

认知基础

Cognitive Foundation

第十二章

意识

Consciousness

史忠植

中国科学院计算技术研究所
<http://www.intsci.ac.cn/>

内容提要

- 意识的概念
- 意识的研究史
- 意识理论概述
- 全局工作空间理论
- 整合信息理论
- 机器意识系统
- 显意识思维与潜意识思维

意识

意识的起源与本质是最重大的科学问题之一。在智能科学中，意识问题具有特别的挑战意义。存在如何决定意识，客观世界如何反映到主观世界中去，既是哲学研究的主题，也是当代自然科学研究的重要课题。意识涉及知觉、注意、记忆、表征、思维、语言等高级认知过程，其核心是觉知(awareness)。近年来，由于认知科学、神经科学和计算机科学的发展，特别是新的无损伤性实验技术的出现，意识的研究再度被提到日程上来，并且开始成为众多学科共同研究的热点。在21世纪，意识问题将是智能科学力图攻克的堡垒之一。

意识

- 当代著名思想家丹尼特（D. C. Dennett）认为，“人类的意识大概是最后一个难解的谜。……对意识，我们至今如坠五里云雾中，时至今日，意识是唯一常常使最睿智的思想家张口结舌、思维混乱的论题。”在人类已建立的众多概念与范畴体系中，意识是最为混乱的概念之一。

意识的理解

在威廉·卡尔文的《大脑如何思维》一书中，对意识的八种不同理解：

- 1、是一种自我感觉，例如处于内疚状态；
- 2、是能思想、有意愿或有感知觉的标志；
- 3、能感知、能理解或是在一定程度上能调控自己的思维；
- 4、在睡眠或昏迷状态下受到抑制的一种心智机能；
- 5、意识表现为经过深思熟虑以后的措施或行为；
- 6、意识的含义与“注意”或“考虑”相近；
- 7、意识表现为对某种事物的关注或兴趣；
- 8、意识以强烈的感情色彩或明确的想法为标志。

希尔斯特的意识定义

威廉·希尔斯特 (William Hirst) 在其颇有影响的论文《意识的认知层面》中，把意识定义为：“人们对心理对象（如知觉、表象、或感受）的觉知”。希尔斯特在用觉知这个词时，其含义包括能够用言语报告这种觉知。他认为“人们不仅能觉知到感觉、表象和感受，同时也能觉知到他们正在体验到它。他们知道他们自己正在看、正在想象和正在感受。任何意识活动都包括对外部世界和心理表象的觉知，还包括自我意识。”

斯坦福哲学百科全书

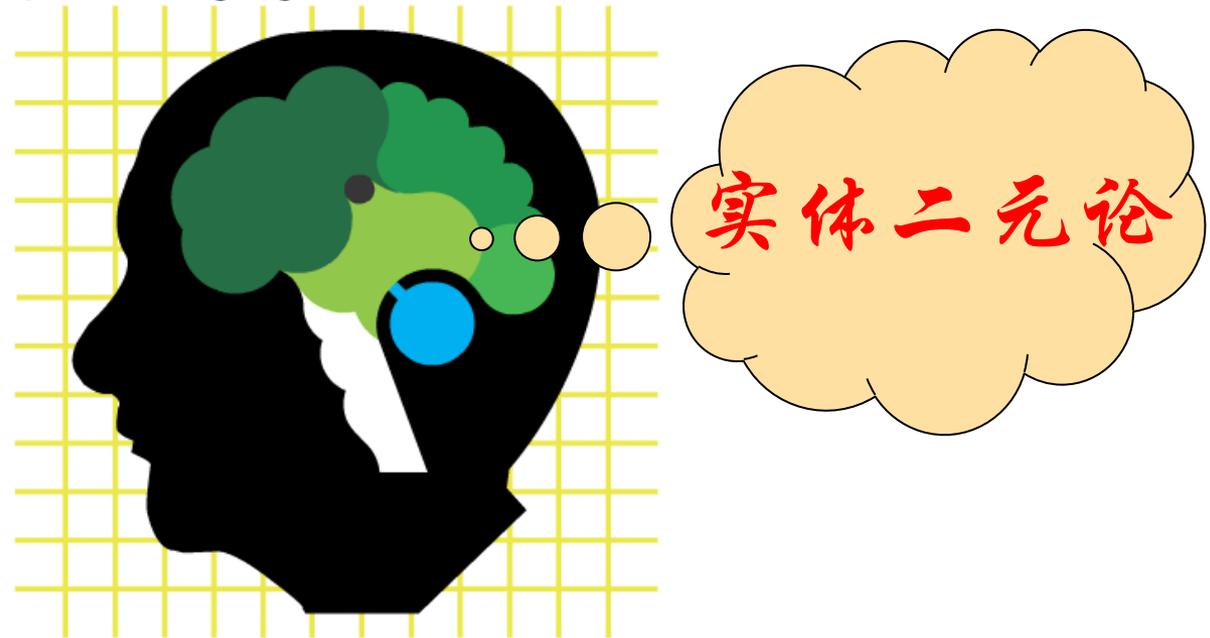
Stanford Encyclopedia Of Philosophy

- 《斯坦福哲学百科全书》是一部由斯坦福大学营运的免费在线哲学百科全书，内容主要以经同行评审认可的论文为主。该百科内的每一篇论文均由一位该领域的专家撰写并维护，作者涵盖来自世界各地学术机构的教授。
- 该百科于1995年由斯坦福大学语言资讯研究中心（英语：Center for the Study of Language and Information）资深研究员爱德华·萨尔塔（英语：Edward N. Zalta）创办，其运作目标为提供一部时常更新且不会像传统的纸本百科全书一样过时的互动式百科全书。

<http://www.informationisbeautiful.net/play/what-is-consciousness/>

实体二元论

实体二元论认为多样性世界有两个不分先后、彼此独立、平行存在和发展的本原的哲学学说。



A field that exists in its own parallel “realm”
of existence outside reality so can’t be seen
(*Substance Dualism*)

<http://www.informationisbeautiful.net/play/what-is-consciousness/>

笛卡尔的实体二元论

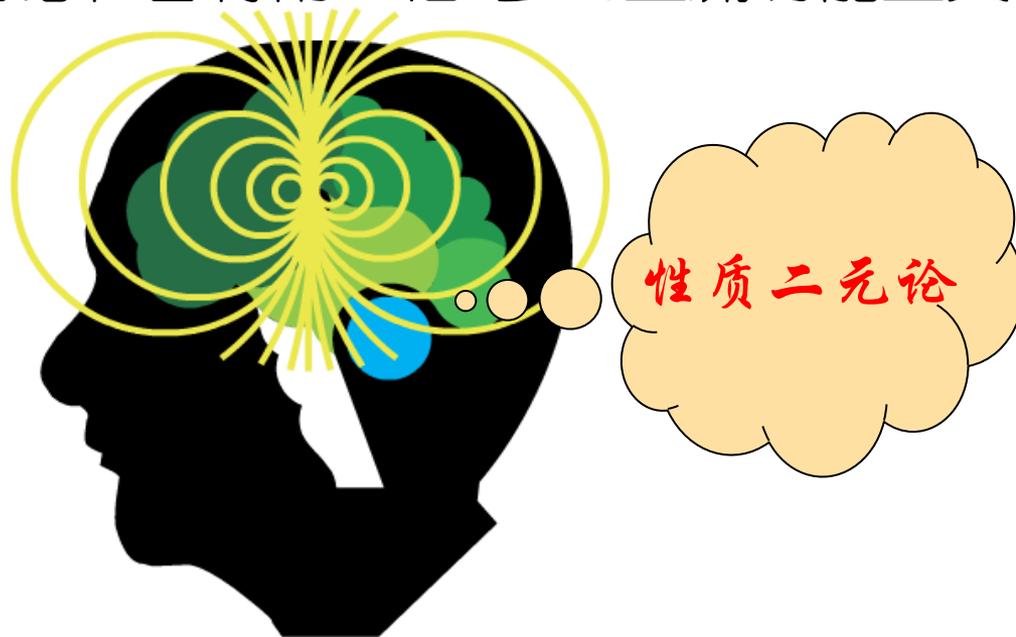
Descartes' dualism



- 哲学史上典型的二元论者是17世纪法国哲学家R.笛卡尔。他以其二元论的形式比较充分地揭示了心智(Mind)与身体(Body)的关系，他把心身严格地区分为两个不同的实体。二元论割裂了物质和精神的关系，不能科学地解决世界的本质问题。
- 笛卡尔怀疑集中体现在其认识论之核心命题“我思故我在”上。奠定了主题性原则和理性主义的基本特征；同时也使之始终难以摆脱二元论的限制。

性质二元论

作为一种哲学本体论, 性质二元论声称所有实体都是物理实体。戴维森的非则一元论和普特南-福多的主流功能主义都属于性质二元论



A physical property of matter, like
electromagnetism, just not one we know about
(*Property Dualism*)

<http://www.informationisbeautiful.net/play/what-is-consciousness/>

物质主义与心理属性

Materialism and mental properties

- 物质可以有不同的性质
- 属性二元论：心理属性不是物理属性
- 类型同一理论：心理属性是物理属性



Hmm...



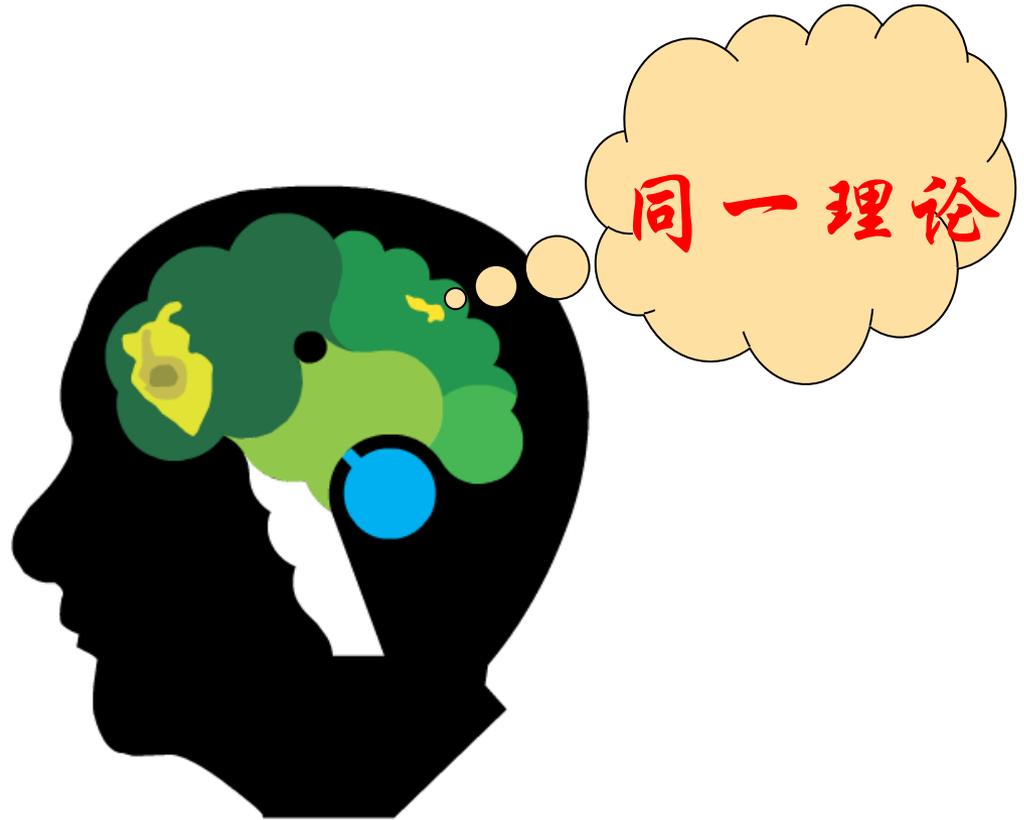
二元论的问题

- 如果头脑只是思想，而不是空间，物质只是在空间中的延伸，那么一个人怎么可能会对另一个产生因果关系？
- 所有的物理效应都有充分的物理原因。任何物理事件都不需要一个非物理解释。
- 心理因果会违反物理学的规律，例如能量守恒定律。

心身同一性与还原

Mind-body identity and reduction

- 如果心智=身体，那么心理因果关系并不违反物理定律，我们可以理解它是如何发生的（通常是物理方式）。
- 本体还原：一个领域中的事物（如精神事物）与另一个领域中的某些事物是相同的。

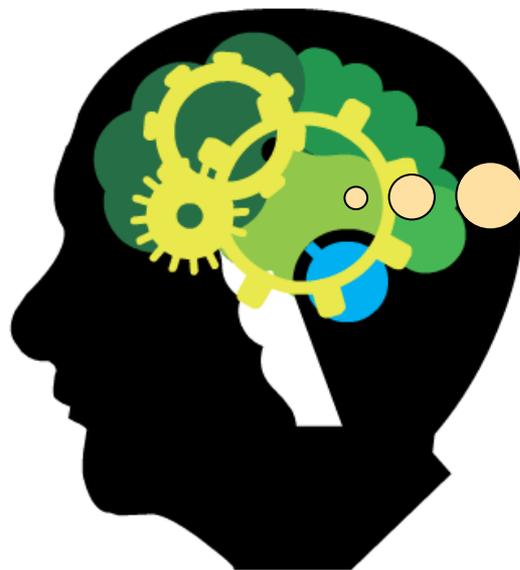


Simply, mental states are physical events that we can see in brain scans
(Identity Theory)

<http://www.informationisbeautiful.net/play/what-is-consciousness/>

功能主义

对有机体来说，理解一个生物现象问题（行为或生理现象过程）的最佳方法是试图理解它的有用功能。



Consciousness and its states (belief, desire, pain) are simply functions the brain performs
(*Functionalism*)

<http://www.informationisbeautiful.net/play/what-is-consciousness/>

泛心理主义

意识从何而来？哲学家菲利普·戈夫在他的新书《伽利略的错误：新意识科学的基础》中提出了泛心理主义：如果意识不是大脑所特有的东西，而是所有物质所固有的一种品质。



All matter has a psychic part. Consciousness:
is just the psychic part of our brain
(Pan Psychism)

<http://www.informationisbeautiful.net/play/what-is-consciousness/>

涌现二元论

意识是由物理组分的复杂组织方式涌现产生的, 但涌现的产物超出了其物理组分本身, 并且这种涌现现象既不是先验可预测的, 也不能单纯从物理属性解释。

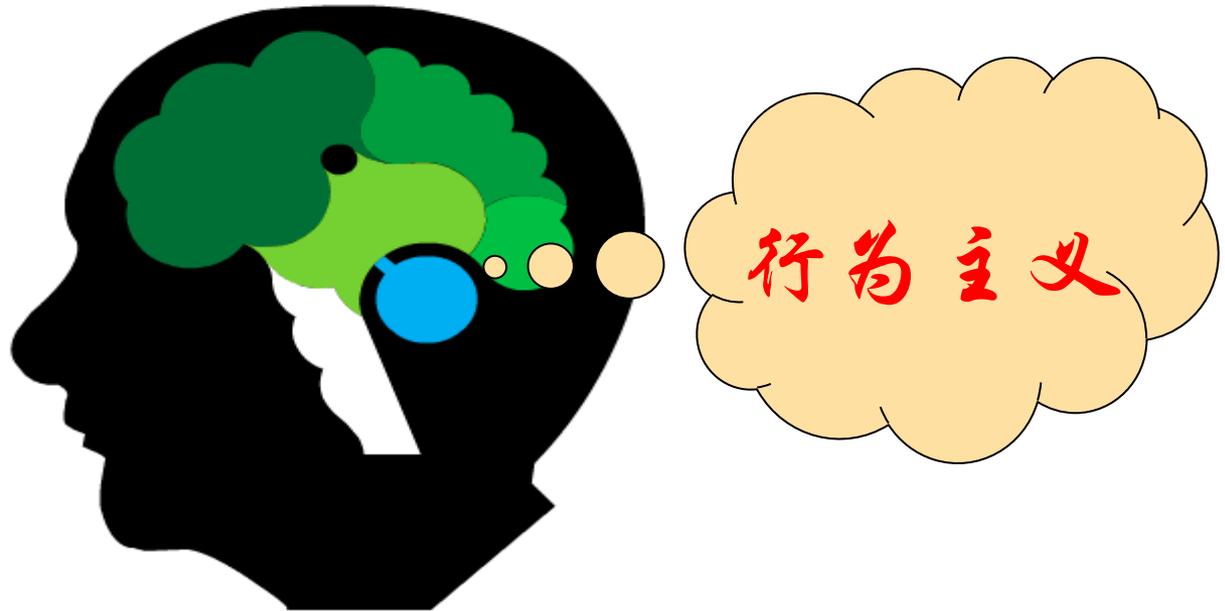


A sensation that “grows” inevitably out of
complicated brain states
(*Emergent Dualism*)

<http://www.informationisbeautiful.net/play/what-is-consciousness/>

行为主义的意识观

行为主义的主要观点是认为心理学不应该研究意识, 只应该研究行为, 把行为与意识完全对立起来。

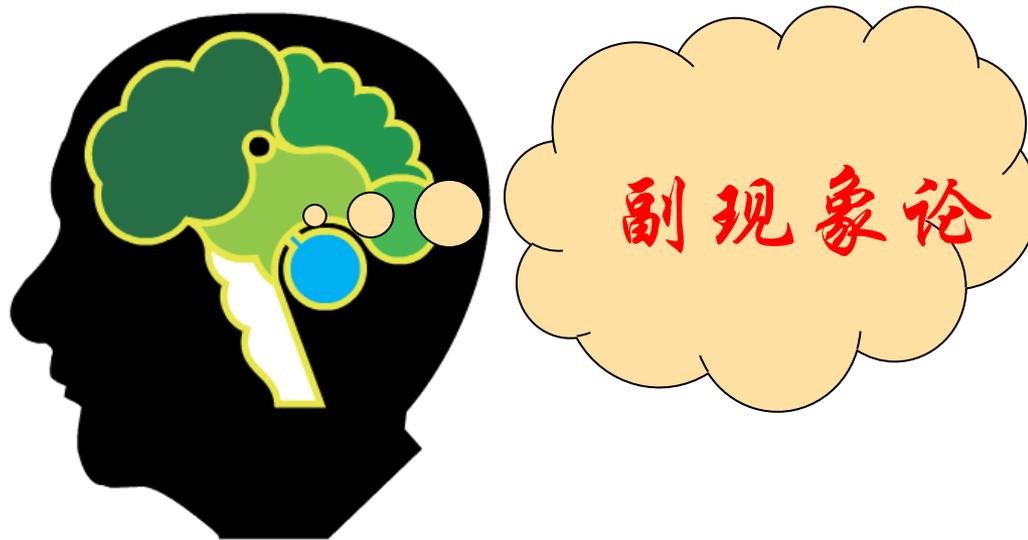


Literally just behaviour. When we behave
in a certain way, we appear conscious
(*Behaviourism*)

<http://www.informationisbeautiful.net/play/what-is-consciousness/>

副现象论

17 世纪英国哲学心理学家霍布斯，从机械唯物主义出发，主张作为物体的身体是实在的，心灵事件或心理事件是身体机械过程的副产品。身心关系是单向的：心智事件

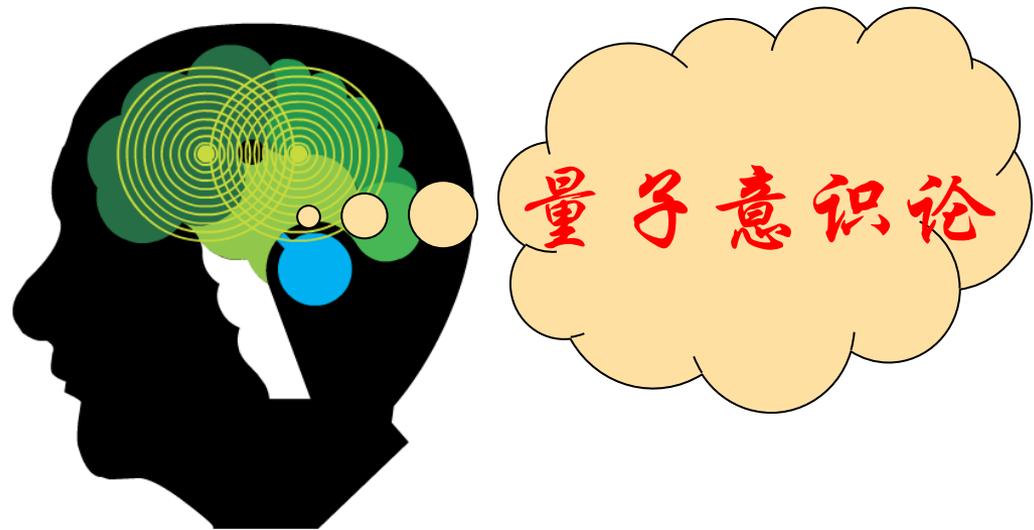


An accidental side-effect of complex
physical processes in the brain
(*Epiphenomenalism*)

<http://www.informationisbeautiful.net/play/what-is-consciousness/>

量子意识论

量子意识理论认为，经典力学无法完整解释意识，意识是一种量子力学现象，如量子纠缠和叠加作用。大脑中存在海量的处于量子纠缠态的电子，意识正是从这些电子的波函数的周期性坍塌中。

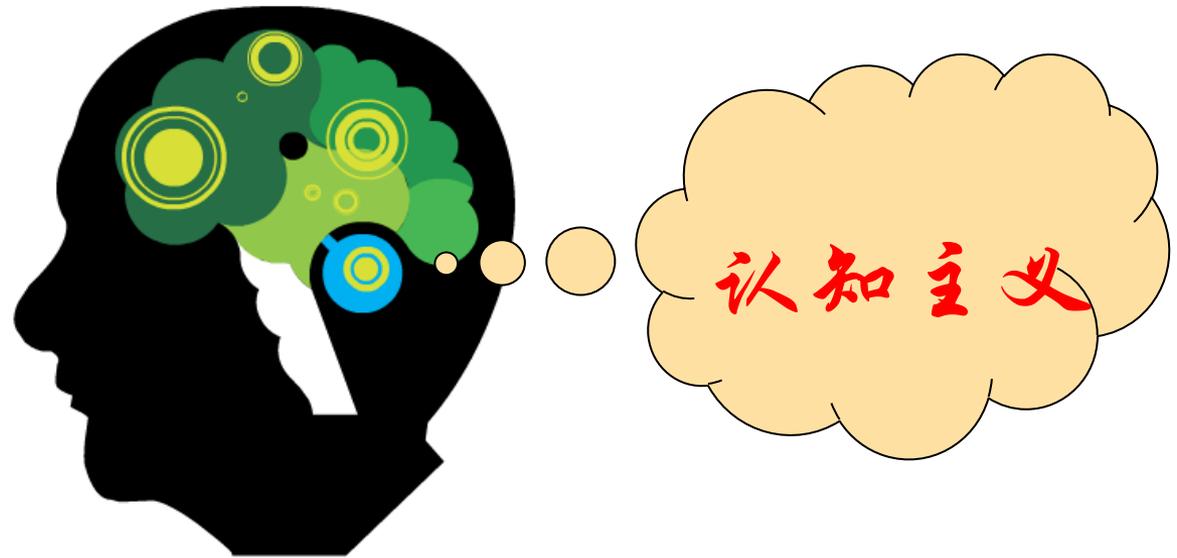


Not sure. But quantum physics, over classical physics, can better explain it
(Quantum Consciousness)

<http://www.informationisbeautiful.net/playing-with-ai-is-consciousness/>

认知主义意识观

许多意识理论侧重于特殊的认知结构或这种结构中的特殊活动模式，例如，巴尔斯（Bernard Baars, 1988）提出的全局工作空间理论。

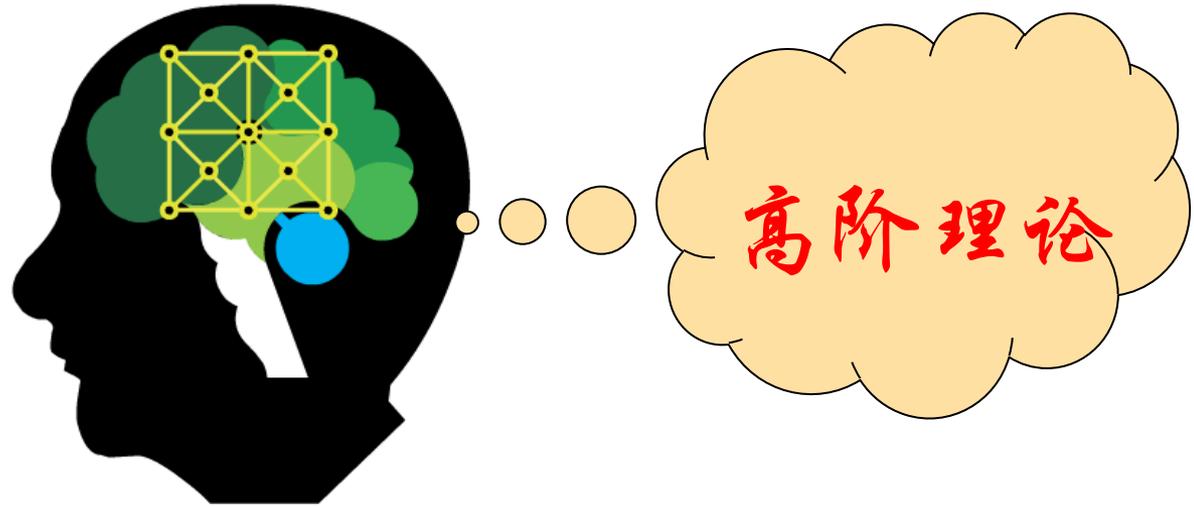


The sensation of your most significant thoughts being highlighted
(Cognitivism)

<http://www.informationisbeautiful.net/play/what-is-consciousness/>

意识的高阶理论

意识的高阶理论能够对有意识的心理状态作出更符合我们日常直觉的解释. 但是, 布洛克的最新论证表明彻底误表征情况的存在会使得高阶理论不融贯.

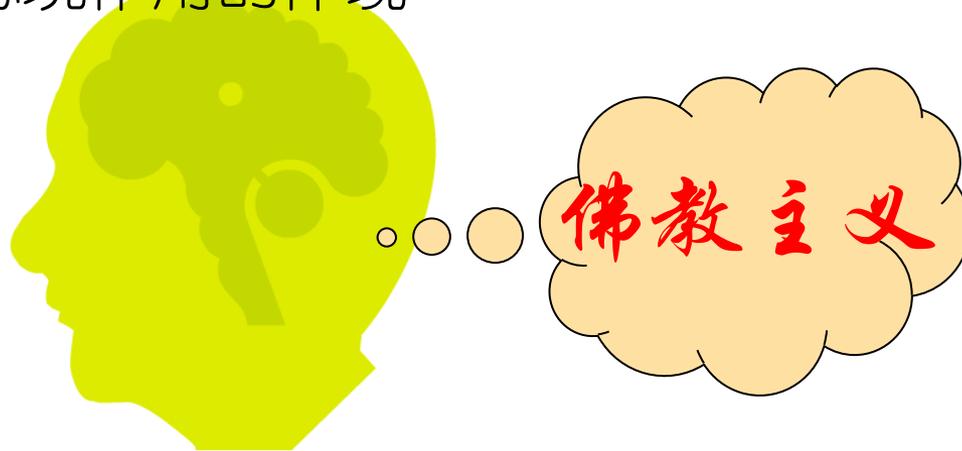


Consciousness is just higher order thoughts
(thoughts about other thoughts)
(Higher Order Theory)

<http://www.informationisbeautiful.net/play/what-is-consciousness/>

佛教主义的意识观

佛教有论意识问题的专著叫《唯识论》，包括《成唯识论》、《唯识二十论》等，认为意和识具有不同含义，“意”指人思量或思考事物；“识”指分辨、区别、明了事物；而意与识都是心的功能，是心境作用的体现



A continuous stream of ever-recurring
phenomena, pinched like eddies into
isolated minds
(Buddhism)

<http://www.informationisbeautiful.net/play/what-is-consciousness/>

意识的定义

意识是指大脑对认知、情感和意志等心理过程的觉知、调节或控制。

其中，认知心理过程具体包括注意、感知、记忆、想象、分析、综合、抽象、概括、判断、推理等心理操作过程。



内容提要

- 意识的概念
- 意识的研究史
- 意识理论概述
- 全局工作空间理论
- 整合信息理论
- 机器意识系统
- 显意识思维与潜意识思维

意识的研究史

- 历史上最早使用意识这个词的是培根 (Francis Bacon) 。他的定义是意识就是一个人对自己思想里发生了什么的认识。 所以，意识问题一直是哲学家研究的领域。
- 冯特 (Wundt WMM) (1832-1920) 创立了实验心理学，用内省法研究意识问题，一时意识问题成为心理学研究的主流。意识问题成为心理学的重要研究对象。
- 1902年加米斯 (James) 提出意识流的概念，指出意识就像流水一样波浪起伏，渊源不断。

意识的研究史

- 弗洛伊德（Sigmund Freud）认为人的感觉和行为受非意识需要、愿望和冲突的影响。根据弗洛伊德的观点，意识流具有深度，意识与非意识加工有不同的认识水平。它不是全或无的现象。
- 20世纪20年代，华生（Watson JB）在心理学中推行行为主义，当时“意识”问题而被打入冷宫。
- 1950年前后心理学与计算机发展互相结合，信息加工学派得到蓬勃发展。20世纪70年代“意识问题开始受到一些关注，到了90年代中期，逐渐形成一个多学科研究“意识”问题的高潮。

意识的研究史

- 1994年4月，在美国Arizona召开了第一次关于“意识”的会议(Toward a Science of Consciousness)，来自心理学、计算机科学、神经网络、机器人学等专业的代表参加。
- 1997年国际意识科学研究学会(Association for the Scientific Studing of Consciousness: ASSC)成立，召开意识问题国际学术会议。
- 2017年6月13至16日，第21届意识科学研究学会会议在北京召开。本次会议由中国科学院生物物理所脑与认知科学国家重点实验室、中国科学院心理所与中国认知科学学会共同承办。近400名国内外研究者参加了此次会议。
- 2000年，汪云九、杨玉芳等召开香山科学会议：意识与脑

克里克和科赫的意识框架

2003年，克里克（Crick）和科赫（Koch）列举意识研究中神经相关物（NCC）的十条框架：

(1) 无意识的侏儒 (homunculus)

脑整体的工作方式，大脑的前部注视着感觉系统，感觉系统的主要工作是在脑的后部进行的。而人并不直接知道他们的想法，而只知道意象中的感觉表象。这时，前脑的神经活动是无意识的。

(2) 刻板 (zombie) 模式和意识

对于感觉刺激，许多反应是快速的、瞬态的、刻板的和无意的，而意识处理的东西更慢、更广，且需要更多时间决定合适的想法和更好的反应。进化上发展出这两种策略以相互补充，视觉系统的背侧道路（大细胞系统）执行刻板的快速反应；腹侧系统（小细胞系统）执行的是有意识的识别任务。

克里克和科赫的意识框架

(3) 神经元联盟 (coalition)

此处联盟是Hebb的集群加上它们之间的竞争。“注意”机制相当于舆论界和选情预测者的作用，试图左右选举形势。皮质第V层上的大锥体细胞好像是选票。但是，每次选举之间的时间间隔并不是有规律的。

(4) 显性 (explicit) 表象

视场中某一部分的显性表象意味着存在一小组神经元，它们对应着这一部分的特性，可以像检测器那样做出反应，而无须复杂的加工。

克里克和科赫的意识框架

(5) 高层次优先

一个新的视觉输入来到后，神经活动首先快速地无意识地上行到视觉系统的高层，可能是前脑，然后信号反馈到低层次，所以，达到意识的第一阶段在高层次，再把意识信号发送到额叶皮质，随后在较低层次上引起相应活动，当然这是过于简单的描述。整个系统中还有许多横向联系。

(6) 驱动性和调制性联系

对皮质锥体细胞而言，驱动性联系多半来自基底树突，而调制性输入来自丛状树突，它们包括反向投射，弥散状投射，特别是丘脑的层间核。从侧膝体到VI区的联系是驱动性的。从背脑到前脑的联系是驱动性的。而逆向联系多半是调制性的。皮质第五层上的细胞（它投射到丘脑）是驱动性的，而第六层则是调制性的。

克里克和科赫的意识框架

(7) 快照 (snapshots) (

视觉觉知过程由一系列静态的快照组成，在离散的时间内有一种运动的感觉。视皮质上有关神经元的恒定发放率，代表有某种运动发生，运动是发生在一个快照与另一个快照之间，每个快照停留的时间并不固定。对于形状和颜色的快照时间可能碰巧一样，它们的停留时间与 α 节律或甚至 δ 节律有关。快照的停留时间依赖于on信号、off信号、竞争相适应等因素。

克里克和科赫的意识框架

(8) 注意和绑定(binding)

把注意分成两类：一类是快速的、显著性驱动的和自下而上的；另一类是缓慢的，自主控制和自上而下的。注意的作用为了左右那些正在竞争的活跃的联盟。自下而上的注意从皮质第五层的神经元出发，投射到丘脑和上丘。自上而下的注意从前脑出发，分散性地反投射到皮质 I，II 和层 III 上神经元顶树突，可能途径丘脑的层间核。普遍认为丘脑是注意的器官。

绑定是把对象或事件的不同方面，如形状、颜色和运动等联系起来。绑定可能有几种类型。如果它是后天造成的，或者经验学得的，它可能具体化在一个或几个节点(nodes)上，而不需要特殊的绑定机制。如果需要的绑定是新的，那么那些分散的基本节点的活动需要联合起来一起活动。

克里克和科赫的意识框架

(9) 发放的风格 (styles)

同步振荡可以在不影响平均发放率情况下增加一个神经元的效率。同步发放的意义和程度仍有争议。计算研究表明其效果取决于输入的相关程度。

(10) 边缘效应 (Penumbra) 和意义

考虑一小堆群神经元，它们对面孔的某些方面有反应。实验者知道这一小群细胞的视觉特性，但是大脑怎么知道这些发放代表的是什么呢？这就是“意义”问题。NCC只是直接关系到所有锥体细胞的一部分，但是它会影响到许多其他神经元，这就是边缘效应。边缘效应由两部分组成，一是突触效应，二是发放率。

埃德尔曼的意识观

- 杰拉尔德·埃德尔曼 (Gerald M. Edelman) 是美国免疫生物学家和神经生物学家, 意识科学研究领域的主要开拓者之一。1972年与英国生物学家罗德尼·波特 (Rodney R. Porter) 分享了诺贝尔生理学或医学奖。埃德尔曼以“三条工作假设”为基本立场, 提出阐释脑工作方式的“神经元群选择理论”与“动态核心假说”等, 在意识研究领域做出了开拓性贡献。
- 埃德尔曼强调意识的整合性和分化性。依据脑的生理病理和解剖学上的事实, 埃德尔曼认为丘脑-皮质系统在意识的产生方面起关键作用。2003年, 埃德尔曼在美国科学院系列 (PNAS) 上发表一篇论文, 一开始就主张掘弃二元论。

意识的特性

- 2012年，巴尔斯(B J Baars) 和埃德尔曼 (G M Edelman) 在刊物《生命物理评论 (Physics of Life Review) 》发表文章 “Consciousness, biology and quantum hypothesis” 中，阐述他们关于意识的自然观，列出了意识状态的17个特性：

(1) 意识状态的脑电图 (EEG) 标记

脑的电生理活动呈现不规则、低幅度和快速的电活动，频率从0.5到400 Hz。意识EEG 看起来与无意识状态（类同沉睡情况）显著不同，癫痫患者和全身麻醉的意识状态呈现规则、高幅度，和慢变化的电压。

意识的特性

(2) 大脑和视丘

意识取决于视丘的复杂性，开启和关闭通过脑干调制，并且与脑皮层下区域没有交互作用，不直接支持意识经验。

(3) 广泛的大脑活动

可报告意识事件与广泛的具体脑活动内容有关。无意识的刺激只唤起局部的脑活动。意识瞬间也对外边专注意识内容引发广泛的影响，表现为隐性学习、情景记忆、生物反馈训练等。

意识的特性

(4) 大范围的可报告内容

意识有特别广泛的不同内容——各种感觉的知觉、内生的形象化描述、感情感觉、内部语言、概念、有关行动的想法和像熟悉的感觉那样的外部经验。

(5) 信息性

当信号变得多余时意识可以消失；信息损失可以导致意识访问的丢失。选择性注意的研究也显示信息更丰富的意识刺激的强烈偏爱。

(6) 意识事件的适应性和飞逝的本质

立即经历感觉输入可以维持到几秒，我们短暂认知的持续存在不到半分钟。相反，庞大的无意识知识可以驻存在长时记忆中。

意识的特性

(7) 内部一致性

意识以一致约束为特征。一般同时给予刺激两个是不一致，只有一个能变得有意识。当一个词多义时任何时候只有一个意义变得有意识。

(8) 有限能力和顺序性

意识的能力在任何规定的片刻好像限制在仅对一个一致景象，和直接同时观察时脑形成的大量并行处理相反，这样的意识景象流是串行的。

意识的特性

(9) 感觉捆绑

感觉大脑就其功能作用是分块的，从而不同的脑区对不同的特征，例如形状、颜色或者目标运动作出反应。一个基本的问题是这些就其功能作用分开的脑区怎样协调它们的活动，产生普遍的有意识的综合完形知觉。

(10) 自我特性

意识经验总是以自我经历为特点，正如威廉·詹姆士称为“观察自我”。自我功能看起来与中央脑区有关，人脑包括脑干、楔前叶(precuneus)和前额叶(orbitofrontal)皮层

意识的特性

(11) 准确可报告性

意识的大多数使用的行为迹象是准确可报告的。全范围的意识内容因为大范围自愿的反应是可报告的，经常有非常高的准确性。可报告不要求完全明确的词汇，因为主体能自动地对意识事件进行比较、对比、指向和发挥作用。

(12) 主观特性

意识以事件私有流方式提供给经历主体为特征。这样的隐私没有违反立法。这表明自我物体综合是有意识认知的关键。

意识的特性

(13) 关注非主流结构

意识被认为倾向于专注明白清楚的内容，“非主流意识”事件，如亲情感、舌尖经验、直觉等同样重要。

(14) 促进学习

几乎没有证据表明学习无需意识。相反，意识经验促进学习的证据是压倒一切的，即使隐性(间接的)学习也需要有意识的注意。

(15) 内容的稳定性。

意识内容给人深刻印象是稳定的。产生的输入和任务的变化需要，例如读者经常使用眼睛运动扫描句子。即使像自身的信念、概念和专题一样的抽象意识内容，可能在几十年内非常稳定。

意识的特性

(16) 关注特性

意识的景象和目标，一般说来是关注外部的来源，虽然他们的形成严重依赖于无意识的框架。

(17) 意识知道和决策

意识对于我们知道周围世界，以及一些我们的内部过程是有用的。意识的表达，包括感觉、概念、判断和信仰，可能特别适于自如的决策。但是，并非全部有意识事件都涉及大范围的无意识设施。这样，意识报告的内容绝不是仅需要被解释的特征。

内容提要

- 意识的概念
- 意识的研究史
- 意识理论概述
- 全局工作空间理论
- 整合信息理论
- 机器意识系统
- 显意识思维与潜意识思维

意识理论概述

1. 法伯的意识观
2. 心理学的意识观
3. 还原论
4. 剧场模型
5. 神经达尔文主义
6. 微管假说
7. 量子意识观
8. 整合信息理论



史忠植：智能科学. 清华大学出版社. 2019

法伯的意识观

法伯 (I. B. Farber) 等人在其《意识与神经科学——哲学与理论问题》一文中，从三个层次讨论了意识概念：

- 第一个层次是“意识觉察”——包括感觉觉察、概括性觉察、元认知觉察和有意识回忆等四种。
- 第二个层次是“较高级的官能”——不仅能被动地感知和觉察信息，还具有注意、推理和自我控制等较高级的官能(即还具有能动作用)。
- 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。

心理学的意识观

普林斯顿大学心理系教授的菲力普·约翰逊-莱尔德 (Philip Johnson-Laird)、布兰迪斯大学 (Brandeis University) 语言学 and 认知学教授雷·杰肯道夫 (Ray Jackendoff)、加利福尼亚州伯克利的赖特研究所的巴尔斯 (Bernard J. Baars) 达成三点共识：

- 并非大脑的全部活动都直接与意识有关，而且意识是一个主动的过程；
- 意识过程有注意和某种形式的短时记忆参与；
- 意识中的信息既能够进入到长时情景记忆中，也能进入运动神经系统的高层计划水平，以便控制随意运动。

还原论的意识观

- 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，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”。
- 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”。他坚信，意识这个心理学的难题，可以用神经科学的方法来解决。

惊人的假说

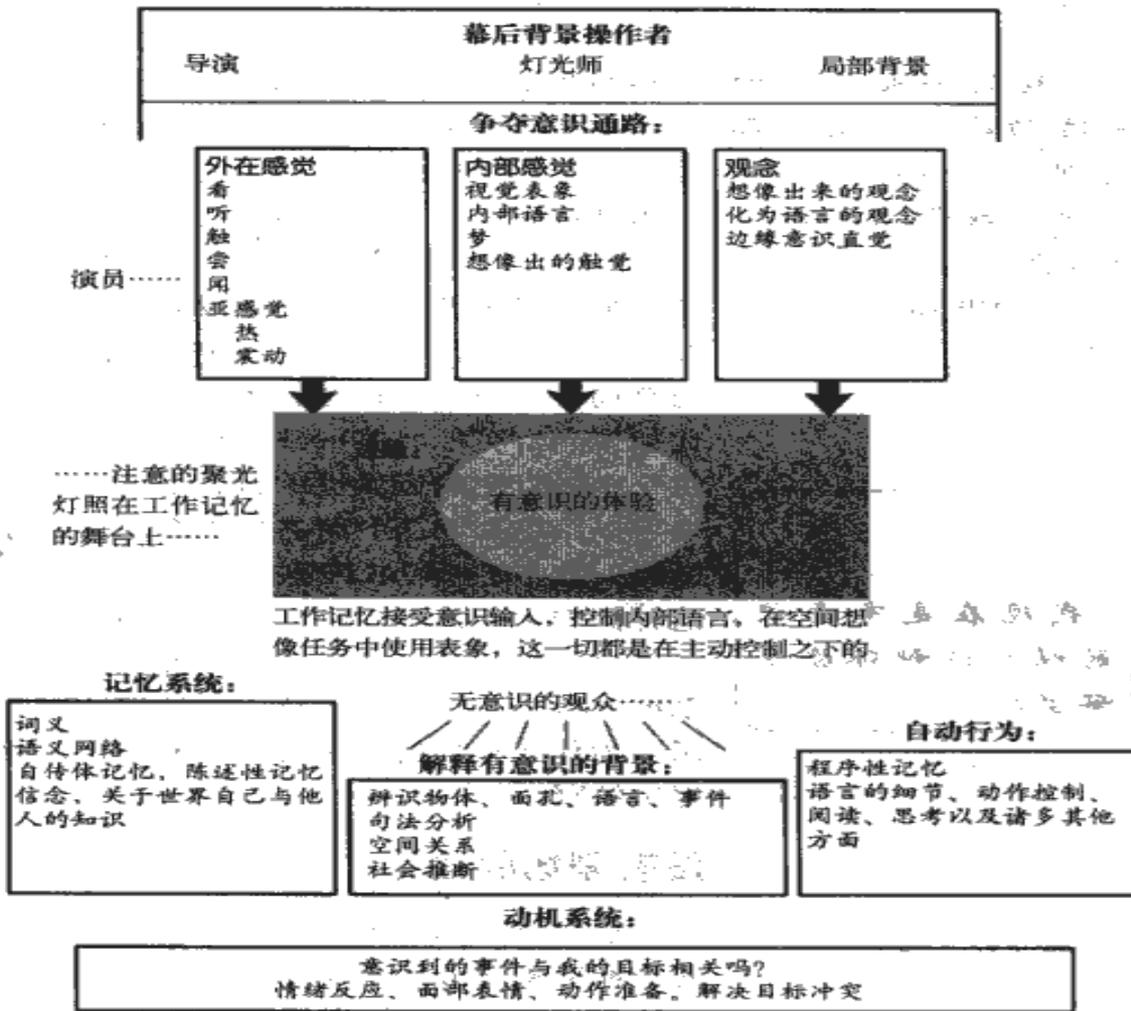
在弗朗西斯·克里克的《惊人的假说——灵魂的科学探索》一书中基于“黑箱”方法的对意识的三种界定：

- 1、并非大脑的全部活动都直接与意识有关；
- 2、意识过程有“注意”和“短时记忆”的参与；
- 3、意识中的信息既能进入长时情景记忆，也能进入运动神经系统以控制随意运动。

剧场模型

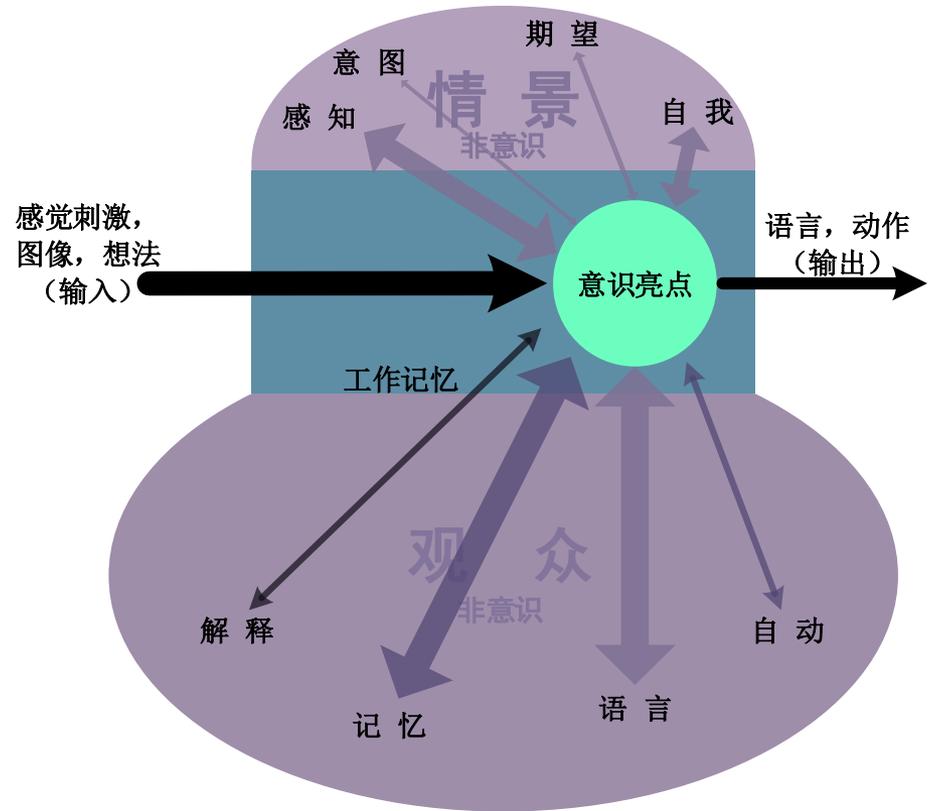
- 关于意识问题，最经典的一个假设即所谓“剧场模型”。在这一个隐喻中，把多个感觉输入综合成一个有意识的经验，比拟为在黑暗的剧场内舞台上聚光灯打出一个光亮点照到某个地方，然后传播给大量的无意识的观众。在认知科学中，关于意识和选择性注意的假设多数来自于这个基本的隐喻。巴尔斯 (Baars B J) 是“剧场隐喻”的最主要的继承和发扬光大者。
- 意识的“剧场模型”隐喻在舞台上同时有许多角色在演出，正像人脑同时接受内外感受器的多种刺激，但是只有少量角色接收聚光灯的照射，这中间有个选择问题，而且聚光灯不是停留在一个地方、一个角色身上，而是随着时间流动，观众代表着脑的无意识的部分。

剧场模型



全局工作空间理论

- 全局工作空间理论详细说明大脑意识活动在认知活动中的角色和作用。观众作为戏院中无意识的处理器，接受舞台上意识焦点的广播。对于这个焦点的控制相当于选择性的注意。在后台，无意识的连接系统形成和导向有意识的内容。有意识的活动是和有限容量的处理过程强烈相关的，表现出有意识内容的内部相容性和低计算效率。



神经达尔文主义

- 神经达尔文主义是由美国洛克菲勒大学的埃德尔曼（G. M. Edelman）提出的，其主要思想借鉴了达尔文的自然选择学说，认为由神经元通过紧密互联组成的神经元群是脑内神经联结的结构和功能模式的选择性活动主体。而我们的意识活动和心智活动就是动态的达尔文过程，所有的行为现象都是神经细胞活动的时空模式决定的，这些时空模式相互竞争中的每一时刻的赢家，就将成为显现的心智活动，特别是意识活动。说到底，意识活动无非就是大量神经活动中模式选择“胜者为王”的结果。

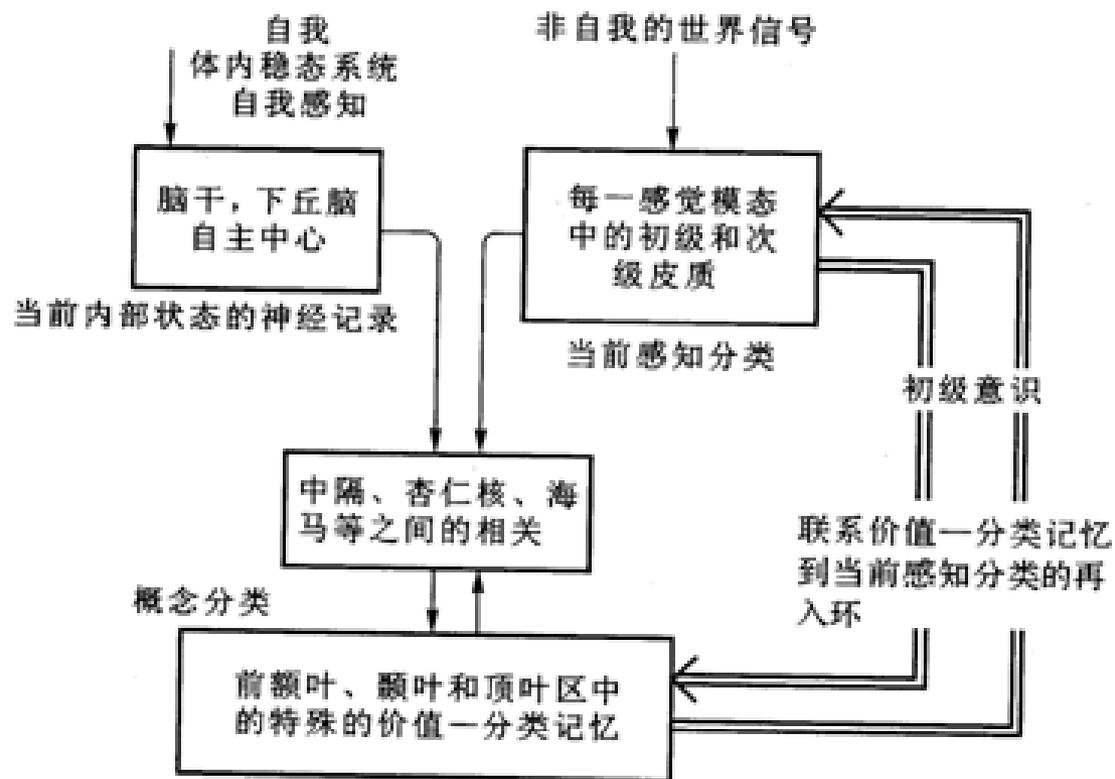
神经达尔文主义

埃德尔曼(2003)在阐述他的意识的自然观时，列出一张表，把意识状态的特性分成一般性、信息角度和主观方面三类：

- 一般性：意识状态是单一的，整合的和被大脑构建的是可以被大量分化的。
- 信息意向性：从时间上考虑它们是顺序的，系列的和可以变化的，可以把不同模式捆绑起来。
- 主观感觉和感受性：神经元群体中动态相互作用的结果，主要是丘脑-皮质系统。

神经达尔文主义

来自两类的信号(自我和非自我)是关键。
关系到价值的信号与来自外界世界的分类,信号进行相关并进入记忆,以便进行概念上的分类。这种价值分类记忆由再入通路联络到外周世界信号的当前感知分类。



微管假说意识理论

- 微管假说以彭罗斯(Penrose)最为著名。他是当代著名的数理科学家。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。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。
- 人们认为从意识角度考虑,这个空间尺度太小,层次太低,从系统论角度考虑用这么小的空间尺度和低层次的角度,无法对如此复杂和高层次的功能做出完美的解释。

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。



Not sure. But quantum physics, over classical physics, can better explain it
(*Quantum Consciousness*)

量子意识观

- 彭罗斯 (Roger Penrose) 和哈梅罗夫 (S R Hameroff) 合作，在20世纪90年代早期共同建立了广受争议的“和谐客观还原模型 (Orch-OR模型)”。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。
- 佩罗斯 (M Perus) 将神经计算与量子意识相结合的设想。在量子理论中量子系统的状态则可以用随时间变化的波函数描述。这样一来，神经元状态是神经元图式的一种迭加就可以变为是量子本征波函数的一种迭加了，并且迭加的量子本征波函数通常具有正交性和正则性。

整合信息理论

- 艾伦脑科学研究所(Allen Institute for Brain Science)所长、首席科学家科赫(Christof Koch)Koch认为,“所有的认知,包括意识,只不过是某种形式的计算”。
- Koch一直在与神经学家朱利奥·托诺尼(Giulio Tononi)合作发展一种与GWT竞争的理论。他们认为意识不是在将输入转化为输出时产生的东西,而是一种就在认知网络中的内在属性,一种网络结构中具有特定特征。Tononi将这一观点命名为整合信息理论(Integrated Information Theory, IIT)。

内容提要

- 意识的概念
- 意识的研究史
- 意识理论概述
- 全局工作空间理论
- 整合信息理论
- 机器意识系统
- 显意识思维与潜意识思维

两种意识理论的争论

- 德克萨斯州大学奥斯汀分校的计算机科学家 Scott Aaronson说研究者制定了一整套实验方案，将各种意识理论无情地暴露在聚光灯下，期望至少能排除其中的一些不合理选项。
- 邓普顿世界慈善基金会（the Templeton World Charity Foundation）的一名高级项目负责人Dawid Potgieter负责协调这项工作，并计划未来五年内建立更多类似这样“有组织的对抗性合作”。他对项目前景看法很实际，表示目前的努力只是持续筛选意识理论的一个开始。

两种意识理论的争论

- 邓普顿世界慈善基金会已经对这项任务拨款2000万美元，用于判定这些意识理论的生死存亡。这是从 Potgieter 所说的 IIT 和 GWT 的“结构化敌对协作” (structured adversarial collaboration) 开始的，因为它们能够做出可验证的对比和预测。
- Potgieter 声言，巴黎法兰西公学院的认知科学家Dehaene 和 Tononi 愿意“把自己置身于危险之中”，是促使选择 GWT 和 IIT 作为第一批理论候选的考量因素之一。

全局工作空间理论

- 关于意识问题，最经典的一个假设即所谓“剧场中的亮点”隐喻。在这一个隐喻中，把多个感觉输入综合成一个有意识的经验，比拟为在黑暗的剧场内舞台上聚光灯打出一个光亮点照到某个地方，然后传播给大量的无意识的观众。在认知科学中，关于意识和选择性注意的假设多数来自于这个基本的隐喻。巴尔(Baars B J)是“剧场隐喻”的最主要的继承和发扬光大者。
- 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。

全局工作空间理论

■ 全局工作空间理论GWT要点：

- 只有少数事件被意识到；
- 聚焦的意识类似起到亮点的作用，意识的内容被限制的亮点之内；
- 舞台、后台与观众的互动是以一个全局工作空间结构为基础。

内容提要

- 意识的概念
- 意识的研究史
- 意识理论概述
- 全局工作空间理论
- 整合信息理论
- 机器意识系统
- 显意识思维与潜意识思维

整合信息理论

- 艾伦脑科学研究所(Allen Institute for Brain Science)所长、首席科学家科赫(Christof Koch)Koch认为,“所有的认知,包括意识,只不过是某种形式的计算”。
- Koch一直在与神经学家朱利奥·托诺尼(Giulio Tononi)合作发展一种与GWT竞争的理论。他们认为意识不是在将输入转化为输出时产生的东西,而是一种就在认知网络中的内在属性,一种网络结构中具有特定特征。Tononi将这一观点命名为综合信息理论(Integrated Information Theory, IIT)。

整合信息理论

- 威斯康星大学麦迪逊分校的神经科学家托诺尼 (G Tononi) 与埃德尔曼等发表了一系列论文，阐明意识的整合信息理论。提出意识量是由复杂元素生成的整合信息量，并由它生成的信息关系规定的体验质量。托诺尼提出整合信息两个测度。
- 托诺尼的第一项测度的工作原理是考虑所有的各种双分区的神经系统（分裂成两部分）：综合信息的能力被称为 Φ ，并且由双分区子集可以交换的最小有效信息给定。托诺尼的方法需要检查所考虑的系统每个子集。

整合信息理论

- 每个双分区分为两个不重叠的部分。假设子集S，和二分为A和B，托诺尼定义了一个测度，称为有效信息（EI）。有效信息使用信息论中的标准度量互信息（MI）。这不是标准的互信息测度，而是考虑A和B之间的连通性的信息增益互信息测度。托诺尼的EI是一个衡量累积的信息增益测度，当A的输出在所有可能的值随机变化时，考虑对B的效果。其目的是将因果关系的一些因素结合起来。互信息MI可以用下面公式描述：

$$MI(A:B) = H(A) + H(B) - H(AB)$$

其中， $H(\quad)$ 是熵，反映不确定性的测度。如果A和B之间没有交互，互信息为零，否则它是正值。

整合信息理论

- 托诺尼和合作者提出 Φ 的修订测度，那就是 Φ_2 。该修订测度 Φ_2 比前面的测度优越，因为它可以处理随时间变化的系统，提供一个瞬时到瞬时变化的 Φ_2 测度，对应于衡量瞬时到瞬时的意识水平。
- 托诺尼认识到，信息综合理论被用来研究系统维持状态的能力，可以说是“智能”。托诺尼以几何的方式，引入形状，体现由系统相互作用产生的一整套信息关系作为感受性 (Qualia)。文献[56]探讨了感受性 (Qualia) 涉及到底层系统的特征和体验的基本特征，提供关于感受性几何神经生理学和现象学几何的初始数学词典。感受性空间 (Q) 是具有复杂性每个可能状态 (活动模式) 的轴线空间。

整合信息理论

托诺尼认为：“意识就是一种体验。”这些体验的基本特征：

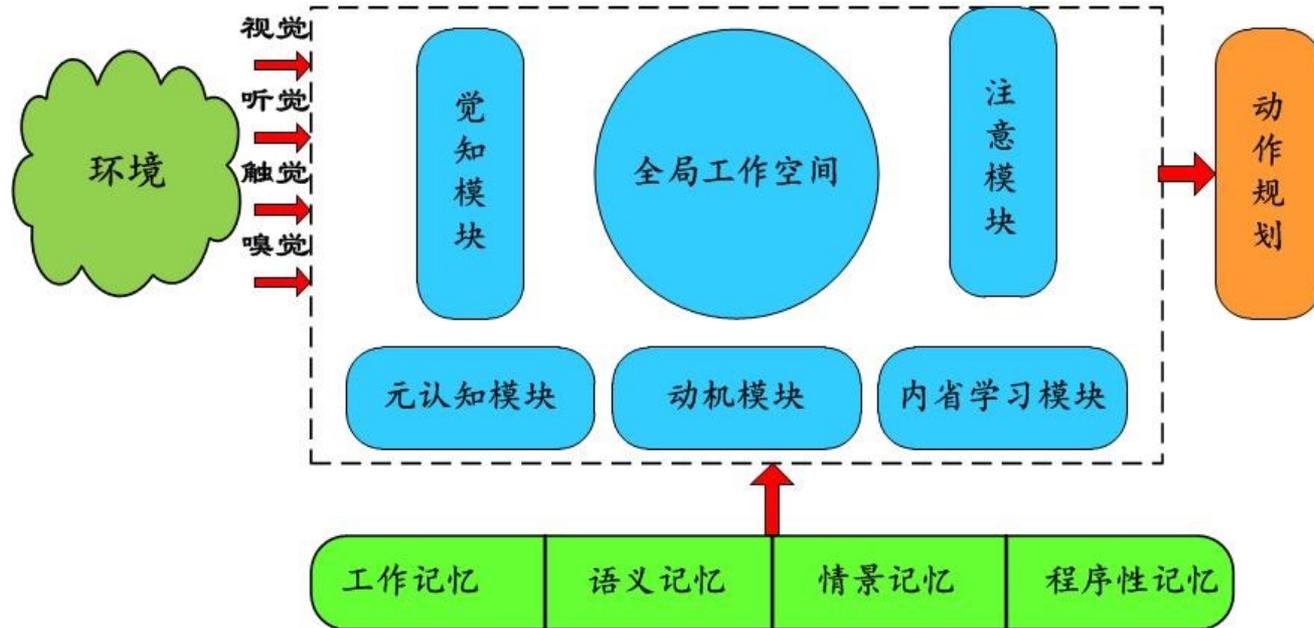
- 主观性 (subjective)：体验只存在于有意识的主体中；
- 结构性 (structured)：意识内容彼此相关，比如“蓝皮书在桌子上”；
- 具体性 (specific)：比如这本书是蓝色的，而不能同时是红色的；
- 统一的 (unified)：一次只有一种整体体验，不能分解为部分之和（红三角 \neq 三角+红色）；
- 有界性 (definitive)：这种体验包含的内容在时空是有边界的。

内容提要

- 意识的概念
- 意识的研究史
- 意识理论概述
- 全局工作空间理论
- 整合信息理论
- 机器意识系统
- 显意识思维与潜意识思维

机器意识系统架构

机器意识系统由全局工作空间、觉知、注意、动机、元认知、内省学习模块构成。



觉知模块

- 觉知开始于外界刺激的输入，激活感知系统的初级特征检测器。输出信号被发送到感觉记忆中，在那里更高层次的功能探测器用于更抽象的实体，如对象，类别，行动，事件等的检测。所产生的知觉移动到工作区，在那里产生本地联系的短暂情景记忆和陈述性记忆会被做线索标记。这些本地联系与知觉结合，产生当前情景模型，用以表示智能体对当前正在发生的事情的理解。

注意模块

检测新事件是任何信号分类方法的一个重要的功能。因为我们不能对机器学习系统训练所有可能遇到的对象类的数据，它就变得很重要，它在测试时能够区分已知和未知的对象信息。新奇检测是一个非常具有挑战性的任务，可以在复杂的、动态的环境中发现新颖的、感兴趣的事件。新奇检测是一个很好的分类或识别系统的基本要求，因为有时候测试数据中包含的对象，训练模型时信息并不知道。觉知的新颖性是关系到认知，而认知的新颖性是关系到知识。基于固定组训练样本从一个固定数量的类别，新奇的检测是一个二元决策任务对每个测试样本确定它是否属于一个已知的类别或不。

选择策略

■ 阈值选择机制 (TSM)

TSM 是一种阈值过滤算法。当响应显著性值大于阈值时，这个神经元的响应被选择进入下一级神经元的处理，否则被忽略。

$$Output(A_i) = \begin{cases} 0 & \text{if } RS(A_i) \leq T_i \\ R_i & \text{if } RS(A_i) > T_i \end{cases}$$

■ 比例选择机制 (PSM)

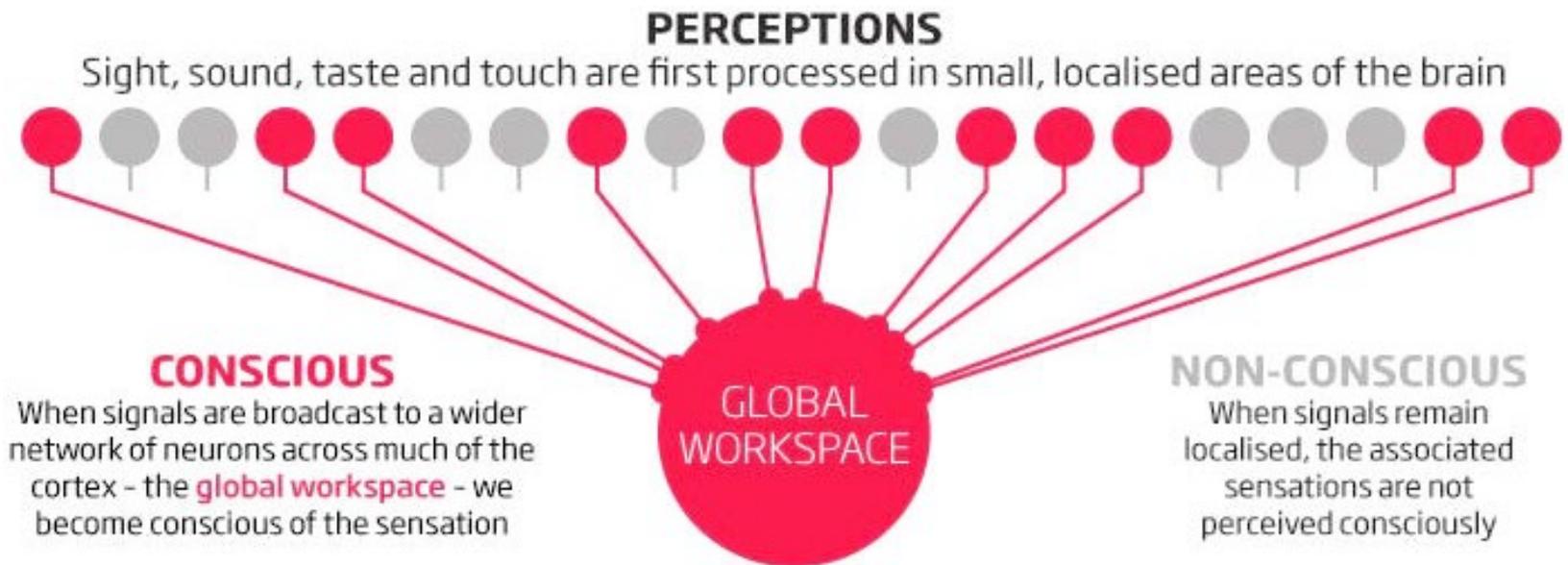
PSM 是一种瓶颈过滤算法。根据响应显著性值进行排序，选择前面的一定比例（假设为p）的神经元进入下一级处理。

全局工作空间模块

全局工作空间是处在工作记忆部位，在这个记忆里不同的系统可以执行它们的活动。全局意味着这个记忆中的符号通过众多的处理器被分配、传递开来。当然，每一个处理器都可能产生一些局部的变量并运行。但它对全局性的符号、信息却是相当敏感，可以及时做出感应。当面对全新的以及与习惯性刺激存在差异的事物时。我们的各种感官都会产生定向反应，同时各种智能处理器会通过合作或竞争的方式在全局工作空间中展示它们对该新事物的认知分析方案，直到获得最佳的结果。正是在这个过程中，我们对此新事物产生了意识。全局工作空间可以看作信息共享的黑板系统，通过使用黑板，各个处理器试图传播全局性的信息，联合建立问题解决的办法。

全局工作空间模块

工作区内部结构是由各种不同的输入缓冲区和三个主要模块组成：当前情景模型，暂存器和意识内容队列。当前情景模型是一个存储代表实际的当前内部和外部事件的结构。



动机模块

- 一个动机可以表示为一个三元组 $\langle N, G, I \rangle$ ，其中 **N**表示need，即需求；**G**表示goal，即目标；**I**表示intensity，强度。
- 对于一个动机的被选中，需要满足：

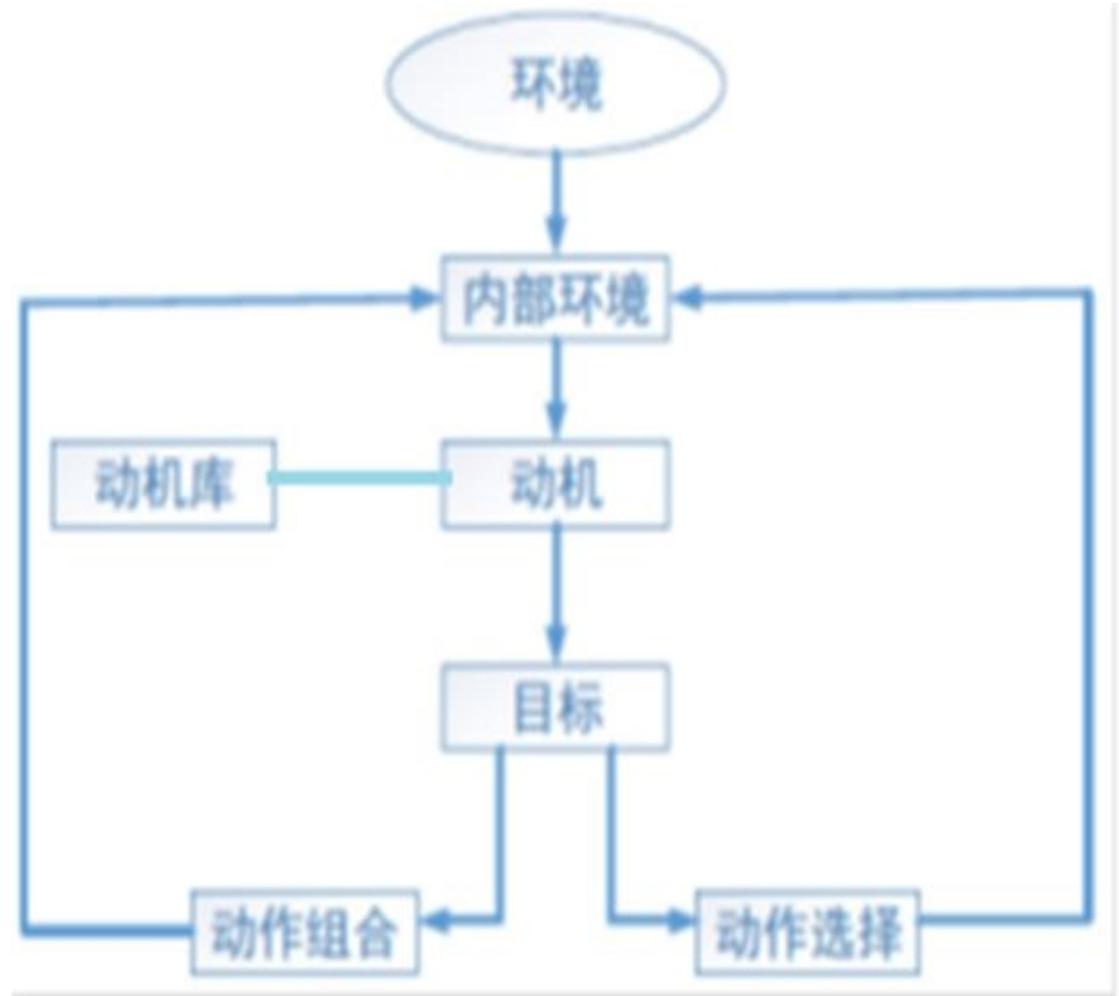
$$R=(P, D, \text{Strength}(P|D))$$

其中，**P**表示实现动机的条件，**D**表示实现动机的动作集合，**Strength(P|D)**的取值为 $[0,1]$ 。

动机系统结构

动机系统逻辑结构主要包含：

- (1) 环境
- (2) 内部环境
- (3) 动机
- (4) 动机库
- (5) 目标
- (6) 动作选择
- (7) 动作组合



动机执行流程

动机执行流程如下。事件接收器从环境中获取信息并将信息放置到动机列表中。分配模块不断地从事件列表中选择事件，并为之提供相应的规划。当目标到达的时候，检测现有的环境是否满足目标实现的条件。分配器选择相应的规划用来完成预期的目标，这个过程中需要推理机的配合。这就意味着，系统会找到一个或者多个我们在以前人为制定的方案。当对现有的目标进行推理的时候，有可能找到的并不是唯一的解决方案。这时候，推理机需要按照其内部的规则进行选择，选择的标准需要事先人为指定。不同的选择标准，会导致智能体在进行决策的时候，有不同的行为响应。选择好的规划方案后，系统会将需要事先的目标和方案联系起来。

元认知模块

元认知为智能体提供关于自己思维活动和学习活动的认知和监控，其核心是对认知的认知。元认知模块具有元认知知识、元认知自我调节控制和元认知体验的功能。元认知知识包括关于主体的知识，任务的知识以及策略的知识。元认知体验指的是对于自己认知过程的体验。在认知过程中，通过元认知自我调节控制，选择合适的策略，实现策略的使用，进程与目标的比较，策略的调整等等。

内省模块

内省学习系统中的一个重要问题就是失败的分类问题。失败的分类是诊断任务的基础，同时它为解释失败和构建修正学习目标提供重要的线索。失败分类需要考虑两个重要的因素，一个是失败分类的粒度，另一个是失败分类、解释失败及内省学习目标的关系。基于本体的知识库是将基于本体的知识表示方式同专家系统的知识库相结合，从而知识库具有概念化、形式化、语义明确、共享等优点。通过利用基于本体的知识库方法解决内省学习中的失败分类问题，使得失败分类更加清晰，检索过程更加有效。

内容提要

- 意识的概念
- 意识的研究史
- 意识理论概述
- 全局工作空间理论
- 整合信息理论
- 机器意识系统
- 显意识思维与潜意识思维

有意识与无意识的判别标志—— 工作记忆的内容

既然意识是人脑对时间逻辑思维 and 空间结构思维的觉察、调节或控制。要判断是否有意识产生，至少要能“觉察”出思维过程。任何思维过程都离不开四个要素：思维加工材料、思维加工方式、思维加工缓存区和思维加工机制。换言之，只要缺少其中任一要素，思维过程都将因难以进行而不能被觉察出来。因此，如果只是要判别是有意识或无意识即仅局限于觉察思维过程的话，问题就简单得多，因为要觉察出某种思维过程，并不需要去仔细分析四个要素中的每一个要素，只要抓住其中最简单又与觉察直接相关的要素就行了——这个要素就是“工作记忆”。这是因为，工作记忆与长时记忆不同，工作记忆中的内容并不长期保留，它只起缓冲存储器作用：在思维加工过程中暂存加工对象和加工结果，加工完成后，其中内容就会很快消失。因此，我们不必考虑思维加工的方式、机制，也不必管思维加工的对象，只需考虑工作记忆中是否有内容，以及工作记忆持续时间的长短就可以觉察出是否有思维过程发生。

显意识与时间逻辑思维

时间逻辑思维是建立在语言符号序列的基础上，所以具有顺序性、持续性的特点。也就是说，在运用基于语言符号的概念来进行判断、推理时，只能按照词语符号出现的先后顺序一步一步地操作，持续时间往往较长，尤其是在分析处理事物之间的复杂关系时更为突出。总之，在逻辑思维情况下，其工作记忆的持续时间是较长的，加上是按线性、顺序方式工作，每一步骤都很清楚、明确。因此这种思维过程很易于觉察，而且还可以由思维主体将思维过程的每一步骤用言语表述出来——使内隐的心理操作过程变成外显的言语活动。正是在这个意义上，通常把时间逻辑思维称为“显意识思维”。换言之，显意识思维是指其过程不仅能被觉察还能用言语加以描述的思维；反之，其过程若不能被觉察或不能用言语加以描述的思维皆属潜意识思维。

形象思维

■ 形象思维（imaginal thinking）是用直观形象和表象解决问题的思维。形象思维是对形象信息传递的客观形象体系进行感受、储存的基础上，结合主观的认识和情感进行识别，并用一定的形式、手段和工具创造和描述形象的一种基本的思维形式。

■ 形象思维的特点

➤ 形象性

形象思维所反映的对象是事物的形象，思维形式是意象、直感、想象等形象性的观念，其表达的工具和手段是能为感官所感知的图形、图象、图式和形象性的符号。形象思维的形象性使它具有生动性、直观性和整体性的优点。

形象思维

➤ 非逻辑性

对信息的加工过程不是系列加工，而是平行加工，是面性的或立体性的。它可以使思维主体迅速从整体上把握住问题。

➤ 粗略性

形象思维对问题的反映是粗线条的反映，对问题的把握是大体上的把握，对问题的分析是定性的或半定量的。所以，形象思维通常用于问题的定性分析。

➤ 想象性

形象思维并不满足于对已有形象的再现，它更致力于追求对已有形象的加工，而获得新形象产品的输出。

显、潜意识与直觉思维

直觉思维是凭借已有知识和经验对事物直接领悟的思维活动，是靠直接的判断解决问题的能力。它不是线形、顺序的慢节奏加工，而是在整体综合、直观透视基础上快速作出的直觉判断。所以我们把这种思维称之为“直觉思维”。

直觉思维通常有两种：简单直觉思维与复杂直觉思维。简单直觉思维的加工对象是与空间视觉定位有关的空间位置表象，即上面所述的空间表象；复杂直觉思维的加工对象则是用来描述复杂事物之间结构关系的“关系表象”。在空间视觉定位情况下，工作记忆内必有关于客体位置的初始特征值，以便根据这些特征值由思维加工机制确定客体的空间位置；而在判断、处理复杂事物关系的情况下，由于事物之间隐含的复杂关系是有待发现的，所以在工作记忆中将不会有初始值。这是两种直觉思维的很大不同之处。

显、潜意识与直觉思维

以空间表象作为加工对象的空间结构思维，由于其特点是整体把握、直观透视、空间整合基础上进行的快速立体思维，其工作记忆必然短暂，因而这种思维过程难以觉察，往往表现为“潜意识思维”。尤其是在复杂直觉思维中，由于事物之间复杂的内在关系一时很难把握住，甚至经过较长时间思索也找不到这种关系。这时将会出现和创造想象过程中的类似现象——有一段时间工作记忆的内容为空白。可见复杂的直觉思维是和创造想象一样的真正的潜意识过程。

直觉思维模型

- 查理·芒格(Charlie Munger)100种思维模型第二个：直觉思维模型。直觉又称为直观，通常被描述为，一种不用经过太多思考过程，很快就能出现的直接想法、感觉、信念或者偏好。当我们有某种信仰，但是不确知它的原因时，通常会将它归于是一种直觉。认知科学认为这是因为生存的演化压力而产生的人类心智能力，让人类可以快速做出判断，采取行动，它通常与右脑连结在一起。心理学与超自然学对这个能力深感兴趣。
- 查理·芒格跳出了传统的金融和会计为基础的投资分析框架，提出了综合多种学科立体地辩证的“思维格栅模型”。

思考题

12-1 什么是意识？

12-2 意识具有哪些基本要素？

12-3 全局工作空间模型的核心思想是什么？

12-4 如何理解意识是复杂系统上“涌现”出来的功能？

12-5 意识的综合信息论如何定量分析意识潜力？

12-6 CAM的机器意识系统由哪些模块组成？试阐述各模块的功能。

Thank You

