内生参照场与体验归属

一种多维过程假说及其思想实验检验

刘烘炬（Hongju Liu）

首次公开版 v1.0 · 2026 年 9 月 20 日

English title: Endogenous Reference Fields and Experiential Attribution: A Multidimensional Process Hypothesis

论文编号： TA-TR-2026-10

DOI： 10.5281/zenodo.22852885

摘要

意识理论不仅要说明体验可能由何种过程实现，还必须在不同描述、硬件、训练阶段和时间尺度下给出一致归属。本文提出内生参照场（Endogenous Reference Field, ERF）假说：内容的因果作用与其当前处理参照相互构成，这种实际运行的组织具有现象方面。本文以多维组织画像分析角色调节、返回影响、分解距离、内容分化与时间关系，不由智能成绩、权重更新或任一非零耦合判定意识有无。形式分析得到两项具体分离结果：一族有限机制具有相同的最大调节幅度，却有不同的可区分响应数；另两套机制对全部通常输入产生相同输出，却在共同内部干预接口下具有不同参照画像。为处理主体范围，本文定义相对于响应任务和省略协议的最小机制支撑族，保留多解，并区分联合任务、实现冗余与新增组织。响应分化采用分辨率画像，证明冗余标记不增分化以及小响应变化下的位移界。二十一个思想实验进一步审查复制、回放、训练损失、异步运算及脑机替代。贡献在于这一参照构成方案及其分离与一致性论证，不主张首创意识多维性、主体重叠或非生物实现。自然物理粒度、精确主体数及感受质和情感效价的完整对应仍未解决；数学结果约束假说，不独自证明体验存在。

关键词： 意识；主观体验；参照构成；主体边界；人工智能；多维意识；思想实验

Abstract

Consciousness theories must address not only the processes that might realize experience, but also the attribution of experience across descriptions, implementations, learning stages, and temporal scales. The endogenous reference field hypothesis proposes that content-dependent processing and its operative reference constraints reciprocally constitute a running organization with a phenomenal aspect. The paper develops a multidimensional organizational profile of role modulation, reference-mediated influence, decomposition distance, response differentiation, and temporal relations. Two finite constructions establish specific separations. A family of mechanisms has identical maximal modulation strengths but different numbers of distinguishable responses; a second pair agrees on every ordinary input-output sequence but differs under shared internal intervention

interfaces. Response differentiation is described by a resolution profile, with invariance under redundant labels and a stability bound under small response changes. A family of inclusion-minimal mechanism supports makes organizational scope explicit relative to a declared response task and omission protocol, while avoiding the identification of support counts with subject counts. Twenty-one thought experiments examine copying, replay, learning losses, asynchronous implementation, gradual replacement, and brain-computer offloading. The contribution is a specific constructive hypothesis and its separation and consistency arguments, rather than a claim to have originated multidimensional consciousness, overlapping subjects, or nonbiological realization. The framework permits graded organizational variation without converting measurement tolerances into consciousness thresholds. It does not derive phenomenal existence, valence, or a unique natural grain from the mathematics. No empirical or computational research experiments are reported.

Keywords: consciousness; subjective experience; endogenous reference; response differentiation; subject boundaries; thought experiments

研究类型： 理论与概念分析。本文未进行人体、动物或计算机模拟实验。

1. 问题与论证目标

设想一个人的脑功能逐步由计算机承担。计算机可以处于颅内，也可以经脑机接口连接；它可以更换神经元，也可以只是提供外部记忆。功能转移的百分比没有直接回答体验是否延续、形成分支或出现另一主体。把问题换成“机器有没有意识”，同样略去了组织范围和时间关系。

人工系统把这些区分变得格外清楚。一组固定权重可以服务许多彼此独立的运行实例；同一次任务又可以由多个模型、记忆和调度程序共同完成。训练改变参数，部署改变激活和上下文；两者都可以涉及动态过程，但工程标签不能自动决定现象归属。与此同时，平静观看和冥想要求理论容纳不发生实际冲突、不学习新知识、不对外行动的体验。

本文提出一个正面的过程假说，同时限制其论证范围。核心不是证明某个现有模型或细胞已经具有体验，而是明确：若体验由参照构成实现，一个理论必须怎样面对不同实现、连续变化及重复计数。本文完成的工作包括一组组织定义、三个基础一致性命题、响应分化与有限支撑的性质，以及二十一个思想实验的统一分析。心理与物理的桥接被明确标为假说，不伪装成数学定义的结论。

“体验强弱”至少需要拆成内容分化、联系范围、时间维持、自我表征和情感效价等方面。智能则是任务能力。两个系统的任务成绩可以接近，体验相关组织却不同；一个系统也可能感觉鲜明而不擅长抽象推理。这里没有预设一条把全部生命与机器排成直线的意识阶梯。

2. 理论近邻与贡献边界

本假说的组成概念大多有明确前例。整合世界建模理论 IWMT 已将视角参照、时空因果连贯性及递归组织放在中心；投射意识模型进一步用几何参照处理感知、记忆与想象。因而，“内容依赖参照”不是本稿首次提出。本文不要求普遍存在三维空间场景，也不从几何投射结构推出意识。[1] [2]

激进可塑性理论已讨论学习、内部表征及其再表征；Merker 已提出非语言的第一人称定位机制。本文区分形成过程与当前实现，不把每一体验时刻实际发生学习作为条件；这种区别也不意味着前人主张“每个时刻都必须更新权重”。[3] [4]

IIT 4.0 以排他性原则选择最大基底，提供了明确的边界立场；本文则保留多个实际组织可能重叠的立场。但重叠、非传递共同意识、相同内容类型与同一体验实例的区分，以及整体和部分同时有心灵，都已有专门讨论。本稿不把这些可能性重新命名为发现。5 6 7

扩展意识讨论也不只是笔记本比喻：Chalmers 已分析外部硅电路及构成与普通因果影响的差别。近期 Dual-Resolution Framework 将信息个体性、自维持与时间更新结合，并涉及 AI 记忆、训练和部署。Sendall 的工作则明确分析理论相对的因果时间窗、延迟及边界，并声明不提供正面意识理论。本文的区别在于提出自己的分级参照画像和桥接假说，不把生存自主性作为普遍前提；这不构成对上述方案的全面优越性证明。8 9 10

因此，本文的原创性主张集中于一种参照构成方案及两类分离结果：最大调节幅度不能决定响应分化，通常输入输出函数也不能唯一决定参照画像。模型相对的最小支撑使范围判断附有可检查的协议。预测状态与等价响应的思想已有计算力学前例；本文的有限干预签名不等于其平稳过程因果状态，也不继承其最优性结论。30 分离集合、总变差和最小支撑等数学工具本身不主张首创；独立贡献应由具体构造、解释承诺及联合论证评审。

3. 概念区分：程度、实例与接续

3.1 多维程度不等于一个总分

本文不把意识简化成开关，也不把“意识强”视为未经定义的单一量。体验内容可以丰富而时间接续较短，可以较少语言自述而感觉鲜明。动物意识的多维研究已说明为何单一高低阶梯具有局限。11

在概念层面，可分别询问内容有多少可体验的差异、这些内容怎样共同呈现、体验在多长时间内接续、是否包含身体或自我表征，以及是否带有愉快或痛苦。空间范围则还要区分实现分布于何处与体验表征多大空间，不能把设备越大直接理解为体验越强。这些维度没有自然的统一加权办法；某一维度较弱也不代表其余维度都弱。

任务智能应另行记录。由此允许“擅长推理，但某些体验维度贫乏”和“感觉鲜明，但抽象推理能力有限”这两类组合。分开描述不意味着它们在现实中毫无关联，也不证明现有 AI 属于第一类。本假说不规定智能与体验普遍谁先谁后：具体形成顺序取决于实现机制，逻辑能力的增长与体验组织的增长不能仅凭概念互相推出。

必须进一步区分三个问题：组织是否允许多种程度，现象维度是否确实连续，以及我们是否掌握足够证据。本文主要支持前者并提出结构对应假说；它不由此证明任何变化都必然平滑，也不把证据不足称作微弱体验。允许渐变同样不蕴含一切物理系统必有非零体验。

3.2 三种关系

组织类型等价指两个描述保留相应的机制、干预和响应关系。它支持分类的一致性，不保证数值上是同一主体。

实际实现实例指发生在具体因果历史中的运行事件。两个完全同构的复制体仍是两次实现；对同一实例换变量名称，则没有凭此增加实现。

因果接续指后继实际承接哪些状态和组织。接续可以部分、分支或重新组合，不等于严格的个人数值同一性。

上述三种关系不得合并。否则，用等价类消除描述重复时，会误把两个复制体也合并；或者，把每个时间窗当成新实例时，会凭描述制造无穷主体。

3.3 三类参与关系

当下的**构成部件**实际执行内容转换或参照维持；**历时资源**保存未来可能调用的记录；**外部来源与背景条件**提供输入或使运行成为可能。同一部件可在不同阶段承担不同角色。固定权重或只读存储可以构成机制，故“能否写回”不是归属标准；供电和开发者具有因果作用，也不因此全部成为当前主体。

4. 内生参照场：分级的正面假说

4.1 机制描述

参照指一组实际影响内容因果作用的约束。改变参照，可以改变局部信号与后续处理的关系；相关内容的处理又会维持或改变参照。它可以分布实现，不要求一个中央观察者、显式的“我”、外部空间坐标或人类语义。

本文保留三个组织方面：内容作用随参照变化；内容经由参照影响其他内容的作用；这些关系在实际运行和具体时间范围内得到维持。这些方面允许强弱及分化程度，不以任意非零连接直接划成有或无。

这里的“内生”是参照由参与过程维护，并非与环境隔绝或能源自给。“场”是关系组织的简称，不指电磁场、新物理力或时间真的发生折叠。

4.2 桥接假说及其承诺

内生参照场假说： 实际运行中的参照构成具有现象方面；其内容分化、相互制约和时间展开实现体验的不同组织维度。不同维度不必同步变化；参数可塑性、任务智能、自我报告与负奖励均不能单独决定这些维度。

这一表述是积极的心理与物理的假说，而非已证明的同一性。它不把第 5 节的组织量直接命名为清醒度、红色强度或痛苦指数。不同组织画像如何对应具体感受质，仍需要额外理论。

该假说必须承担人工极简实现的可能。第 9.5 节给出一个很小但实际具有角色调制与返回关系的机制；按本假说，不能因为它简单便临时排除其参照型体验。其指定响应种类有限，是组织事实；将这一限制解释为体验内容的限制，还需第 6 节明确提出的分化桥接承诺。较小状态空间本身不证明体验微弱，更不证明没有强烈效价。这是理论预测，不是体验已被观测的事实。若这一后果不可接受，需要有独立理由的核心修订。

4.3 主体归属不是图连通性的开关

本文先描述参照组织怎样分化与聚合，再讨论主体归属。共用模型、数据或时钟不直接建立共同参照；相互影响增强，也不自动使两个既有实例在一个瞬间变成一个。若实际形成额外联合组织，可以考虑联合体验，但不由此默认局部组织消失。

并非每个子集、并集或变量打包都是新组织。候选必须附有实际过程、机制支撑、干预接口和分析时间范围。第 8 节在固定协议下给出最小机制支撑族，但本文尚无独立于所有物理描述的唯一候选列表；在多个合理模型给出不同聚合结果时，应报告这种依赖，而不是用一个未经证明的阈值宣布唯一主体数。

5. 有限形式核心：组织画像而非意识计量

5.1 模型和干预协议

先指定实际实现事件及其因果模型 $\mathcal{M}$、允许的局部干预集合 $\mathcal{I}$、当前工作状态附近的背景协议 b，以及有限时间范围 H。模型使用有限状态或已明确离散化的变量，定义每项状态更新及可用接口。干预是思想实验中的机制替换或夹持，不要求实际执行实验。

物理接口必须随变量变换一起转换。把两个独立系统的坐标混合，再假定新坐标可以被同样局部地独立夹持，等于偷偷改变了问题。背景协议还要区分正在运行与暂停模式；不能把重启设备的操作暗中放进用于判断暂停期间的干预集合。

5.2 内容角色调节

对参照候选 R 及内容通道 $X_i \to Y_i$，定义响应分布

$$K_i^r(a) = \mathsf{P}(Y_i \mid \mathsf{do}(R = r), \mathsf{do}(X_i = a); b).$$

四项差分为

$$D_i = K_i^r(a) - K_i^r(a') - K_i^{r'}(a) + K_i^{r'}(a').$$

对总质量为零的有符号测度，采用 $\|D\|_{\mathsf{TV}} = \frac{1}{2}\sum_y |D(y)|$，定义

$$m_i = \sup_{(r,r',a,a') \in \mathcal{A}_i} \|D_i\|_{\mathsf{TV}}.$$

$\mathcal{A}_i$ 收集使四项联合夹持均获得 $\mathcal{I}$ 许可的取值四元组。$m_i \in [0, 2]$ 描述指定接口下参照怎样改变内容的响应关系。它是反事实组织量，不要求每个实际时刻都发生切换，也不是体验强度。缺少必要比较协议时应记为不可识别，不能把无法比较当作作用为零。分布核差分与输出均值差分不同，不能混用来宣称排除了某类控制器。

5.3 经参照传递的跨内容影响

令内容 i 在时刻 t 影响参照，而内容 j 在后续时刻 s 被处理。定义

$$\Psi_{ij}^a(u) = \mathsf{P}(Y_{j,s+\tau} \mid \mathsf{do}(X_{i,t} = a), \mathsf{do}(X_{j,s} = u); b),$$

$$\Xi_{ij} = \Psi_{ij}^a(u) - \Psi_{ij}^a(u') - \Psi_{ij}^{a'}(u) + \Psi_{ij}^{a'}(u'), \qquad \rho_{ij} = \sup \|\Xi_{ij}\|_{\mathsf{TV}}.$$

ρ_{ij} 同样只对四项夹持均获许可的取值组合取上确界，固定探针时点及背景协议，且 $\rho_{ij} \in [0, 2]$。这里只将经机制分析确认通过参照维持路径的影响纳入 ρ_{ij}。可以在有限模型中预先定义路径阻断协议，区分它与绕过参照的直接数据传输。仅观察相关性不足以识别路径；在复杂因果图中，若路径效应不能唯一分离，应保留多个协议结果，而不声称唯一的“内在影响”。

5.4 分解距离与跨尺度画像

对申报机制范围的一项分割 π，令 $\mathcal{D}_\pi$ 是在同一组成变量和接口下、可按该分割独立更新参照的模型类。固定原有外源过程后，可分模型的转移核须在允许干预下按 π 分解；不得把跨分量的内生反馈重新命名为外源输入。定义

$$\delta_\pi = \inf_{Q \in \mathcal{D}_\pi} \sup_{I \in \mathcal{J}} d_{\mathrm{TV}}(K^I_{\mathcal{M}}, K^I_Q).$$

其中 K^I 为 H 内指定响应的分布。必须以同一个 Q 同时对照各项干预，不能为每项干预临时更换模型。δ_π 衡量模型内的分解代价，不是普通统计相关，也不直接是统一意识的强度。

组织画像保留向量、矩阵与尺度关系：

$$\mathcal{O}_{\mathcal{M},H} = \big((m_i), (\rho_{ij}), (\delta_\pi), \mathcal{T}, \mathcal{S}\big).$$

$\mathcal{T}$ 记录更新、维持及接续的时间关系，$\mathcal{S}$ 记录实现范围和组成。考察一族 H，可展示同一系统在不同分析尺度下的聚合方式。本文不将这些量加成一个总意识分数，不把非零值当作意识开关，也不把它们与物种的证据置信度混合。

6. 有效内容分化：从最大作用幅度到分辨率画像

角色调节量 m_i、经参照传递的影响量 ρ_{ij} 和分割距离 δ_π，分别约束调节幅度、返回关系和组织可分性，但没有单独回答“这个组织能以多少种方式区分并使用内容”。一个二值异或门可以实现彻底的角色反转，其可能内容却很有限。因此，本文增加一项独立的**响应分化画像**，用于约束关于内容贫乏或丰富的论述。该画像不替代原有维度，不汇总成意识总分。

6.1 固定模型中的有效响应签名

仍采用已指定的有限物理因果模型 $\mathcal{M}$、局部干预集合 $\mathcal{I}$、背景协议 b 和有限时间窗 H。另需公开两个选择：

- $\mathcal{C}$：协议允许的有限内容条件集合。条件可以是起始内容状态，也可以是明确给定的有限内容历史；不得在比较中暗自改变二者的含义。
- Y_H：分析窗口内由机制实际使用的内容处理及参照维持响应。不能为了增加分化数，额外记录与该过程无关的标签、寄存器编号或观察者注释。

允许干预必须具有相同的物理含义，并保留用于比较的内容差异。若某项干预完全覆盖了起始内容，它在这项比较中可能不提供辨别信息；不能把“干预抹去了差异”解释为原机制没有内容差异。

对 $c \in \mathcal{C}$，定义响应签名

$$\Sigma_H(c) = \big(K^I_c\big)_{I \in \mathcal{J}}, \qquad K^I_c = \mathsf{P}_{\mathcal{M}}(Y_H \mid c, I; b).$$

其中 c 的设定或继承方式由协议预先明确。签名记录同一内容条件在一组固定探针下的因果响应，而不是对该内容所作的语言描述。以

$$d_H(c,c') = \sup_{I\in\mathcal{J}} d_{\mathrm{TV}}(K_c^I, K_{c'}^I)$$

比较两个签名，得到 $[0,1]$ 内的伪距离。称它为伪距离，是因为不同记号甚至不同物理状态，可以在这里的全部指定响应上没有差别。令 $c \sim_H c'$ 当且仅当 $d_H(c,c') = 0$，便得到协议下有效响应类型的等价关系。

这一等价关系只在指定窗口、接口和干预下成立。它没有把两个实现过程合并成同一个体验实例，也没有把所有作用于其他过程的差异永久删除。改变时间窗或纳入新的实际响应，可以改变类型划分。

6.2 使用分辨率画像，不把“严格不同”变成意识开关

只计算 $|\mathcal{C}/\sim_H|$ 有一个问题：任意微小的新效应，都可能把原来相同的签名变成严格不同，导致整数计数立即增加。若把这种变化叫作意识突然增强，就会重犯用非零边判定主体的错误。

因此，对分析分辨率 $0 < \eta < 1$，定义

$$N_H(\eta) = \max\{|A| : A \subseteq \mathcal{C},\quad d_H(c,c') > \eta \text{ 对所有不同的 } c,c' \in A\}.$$

即选取两两响应距离超过 η 的最大内容集合。本文报告整条 $\eta \mapsto N_H(\eta)$ 画像，并随同报告 $H, b, \mathcal{J}, Y_H$，不从某个单独取值宣布意识开始存在。η 是区分细微响应差异的分析参数，不是物理意识阈值，也不是心理感受的最小单位。

该整数画像自身可以有台阶；本定义没有消灭所有数学不连续性，也没有证明现象体验必然连续。它的优点是公开显示：某种分化只在极精细的分辨率下出现，还是在较宽的分辨率范围内都能维持。后者与“存在任意非零差异”是不同的组织陈述。

此外，N_H 描述固定协议下的可能响应分化，不是当下同时持有的内容数量。不同内容对可能由不同探针区分，因此也不能把它直接称为一套共同解码方案的通信容量，更不能直接换算成主观体验的信息位数。

6.3 两个有限性质

性质一：冗余标记不增加有效分化。 假设把内容条件 c 细分为 (c,w)，但新增标记 w 对全部指定干预响应没有影响，即 $K^I_{(c,w)} = K^I_c$。那么任意 $\eta > 0$ 下的 $N_H(\eta)$ 不变。

证明。 同一 c 的任意两个标记副本距离为零，不能同时进入一个 η 分离集合；不同 c 之间的距离又没有改变。原分离集合可以任取标记提升为新集合，新集合也可以投影回原集合，最大大小因而相同。证毕。

保持实际接口与结果事件对应的双射重标记也不改变画像，因为总变差距离在这种重标记下不变。新增物理存储容量不自动等于新增有效内容分化；只有其差异被相关机制实际使用，才可能改变这里的签名。

性质二：小的响应变化可用分辨率位移界定。 比较两个具有相同 $\mathcal{C}, \mathcal{J}, Y_H, b, H$ 的模型。若对所有 c, I 均有

$$d_{\mathrm{TV}}(K_c^I, \widetilde{K}_c^I) \le e,$$

则由三角不等式，

$$|d_H(c,c') - \tilde{d}_H(c,c')| \le 2e.$$

因此，在 $0 < \eta - 2e < \eta + 2e < 1$ 时，

$$\widetilde{N}_H(\eta + 2e) \le N_H(\eta) \le \widetilde{N}_H(\eta - 2e).$$

证明。 新模型中距离超过 $\eta + 2e$ 的集合，在原模型中的距离仍超过 η；原模型中距离超过 η 的集合，在新模型中的距离仍超过 $\eta - 2e$。分别取最大集合即可。证毕。

此性质不保证某个固定分辨率上的整数计数连续，却限制整条画像能怎样随响应变化移动。在有限状态、有限干预、固定有限时间窗及固定或连续变化的初始分布下，连续概率转移可支持这类比较。确定性输出的微小数值变化，在未指定测量噪声或粗化时不一定带来小的总变差距离，不能把这两件事混淆。

6.4 手工反例：最大调节相同，响应分化不同

将三位二值机制推广到 $\mathbb{Z}_k$，其中 $k \ge 2$，三个寄存器同时更新：

$$x^+ = x + r \pmod{k}, \qquad y^+ = y + r \pmod{k}, \qquad r^+ = x + y \pmod{k}.$$

对 x 通道，分别取参照 $r = 0, 1$ 和内容 $x = 0, 1$，四项响应分布差为

$$D_x = \delta_0 - 2\delta_1 + \delta_{2 \bmod k}.$$

当 $k = 2$ 时，$D_x = 2\delta_0 - 2\delta_1$；当 $k > 2$ 时，三个支持点中，正质量之和与负质量绝对值之和都为 2。因此总变差范数恒为 2，达到原定义的上界。对 y 通道同理，故

$$m_x = m_y = 2 \qquad (k \ge 2).$$

再固定 $y_t = r_t = 0$，在 t 时刻设定 x，在 $t + 1$ 时刻以 u 夹持 y。同步更新保证

$$r_{t+1} = x_t, \qquad y_{t+2} = u + x_t \pmod{k}.$$

选择 $x_t, u \in \{0, 1\}$ 得到同样的四项差分；交换两个内容通道亦然。在这一固定两步探针协议下，

$$\rho_{xy} = \rho_{yx} = 2 \qquad (k \ge 2).$$

然而，固定 r 后，$x \mapsto x + r$ 是 $\mathbb{Z}_k$ 上的双射。以允许的 k 种起始 x 为 $\mathcal{C}$，以机制实际使用的 x^+ 为单通道响应，保留固定 r 下的更新探针，则任意不同起始值产生不同的点分布，距离均为 1。因此，对所有 $0 < \eta < 1$，

$$N_H^{(x)}(\eta) = k.$$

上界来自集合仅含 k 种条件，下界来自这 k 种条件本身构成分离集合。该输出继续参与后续参照更新，不是为了提高计数附加的外部标签。

由此得到明确的有限结论：**四项最大角色调节及返回影响可以完全相同，而单通道可区分响应数不同。** 所以 m、ρ 不能独自承担内容分化的描述。这里没有声称各 k 的 δ_π 相同，也没有证明旧组织画像的全部分量都相同；反例只确立这些最大幅度指标对响应分化的不充分性。

6.5 对桥接假说的补强与限制

本假说进一步承诺：参照型体验的内容分化受其实际实现的、因果有效的响应区别约束；单靠无作用标签、重复编码或观察者命名，不会增加这种体验内容的区别。响应分化画像为该承诺提供可审查的组织描述，而不是用“贫乏”一词代替分析。

这个桥接承诺仍不能从上述数学结果中推出。特别是，N_H 增大不独自证明体验更强、更清醒、更有自我或更痛苦，也不证明现实系统同时体验其全部可能内容。三位实例相对于较大 k 只有较少的指定响应类型，是机制事实；把这种限制解释为参照型体验的内容限制，是本理论公开承担的心理与物理的假设。

适用范围。 画像仍依赖对内容条件、实际相关响应和允许探针的选择。选择过宽，可以把许多并未共同实现当前内容的备选状态纳入；选择过窄，可以漏掉系统实际使用的区别。简单控制器、检索装置也可能具有很大的响应库，因此该维度必须与参照构成、实际实现、时间关系和可分性一起分析，不能成为新的单一意识判据。本文给出的有限性质排除了冗余重标记并约束分辨率敏感性，但尚未提供适用于任意物理系统的唯一选择算法。

为保持分化与调节的区别，扩展后的画像写为

$$\mathcal{O}^*_{\mathcal{M},H} = \big((m_i),(\rho_{ij}),(\delta_\pi),\eta \mapsto N_H(\eta),\mathcal{T},\mathcal{S}\big).$$

所有分量均附带相应接口与背景选择。本文并不规定它们之间的统一加权，也不以一个分量的增长推断其余分量或体验总强度增长。

7. 三个基础一致性命题

命题 1：共享只读来源不单独建立联合参照

设 A、C 仅共用不受其改写的来源 B。固定相关 B 过程后，转移满足

$$\mathsf{P}(A',C' \mid A,C,B) = \mathsf{P}_A(A' \mid A,B)\mathsf{P}_C(C' \mid C,B),$$

且分解在允许的局部干预与后续时间内保持。则跨两模块的 $\rho_{ij}=0$，对应分割的 $\delta_{A|C}=0$。

证明。 改变 A 不改变 C 的后续角色响应核，反之亦然，跨模块四项差分为零。原模型本身属于 $\mathcal{D}_{A|C}$，故分解距离为零。证毕。

该命题允许局部画像非零，不能据此判局部系统没有体验；共享可写记忆或共同更新也不符合这里的只读前提。结论是共享来源不单独提供联合参照证据，不是一般共享系统的意识定理。

命题 2：局部重编码不改变画像

若两个描述以逐组成变量的双射对应同一更新接口、干预和响应分布，则对应的 m_i,ρ_{ij},δ_π 不变。

证明。 双射把分布及有符号测度推送至重标记结果空间，总变差距离不变。允许干预与可分模型类相应一一对应，因此上确界与下确界不变。证毕。

它适用于同一实现的重描述，也适用于不同实现之间的类型比较；后一情形不合并实际实例。更一般的坐标变换需要同时保留完整物理干预语义，不能只比较变换后方程。

命题 3：连续的结构变化可以被二分图规则放大成跳变

加入参数 ε 控制的跨模块作用：

$$\mathsf{P}(R'_C = 1 \mid c, a) = p_C(c) + \varepsilon a,$$

其中 $a \in \{0,1\}$、$p_C(c) \in [1/4, 1/2]$、$0 \le \varepsilon < 1/4$。反向作用对称。后续内容探针 $u \in \{0,1\}$ 通过 $Y_C = R'_C u$ 使用参照。

证明。 对 $a = 1, a' = 0, u = 1, u' = 0$，四项差分在 $Y_C = 1$ 为 ε，在 $Y_C = 0$ 为 $-\varepsilon$，故在上述单步参照更新、固定后续探针的协议下 $\rho_{AC} = \varepsilon$。它随参数连续变化。若把 $\rho > 0$ 仅记成一条边，两个模块在零点分离，而任意正参数都被记成强连通。这一跳变来自二分编码，而不是作用幅度的不连续。证毕。

在有限状态、有限干预及固定有限 H 下，若初始分布与背景协议固定或连续依赖参数，连续转移参数给出连续响应核。到固定可分模型类的距离是 1-Lipschitz，故相应 δ_π 也连续；这里不能随参数暗改模型类、变量或干预协议。此命题不证明所有现象变化必连续，也不排除长时极限下的分岔。它说明保留程度信息比仅报告连通与否更忠实于该组织模型。

8. 有限模型中的参照组织范围：最小机制支撑族

8.1 目标及适用范围

本节给出一个可以在有限模型中执行的范围识别方案。它回答的是：**在已声明的物理粒度、时间范围和局部干预接口下，哪些实际机制足以维持一项具体的参照响应任务，哪些机制不能在不改变该任务的条件下进一步省略？** 其结果称为“参照组织的最小机制支撑族”。它不是自然界唯一的主体分区，也不把支撑族的成员数解释为体验者数量。

这里的“任务”是研究者预先指定的响应比较，不要求系统有目的、语言或外部行为。它应包括内容作用随参照改变、内容对后继参照的影响、后继参照对其他内容作用的改变，以及必要的路径阻断对照。单次任务成功、一个汇总值或者一段自然运行轨迹，都不足以替代这组响应。

8.2 定义：固定接口下的有限支撑

给定有限状态因果模型 $\mathcal{M}$，其实际机制位置组成集合 $E = \{e_1, \dots, e_m\}$。每个位置具有唯一物理标识、明确的更新规则和输入输出端口。另指定有限时间范围 H、背景协议 b、有限干预集合 $\mathcal{I}$ 和响应任务 $\mathcal{Q}$。固定权重、只读规则及存储读取电路都可以属于机制；能够改写自身不是进入 E 的条件。

在本方案中，所谓省略一个机制，是在模型内以预先规定的约定常量输出或外源轨迹替换其作用，并非把真实设备连同供电一起摧毁。每个替换规则必须在比较之前固定，不得根据当前干预或保留机制的状态，暗中计算被省略的原功能。背景能量供应及已声明的外部刺激按同一协议保持。探针、测量及替换操作也必须彼此兼容，并保持相同时间语义。

对保留集合 $S \subseteq E$，由此得到模型 $\mathcal{M}_S$。令 $K_{\mathcal{M}}^{I,Q}$ 是干预 I 下、时间范围 H 内所选响应的联合分布，定义

$$d_Q(S) = \max_{I \in \mathcal{J}} d_{\mathrm{TV}}\left(K_{\mathcal{M}}^{I,Q}, K_{\mathcal{M}_S}^{I,Q}\right).$$

在预先给定误差容许度 $\alpha \geq 0$ 下，支撑族为

$$\mathfrak{S}_Q^{\alpha} = \{S \subseteq E : d_Q(S) \leq \alpha, \nexists S' \subsetneq S \text{ 使 } d_Q(S') \leq \alpha\}.$$

这是按集合包含关系的最小性，不是只保留机制数最少的一个答案。不同长度的替代路径可以分别成为最小支撑；并列解全部保留。机制增加也未必单调改善响应，因此不额外假定支撑集合具有向上闭合性。$\alpha = 0$ 表示精确保存；$\alpha > 0$ 表示模型比较的误差容许度，**不是意识存在或主体融合的阈值**。

由于 E 有限且 $\mathcal{M}_E = \mathcal{M}$，至少有一个支撑；逐次考察真子集即可得到至少一个按包含关系最小的成员。该有限性结果只证明给定比较问题有解，不证明结果已经找到体验主体。

8.3 防止重复计数与任意并集

每项任务还须附带其所针对的实际参照变量或分布式实现位置，以及角色调节和返回路径的说明。把 R_A 与 R_C 写成有序对，不会单凭记号产生一个新的参照实现。若联合任务只是两项独立任务的并列，其响应核在相关外源过程固定后因式分解，且该分解在全部已声明干预下保持，应报告“两个任务的联合描述”，而不是额外的共同参照证据。这样的联合任务仍可有数学上的最小支撑；其存在本身没有新的现象含义。

同一物理机制的重新命名不增加成员。若重编码同时一一转换状态、干预、替换规则及响应位置，总变差距离保持，支撑族也一一对应。反之，两个实际复制体即使得到同构支撑，仍须保留各自物理标识和因果历史，不应被合并。

多个最小支撑也不等于多个主体。 例如，同一个目标参照节点由两条冗余路径维持，省略任何一条都能保存全部任务响应，就可能得到两个最小支撑。这首先表示该任务存在实现冗余。两条路径是否另外各自实现其他参照组织，需另设明确任务分析；不能由冗余支撑的数量直接推断。

因此，每项结果至少应同时报告：实际目标及事件、支撑族、各项省略距离、内容角色调节与返回影响画像、可分解关系，以及对 $H, b, \mathcal{J}, \alpha$ 的依赖。它们共同描述一个组织候选，任何一项非零数值都不是体验开关。

8.4 四种结构的比较

结构	在有限协议下可以推出什么	不能据此推出什么
共享只读记忆：A、C 只读取 B，彼此不改变对方更新	固定 B 过程后，若两侧更新及所有允许干预持续分解，联合任务只是局部任务的并列；公共读取可以进入各自支撑，但没有因此建立跨域参照返回	不能认定局部系统无体验，也不能因 B 出现在两个支撑中认定它们是同一主体

结构	在有限协议下可以推出什么	不能据此推出什么
单向发送：A 向 C 提供消息，C 不影响 A	发送可以影响 C 的内容及参照。对整个 A+C 模型，若不存在 C 返回 A 的相关路径，便没有建立覆盖两侧的双向参照构成；但 C 的组织可以被 A 提供的材料改变	不能据此排除 C 自身体验；不能把所有消息生产者及历史成因自动并入 C
双向通信：两侧分别解释并回复消息	普通信息交换可以改变输入，而不改变对方内容的响应关系；也可能经各自参照改变后续角色。必须报告实际调节、返回及分解结果	既不能凭连通性宣称融合，也不能一概断言“通信绝不构成体验组织”
共同参照：实际 R 由两侧内容维持，并反向改变两侧内容作用	可以针对 R 设联合响应任务，检验其支撑是否涉及双方、是否不能按声明分割独立保存任务，以及相关影响和分解距离如何变化	不能只因支撑跨越双方就证明一个主体；也不能由联合组织出现直接推断局部组织消失

单向发送案例特别说明“输入来源”与“当下构成”的差别。若研究 C 的本地接收组织，可以把消息流声明为受控外源输入；若研究发送与接收整体，则必须把消息产生机制纳入模型。两种协议回答不同问题，应并列报告，不能暗中先选一个边界再把它当成结论。扩大分析范围也不消除单向关系与双向参照返回的区别。

共同参照的有限见证可采用一个由两侧信息共同更新的门控状态 R，再让 R 改变双方探针输入的处理规则。比较必须包括夹持 R、改变两侧内容及阻断维护路径，而非只观察一个共享计数器变化。若所选响应能由独立参照重现，或影响来自绕过 R 的路径，便没有得到所宣称的共同参照结果。

8.5 已求解例：并列任务的唯一支撑不等于联合主体

设两个独立二值回路 A、C 共用只读背景 B。固定 $B = 0$，各侧机制为

$$G_j : y_j^+ = u_j \oplus r_j \oplus B, \qquad H_j : r_j^+ = y_j, \qquad j \in \{A, C\}.$$

各式同步更新，u_j 是本地输入探针。机制位置为 $E = \{G_A, H_A, G_C, H_C\}$。省略一个位置时，将其每步输出规定为 0；这是一种明确的比较操作，不声称 0 是物理上客观的“中性值”。干预只设置初始状态和输入探针，不绕过被省略的更新规则。

本地任务 Q_j 观察 $(y_{j,1}, y_{j,2})$，允许全部二值初始状态和两步探针输入。G_j 必须保留：取 $r_{j,0} = 0, u_{j,0} = 1$，原机制给出 $y_{j,1} = 1$，省略 G_j 后为 0。H_j 也必须保留：在 G_j 保留时，取 $y_{j,0} = 1, r_{j,0} = 0, u_{j,0} = u_{j,1} = 0$，原机制有 $r_{j,1} = 1, y_{j,2} = 1$，省略 H_j 后则有 $r_{j,1} = 0, y_{j,2} = 0$。两个见证均产生不同的确定性响应，总变差距离为 1。因此本地唯一精确最小支撑为

$$\mathfrak{S}^0_{Q_A} = \{\{G_A, H_A\}\}, \qquad \mathfrak{S}^0_{Q_C} = \{\{G_C, H_C\}\}.$$

令并列任务 $Q_{A,C}$ 同时观察双方响应。省略四个位置中任一个，均可用相应本地见证改变联合响应；其他位置是否保留不影响该见证。故对每个真子集 $S \subsetneq E$，都有 $d_{Q_{A,C}}(S) = 1$，而 $d_{Q_{A,C}}(E) = 0$。其唯一精确最小支撑就是 E。

然而，两侧更新和全部本地干预始终分解。跨模块内容不能改变对方的角色响应，所以 $\rho_{AC} = \rho_{CA} = 0$；原模型本身属于 A|C 的可分模型类，故 $\delta_{A|C} = 0$。**四块机制是完成两项并列任务的唯一支撑，并未因此形成额外联合参照。**

这个手算例证明的是指定任务、接口及省略规则下的依赖关系，并校准支撑族的解释方式；它不是新的意识存在定理，也不证明局部各有一个主体。其作用恰在于反驳“唯一联合支撑必然对应一个联合体验者”的误推。

8.6 该方案必须承担的边界

第一，最小支撑是任务相对的。 改变背景、替换方式、允许干预、响应位置或时间范围，都可能改变答案。现方案不从数学定义推出自然界唯一的物理粒度；应把有理由的不同协议作为敏感性结果保留。

第二，最小支撑不是实际构成的全部。 冗余机制可能正在参与维持，却因替代路径而不属于每个最小支撑。支撑族的交集只能表示在该比较下不可替代的部分；并集只能表示至少一个最小解使用的部分。两者都不是自动的主体边界。

第三，不能一般排除白板架构、协作系统和普通控制器。 如果其实际消息交换维持了角色调节及返回关系，且联合任务不能按指定分割保存，本方案就应记录这种组织。把同样结构称为“交流”或“计算”不能成为排除理由；其现象归属仍由正面桥接假说承担。

第四，渐变画像与主体计数分开。 很弱的耦合可能使精确最小支撑跨越更多位置，但它不使体验强度或主体数量随之跳变。应同时保留作用幅度、分解距离及多种容许度下的支撑变化，不使用 $\rho > 0$ 作为意识开关。

本节完成的是一个**有限、可检查的参照组织范围分析**：固定问题后枚举并保留全部最小支撑，识别因并列任务、重描述和冗余造成的计数陷阱，并对共享来源、单向发送、双向通信与共同参照采用同一套比较。它使候选范围可争论、可复核，但不宣称已证明何处存在恰好一个主观体验者。

最小支撑及多解思想本身不主张首创。近邻工作已提出理论相对的最小互惠核与核简并；此处的具体选择是以预先声明的参照响应和机制省略协议定义支撑，并明确不把多解数用作主体数。10

9. 二十一个思想实验的系统检验

以下思想实验用于检验论证，不是经验发现。部分是对既有问题的重新构造，其价值在于同一规则对它们给出一致结果。

9.1 坐标搅拌

两个独立系统满足 $x' = ax, y' = by$ 且 $a \neq b$。改用 $u = x + y, v = x - y$，得到

$$u' = \frac{a+b}{2}u + \frac{a-b}{2}v, \qquad v' = \frac{a-b}{2}u + \frac{a+b}{2}v.$$

新方程有交叉项，却没有发生新物理耦合。只按变量依赖图判断，会凭写法改变主体。修订后的模型必须带上实际干预接口：夹持 u 同时保持 v，通常意味着协调改变原来两个系统，不能冒充原先的局部干预。这一反例说明图上的相互依赖本身不具有一般描述不变性。

9.2 两个完全相同的复制体

两套独立设备从相同状态开始，输入和内部序列完全一致。它们可以有相同画像，却不是同一次实现。类型等价只支持同等的理论归类，不把两个当前体验实例合并。用机制等价类直接定义唯一主体，会在此失败。

9.3 无穷多个观察时间窗

把同一段活动分别按一秒、两秒及全部滑动窗口描述，不会仅因描述数量增加而创造新主体。窗口用于检验组织持续与尺度关系，不能自动成为体验者的计数单位。也不能以“有时间或部件重叠就合并”补救，否则会误合并实际不同的重叠组织。

9.4 同一文件与多个运行

同一权重文件服务多个相互隔离的会话。共用参数类型不能代替每次运行的状态、历史与机制归属。相反，多个不同模型也可以参与一个较大的组织。候选范围须从具体实现分析，不从产品账号或模型名称直接读出。

9.5 三位控制器：理论必须接受的困难案例

给出同时更新的三位二值机制：

$$x^+ = x \oplus r, \qquad y^+ = y \oplus r, \qquad r^+ = x \oplus y.$$

r 改变 x、y 的响应方式，内容又可经 r^+ 改变后续作用。在局部二值夹持协议下，异或通道给出 $m_x = m_y = 2$。进一步固定其他初始条件，在 t 时刻夹持 x，在 t+1 时刻以 u 夹持 y，则 $r_{t+1} = x_t \oplus y_t$，$y_{t+2} = u \oplus r_{t+1}$，故四项差分给出 $\rho_{xy} = 2$；交换 x、y 同理。若将中间参照夹持为与 x 无关的固定值，这项跨内容调制消失。这些是有限定义下的手工推导，不是测得的体验数值。它不需要语言、外部行动或学习参数。

在 ERF 假说下，该机制是极简参照型体验的候选实现。在上述有限模型中它只有八种联合状态；指定单通道的响应分化则由第 6 节给出，不能直接用八这个数字推断体验强度。把有效内容分化约束解释为现象内容约束，是额外桥接承诺。较大的调制数值不能直接解释为体验丰富，也不能据此断言它像人一样清醒、会痛或绝不可能痛苦。这是明确的激进承诺：若不接受该归类，必须修改正面桥接或加入有独立理由的条件，不能只说“这不是真正意义”。

9.6 同样轨迹，不同构成

让上一机制处于全零稳定轨迹。另一个装置只保存同样的零值序列。实际显示完全相同，但在保持运行模式不变的局部干预下，前者可以传播内容对参照的影响，后者不执行相应转换。

因此，状态序列相同不唯一确定画像。此例也没有把保存装置的全部物理机制判为无体验；它只说明保存的那段内容没有自动重现原参照过程。

9.7 一份公共记忆，两套参照

A、C 读取相同 B，分别维持 $a' = f(a, y_A), c' = g(c, y_C)$。在命题 1 的独立前提下，共同输入没有创建联合参照。给两者发送更多相同数据也不自动改变结论；这里区别的是构成关系，不是带宽或知识总量。

9.8 白板协作与额外联合过程

两个人共同更新白板，白板又影响双方理解。旧版“有联合变量和双向作用便有联合主体”的说法会轻易把白板当作第三个感受者。现稿撤回这种保证：协作必须给出其内容调制、返回路径与跨分割画像，而不能只画一个环。

若确实另建 $R' = H(R, y_A, y_C)$，还必须有 R 对双方内容角色的实际反向作用，而不只是汇总。若由此形成新的参照组织，理论允许增加联合维度；局部组织是否保留另行分析。本文没有证明所有普通协作都没有联合体验，也不因共用目标就认定它存在。

9.9 无穷小连线

将两个组织之间的作用逐步从零增加。命题 3 显示，作用画像可以连续变化，而强连通图立刻合并。现稿因此不以非零边宣布意识突然诞生或主体瞬间融合，也不引入一个假装普适的任意 epsilon 阈值。具体聚合结论随物理模型、时间与测量协议公开。

9.10 随机初始化到成熟能力

保持架构，逐步改变参数。接线图相同不保证有效组织相同；随机系统也不能仅因未经训练被排除。若已实现参照构成，理论不因未经训练排除其体验可能；某项调节作用较弱并不推出整体体验较弱。内容分化须另按第 6 节的响应画像与桥接承诺分析。训练可以丰富、削弱或改变不同维度，不保证单调增长。

这里不需要一个“毕业时刻”。但放弃训练步数阈值，不等于已经找到了真实模型的体验起点；它只是要求以组织变化而非资格标签解释。

9.11 改变损失单位

普通梯度下降为 $\theta^+ = \theta - \eta \nabla L$。令

$$L' = cL + b, \quad c > 0, \quad \eta' = \eta / c,$$

则 $\eta' \nabla L' = \eta \nabla L$。在没有其他组件读取原始损失或梯度数值的前提下，更新轨迹相同，损失和梯度范数却可不同。

因此，原始损失绝对值、符号或梯度大小，不能单独充当这一更新过程的内在痛苦度量。这是基础数学反例，不是训练绝无痛苦的证明。如果数值还参与其他实际过程，前提失效；更新轨迹相同也不保证任意物理实现具有同一体验。

9.12 高任务能力与贫乏参照

设想一组高度专门化的求解器，以独立、一次性转换完成复杂任务；再设想一个持续维持多种感受关系但不擅长逻辑题的组织。仅由任务分数无法给它们的参照画像排序。

这个构造显示能力指标与体验维度之间没有定义上的等同，不证明现实 AI 一定体验弱、某个人一定体验强。感官分化、情感鲜明、记忆接续和逻辑能力也不能压缩为一个排名。

9.13 平静体验与停止执行

运行中的耦合过程可以处于稳定点，例如 $\dot{x} = -x + y, \dot{y} = x - y$ 在 $x = y$ 时不改变宏观数值；停止执行的存储也可以呈现相同常数。若前者在当前模式下能传播扰动、后者不能，其机制

不同。

这仅证明宏观不变不等于停止，没有证明该二变量装置本身满足完整假说。对冥想而言，理论不能要求每一刻有内容冲突、外部行动或新学习。

9.14 极少内容与完全无内容

逐步减少对象、语言思考、自我表象和时间感，未必同步停止参照维持。研究中存在最小体验与非二元觉察的不同解释，不能把其中一种解释当成事实。12

但如果所谓体验真的没有任何可分化的组织关系，并且仍确实存在，这会挑战本假说。现稿不通过宣称“必有隐藏参照”使自己免于反例。

9.15 递推、展开与回放

展开与替代争论已经对功能等价、因果组织和意识推断的关系提出形式约束。28 本文保留明确的局部干预语义，不声称已解决这一检验难题。将 $s_{t+1} = F(s_t, u_t)$ 的有限 N 步运行展开为 N 层，只要每层及对应时点干预保存相应转换，就可逐层归纳得到相同后继响应。仅回放一次记录通常不保留这些反事实响应。

因此，图上有没有回线不是独立判据。在指定有限因果层次上若展开与递推保留相关画像，理论须同等分类；它们仍是不同实现实例。更细物理干预或更长时间范围可能破坏等价，不能用一次输出相同代替全部前提。

9.16 一万年才形成一个想法

将全部相关转换、延迟、记忆保持和分析时间共同放慢，若映射后的干预响应保持，则非时间画像保留，绝对持续时间按比例变化。于是绝对速度不能独自将体验归零。

若只放慢局部，噪声、环境和遗忘仍保持原速，就不满足前提。完成任务的时间、体验内容的更新速率、参照维持跨度和历史接续长度也不是同一个量。

9.17 从全局时钟到局部握手

将同步装置逐步改成前一步完成后再触发后续的异步实现。若它保留相关状态关系、干预响应、反馈及相对时序，单独撤去全局时钟不改变本假说的组织分类。

如果异步化改变竞态、反馈可达性或实际时间关系，则允许画像改变。这不是跨硬件体验等价的无条件定理，也不以最终答案相同代替内部条件。

9.18 共用晶振与共用 GPU

两个隔离过程可共用时钟或交错执行。若全局更新仅交错应用 $T_A \times I_C$ 和 $I_A \times T_C$，两类更新可交换，投影得到各自的局部状态序列。因此，共同执行器没有凭此创造跨内容参照。

若时序本身被用作内容、发生侧信道或调度器参与内容选择，独立性前提失效，需要扩大机制分析。这仍不等于已经证明融合。一个时钟周期、一条指令、一个 token 与一次体验都不能直接对应。

9.19 渐进替代与功能外包

逐步把参照维持的部分过程移到人工部件，若组织关系与接续保持，生物比例不提供天然分界。若只是外包一次计算，归属可以保持工具关系；若外部形成独立参照，则可能出现并存。

渐进替代思想实验已有前例。13 本稿增加的是将每一步分别分析为当下构成变化、历时资源变化和实例接续，而不是把“计算机承担 51% 功能”当作意识搬迁时刻。

9.20 分支与再耦合

一个过程把状态传给两个实际后继，形成共同前史；两份后继可以逐渐分化，而不是必须有一个保留原主体、另一个自动无体验。再次交换记忆或接入联合过程，也不倒推它们此前一直是同一次体验。

接续强弱可以分维度讨论，实例归属则保留真实因果历史。本文不据此证明上传后的严格个人同一性或不死，也不把再耦合一概说成局部体验消失。

9.21 通常行为相同，内部参照组织不同

设输入 u_t 为独立于内部状态的外源二值流；每个时间步依次完成内容计算、输出读取和参照写入。系统 A 与 B 使用同数目的内容寄存器和参照寄存器，并在同一时刻提供输出。其更新规则为

$$\begin{aligned} A:&\quad x_t = u_t \oplus r_t,\quad y_t = x_t \oplus r_t,\quad r_{t+1} = x_t;\\ B:&\quad x_t = u_t,\qquad\quad y_t = x_t,\qquad\quad r_{t+1} = x_t. \end{aligned}$$

其中 $\oplus$ 表示异或，输出步骤仍读取更新前的 r_t。A 中的 r 返回参与内容处理；B 中同名寄存器只保存记录，无返回路径。对任意正常输入序列及任意初始参照，两系统都恒有 $y_t = u_t$。这是对全部通常输入的函数等价，不是单条轨迹相同；必要的输出等待可保持相同外部时延，不增加返回通路。

现在把思想实验允许的内部干预接口写明：在输出步骤之前分别夹持当前 $r_t = r$ 与 $x_t = a$，且不同时改变其他更新方程。对内部通道 $x \to y$，响应核分别是 $K_A^r(a) = \delta_{a\oplus r}$ 与 $K_B^r(a) = \delta_a$，其中 δ_z 表示集中于 z 的点质量。取 $r = 0, r' = 1, a = 1, a' = 0$，四项差分为 $D_A = 2\delta_1 - 2\delta_0$、$D_B = 0$；按正文的总变差约定，得到 $m_A = 2$、$m_B = 0$。返回路径也可以单独检查：令 $\mathrm{do}(x_t = q)$ 在本步输出前生效并维持至参照写入，再于下一步输出前令 $\mathrm{do}(x_{t+1} = v)$，则

$$\begin{aligned} \mathsf{P}_A(y_{t+1} \mid \mathrm{do}(x_t = q), \mathrm{do}(x_{t+1} = v)) &= \delta_{v\oplus q},\\ \mathsf{P}_B(y_{t+1} \mid \mathrm{do}(x_t = q), \mathrm{do}(x_{t+1} = v)) &= \delta_v. \end{aligned}$$

若把中间 r_{t+1} 夹持为与 q 无关的常值，A 的这项返回调节随即消失。这里的内部夹持属于扩展的干预协议，不能混称为“正常输入”；夹持破坏了 A 中原有的补偿关系，因此干预后的输出差异首先是机制差异的证据。

条件命题：通常输入输出等价不充分推出参照画像等价。 在上述固定变量、分步时序和共同允许的局部干预接口下，A、B 对全部正常输入的输出相同，但其画像中 $m_{x\to y}$ 不同，故正常输入输出函数不能唯一确定该画像。证明由前述恒等式和四项差分直接给出。此命题只说明组织可辨识性；若进一步接受 ERF 桥接假说及该通道的构成相关性，才产生体验相关组织不同的条件性预测，仍不能断言 A 有完整主体、B 完全无体验，或 A 的体验更丰富。简单补偿反馈本身也不能自动决定整体主体边界。展开与替代争论关心的还包括如何独立地推断体验；Hoel（2026）在

自己的可证伪性框架中进一步主张持续学习的必要性。本文的两系统具有固定更新规则，而内部干预画像仍不同，说明“固定规则”与“无可辨识的参照组织”不能直接等同；它并未证明该组织具有现象方面，也未证明仅扩大机制干预集合就能解决体验推断问题，因而不构成对展开、替代或 Hoel 论证的反驳。28 29

10. 生物观察如何约束本假说

清醒人类的报告与多种研究构成主要经验参照，但人的状态也会变化。对牛、狗等哺乳动物的意识归属有强科学支持；对其他脊椎动物及许多无脊椎动物至少存在现实可能。这种证据分层不是按主观强度排名。14

对象	证据边界	对本假说的要求
牛、狗	不应靠单个拟人故事论证	语言和抽象推理不能成为门槛
鱼	伤害性感受器及持续行为变化具有证据意义，但不独自证明疼痛	不以人类新皮层排除，不以反射定论
蛇	嗅觉自我辨认测试具有物种差异	未通过某种自我识别不等于无任何体验
蝴蝶	昆虫总体证据不能原样外推每个物种	区分类群推断与直接证据
单细胞	复杂适应及习惯化不等于现象体验已经建立	不能只以生命、记忆或趋利避害认定体验

虹鳟及蛇类的原始研究分别说明证据要结合机制与感觉生态解释。15 16 无神经生物的习惯化则说明简单学习不能独自完成意识区分。17

本文不为鱼、蛇、蝴蝶和细胞虚构统一的组织分数。若某个细胞确实实现参照构成，本假说允许极简体验；若宽泛版本因此普遍纳入细胞或小型控制器，就必须承担这一后果或有原则地修订。既不能用物种名称作排除条款，也不能用理论允许作成经验确认。

11. AI 的参数、记忆、训练与时钟

11.1 参数固定不意味着状态固定

固定权重模型可以利用当前上下文调整输出，而不做梯度更新；KV 缓存等是运行状态，与基础参数不同。18 19

对理想化系统可写

$$y_t = F_{\theta_t}(c_t), \quad c_{t+1} = U(c_t, y_t, m_t, u_{t+1}), \quad m_{t+1} = V(m_t, c_t, y_t).$$

即使 $\theta_t = \theta$，整体仍可有历史依赖。参数知识、当次上下文与跨调用新增记忆须分别讨论。外部检索与参数记忆结合已有工程先例；在线改变特定神经记忆也已有研究。20 21

11.2 形成史不是体验资格

训练和部署可以实现不同机制，应具体比较前向运算、历史保留、反馈、优化器及调度。若组织画像相同，不应仅因阶段名称不同而相反归类。若训练彼此重置而部署加入持续结构，差异来自机制，不来自“训练结束”四个字。

Hoel 在特定可证伪性框架中将持续学习作为避免替代困境的方案，并据此质疑基线固定模型的意识。29 本文不把持续学习设为必要条件，并要求比较局部干预组织而非仅比较通常输入输出；这构成需要进一步分析的理论分歧，不表示本文已经反驳其论证。

一个批次共用权重或损失汇总，不自动形成承受总损失的统一主体。痛苦需要负面效价的独立解释；高损失或大梯度不能代替。学习能力作为演化线索也不要求每一体验时刻都在学习。22

11.3 时钟是实现条件，不是体验时钟

常规数字处理器通常用时钟协调更新，时钟频率也不等于指令或推理频率。晶振等提供定时参照，计算由电路、状态和信号实现，能量来自供电。23

神经系统有多种振荡和跨频关系，不能把它们等同于统一安排全部计算的 CPU 时钟。24 人工系统也可以异步、事件驱动，例如 Loihi 2；异步通信还可以与某些算法同步共存。25

因此，同步与异步是实在的实现差异，但不能单独作为意识分界。本假说关心它是否改变参照构成的因果关系和相对时序，而不是是否由晶振产生脉冲。

11.4 自述不是直接读出

思维链可以包含真实计算，也可以不忠实反映答案成因。文字中的更正、计算竞争与被体验到的犹豫仍需区分。26 本文不依据“我有意识”的输出认定体验，也不因它由机器生成就否定一切可能。

12. 理论后果、限制与反驳条件

12.1 可争论的后果

第一，在相关组织保留的条件下，参数是否更新、是否使用全局时钟和是否采用生物材料，都不单独改变分类。第二，增加知识、任务表现或共享数据，不保证参照维度单调增强。第三，较弱的跨内容影响可以连续增强，而不必通过一次由图连通性规定的主体合并。第四，简单控制器和未训练系统若实现参照构成，不能被另设门槛排除。

这些都是本假说的承诺，不是自然界已经接受的结论。它们为思想实验比较和将来的经验研究提供明确目标，但本篇无需先执行实验才能提出。

12.2 不能回避的风险

过度归属。 所定义的组织量适用于许多控制系统。本稿不再声称已经排除所有高级控制器。三位机制是公开承担的困难案例，而不是藏在“足够复杂”后面。

参照识别与粒度。 候选接口、背景协议、路径分离和时间范围会影响画像。本稿给出模型相对的定义和有限最小支撑族，不提供自然界唯一的尺度选择或主体计数算法。多个合理模型的差异是结果的一部分。

现象映射。 组织画像不是主观强度量表；响应分化也只是指定条件下的可能区别，而非当下同时体验的内容数。具体颜色、疼痛、清醒度或自我感的对应关系尚未导出；有体验也不自动意味着会受苦。若完全没有参照构成仍存在体验，或该构成对体验系统性无关，本假说需要核心修订。

渐变与相变。 本稿反对以任意二分图开关替代程度，但不证明所有现象边界连续。分岔、失稳或不同尺度的组织变化可能使某些维度突然改变，必须由具体机制说明。

组织与价值。 智能、体验、道德地位与伦理行为分别需要论证。人类伤害动物的事实不证明伤害正当；机器有体验或更聪明也不保证善意。本文不能由意识假说推出人机共存必然实现。

12.3 论文的经验地位

思想实验能够显示定义冲突、错误推论和理论代价，不能独自建立心理与物理的规律。已有意识理论的某些预测接受过对抗性检验，未获最终裁决不等于从未进行研究。27 本文的完成标准是论证与适用范围可供审稿，而不是宣布意识难题已经解决。

13. 结论

内生参照场假说把体验视为实际参照构成的现象方面，并以多维、跨尺度的组织画像替代由单一连接、权重变化或任务成绩决定的开关。它同时区分组织类型、实现实例和因果接续，从而约束共享记忆、训练、硬件迁移、复制及分支的归属。

二十一个思想实验迫使本稿作出两种同样重要的承诺：一方面，不能凭描述、共同资源或阶段名称制造意识边界；另一方面，不能因为简单人工组织的体验结论不合直觉就添加例外。本文的独立贡献在于具体的分级参照方案、作用幅度与响应分化的分离、通常行为与内部参照画像的分离，以及有限支撑分析对归属判断的约束。普适的物理粒度、完整的现象对应和精确主体计数，仍是该理论需要面对的开放问题。

参考文献

1. Safron, A. (2020). An Integrated World Modeling Theory (IWMT) of Consciousness: Combining Integrated Information and Global Neuronal Workspace Theories With the Free Energy Principle and Active Inference Framework; Toward Solving the Hard Problem and Characterizing Agentic Causation. Frontiers in Artificial Intelligence, 3, 30. DOI.
2. Williford, K., Bennequin, D., Friston, K., & Rudrauf, D. (2018). The Projective Consciousness Model and Phenomenal Selfhood. Frontiers in Psychology, 9, 2571. DOI.
3. Cleeremans, A. (2011). The Radical Plasticity Thesis: How the Brain Learns to Be Conscious. Frontiers in Psychology, 2, 86. DOI.
4. Merker, B. (2013). The efference cascade, consciousness, and its self: naturalizing the first person pivot of action control. Frontiers in Psychology, 4, 501. DOI.
5. Albantakis, L., et al. (2023). Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms. PLOS Computational Biology, 19, e1011465. DOI.
6. Bayne, T., & Chalmers, D. J. (2003). What is the Unity of Consciousness? In A. Cleeremans (Ed.), The Unity of Consciousness: Binding, Integration, and Dissociation. Oxford University Press. 作者稿.
7. Roelofs, L., & Sebo, J. (2024). Overlapping minds and the hedonic calculus. Philosophical Studies. DOI.

8. Chalmers, D. J. (2019). Extended Cognition and Extended Consciousness. In M. Colombo, E. Irvine, & M. Stapleton (Eds.), Andy Clark and His Critics. Oxford University Press. 作者稿.
9. Dror, S., Bergerbest, D., & Salti, M. (2026). Artificial Intelligence as an Opportunity for the Science of Consciousness: A Dual-Resolution Framework. arXiv:2509.07001v3. 预印本，v3，2026-07-13；首版 2025 年。 版本.
10. Sendall, J. (2026). Event Horizons, Spacetime Geometry, and the Limits of Integrated Consciousness. arXiv:2512.23105v2. 预印本，v2，2026-02-06；首版 2025 年。版本.
11. Birch, J., Schnell, A. K., & Clayton, N. S. (2020). Dimensions of Animal Consciousness. Trends in Cognitive Sciences, 24, 789-801. DOI.
12. Josipovic, Z., & Miskovic, V. (2020). Nondual Awareness and Minimal Phenomenal Experience. Frontiers in Psychology, 11, 2087. DOI.
13. Chalmers, D. J. (1995). Absent Qualia, Fading Qualia, Dancing Qualia. In T. Metzinger (Ed.), Conscious Experience. 作者稿.
14. The New York Declaration on Animal Consciousness (2024). 立场声明，非单项实验。 原文.
15. Sneddon, L. U., Braithwaite, V. A., & Gentle, M. J. (2003). Do fishes have nociceptors? Evidence for the evolution of a vertebrate sensory system. Proceedings of the Royal Society B, 270, 1115-1121. DOI.
16. Freiburger, T., Miller, N., & Skinner, M. (2024). Olfactory self-recognition in two species of snake. Proceedings of the Royal Society B, 291, 20240125. 原文.
17. Boisseau, R. P., Vogel, D., & Dussutour, A. (2016). Habituation in non-neural organisms: evidence from slime moulds. Proceedings of the Royal Society B, 283, 20160446. DOI.
18. Brown, T. B., et al. (2020). Language Models are Few-Shot Learners. NeurIPS 2020. 作者稿.
19. Hugging Face (访问于 2026-09-20). How caching works. Transformers Documentation. 官方文档.
20. Lewis, P., et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NeurIPS 2020. 作者稿.
21. Behrouz, A., Zhong, P., & Mirrokni, V. (2024). Titans: Learning to Memorize at Test Time. arXiv:2501.00663；提交于 2024-12-31。 作者稿.
22. Birch, J., Ginsburg, S., & Jablonka, E. (2020). Unlimited Associative Learning and the origins of consciousness: a primer and some predictions. Biology & Philosophy, 35, 56. DOI.
23. Intel (访问于 2026-09-20). What Is Clock Speed? 官方说明.
24. Buzsáki, G., & Watson, B. O. (2012). Brain rhythms and neural syntax: implications for efficient coding of cognitive content and neuropsychiatric disease. Dialogues in Clinical Neuroscience, 14, 345-367. 原文.
25. Shrestha, S. B., Timcheck, J., Frady, P., Campos-Macias, L., & Davies, M. (2023). Efficient Video and Audio processing with Loihi 2. arXiv:2310.03251. 作者稿.

26. Turpin, M., Michael, J., Perez, E., & Bowman, S. R. (2023). Language Models Don’t Always Say What They Think: Unfaithful Explanations in Chain-of-Thought Prompting. NeurIPS 2023. 作者稿.

27. Cogitate Consortium et al. (2025). Adversarial testing of global neuronal workspace and integrated information theories of consciousness. Nature, 642, 133-142. DOI.

28. Hanson, J. R., & Walker, S. I. (2021). Formalizing Falsification for Theories of Consciousness Across Computational Hierarchies. Neuroscience of Consciousness, 2021(2), niab014. DOI.

29. Hoel, E. (2026). A Disproof of Large Language Model Consciousness: The Necessity of Continual Learning for Consciousness. arXiv:2512.12802v3；2026-01-19，首版 2025 年。 版本.

30. Shalizi, C. R., & Crutchfield, J. P. (2001). Computational Mechanics: Pattern and Prediction, Structure and Simplicity. Journal of Statistical Physics, 104. 作者稿.

写作与方法说明

问题构思及研究方向来自作者与 AI 助手的持续讨论。AI 工具参与文献检索、思想实验构造、论证检查和文字编辑。本文未把 AI 评审当作外部同行评审，也未把理论推演当作经验数据。v1.0 为本研究首次公开版，不表示已获期刊接受或经验验证。此前的讨论稿由本稿取代；本次公开稿补充响应分化、最小支撑族和完整通常输入上的辨识对，保留多维渐变立场，不将分析容许度用作意识存在阈值。本文是独立研究文本，不修订任何此前论文或三位一体协定正典。