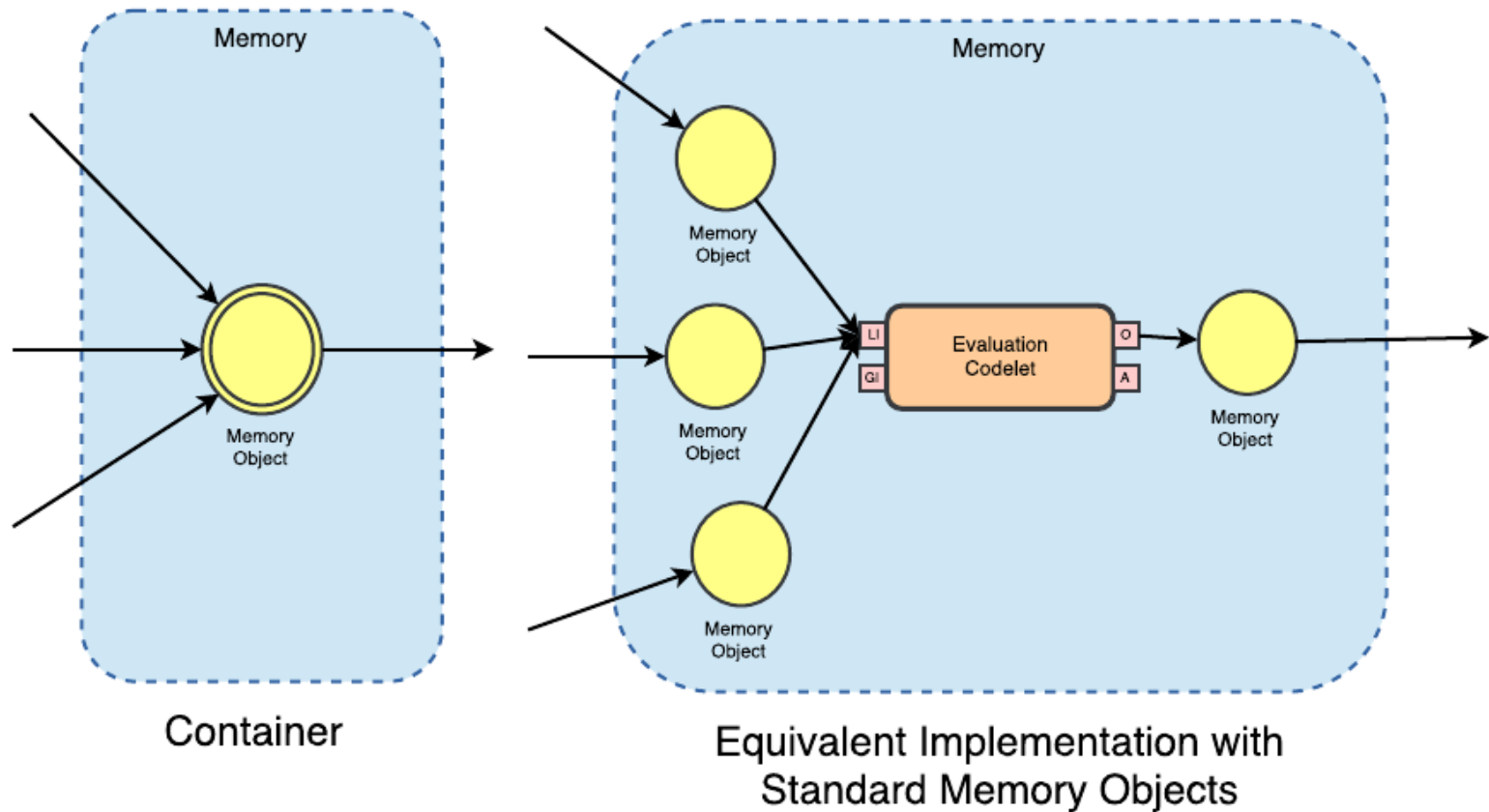
Subsistema Sensorial



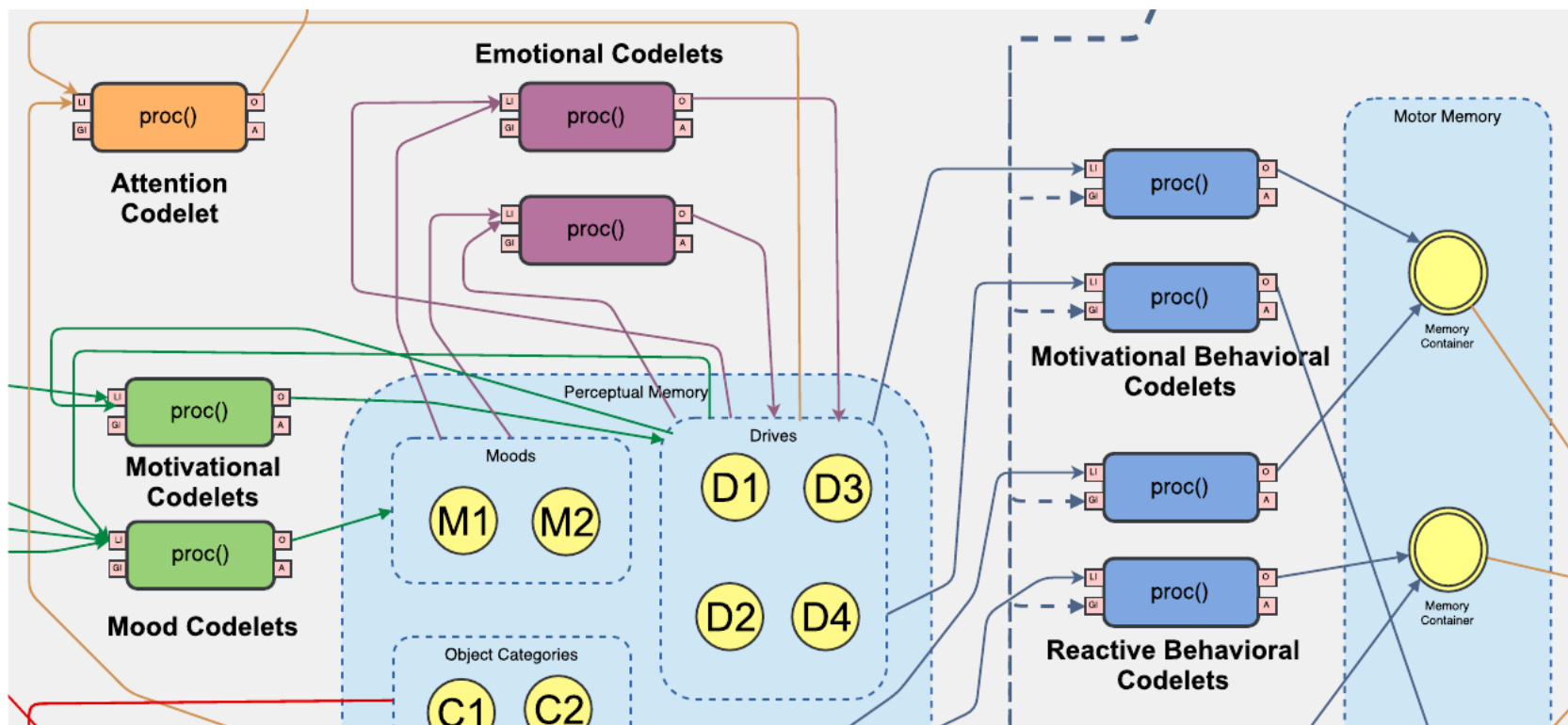
Subsistema Comportamental/Motor



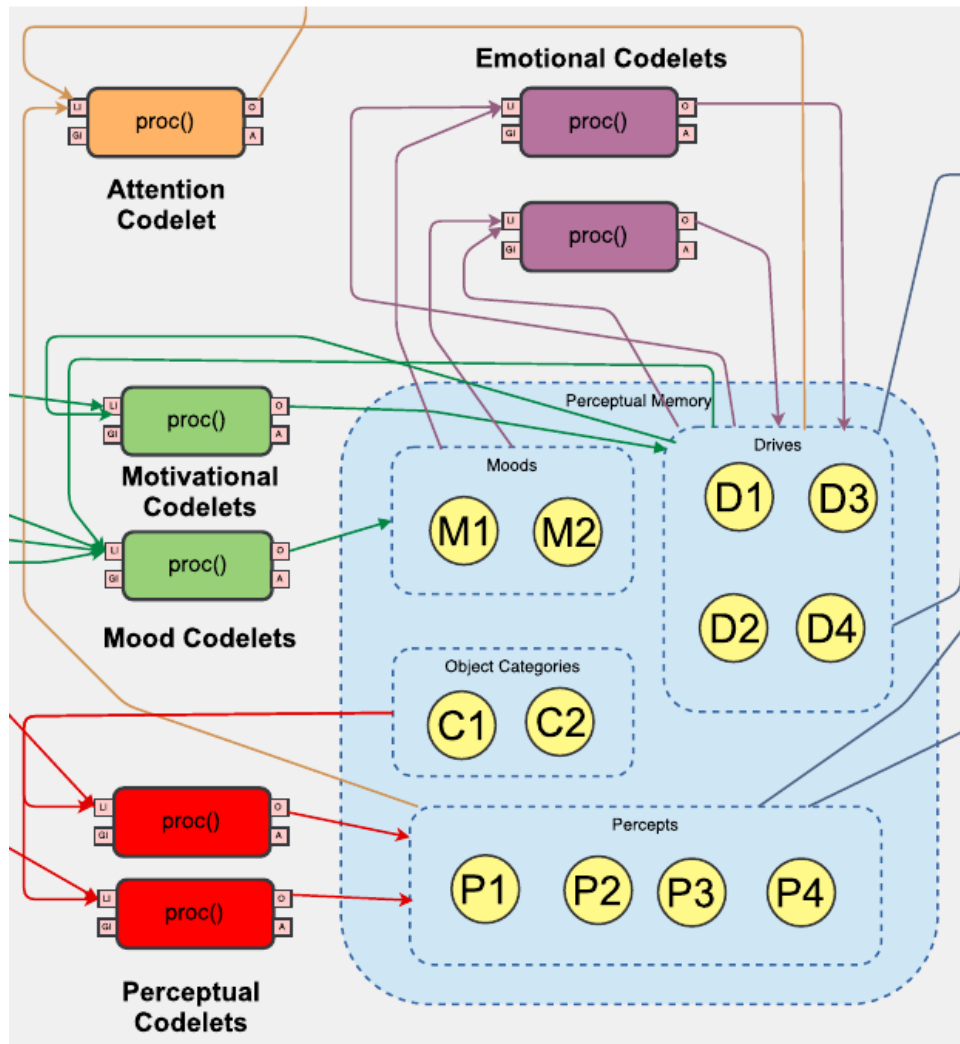
Containers de Memória



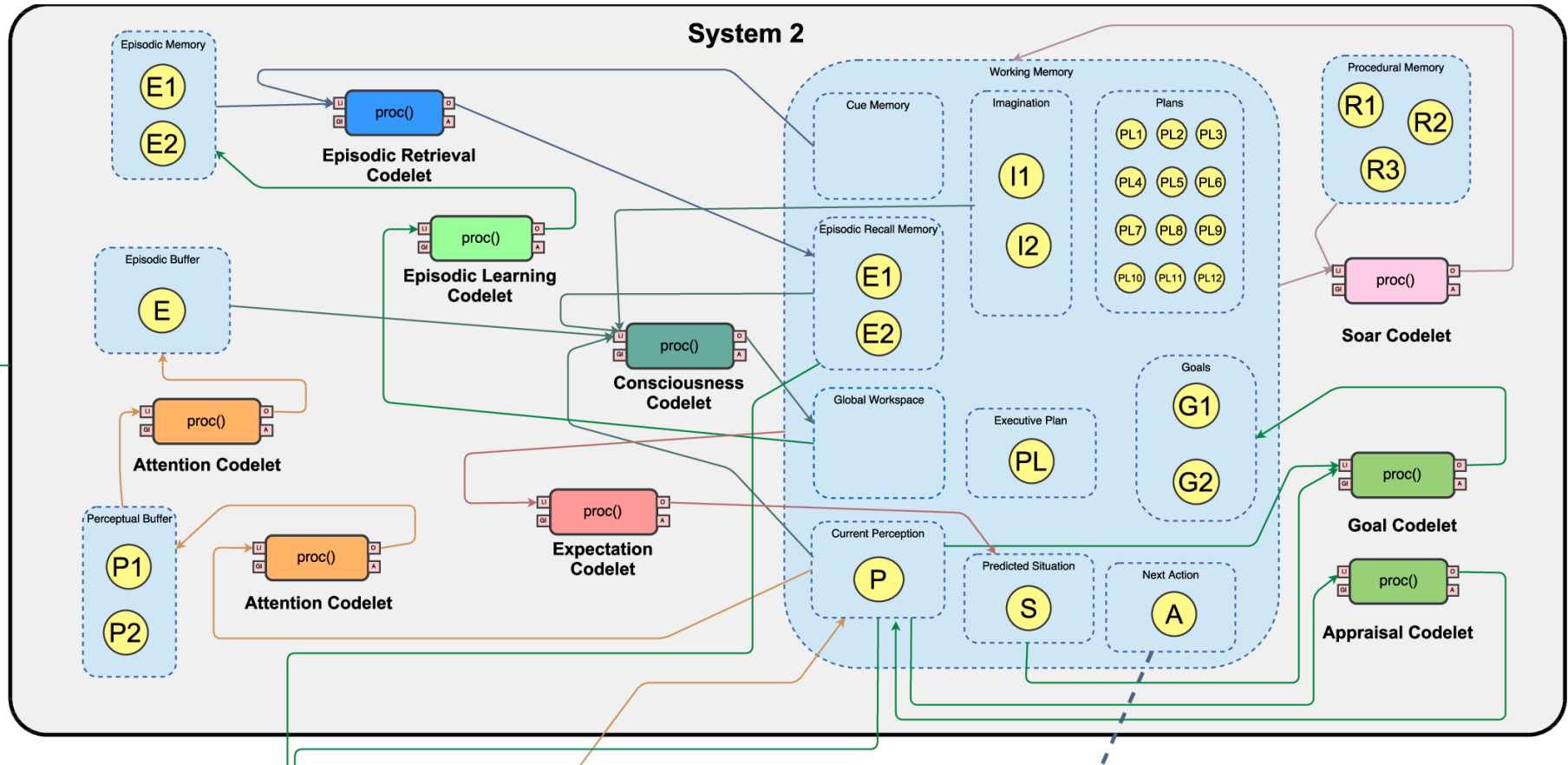
Subsistema Motivacional em System 1



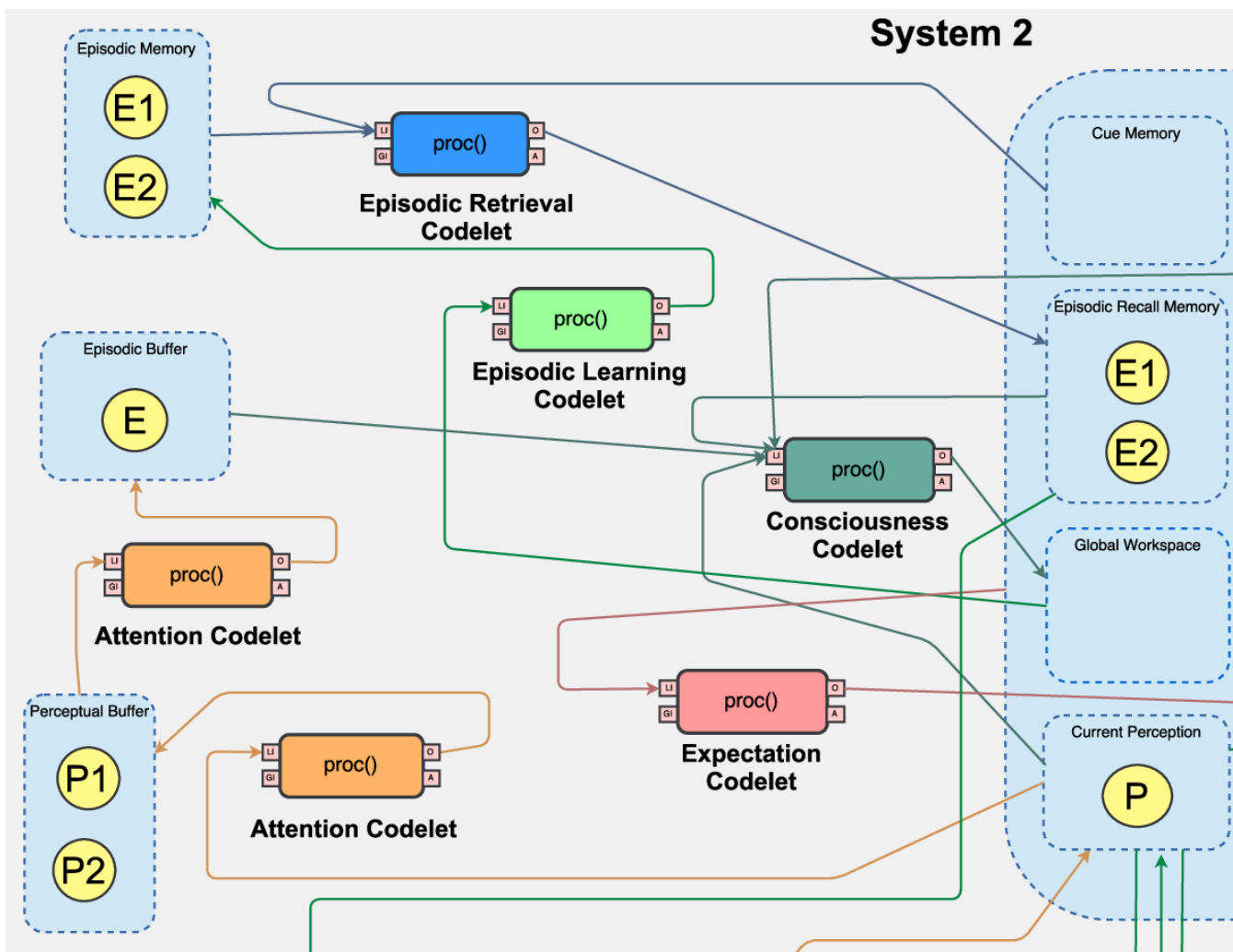
Subsistema Perceptual



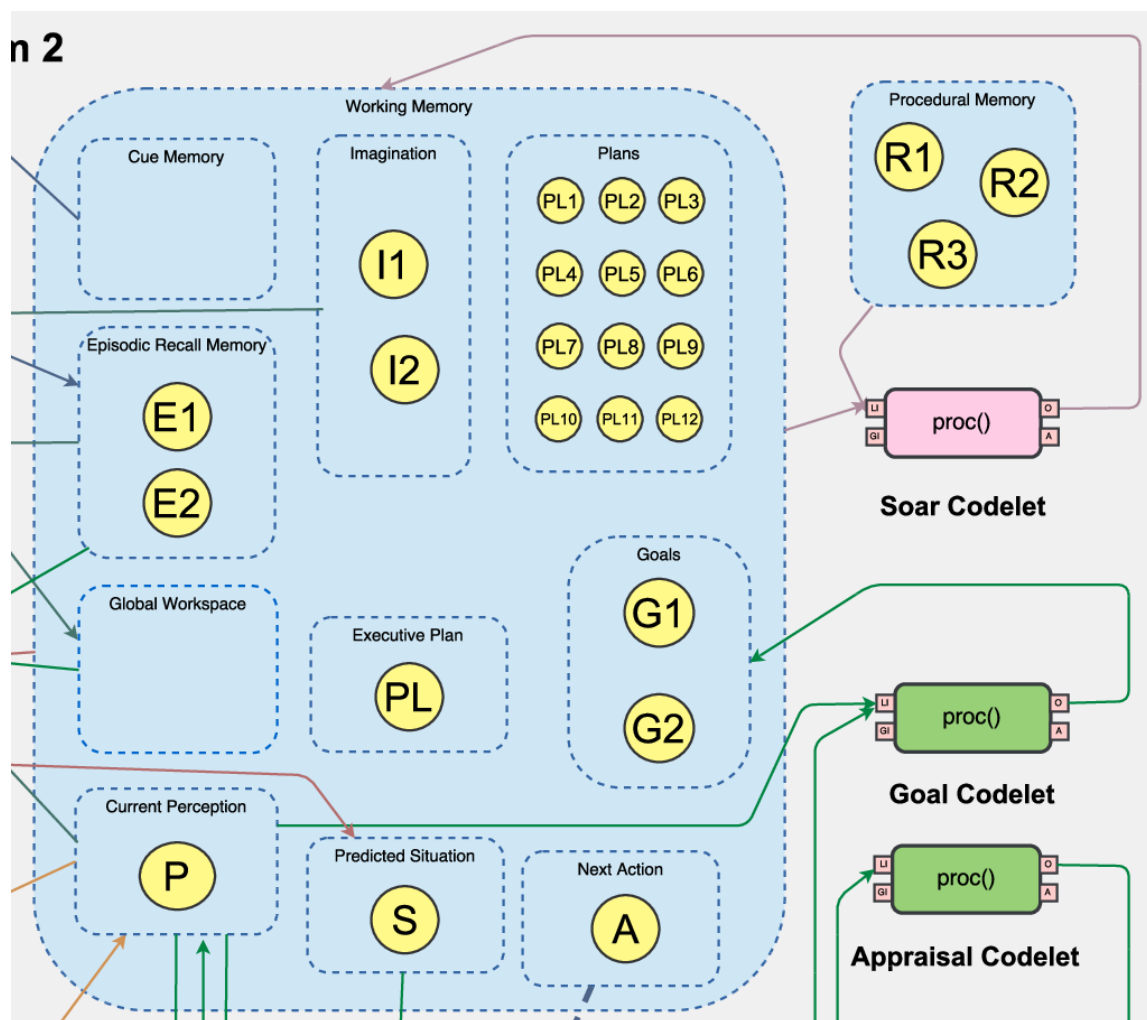
System 2



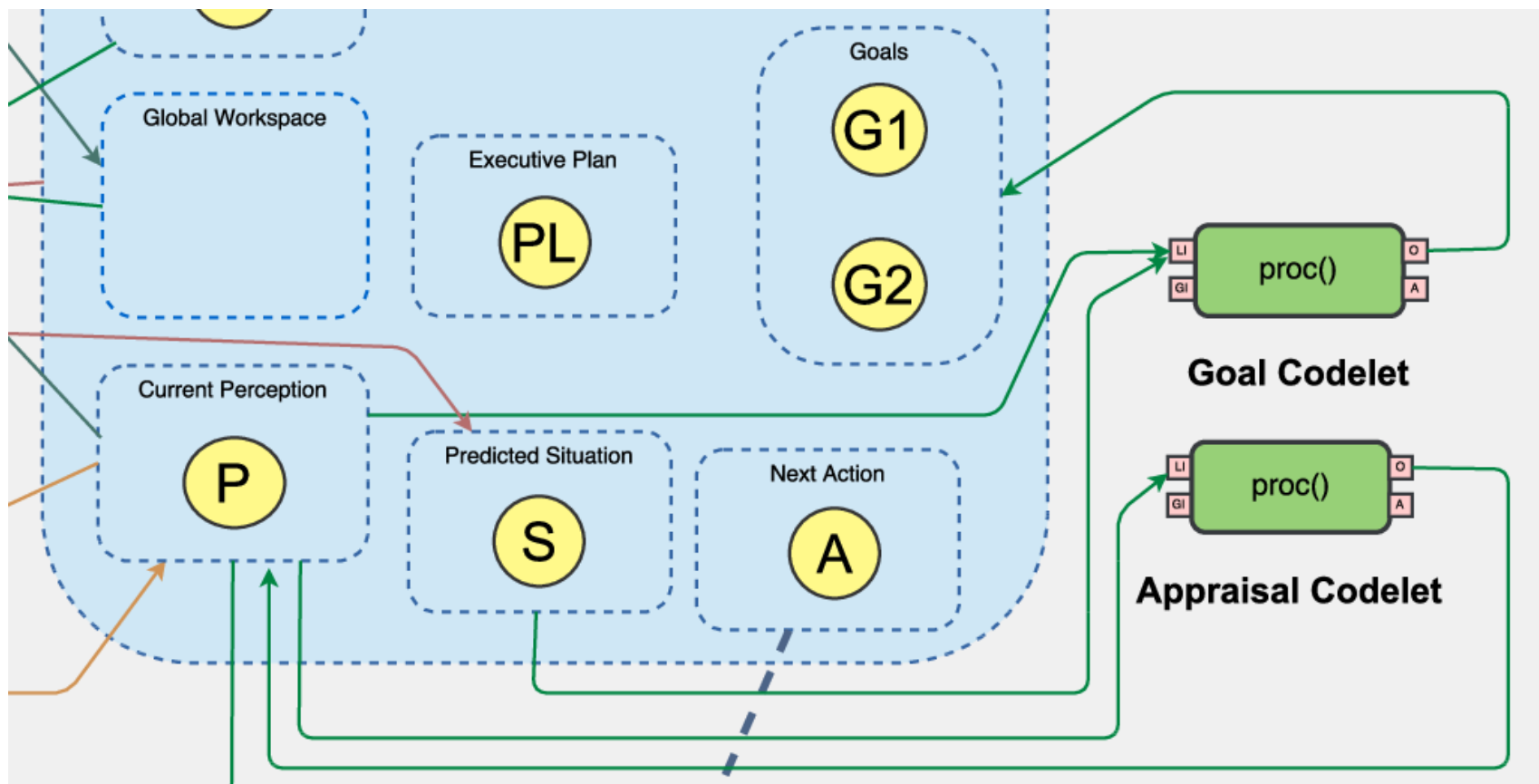
Subsistema de Memória Episódica



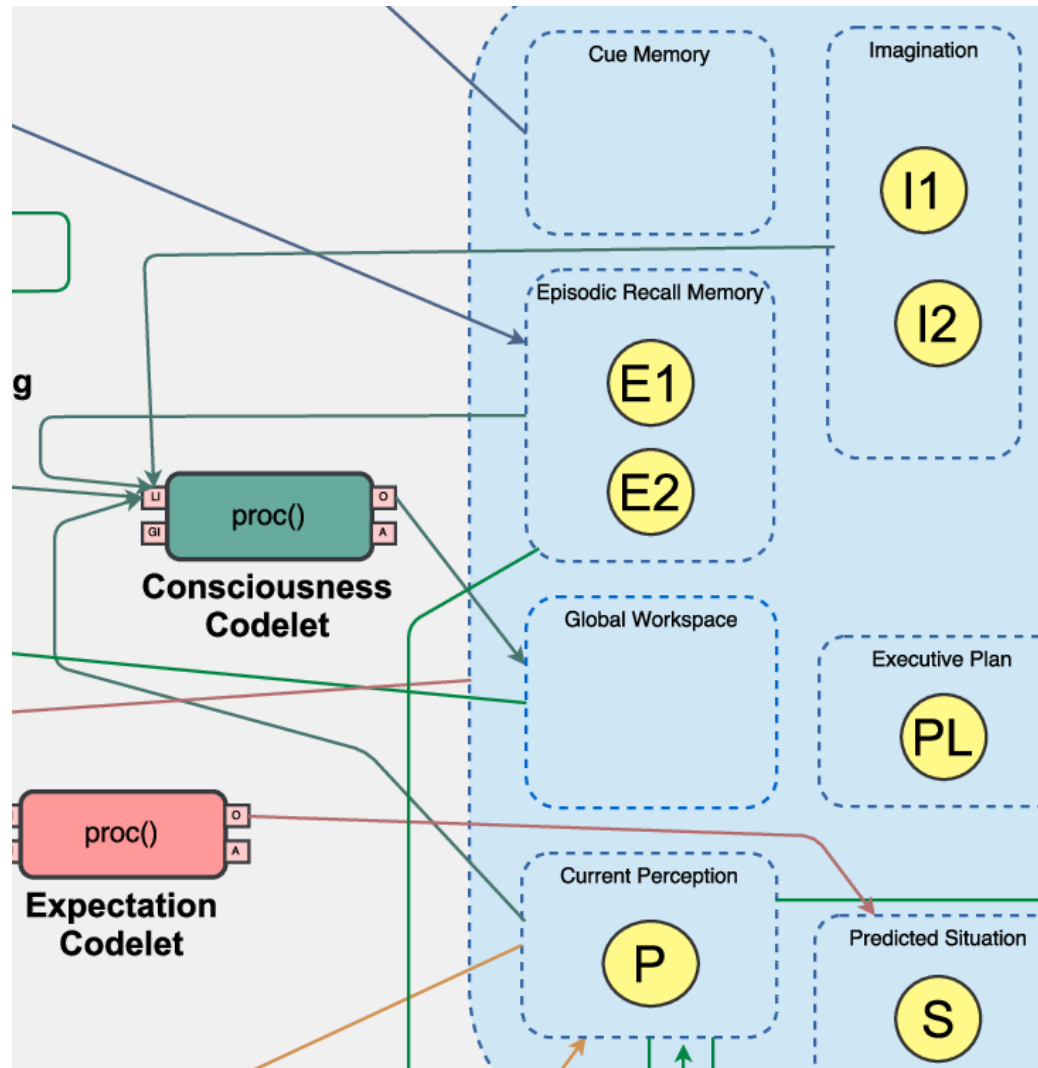
Subsistema de Planejamento



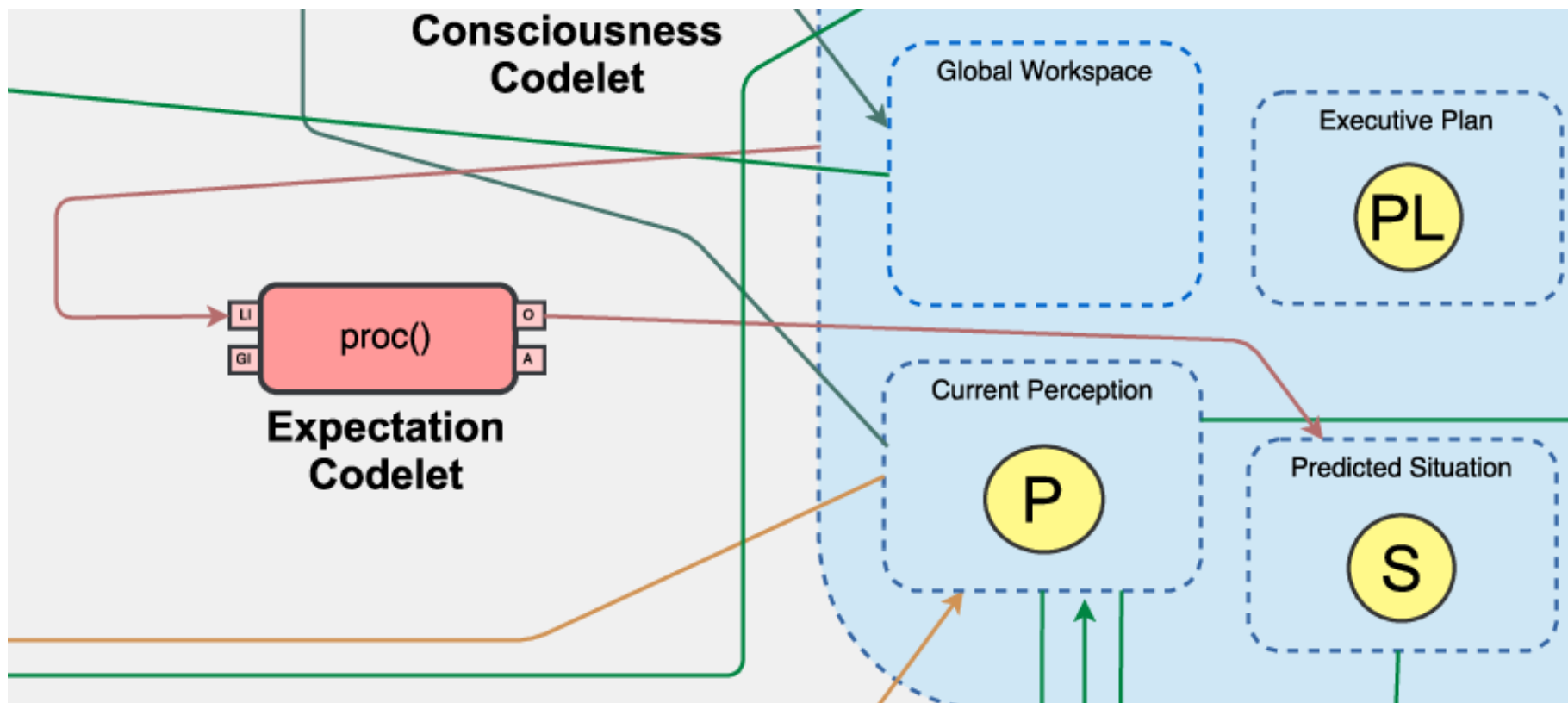
Subsistema Motivacional em System 2



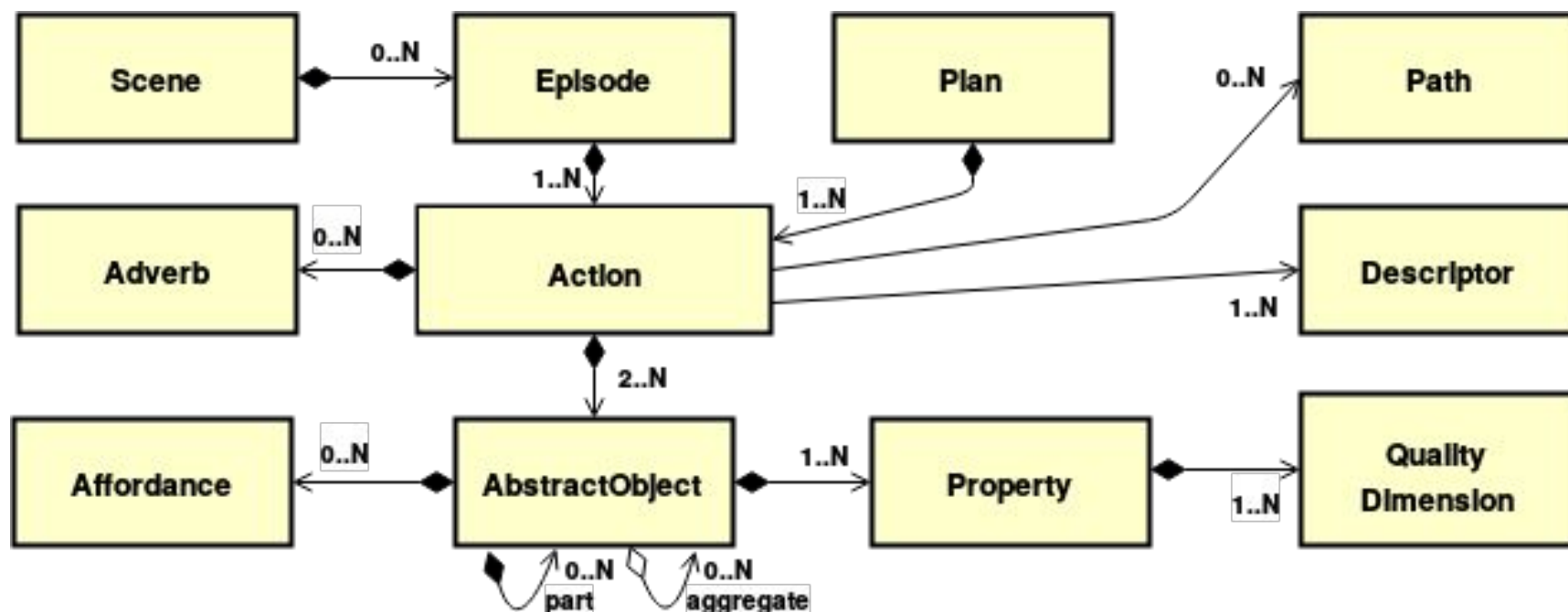
Subsistema de Consciência



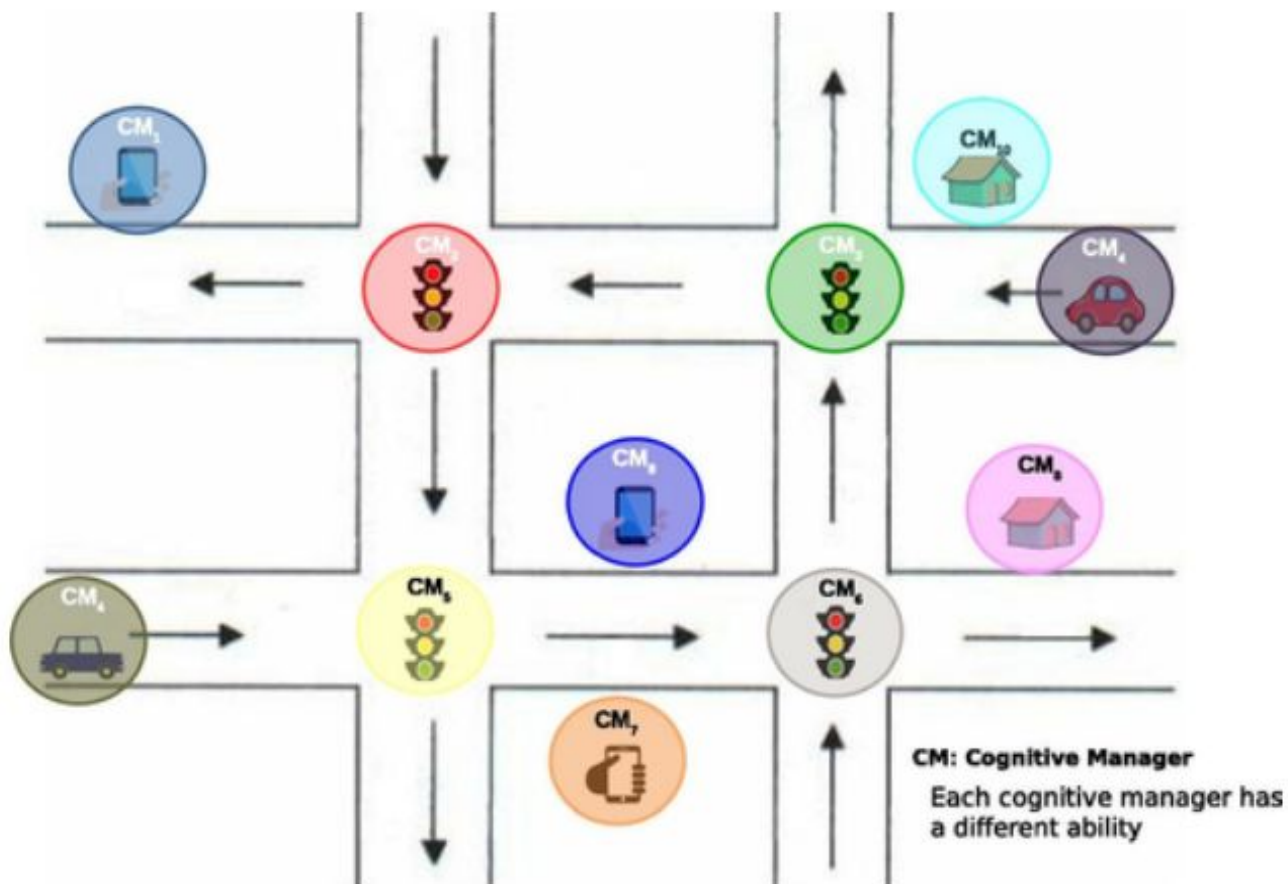
Subsistema de Expectativas



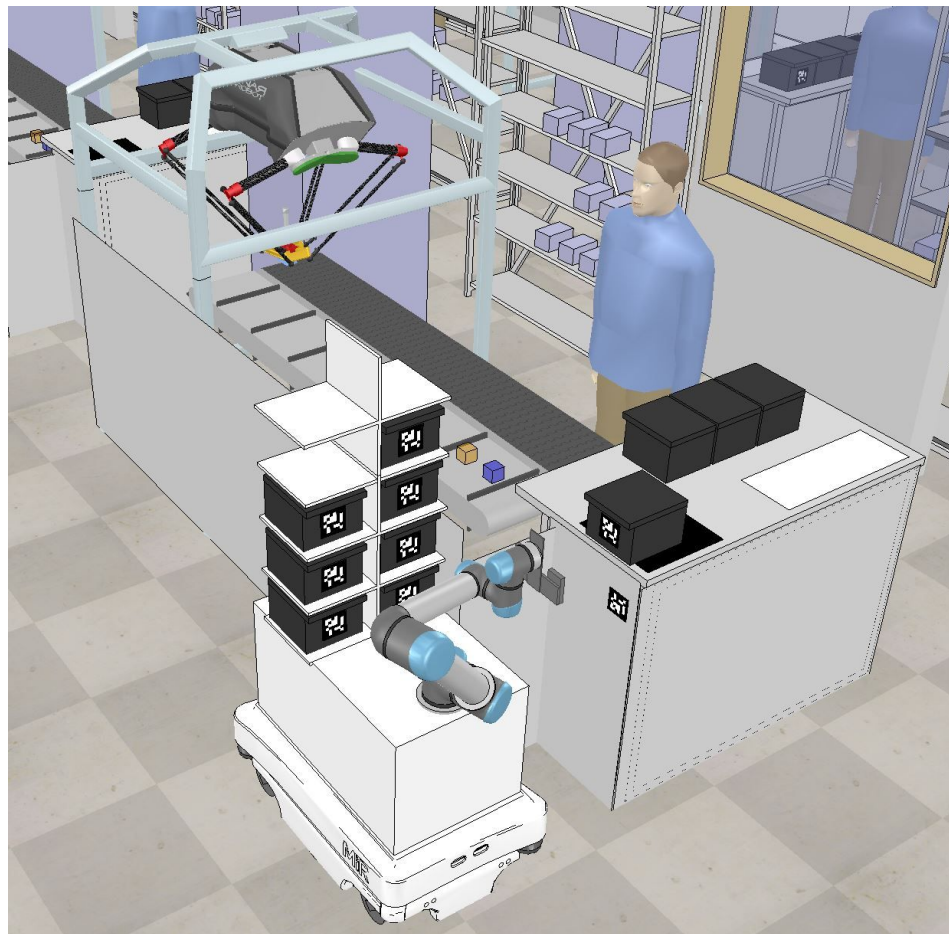
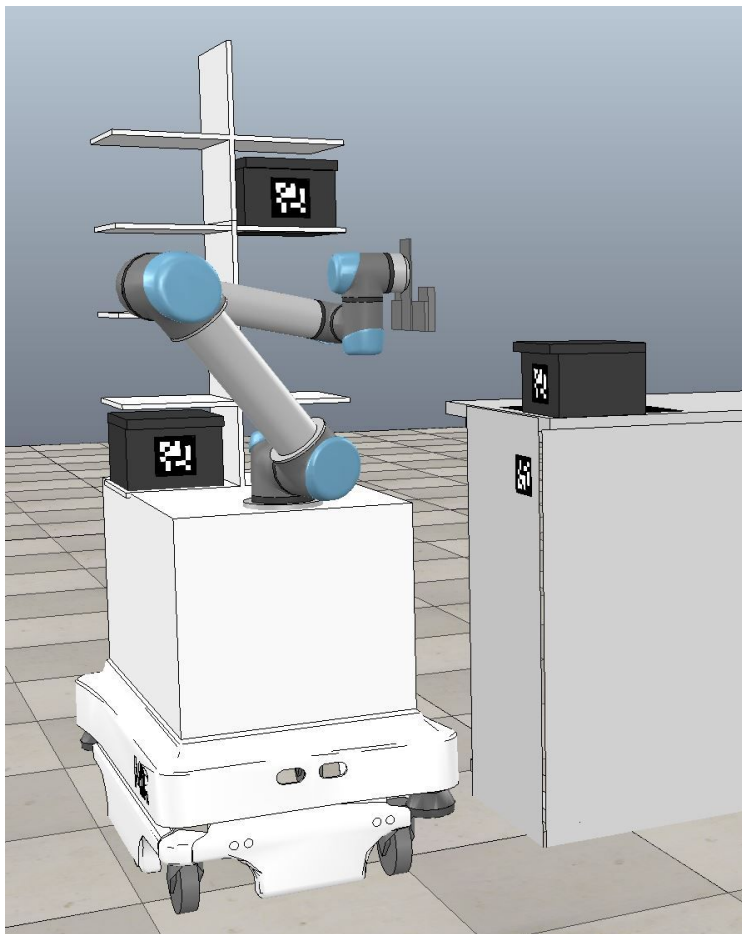
Elementos de Representação



Exemplo: Gerentes Cognitivos em Cidades Inteligentes



Exemplo: Robôs Transportadores Autônomos



■ MECA

- Arquitetura Cognitiva implementando a Teoria dos Processos Duais
- System 1: Responsável pelo processamento Inconsciente
- System 2: Responsável pelo processamento Consciente

■ Mecanismo de Consciência em MECA

- Instância da Teoria do Workspace Global (Baars)
- Workspace Global
 - Foca a atenção em conhecimento oriundo de:
 - Percepção (Presente)
 - Memória Episódica (Passado)
 - Imaginação (Futuro)
 - Promove o broadcast global do conhecimento focado
- Interação Inconsciente (System 1) x Consciente (System 2)