

# 智能科学

---

— Chapter 4 —

## 心智模型

史忠植

中国科学院计算技术研究所

[shizz@ics.ict.ac.cn](mailto:shizz@ics.ict.ac.cn)

<http://www.intsci.ac.cn/>

# 内容提要

---

- 一、概述
- 二、物理符号假设
- 三、诺尔曼模型
- 四、认知记忆模型
- 五、ACT模型
- 六、SOAR系统
- 七、心智社会
- 八、构建复杂心智联机系统
- 九、意识

# 概述

---

心智 (mind) 是脑的精神活动的总称，包括感觉、知觉、学习、记忆、注意、情感、意志、兴趣、性格、思维等大脑活动的一切方面。神经系统看成是脑的硬件，心智可以看成脑的软件。

# 概 述

---

心智以人类全部精神活动（包括情感、意志和感觉、知觉、表象、思维、直觉等）为研究对象，用现代科学方法来研究人类非理性心理与理性认知融合运作的形式、过程及规律。心智是一个庞大的学科群体，它涉及生物学、心理学、细胞学、脑科学、遗传学、语言学、逻辑学、认知科学、人工智能等多个领域。其目标在于探讨人类生命和精神方面的问题，例如，信息如何经由感官输入，如何处理、储存与运用，及与遗传、进化的关系等等。

# 概述

---

人类学家认为，一切生物都有某种心智。心智的发展经历了简单反射、条件反射、工具使用、语言符号四个阶段。使用语言符号与外界进行交流，只有人类才能做到，因此，人类心智的产生与发展是和人类所特有的符号语言分不开的。人类的符号语言以文字为载体，文字虽然产生很晚，但它对心智的发展和文明进步所起的作用是无以伦比的。250万年前人类意识出现，30万—10万年前人类开始使用符号语言，中国最早的文字甲骨文有3000多年的历史。在这个漫长的进化过程中，人类使用符号语言和符号文字不过是很短的一段时间，但人类心智的发展、人类文化和文明的进步却取得如此辉煌的成就，足见语言与心智研究具有何等密切的关系。

# 概 述

心智问题是一个非常复杂的非线性问题，我们必须借助现代科学的方法来研究心智世界。心智科学研究的是心理或心智过程，但它不是传统的心理科学，它必须寻找神经生物学和脑科学的证据，以便为心智问题提供确定性基础。心智世界与现代逻辑学和数学所描述的可能世界也有明显的区别：逻辑学和数学所描述的可能世界是一个无矛盾的世界，而心智世界则处处充满了矛盾；逻辑和数学对可能世界的认识和把握只能用演绎推理和分析方法，而人的心智对世界的把握则有演绎、归纳、类比、分析、综合、抽象、概括、联想和直觉等等多种手段。所以心智世界比数学和逻辑学所描述的可能世界要复杂广大得多。那么，我们应该如何从有穷的、无矛盾的、使用演绎法的、相对简单的可能世界进入无穷的、有矛盾的、使用多种逻辑和认知方法的、更为复杂的心智世界呢？这是心智研究要探索的基本问题之一。

Model Name	Acronym/ Abbreviation	Pew and Mavor (1998)	Ritter, Shadbolt et al. (2001)	Present IDA Study
Atomic Components of Thought	ACT	Yes	No	Yes
Adaptive Resonance Theory	ART	No	No	Yes
Architecture for Procedure Execution	APEX	No	Yes	Yes
Artificial Neural Networks	ANNs	Yes	No	No
Business Redesign Agent-Based Holistic Modeling System	Brahms	No	No	Yes
Cognition and Affect Project	CogAff	No	Yes	Yes
COGnition As a NETwork Of Tasks	COGNET	Yes	No	Yes
Cognitive Complexity Theory	CCT	No	No	Yes
Cognitive Objects within a Graphical EnvironmenT	COGENT	No	Yes	Yes
Concurrent Activation-Based Production System	CAPS	No	No	Yes
Construction-Integration Theory	C-I Theory	No	No	Yes
Distributed Cognition	DCOG	No	No	Yes
Elementary Perceiver And Memorizer	EPAM	No	Yes	No
Executive Process/Interactive Control	EPIC	Yes	No	Yes
Human Operator Simulator	HOS	Yes	No	Yes
Belief-Desire-Intention architecture	BDI	No	Yes	No
Man-machine Integrated Design and Analysis System	MIDAS	Yes	No	Yes
Micro Systems Analysis of Integrated Network of Tasks	Micro SAINT	Yes	No	Yes
MIDAS Redesign		Yes	No	Yes
Operator Model ARchitecture	OMAR	Yes	No	Yes
PSI	PSI	No	Yes	Yes
Situation Awareness Model for Pilot-in-the-Loop Evaluation	SAMPLE	Yes	No	Yes
Sparse Distributed Memory	SDM	No	Yes	No
State, Operator, And Result	Soar	Yes	No	Yes

2007-7-15

史忠植 心智模型

# 心智的计算理论

---

- Pinker (as well as others such as Lindsay and Norman, 1977; Minsky 1986) argues that “the mind is a system of organs of computation.”<sup>1</sup>
- “The mind is what the brain does; specifically, the brain processes information, and thinking is a kind of computing.”<sup>1</sup>

1) Pinker, Steven, 1997, *How the Mind Works*, W.W. Norton, New York, NY.

2) Lindsay, P.H., Norman, D.A., 1977, *Human Information Processing (2<sup>nd</sup> Edition)*, Academic Press, New York, NY

3) Minsky, Marvin, 1986, *The Society of Mind*, Simon and Schuster, New York, NY.



# 心智的计算理论

---

- CTM holds that thought is a kind of computation.
- CTM holds that the mind is a digital computing device.

# 心智的计算理论

---

- “The mind is organized into modules or mental organs, each specialized in one way of interaction with the world.”<sup>5</sup>

# 希尔伯特



David Hilbert  
(1862-1943)

希尔伯特，德国数学家。1862年1月23日生于柯尼斯堡，1943年2月14日在格丁根逝世。希尔伯特1880年入柯尼斯堡大学；1885年获博士学位；1892年任该校副教授，翌年为教授；1895年赴格丁根大学任教授，直至1930年退休。他自1902年起，一直是德国《数学年刊》主编之一。

希尔伯特是20世纪最伟大的数学家之一，他的数学贡献是巨大的和多方面的。他典型的研究方式是直攻数学中的重大问题，开拓新的研究领域，并从中寻找带普遍性的方法。1900年，希尔伯特在巴黎举行的国际数学家会议上发表演说，提出了新世纪数学面临的23个问题。对这些问题的研究有力地推动了20世纪数学发展的进程。

# 希尔伯特



David Hilbert  
(1862-1943)

希尔伯特23个问题及其解决情况：

1. 连续统假设

2. 算术公理的相容性 欧几里

得几何的相容性可归结为算术公理的相容性。希尔伯特曾提出用形式主义的证明论方法加以证明。1931年，哥德尔发表的不完备性定理否定了这种看法。1936年德国数学家根茨在使用超限归纳法的条件下证明了算术公理的相容性。

1988年出版的《中国大百科全书》数学卷指出，数学相容性问题尚未解决。

# 希尔伯特



David Hilbert  
(1862-1943)

希尔伯特23个问题及其解决情况：  
：

6. 物理学的公理化 希尔伯特建议用数学的公理化方法推演出全部物理，首先是概率和力学。1933年，苏联数学家柯尔莫哥洛夫实现了将概率论公理化。后来在量子力学、量子场论方面取得了很大成功。但是物理学是否能全盘公理化，很多人表示怀疑。

# 图灵



Alan Turing  
(1912-1954)

## • *On computable numbers with an application to the Entscheidungsproblem* (1936)

230

A. M. TURING

[Nov. 12,

ON COMPUTABLE NUMBERS, WITH AN APPLICATION TO  
THE ENTSCHEIDUNGSPROBLEM

By A. M. TURING.

[Received 28 May, 1936.—Read 12 November, 1936.]

The “computable” numbers may be described briefly as the real numbers whose expressions as a decimal are calculable by finite means. Although the subject of this paper is ostensibly the computable *numbers*, it is almost equally easy to define and investigate computable functions of an integral variable or a real or computable variable, computable predicates, and so forth. The fundamental problems involved are, however, the same in each case, and I have chosen the computable numbers for explicit treatment as involving the least cumbersome technique. I hope shortly to give an account of the relations of the computable numbers, functions, and so forth to one another. This will include a development of the theory of functions of a real variable expressed in terms of computable numbers. According to my definition, a number is computable if its decimal can be written down by a machine.

In §§ 9, 10 I give some arguments with the intention of showing that the computable numbers include all numbers which could naturally be regarded as computable. In particular, I show that certain large classes of numbers are computable. They include, for instance, the real parts of all algebraic numbers, the real parts of the zeros of the Bessel functions, the numbers  $\pi$ ,  $e$ , etc. The computable numbers do not, however, include all definable numbers, and an example is given of a definable number which is not computable.

Although the class of computable numbers is so great, and in many ways similar to the class of real numbers, it is nevertheless enumerable. In § 8 I examine certain arguments which would seem to prove the contrary. By the correct application of one of these arguments, conclusions are reached which are superficially similar to those of Gödel†. These results

† Gödel, “Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I”, *Monatshefte Math. Phys.*, 38 (1931), 173–198.

# 图灵



Alan Turing  
(1912-1954)

1936年，年仅24岁的英国人图灵发表了著名的《论应用于决定问题的可计算数字》一文，提出思考实验原理计算机概念。图灵把人在计算时所做的工作分解成简单的动作，与人的计算类似，机器需要：

- (1) 存储器，用于贮存计算结果；
- (2) 一种语言，表示运算和数字；
- (3) 扫描；
- (4) 计算意向，即在计算过程中下一步打算做什么；
- (5) 执行下一步计算。

# 图灵



Alan Turing  
(1912-1954)

具体到一步计算，则分成：

- (1) 改变数字可符号；
- (2) 扫描区改变，如往左进位和往右添位等；
- (3) 改变计算意向等。图灵还采用了二进制制。这样，他就把人的工作机械化了。这种理想中的机器被称为“图灵机”。图灵机是一种抽象计算模型，用来精确定义可计算函数。图灵机由一个控制器，一条可以无限延伸的带子和一个在带子上左右移动的读写头组成。这个概念如此简单的机器，理论上却可以计算任何直观可计算函数。图灵在设计了上述模型后提出，凡可计算的函数都可用这样的机器来实现，这就是著名的图灵论题。



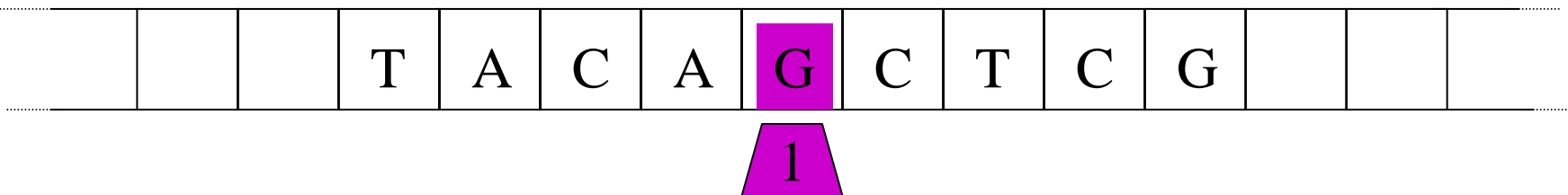
# 图灵机

## The Turing Machine

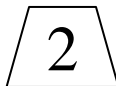


1

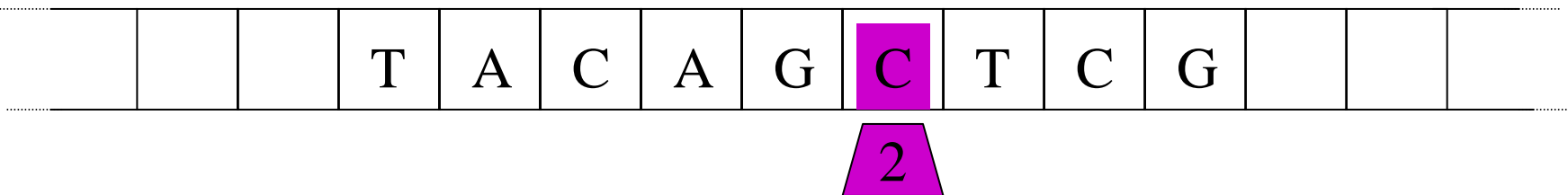
	-	A	C	G	T
0	HALT	HALT	HALT	HALT	HALT
1	-,<=,0	A,=>,1	C,=>,1	G,=>,2	T,=>,1
2	-,<=,0	A,=>,1	C,<=,3	G,=>,2	T,=>,1
3				T,=>,4	
4			A,=>,1		



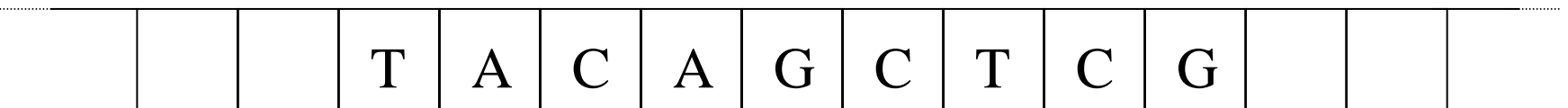
	-	A	C	G	T
0	HALT	HALT	HALT	HALT	HALT
1	-,<=,0	A,=>,1	C,=>,1	G,=>,2	T,=>,1
2	-,<=,0	A,=>,1	C,<=,3	G,=>,2	T,=>,1
3				T,=>,4	
4			A,=>,1		



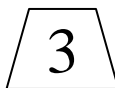
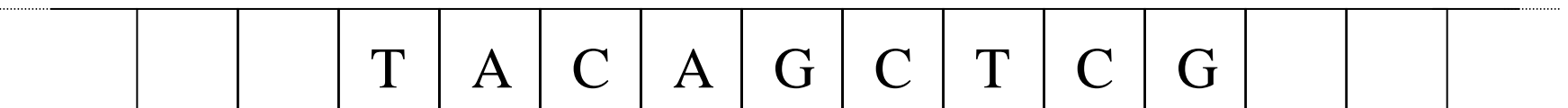
	-	A	C	G	T
0	HALT	HALT	HALT	HALT	HALT
1	-,<=,0	A,=>,1	C,=>,1	G,=>,2	T,=>,1
2	-,<=,0	A,=>,1	C,<=,3	G,=>,2	T,=>,1
3				T,=>,4	
4			A,=>,1		



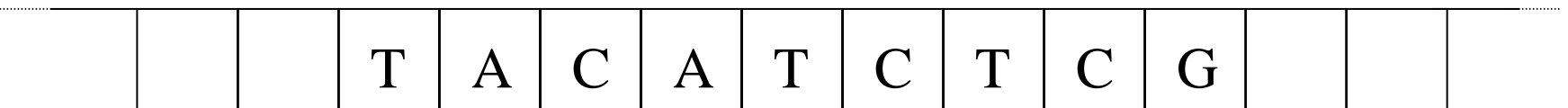
	-	A	C	G	T
0	HALT	HALT	HALT	HALT	HALT
1	-,<=,0	A,=>,1	C,=>,1	G,=>,2	T,=>,1
2	-,<=,0	A,=>,1	C,<=,3	G,=>,2	T,=>,1
3				T,=>,4	
4			A,=>,1		



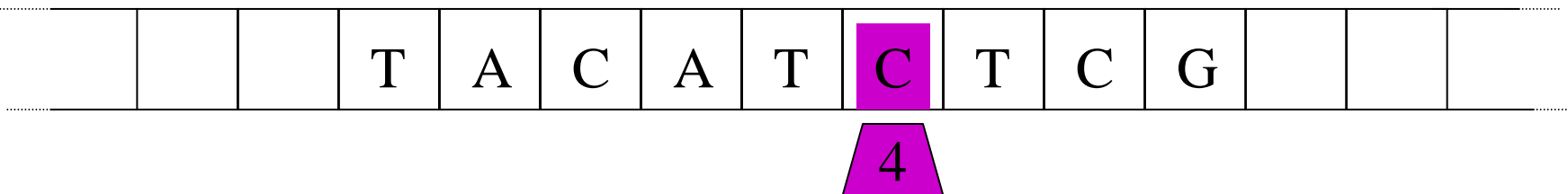
	-	A	C	G	T
0	HALT	HALT	HALT	HALT	HALT
1	-,<=,0	A,=>,1	C,=>,1	G,=>,2	T,=>,1
2	-,<=,0	A,=>,1	C,<=,3	G,=>,2	T,=>,1
3				T,=>,4	
4			A,=>,1		



	-	A	C	G	T
0	HALT	HALT	HALT	HALT	HALT
1	-,<=,0	A,=>,1	C,=>,1	G,=>,2	T,=>,1
2	-,<=,0	A,=>,1	C,<=,3	G,=>,2	T,=>,1
3				T,=>,4	
4			A,=>,1		

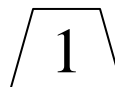
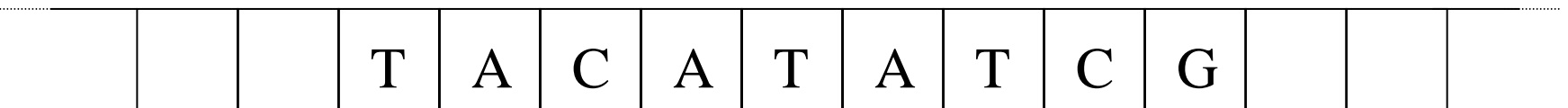


	-	A	C	G	T
0	HALT	HALT	HALT	HALT	HALT
1	-,<=,0	A,=>,1	C,=>,1	G,=>,2	T,=>,1
2	-,<=,0	A,=>,1	C,<=,3	G,=>,2	T,=>,1
3				T,=>,4	
4			A,=>,1		



	-	A	C	G	T
0	HALT	HALT	HALT	HALT	HALT
1	-,<=,0	A,=>,1	C,=>,1	G,=>,2	T,=>,1
2	-,<=,0	A,=>,1	C,<=,3	G,=>,2	T,=>,1
3				T,=>,4	
4			A,=>,1		





	-	A	C	G	T
0	HALT	HALT	HALT	HALT	HALT
1	-,<=,0	A,=>,1	C,=>,1	G,=>,2	T,=>,1
2	-,<=,0	A,=>,1	C,<=,3	G,=>,2	T,=>,1
3				T,=>,4	
4			A,=>,1		

# 图灵测试

- 半个世纪以来，数学家提出的各种各样的计算模型都被证明是和图灵机等价的。**1945**年，图灵到英国国家物理研究所工作，并开始设计自动计算机。
- **1950**年，图灵发表了题为《计算机能思考吗？》的论文，给人工智能下了一个定义，而且论证了人工智能的可能性。**1951**年，他被选为英国皇家学会会员。

# 物理符号系统



## Herbert A. Simon

was born in 1916 in Milwaukee, Wisconsin. He began his career at the Cowles Commission. In 1978 he was a Nobel Laureate in Economics. He is currently serving as a Richard King Mellon University Professor of Computer Science and Psychology at Carnegie Mellon University.

# 物理符号系统

- 我们把人看成一个信息加工系统，常称作物理符号系统。用物理符号系统主要是强调所研究的对象是一个具体的物质系统，如计算机的构造系统，人的神经系统、大脑神经元等。所谓符号就是模式；任何一个模式，只要它能和其它模式相区别，它就是一个符号。不同的英文字母就是不同的符号。对符号进行操作就是对符号进行比较，即找出哪个是相同的符号，哪几个是不同的符号。物理符号系统的基本任务和功能就是辨认相同的符号和区分不同的符号。符号既可以是物理的符号，也可以是头脑中的抽象的符号，可以是计算机中的电子运动模式，或可以是头脑中的神经元的某种运动方式。纸上的文字是物理符号系统，但这是一个不完善的物理符号系统，因为它的功能只是存储符号，即把字保留在纸上。一个完善的符号系统还应该有更多的功能。

# 物理符号系统

- 它由记忆、一组操作、控制、输入和输出构成。它的输入是确定部位的客体；它的输出是确定部位的客体的修改或建立，后者部位常与输入是不同的。那么，它的外部行为就由输出组成，它们的产生是输入的函数。大的环境系统加上物理符号系统就形成封闭系统，因为输出客体变成后面的输入客体，或者影响后面的输入客体。物理符号系统的内部状态由它的记忆和控制的状态构成。它的内部行为是由这些内部状态全部变化构成。

# 物理符号系统

- 记忆是由一组符号结构  $\{E_1, E_2, \dots, E_m\}$  组成，在整个时间里它们在数量和内容上是变化的。符号结构的内部改变称作表达。为了定义符号结构给出一组抽象符号  $\{S_1, S_2, \dots, S_n\}$ 。每种符号结构都具有给定的类型和一些不同的作用  $\{R_1, R_2, \dots\}$ ，每种作用包括一个符号。采用显式表示可以写成  $(\text{Type: } T \ R_1: S_1 \ R_2: S_2, \dots, R_n: S_n)$  若用隐式表示，则写成：
  - $(S_1 S_2, \dots S_n$

# 物理符号系统

纽威尔 (Allon Newell) 规定了10种操作符。每一个在图中表示一块。这10种操作符的功能如下：

- (1) 赋值符号 (Assign a symbol): 建立符号与项之间的基本关系。对项赋值，称之为存取。符号可以赋给项，而不能赋给表达式。存取一个操作符意味着存取它的输入、输出和唤醒机制。存取给定类型的作用意味着存取作用的符号，这种作用是给定的类型的表达式有关，并在那种作用写入新的符号。
- (2) 复制表达式 (Copy expression): 将表达式和符号加到系统里，新的表达式是输入表达式准确的复制，即在各种作用中具有完全相同的类型和符号。
- (3) 写表达式 (Write an expression): 建立任何规定内容的表达式。它并不建立任何新的表达式，而是修改它的输入表达式。
- (4) 写 (write): 在给定的作用建立一个符号。
- (5) 读 (Read): 在给定作用下读符号。

# 物理符号系统

- (6) 执行序列 (Do sequence): 使系统按规定的序列执行任何动作。
- (7) 条件退出和条件继续 (Exit-if and Continue—if): 系统行为有条件地继续执行一个序列, 或从中退出。
- (8) 引用符号 (Quote a symbol): 控制自动地解释被运行的表达式。
- (9) 外部行为 (Behve externally): 符号系统可控的外部行为的集合。
- (10) 环境输入 (Input from enuironment): 利用记忆中新建立的表达式将外部环境的输入录入到系统中。



# 物理符号系统

我们可以将物理符号系统的功能简化成6种，即：

(1) 输入符号。

(2) 输出符号。

(3) 存储符号。

(4) 复制符号。

(5) 建立符号结构：通过找到各种符号之间的关系，在符号系统中形成符号结构。

(6) 条件转移：如果在记忆中已经有了一定的符号系统，再加上外界的输入，就可以继续完成行为

。

# 物理符号假设

- 1976年纽威尔和司马贺提出了物理符号系统假设，说明物理符号系统的本质。主要假设内容如下：
- 物理符号系统假设，物理系统表现智能行为必要和充分的条件是它是一个物理符号系统。
- 必要性意味着表现智能的任何物理系统将是一个物理符号系统的例示。
- 充分性意味着任何物理符号系统都可以进一步组织表现智能行为。
- 智能行为就是人类所具有的那种智能：在某些物理限制下，实际上所发生的适合系统目的和适应环境要求的行为。

# 诺尔曼模型

- 人的心理活动是一个整体，是认知因素(感知、表象、记忆、思维等)和非认知因素(情感、意志、动机、兴趣等)的辩证统一，因此必须考虑情感等因素对认知和行为的影响。诺尔曼在1979年提出诺尔曼模型。这个模型是以控制系统为中心，感觉输入和运动输出都直接与它相联。在控制系统和认知系统之间是情感系统，指出情感等因素的重要性。

# The Basics of the Model

---

**Based on Norman's (1980) model of the mind and mental functioning:**

**Cognitive  
system**

**Encodes, processes, stores,  
retrieves, processes  
information; purpose is  
manipulation of information**

**Norman, D. (1980). Twelve issues for cognitive science. Cognitive Science, 4, 1-32.**

# The Basics of the Model

---

**Based on Norman's (1980) model of the mind and mental functioning:**

**Affective  
system**

**Connects the regulatory  
system and cognitive system;  
purpose is arousal**

**Norman, D. (1980). Twelve issues for cognitive science. Cognitive Science, 4, 1-32.**

# The Basics of the Model

---

**Based on Norman's (1980) model of the mind and mental functioning:**

**Regulatory  
system**

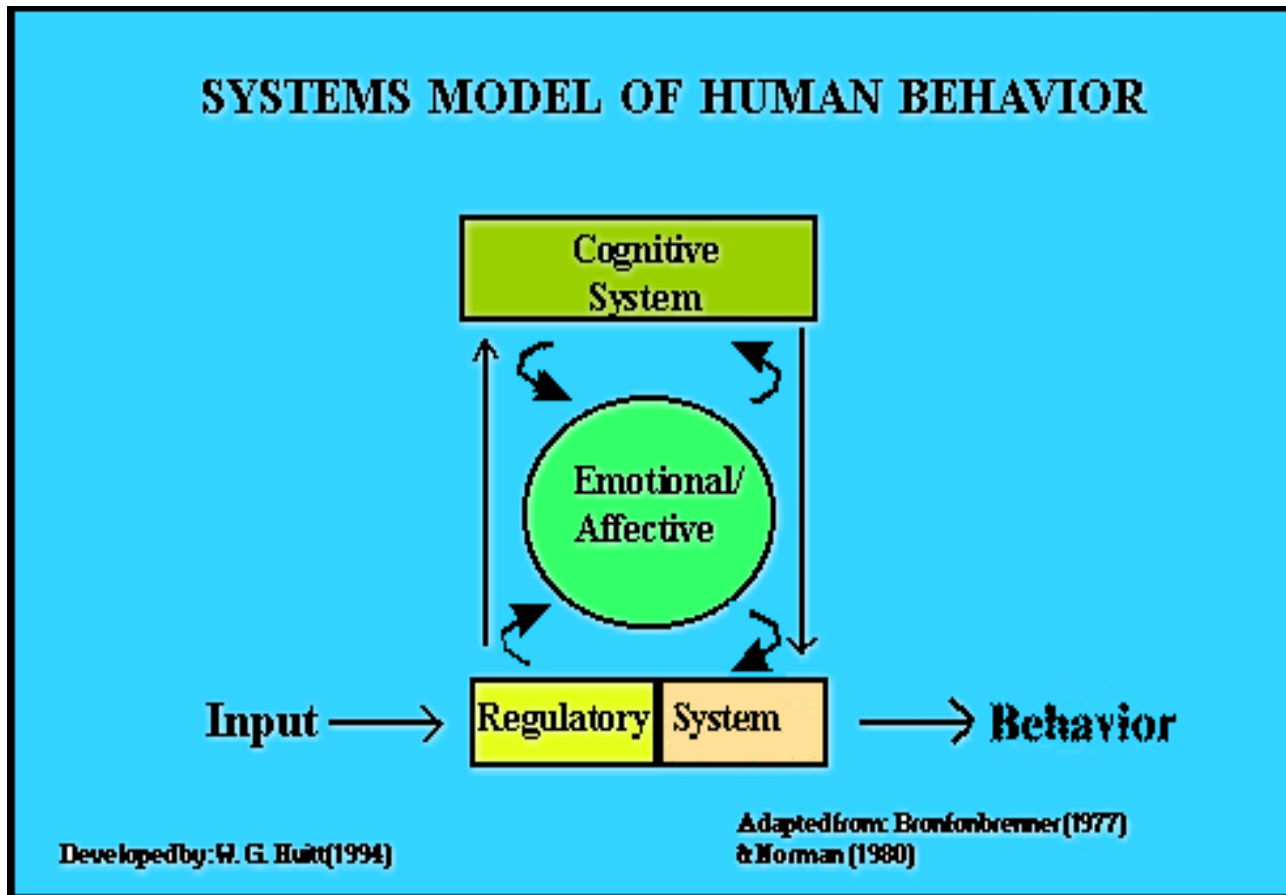
**Biological structures that regulate the processing of stimuli and activation of responses; primary purpose is maintenance of the organism**

**Norman, D. (1980). Twelve issues for cognitive science. Cognitive Science, 4, 1-32.**

---

# 诺尔曼模型

## The Basics of the Model



# The Basics of the Model

---

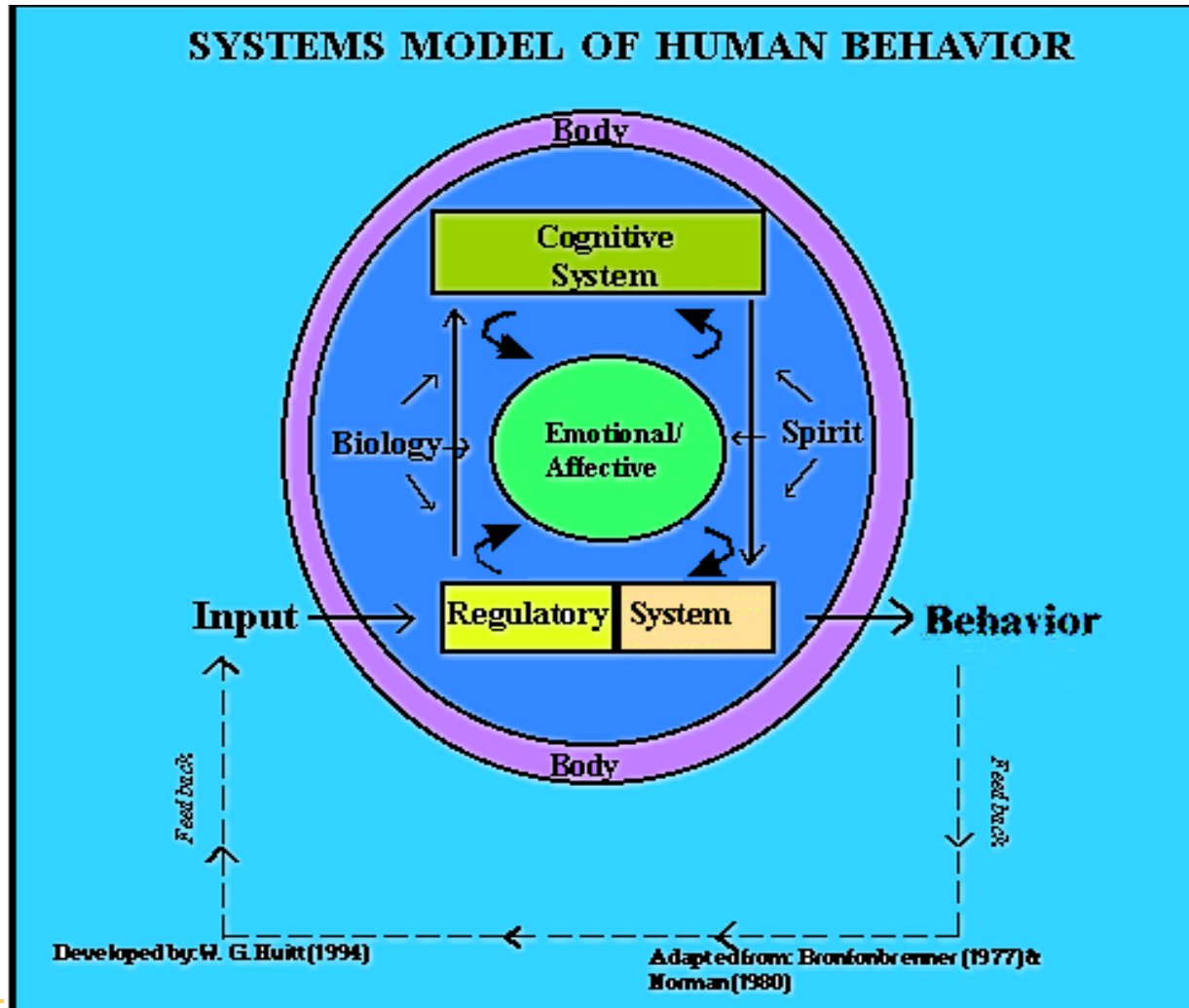
**The regulatory system is redefined as the aspect of the mental system that regulates input and output functions (not just biological functions).**

**Other modifications:**

- mind receives information and displays action through the body**
- adds a biological and spiritual influence**
- adds a feedback loop**



# The Basics of the Model



# The Basics of the Model

---

**It is hypothesized that an individual's thinking (cognition), feeling (affect), and willing (conation) as well as overt behavior develop as a result of:**

- **transactions among the various components of mind as**
- **influenced by biological maturation, bodily functioning and the spiritual dimensions,**
- **the environment or context of the individual, and**
- **feedback from the environment as a result of an individual's overt behavior.**

# 认知记忆模型

- 日本的小谷津孝明于1982年提出的认知——记忆信息处理模型
- 为了理解人的高级认知功能，仅仅是理解输入信息的分析过程是不充分的，必须要与长期记忆中已有的世界知识密切配合是非常必要的。那种不是信息单方向流动的结果，而是根据输入信息记忆内部状态的变化，把内部状态有关的变化作为整体形成一种认知活动，就是概念化的过程。这种类型的模型的开发，多数是与自然语言理解的研究有关，是以知识表示和利用为中心。

# 认知记忆模型

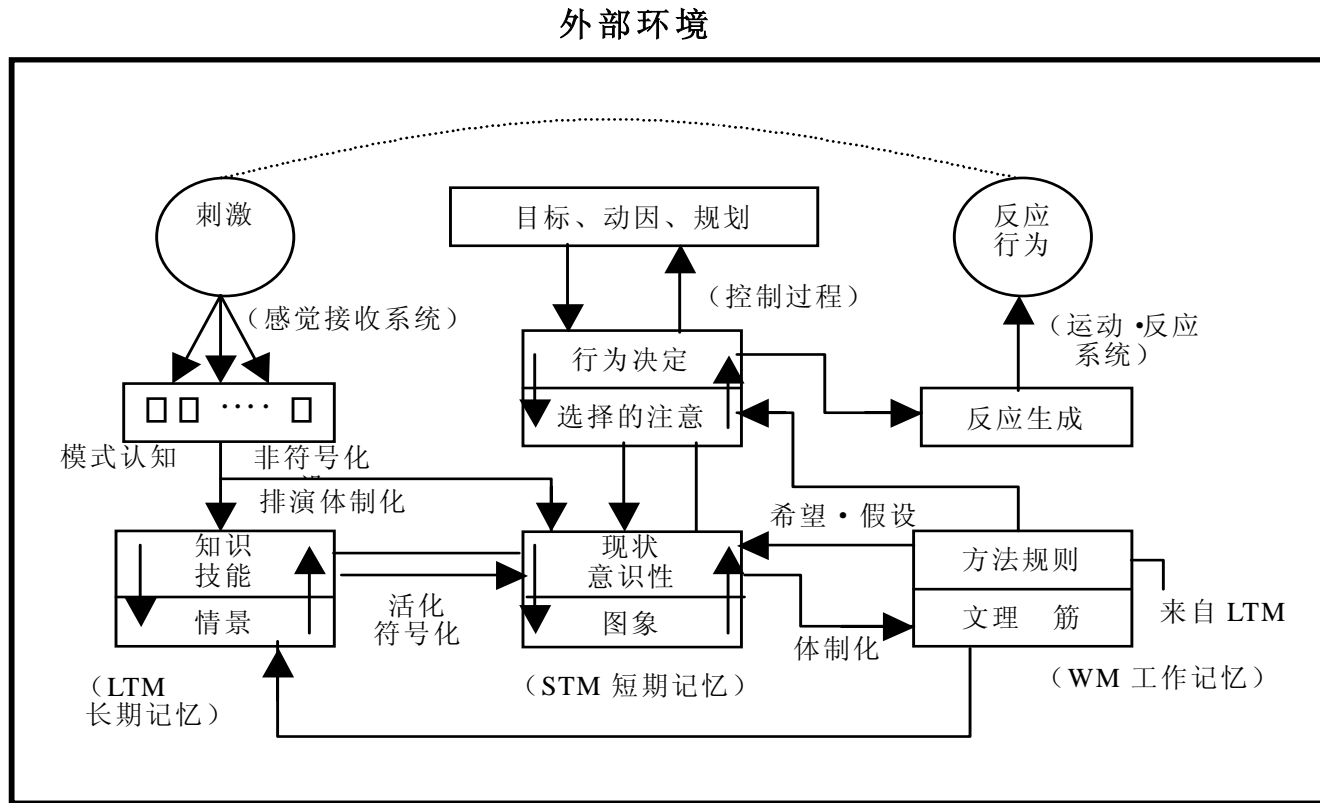


图 认知——记忆信息处理模型

# ACT模型

---

- 美国心理学家安德森(**J.A. Anderson**)于**1976**年提出系统的整合理论与人脑如何进行信息加工活动的理论模型,简称**ACT模型[Anderson 1976]**,原意为“思维的适应性控制”。安德森将人类联想记忆模型(**HAM**)与产生式系统的结构相结合,模拟人类高级认知过程的产生式系统,在人工智能的研究中有重要意义。

# History of the ACT-framework\*



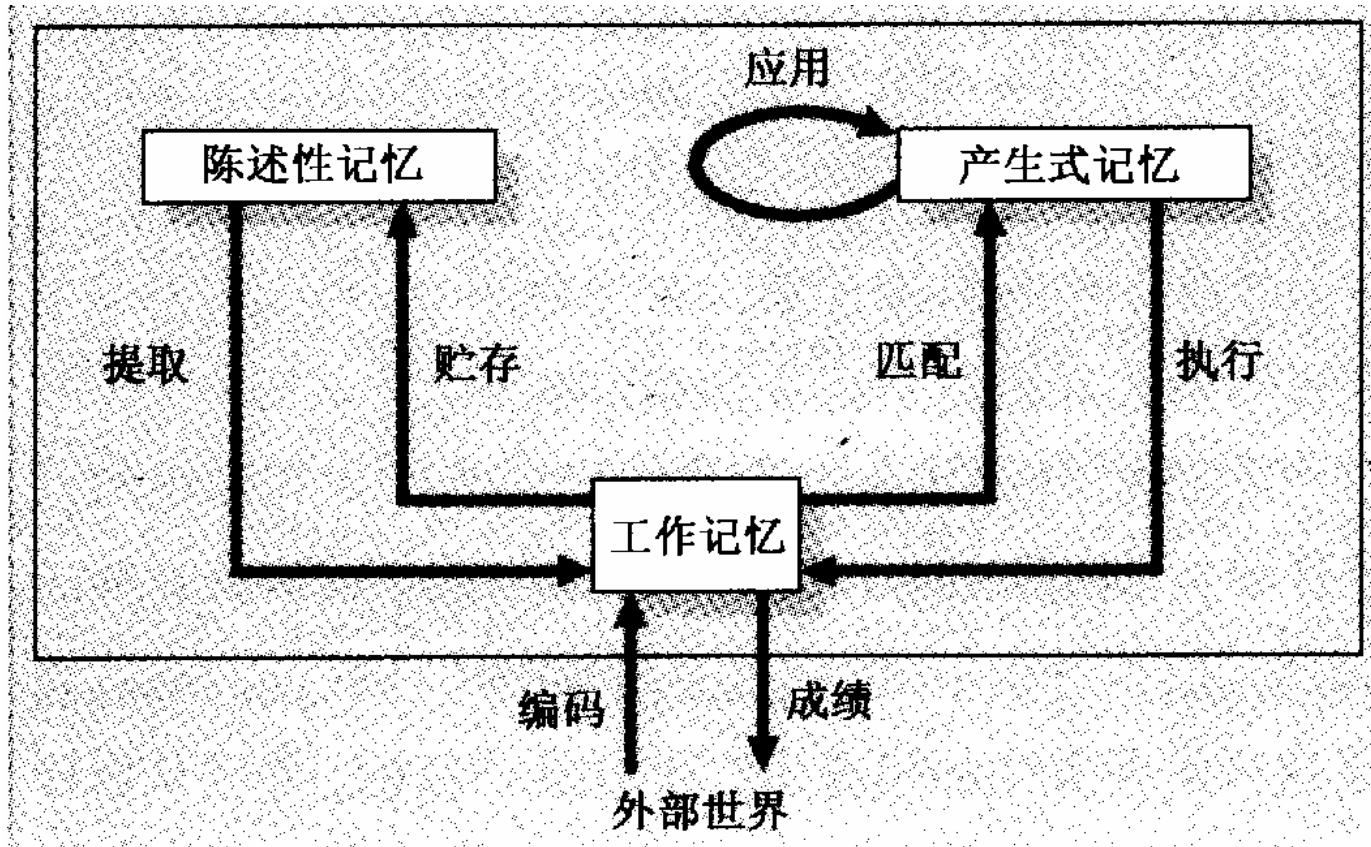
Predecessor	HAM	(Anderson & Bower 1973)
Theory versions	ACT-E	(Anderson, 1976)
	ACT*	(Anderson, 1978)
	ACT-R	(Anderson, 1993)
	ACT-R 4.0	(Anderson & Lebiere, 1998)
	ACT-R 5.0	(Anderson & Lebiere, 2001)
Implementations	GRAPES	(Sauers & Farrell, 1982)
	PUPS	(Anderson & Thompson, 1989)
	ACT-R 2.0	(Lebiere & Kushmerick, 1993)
	ACT-R 3.0	(Lebiere, 1995)
	ACT-R 4.0	(Lebiere, 1998)
	ACT-R/PM	(Byrne, 1998)
	ACT-R 5.0	(Lebiere, 2001)
	Windows Environment	(Bothell, 2001)
	Macintosh Environment	(Fincham, 2001)
	ACT-R 6.0	(Bothell, 2004??)

# ACT-R



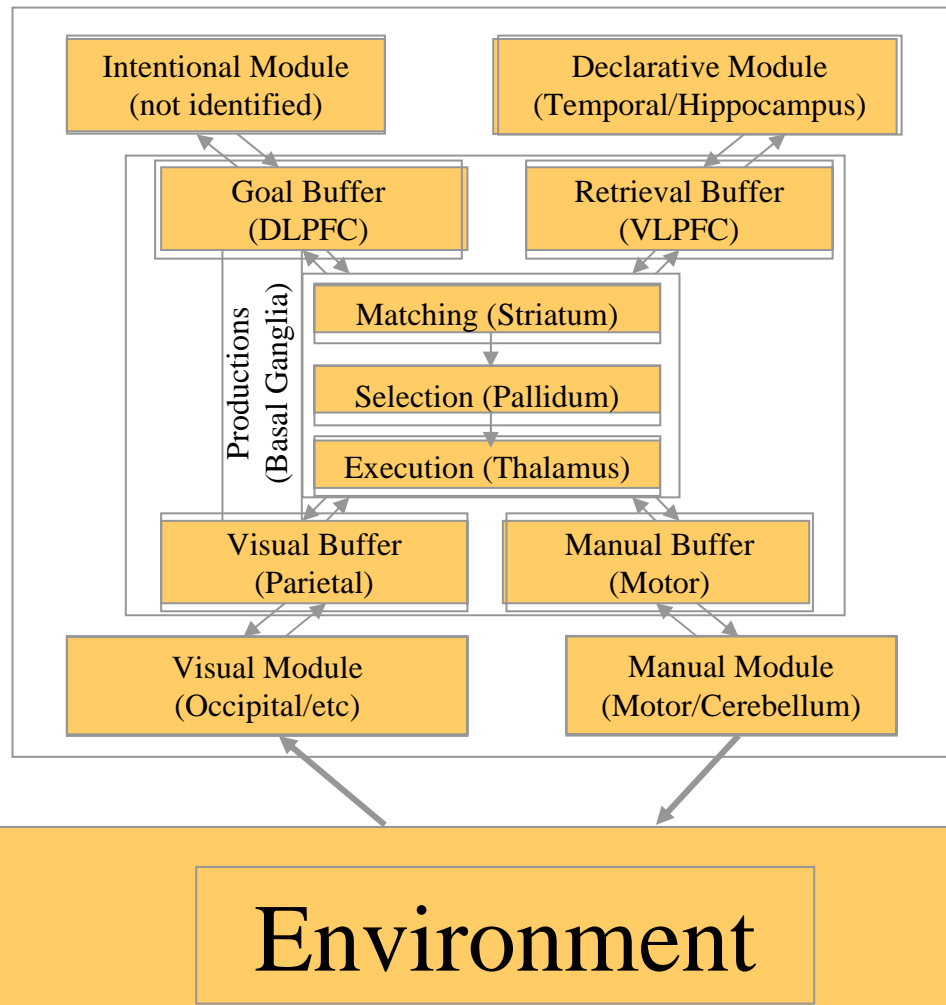
- Procedural vs Declarative Knowledge
- Declarative
  - Aware of
  - Can describe to others
  - Facts
- Procedural
  - Display in behavior
  - Not conscious of
  - e.g. language

# ACT-framework\*

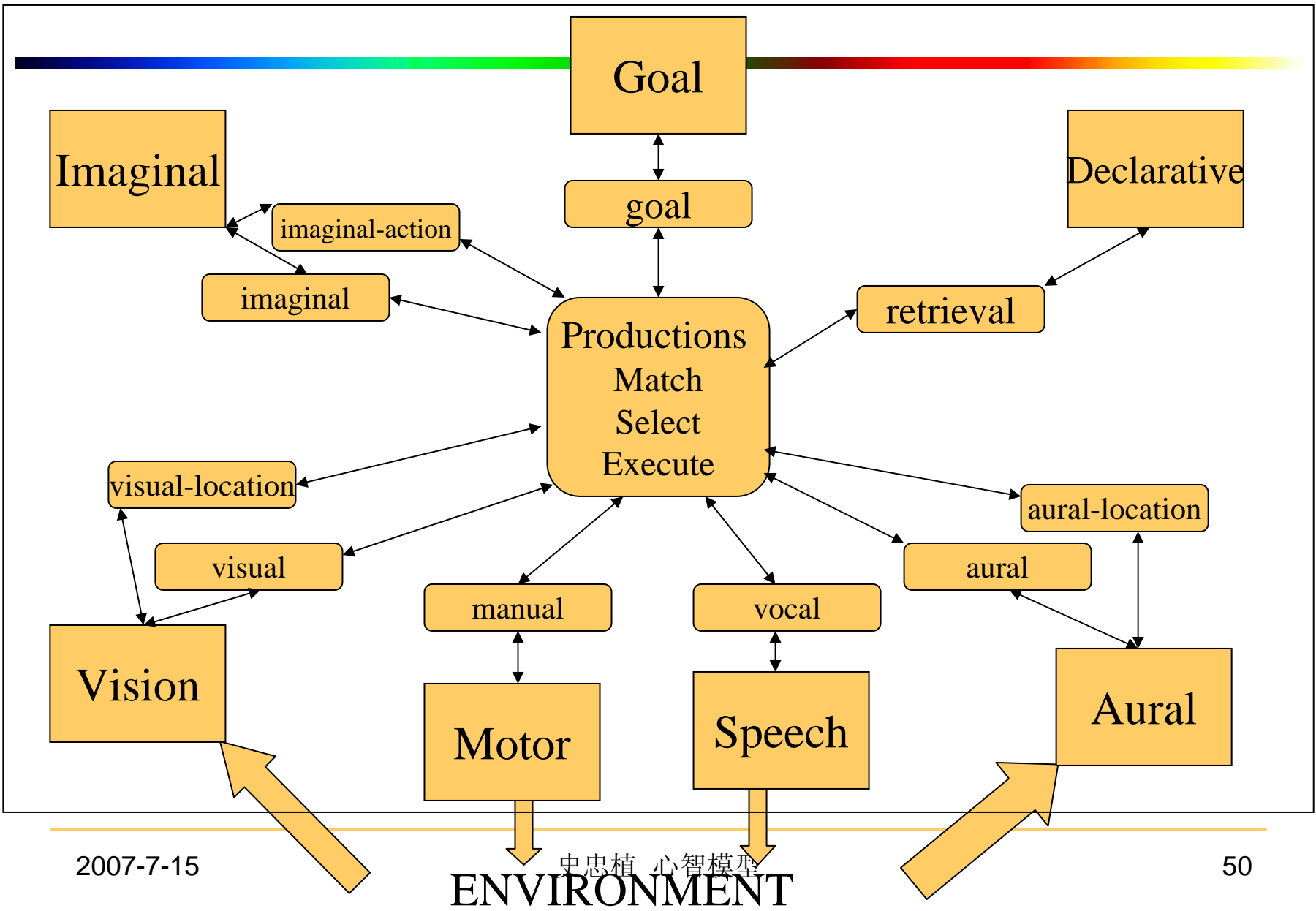




# ACT-R 6.0 Mapping to the Brain\*

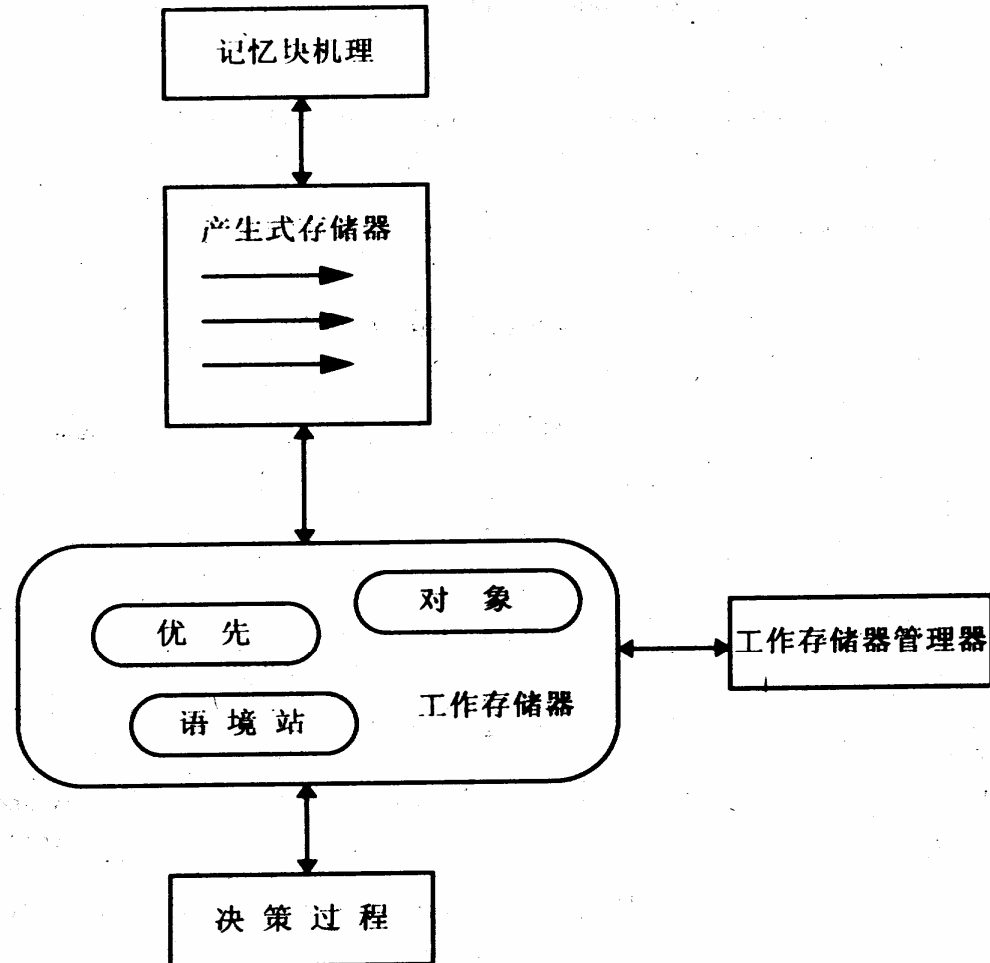


# ACT-R 6.0



# SOAR

- 开发智能系统的统一架构



# SOAR



- 确立计算结构：
  - 编码知识Encode knowledge
  - 访问知识Access encodings
  - 产生动作Produce actions
  - 达到目标Achieve goals

# 理论基础

---

- Physical Symbol Systems Hypothesis
- Problem Space Hypothesis
- Assumption:
  - Parsimonious functional mechanisms provide the basis for a unified theory of cognition
  - Humans are agent systems/agent systems need to be human-like

# 知识表示

---

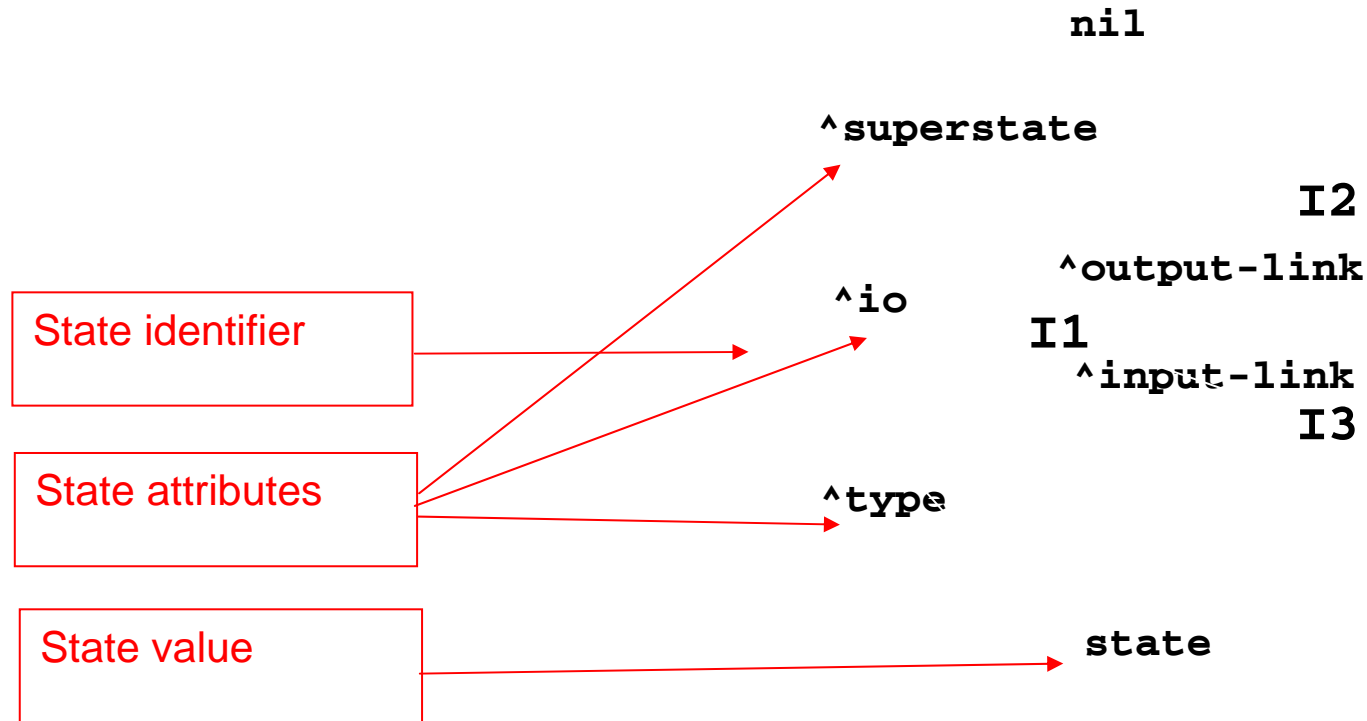
- Long-term knowledge
  - Relational pattern recognizers/producers
    - “Productions”
- Working knowledge
  - Blackboard of frames or objects (with associated attributes and values)
  - First-class objects
    - States
    - Operators

# 实现问题

---

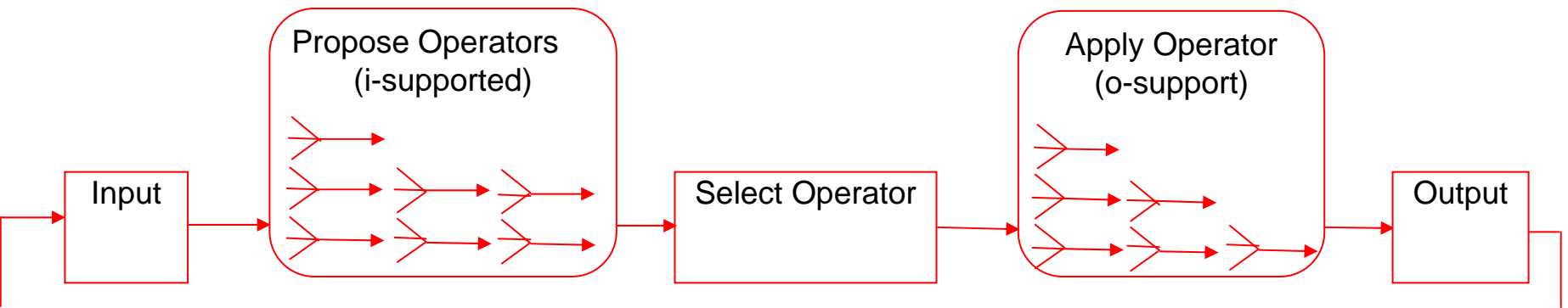
- Long-term memory
  - Production rules; relational stimulus-response patterns
  - Variablized graph structure
  - Productions match and fire in parallel
- Working memory
  - Graph of objects linked by attributes
  - Stored as triples
  - Special objects
    - *State*: Created automatically by Soar
    - *Operator*: Created by productions
  - Only one operator at a time per state
  - Contrast with other rule-based systems

# 工作记忆





# 决策周期



# 产生式是基本表示

---

- Soar uses productions to represent all knowledge
  - Inference rules
  - Action selection rules, operators
  - Metacognitive rules
  - Declarative knowledge
- This doesn't preclude use of higher-level language for any of these purposes – as long as it can be “compiled” into productions

# 问题空间搜索

---

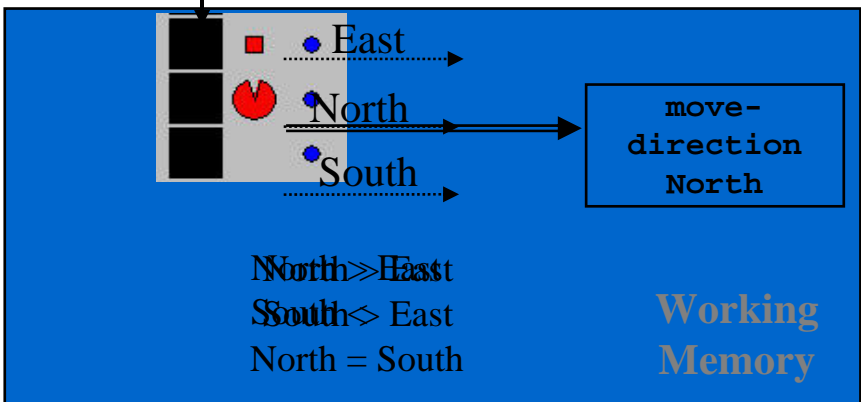
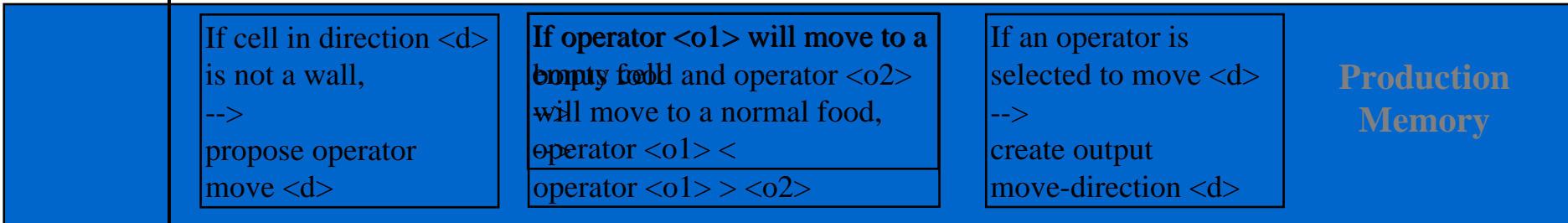
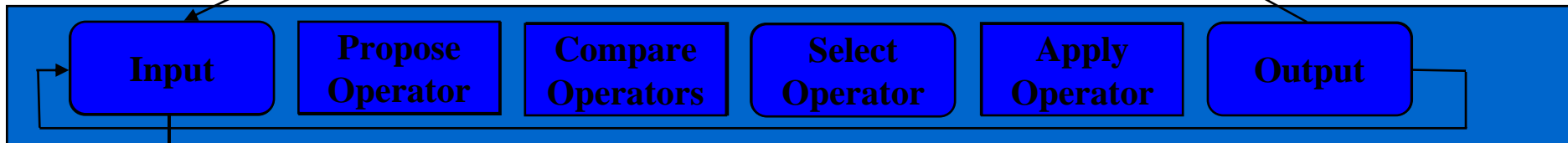
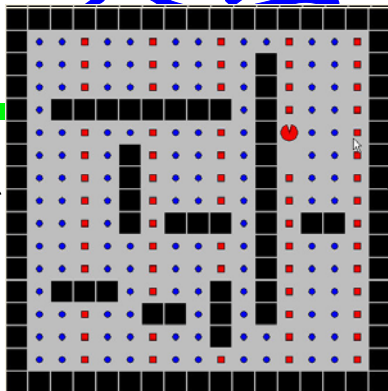
- Problem space:
  - Start state, goal state
  - Operators (actions that transform state)
- Soar assumption: every cognitive behavior can be characterized as search for a sequence of operators leading from a start state to a goal state
  - Routine behavior
  - Problem-solving
  - Meta-cognition / search control
- “Universal weak method”

# 难局驱动的审议和学习

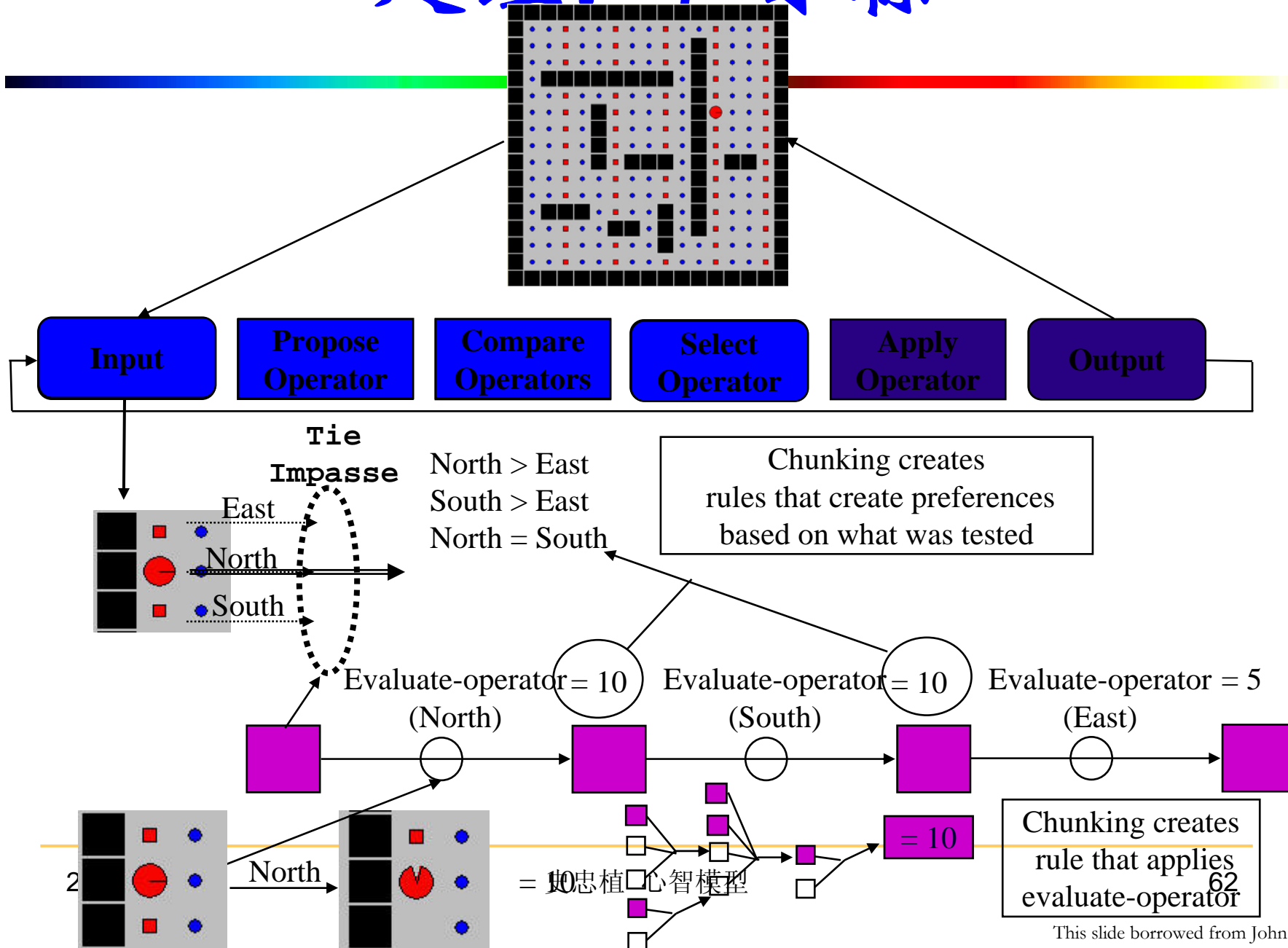
---

- Soar's standard process for selecting action includes several intermediate decision points
- If each such decision is completely determined by existing productions, the action is “routine”
- Underdetermined decisions (zero or multiple conflicting productions) create an “impasse”
  - Results in deliberation (e.g. problem space search)
  - Results in learning a new rule that prevents impasse in future similar situations

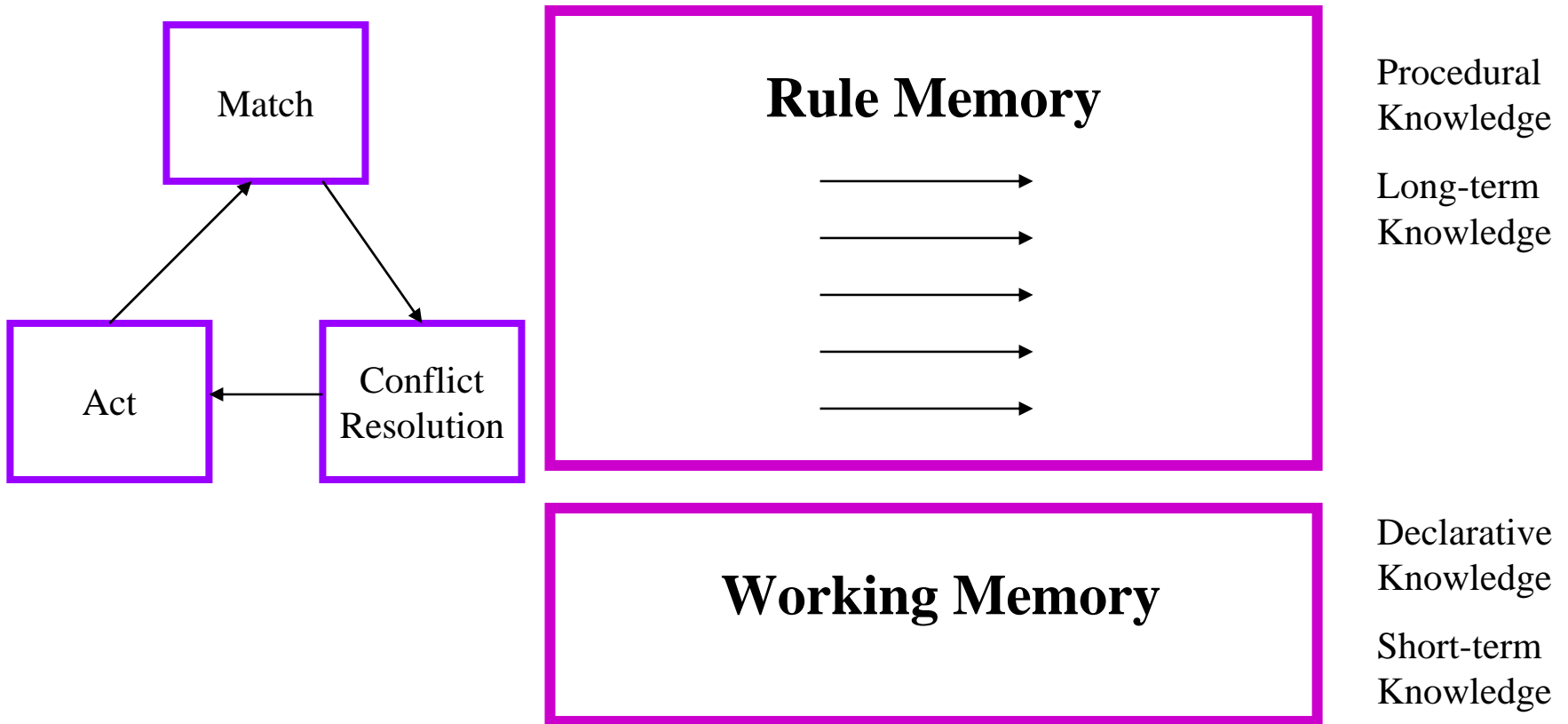
# 处理



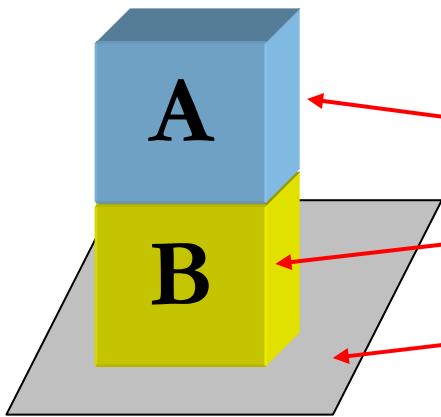
# 处理：子目标



# 表示和方法论Memories



# 表示和方法论 Working Memory



```
(s1 ^block b1 ^block b2 ^table t1)
(b1 ^color blue ^name A ^ontop b2 ^size 1
 ^type block ^weight 14)
(b2 ^color yellow ^name B ^ontop t1 ^size 1
 ^type block ^under b1 ^weight 14)
(t1 ^color gray ^shape square
 ^type table ^under b2)
```



# 表示和方法论 Rule Syntax

If I exist,  
then write "Hello World" and halt.

```
sp {hello-world
    (state <s> ^type state)
-->
    (write "Hello World")
    (halt)}
```

# 表示和方法论

## Operators

Propose\*hello-world:

If I exist, propose the hello-world operator.

Apply\*hello-world:

If the hello-world operator is selected, write "Hello World" and halt.

```
sp {propose*hello-world
    (state <s> ^type state)
-->
    (<s> ^operator <o> +)
    (<o> ^name hello-world)}
```

Creating acceptable preference for operator

```
sp {apply*hello-world
    (state <s> ^operator <o>)
    (<o> ^name hello-world)
-->
```

Testing selected operator

```
(write 15 |Hello World|)
(halt)}
```

# 可用性

- Getting a cognitive architecture to generate behavior is essentially a programming problem
  - Productions are low-level (“atomic”)
  - High-level language (TAQL) had limited success
- Cognitive architecture pose many usability problems
  - Installation
  - Learnability
  - Performance
  - Debugging support
  - Interoperability facilitation
  - Verification and validation

Etc.

# 可用性

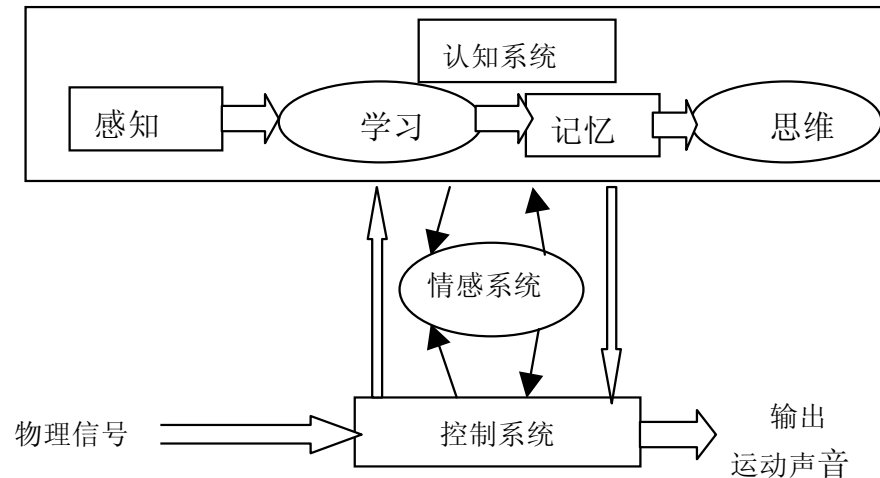
Dozens of usability problems identified at 23<sup>rd</sup> Soar workshop

---

- Informationally undense code with distal connections
- Syntax is difficult and not supported
- Mismatch in language names to AI
- Basic interface problems (lack of, bugs)
- Lack of example models
- Need graphical view of working memory
- System can't be learned without understanding theory behind it all
- Need ability to set breakpoints
- Need ability to trace / undo
- Graphical view of working memory?
- Representational compactness
- Operator-level profiling (why is operator so expensive?)

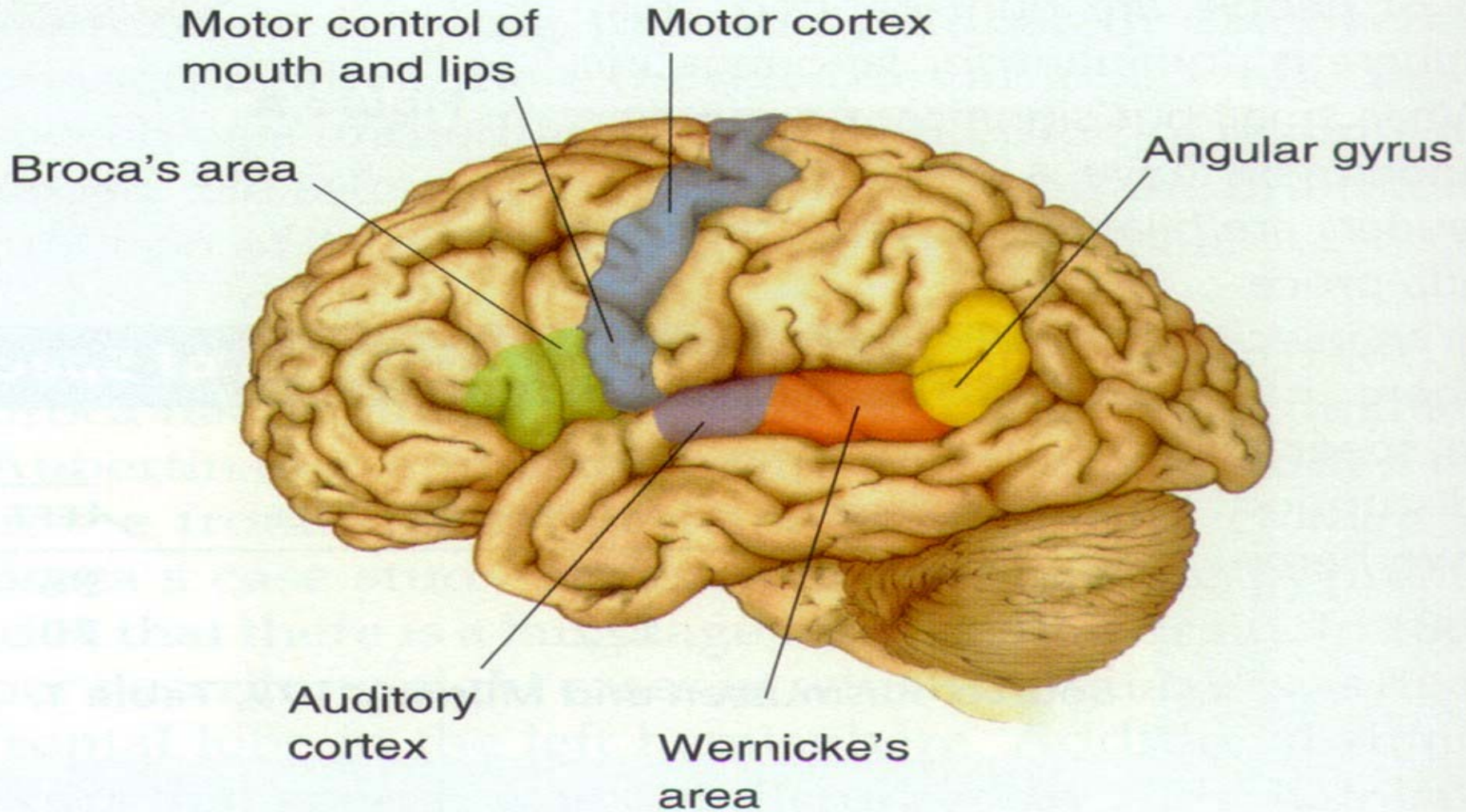
# 心智模型

心智  
(mind)  
是人脑  
的软件  
系统。



心智模型

# 脑语言处理



# Wernike-Geschwind 模型

VI --> angular gyrus 角回



弓状神經束

A1 --> Wernike's area -----> Broca's area



M2,MI

# W-G 模型的解釋

---

- 讀字或聽字複誦。
- Broca's 失語能聽話，難說話。
- Wernike's 失語聽不懂，口若懸河。
- 弓形神經束受損，聽得懂，說得流利，也知道自己在講啥；但不能聽字複誦。



# W-G 模型的局限

---

- 各單位的「功能」，不見得像所假設的那樣單純。
  - 如，**Broca's area**不只處理如何發音而已。
- **VI**另有通道直達**Broca's area**。
- 中風後，失語症狀會有改善，代表其他部位有補償功能，即，不限於所假設之部位。

# Marvin Lee Minsky



- **1927: Geboren in New York**
- **1950: Abschluss als Bachelor of Arts**
- **1954: Abschluss Studium der Mathematik mit PhD.**
- **1959: Gründung des M.I.T AI Laboratory**

# Marvin Lee Minsky



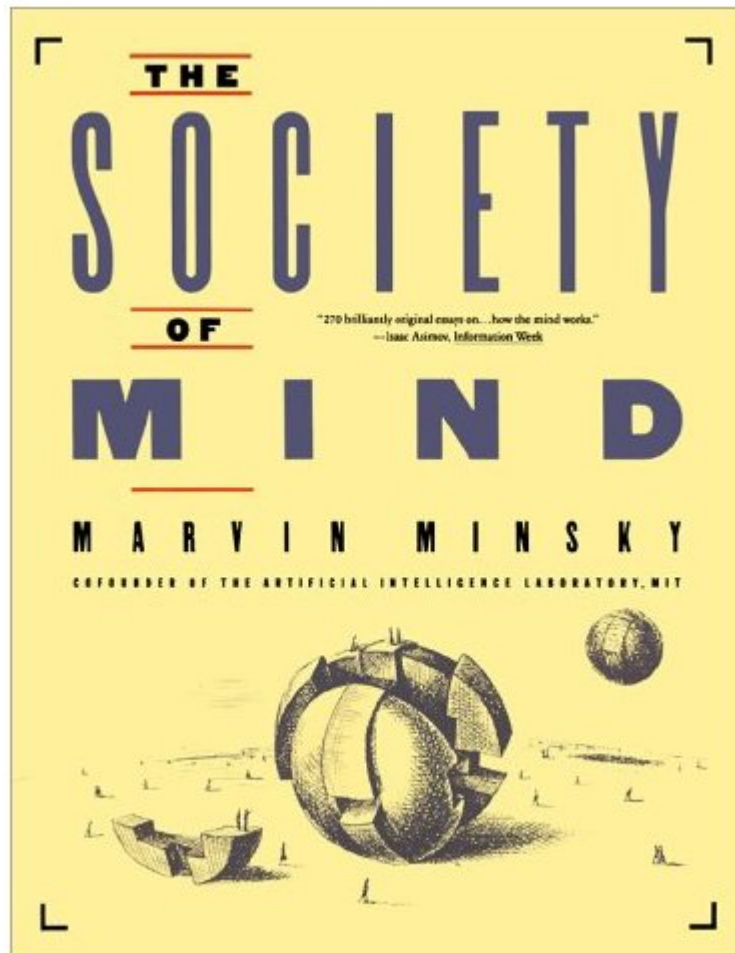
- **1969: Entwicklung der ersten Roboterhand**
- **1969: Veröffentl. v. „Perceptrons“ (mit S. Papert)**
- **1987: Veröffentl. v. „The Society of Mind“**

# Marvin Lee Minsky



- **1969: Erhalt des ACM Turing Awards**
- **1990: Erhalt des Japan Preises**
- ...

# 心智社会



明斯基(M. Minsky)于1985年出版了《心智的社会》一书。他在这本书中指出，智能并非存在于中央处理器中，而是在许多具有专门用途、彼此紧密联结的机器的集体行为中产生的。明斯基指出：心智是由许多称作智能主体(agent)的小处理器组成；每个主体(agent)本身只能做简单的任务，他们并没有心智；当主体(agent)构成社会，就得到智能。

# Society Theory of Mind (STM) 6

---

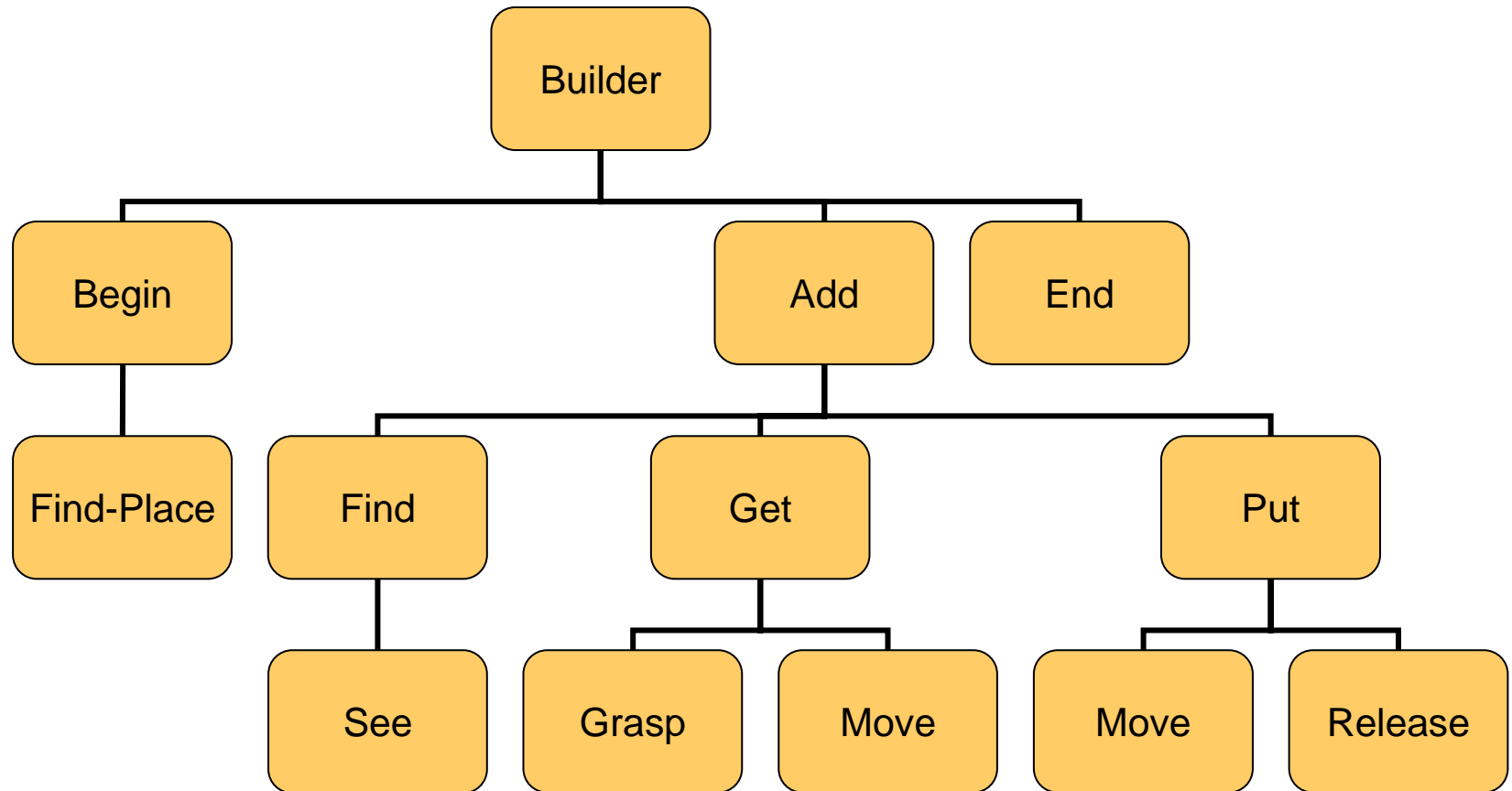
- Mind is made of many smaller processes, called *mental agents*.
- Each agent can only do a simple (almost *mindless*) task by itself.
- When agents are joined in special ways (in *societies*) we get intelligence.

# Mental Agents

---

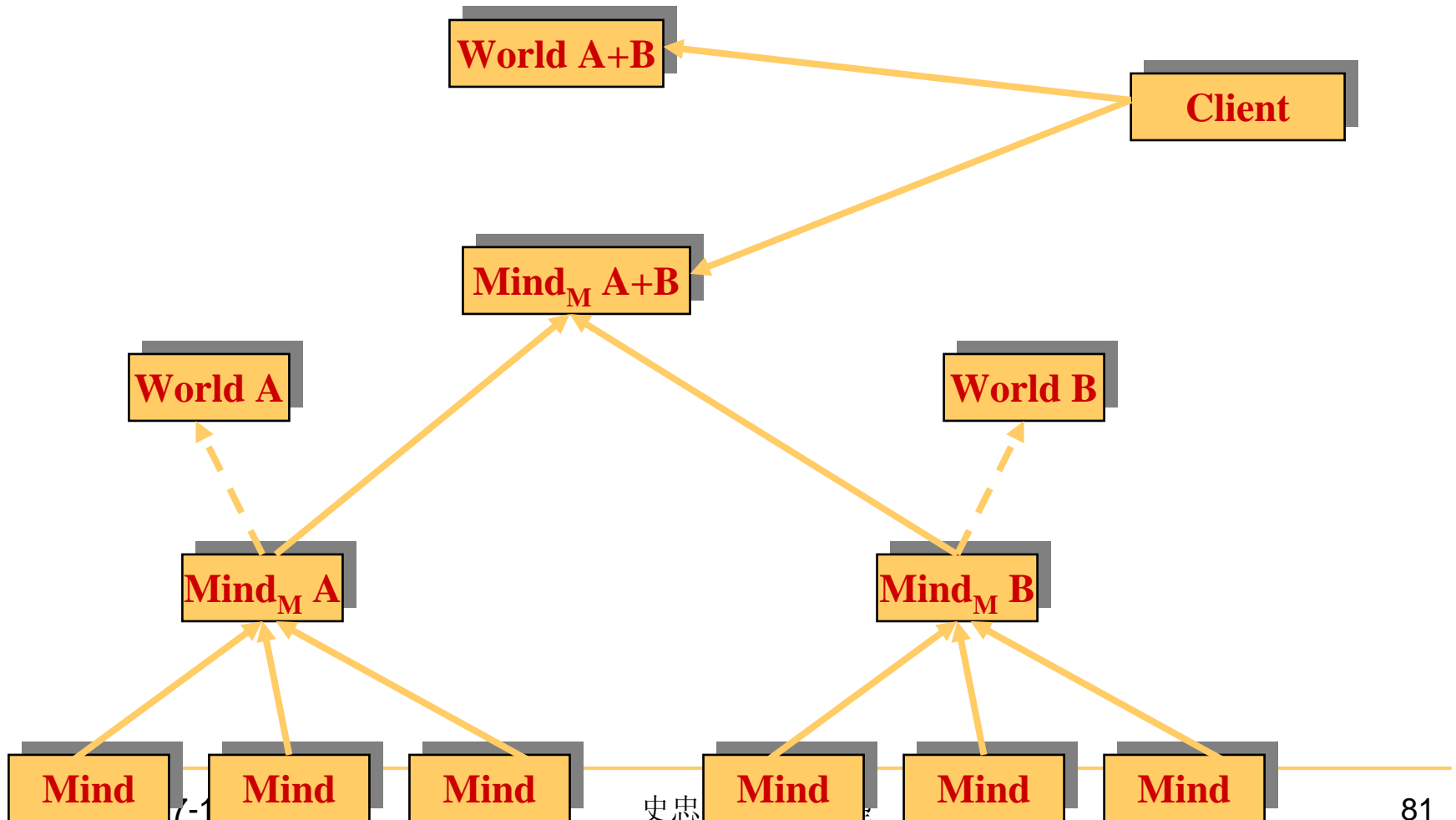
- Task: Make a tower from blocks
- Some of the agents involved in task:
  - Builder
  - Begin
  - Add
  - End
  - Find
  - Get
  - Put
  - See
  - Grasp
  - Move
  - Release
  - Find-Place

# Mental Agents in a *Bureaucracy*





# 心智社会



# 构建复杂心智联机系统

Constructing complex minds, online

---

Ciarán O'Leary  
Dublin Institute of Technology  
22nd May 2003

**Technology for automated assessment:  
The World-Wide-Mind**

# The Jolly Little Creature's Mind

---

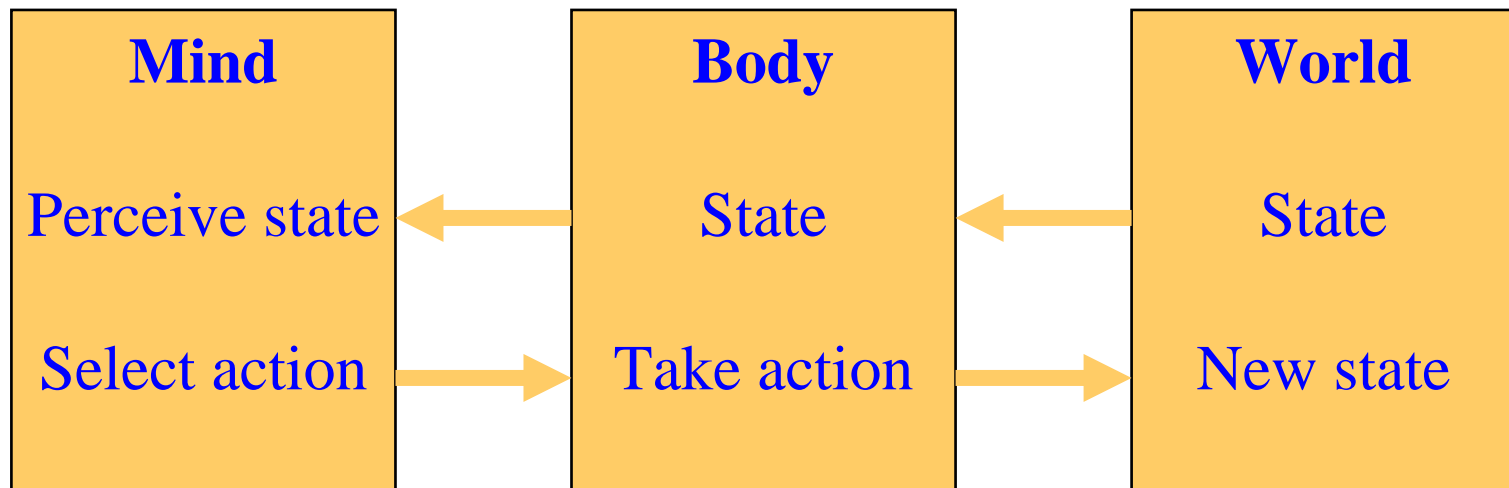
- Multiple (massively diverse) modules
- *Since, according to faculty psychologists, the mental causation of behaviour typically involves the simultaneous activity of a variety of **distinct psychological mechanisms**, the best research strategy would seem to be divide and conquer: first study the **intrinsic characteristics of each of the presumed faculties**, then study the **ways in which they interact***
  - Jerry Fodor – The Modularity of Mind, p1.

# Building JLC's Mind

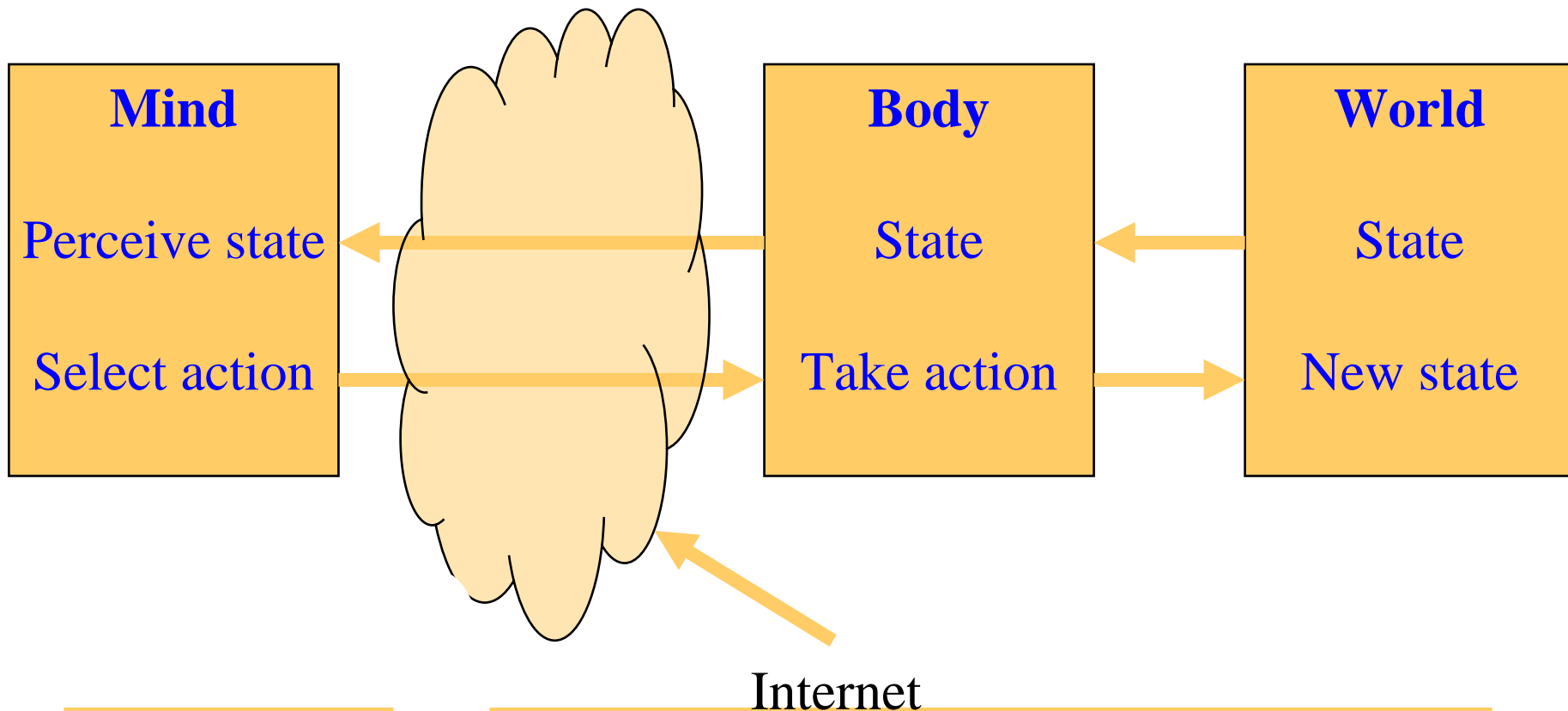
---

- Requirements
  - Architecture
  - Standard/protocol to use to integrate components
- Requirement for protocol
  - No barriers to entry
  - Simple
  - Not tied to any platform

# JLC's Mind

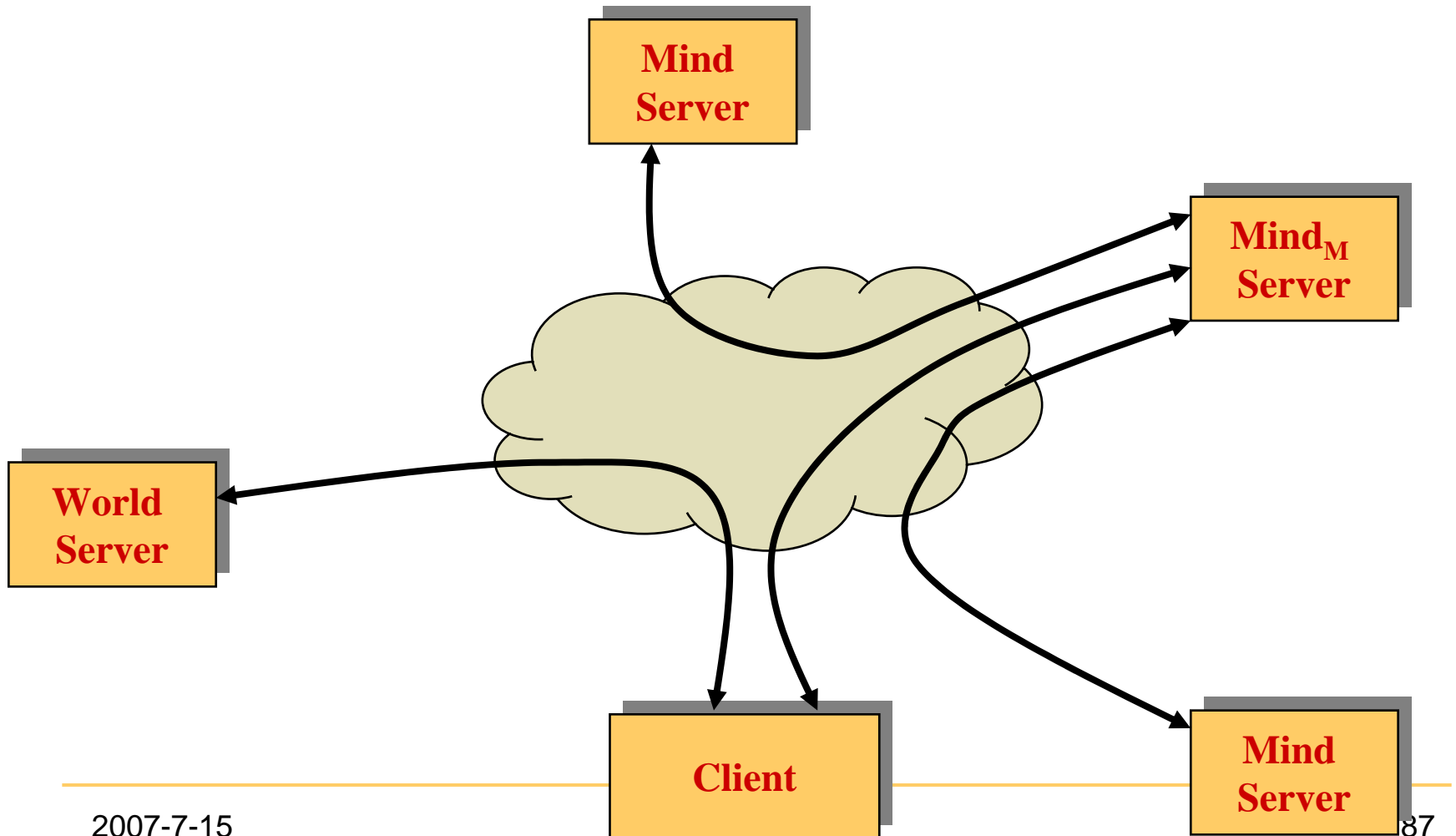


# JLC's Mind Online



# 构建复杂心智联机系统

Constructing complex minds, online



# Service Methods

---

## World Service

- newrun
- getstate
- takeaction
- endrun

## Mind Service

- newrun
- getaction
- endrun



# World-Wide-Mind Protocol

---

- “The term **Web services** describes a **standardized way of integrating Web-based applications** using the XML, SOAP, WSDL and UDDI open standards over an Internet protocol backbone. XML is used to tag the data, SOAP is used to transfer the data, WSDL is used for describing the services available and UDDI is used for listing what services are available”
  - <http://www.webopaedia.com>
- Society of Mind Markup Language (SOML)

# 技术要求

---

- Web Server space
- One of...
  - Java Servlets
  - ASP
  - JSP
  - PHP
  - CGI – C++/Perl/any language
- To write own client (or to integrate another mind), need to know how to connect to URL
  - Simple code in most languages

# Intelligence Is *In* the I/O- Winston

## The Explanation

**Linguistic  
Reasoner**

**Motor  
Reasoner**

**Visual  
Reasoner**

# 霍金斯模型

---

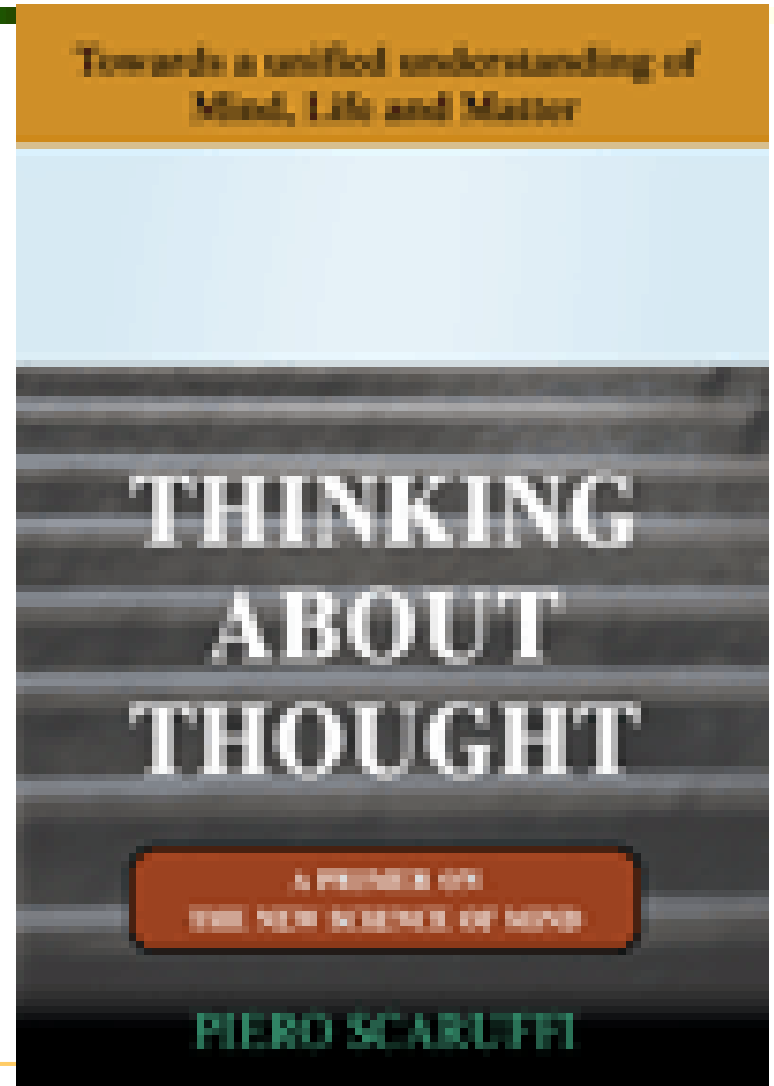
2004年的10月，霍金斯(Jeff Hawkins)发表了《On Intelligence》一书，提出了对大脑新的理解。

霍金斯提出智能是大量群集的神经元涌现的行为，用基于记忆的世界模式产生连续不断的对未来事件的一系列预测。时间是大脑做什么和怎样做最至关重要部分。大脑的行为有三个至关重要的组成：大脑工作于输入的时间序列流，神经网络的许多反馈，真实网络的层次结构所具有的模式，对它们的功能十分重要。

# Piero Scaruffi



- Latest: “Thinking About Thought” (2003)
- [www.scaruffi.com](http://www.scaruffi.com)
- [www.thymos.com](http://www.thymos.com)
- E-mail: [editor@scaruffi.com](mailto:editor@scaruffi.com)



# Piero Scaruffi



- **Mind = cognition + consciousness**
- **Cognition is ubiquitous in nature**
- **The mind translates a world of particles and waves into a world of colors, sounds and smells**

Towards a unified understanding of  
Mind, Life and Matter

## THINKING ABOUT THOUGHT

A PRIMER ON  
THE NEW SCIENCE OF MIND

PIERO SCARUFFI

# 意识和心智

- **What is the mind?**
  - The set of all brain processes
  - The set of all conscious processes
  - Something in between
- **Mental faculties: the mechanism that allows a piece of living matter to remember or learn something (in the sense of being able to perform future actions based on it) + the awareness of remembering or learning something**
- **Mental properties = properties of matter + properties of consciousness**
- **Mind = cognition + consciousness**

# 意识

---

## 意识系统

意识的原理是21世纪科学面临的一大难题，也是心理学历来关注的问题。近年来，国际学术界从事意识研究的自然科学人员队伍迅速扩大，如何用自然科学的方法研究意识的核心问题，是目前值得探讨的



# 法伯的意识观

三个层次讨论了意识概念

第一个层次是意识觉知。包括：感觉觉知、概括性觉知、元认知觉知和有意识回忆（能觉察到过去发生的事情）等四种。

第二个层次是高级能力，即不仅能被动地感知和觉知信息，还具有能动作用或控制等高级功能，这些功能包括注意、推理和自我控制（如理性或道德观念对生理冲动的抑制作用）。

第三个层次是意识状态，可理解为一个正在进行的心理活动，包括了意识概念中最常识性的、也是最困难的环节，这种状态可以分为不同的层次：有意识与非意识、综合性调节、粗略的感觉等。

# 克里克还原论

---

诺贝尔奖获得者，DNA双螺旋结构的提出者克里克(F. Crick)是这方面的典型代表之一。他认为意识问题是整个神经系统高级功能中的关键问题，所以他于1994年出版了一本高级科普书，名为“The Astonishing Hypothesis”(惊人的假设)，副标题为“用科学方法探索灵魂”[95]。他大胆地提出了一个基于“还原论”的“惊人的假说”。他认为“人的精神活动完全由神经细胞、胶质细胞的行为和构成及影响它们的原子、离子和分子的性质所决定”。

# 克里克还原论

---

认为意识问题与短时记忆和注意的转移有关，他还认为意识问题虽然牵涉到人的许多感觉，但他想从视觉意识着手，因为人是视觉性动物，视觉注意容易进行心理物理实验，而且神经科学在视觉系统研究方面积累了许多资料。20世纪80年代末90年代初在视觉生理研究方面有一个重大的发现：从不同的神经元的发放中记录到同步振荡现象，这种大约40Hz的同步振荡现象被认为是联系不同图像特征之间的神经讯号。克里克(F. Crick)等提出视觉注意的40Hz振荡的模型。并推测神经元的40Hz同步振荡可能是视觉中不同特征进行“捆绑”的一种形式。

# 剧场假设

巴尔 (Baar B J) 是“剧场隐喻”的最主要的继承和发扬光大者。1984年克里克 (Crick) 借用它在选择性注意方面提出“探照灯”假设，这个假设具体指丘脑-皮层相互作用，在选择性注意中的功能。由于大脑皮层分化为视、听、体感、运动、前额叶等区域，而丘脑也有相应的神经核团与之相对应，因此“探照灯”在它们的相互关系中起作用。在这个假设中的“观众”是指无意识的脑区，像部分皮层、海马、基底核、杏仁核以及运动执行系统和解释系统。

意识的“剧场假设”隐含着，在舞台上同时有许多角色在演出，正像人脑同时接受内外感受器的多种刺激，但是只有少量角色接收聚光灯的照射，这中间有个选择问题，而且聚光灯不是停留在一个地方、一个角色身上，而是随着时间流动，观众代表着脑的无意识的部分。

# 意识的主动模式和感知模式

---

1977年奥恩斯泰(Ornstein)提出意识存在的两个模式：主动—言语—理性模式(active-verbal-rational mode)与感知—空间—直觉—整体模式(receptive-spatial-intuitive-holistic mode)，分别简称为主动模式和感知模式。他认为两个模式分别被一侧大脑半球所控制，对主动模式的评价是自动进行的，人类限制了觉知的自动化以阻挡与其生存能力无直接相关的经验、事件和刺激。当人们需要加强正在进行的归纳与判断时，通过感知模式增加了正常的觉知。

# 微管假说

微管假说以朋罗塞 (Penrose) 最为著名。他是当代著名的数理科学家，曾提出一种复杂的几何图形可以完全覆盖全平面而不留空隙，而且在黑洞和引力方面做出创造性的工作。他在计算机、思维和意识方面，先后发表了2本著作。他的观点认为：“意识起源于神经元中特殊的蛋白质结构(微管)的量子物理过程”。Penrose认为神经元的细胞组织中细胞骨架在传递信息上起重要作用，细胞骨架是由“微管”构成，由于其空间尺度很小，应当用量子力学来考虑，电磁波有可能在其中传播。

# 量子意识观

由于量子特性与意识特性都有超越常规计算局限性的共同性，可以期待通过量子计算方法去解决意识的自明性问题。换句话说讲，可以利用量子迭加性、纠缠性、不确定性等特性，具体给出能够实现意识计算的量子算法。

在量子力学与意识之间，也许存在10来个组构层次：化学键、分子及其自组织、分子生物学、遗传学、生物化学、膜及其离子通道、突触及其神经递质、神经元本身、神经回路、皮层柱和模块、大规模皮层的动态活动等等。因为在相邻层次上工作的神经科学家之间有着强烈的竞争，因此在神经科学研究中人们总是意识到这些层次的。

# 建构理论

哲学家兼计算机专家布克斯(A. W. Burks)提出的计算机建构理论[71]是这样给意识下定义的：

(1)意识要有意向(**intention**)，即能选择一个目标，有方向性地进行搜索。

(2)意识要有直接经验，如能感觉到疼痛。

(3)意识要有改正和修复的功能。当计算机的局部系统出现故障时，它能自动地转到另一系统，或自动设计新的程序，以达到自身的完善化。

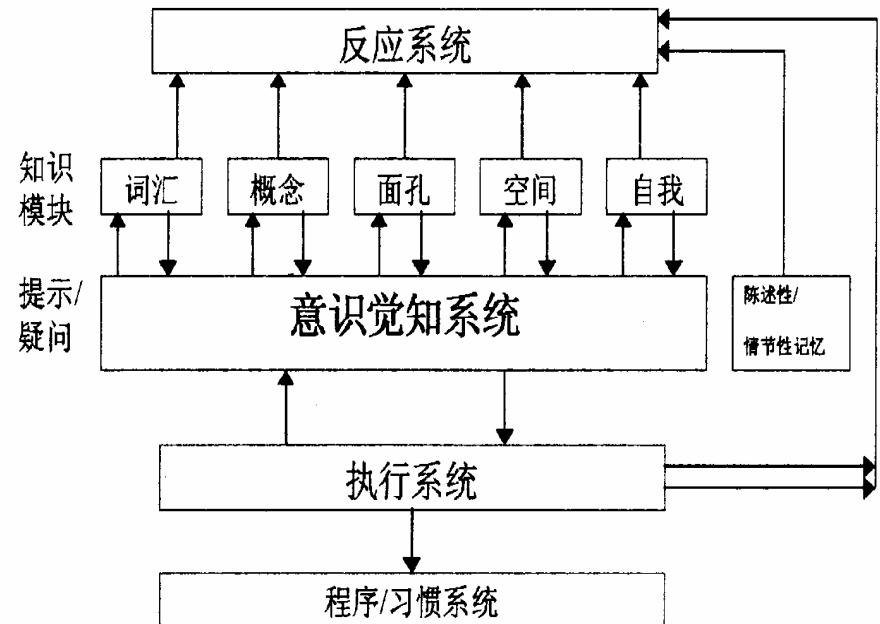
(4)意识要有觉醒功能。人在睡眠时相当于系统的不活动期，他在觉醒时才有自觉意识。同样，计算机具有保护系统，在其不活动期，也能处于工作状态。

(5)意识可以进行短时控制和制定长时计划。从建构角度来看，意识是一种特定的计算机控制系统，是一种相对简单的实时控制机构，当系统处于觉醒状态下，它能指导短时活动并能实施长时计划。

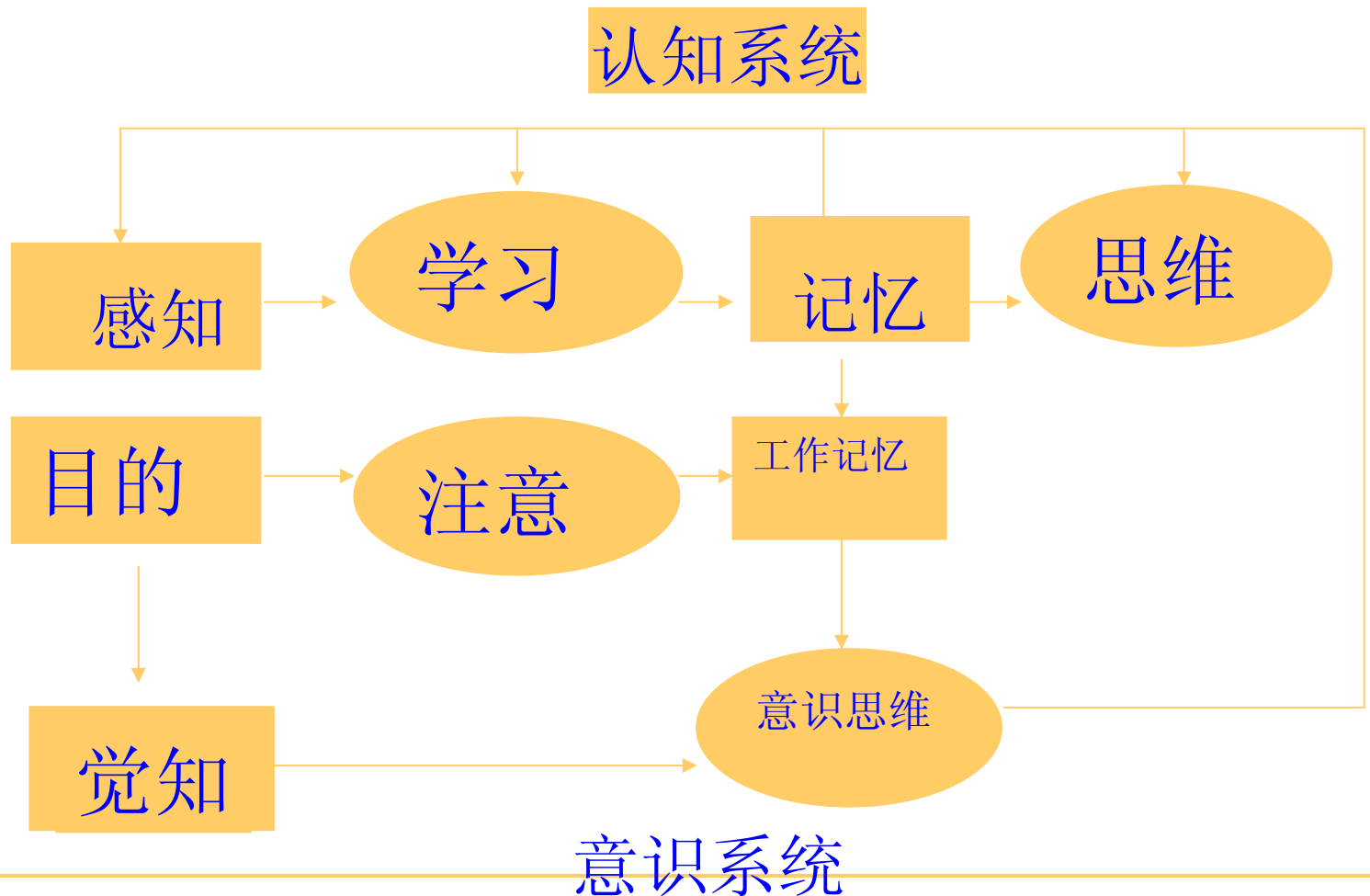


# 意识的方块模型

1988年萨克特(D.L. Schacter)等提出了意识的方块模型。萨克特(D.L. Schacter)没有给意识觉知系统(Conscious Awareness System, CAS)具体定位,但他认为确实存在这样的意识觉知系统,其活动是意识体验所必需的,是与知觉、认知和行动有关的脑区相分离的。



# 意识模型



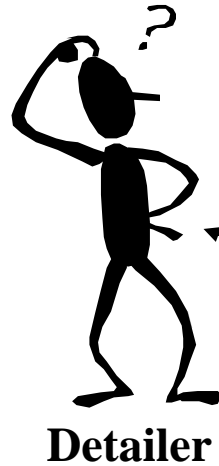
# Why a 'Conscious' Agent?

---

- Flesh out the theory with detailed architecture and mechanisms
- Hypotheses for cognitive scientists and neuroscientists
- Produce flexible, adaptive, human-like software
- Want smart agents?  
Model them after humans.

# IDA: an Intelligent Distribution Agent

Read personnel data  
Check job requisition list  
Adhere to Navy policies  
Choose jobs to offer members  
Negotiate with members  
Write orders



Telephone



Internet

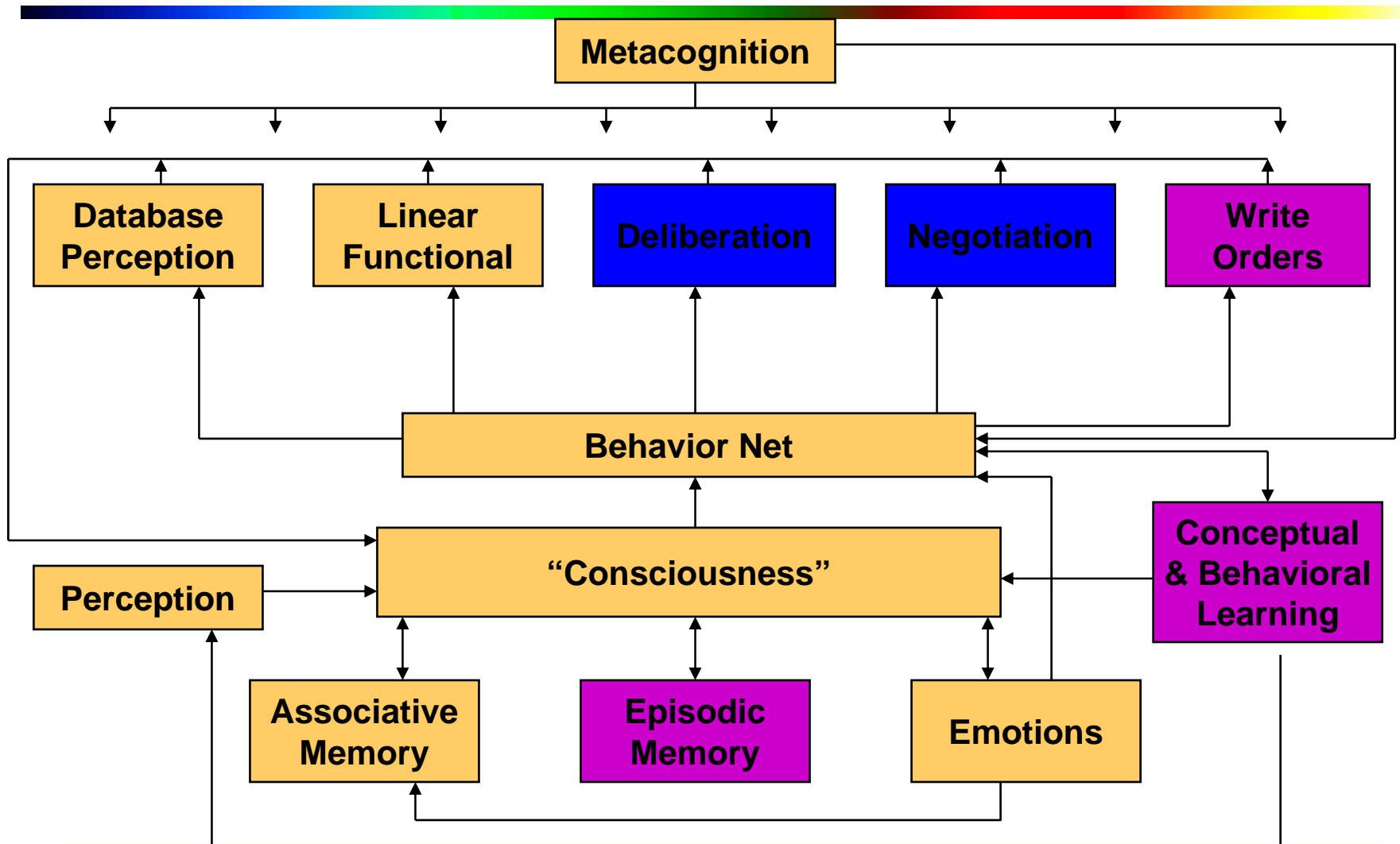


# Modules and Mechanisms

---

- **Perception—Copycat Architecture—Hofstadter**
- **Action Selection—Behavior Net—Maes**
- **Associative Memory—Sparse Distributed Memory—Kanerva**
- **Episodic Memory—Case-based Memory**
- **Emotions—Pandemonium Theory—Jackson**
- **Metacognition—Fuzzy Classifier Systems—Holland, Zadeh**
- **Learning—Copycat Architecture, Case-based Reasoning**
- **Constraint Satisfaction—Linear Functional**
- **Language Generation—Pandemonium Theory**
- **Deliberation—Pandemonium Theory**
- **“Consciousness” —Pandemonium Theory**

# IDA's Architecture



# Levels of abstraction



- High level
  - behaviors
  - message type nodes
  - emotions
  - metacognitive actions
  - etc.
- Low level
  - codelets

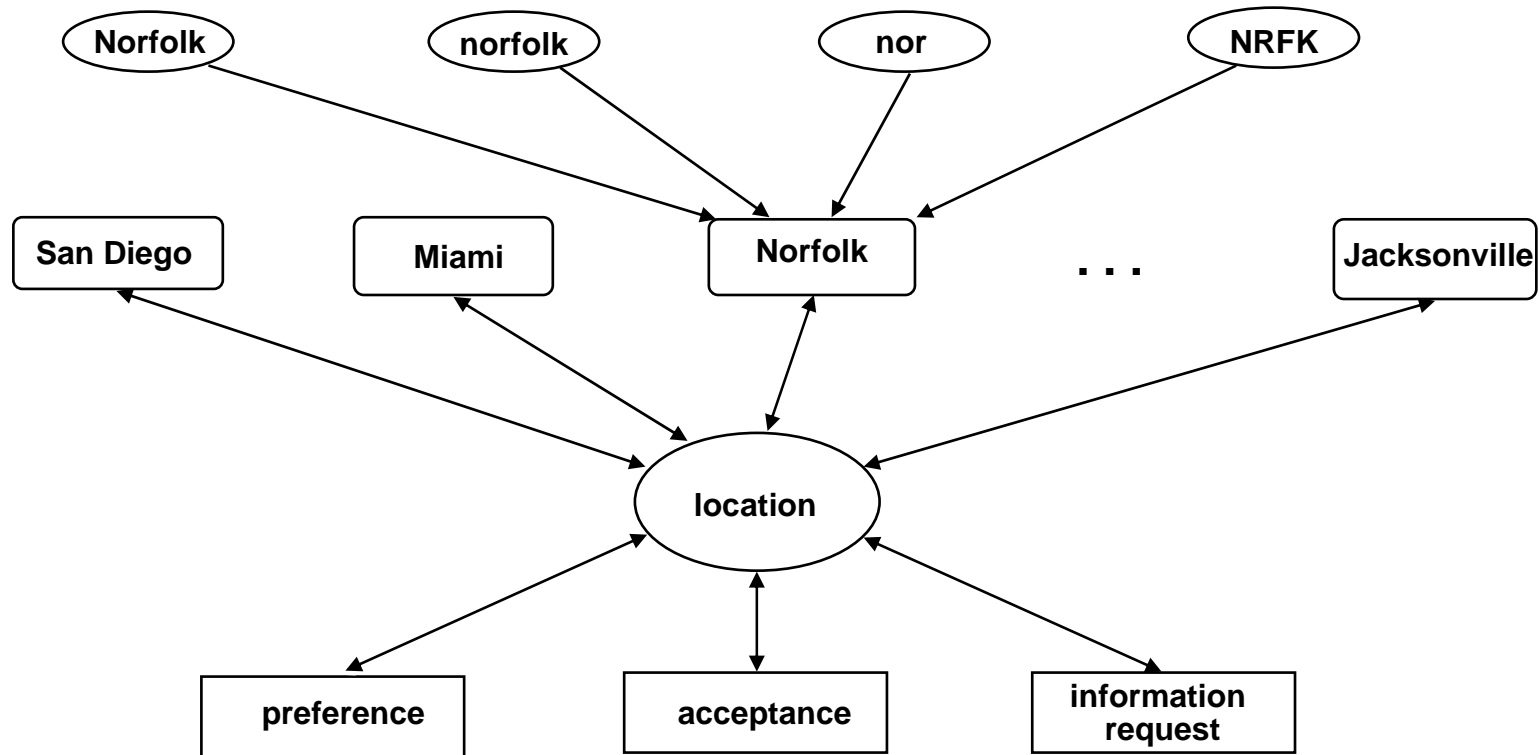
# Codelets

---

- Small pieces of code each performing a simple, specialized task
- Acts as a demon, always watching for a chance to act
- Most subserve some high level entity, e.g.
  - behavior
  - slipnet node
  - metacognitive action
- Some codelets work on their own, e.g.
  - watching for incoming mail
  - checking for time and place conflicts
- Codelets do almost all the work
- IDA is a multi-agent system



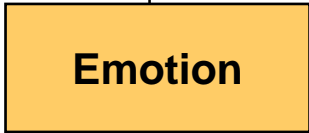
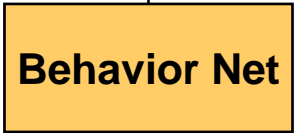
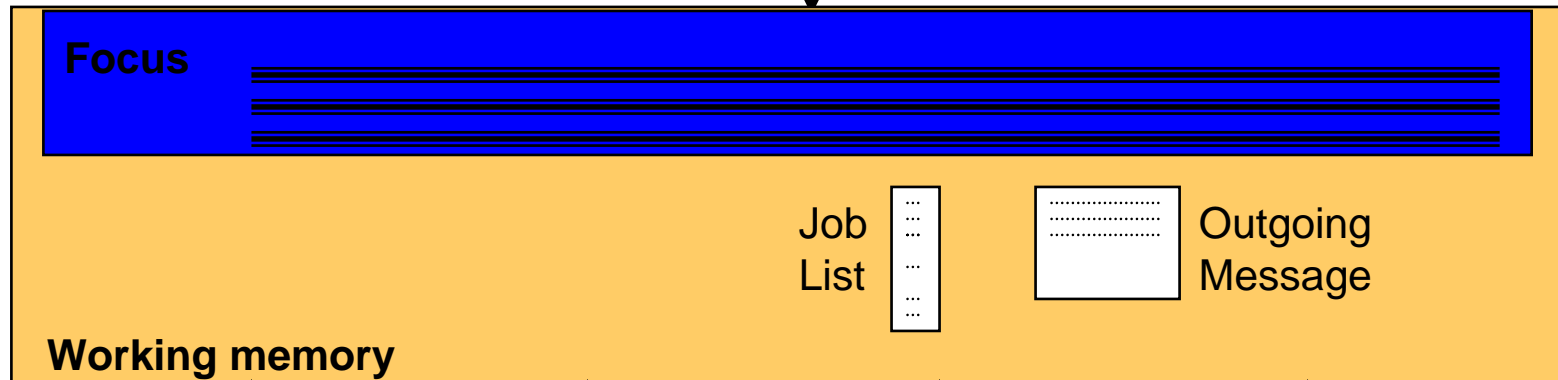
# Perception via a Slipnet



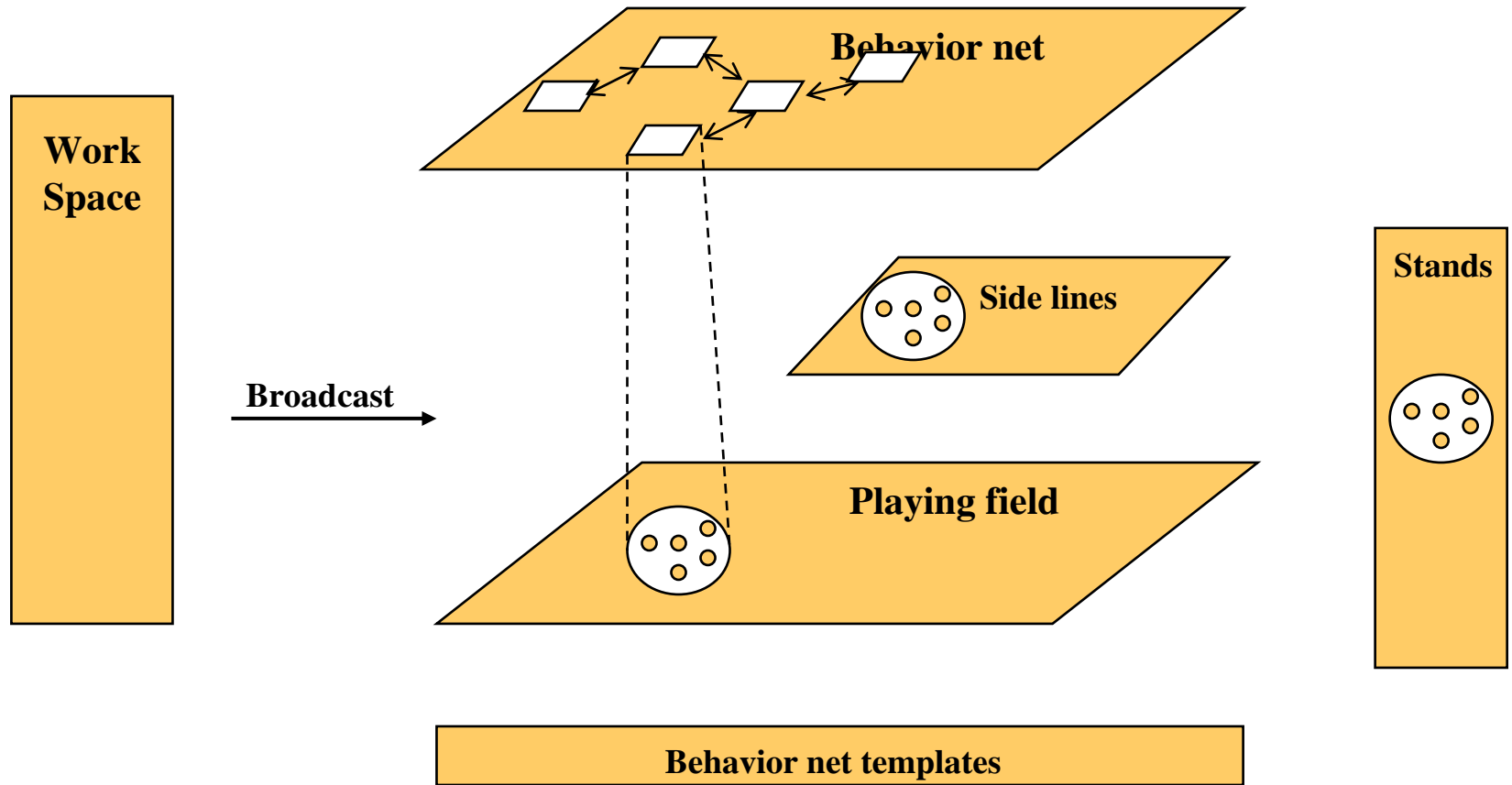
# Associative Memory

**Sparse Distribute Memory — Boolean Space — dim = N (enough to code features)**

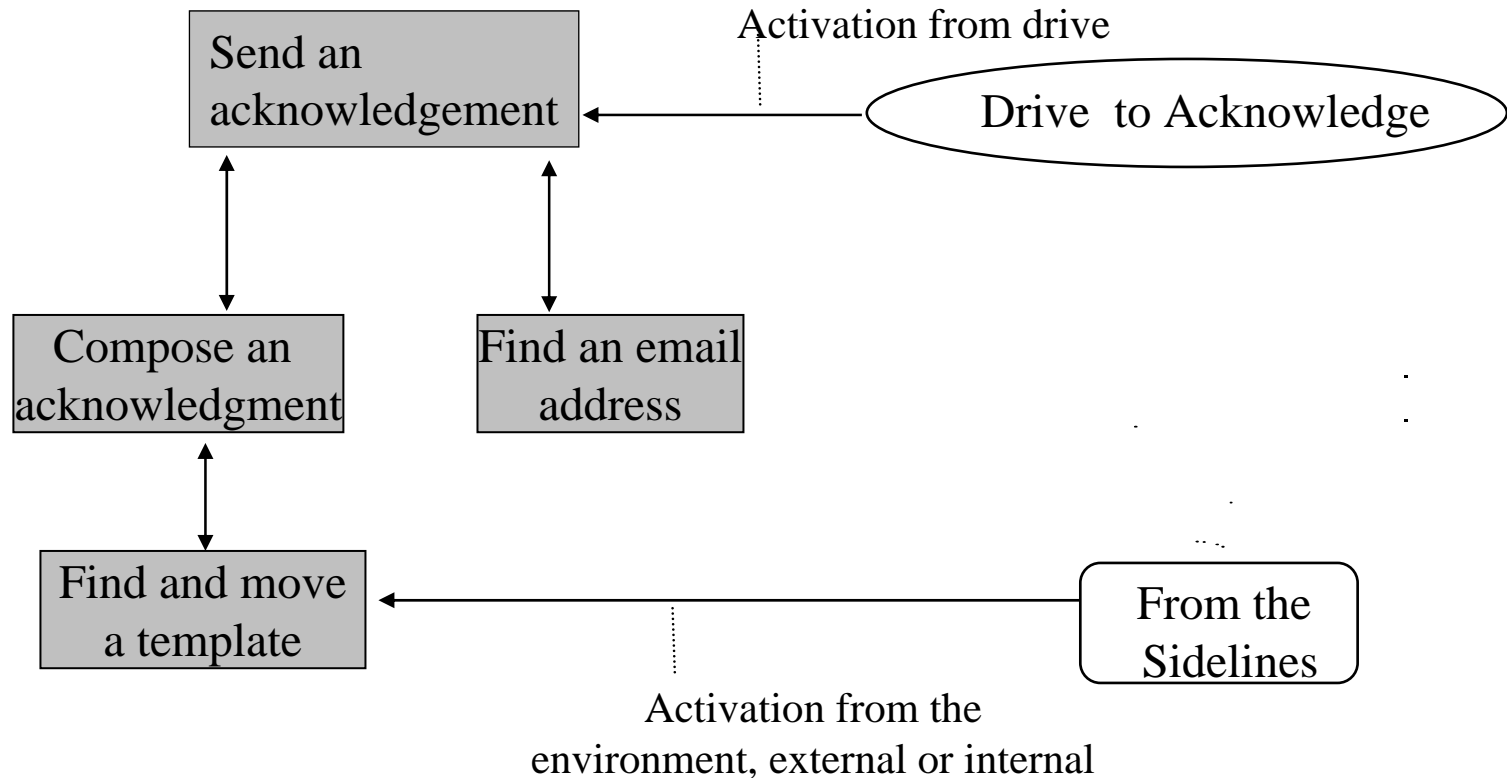
bit vector



# Behavior Net in Action



# A Behavior Stream



# Deliberation

---

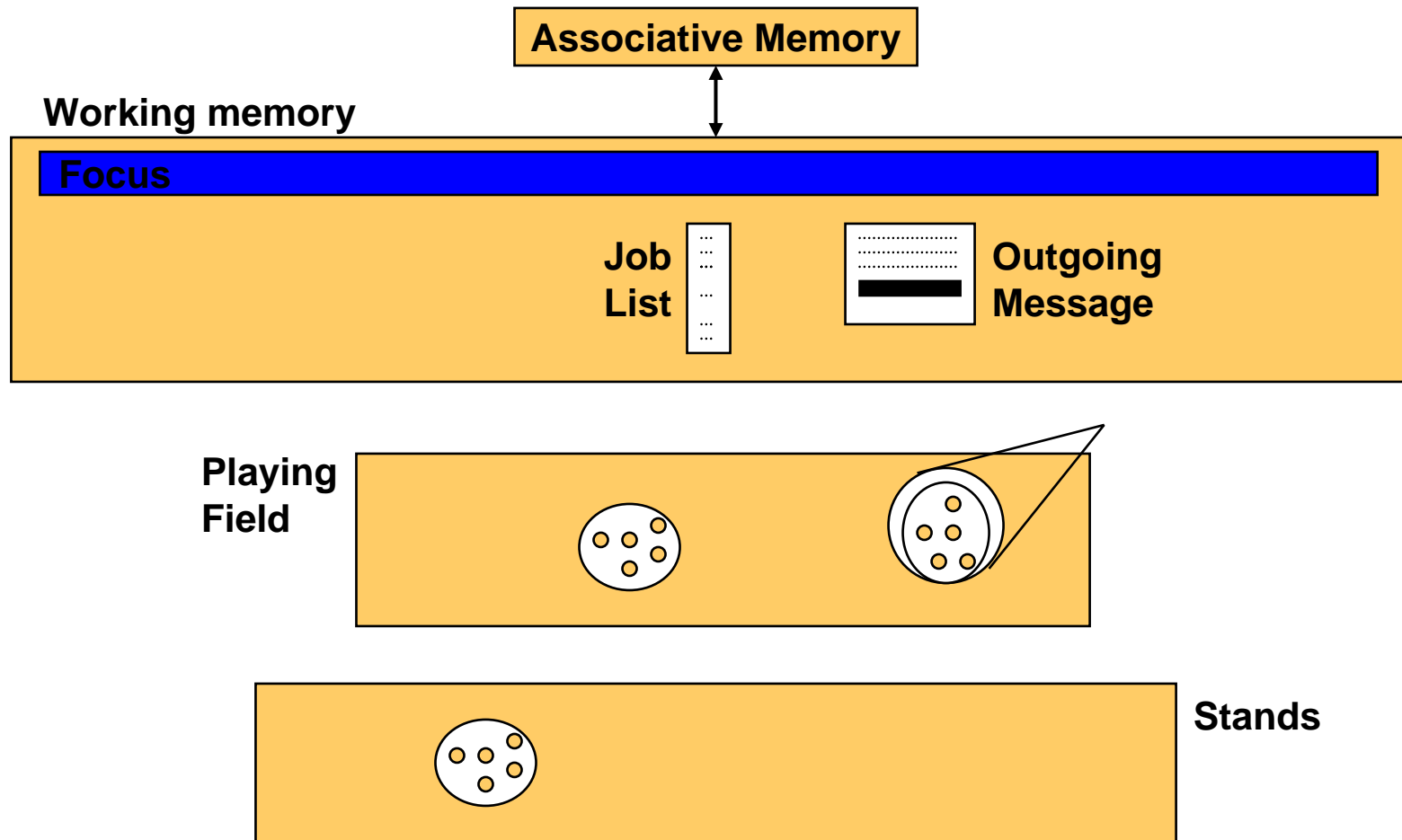
- Faced with a goal or problem
- Imagine possible plans or solutions
  - Scenarios
  - Routes
  - Internal virtual reality—Dawkins
- Evaluate them
  - Using reason
  - Using emotions
- Choose among them

# IDA's Deliberation

---

- Create scenes
  - May require objects, actors, concepts, relations, frames
  - Organized around events
- Build scenarios as sequences of scenes
- Choose between scenarios, discarding some
- Using Barsalou's perceptual symbol systems as a guide

# “Consciousness” in Action



# Metacognition

---

- Thinking about thinking
- What Sloman calls meta-management
- Influences action selection strategies
  - More or less opportunistic, thoughtful or goal-oriented
- Influences resource allocation
- Fuzzy classifier system



# Learning

---

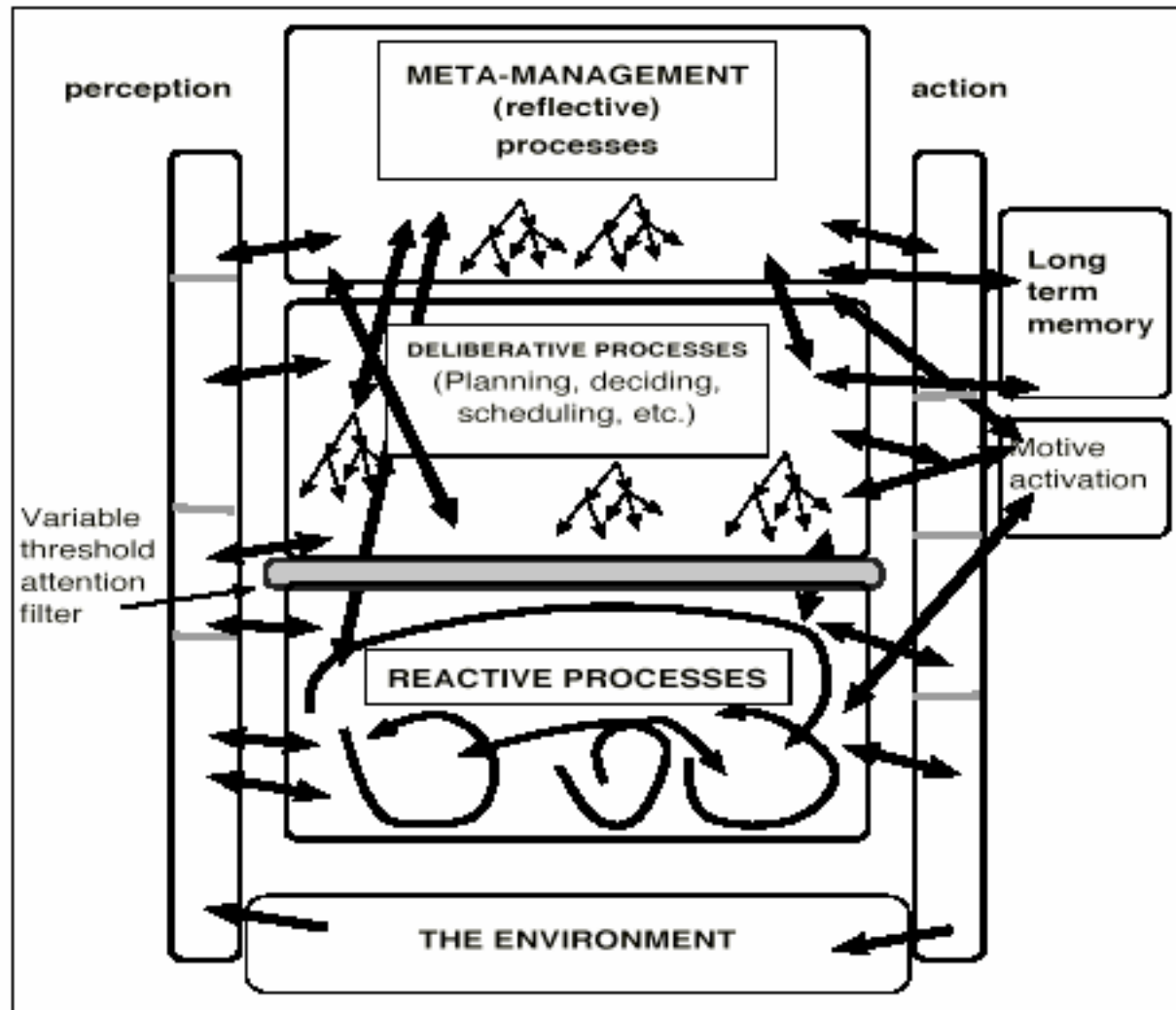
- Associative learning via pandemonium theory
- Associative learning via sparse distributed memory
- Perceptual learning via case-based reasoning
- Behavioral learning via case-based reasoning
- Metacognitive learning via classifiers

# Modeling Cognition



- Situated (embodied) cognition—Varela, Thompson & Roach
- Perceptual symbol systems—Barsalou
- Memory via actions—Glenberg
- Global workspace theory—Baars
- Cognitive architecture—Sloman

# Sloman's Architecture



# 结束语

---

心智模型的研究目标，就是要揭开人类心智的奥秘，它的研究不仅能够促进人工智能的发展，揭示生命的本质和意义，在促进现代科学特别是心理学、生理学、语言学、逻辑学、认知科学、脑科学、数学、计算机科学甚至哲学等众多学科的发展上，都有非同寻常的意义。

# Web and Email Addresses

---

- **Stan Franklin**    [stan.franklin@memphis.edu](mailto:stan.franklin@memphis.edu)—  
[www.msci.memphis.edu/~franklin](http://www.msci.memphis.edu/~franklin)
- **“Conscious” Software Research Group**  
[www.msci.memphis.edu/~csrg](http://www.msci.memphis.edu/~csrg)
- **CMattie Project**—[www.msci.memphis.edu/~cmattie](http://www.msci.memphis.edu/~cmattie)
- **Ciarán O’Leary**。 Technology for automated assessment:  
The World-Wide-Mind. Dublin Institute of Technology  
22nd May 2003
- **IDA Project**—[www.msci.memphis.edu/~ida](http://www.msci.memphis.edu/~ida)

# Thank You

---

Question!

Intelligence Science

<http://www.intsci.ac.cn/>

