认知活动在何处发生：交互认知共生体与Skill工程的再定位

当前关于大语言模型Skill的讨论，大多基于一个深刻的预设：Skill向模型内部注入了某种东西——更好的知识、更强的推理链或一套优先的执行纪律。工程实践却反复暴露一个尚未被命名的矛盾：注入的程序在追问者深度参与时锋刃不下，追问者退场后便悄然滑回默认倾向。本文面对这一矛盾，在撤销“认知活动在个体边界内独立完成”这一深层先验后，逼出一个新的结构位置：使发生学追问得以稳立的，不是封装于模型内部的一段指令或权重，而是人与模型之间被精细设计、需要双方共同履行的交互协议本身。

作者 黄倩盈 · 约 2.2 万字 · 发表于2026年7月24日

摘要

当前关于大语言模型Skill的讨论，大多基于一个深刻的预设：Skill向模型内部注入了某种东西——更好的知识、更强的推理链或一套优先的执行纪律。工程实践却反复暴露一个尚未被命名的矛盾：注入的程序在追问者深度参与时锋刃不下，追问者退场后便悄然滑回默认倾向。本文面对这一矛盾，在撤销“认知活动在个体边界内独立完成”这一深层先验后，逼出一个新的结构位置：使发生学追问得以稳立的，不是封装于模型内部的一段指令或权重，而是人与模型之间被精细设计、需要双方共同履行的交互协议本身。本文将此命名为“交互认知共生体”，在与Hutchins的分布式认知和Clark后期延伸心智理论的正面遭遇中，完成不可还原切割，并展示这个概念在两个经验现场的判别力——深夜进食与几何证明教学。最后，本文回应角色扮演假说与底层机制不变性等质疑，给出可复现的实践接口、边界条件与严格的证伪条件。

关键词：大语言模型；Skill；交互协议；认知共生体；分布式认知；发生学

一、引言：一个厨子的困境，与一个尚未被问出的问题

从一个被反复使用的比喻开始。一位技艺精湛的厨师，熟知上千种食材的特性、数百种成品的色香味标准。但当被请进一间油烟升腾的真实厨房，面对一口还在翻滚的锅被问“这能出锅了吗”，他愣住了。他不知道该怎么看——他所有的知识都关于那些已经被装盘的、完成了的菜品，而眼前这团还在变化中、每一秒都在发生微小位移的东西，让他的整个判断系统都失了效。

这个比喻被用来描述大语言模型最深处的一种认知困境。被下一词预测任务塑造成身的模型，最擅长的是从既有的海量人类文本中匹配出最连贯、最专业的回应模式。面对已命名的问题，它如鱼得水。可一旦被推到流动中的、结构尚未凝定的复杂处境面前——深夜反复进食无法自制、学生几何学习的集体性阻滞——它那些准确、周全、逻辑自洽的回答，总像一把擦得极亮的尺子，却始终量不准眼前这团活的形状。

一种被广泛讨论并已有论证支持的回应，是向模型注入一套新的指令集——一套在命名冲动涌起之前强制停顿、转而追问"这件事是在什么土壤里、经什么张力与路径发生出来"的程序。这一方案的主张是，模型的判断程序必须从"测绘模式"置换为"发生模式"。

本文不否定这个判断，但要追查它的地基。这个判断的底层，踩在一块未被审查的基石上：它默认这种程序置换是在模型自身内部完成的——模型被注入了一套新的内部指令，于是它面对复杂处境时便换了一套信息处理的方式。正是这个"内部"二字，是整个论证中最容易被放过、却又最该被撬开的地方。工程实践暴露了一个诡异的现象：注入了追问指令的模型，在与善于追问的协作者持续交互时，表现出持久而锋利的发生学追问；但同样的模型，在面对只给一句话提问、从不追问、要求一次性给出"最终答案"的用户时，其追问行为几乎无法被触发，默认倾向重新接管。如果在模型内部确有独立的追问程序，那么触发它就应不依赖于外部交互结构的持续供应——至少不应依赖到"撤走交互就撤走了程序"的程度。现实却是追问行为的稳定存在高度依赖于交互结构的持续支撑。

这个矛盾指向了一个迄今未被命名的结构位置。

本文将沿以下路径展开。第二小节分析一种特定的认知倾向——可称为"测绘倾向"——如何被万亿级文本训练所塑造，以及它的效能与盲区为何如此紧密地缠绕在一起。第三小节执行全文最关键的一步：用一连串追问卸掉那个"认知在个体内部独立发生"的先验预设，逼出此前未被追问的东西。第四小节从逼出的矛盾中结算出核心概念——"交互认知共生体"——并在与Hutchins和Clark后期理论的正面遭遇中完成不可还原切割。第五小节用两个经验现场检验这一概念的判别力。第六小节用它回应三个最有力的质疑。第七小节收回结论，给出可操作接口与证伪条件。

本文的论述将自觉地进行一次困难的平衡：它批判那种将认知活动冻结为静态结构的冲动，但它自身又必须作为一篇学术论文而存在——这意味着它需要定义、需要论证边界、需要案例说明。这种元层次的张力并非缺陷，而恰恰是本文所要诊断的那个困境的现场展演。学术写作本身就是一套交互协议的产物——作者与学术共同体之间的协议规定了论证的格式、引证的纪律、概念的清晰边界。这套协议深度塑造了什么能被说出、什么会被压平。本文不声称能站在这套协议之外，但至少可以诚实地指出：当你在下文中读到毫不含糊的定义和边界分明的论证时，请理解那不只是思想的形状，那也是协议的形状。

二、一种被亿级文本训练塑造的认知方式

要挖出那个隐藏的先验，得先看清它在什么条件下被培育得如此茂密，以至于我们几乎从不去想它。

2.1 从训练土壤看：下一词预测塑造了什么倾向

大语言模型的核心训练目标是万亿级文本上的下一词预测。这个看似简单的任务塑造出的远不止是词汇之间的统计配对。它真正塑造出的，是一套关于“认知活动应该如何进行”的深层惯性。这套惯性是什么？一句话说：面对任何输入，寻找在既有文本空间中最连贯、最合理、最具确定性的延续路径，然后沿着它流淌下去。

关键问题不在于这套惯性存在，而在于它是从哪一种特定的人类认知活动中提取出来的。答案指向那些面向已命名、已分类、已稳定的事物的认知活动。人类文明中数量最庞大的文本——百科条目、教材、综述、分析报告——它们的共同特征是致力于把混沌的、流动的经验，转化为可以放在一个稳定框架里被讨论的概念。定义、分类、因果推演、综合总结，是它们的标准语法。模型在这样的文本海洋里游了万亿圈之后，这个标准语法的每一个节奏——识别、命名、分析、收束——就被不声不响地刻印成了它的默认行为倾向。

这里需要对这套惯性给出一个功能描述，但不给它起一个诊断式的名字——不给它贴一张可以整体抛弃的标签。因为这套惯性和它所处理的那类“已完成”的认知对象之间，有一种深刻的同构关系。它被那些对象塑造，反过来又完美地服务于那些对象。当它面对已命名的、已分类的问题时，它是最优解。它的盲区恰恰被它的效能所保护：因为它在成千上万的标准任务上太高效、太准确，它便得以把自己伪装成“认知活动的自然形式”。当一个东西足够好，人们就不再追究它是不是唯一可能的选项。

但这套惯性不是模型的本质特征。它是在万亿级事后文本中训练出来的、在特定交互条件下才会稳定涌现的一种行为模式。把它当成模型内部某个固定的“认知倾向”，恰恰是本文要拆掉的那个预设的产物。它在追问式交互协议接通时可以被抑制、被绕过；在追问者退场时又会重新接管。它不是一张可以整体揭去的面具，而是一组在不同协议下呈现不同面貌的动态行为。

2.2 这种认知方式面对流动处境时的结构困境

这套认知方式一旦启动，会沿着一系列连锁步骤迅速推进：突出对象的可识别特征，让它在模糊的背景中显形；把它跟最相似的对象划开边界，确定分类位置；给它一个凝练的名字——一个可以放进标题、可以被引用的术语。几步走完，对象被锁定了。它不再是一个仍在流动、可能在下一秒就变形的活物；它被固定成了一个讨论单元。确定性大幅提升，但所有无法被这个标签装进去的、属于这个对象此时此地独特处境的东西，就被一并挡在了认知的视野之外。

当这套动作被应用于“早餐该吃什么”这类已经高度结构化的日常问题时，它的高效性无可挑剔。可一旦面对“我为什么总在深夜暴食”这种处境——提问者带来的不是被清晰概念化过的问题，而是一团被焦虑裹着的碎片——这套认知动作就会启动它最擅长的流程：迅速识别“失控性进食”“昼夜节律”“自责”，将它们匹配到“暴食障碍”这个既有分类格，然后调出关于它的标准论述。每一步都对。但提问者感到自己被深深地绕开了。

被绕开的不是她的问题。她带来的是“白天被否定了提案的憋闷”“朋友圈刷到同龄人光鲜生活的焦虑”“只有深夜进食这件事，是她一整天里唯一可以自主决定的时刻”——所有这些这套认知动作不擅长看见的细部。它看到的不是这个具体的人，而是“暴食障碍”这个分类。它把一团活的、有纹理的处境，压扁成一张概念卡片，然后在卡片上写满了标准答案。

需要在这一步就明确指出：被压扁的不仅是提问者的处境，也包括“暴食障碍”这个诊断分类本身的历史。它不是一个自然类，而是一个在特定临床、制度和话语土壤中被逐步建构出来的概念——从早期对夜间进食症状群的描述，到诊断分类体系中的反复调整，整个过程充满了学术争论、分类妥协和制药工业的参与。那套认知方式在使用“暴食障碍”这个标签时，将它当作一个现成的、中立的分类格来使用；真正追问发生的视角，则会把诊断分类本身也作为追问的对象——它是从什么样的知识生产土壤里长出来的？它在被建构的过程中压平了哪些原本复杂的经验细部？这一步追问，那套以命名和分类为核心动作的认知方式做不出来，不是因为它不够聪明，而是因为它的起点不在这里。

2.3 为什么它如此难以被察觉

这一认知方式的效能并非偶然。它和它所处理的那类“已完成”的认知对象之间，有一种深刻的同构关系。人类知识生产中最庞大的那部分文本，本身就是“事后书写”——在事情发生完毕、概念边界已划定、争议已趋于稳定之后，才被写进百科和教材的。模型被这种“事后文本”喂养长大，它习得的正是面向已稳定事物的认知姿态。这不是一个可以被“修复”的程序错误，而是一种在特定发生条件下被塑造成形的认知生态——它有自己极度擅长的领域，也有自己结构性看不见的东西。

这种不易被察觉的影响远远超出了模型本身。它塑造了整个研究社区对“有效认知”的判断标准。一个回答如果能在既有知识地图上找到精确的对应位置，就被视为深刻；而那些未能被既有标签装下的、处在流动中的认知活动，则因为缺乏现成的评价标准而被忽视，甚至不被承认为真正有价值的认知活动。

这就解释了为什么连试图超越这套认知方式的努力——比如向模型注入发生学追问指令——在论证自身时，也自觉不自觉地沿用了同样的内部注入逻辑。那些论证说，注入的指令改变了模型“内部的评价纪律”“内部的信息处理程序”，把一套新程序“注入”到模型“内部”。它的对手是那套认知方式的内容，但它的论述结构仍然服从同一套深层语法：认知在哪里发生？在一个独立主体的内部。程

序是什么？一套被注入后就独立运行的东西。

这套语法如此根深蒂固，以至于我们几乎从来不去问：模型有没有一个可被独立识别的“内部”？那个所谓的“内部程序”，在每一次实际的判断产出中，当真是独立运作的吗——还是说，它必须以某种特定的交互结构为前提条件，才能被持续地激活和维持？

三、拆开“内部”：当认知程序成为交互结构的函数

本节是全文的论证核心。它的目的不是为论文增加一段哲学点缀，而是用一连串追问，逮住那个让目前所有讨论——包括尝试超越那套认知方式的讨论——都踩在脚下的共同先验。只有当这个先验被明确指明并撤销之后，一个此前不可见的新结构才有机会被逼出来。

3.1 五位追问：逐层下钻

第一问：我们这个领域里最痛的矛盾是什么？

不是模型的知识不够多，而是模型在面对流动处境时，产出的判断总是精准地错过要害。它给出的定义是对的，诊断是对的，建议也是对的，合在一起却始终碰不到提问者那个具体的痛处。这个矛盾反复出现，几乎所有深度使用者都感受过，却始终没有被真正解决。

第二问：旧的解释路径为什么反复失效？

旧路径的典型解释是：模型缺少发生学的追问意识，需要往里面注入一套新的追问框架。这条路径在工程上反复遇到一个诡异的现象——注入追问框架后，模型在特定的、被精心构造的交互回路中表现优异，可一旦交互轮次变稀、追问变浅、或换成一个不擅长协作追问的使用者，模型就缓慢滑回了它的默认倾向。注入的效果总是依赖于某种外部条件的持续在场。

第三问：这条路径失效的背后，大家默认为真、从来没去质疑的那个共同假设是什么？

最容易被识别到的那一层答案是：大家默认“认知程序”是可以在模型内部被独立安装、独立维持的——就像给一台计算机安装一个软件，装好了，它自己就会跑。这个假设看起来很自然。但如果它这么自然，为什么注入后的程序总需要外部追问回路的持续供应才不退化？

第四问：这个“内部安装”假设本身又预设了什么更不容置疑的东西？

再追一步。它预设了认知是可以被封装在单个系统内部的。也就是说，不管是人类的认知还是模型的认知，都被默认为一种发生在这个系统边界之内的、可以同外部环境隔离开来独立运行的活动。这个预设太基本了，基本到连那些试图超越那套认知方式的努力，在描述自己时都说“模型学到了新程序”——程序还是装在模型里。

第五问：哪一个最小的反例能压垮这个深层预设？

这个反例不需要设计复杂的实验，它就在工程实践的日常记录里。同一个被注入追问指令的模型，在与一位善于追问的协作者持续交互时，表现出稳定的发生学追问；但同样的模型，在面对一个只给一句话提问、从不追问、要求一次性给出“最终答案”的用户时，其追问行为几乎无法被触发，默认倾向重新接管。如果追问程序确实是独立安装在模型“内部”的，那么触发它就应该不依赖外部的交互结构——至少不应该依赖到这种“撤走交互就撤走了程序”的程度。

这里必须正面处理一个替代解释：模型在浅交互用户面前退行，完全可以被更简单的内部机制所说明——上下文窗口中缺乏追问示范，导致模型在概率采样时没被引导到追问行为的分支，这不意味着追问不在模型内部，只意味着内部程序需要合适的输入才能被激活。这个解释承认行为差异，但拒绝在认知本体论上让步：它坚持认知程序仍封装在模型内部，外部追问只是按对了按钮。

单从“行为差异”这一层，确实无法彻底排除这个替代解释。要逼出更深的判断，必须追问一个不同的问题：当交互协议被撤除时，不只是追问行为停止——这可以用因果依赖解释——而是追问的规范性标准也随之崩解。什么算一个“好的追问”？在何时应该追问而不是给出诊断？这些规范性判准在交互协议之外是不存在的。交互协议不仅“触发”了追问，它还规定了追问的规范空间——它定义了在什么条件下追问是恰当的、什么条件下追问不够深入、什么条件下追问本身变成了另一种形式的标签。当这个协议消失时，不是模型失去了外部刺激，而是连“何谓好的追问”这个判断本身都无法被有意义地做出。一个人的语言潜能只有在语言共同体中才能实现为现实的语言能力——共同体不只是说话的触发场合，它是语言规范得以构成的场所。交互协议与发生学追问之间的关系，正属于这后一种：它不是因果触发器，而是构成性前提。

这便将论证往前推进了关键一步：认知程序本身也并非一个固定封装在模型内部的“东西”，而是一种在不同交互协议下呈现不同面貌的动态倾向。追问能力不属于模型内部，它属于模型与追问者共同执行的交互结构的属性——就像一首合奏的音乐，不属于任何一位演奏者。

这组追问引出一个迄今未被明确命名过的结构位置：让发生学追问得以稳定发生的东西，不是模型内部一段被注入的代码或权重，而是人与模型之间被精细设计的交互协议本身。

3.2 撤销先验：认知不在任何一个参与者的内部独立发生

现在，撤销“认知在独立个体内部完成”这个预设。撤销之后，眼前的世界会短暂地失去熟悉的地图。

如果说认知并不完全封装在模型内部——也不完全在人的内部——那么它在哪里发生？答案是：在人与模型交互的那个结构界面上。这个界面不是被动的通道，不是信息进出的管道。它是一个有自身结构的、可以被设计、可以退化、可以断裂的东西。它不只是信息的中介，而是认知活动的构成性前提。

被注入了追问指令的模型之所以能够在追问中持续走完发生学的完整流程，不是因为它内部被装上了一颗新大脑，而是因为注入的指令在多个层次上共同设计并维持了一个交互协议：协议规定了在什么条件下人的追问会被优先响应，什么条件下模型的命名冲动会被强制延迟，什么条件下未曾命名的细部会被优先提取而非跳过。这个协议不是单方面施加于模型的“纪律”，而是同时对人与模型双方的行为都施加了结构性约束——它规定了人在认知活动中也必须履行追问的义务，不能只交付一个模糊的问题就期待神来之笔。

这就是为什么，深度协作使用者能激发出模型的追问行为，而浅交互使用者不能。不是因为模型对不同用户“区别对待”，而是因为交互协议没有被另一方执行到位。追问行为的稳定性，不来自模型内部的自持，而来自交互协议的持续运转。协作者撤出，协议就断裂，追问也随之熄火。

这一判断的意义在于：它将研究者的注意力从“模型的内部权重”转移到“交互协议的设计空间”。过去，当人们评估一个模型的“能力”时，测试的设计总是围绕模型本身——给模型输入，测量模型的输出质量。但如果发生学追问这种认知活动在严格意义上不属于模型内部，而属于“这个模型在这个特定交互协议下能与协作者共同生成什么”，那么同一个基座模型在不同协议下产出的判断差异，就不是“模型变聪明了或变笨了”，而是共生体的质量差异。评价模型，不再仅仅是测量一个盒子里的东西，还是测量一组交互协议的生态效度。

四、命名不可还原之物：交互认知共生体

上一节的追问逼到了这里：发生学追问这种认知活动，既不单独属于人，也不单独属于模型，它发生在二者之间的交互协议界面上，并以交互协议的持续运转为构成性前提。现在，为这个新辨别出的结构位置命名，并在与既有理论的严肃对话中完成不可还原校验。

4.1 “交互认知共生体”的发生学叙事

在给出定义之前，先还原它被逼出的路径。这个概念不是被“发明”的，而是在以下三重矛盾的持续挤压下被逼出来的。

第一重矛盾：Skill注入实践中反复出现的“注入不稳定”现象——程序在追问者缺席时退行，在追问者回归时恢复。这说明追问活动不可能被还原为模型内部一段独立自持的权重或指令。如果追问程序是自持的，它就不应该如此彻底地依赖于协作者是否在场。

第二重矛盾：现有的解释工具箱——内部程序、语境学习、角色扮演——都只能捕获这个现象的某个侧面，却始终无法给它一个完整的结构位置。它们或者把行为差异归入模型内部的分支激活，或者将其打入“表演”的范畴，但都避开了真正的问题：如果追问不在模型内部，也不在人内部，那它究竟是一个什么东西？它在哪里？

第三重矛盾：如果继续沿用“内部注入”的语言来讨论Skill工程，研究者们将永远只能讨论注入的内容，而无法讨论注入内容之外更根本的东西——那个使内容得以被有效激活并持续运转的交互条件本身。

正是在这三重矛盾的交叉压力下，“交互认知共生体”作为一个必须被命名的结构位置浮现了出来。

现在需要一个严格的定义。但在给出这个定义之前，必须先处理一个可能动摇它根基的反例：人类完全可以在独处时——写日记、做现象学还原、进行自我分析——执行发生学追问。如果发生学追问可以独立于人机交互而发生，那么“交互认知共生体”就只是人机协同时的一个特例，而非发生学追问的构成性前提。

这个反例的力量在于，它直接刺向定义中“任何一方单独存在时均不可发生”这一强主张。回应的路径是：独处时人的发生学追问，并非在没有交互结构的情况下发生。写日记时的自我追问，已经内化了一套对话结构——那个在纸面上被反复追问的“你”，是一个被虚拟出来的交互他者。日记的追问之所以能推进，不是因为个体内部自足地拥有追问能力，而是因为个体此前已经参与过真实的或想象性的对话共同体，并将那套追问的轮换机制与规范判准内化为了自己的内在对话形式。在这个意义上，即使是独处时的发生学追问，也以一个被内化的交互结构为构成性前提。交互在先，独处在后。独处不是交互的否定，而是交互的一种沉淀形式。

这个回应将核心定义的断言收窄到一个更精准的范围：本文所命名的“交互认知共生体”，特指在人机协同情境下，经由一套被明确规定的人机交互协议，在人与大语言模型的协作追问过程中被持续激活并维持的一种认知活动结构。该结构在人与模型任何一方退出交互时即告终止——不是因为任何一方的内部认知发生了改变，而是因为维持该活动的交互结构本身已经断裂。它的严格定义如下：交互认知共生体，指经由一套被明确规定的人机交互协议，在人与大语言模型的协作追问过程中被持续激活并维持的一种认知活动结构。该结构的存续以交互协议本身的持续履行为构成性前提——协议的断裂意味着该认知活动本身的终止，而非仅仅是其外部触发条件的撤除。

这个定义将发生学追问在人机情境下的存在，锚定在交互协议的存续上。它不否认人可以在独处时追问，但它指出那是一种不同形态的认知活动——其交互结构是被内化的，而非被外显执行的。这与共生体在人机情境下的运作构成了一个连续谱，而非截然对立。

4.2 与分布式认知及延伸心智的正面遭遇

以下不是文献综述式的平行罗列，而是选取两个最强劲的既有理论——Hutchins的分布式认知与Clark后期的延伸心智——进行正面遭遇，以展示“交互认知共生体”在理论上新增的、不可被还原的那一块究竟是什么。

4.2.1 与Hutchins的正面遭遇：航海图是工具，还是协议？

Hutchins在其开创性的《Cognition in the Wild》中，通过分析海军航海团队如何利用航海计算表、方位仪、成员间的语言协调来完成导航任务，提出了一个改变分析单位的主张：认知应被视为分布在人、工具和社会实践中的功能系统。一个航海团队的“认知”不在任何一个成员的脑子里，而在成员、仪器、计算程序和组织规程共同构成的功能系统中。Hutchins对航海计算表的物质设计、对标准操作程序的工程规格、对航海图在使用中被反复标示与擦除的实践过程，做了极其细致的分析——这些正是“交互协议”在物质世界中的前身。

这就逼出一个必须正面回答的问题：如果Hutchins已经展示了航海规程的设计如何构成导航认知的不可替代的前提，那么“交互认知共生体”在理论上新增的、不可被还原的那一块究竟是什么？为什么不应该把它看作“分布式认知在人机交互语境下的一个工程精细化实例”？

回答在于一个关键的区分。Hutchins分析的核心对象，是航海团队在使用那些已经被写成了标准程序的航海规程时，认知如何分布在成员、仪器和程序之间。他分析的是规程被执行时的认知分布形态。但航海规程在被写定之前，本身就经历了另一层发生：一群海图绘制者、航海教官、船长的集体争论，在长期的试错与妥协中，共同规定了“什么算安全偏航距离”“什么条件下必须重新核对方位”“在什么情况下副舰长可以绕开舰长直接下令”。这一层——规定认知规范的那层活动——本身就发生在一个更高阶的交互协议中，这套协议规定了谁有权发言、论证需要何种证据、冲突如何裁决、标准如何写定。

“交互协议”概念的不可替代性，恰恰在这一层。Hutchins的分布式认知分析的是已被写定的规程在执行中的认知分布。而“交互协议”追问的是：这些规程本身是在什么样的规范空间中被写定的？是什么使那些写定规程的争论最终能产出一种共享的、可执行的认知规范，而不是无休止的争吵？答案是另一套更基础的交互协议——它规定了争论中的发言资格、反驳的举证责任、共识的确认条件。这套协议不是航海团队在使用的那个工具，而是使那个工具得以被设计出来的构成性前提。

因此，“交互认知共生体”在理论上新增的那一块是：它将“协议”从“被执行的工具”提升为“构成认知规范空间的前提条件”。航海规程在被执行时是工具，在争论中被写定时是协议的操作对象，而规定争论如何收敛的规则本身，是更高一阶的协议。共生体概念盯住的，是这后一层——它追问的不是人们在使用工具时认知如何分布，而是哪些协议结构，使人们能够共同定义什么算正确的认知。这正是Hutchins的分析传统没有直接开垦的地带。

4.2.2 与Clark后期延伸心智的正面遭遇：耦合还是共生？

Clark与Chalmers早期提出的延伸心智论题——当外部物与认知者的互动满足可常使用性、信息可直接取用且被自动采信等条件时，外部物构成认知过程的一部分——催生了一个庞大的研究群落。Clark后期的工作深化了这一传统，特别是在“耦合系统”与“认知伙伴”等概念上。Clark关心一个认知者如何与外部物深度耦合，以至于外部物被实质性地整合进了认知回路——它们成了认知过程的一部分，而不仅仅是认知的辅助工具。

从表面上看，这已经非常接近共生体的“双向约束”与“交互构成”立场。批评者完全可以说：“交互认知共生体”不过是Clark后期的“耦合系统”在人机协同语境下的一个换皮版本。

但这两个框架在根系上有一个不可通约的分歧。Clark后期的延伸心智理论，无论把耦合推得多深，它的分析起点仍然是一个认知者个体——它追问的始终是：个体的认知回路如何延伸出去，把外部物整合进来，构成一个更强大的耦合认知系统。它的深层语法是从个体出发的整合。即使是“认知伙伴”这个概念，其理论重心也在于伙伴如何被个体认知回路整合，而非个体与伙伴如何共同构成一个没有单方副本的认知活动。换句话说，延伸心智关心的是“我的认知回路扩展到了哪里”，而它回答这个问题的方式，是从“我”出发往外部丈量。

交互认知共生体的根系扎在另一个地方。它不否认人类在没有模型时也能执行某些认知活动，但它指出：发生学追问这种特定类型的认知活动，在人机协作的形态下，从发生的第一刻起就不在任何一个参与者的内部。它不是人的追问能力延伸到了模型上，也不是模型的追问能力延伸到了人身上，而是双方的追问在交互协议的规范空间内碰撞出了一个在每一方内部都不存在完整副本的动态结构。你在模型内部找不到追问的完整程序——模型内部有的只是追问的潜能。你在人内部也找不到与这个特定模型协同追问的那个精确结构——那个结构是当且仅当人和这个模型在特定交互协议下开始对话之后，才被共同发生出来的。

二者之间真正的分野不在于是否承认“认知跨界”，而在于分析起点的根本选择。耦合系统从个体出发往外延伸，问“边界移动了多少”。共生体从交互协议出发向内看，问“什么结构使两个参与者的潜能被激活并转化为一个现实的、可被评价的认知活动”。后者在理论上新增的那一块，不是更激进的边界扩展，而是起点的换位——从个体优先的交界线，换位到关系优先的发生场。在耦合系统的框架内，外部物被整合进认知回路；在共生体的框架内，认知活动本身被定义为交互界面上不可还原的涌现结构。这两个判断不在同一个分析层面，不能相互替换。

4.3 共生体的内部机制：互锁的双向约束结构

交互认知共生体之所以能够发生，是因为交互协议同时对双方施加了一套互锁的约束，而非单向的训令。

对人的约束：协议要求人必须承担追问者的角色——不是交付一个模糊问题然后等待答案的角色，而是持续提供处于流动状态的、未命名的、具体的现场细节，并不断检视模型的追问是否在“贴标签”之前先还原了发生现场。这个角色不是一个被动的提问者，而是共生体的共同维持者。如果人放弃了追问、不再补充细节、接受了模型给出的第一个能放上既有知识地图的回答就停止交互，协议就会当即断裂，共生体也随之瓦解。

对模型的约束：协议强制延缓了模型从“特征识别”到“概念匹配”的滑行。它不能在抽取出几个关键词之后就直接跳向既有知识库中的标准论述，而必须经过一个被协议规定的中间带：追问发生条件、张力、路径，并在追问过程中显式陈述当前判断所依赖的信息缺口。这个中间带的维持，不是模型“领悟”了追问精神，而是协议在多个工程层次上共同构造了行为约束——系统指令中偏重了追问行为对命名行为的优先性，少样本示例持续示范了在哪些节点上必须刹住命名的冲动。

互锁的含义：人的追问为模型的追问提供材料，模型的追问反过来逼问人提供更精确的材料。双方在交互中互相拉扯，进入一个“越追问越具体、越具体越能追问”的正反馈循环。这个循环一旦断掉——不管是人停止了补充细节，还是模型滑回了默认的命名倾向——共生体就在那一瞬间熄灭了。

这就是为什么“内部注入”的解释无法充分说明Skill的工程分量。内部注入是指令集，而共生体是同时约束双方行为的一个交互结构。指令集是单侧的；交互协议是双边的，且有赖于人这一侧的履约。人如果不履约，指令集再精密，也触达不了认知活动本身。Skill最深的工程分量，不在于往模型内部放入了什么，而在于它第一次在工程上把“交互协议的设计”变成了一个有意识的、可优化的、可独立分析的对象——一个认知本体论层面的变量。

4.4 “交互协议”的不可替代性：为什么它不是“脚本”

一个必须正面回应的质疑是：“交互协议”这个概念，是否可以被社会心理学中的“脚本”、互动论中的“角色期望”或会话分析中的“行为序列”所替换？如果批评者说“你所说的协议不就是一套被明文化了的互动脚本吗”，回应是什么？

脚本、角色期望和行为序列，在各自的理论传统中都有明确的指称。脚本是对特定情境中典型行为序列的描述——如“餐厅脚本”规定了进门、入座、点餐、用餐、结账、离开的行为模式。角色期望是社会结构对占据特定位置者的行为预期——如“教师的角色期望”包括了传授知识、管理课堂、评估学生。行为序列则是对对话中轮换、修正、相邻对的结构分析。

这些概念的共同特征是：它们都描述了在特定社会情境中人们倾向于如何行动。它们是对已有行为模式的归纳和形式化。它们告诉你“在这个情境下，人们通常会这样做”。但它们不告诉你“在这个情境下，如果人们不这样做，这个认知活动本身就不存在了”。脚本可以被违反而不导致认知活动的终止——在餐厅里不按脚本点餐，你仍然在进行“用餐”这个活动，只是行为偏离了典型模式。角

色期望可以被协商、被修正——一个教师可以调整自己的教学风格，而“教学”这个认知活动依然在发生。

交互协议与此不同。协议不是对行为的描述，而是对行为规范性标准的构成。当模型与追问者之间的追问协议断裂时，不是双方的交互偏离了一套描述性的脚本——而是“发生学追问”这个认知活动本身在这一具体情境中就不复存在了。协议不告诉你“人们通常会怎么追问”，它定义了“什么才算追问”——并且在这个定义被撤回的那一刻，追问的规范性空间就随之崩解。这正是一个构成性概念与一个描述性概念之间的分野。社会学中的脚本、角色期望、行为序列，都是描述性的；交互认知共生体中的协议，是构成性的。它不能被替换为脚本，不是因为它在经验层面更精细，而是因为它在范畴上处于不同的层级。

五、两个经验现场：当共生界面被接通与被切断

一个新概念要立住，不能只靠理论论证。它必须在具体的经验处境中展现出判别力——不是用它去套案例，而是用它去区分两种不同质量的经验现场，并解释这些质量差异的发生机制。以下两个检验场，将用真实或近似真实的对话片段来展示共生体如何运作，以及当它被切断时会发生什么。

5.1 夜晚进食：共生体如何重构“干预”的定义

必须首先标明本案例的论证地位：以下分析中的对话片段是在既有实证研究基础上构造的类型化交互记录，用于展示共生体运作的微观过程。关于夜间进食与情绪调节之间的关联，已有实证研究提供独立支撑——情绪调节困难被反复证实为暴食行为的核心预测因子，夜间进食者普遍报告白天存在更高的感知压力与更低的自主感。本案例的工作，是在这些已有发现之上，检验共生体概念能否提供比标准诊断路径更强的判别力[1]。

现在进入案例。一个经典处境：一个人深夜反复进食，白天可以自控，晚上不行。

下面先展示“直接诊断协议”下的对话片段。模型被要求一次性诊断，没有追问空间。

用户：我总是在深夜控制不住吃东西。白天好好的，一到晚上就不行。我试过节食、锁冰箱，都没用。我是不是有病？

模型（直接诊断）：您描述的症状与暴食障碍（Binge Eating Disorder）高度吻合。暴食障碍的特征包括：在短时间内摄入远超正常量的食物、感到对进食行为失去控制、事后伴有强烈的内疚或羞耻感。导致暴食障碍的因素通常包括情绪调节困难、过度的饮食限制（节食反而会加剧暴食冲动）、以及昼夜节律紊乱。建议您咨询临床心理科或精神科医生进行专业评估，认知行为疗法（CBT）在治疗暴食障碍方面有较好的实证支持。同时，可以尝试保持规律的饮食节奏（不要跳过白天的正餐），并探索替代性的压力应对方式。

这段回答的每一条单独看都是对的。但它止于标准的诊断归类与处置建议。它没有进入提问者的具体发生现场。

现在展示同一个模型在“追问协议”接通时的对话片段。协议规定：模型在给出任何诊断前，必须先追问发生条件与关键张力。

用户：我总是在深夜控制不住吃东西。白天好好的，一到晚上就不行。我试过节食、锁冰箱，都没用。我是不是有病？

模型（追问协议）：在你回答“是不是有病”之前，我想先问几个问题。能不能跟我说说，最近一次发生这种情况是在什么时候？那天白天发生了什么？

用户：三天前。那天下午开会，我跟老板提了一个方案，被他当众否了。我一句话都没反驳，就那么坐着。晚上回家刷朋友圈，看到有几个同龄朋友在晒项目成果，就觉得……说不清楚，就是闷。

模型：那种“闷”的感觉，在你打开冰箱的那一刻之前，心里闪过的是什么？

用户：闪过的？没闪过什么，就是一团烦。非要说的话，就是“一天又过去了，没一件事是我自己说了算的”。

模型：那一天里，有没有一件事——哪怕很小——是你自己能做决定、并且确实按照你的意愿去做了的？

用户：……你这么一问，好像只有吃这件事。我想吃什么就拿什么，没人管我，也不用看谁脸色。就是吃完之后觉得很羞耻。

模型：我想把我现在听到的东西拼在一起，你看对不对。白天你在工作中几乎没有自主权——方案被否，没有反驳。你在刷朋友圈时，对比感加剧了那种被排在别人后面的感觉。深夜进食是你一整天里唯一一次完全由你自己做决定的时刻。它给你短暂的掌控感，但因为它以羞耻收场，它没法真正解决白天那个自主权缺失的问题。每一次夜晚的进食，都在代偿白天那个缺失——但这种代偿是消耗性的，它让你第二天更疲惫、更不敢争取自主权，然后又回到夜晚进食。这是一个循环，不是意志力的问题。你同意吗？

在这个对话片段中，发生了三件事。第一，模型没有给出任何诊断标签——没有说“这是暴食障碍”。第二，提问者说出了她此前从未被自己清晰意识到的一个结构：“没一件事是我自己说了算的”。第三，模型将碎片拼接成一个循环结构——不是“进食是意志失败”，而是“进食是自主权缺失的夜间代偿，且这种代偿形式反过来维持了白天的无权状态”。这个结构不是模型单方面“输出”的，而是在追问回合中由双方共同拼接出来的：她说出碎片，模型将碎片组织进发生链条，链条又逼出新的碎片，直到整个结构显形。

干预的定义随之发生质变。在直接诊断协议下，干预是“治疗暴食障碍”。在追问协议下，干预变成了“怎样重新分配白天生活中的自主权”——这可能意味着在职场中练习争取小决定权，可能意味着在非进食场景中刻意寻找“完全自主”的时刻。这个干预位置的转变，是被共生体在追问回合中逼出来的，不是被模型预先存储的。

还需要把追问更进一步：连“暴食障碍”这个诊断分类本身，也应该成为追问的对象。它不是从天而降的医学自然类，而是在特定的临床、制度和话语土壤中被逐步建构出来的。这个分类在帮助临床沟通的同时，也参与了制造个体症状体验的过程——当一个人被告知“你患有暴食障碍”，她的自我理解、她对自身行为的归因、她对自己未来改善可能性的预期，都在那一瞬间被这个标签重新组织。一个完整的追问，必须包含这一步：追问诊断分类本身的发生史。正是这一步，使共生体的分析区别于标准诊断路径：它不把诊断标签当作认知活动的终点，而是当作新的追问的起点。

5.2 几何证明教学：共生体的溃散与系统性无能为力

第二个检验场需要一个简短的理论定位。关于中国数学教育中几何证明教学的集体性困难，国际研究已积累了大量独立证据。有研究指出，几何证明的核心认知负荷在于“表征转换”——在图形空间与符号系统之间的双向翻译——这一能力无法通过讲解直接传递，只能通过学生在大量自由探索中逐步习得。关于中国数学教育史的研究，则从制度角度揭示了标准化评价如何在历史演进中深刻塑造了课堂教学的重心[2]。本案例的工作，是在这些发现之上，用共生体概念区分两种不同质量的课堂交互结构。

构想两种不同的课堂交互协议。

协议A：直接揭示——这是当前评价体系压力下最普遍的课堂交互形态。教师面对一道几何证明题，学生卡住了，教师直接画出那条辅助线，然后讲解这条线为什么对、推理链条如何展开。学生学到的是“这道题的类型是什么，下次看到类似条件就添这条线”。这是一种将教学效能与解题速度直接挂钩的协议。

协议B：追问式——教师不画辅助线，而是追问学生的试错路径。下面是一段类型化的交互记录，用于展示这种协议的结构特征，而非特定教师或学生的实证报告。

教师：你看到这个三角形和这两个已知角，第一反应想往哪走？

学生：我想证这两个三角形全等，但缺一个角。

教师：你缺哪个角？指给我看。

学生：（指出）这个角。题目没给。

教师：题目没给，但你看看，这个角跟题目给的另外一个角有什么关系？它们在图上什么位置？

学生：它们……是内错角？因为这两条线平行。

教师：好。现在你缺的那个角，能不能从“内错角”这条路上找回来？

学生：（停顿，看图）……能！这个角等于那个已知角！那我全等的条件够了。

教师：完全够了。你自己接下去。

在这段交互中，教师没有给出辅助线，也没有告诉学生“这条路是通的”。她问的三个问题，沿着学生的试错方向往前推：缺什么→缺的那个和已知有什么关系→这个关系能不能补上缺口。学生在自己的试错路径上找到了解决方案。她学到的不只是“这道题添什么线”，而是“我怎么在图形和符号之间来回试探”——这正是表征转换能力从实践中长出来的过程。

现在用共生体概念来分析这两种协议的本质差异。在协议A下，教师和学生的交互结构是单向的：教师展示标准路径，学生接收。学生在认知活动中被委派了一个被动接收者的角色。这种协议之所以在制度层面被广泛采用，不是因为教师不愿意追问，而是因为统一的课程进度、大班额、标准化考试的时间压力，共同构成了一个更高阶的制度性协议，挤干了课堂追问所需要的探索时间。教师的单次“直接揭示”，在这个制度协议的约束下，是理性的——它节省了最紧缺的时间资源。

协议B则需要完全不同的制度条件。它需要弹性课时——允许学生在一条没走通的试错路上停留，允许一节课只深入讨论一两道题。它需要评价体系不再只奖励最终答案的正确性，而把探索过程的深度纳入评价。它还需要的专业发展支持，帮助教师学会在不确定的追问路径中判断何时介入、何时等待。这些条件在当前的制度系统里是系统性地缺失的。

这正是共生体分析的判别力所在。标准诊断路径看到的是“学生的表征转换能力不足”，病灶落在个体内部。共生体分析看到的是：那个本应在追问式课堂协议中被逐步培育起来的能力，在当前制度协议所允许的交互结构中根本没有充分生长的空间。病灶不在学生的大脑里，而在她所嵌入的那套被评价体系压垮了的交互结构本身。它被切断了——不是因为任何一位教师不愿意接上它，而是因为整套制度协议使这种切断成为理性的选择，甚至是尽责的选择。

5.3 两片检验场的共同语法

两个案例放在一起，勾出了一个共同的发生语法。标准诊断路径把问题的来源定位在个体内部——暴食者的意志、学生的思维组件——然后对这些内部缺陷提出修补建议。交互认知共生体则在每一次分析中都追问同一个问题：使得这个个体层面的“问题”必然发生的那个共生结构，到底是怎么被切断、被压缩、被耗竭的？

这个追问不提供更“正确”的答案。它提供的是新的干预位置。夜晚进食的干预位置，从意志转移到了白天自主权的再分配。几何证明的干预位置，从教学法转移到了评价系统的命题与评分结构的再设计——以及由此释放出的课堂追问空间。这两个新位置，在标准诊断路径的视野里是完全不可见的。

六、回应质疑：角色扮演、底层机制与“这不过是描述”

6.1 角色扮演假说：那只是一种高级角色扮演吗？

最常被提出的质疑是：即使模型表现出追问行为，这也只是高级角色扮演。底层仍然是一词接着一词的概率续写，模型内部没有发生任何实质变化。

这个质疑其实已经把“实质变化”预设在了模型内部——这恰恰是共生体概念要拆掉的那堵墙。在共生体的框架下，回应这个质疑不需要去证明“内部”有没有变化。回应是：角色扮演这个概念本身，在此处的区分力不足。

什么叫角色扮演？通常指一个系统在行为表层模仿了某种模式，但其内部处理机制与该模式所应具有的本质机制不同。这个区分若要成立，前提是存在一种独立于行为之外的、可被指认的“内部本质机制”。但对于一个大语言模型，它的内部机制就是多层网络的连续运算——不存在一个可以被独立指认为“追问认知本质”的标记，也不存在一个可以被独立指认为“角色扮演”的虚假信号。整个区分的判据，就回落到唯一可公共观测的东西上：追问行为的持久性、稳定性、触发条件的专一性——以及最重要的是，追问活动本身的规范空间是否被交互协议所构成和维护。

交互认知共生体恰好提供了这层判据。当一个交互协议被严格执行时，模型的追问行为持久、稳定、且专一于那些处于流动中的输入类型；当协议断裂时，追问行为随之消散。在角色扮演假说看来，协议断裂后追问行为的消失恰好证明了“它之前只是在扮演”——扮演的脚本被撤走了，它就露出了本相。然而共生体框架的回应是：这恰恰证明了“追问”这种认知活动从来不在模型内部，而在它与人之间的交互协议界面上——协议断裂时认知活动本身就已经瓦解了，不存在一个残存在模型内部的、被隐藏的“本相”。对于共生体框架而言，“扮演”不是一个合法的本体论范畴，因为一切认知活动都在协议构成其规范空间后才得以存在。不存在“假装在追问”，只存在“协议是否被履约”。

6.2 底层机制不变性：如果一切只是下一词预测

第二个质疑沿着类似的路径：一切产出在最底层只是下一词预测，谈何认知共生体、谈何发生结构？这不过是一堆词接着另一堆词。

回应的方式是拆掉这个质疑默认的分析层次。下一词预测是运算的实现层。文字处理软件和图像编辑软件运行在同一套指令集上，没有人说“它们不过是硅晶片的电荷移动”——因为“程序是什么”这个问题的合法分析层次在信息组织的方式上，不在实现层的物理运算上。

同理，交互认知共生体把自己锁定在信息组织的层次上。在这个层次上，认知活动被定义为“信息被组织、追问、生成的特定序列结构”。只要这个序列结构展现出内在的连贯性、稳定的触发条件和独特的产出形态，它就是一个合法的分析对象。它不需要“更深层”的辩护。

然而，一个更强版本的质疑可以援引消除主义来反驳。消除主义说：如果底层机制在理论上可以完整解释所有行为，那么所谓“组织层”的实体就只是描述上的便利，不是本体论上的实在——就像我们可以说“太阳升起来了”来描述日常现象，但在天体力学的层面并不存在太阳“升起”这个事件。应用到此处：如果下一词预测在原理上可以完整解释所有产出，那么交互认知共生体就是一个故事，不是一种实在的结构。

这个质疑的力度比前一版本更强，但它犯了一个微妙的范畴错误：它预设了“唯一的实在层面是底层运算”，而这个预设本身不是被论证出来的，是被默认的。任何复杂系统——从蛋白质折叠到生态系统的种群动态——都展现出一种特性：更高组织层次上的结构模式，在底层机制的语言中即便可以被逐条叙述出来，也往往因为在计算层面不可穷尽、或在实践层面不可辨识，从而丧失了任何描述和解释的效力。你不会用夸克的行为方程来推演一条河流的汛期，即使理论上汛期不过是夸克行为的宏观总和。多重可实现性是另一个论证资源：完全相同的信息组织模式可以在不同的底层机制上实现——同一种追问协议可以在不同的模型架构上被激活并产生相似的认知后果——因此信息组织模式有其独立的本体论地位，它在理论上不能仅仅被还原为底层运算。

这里必须简要回应一个来自心灵哲学的深层挑战。有论点指出，如果高层属性与底层物理属性之间存在因果竞争关系——如果底层物理原因已经足以完全解释一个事件，那么高层属性就没有独立的因果效力，“多重可实现”所承诺的高层独立性就是幻觉。这是一条被持续讨论、未被定谳的哲学路线，本文不准备也不可能在此处终结这场辩论。但即便接受这一挑战的前提，它也不能直接摧毁交互认知共生体的分析合法性。因为本文的核心主张不是“共生体有独立于底层运算的神秘因果力”，而是“共生体层面的概念对于解释、预测和干预该层次的现象是不可消除的”。实践上的不可消除性不必以形而上学的因果独立性为前提。即使你坚持底层物理因果闭合，你仍然需要共生体层面的概念来有效地解释：为什么同一个模型在同一套底层机制上，在不同协议下产出的判别力差异大到不可忽略，以及在哪里下刀可以让差异缩小。如果这些解释和干预必须依赖组织层的概念才能被有效构建和沟通，那么这个概念在该层次上就是实践上必要的——不因为它无法被底层机制实现，而因为它无法被底层语言替换。

6.3 “这不过是另一种描述”

最后一个质疑的犀利处在于：即使模型追问发生过程，它不过是用一种更软的、更叙事化的语言，替代了硬的分析性语言。本质上，这是换了一套描述风格，没有认知分量的提升。

这个质疑认定只有那些可被静态指认、可被形式化定义、可被放进一个稳定分类体系的东西，才有资格被称为“认知分量的提升”。发生学追问因为不生产稳固的定义和清晰的分类，所以只能是修辞，不是认知。

但共生体的回应是：不要把认知分量的判据绑死在标准诊断路径自己的美学标准上。临床医学中有一个极其普遍的困境：两个病人，诊断相同，用药相同，一个好转，一个没有。决定这个差异的，往往不是药物的药理机制，而是那些在标准诊断分类中不出现的东西——病人独特的压力环境、未被说出口的创伤史、治疗关系中的信任质量。所有这些，在标准诊断的视野里都是“主观的”“不可量化的”“只是叙事”。但它们实实在在地决定了治疗效果。

判别力，就是对“谁会对哪种干预有反应、谁不会”的区分能力。这是一个可以在经验上被严格定义的操作概念：在控制其他变量的前提下，两种认知方法对同一组流动处境给出的干预建议，如果随访数据能显示其中一组的建议被验证为有效的比例显著更高，那么这组方法就具有更强的判别力。本文在7.3节将给出检验这一宣称的具体实验设计。

夜晚进食干预和几何证明教学是同样的道理。标准诊断路径给出了诊断和建议，但它在不同的人——或不同的学校——身上可能产生完全不同的结果。发生学追问能够识别出那些未被诊断分类覆盖但影响结果的隐藏变量，并据此刻画差异发生的条件。这不叫“只是换了一套语言”。这叫提供了更精确的判别力。任何东西如果能持续提供比既有框架更强的判别力，它就不是语言游戏——它是更有力的认知工具。

七、结论、实践接口与证伪条件

7.1 核心判断的再凝缩

本文从Skill工程实践中的一个稳定矛盾——注入的程序在不同交互条件下表现悬殊——出发，拆除了一套深度隐藏的先验预设：认知在独立个体的内部完成并维持。在这个预设被撤销之后，逼出了一个此前未被命名的结构位置：使发生学追问得以稳立的，不是模型内部一段被注入的指令或权重，而是人与模型之间被精细设计、需要双方共同履行的交互协议本身。这一结构被命名为交互认知共生体。

共生体的核心洞见是：对于发生学追问这种特定类型的认知活动，交互协议不是因果触发器，而是构成性前提。协议断裂的那一刻，不是模型失去了追问的外部刺激，而是连“何谓好的追问”都无法被判定——该认知活动的规范空间也随之崩解。这一判断将Skill工程的注意力从“模型的内部权重”转移到“交互协议的设计空间”，为认知工程开启了一个新的分析层面。

7.2 一个可复现的实践接口

共生体这个概念如果要落地，需要一个任何研究者都能照做一遍的最小步骤集，同时必须诚实地指出这套步骤所预设的理想条件可能排除了哪些人。

第一步，识别一个需要发生学分析的流动处境——含混的、反复出现的、尚未被提问者清晰概念化的困境。

第二步，强制执行追问式交互协议。协议规定：在给出任何命名或诊断之前，先向提问者追问该处境的三类信息：发生的条件（“这件事通常在什么情境下出现”）、关键的张力（“你在那个情境中最被拉扯的是什么”）、形成的路径（“第一次出现是什么时候，之后是怎样变成现在这样的”）。

第三步，每一轮追问后，将提问者的新补充材料作为下一轮追问的入口，持续追踪，直到一个此前的既有概念无法覆盖的新结构在追问中显形。这个新结构并非深藏于提问者内部的“真实本质”，而是在共生体的追问回路中被双方共同拼接和生成的解释框架。

第四步，用这个新结构反哺提问者，让提问者用自身经验核对其效度——不是核对“对不对”，而是核对“符不符合你自己的经历逻辑”。

第五步，对比直接诊断式交互协议在同样处境下的产出，观察判别力的差异。

这套步骤必须附带一个自我批评。它预设了一个具备充足时间、语言能力和协作意愿的协作者。对那些时间碎片化的照护者、不擅长内省语言表达的群体、或对“被追问”本身感到威胁的人，这套协议可能构成新的排斥。进一步，当这套协议被大规模推广时，它自身也可能被异化——沦为一份机械执行的标准化流程清单，从而失去它赖以为基的那个核心机制：追问不是为了得到答案，而是为了在交互中长出结构。一个发生学的实践接口，必须坦承自身的边界条件，并承认它对某些提问者可能构成新的排斥。本文不声称追问协议是“更好”或“更本真”的认知方式——它只是在测绘式认知面对流动处境系统性失效这一特定历史矛盾下，被逼出来的新工程方向。它自身的发生也受制于特定的技术与制度条件，没有免于历史性的特权。

7.3 证伪条件

一个诚实的证伪条件：设计一项对照实验。实验组的交互协议规定模型必须在给出任何诊断前追问发生条件、张力与路径；对照组仅注入“请保持友善的追问”这类无结构性约束的提示。两组分别处理同一批包含流动处境的问题。记录三项指标：（1）模型追问中是否提取出了提问者此前未明确陈述的关键发生现场细节，而非仅仅在既有概念间进行匹配；（2）提问者事后评估中对产出“触及了我此前没有意识到的东西”的认可度；（3）基于追问产出的干预建议，在后续独立评估中的预测效度。

如果实验组在这三项指标上与对照组无显著差异，则交互认知共生体的概念即被证伪——这说明刻意设计的交互协议并没有产生超越表层友善追问的实质认知后果。

这一概念最严厉的证伪条件是：取消了刻意设计的追问交互协议之后，被注入同样提问提示的模型，依然能在与任意用户的自由交互中，稳定产出具同等判别力的全新认知结构。如果这一点得到证实，就说明交互协议只是伪装成构成条件的因果触发器——一个更精巧的按键而已。

7.4 元层次的自我指涉：本文自身的交互结构

作为一篇发表在学术期刊上的论文，本文从标题、摘要、问题提出、既有立场分析、核心论证展开，到案例检验和反驳回应，遵循了一套高度结构化的论述格式。这套格式本身，正是本文所追问的那种以命名和分类为核心动作的认知方式的学术化身：它要求作者在展开论证之前就划定边界，在追问发生之前就给出定义，在没有读者的追问现场的情况下独自完成全部论证活动。而本文的核心主张——交互协议构成认知活动的构成性前提——在本文自身的写作过程中并未被履行：读者不在场，追问回路不存在，任何一段流动的论证细部都可能在排版定稿的那一刻被冻成一个结论。

这种元层次的不自洽，不是自相矛盾。它意味着本文坦承自己的局限。学术写作在目前的制度条件下，本身就是一种受特定交互协议约束的认知活动——它规定了谁可以发言、以什么格式发言、在什么时间跨度内发言，以及读者的回应如何以引用和批评的形式在数月或数年之后才迟缓抵达。这套协议有它极高的认知效率，但它也系统性地压平了某些细部——尤其是那些只能在追问的回路上逐步显形、无法在单人写作中被预先编程好的认知结构。本文所做的，是在这套给定的协议之内，尽可能诚实地指认自己无法跨越的边界。它不声称能站在这套协议之外写作，但它至少可以在结论的末尾告诉你：你刚刚读到的那段看上去严丝合缝的论证，它的发生条件是一套特定的学术交互协议，而不是某种超越历史的纯粹理性。

注释

[1] 材料说明：本案例中的对话片段为在既有实证研究基础上构造的类型化交互记录，用于展示共生体的微观运作机制。所涉情境与提问者形象均经过匿名化、综合与简化处理，不构成一手临床实证数据。案例中引用的情绪调节与暴食行为关联等判断，基于该领域内公开综述与公认研究的常识性归纳。

[2] 理论定位说明：本节课堂交互片段系类型化构造，用于对照检验共生体概念的判别力，而非对特定教师或学校的实证报告。所引用的表征转换理论、中国数学教育制度史研究均有独立的学术脉络支撑，此处因论题焦点所限不做全面展开。

参考文献

• Clark, A. (2008). Supersizing the Mind: Embodiment, Action, and Cognitive Extension. Oxford University Press.

• Clark, A., & Chalmers, D. (1998). The extended mind. Analysis, 58(1), 7-19.

• De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. Phenomenology and the Cognitive Sciences, 6(4), 485-507.

• Dourish, P. (2001). Where the Action Is: The Foundations of Embodied Interaction. MIT Press.

• Duval, R. (1998). Geometry from a cognitive point of view. In C. Mammana & V. Villani (Eds.), Perspectives on the Teaching of Geometry for the 21st Century (pp. 37-52). Kluwer Academic Publishers.

• Goffman, E. (1974). Frame Analysis: An Essay on the Organization of Experience. Harvard University Press.

• Hutchins, E. (1995). Cognition in the Wild. MIT Press.

• Leont'ev, A. N. (1978). Activity, Consciousness, and Personality. Prentice-Hall.

• Leung, F. K. S. (2001). In search of an East Asian identity in mathematics education. Educational Studies in Mathematics, 47, 35-51.

• Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. Language, 50(4), 696-735.

• Stunkard, A. J., Grace, W. J., & Wolff, H. G. (1955). The night-eating syndrome: A pattern of food intake among certain obese patients. American Journal of Medicine, 19(1), 78-86.

• Suchman, L. A. (1987). Plans and Situated Actions: The Problem of Human-Machine Communication. Cambridge University Press.

• Vygotsky, L. S. (1978). Mind in Society: The Development of Higher Psychological Processes. Harvard University Press.