整合信息论简介

——探索意识的机制

解读：何真

2021年11月

提纲

1 整合信息论简介

2 关于意识的公理

3 对意识的物理基质的假定

4 机制的系统与概念结构

5 整合信息论的局限性

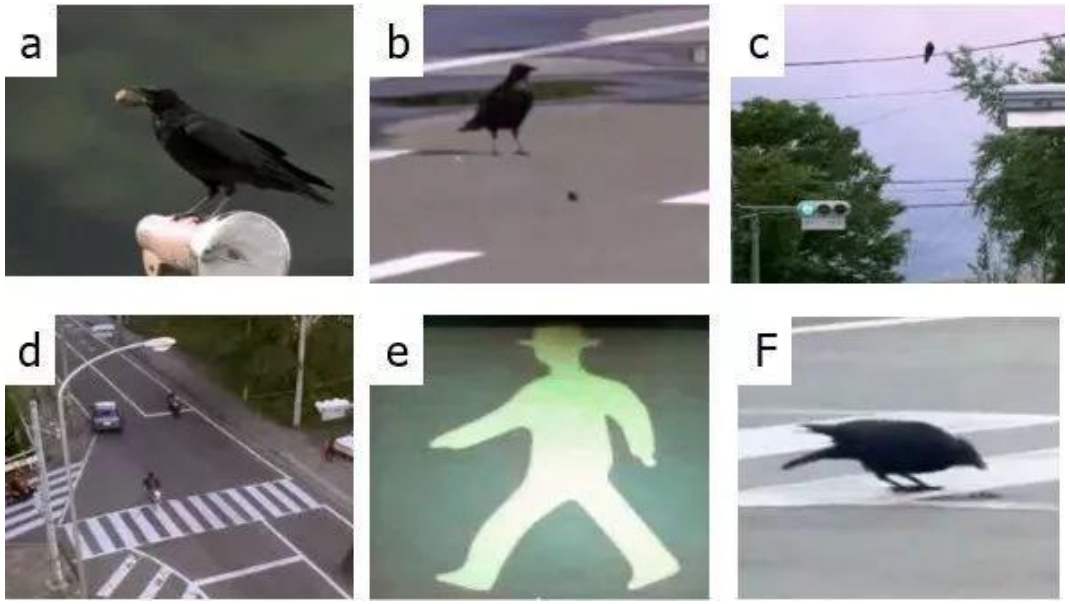
整合信息论（IIT）是什么？

- 意识是主观经验——例如，理解一个场景的感觉，反思经验本身的感觉。当意识消退的时候，就像在不做梦的睡眠中，从经验主体的内在视角看，整个世界消失了。
- 意识依赖于大脑的某些区域，一个经验的特定内容依赖于大脑皮层的某些部分的神经元的活动。
- 然而，尽管有越来越多的临床和实验研究，对意识和大脑之间的联系的正确理解仍然有待建立。
  - 例如，我们不知道为什么大脑产生意识，而小脑不会，尽管小脑的神经元数目是大脑的四倍。
  - 又如，为什么深度睡眠时意识消退了，但大脑皮层依旧活跃？

整合信息论（IIT）是什么？

- 更多关于意识的问题：
  - 新生的婴儿有意识吗？
  - 有些动物表现出复杂行为，但具有与人类非常不同的大脑，它们有意识吗？
  - 智能机器会有意识吗？
  - 难问题：为什么有些神经机制与意识相关，而其它的与意识无关

乌鸦头有多大？不到人脑的1%大小。 人脑功耗大约是10-25瓦，它就只有0.1-0.2瓦。
——朱松纯教授浅谈人工智能：现状、任务、构架与统一



整合信息论（IIT）是什么？

- 对意识有两种研究途径
  - 经验主义研究
  - 理论路径

- 整合信息论，Integrated information theory （IIT）
  - 以一种新的方式探讨“难问题”
  - 不是从大脑出发，不直接研究大脑如何产生经验
  - 它从经验的基本现象学性质出发

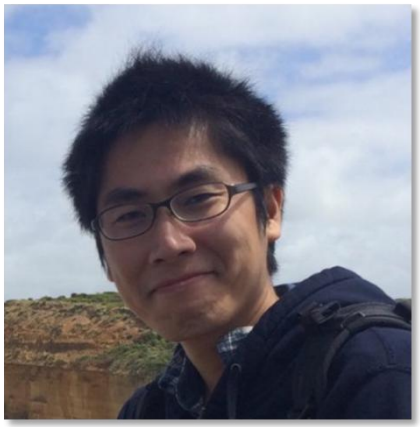
整合信息论（IIT）提出者

朱利奥·托诺尼
（Giulio Tononi）
威斯康星大学麦迪逊分校
神经科学家和精神病学家

主要合作者：

Larissa Albantakis

Department of Psychiatry, UW Madison Computational Neuroscience



Masafumi Oizumi

University of Tokyo

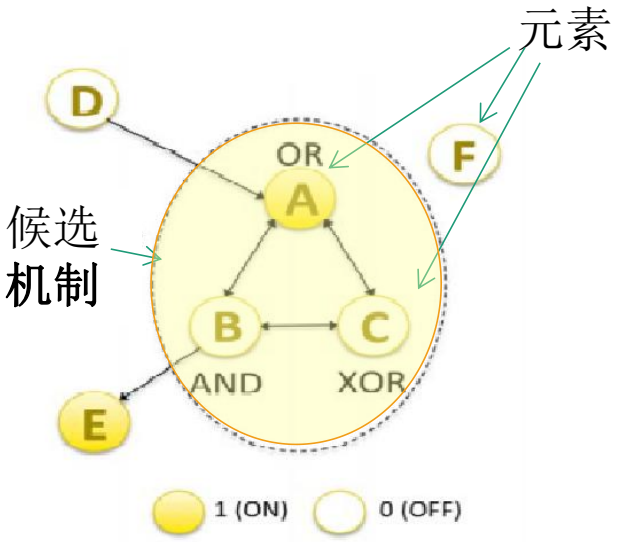
参考资料

- Integrated information theory: from onsciousness to its physical substrate
  Nature Reviews Neuroscience (2016)
- From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0
  PLoS Computational Biology (2014)
- Integrated information theory (百科)
  Scholarpedia (2015)

- http://integratedinformationtheory.org/
  （论文、代码）

整合信息论（IIT）的主体框架

- 整合信息论，Integrated information theory （IIT）
  - 它从经验的基本现象学性质出发，提出**公理**
  - 推断出意识的物理基质需要满足的特性（**假定**）
  - 从系统内在的角度分析候选机制的**因果作用**
  - 给出定性与定量评估意识的数学框架



整合信息论考察候选机制：

- 是否具有**因果作用**？
- 是否构成**概念**？
- 是否形成**意识**？

提纲

1 整合信息论简介

2 关于意识的公理

3 对意识的物理基质的假定

4 机制的系统与概念结构

5 整合信息论的局限性

关于意识的公理

经验（experience）的基本性质。

- IIT公理（axioms）应该能捕捉经验的基本性质，具体而言：
  - 关于经验本身的；
  - 明显的：不需证明；
  - 基本的：适用于所有经验；
  - 完整的：包含意识的所有的基本性质；
  - 一致的：公理间没有矛盾；
  - 独立的：公理不能互推。

关于意识的公理

经验（experience）的基本性质。

1. Intrinsic existence（内在的存在）

意识是存在的：每个经验都是真实的——实际上，我在此时此刻的经验存在（即该经验是真的），这是我唯一能够立即且绝对地确信的事情。

内在真实的（ intrinsically real or actual ）：我的经验的存在是独立于外部观察者的，是从它自身固有的角度来看的。

关于意识的公理

经验（experience）的基本性质。

2. Composition（组合）

意识是结构化的：每个经验由多重初级的或高阶的现象特质（*phenomenological distinctions*）组成。例如，在一个经验中，我能够区分书、蓝色、一本蓝色的书、左边、左边的一本蓝色的书，等等。

关于意识的公理

经验（experience）的基本性质。

3. Information（信息）

意识是特定的：每个经验都以它特殊的方式存在——是由特定的现象特质的一个特定集合组成的——因而区别于其它可能的经验（*differentiation*）。例如，当前我的经验的内容是看见一个房间，书等等，而与其它经验区别开（例如，没有书的场景）。

关于意识的公理

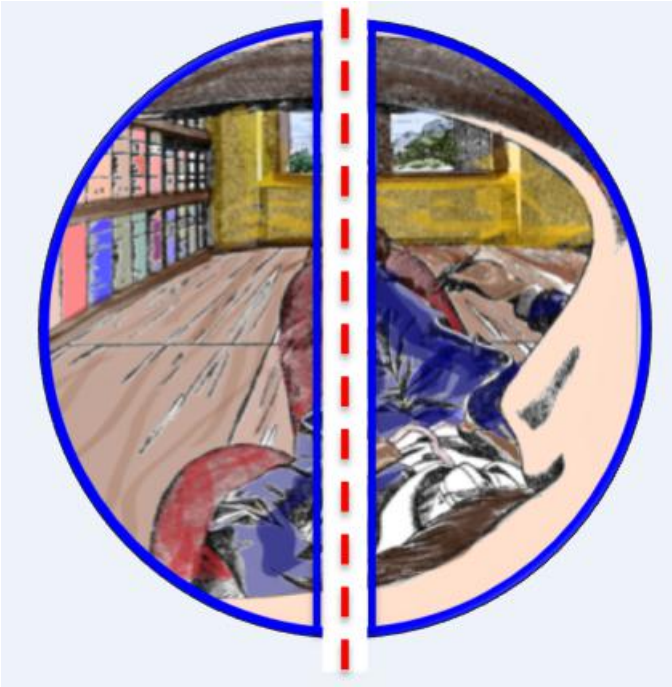
经验（experience）的基本性质。

4. Integration（整合）

意识是统一的：每个经验都不能还原为非相互依存的不相交的现象特质子集。

例如，看见空白纸中间写着一个单词“BECAUSE”的经验，不能还原为“在左侧看见‘BE’的经验”加上“在右侧看见‘CAUSE’的经验”。又如，看见一本蓝色的书。

整体大于部分和。

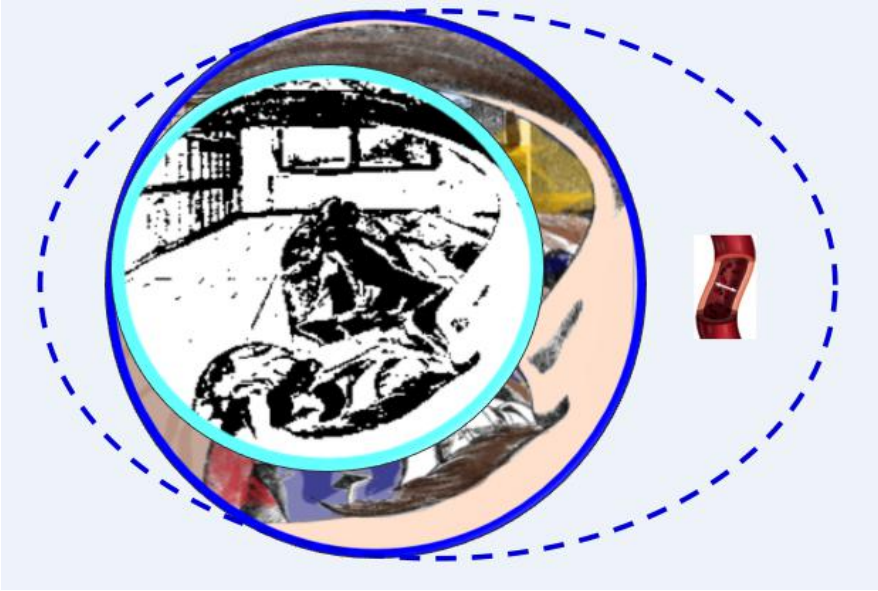


关于意识的公理

经验（experience）的基本性质。

5. Exclusion（排他）

意识在内容和时空粒度上是确定的：每个经验都有一个确定的现象特质集合，不会更少也不会更多；其流动的速度也确定，不会更快也不会更慢。



提纲

1 整合信息论简介

2 关于意识的公理

3 对意识的物理基质的假定

4 机制的系统与概念结构

5 整合信息论的局限性

对物理基质的假定 经验的物理基质（substrate）所需具备的性质

- **假定**（postulates）描述了，经验的物理基质（substrate）所需具备的性质。
- 对经验的每个基本性质，都有一个物理基质的因果性质与之对应。
- 这些假定，都是从现象学（phenomenology）到物理的推断，而不是反过来。因为，意识及其基本特性是确定的，而物理世界的存在和性质是我们意识中的猜想。

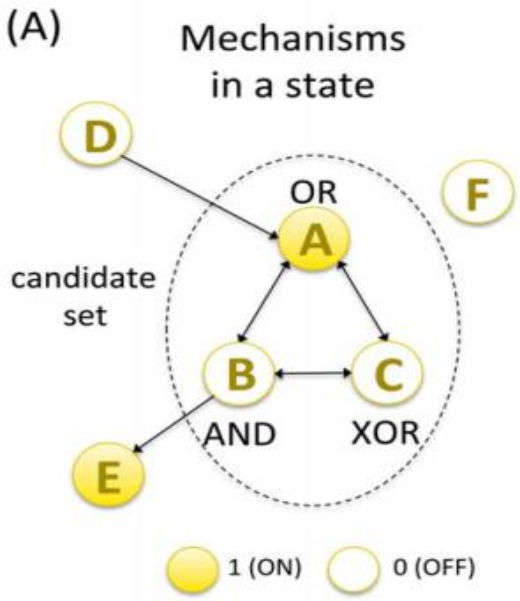
对物理基质的假定 经验的物理基质（substrate）所需具备的性质

- 为了简化讨论，在下面研究中，考虑由具有状态的元素（如，神经元或逻辑门）组成的物理系统。
- 这些“元素”应具备的特点：

  具有两个及以上的内部状态；

  具有能影响状态的输入；

  具有能被状态影响的输出。



回顾：

整合信息论中的公理　经验（experience）的基本性质。

回顾：

1. Intrinsic existence（内在的存在）

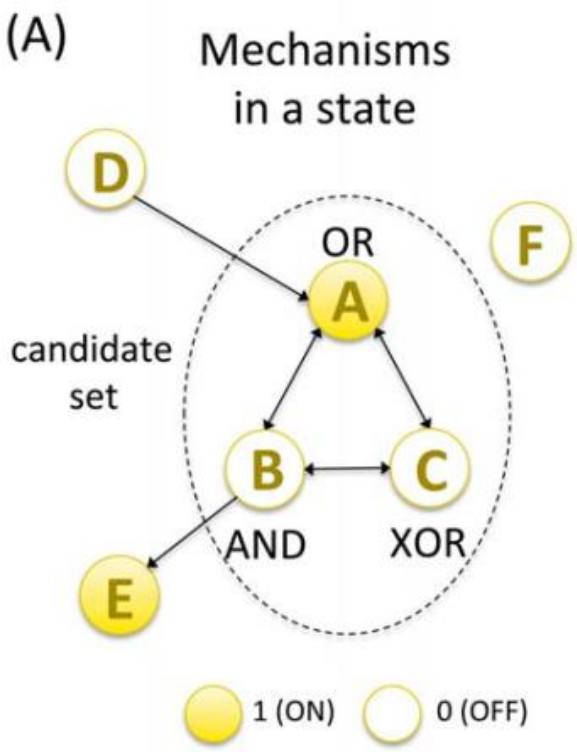
意识是存在的：每个经验都是真实的——实际上，我在此时此刻的经验存在（即该经验是真的），这是我唯一能够立即且绝对地确信的事情。

内在真实的（ intrinsically real or actual ）：我的经验的存在是独立于外部观察者的，是从它自身固有的角度来看的。

对物理基质的假定 经验的物理基质（substrate）所需具备的性质

1. Intrinsic existence（内在的存在）对应的假定

该物理系统必须内在存在（真的）：确切地说，它必须具有因果力。“make a difference”。 而且要从内在视角存在。



(B) Transition Probability Matrix (TPM)

t_0 \ t_1 ABC	000	100	010	110	001	101	011	111
000	1	0	0	0	0	0	0	0
100	0	0	0	0	1	0	0	0
010	0	0	0	0	0	1	0	0
110	0	1	0	0	0	0	0	0
001	0	1	0	0	0	0	0	0
101	0	0	0	0	0	0	0	1
011	0	0	0	0	0	1	0	0
111	0	0	0	1	0	0	0	0

current state s_0(ABC) = 100

图1

对物理基质的假定 经验的物理基质（substrate）所需具备的性质

小结：

1.内在存在：存在物理基质，有状态变化；

回顾：

整合信息论中的公理

经验（experience）的基本性质。

回顾：

2. Composition（组合）

意识是结构化的：每个经验由多重初级的或高阶的现象特质（*phenomenological distinctions*）组成。例如，在一个经验中，我能够区分书、蓝色、一本蓝色的书、左边、左边的一本蓝色的书，等等。

对物理基质的假定 经验的物理基质（substrate）所需具备的性质

2. Composition（组合）对应的假定

基础（elementary）的机制能够以各种各样的组合，结构化地形成更高阶的机制。

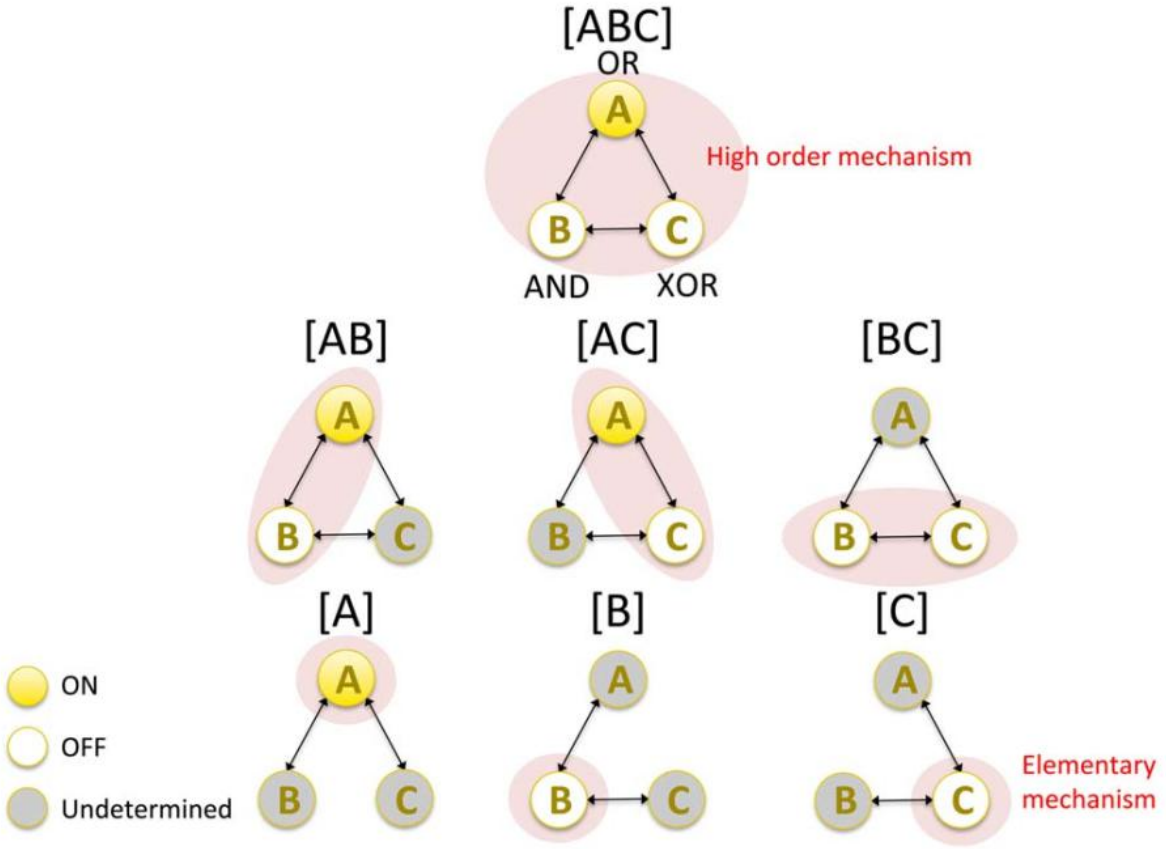


图2

对物理基质的假定 经验的物理基质（substrate）所需具备的性质

小结：

1.内在存在：存在物理基质，有状态变化；

2.组合：元素的幂集组成不同阶次的机制；

回顾：

整合信息论中的公理

经验（experience）的基本性质。

回顾：

3. Information（信息）

意识是特定的：每个经验都以它特殊的方式存在——是由特定的现象特质的一个特定集合组成的——因而区别于其它可能的经验（*differentiation*）。例如，当前我的经验的内容是看见一个房间，书等等，而与其它经验区别开（例如，没有书的场景）。

对物理基质的假定 经验的物理基质（substrate）所需具备的性质

3. Information（信息）对应的假定

从系统自身的视角，捕获“产生影响的差异（differences that make a difference）”。是因果的也是内在的。与香农的信息不同。

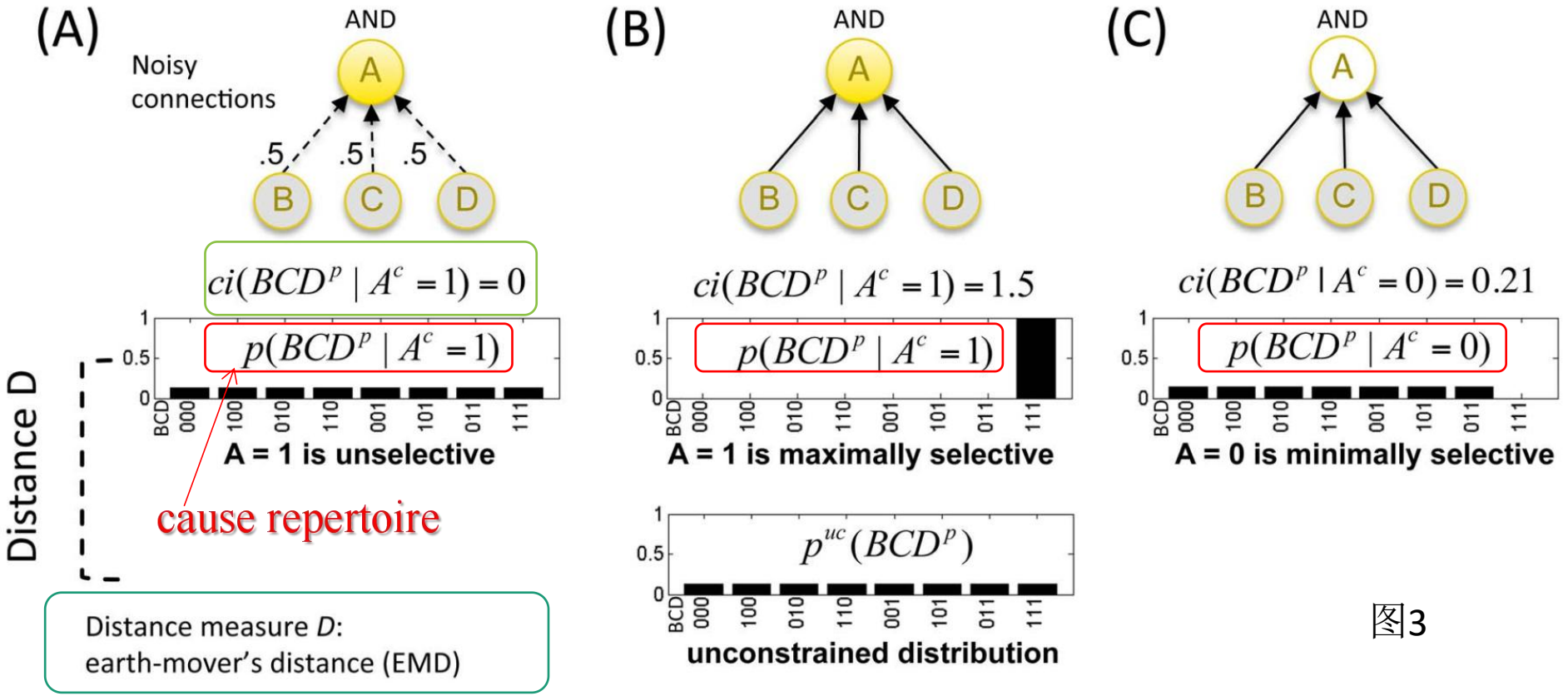


图3

对物理基质的假定 经验的物理基质（substrate）所需具备的性质

3. Information（信息）对应的假定

1. 算出概率分布
2. 计算ci和ei
3. 计算cei（因果信息）

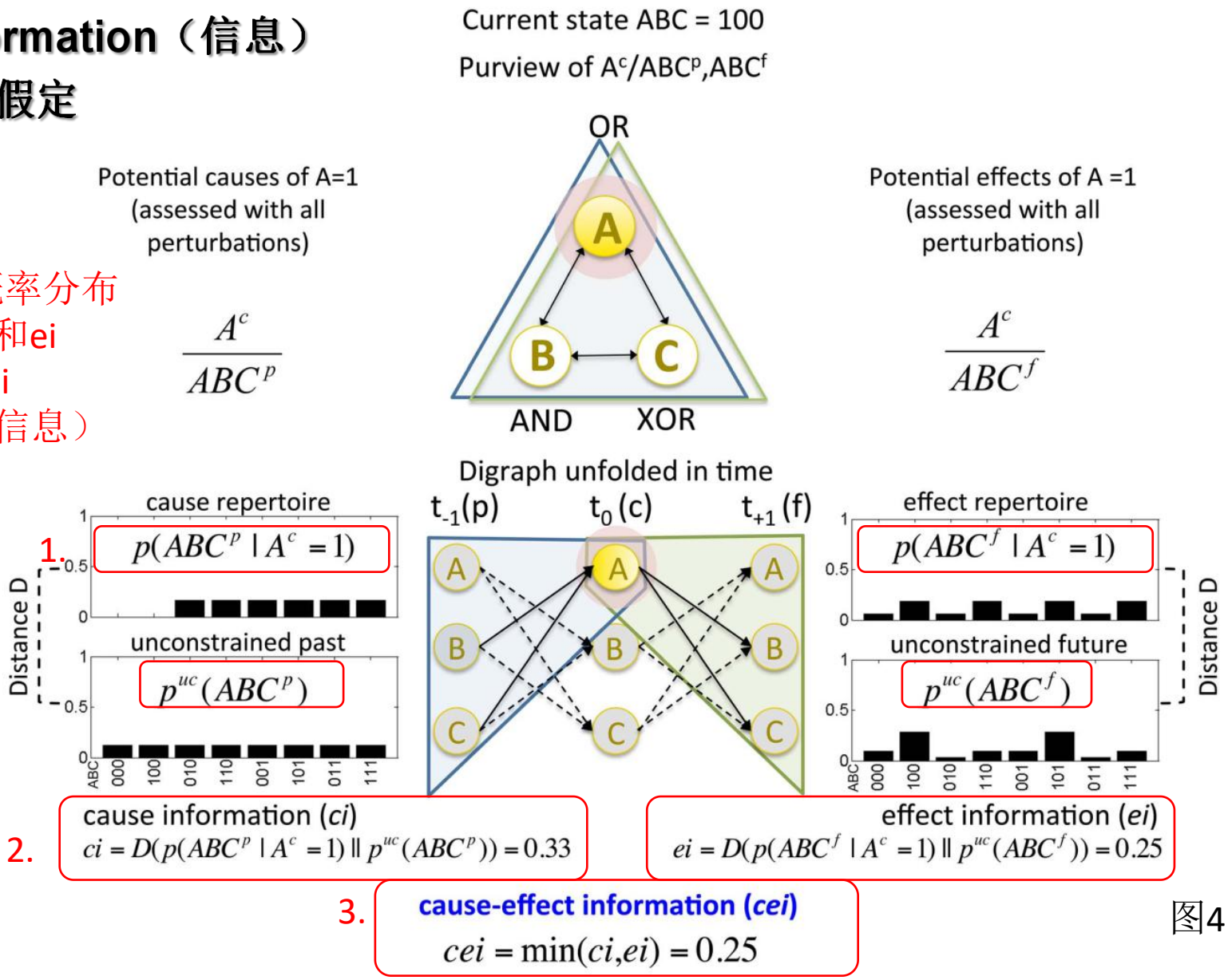


图4

对物理基质的假定 经验的物理基质（substrate）所需具备的性质

3. Information（信息）对应的假定

从系统自身的视角，捕获“产生影响的差异（differences that make a difference）”。是因果的也是内在的。与香农的信息不同。

1. 算出概率分布
2. 计算因信息ci和果信息ei
3. 计算cei （因果信息）

对物理基质的假定 经验的物理基质（substrate）所需具备的性质

小结：

1.内在存在：存在物理基质，有状态变化；

2.组合：元素的幂集组成不同阶次的机制；

3.信息

针对某个机制分析内在因果关系，通过概率分布的推土距离，计算**因果信息cei**。

回顾：

整合信息论中的公理

经验（experience）的基本性质。

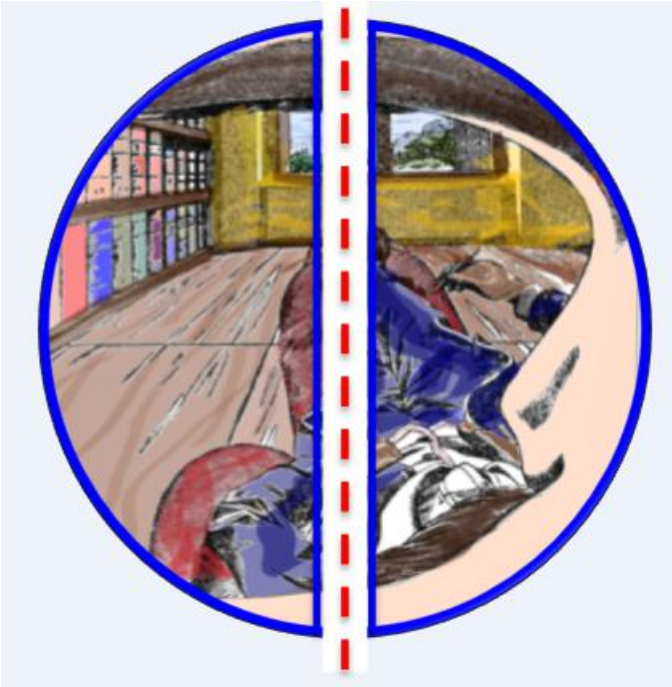
回顾：

4. Integration（整合）

意识是统一的：每个经验都不能还原为非相互依存的不相交的现象特质子集。

例如，看见空白纸中间写着一个单词“BECAUSE”的经验，不能还原为“在左侧看见‘BE’的经验”加上“在右侧看见‘CAUSE’的经验”。又如，看见一本蓝色的书。

整体大于部分和。



对物理基质的假定 经验的物理基质（substrate）所需具备的性质

4. Integration（整合）对应的假定

整合信息，是由整个机制产生的信息，超越了部分（parts）产生的信息。意味着，对信息而言，机制是不可规约的。

计算流程：

1. 对机制进行分割（很多种分割方式）
2. 对每个分割方式计算概率分布
3. 选取与整体机制的概率分布距离最小的分割方式，（minimum information partition, MIP）
4. 计算 φ^{MIP}（**整合信息**）

对物理基质的假定 经验的物理基质（substrate）所需具备的性质

4. In

信息

1. 对机制进行分割

Purview of $ABC^c/ABC^p,ABC^f$

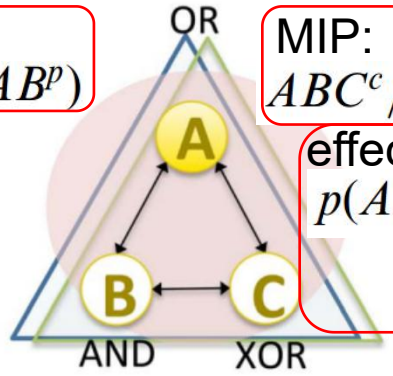
MIP:

$$ABC^c/ABC^p \rightarrow (AB^c/C^p)\times(C^c/AB^p)$$

cause repertoire:

$$p(ABC^p|ABC^c=100/\mathrm{MIP})=$$

$$p(C^p|AB^c=10)\times p(AB^p|C^c=0)$$



MIP:

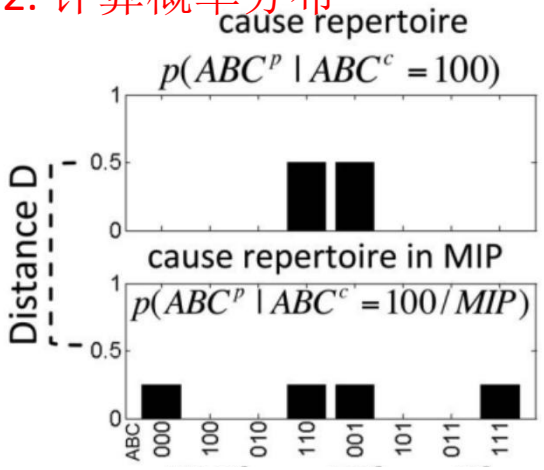
$$ABC^c/ABC^f \rightarrow (ABC^c/AC^f)\times([]/B^f)$$

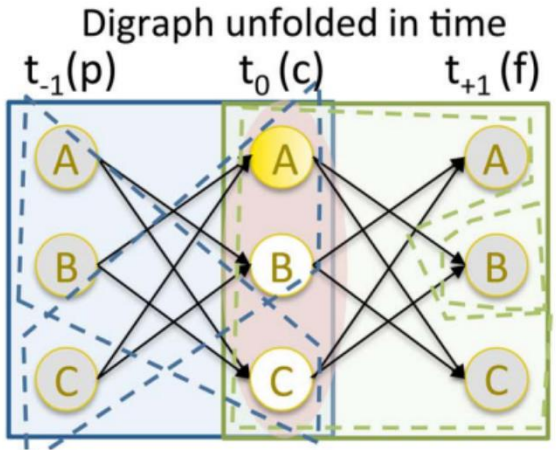
effect repertoire:

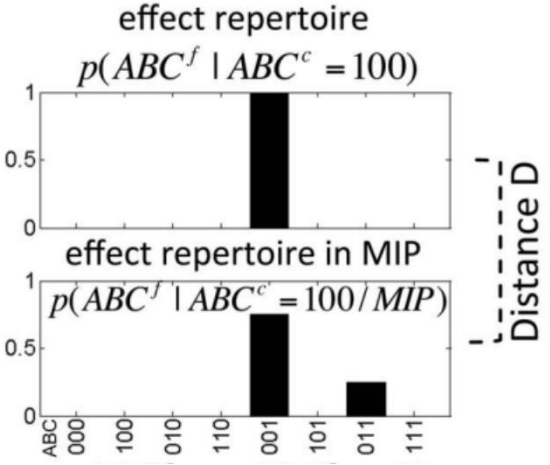
$$p(ABC^f|ABC^c=100/\mathrm{MIP})=$$

$$p(AC^f|ABC^c=100)\times p(B^f)$$

2. 计算概率分布







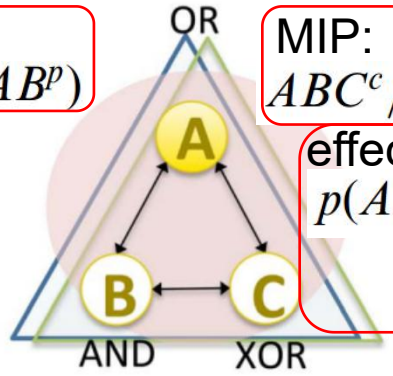
对物理基质的假定 经验的物理基质（substrate）所需具备的性质

1. 对机制进行分割

MIP:
$ABC^c/ABC^p \rightarrow (AB^c/C^p)\times(C^c/AB^p)$

cause repertoire:
$p(ABC^p|ABC^c=100/\mathrm{MIP})=$
$p(C^p|AB^c=10)\times p(AB^p|C^c=0)$

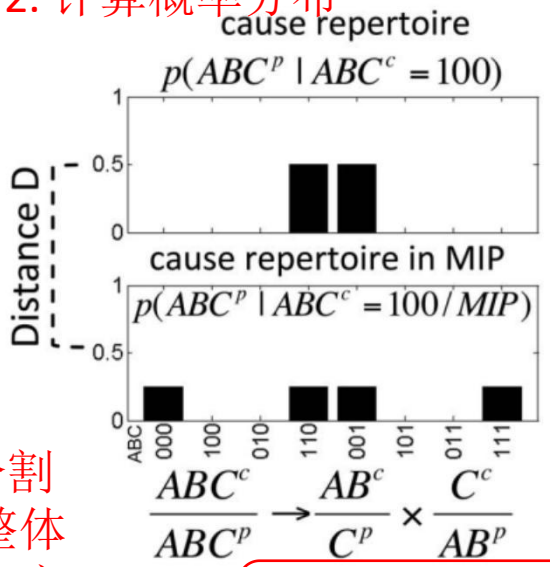
Purview of $ABC^c/ABC^p,ABC^f$

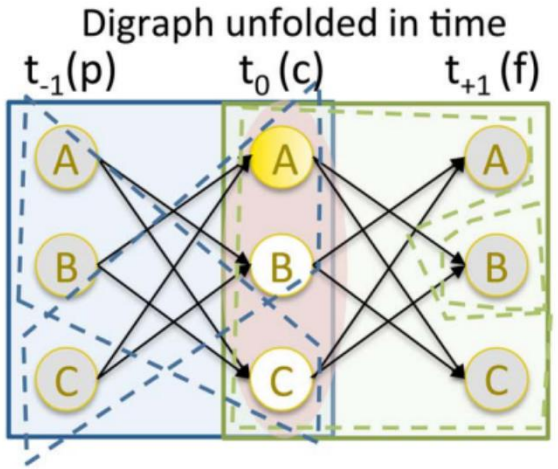


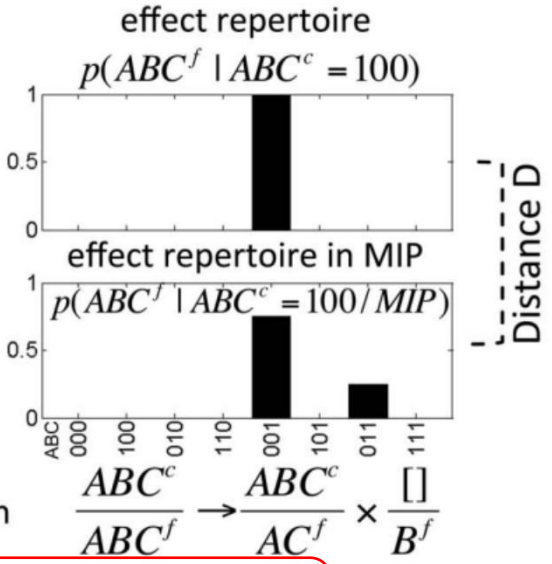
MIP:
$ABC^c/ABC^f \rightarrow (ABC^c/AC^f)\times([]/B^f)$

effect repertoire:
$p(ABC^f|ABC^c=100/\mathrm{MIP})=$
$p(AC^f|ABC^c=100)\times p(B^f)$

2. 计算概率分布







3.计算分割机制与整体机制的概率分布距离

$\varphi^{MIP}_{cause}(ABC^p \mid ABC^c=100)=D(p(ABC^p \mid ABC^c=100)\,\|\,p(ABC^p \mid ABC^c=100/MIP))=0.5$
$\varphi^{MIP}_{effect}(ABC^f \mid ABC^c=100)=D(p(ABC^f \mid ABC^c=100)\,\|\,p(ABC^f \mid ABC^c=100/MIP))=0.25$

4. 计算整合信息

integrated information (φ^{MIP})
$\varphi^{MIP}(ABC^{p,f} \mid ABC^c=100)=\min(\varphi^{MIP}_{cause}(ABC^p \mid ABC^c=100)),\varphi^{MIP}_{effect}(ABC^f \mid ABC^c=100))=0.25$

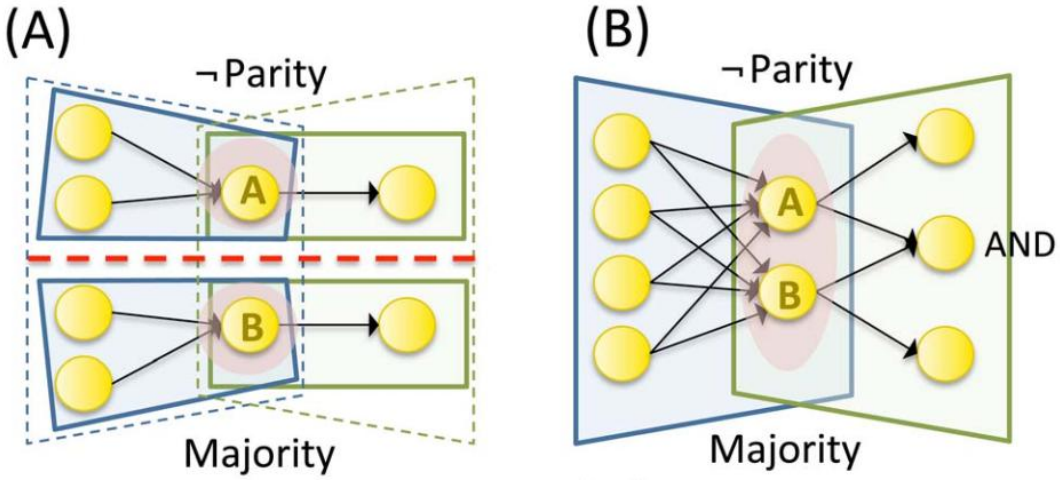
图6

对物理基质的假定 经验的物理基质（substrate）所需具备的性质

4. Integration（整合）对应的假定

整合信息，是由整个机制产生的信息，超越了部分（parts）产生的信息。意味着，对信息而言，机制是不可规约的。

讨论：根据IIT，不产生整合信息的机制，从系统的内在视角来看不存在。



Mechanism AB does not exist.

Mechanism AB exists. it plays an irreducible causal role: it picks up a difference that makes a difference to the system in a way that cannot be accounted for by its parts.

对物理基质的假定 经验的物理基质（substrate）所需具备的性质

小结：

1.内在存在：存在物理基质，有状态变化。

2.组合：元素的幂集组成不同阶次的机制。

3.信息

针对某个机制分析内在因果关系，通过概率分布的推土距离，计算**因果信息cei**。

4.整合

针对某个机制进行分割，评估获得最小信息分割方式，计算**整合信息** φ^{MIP} 。

回顾：

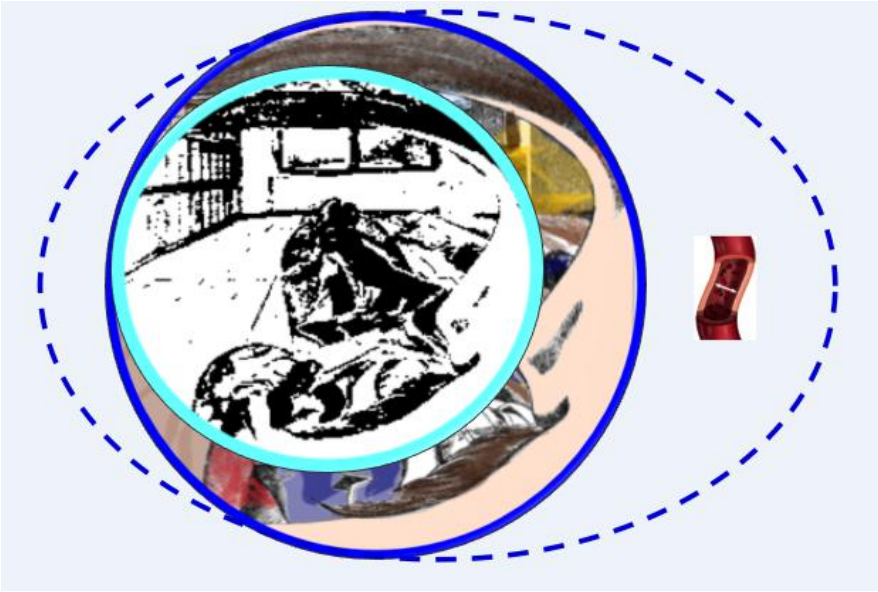
整合信息论中的公理

经验（experience）的基本性质。

回顾：

5. Exclusion（排他）

意识在内容和时空粒度上是确定的：每个经验都有一个确定的现象特质集合，不会更少也不会更多；其流动的速度也确定，不会更快也不会更慢。



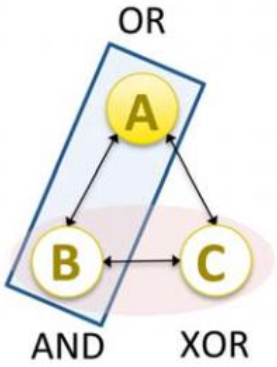
对物理基质的假定 经验的物理基质（substrate）所需具备的性质

5. Exclusion（排他）对应的假定

机制层面的排他假定是指，每个机制只有一个因和一个果（核心因和核心果），即最大不可规约的那个。

例：
针对机制BC，当前值BC=00。

在过去状态的幂集上计算因的整合信息。



Purviews of mechanism BC (Past)

$$BC^c / A^p \rightarrow (C^c / []) \times (B^c / A^p) \quad \varphi_{cause} = 0$$
$$BC^c / B^p \rightarrow (C^c / []) \times (B^c / B^p) \quad \varphi_{cause} = 0$$
$$BC^c / C^p \rightarrow (C^c / []) \times (B^c / C^p) \quad \varphi_{cause} = 0$$
$$BC^c / AB^p \rightarrow (C^c / AB^p) \times (B^c / []) \quad \varphi^{Max}_{cause} = 0.33$$
$$BC^c / BC^p \rightarrow (C^c / []) \times (B^c / BC^p) \quad \varphi_{cause} = 0$$
$$BC^c / AC^p \rightarrow (C^c / []) \times (B^c / AC^p) \quad \varphi_{cause} = 0$$
$$BC^c / ABC^p \rightarrow ([] / C^p) \times (BC^c / AB^p) \quad \varphi_{cause} = 0.17$$

Core cause 核心因
具有最大整合信息的因

图8

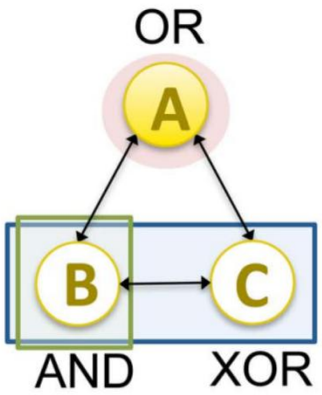
核心果用同样的方法确定。
维数开始大了。

对物理基质的假定 经验的物理基质（substrate）所需具备的性质

5. Exclusion（排他）对应的假定

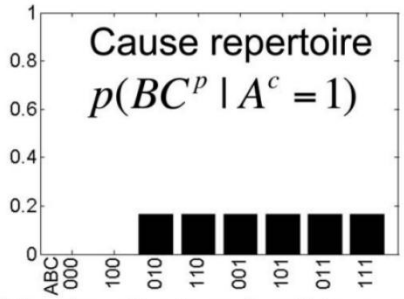
具有最大不可归约因果（MICE）的机制，构成了概念(Concept)。

每个概念有对应的 φ^{Max}



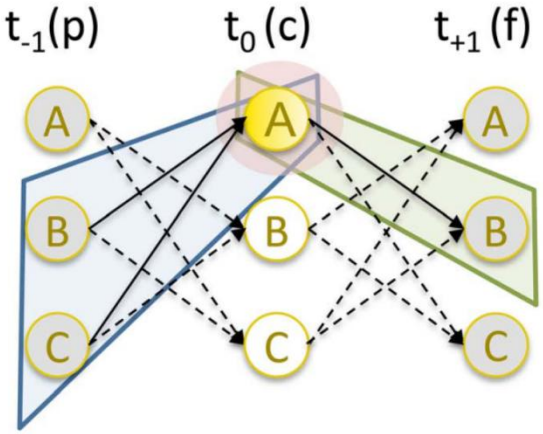
$\frac{A^c}{BC^p}$

Core cause



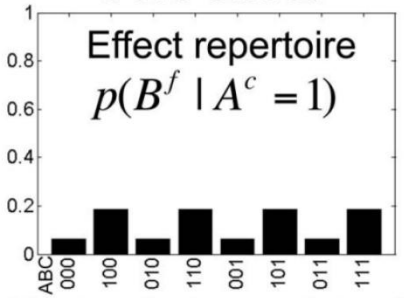
Maximally irreducible cause

$\varphi^{Max}_{cause}(P \mid A^c = 1) = 0.17$



$\frac{A^c}{B^f}$

Core effect



Maximally irreducible effect

$\varphi^{Max}_{effect}(F \mid A^c = 1) = 0.25$

maximally integrated information (φ^{Max})

$$\varphi^{Max}(A^c = 1) = \min(\varphi^{Max}_{cause}(P \mid A^c = 1), \varphi^{Max}_{effect}(F \mid A^c = 1)) = 0.17$$

P/F: Power set of all past/future purviews

all other functions are excluded (treated as extrinsic "noise")

图9

对物理基质的假定 经验的物理基质（substrate）所需具备的性质

5. Exclusion（排他）对应的假定

机制层面的排他假定说的是，每个机制只有一个因和一个果（核心因和核心果）。最大不可规约的那个。

计算流程：

1. 在因/果幂集上取一个元素；
2. 对机制进行分割，计算得到最小分割MIP与对应的整合信息（即，假定4的计算过程）；
3. 比较幂集中每个因/果的整合信息，得到核心因/果；
4. 计算得到**最大不可归约因果——概念。**

对物理基质的假定 经验的物理基质（substrate）所需具备的性质

小结：

1.内在存在：存在物理基质，有状态变化。

2.组合：**元素的幂集**组成不同阶次的机制。

3.信息

针对某个机制分析内在因果关系，通过概率分布的推土距离，计算**因果信息cei**。

4.整合

针对某个机制进行分割，评估获得最小信息分割方式，计算**整合信息** φ^{MIP} 。

5.排他

针对某个机制，在过去（将来）状态的幂集上计算因（果）的整合信息，选最大的得到核心因/果，进而得到**最大整合信息** φ^{Max}，对应的机制即为**概念**。

提纲

1 整合信息论简介

2 关于意识的公理

3 对意识的物理基质的假定

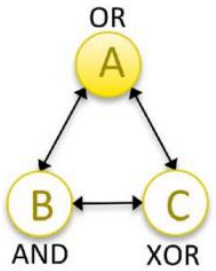
4 机制的系统与概念结构

5 整合信息论的局限性

机制的系统（Systems of mechanisms） 1. 信息

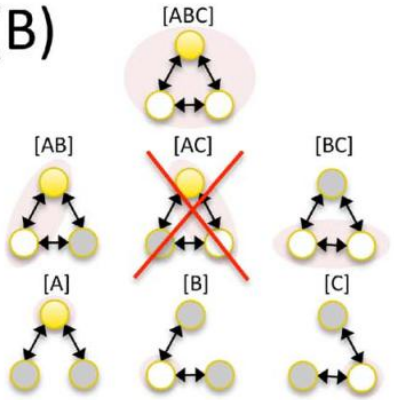
ABC的机制集合产生的所有概念（维数再增大，分割_^幂集_^幂集）

(A)



A set of elements
in their current state s_0

(B)

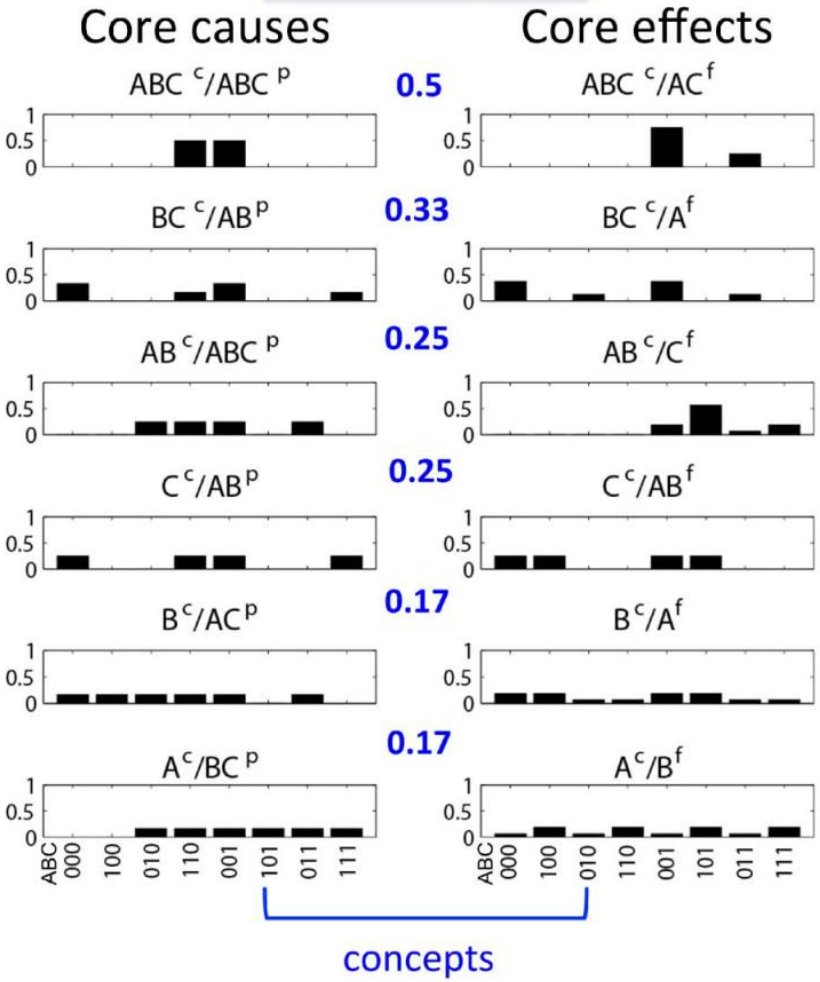


Power set of mechanisms

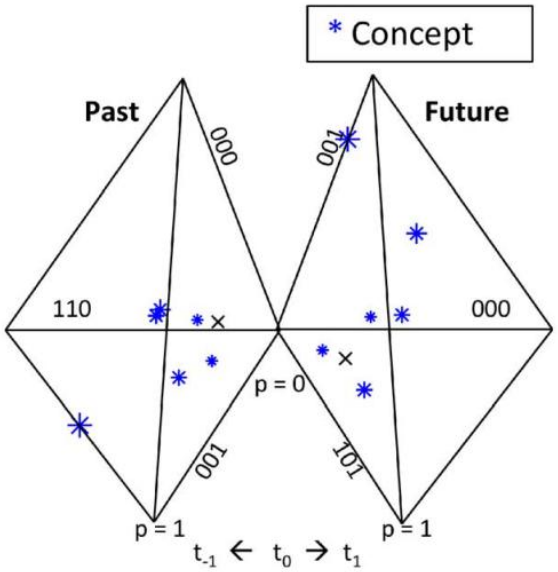
只有机制AC不能产生概念，因为其整合信息=0

(C)

$$\varphi^{Max}(P,F \mid X = s_0)$$



(D)



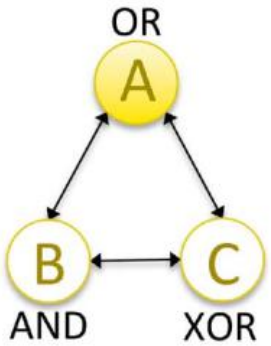
Conceptual structure
plotted as a **constellation**
of concepts
in concept space

概念结构与概念空间

图10

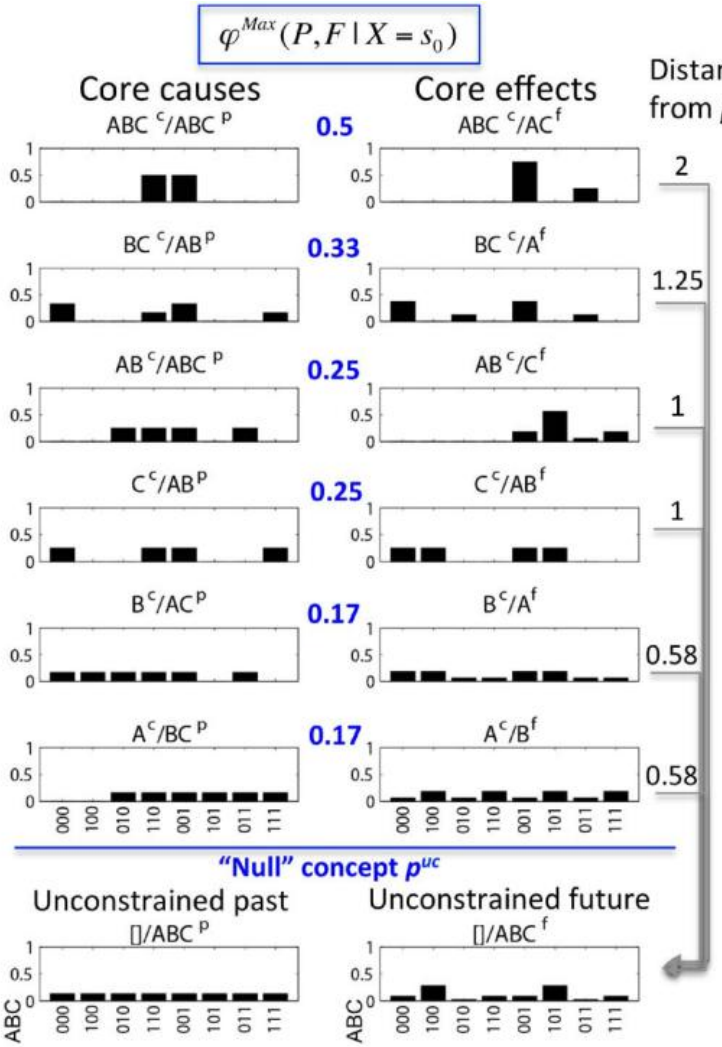
机制的系统（Systems of mechanisms） 1.信息

(A)



$s_0(ABC) = 100$

$$CI(C|s_0) = \mathrm{D}(C\|p^{uc}) = 2.11$$



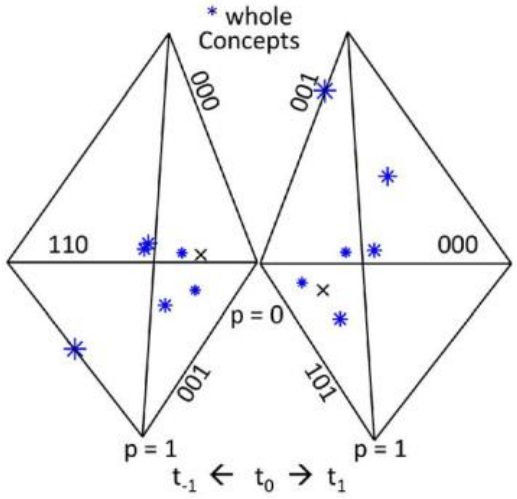
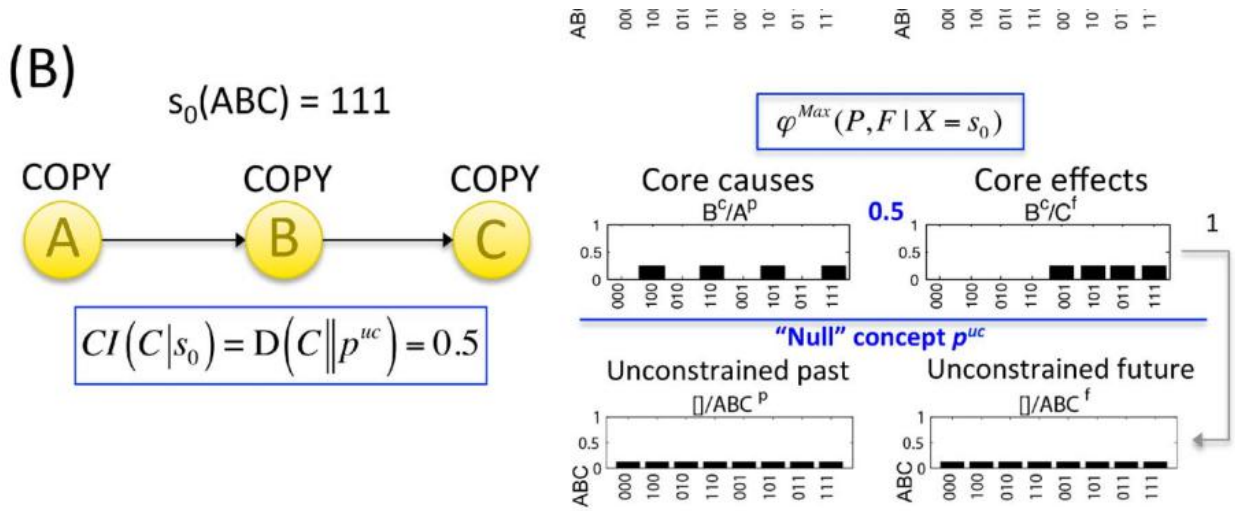


图11A

在系统层，与机制层因果信息（cei）对应地计算**概念信息（CI）**。

CI=叠加（每个概念与空概念之间的EMD距离* φ^{Max} 值）

机制的系统（Systems of mechanisms）　1. 信息



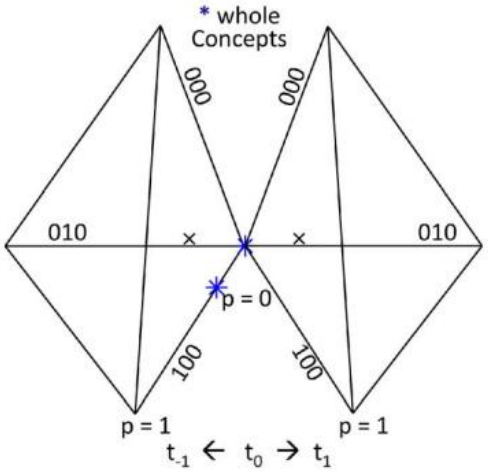


图11B

在系统层，与机制层因果信息（cei）对应地计算概念信息（CI）。
CI=叠加（每个概念与puc之间的EMD距离* φ^{Max} 值）

初级或更高阶概念丰富的星座，产生高CI。
相反，由一个初级机制组成的系统产生最少的CI。

机制的系统（Systems of mechanisms） 2. 整合

只有整合的概念结构才能产生意识。

整合概念信息 Φ^{MIP}：最小分割概念星座与完整概念星座的距离。

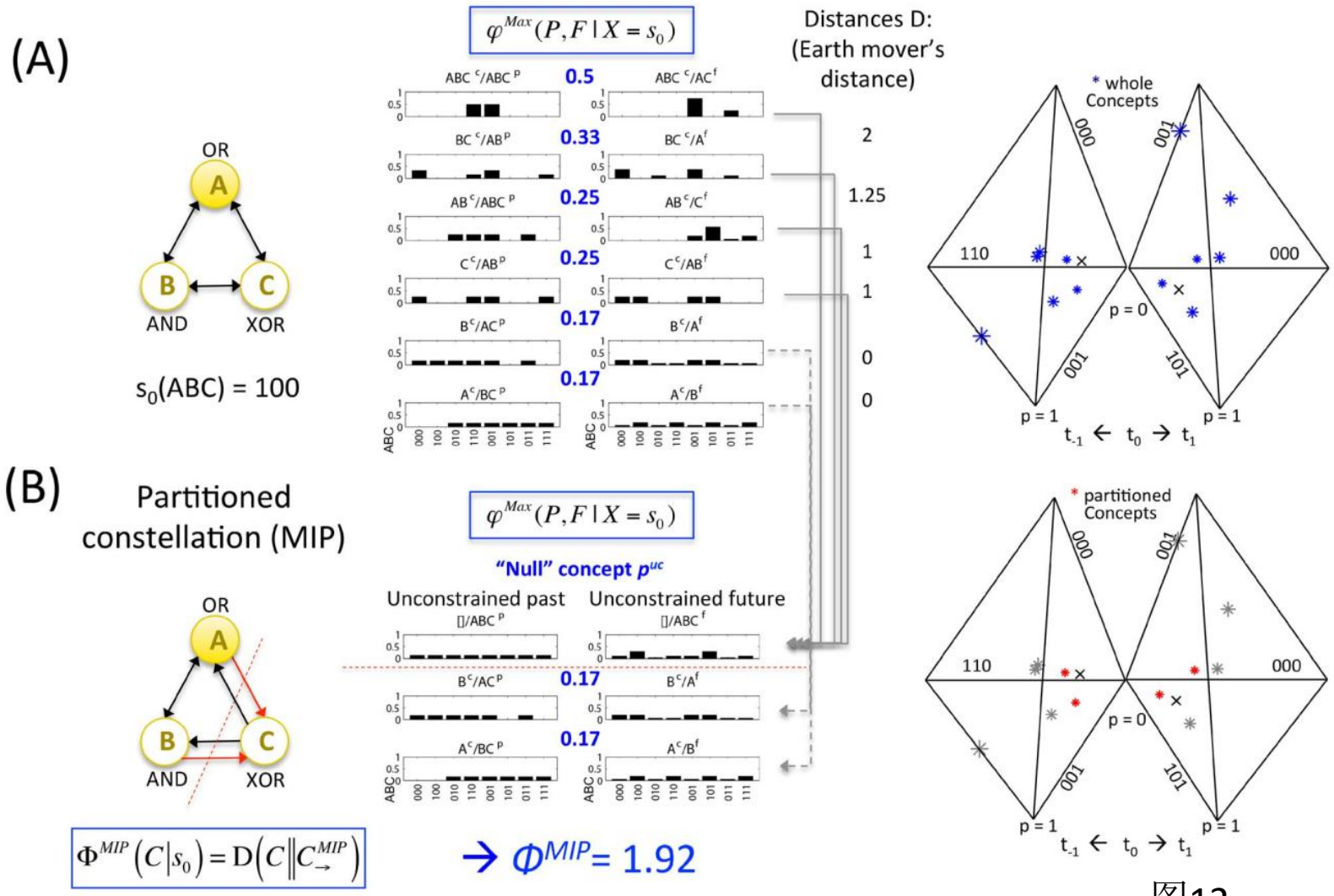
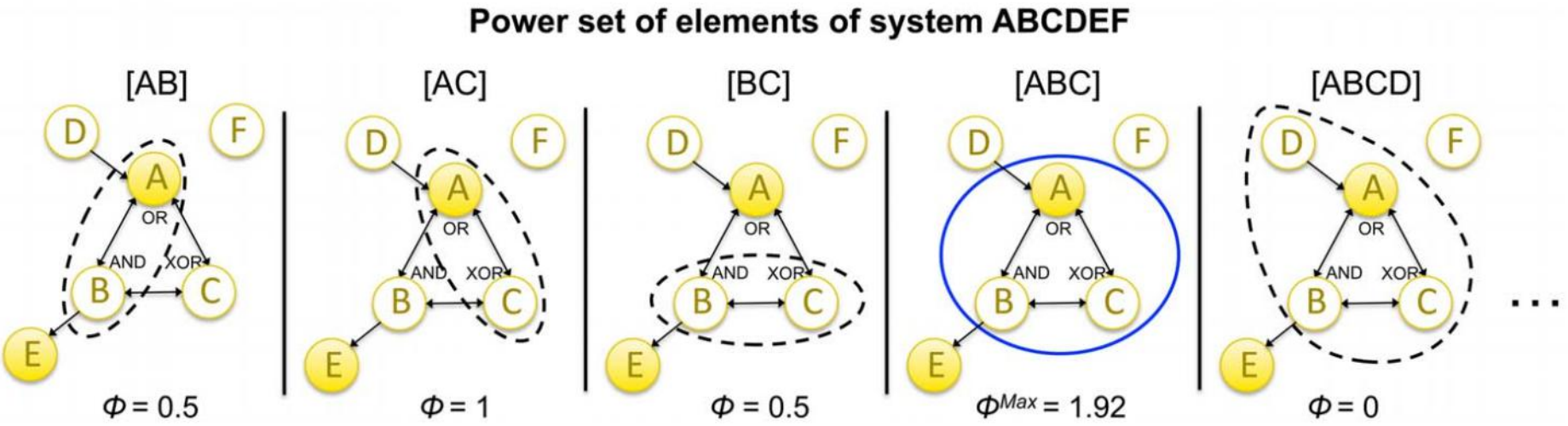


图12

机制的系统（Systems of mechanisms） 3. 排他

复合体（complex）：系统中产生最大整合概念信息的元素集合。
形成**最大不可归约概念结构（MICS）**。 Φ^{Max}



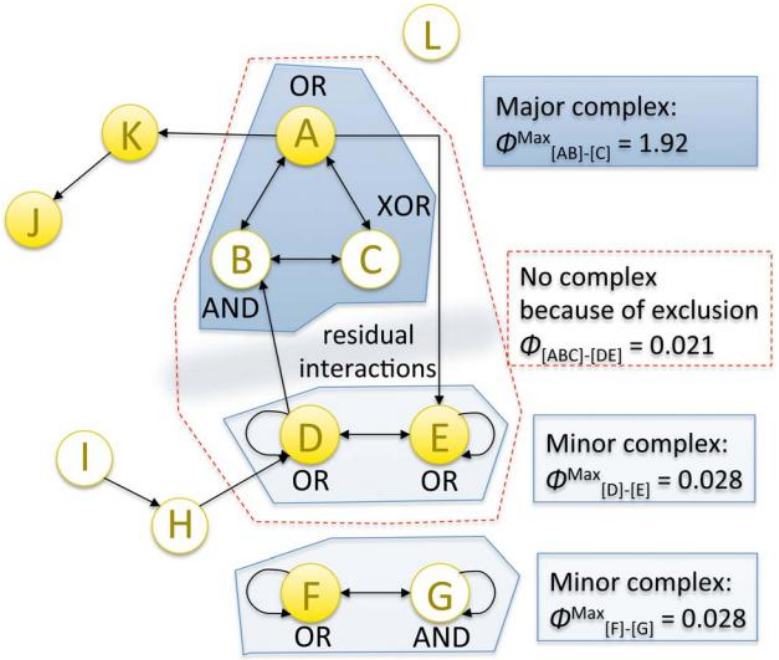
对系统元素的幂集，计算整合概念信息。
只有一个元素集合会构成复合体（complex），具有**最大整合概念信息** Φ^{Max} 。（图中为集合ABC）
经验（意识）即为一个complex的内在性质。经验等同于complex的最大不可归约概念结构（MICS）。

图13

机制的系统（Systems of mechanisms）

举例与讨论：

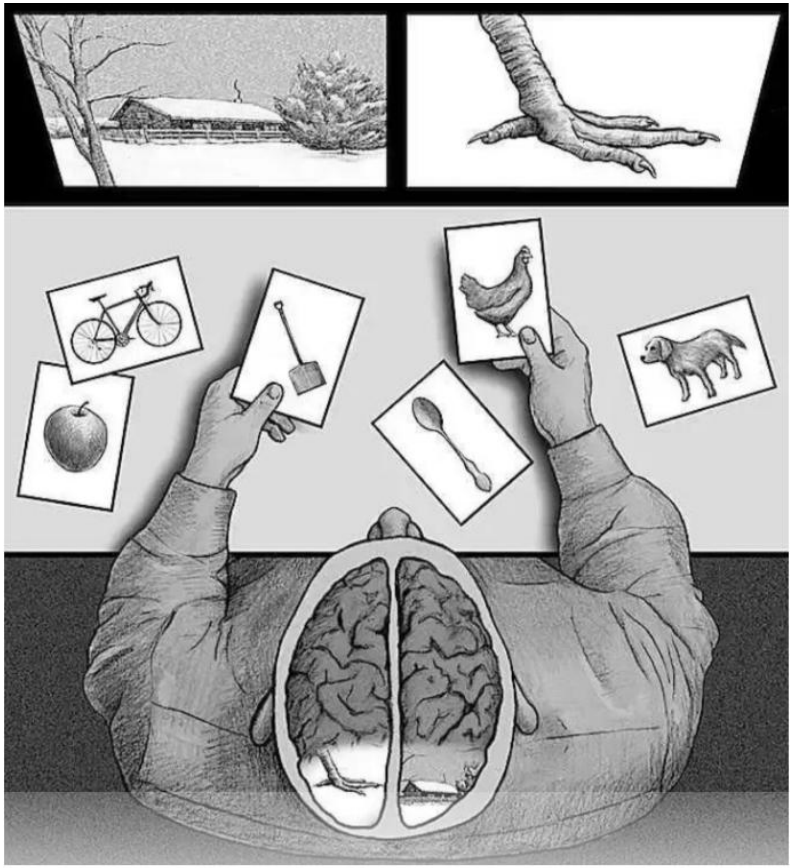
一个系统可以压缩为一个主要复合体和数个次要复合体。



机制的系统（Systems of mechanisms）

举例与讨论：

一个系统可以压缩为一个主要复合体和数个次要复合体。

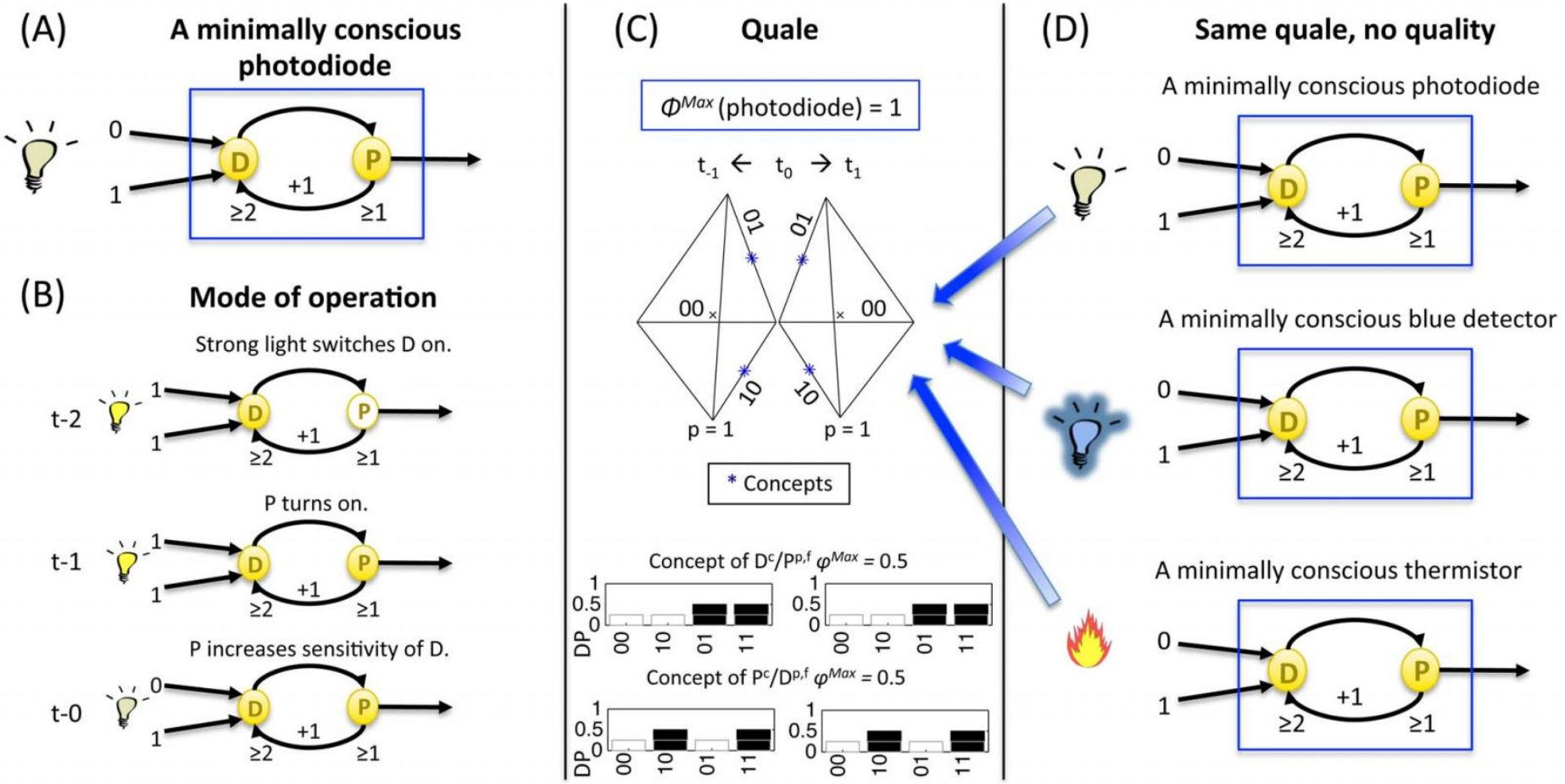


大脑的例子。
经过裂脑手术，原来的一个主要复合体会分割为两个主要复合体（均具有高的最大整合概念信息　　。
有实例证实，这种情况下意识会被分割为两个独立的互不知晓的意识。

机制的系统（Systems of mechanisms）

举例与讨论：

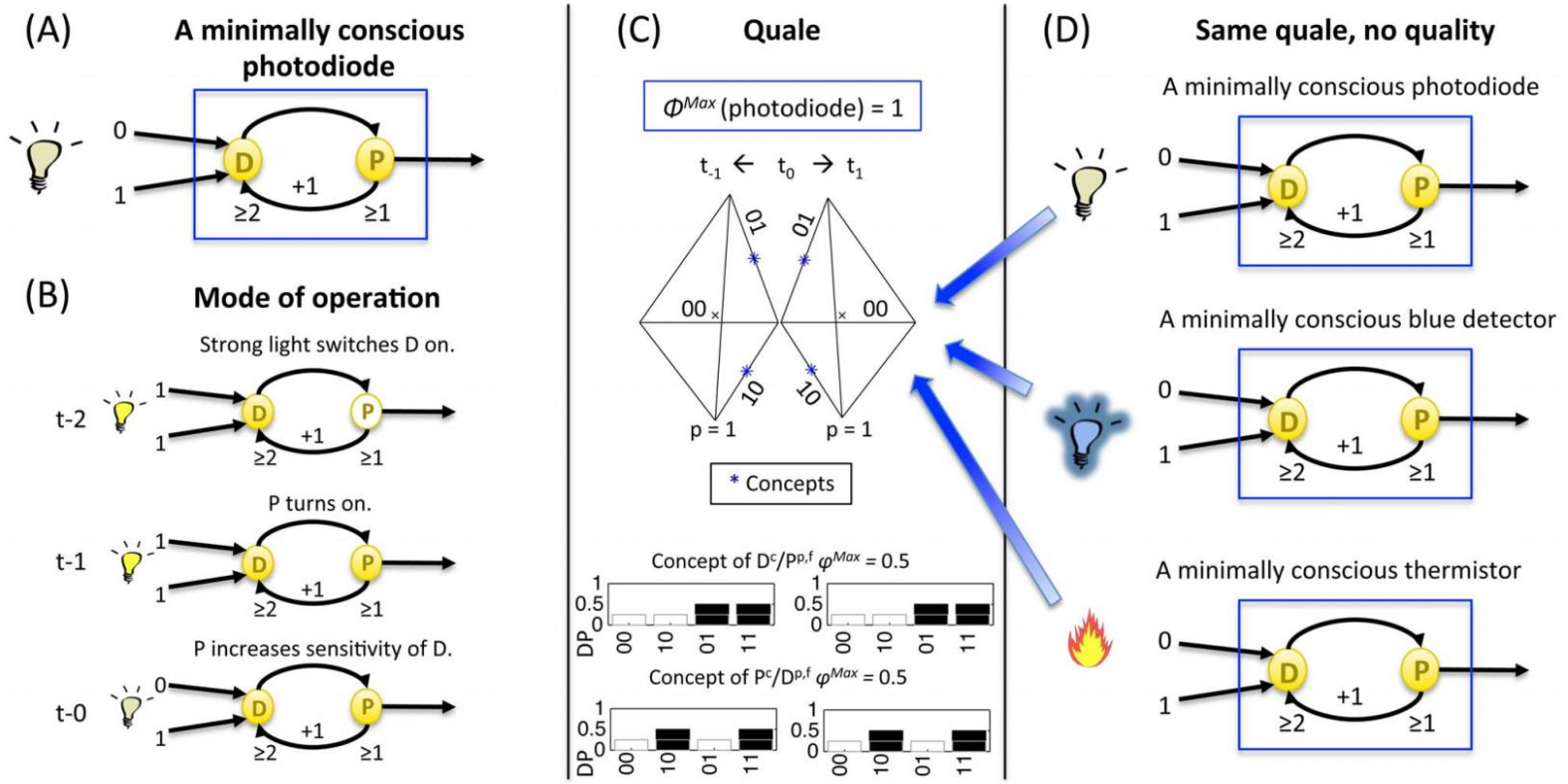
简单系统也可以有意识：“最小意识”光电二极管



机制的系统（Systems of mechanisms）

举例与讨论：

简单系统也可以有意识：“最小意识”光电二极管

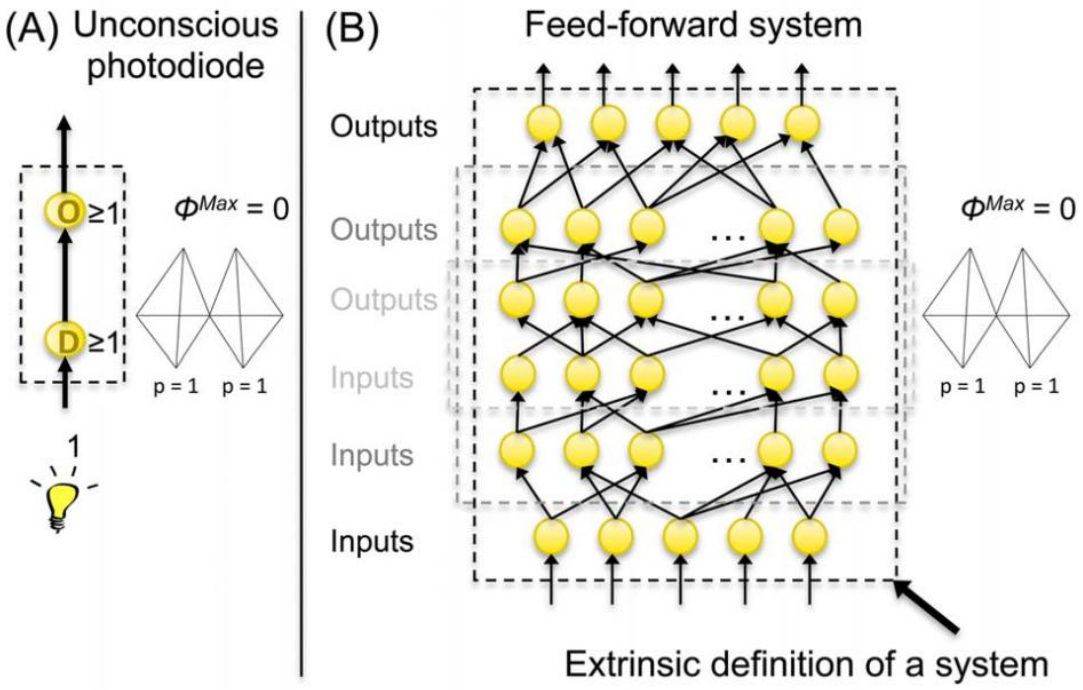


从外部观测者的角度，光电二极管检测了光，从内在角度看，
并没有达到这个程度的意识，它只是反映了元素之间的因果关系。

机制的系统（Systems of mechanisms）

举例与讨论：

复杂的系统可能没有意识：一个“僵尸”前馈网络。



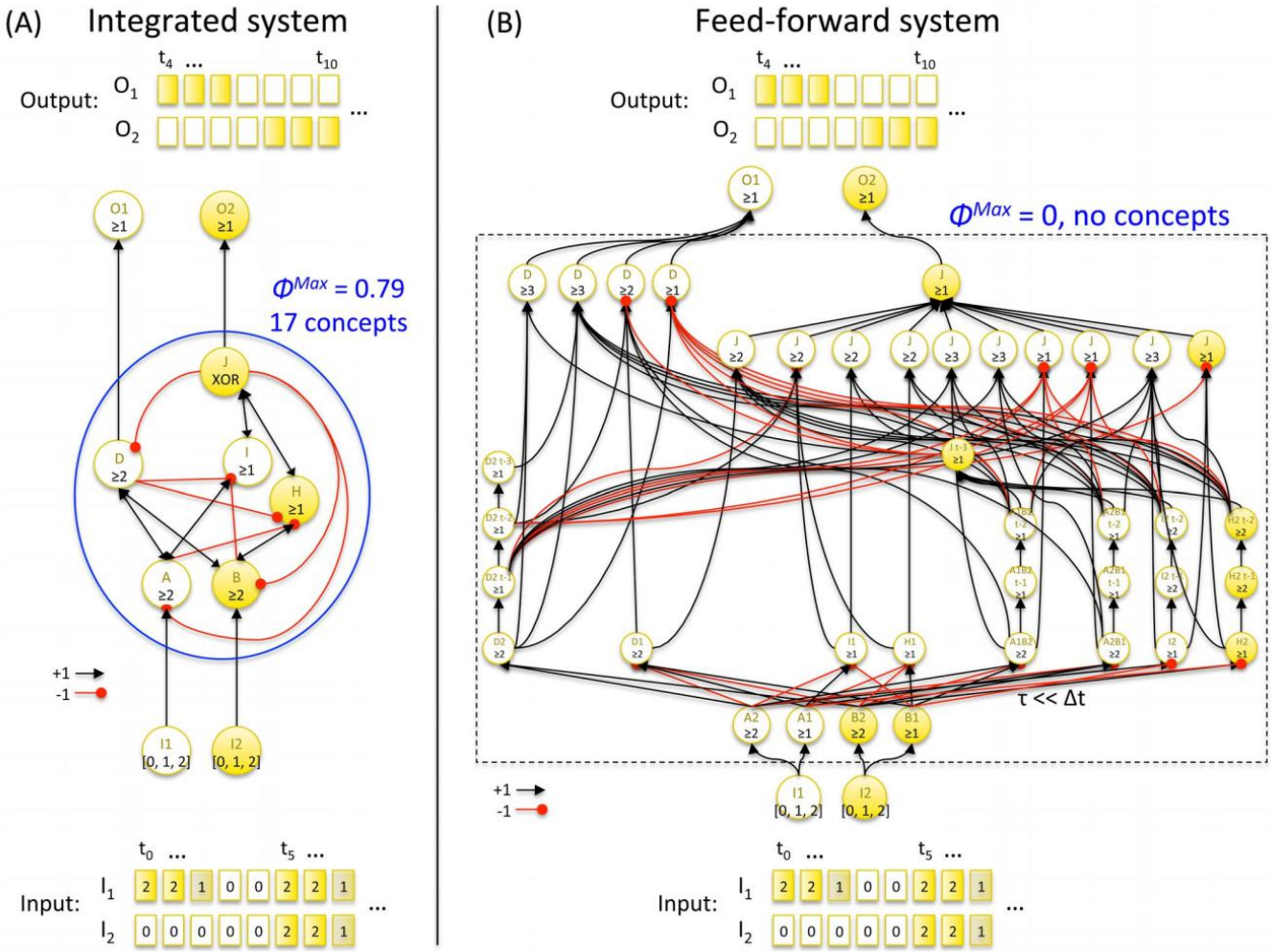
反馈对意识很重要。

人工神经网络？深度学习？

机制的系统（Systems of mechanisms）

举例与讨论：

有意识的复合体与无意识的“僵尸”系统可能在功能上等价。



提纲

1 整合信息论简介

2 关于意识的公理

3 对意识的物理基质的假定

4 机制的系统与概念结构

5 整合信息论的局限性与展望

整合信息论的局限性与展望

- 理论本身还有待完善的地方：
  - 没有讨论MICS与现象学特定方面之间的关系；
  - 时空颗粒度尺寸；
  - 将整合信息与因果作为同一个事物，有待深入考察

- 机制在时间和空间上是离散的。
- 转移概率已知。生物系统限于可观测的状态。
- 目前的分析对大于一打元素的系统不可行。计算量大，组合爆炸。

可能的解决方法；启发式方法，实验方法可能能用于测试该理论的预测。