学 科 通 融 · 之 二

认不出来的东西，争不出结果

为什么有些架吵一百年也吵不完，而且吵的人全都在认真做实验

意识科学 × 代谢医学 × 人工智能评估

作者 王德生 + Claude

栏目 学科通融 · 站外碰撞
篇幅 约1.9万字 · 25页
发表 2026年7月27日

SDE Universes · 德麦国际 Demai International · 本篇由三个分属不同学科、且互相冲突的理论体系碰撞而成；三家来源均为站外公开文献

摘　要

三场架正在打，凶到其中一场有一百多人联名说对方不算科学。一场吵人为什么有意识，一场吵人为什么会胖，一场吵机器是不是真在思考。把三场架的形状画出来会看到同一个东西：双方都同意发生了什么，分歧在于这件事该被算作什么——而这一步在任何测量开始之前就已经做完了，做的时候手上一份数据也没有。本文由此给出一条判断、三个可随身带的问题、十一个行当里的同一件事，以及四条可以让自己失败的判据。

这一篇由哪三个学科的理论体系撞成

第一场里双方事先签字接受被推翻，结果两边核心预测都没活下来，两边都不认输；第二场把因果箭头整个掉了头——一边说吃多了才胖，一边说脂肪被锁进去才导致吃多；第三场同一次停止输出，一边读作能力崩溃，一边读作明智止损，而反驳论文的共同作者是一个模型本身。　三家均为站外的公开文献，链接直达原始出处，可自行核对。

意识科学 · 神经科学与心灵哲学

全局神经工作空间理论对阵整合信息理论 —— Cogitate 预注册对抗实验，《Nature》2025，642(8066):133-142

两百五十六名参与者、三种脑成像、中立第三方主持，双方事先写好什么算自己被推翻；结果两边的核心预测都没活下来，而两边都保留了「被测的操作方式不等于理论本身」这个退路。

https://www.nature.com/articles/s41586-025-08888-1

代谢医学 · 肥胖病因学

能量平衡模型对阵碳水化合物-胰岛素模型 —— 因果方向的反转

同一个观察事实——胖的人吃得多——一边读作原因，一边读作结果。而中立的评估指出，两套模型在它们想要解释的东西上根本不是同一件事。

https://royalsocietypublishing.org/rstb/article/378/1888/20220211/109323/Carbohydrate-insulin-model-does-the-conventional

人工智能 · 推理能力评估

[《思考的幻觉》对阵《思考的幻觉的幻觉》—— 同一批数据的两种读法](https://arxiv.org/html/2507.01231v1)

一边读作推理能力的固有崩溃，一边读作输出格子塞满了；改掉输出方式之后，原来「完全失败」的十五盘汉诺塔被正确解出。

https://arxiv.org/html/2507.01231v1

目　　录

为什么有些架吵一百年也吵不完，而且吵的人全都在认真做实验

有三场架正在打，打得很凶。

不是"学术观点不同"那种客气的凶，是"如果你是对的，那我这辈子的工作全是错的"那种凶。其中一场里，有一百多个研究者联名写信，说对方那套东西根本不算科学。

三场架分属三个毫不相干的领域：一场在研究人为什么会有意识，一场在研究人为什么会变胖，一场在研究机器是不是真的在思考。

这篇文章要说的不是谁对谁错。我没有能力裁决它们中的任何一场，而且我认为**任何人都没有这个能力**——这正是要说的事情。

要说的是：**这三场架为什么吵不完。** 而它们吵不完的原因是同一个，三场架里没有任何一方意识到这一点，因为每一方都以为这是自己领域特有的困难。

一、三场正在打的架

第一场：一个人有没有意识，怎么知道

这场架的两方，是当今研究意识最有影响的两套理论。

一方说：意识就是信息被广播出去。

这一派的图景是这样的：大脑里有大量并行的加工在同时进行，绝大部分你意识不到。当某一路加工赢得了竞争，它的内容就被"广播"到整个大脑——特别是前额叶那一带——被其他所有部分共享。**被广播了，就是意识到了；没被广播，就是没意识到。**

这套说法有个很实用的推论：意识和"能报告出来"是绑在一起的。你能说出你看到了什么，正是因为那个内容已经被广播到了能说话的那部分。

另一方说：意识就是一个系统内部整合的程度，跟广播没关系。

这一派的图景完全不同：意识不在于信息被送到哪儿，而在于这个系统的各部分互相纠结、互相约束到什么程度。整合的程度可以被算出来，是个数。**这个数够大，就有意识；不够大，就没有。**

这套说法有个非常刺激的推论：一个系统可以有意识，但完全不报告——因为报告需要的是别的机器，跟意识本身无关。而且照这个算法推下去，很多不是大脑的东西也可能有一点点意识。

这两套说法不是互补的，是互相排斥的。 一个说意识在大脑后部持续地整合着，一个说意识需要前面那一下广播。它们对同一次实验会给出方向相反的预测。

于是发生了一件在科学史上很少见的事：**两边坐下来，一起设计了一个实验，用来搞死其中一方。**

这不是比喻。二〇二五年六月，一份研究发表在《自然》上——两百五十六名参与者，三种不同的脑成像手段，多个实验室，而且是预注册的：**双方在实验开始之前就白纸黑字写好了，什么样的结果算自己这一方被推翻。**

主持这件事的是一个中立的第三方团队，双方都不掌握数据分析。这在人类做科学的历史上是极其罕见的安排，因为它要求参与者接受一个可能性：自己一辈子的工作扛不住检验。

结果是：两边的核心预测都没活下来。

整合那一派预测的、大脑后部持续的同步，没找到。广播那一派预测的、前额叶应该出现的身份信息，也没找到。有一部分数据偏向这边，有一部分偏向那边，还有一整套数据对两边的预测都不支持。

然后，两边都没有认输。

其中一派的代表人物在实验之后说：现在的挑战是想出办法来解释那个零结果，我们正在做。

请把这句话读两遍。事先讲好了什么算自己输，结果出来了，然后要"想办法解释"那个结果。

这不是耍赖——他有真实的理由。他可以说，实验里用的那个具体的操作方式，不完全等于理论最深处的主张。这个申辩在技术上是站得住的，而且不只他一方能用，另一方也能用，也确实用了。

而在这之后不久，一百多名研究者联名发信，宣称整合那一套理论是伪科学——理由是它的核心主张在实践上无法被推翻。

一场花了几百万、动员了几十个实验室、双方事先签字画押的对抗实验，结束之后，这个领域比开始之前更分裂了。

第二场：人为什么会胖

这场架安静得多，但结构上更极端——因为两边不是对同一个原因估计不同的权重，而是**把因果箭头整个掉了个头。**

主流的说法是能量账本：吃进去的比消耗掉的多，多出来的存成脂肪。所以对策是吃少点、动多点。

这套说法背后是热力学第一定律，没有人能反驳它。而且它符合每一个人的直觉：你看那些胖的人，确实吃得比较多。

另一方说：你把因和果搞反了。

这一派的说法是：不是"吃多了所以胖"，而是"**脂肪被存进去了，所以才吃多**"。

链条是这样的：吃了大量快速消化的碳水，血糖冲高，胰岛素跟着冲高；胰岛素的作用是把能量往脂肪组织里塞，同时不让脂肪往外放；于是餐后没多久，血液里可用的能量反而变少了；大脑感知到这个"缺能量"的信号，就让你饿，让你少动。**你吃多不是因为你嘴馋，是因为你的细胞真的在挨饿——尽管你的脂肪组织里堆着用不上的能量。**

这一派对主流那一套的批评是致命的，而且很难反驳：**你说的那句话是对的，但它不是一个解释。** "吃进去的多于消耗的"这句话，只是把热力学第一定律换了个说法重述了一遍——它没有说任何关于身体里发生了什么的事。一个人处于正能量平衡，这是他在变胖这件事的**定义的一部分**，不是它的原因。

主流那边当然有回应，而且回应也很硬：控制条件下的喂养实验，并没有稳定地显示出这一派预测的那些效应。

这场架的要命之处在于：双方看着同一个事实——胖的人吃得多——一边读作原因，一边读作结果。而这个事实本身，两边都承认。

有中立的评论者试图调停，结论是：这两套模型**在它们想要解释的东西上根本不是同一件事**，所以争谁对谁错意义有限。

这句话说得比它自己意识到的更重。我们后面会回来。

第三场：机器是不是真的在思考

这场架最新，也最热闹，因为它直接关系到很多钱和很多人的判断。

二〇二五年六月，苹果的研究团队发表了一篇标题很不客气的论文：《思考的幻觉》。

他们做的事情很干净：找了几类经典的逻辑难题——汉诺塔、过河问题、积木世界——这些题的难度可以精确地一级一级往上加，而且几乎不可能在训练数据里见过。然后把当前最强的那几个"会推理"的模型放进去。

结果是：**难度加到某个点，模型的表现不是逐渐下降，是直接塌了。** 从做得挺好，到完全做不出来，中间没有过渡。

他们的结论：这些模型不是在推理，是在做模式匹配；看起来像思考的那个东西，在需要长链条、多步骤的时候会露馅。

这篇论文引爆了整个领域。有人拿它当作"大模型不可能通向真正的智能"的铁证。

几天之后，一篇反驳出现了，标题更不客气：《思考的幻觉的幻觉》。

反驳者做的事情是：他去看模型到底为什么"塌"了。

他发现的是这样：汉诺塔十五个盘，完整的解法要输出三万两千多步。而模型有输出长度上限。**模型不是不会解，是纸不够写。** 而苹果的评分脚本把"输出被截断"判成了"答错"。

更有意思的是第二点：在有些情况下，模型是**认出了自己写不完，于是主动停了**——它没有傻乎乎地开始写一个注定写不完的答案。

反驳者接着做了一件事：他改掉了输出方式——不让模型一步步列出全部动作，而是让它给出解法本身。**改完之后，那几个模型都能正确处理十五个盘的汉诺塔**，远远超过苹果报告的"完全失败"的那个难度。

所以，**同一个现象——模型在某个难度上停止输出正确答案——一边读作"推理能力的固有崩溃"，一边读作"输出格子塞满了"**。

而这场架里最耐人寻味的一笔是：那篇反驳论文的共同作者，是一个大语言模型本身。

一个被质疑不会思考的东西，参与撰写了论证自己会思考的论文。 这件事本身应该算作什么证据，没有任何一方说得清。

后来还有人写了《思考的幻觉的幻觉的幻觉》。这个玩笑本身就说明了问题：这一路走下去没有尽头。

二、三场架其实是同一场

三个领域，三场架，三套完全不同的专业词汇。

但如果你把三场架的**形状**画出来，会发现是同一个形状。

每一场里，双方都同意发生了什么。

意识那场：两边都同意实验数据是什么。后部有什么信号、前额叶有什么信号，数字就在那儿，没人质疑。 肥胖那场：两边都同意胖的人吃得多。这个事实没有争议。 机器那场：两边都同意模型在那个难度上停止输出正确答案。这件事双方都看到了。

每一场里，双方分歧的是：这件事该被算作什么。

意识那场：后部持续的整合，算不算"意识在这里"？还是必须被前面广播出去才算？ 肥胖那场：吃得多，算原因还是算结果？ 机器那场：停止输出，算"不会"还是算"没地方写"？

而最要命的是第三层：每一场里，"该被算作什么"这个问题，都不能靠做更多实验来回答。

因为任何一个新实验，都得先回答同一个问题才能开始设计。你要测"模型会不会推理"，你得先定下什么算会推理；你要测"意识在不在这一刻",你得先定下什么算意识在场；你要查"什么导致了肥胖",你得先定下什么算原因。

这一步在任何测量开始之前就已经做完了。而做这一步的时候，你手上一份数据也没有。

得先替三家各自说句公道话

上面把三场架写得像是各方在耍赖。不是的。三方各自都有硬的自辩，不先摆出来，后面撞出的东西就不算数——撞倒稻草人不叫碰撞。

意识那边会说：我们做了科学史上罕见的一件事——事先签字，接受自己可能被推翻。有几个领域敢这么干？而且申辩"被测的操作方式不等于理论本身"，这在科学里是完全正当的：牛顿力学在水星轨道上算不准，当时的人没有立刻废掉牛顿，而是去找有没有别的行星。后来这个思路在海王星上是对的,在水星上是错的。**你不能因为一次不符就要求人放弃理论，科学从来不是这么运转的**。

这条自辩非常有力。它同时也是一把双刃刀：正因为它这么有力，它才可以被永远用下去。

肥胖那边会说：我们不是在耍赖，我们指出的是对方那套说法在逻辑上的一个真实缺陷。"吃多于耗"不是一个可以被检验的因果主张，它是一个恒等式。你不能用一个恒等式来解释任何事情。而且我们给出了具体的、可以被推翻的生理链条：血糖、胰岛素、脂肪组织、能量可用性——每一环都能测。

这条自辩也很有力。它反过来质疑对方：你说我的预测在喂养实验里没被证实，可你自己那套东西压根就没有可以被证实或推翻的预测——它只是复述了一条物理定律。

机器那边会说：我们不是在给模型找借口。我们改掉了一个具体的实验设置，然后模型的表现变了。**这是一个真实的、可以被别人重复的实验结果**，不是嘴上的辩解。你原来说的"崩溃"，在改掉输出限制之后不出现了——那么原来那个崩溃是关于什么的?

这条自辩最有力，因为它兑现了。它不是说"你测的不是我"，它是说"你测的不是我，而且我做了个实验证明这一点"。

三条自辩摆出来之后，问题反而更尖锐了：

三方都有真理由，三方的申辩在各自的语境里都是正当的。**它们之所以撞不出结果，不是因为哪一方不诚实，而是因为它们在争的那个东西，本身就不是证据能够裁决的类型**。

这一点是下一节的入口。

三、把那条线拽出来：先认后测

要研究一样东西，你得先能认出它什么时候在场。

这听起来是句废话，但它是全篇的地基，所以要说慢一点。

假设你要研究温度。你有温度计。温度计不是一个关于温度的理论——它是一个所有各方都接受的、能告诉你此刻温度是多少的装置。**因为有它，关于温度的争论可以被数据裁决**：你说这个反应放热，我说不放，我们把温度计插进去，看读数。

现在假设你要研究意识。**你有什么?**

你能拿到的最直接的东西是报告——问他"你看见了吗"，他说"看见了"。但报告能不能算作意识在场的标志，**恰恰就是这场架的内容之一**。一方说报告是意识的构成部分，另一方说一个系统可以有意识而完全不报告。

所以你没有温度计。你手上唯一那个像温度计的东西，它算不算数，正是争论的对象。

肥胖那场也是。你要研究"什么导致了变胖"，你得先能认出什么算"原因"。而一方说能量平衡是原因，另一方说能量平衡是这件事的定义的一部分、根本不能当原因用。**"什么算原因"这个问题，没有一个独立于两套理论的答案。**

机器那场最清楚。你要研究"模型会不会推理"，你得先能认出什么算推理在场。苹果那边的标志是"输出了完整正确的解"，反驳那边的标志是"内部策略正确"。**这两个标志会给出不同的答案，而没有第三个标志可以裁决它们。**

于是有了这条判断：

> **一样东西如果没有一个独立于各家理论的在场标志，那么关于它的争论，在原理上就不能被证据裁决。而各方仍然会不断地做实验、产出数据——因为产出数据是这个领域里唯一能做的事。**

我把这件事叫作**先认后测**：**认出它在场，这一步在测量之前；而当"怎么认"本身是争论的内容时，测量就不再是裁判，只是各方各自的复述。**

注意这条判断的形状。 它不是说这些领域的研究者不认真——恰恰相反，意识那场架里的人认真到愿意事先签字接受自己被推翻，这是很少有人做得到的。它也不是说这些数据没价值——那些数据是真的，昂贵的，而且会长期有用。

它只说一件事：**那些数据不会结束那场架。而各方在做数据的时候，都以为自己在朝结束那场架前进。**

四、为什么"什么算作它在场"这一步，不能被证据决定

上一节的判断如果只是一个观察，它是软的。得说清楚它凭什么。

这里有一个绕不过去的圈，说穿了很简单：

要用证据来决定"什么算作它在场"，你得先有一批确定它在场的案例，去看那些案例有什么共同点。而要有那批案例，你得先知道什么算作它在场。

圈就在这儿。

拿意识举例。你想用实验来确定"报告算不算意识在场的标志"。怎么做？你需要一批"确实有意识"的时刻和一批"确实没意识"的时刻，然后看报告在两组里有没有区别。

但你怎么划出那两组？ 唯一的办法是用某个标志去划——而你正在测的就是那个标志。

这个圈在有温度计的领域里不存在，因为温度计是先于理论争论就已经被所有人接受的。**注意"先于"这两个字：不是逻辑上更根本，是历史上更早、且在所有各方分歧之前就已经共同接受了。** 你和我可以在热力学上分歧很大，但我们都读同一支温度计。

而意识、推理、肥胖的原因这些东西，**在所有各方分歧之前，没有一个共同接受的标志被建立起来**。分歧和标志是一起长出来的，甚至是分歧先长出来的。

一个特别清楚的例子

机器那场架把这件事暴露得最彻底，因为它最新、最没有历史包袱。

苹果那边和反驳那边看着同一件事：模型在某个难度上停了。

苹果说：停了就是不会。 反驳说：停了是因为它知道自己写不完，这恰恰是聪明。

请注意，这两种读法都需要先有一个关于"什么算会推理"的判据，然后才能读出结论。 而这个判据不在数据里——数据只显示"停了"。

更狠的是：反驳方做了一件真正的实验工作，改掉输出限制，模型解出了十五盘。**这看起来像是用证据裁决了。** 但仔细看，它裁决的是一个更小的问题："那次崩溃是不是输出限制造成的"——这个小问题有独立的标志（改掉限制，看结果变不变），所以它可以被裁决，而且确实被裁决了。

而那个大问题——模型是不是在推理——没有被裁决，一步也没有。 因为"解出了十五盘"算不算推理，仍然取决于你用哪个判据。一个坚持"推理必须是可泛化的"的人，可以说这只是更长的模式匹配。

这一节最要紧的一笔在这里：一场关于"什么算作它在场"的争论里，仍然可以有大量真正被裁决的小争论。 而这些小争论的解决，会给所有人一种正在进步的感觉。

数据在累积，方法在改进，一个个技术细节被澄清——而那个大问题，一步没动。

五、有标志和没标志，差别有多大

为了让这条判断不显得空，得看看有标志的领域长什么样。

有标志的例子：胃溃疡。

上世纪八十年代之前，主流认为胃溃疡是压力和胃酸造成的。有两位澳大利亚研究者提出是细菌造成的，被嘲笑了很多年。

这场架结束得干净利落：**因为"这个人的胃里有没有那种细菌"这个问题，有一个独立于两套理论的答案。** 你把胃黏膜取出来，培养，看长不长。这个动作跟你信哪套理论完全无关。

于是当证据足够多时，主流让步了，两位研究者拿了诺贝尔奖。**整个过程二十年,不算快,但它结束了。**

没有标志的例子：意识。

这场架已经吵了三十多年，方法越来越精密，数据越来越多，钱越来越多，而分歧一点没缩小——它变得更精细、更技术化，但没有变小。

这两个案例的差别不在于哪个问题更难，而在于有没有一支各方都读的温度计。

胃溃疡那场架里，两边对细菌的意义分歧很大，但对"有没有细菌"这个事实没有分歧。 意识那场架里，两边对数据没有分歧，但对"这个数据算不算意识在场"分歧到底。

分歧的位置不一样，结局就完全不一样。

再加一个更微妙的例子：**温度本身也不是一开始就有温度计的。**

早期关于冷热的争论，用的是人的感觉，而人的感觉是不可靠、不可比较的——同一盆水，一只手觉得热一只手觉得冷。在温度计被造出来并被广泛接受之前，关于冷热的争论跟今天关于意识的争论一样，吵不出结果。

温度计做的那件事，不是让人测得更准，是让所有各方第一次拥有了一个大家都读的、独立于各自看法的标志。

这一点很要紧，因为它意味着：**没有标志不一定是永久的。** 一个领域可以从没有标志变成有标志。而这正是第九节要讲的东西。

六、为什么这三场架偏偏在这两年同时炸开

三场架都不新。意识那场吵了三十多年，肥胖那场吵得更久，机器那场看起来最新——但"这东西是真懂还是装懂"这个问题，从上世纪五十年代就有人问了。

可它们**同时炸开**是这两年的事。这不是巧合，有三件事凑到了一起。

第一件：方法变严了。

过去二十年，各个领域都在推同一套东西：预注册、公开数据、独立复现、事先讲好什么算失败。这套东西是为了对付一个真实的毛病——大量结果重复不出来。

这套东西起作用了。但它有一个没被预料到的后果：**它把"这场架能不能被裁决"这个问题，第一次逼到了台面上。**

从前一场架可以永远吵下去，因为双方各做各的实验，各自都能得到支持自己的结果，谁也不必正面撞上谁。**而预注册的对抗实验取消了这个避风港。** 双方必须坐在同一张桌子上，事先写下什么算自己输，然后一起看结果。

意识那场架之所以在二〇二五年炸开，正是因为有人真的把这件事做了。**做完之后大家才发现：即使把所有程序上的漏洞都堵死，这场架仍然结束不了。** 而这个发现，比任何一方赢了都更重要。

第二件：钱和决策压过来了。

三场架都不再是纯学术的。

肥胖那场直接关系到几亿人的饮食建议和一整个食品工业。 意识那场关系到一系列很实际的判断——植物人状态的病人有没有感受、麻醉深度怎么算够。 机器那场关系到最多的钱和最急的决定：该不该管制、该不该信任、该不该给权限。

当一场架被要求马上给出答案时，它的所有缝都会被拉大。 因为决策者要的是"研究表明"，而各方能给的只有"依据我们的判据"。这个落差从前可以被学术语言的谨慎盖住，现在盖不住了。

第三件，也是最要紧的：出现了一个不肯配合的对象。

前两场架的对象——意识、代谢——都不会说话，不会反驳，不会在你测它的时候改变策略。

而第三场架的对象会。

大模型是人类第一次面对这样一个东西：它在几乎所有可观察的方面都像是在做某件事，而你没有任何独立的办法确认它是不是真在做。 而且它还会针对你的测试调整——你出一套题，它下一版就在这套题上表现更好，而你无法确定这是因为它更会推理了，还是因为它更会应付这套题了。

这个对象把"先认后测"这个问题变成了一个**每天都要处理的实务问题**，而不再是哲学系里的思辨。

于是三件事叠起来：**方法变严，逼出了裁决的极限；利害变大，容不下悬而不决；而新出现的对象把这个问题从思辨变成了日常。**

这就是为什么这三场架同时炸开，也是为什么现在把它们放在一起看，能看到单独看任何一场时看不到的东西。

七、那个万能退路：这个测试测的不是我

三场架里，每一方在证据不利时都用了同一招。

这一招的形状是：**"你测的那个东西，不是我说的那个东西。"**

意识那边：预注册实验里的操作方式，不完全等于理论最深处的主张。 肥胖那边：喂养实验的条件，没有真正制造出我们说的那种代谢状态。 机器那边：你的输出限制，不是模型的推理限制。

这一招之所以要命，是因为它常常是对的。

反驳方那次是对的——他改了设置，模型真的解出来了。 牛顿力学在水星轨道上算不准，当时坚持"这不是牛顿的问题"的人，在海王星那件事上是对的。 一个新药在临床试验里失败，"病人选得不对"这个申辩有时候是真的。

所以不能简单地说这一招是耍赖。 问题是：怎么区分正当的申辩和万能的托词？

一条可能的分界线

我想到的分界线只有一条，而且它不完美：

> **一次"你测的不是我"的申辩是否正当，不取决于它听起来多合理，而取决于它在被提出的同时，有没有交出下一个可以让自己输的地方。**

拿三场架来对照：

机器那边是正当的，而且是三场里最干净的。反驳方不只是说"你测错了"，他做了一件事：改掉输出限制，然后模型解出了十五盘。**他的申辩带来了一个新的、当场就被检验了的预测。** 而且这个预测如果失败——改了限制模型还是解不出来——他就输了。他把自己重新架到了火上。

肥胖那边处在中间。 他们的申辩带来了新的可检验预测：同等热量下，不同血糖负荷的饮食应该导致不同的能量消耗。这些预测被检验过，结果混杂——有的支持，有的不支持。**这至少是在火上的，只是火不够旺。**

意识那边目前最弱。 "我们正在想办法解释那个零结果"——这句话没有交出任何新的可输的地方。它是一张欠条。欠条可以被兑现，兑现了就正当；但在兑现之前，它跟托词的形状是一样的。

这条分界线有个好处：它不需要你先判断谁对谁错。 你不需要懂意识理论，也能看出"我们正在想办法"和"我改了设置然后重做了实验"是两种不同的东西。

它也有个明显的弱点，必须说出来：交出新的可输之处，可以是廉价的。一个人可以每次被推翻就交出一个新的、更难被检验的预测，永远走在证据前面半步。这一招在历史上被用得非常成功过。

所以这条线不是一道墙，只是一个可以看的地方。**但有一个地方可以看，比没有强很多。**

一个很少被说的观察

三场架里，用这一招用得最理直气壮的，往往不是弱的一方，而是**理论最丰富的那一方**。

理由不难懂：一套理论如果数学上足够丰富、能推出很多不同的预测，那么任何一个具体的预测被推翻时，它都可以说"那不是我最核心的那条"。**理论越丰富，可退的余地越大。**

这就有了一个很不舒服的推论：**一套理论的丰富程度，跟它的可被推翻程度，是反着走的。** 而我们通常把理论丰富当作优点。

这一点对那封"伪科学"联署信提供了一个更精确的说法。那封信说对方的理论在实践上无法被推翻。而更准确的描述可能是：**它足够丰富，以至于任何单次推翻都可以被吸收。** 这不是造假，是结构。

八、判断一场架能不能被裁决：三个问题

前面都是诊断。这一节给一把可以随身带的尺子。

碰到任何一场长期不决的争论——学术的、公司里的、家里的——问三个问题。

第一问：双方对"发生了什么"有没有分歧?

如果有——那这是个事实问题，去查就行了，多半能结束。谁的销量更高、这台机器有没有坏、他那天有没有去过那儿。

如果没有，双方都同意发生了什么，那就往下走。**注意：绝大多数长期不决的争论，都在这一步就该停止被当作事实问题来吵了。**

第二问：双方对"这件事该被算作什么"有没有分歧?

如果有——比如同一个动作，一边算作关心一边算作控制；同一份财报，一边算作稳健一边算作停滞；同一次沉默，一边算作尊重一边算作冷暴力——那么继续摆事实是没用的。**你们不缺事实，你们缺的是一个共同的判据。**

第三问：这个判据，有没有一个独立于双方立场的来源?

这是决定性的一问。

有 ——比如"合同上是怎么写的""行业标准是多少""第三方检测的结果是什么"——那么这场架可以结束，只要双方都愿意去看那个来源。

没有 ——比如"他到底爱不爱我""这家公司有没有创新精神""这个模型是不是真的在推理"——那么**再多的事实都不会结束它**，而且往往越吵越细、越吵越技术化，给人一种正在接近答案的错觉。

三问走完，你会得到三种处境中的一种，而每一种该做的事完全不同。

第一种：去查。 第二种：先谈判据，别谈事实。 第三种：接受它不会被裁决,然后做第十节说的三件事之一。

最常见的错误是把第三种当成第一种来处理——不停地摆事实，摆了十年，双方对事实的了解都增加了一百倍，而分歧一步没动。

九、十一个地方的同一件事

下面把这条判断放到十个具体的地方去试。

必须先说一句：**这条判断不是说"测不了的东西就别管了"**。　恰恰相反，很多最要紧的东西都在这一类里。它只说：别指望用摆事实的方式结束关于它们的争论，因为那个方式在这里不工作。

一、精神科的诊断

一个人算不算抑郁症、算不算注意力缺陷、算不算人格障碍——这些没有独立于诊断手册的标志。没有血检，没有影像，判据就是那本手册，而手册本身是专家投票改出来的。

于是关于"这个孩子是不是真的有注意力缺陷"的争论，可以无限进行下去，而且双方都能拿出真实的观察。这不是精神科不努力——他们比大多数领域更努力地想找标志。**这是这一类对象的性质**。

能做的是：别把诊断当成发现了一个事实，把它当成一个当前最有用的分类工具——有用是它的全部辩护，也是它唯一需要的辩护。

二、疼痛

疼痛是这一类里最残酷的例子。一个人说他很痛，你没有任何独立的标志去核实。

这不是理论问题，是每天在医院里发生的事：**在没有标志的情况下，判断权会落到不痛的那个人手上**。于是有大量真实的疼痛被判定为夸大，尤其是那些说话不够有说服力的人的疼痛。

能做的是：既然没有标志，就把默认值定在信的那一边——这不是科学判断，是在没有裁决可能时的一个道德选择,而且必须承认它是道德选择,不能伪装成科学。

三、公司里的"有没有创新精神"

这个东西没有标志。专利数不是它，研发投入不是它，员工问卷更不是它。

于是关于"我们公司有没有创新精神"的讨论会永远进行，而且会随着公司规模变大而变得越来越激烈、越来越无解。

能做的是：别开那个会。　改成问一个有标志的问题——去年有几个项目是没人要求就做起来的？这个问题有答案，而且答案会带出真问题。

四、教育里的"有没有真懂"

考试成绩不是它，能复述不是它，能应用也只是接近它。

这一条有个特别的地方：**老师大多知道谁真懂了。** 但这个"知道"没法被外部核验,所以在任何一个需要交代的场合，它都不算数。

能做的是：承认它不能被核验，然后决定要不要在某些地方给它留位置——比如给一部分评分权给那个知道的人，同时接受这会带来偏私的风险。这是一个交换，不是一个改进。

五、经济学里的"生产率"

看起来这个有标志——产出除以投入嘛。但一旦进入服务业、软件、研究这些领域，"产出"是什么就成了争论对象。

于是关于"技术有没有提高生产率"这场架，已经吵了三十多年，数据越来越多，分歧没有缩小。

能做的是：注意那些声称已经测出来的人，用的是哪个产出定义——分歧几乎全在那里，不在数据里。

六、法律里的"故意"

一个人做那件事的时候，心里是怎么想的——这个东西没有任何独立标志。

法律用一套代理判据来处理：他有没有预备、有没有掩饰、有没有从中获利。这些是有标志的。但它们不是"故意"，它们是"故意"的可观测替身。

法律其实很早就承认了这一点，所以它不说"我们查明了他的想法"，它说"依据在案证据，可以认定"。**这个措辞上的谨慎，是这一类问题里做得最好的处理之一**。别的领域可以学。

七、动物有没有感受

跟意识那场架是同一个问题，只是换了对象。而且它有直接的后果——决定了几十亿个生命被怎么对待。

这一条最能说明"没有标志"不等于"可以不管"。 恰恰因为不能裁决，所以只能做一个明确的选择,而且要承认那是选择。

八、模型有没有在思考

前面讲过了。这里只补一句最要紧的：**这个问题正在被越来越多地拿来做重大决定**——该不该管制、该不该信任、该不该给权限。

而它没有标志。所以那些决定实际上是基于判据的选择，不是基于证据的发现。**把它说成"研究表明"，是把一个选择伪装成了一个发现。**

九、一个人有没有变好

心理治疗、戒瘾、教育矫正、婚姻修复——全都撞在这一条上。

量表分数不是它，行为记录不是它，本人自述也不是它。

而这一条比前面所有条都更常见，因为几乎每个人一生都要在某个时刻判断一次：他真的变了，还是只是这段时间表现好？

能做的是：把判据事先说出来，说给对方听。 "我怎么才算认为你变了"这句话如果事先说清楚，后面的争吵会少一大半——不是因为它更准确，是因为双方至少在读同一支温度计，哪怕那支温度计是自己造的。

十、这篇文章有没有价值

放在最后，因为它是自指的。

一篇文章有没有价值,没有独立标志。阅读量不是它，转发不是它，评分更不是它。

而这篇文章会被打一个分数。那个分数会成为很多人判断它的依据——不是因为分数准，是因为**没有别的东西可以拿来判断，而人总得判断**。

这正是本文所说的那件事：当没有标志时，任何一个现成的数字都会被当成标志用。 不是因为有人相信它准，是因为空位必须被填上。

十一、婚姻里的那句「你到底有没有在听」

放在前十条之后，因为它是所有人都亲身经历过的那一条。

「你有没有在听我说话」——这个问题没有独立标志。他能复述你说的每一个字，这不算；他看着你的眼睛，这也不算；他说「我在听」，更不算。

而这场架的形状跟意识那场一模一样：**双方对事实没有分歧**（他确实坐在那儿，确实没打断，确实能复述），分歧在于**这算不算「在听」**。

于是就有了那种最消耗人的争吵：一方不断举证——我复述给你听、我记得你说的每一句；另一方不断说不是这个意思，但说不出是哪个意思。

因为她要的那个东西没有标志，所以他的每一次举证都打不中。 而他的举证越充分，她越确信他没在听——因为一个真在听的人，不需要举证。

这一条给出的处置跟第九条是同一个：**把判据事先说出来。** 不是「你要多听我说话」，而是「我什么时候会觉得你在听」——比如你会问一句我没说到的地方，比如第二天你还记得。

这个判据未必准，但它把一场不能被裁决的架，换成了一场可以被裁决的小架。**而可以被裁决的小架，是能吵完的。**

十、没有标志的领域，还能做什么

诊断到这里就完了。这一节讲能做什么，因为只诊断不给出路是不负责任的。

三条出路，都是真的有人走过的，而且各有各的代价。

出路一：不找它本身，找它绕不过去的后果

这条最实用。

疼痛没有标志，但"这个人愿意为止痛付出什么代价"是有标志的——他愿不愿意接受有副作用的治疗、愿不愿意改变生活安排。这不是疼痛，但它是疼痛绕不过去的后果。

意识有没有标志？没有。但"这个系统在被剥夺某种信息之后行为会不会改变"是有标志的。

这条路的代价是：**你换掉了你原来要研究的东西**。　你研究的不再是疼痛，是疼痛的痕迹。这个交换有时候值,有时候不值,得逐案判断,不能一概而论。

而且必须承认它有一个失败模式:一旦那个后果被用作标志用久了,人们就会忘记它只是替身,开始把它当成本尊——这正是我在另一篇文章里写过的那件事。

出路二：不问"是不是"，问"在什么条件下两种说法会分开"

这条最聪明，也最少被用。

不要问"模型是不是在推理"，问："如果它是在推理，在什么情况下它的表现会跟纯粹的模式匹配**明显不同**？"

这个问法的好处是：它把一个不能被裁决的问题，换成了一串可以被裁决的小问题。你不再需要判据，你只需要找到两种说法预测不同的那些点，然后去看。

机器那场架里，反驳方走的正是这条路，而且走通了一小段——他找到了一个两种说法预测不同的点（改掉输出限制），去看了，得到了答案。那个答案没有解决大问题,但它真的解决了一个小问题,而且是永久解决的。

这条路的代价是：**你可能永远回答不了那个大问题**。　你会积累一大堆被解决的小问题,而大问题始终悬着。有些人受不了这个。

出路三：承认不能裁决，改成比谁做得出东西

这条最粗暴，但在工程领域里最有效。

不争"哪套理论对"，争"谁能用它造出更管用的东西"。

肥胖那场架里，这条路已经在被走了：不再争哪个模型对，而是看不同的饮食方案在具体人群里效果如何。有中立的评论者明确建议这么干——别再争模型了，去在特定患者群体里找具体的作用方式和可行的干预。

这条路的代价最大：**它放弃了理解。** 一个方案管用，不等于你知道为什么管用；而不知道为什么，你就没法预测它在新情况下会不会失效。这是个真代价,不是小事。

三条路都不是解决方案，都是绕行。 而它们共同的前提是同一个动作：**先承认这场架不会被证据结束。** 在承认之前，任何一条路都走不通,因为人会一直期待下一个实验解决问题——而下一个实验永远不会。

温度计是怎么被造出来的

第五节说过，温度也曾经没有标志。这一节要把那段历史讲细一点，因为它是三条出路之外的第四种可能——**也是唯一一种真正解决问题的可能**，虽然它不能被计划。

温度计的故事有一个常被忽略的地方：**它一开始并不比人的手更可靠。**

早期的测温装置，不同的人做出来读数不一样，同一支在不同气压下读数也不一样。那时候的争论跟今天关于意识的争论很像：你说这个装置测的是热，我说它测的是空气膨胀，我们各有各的道理。

它是怎么变成所有人都读的那一支的?

不是因为有人证明了它测的就是温度——**这件事没法证明，因为要证明它，你得先有一个独立的温度标准，而那正是你要造的东西。**

它变成标准，靠的是三件完全不同的事凑到一起:

其一，找到了不依赖于个人的固定点。 水的冰点、沸点——这两个点在任何人手上都出现在同一个位置。这不是理论论证，是一个可以被任何人重复的事实。有了两个固定点，中间就可以刻度，而刻度一旦刻上，不同的装置就可以互相校准。

其二，它被用于别的事，并且在那些事上管用。 温度计的可信，很大程度上来自于它在化学反应、蒸汽机、医学诊断上都给出了有用的读数。**一个标志的可信度，不是在关于它自己的争论里赢来的，是在别的地方兑现出来的。**

其三，它先被普遍使用，然后才被理论辩护。 热力学是十九世纪的事，而温度计在此之前已经用了两百年。**人们先是把它当工具用顺手了，事后才有人说清楚它测的到底是什么。**

这三条对今天那三场架意味着什么?

意味着一个不太好听、但可能是对的结论：**一个领域拿到独立标志，往往不是靠这个领域内部的争论解决的，而是靠这个东西在别的地方被用出来的。**

意识那场架，如果有一天会结束，很可能不是因为哪个实验裁决了它，而是因为某个从这套理论里长出来的东西——某种诊断装置、某种麻醉监测——在临床上被大量使用并且确实管用，于是它背后那套判据就跟着被接受了。

而这条路有一个残酷的性质：它不能被计划，也不能被加速。 你没法开一个会决定"我们要找到独立标志"。固定点是被撞见的，用处是在别处长出来的，理论辩护是最后才来的。

这就是为什么第十节那三条出路都只是绕行。 真正的解决只有这一条，而这一条不在任何人的控制之内。

能做的只有一件事：**在标志出现之前，别假装它已经出现了。**

十一、这条判断最容易被拿去干坏事的地方

必须专门写一节，因为这条判断有一个非常方便的滥用形式。

滥用形式是：**"这个问题没有客观标准，所以你的看法不比我的更有根据。"**

否认气候变化的人会说：气候模型没有独立标志。 拒绝接受诊断的人会说：精神疾病的判据是投票投出来的。 不想改的人会说：这事没有客观衡量标准，各说各话罢了。

这些话全都在借用本文的判断，而它们全都是错的。

判别方法只有一条,而且很清楚：

本文说的是"关于什么算作它在场，没有独立标志"。本文没有说"关于事实，没有独立标志"。

气候变暖是不是在发生——这是个事实问题，有标志（温度记录、冰芯、海平面），已经被裁决了。有争议的是别的（多快、多少人为、该怎么办），而且争议的性质跟"有没有在变暖"完全不同。

一个人有没有服药、有没有到岗、有没有转账——全是事实问题，有标志。

第八节那三个问题就是用来防这个的：先问双方对"发生了什么"有没有分歧。 绝大多数被说成"没有客观标准"的争论，其实在第一问就该结束了——它们是有事实分歧的，只是有一方不想去查。

还有第二种滥用，更隐蔽：**用它来给自己的理论免疫。**

"你的实验测的不是我说的那个东西"——这句话本文承认它常常是正当的。但第七节给了那条线：**说这句话的时候，有没有同时交出下一个可以让自己输的地方？** 没交的，就是欠条。欠条可以打，但不能当钱花。

十二、什么情况下这篇文章是错的

一个说法如果怎么都推不翻，那它也就永远证明不了自己对。何况这篇文章讲的正是"有些说法推不翻"——它更不该给自己留后门。

以下四种情况，任何一种成立，本文的核心判断就该被放弃。

第一种：三场架里有任何一场，被一个新实验干净地结束了。

这是最直接的一条，而且是可以等的。

如果在未来若干年内，意识那场架因为某个新实验而收场——一方公开承认自己错了，领域内形成新的共识——那么本文关于"没有独立标志则不能被证据裁决"的判断就是错的，或者至少是过强的。

要注意排除一种假结束：一方的支持者陆续退休或转行，导致争论消失。那不是被裁决，那是被时间耗掉的。真结束的标志是：**有人公开说"我原来的看法错了，因为这个证据"**。

第二种：找到一场没有独立标志、却被证据结束了的历史争论。

本文说没有标志就不能被裁决。那么只要找到一个反例就够：某场争论，各方对"什么算作那个东西在场"始终没有共同的独立判据，而它仍然被数据干净地终结了。

我自己想过这个问题,还没找到。但我找过的范围很有限,这一条是留给读者的。

第三种，也是我最担心的：也许起作用的不是"没有标志"，而是"有人不想它结束"。

有一个解释比我的解释更朴素，可能也更对：**这些架吵不完，不是因为原理上不能裁决,而是因为吵下去对某些人有好处**。

经费、职位、名声、整个研究团队的存续——这些东西都绑在争论持续上。一场结束了的争论不再有经费。

按这个解释，那些"你测的不是我"的申辩不是结构性的必然，只是利益驱动的抵抗；而如果没有这些利益，同样的证据早就足以让一方让步了。

这两个解释在大多数情况下预测一样。 要分开它们，得找一个**没有任何利益绑定的争论**：没有人靠它拿经费、没有人的职位系于它、双方都是纯粹出于好奇。

如果在这种条件下，一场缺乏独立标志的争论**仍然**吵不出结果——那说明是结构问题，本文站得住。 如果这种争论都能干净地收场——那说明本文找错了对象，真正起作用的是利益，我该收回这个说法重写一篇。

我倾向于前者，理由是第四节那个圈：**你需要先有一批确定在场的案例，才能确定标志；而要有那批案例，你得先有标志。** 这个圈跟有没有人拿经费无关。但我倾向于哪一种不构成证据。这是本文最软的一处,我把它指出来,而不是绕过去。

第四种：三个来源里有一个被推翻。

本文撞的三场架，每一场自己都还在打。如果其中任何一场的一方被干净地推翻——比如证明模型确实只是模式匹配、或者确实在推理——那么本文的三角就塌了一个角。

塌一个角之后，剩下两家会变成互相印证而不是互相打架，而本文这条判断恰恰是从打架处捡出来的。到那时它需要被重新撞一次。

十三、文中提到的说法，出处在这里

本文不是学术论文，正文里没有加注，但把话说到了别人头上，就该把出处交代清楚。

意识那场架。两套理论分别是全局神经工作空间理论（Stanislas Dehaene 一系）与整合信息理论（Giulio Tononi 一系）。本文用到的那场对抗实验，是 Cogitate Consortium 二〇二五年六月五日发表于《Nature》第六四二卷第八〇六六期第一三三至一四二页的《对全局神经工作空间与整合信息两种意识理论的对抗性检验》。这项研究由中立的第三方团队主持，两百五十六名参与者，同时使用功能磁共振、脑磁图与颅内脑电，且实验设计、分歧预测与结果解释全部事先预注册。Christof Koch 关于"想办法解释零结果"的表态，见二〇二五年四月三十日 GeekWire 的报道。那封由一百余名研究者联署、称整合信息理论为伪科学的公开信，发生在此前后，是这场争论中最激烈的一笔。

肥胖那场架。碳水化合物-胰岛素模型的系统表述见 David Ludwig 等人二〇二一年发表于《美国临床营养学杂志》的《碳水化合物-胰岛素模型：肥胖流行的一个生理学视角》；本文引用的"因果方向反转"这一提法出自该文，其在《英国皇家学会哲学汇刊 B》二〇二三年那篇文章里被表述得最直接——过食并不驱动脂肪增加，而是储存过量脂肪的过程驱动了过食。中立的评估见 Cell Metabolism 二〇二三年那篇《调停"大辩论"》，本文引用的"两套模型在它们想要解释的东西上根本不同"以及"应转向在特定患者群体中确立具体作用方式"，均出自该文。

机器那场架。苹果机器学习团队二〇二五年六月的《思考的幻觉：透过问题复杂度的视角理解推理模型的能力与局限》，以及 Alex Lawsen 与 Claude Opus 4 合著的反驳《思考的幻觉的幻觉》。本文引用的几笔——十五盘汉诺塔需输出三万两千余步、评分脚本把输出截断判为答错、模型在部分情况下识别出无法在限制内完成而主动停止、以及改变输出方式后模型能正确处理十五盘——均出自该反驳及其后续报道。另有第三方在 arXiv 上做了复现与澄清工作。

胃溃疡那个对照例子，是 Barry Marshall 与 Robin Warren 关于幽门螺杆菌的工作，二〇〇五年获诺贝尔生理学或医学奖。

需要说明的是，本文与这三场架的任何一方都没有站队，也没有能力站队。**本文关心的那件事，恰好落在三场架的边界之外**——三场架里的人各自都在诊断自己领域的困难，而没有一方把它当作三个领域共有的东西来处理。如果读完之后你觉得本文说的其实就是其中某一场，那么第二节没有写成功，责任在我。

十四、这一篇是怎么撞出来的

按本栏的规矩，把工序和作废数一并交出来。

三个来源，分属三个互不相干的学科，且各自内部正在发生激烈的、未决的冲突：

- **意识科学**：全局工作空间对阵整合信息。
- **代谢医学**：能量平衡模型对阵碳水化合物-胰岛素模型。
- **人工智能评估**：推理崩溃论对阵输出限制论。

选这三个的理由是它们**冲突的烈度**：不是侧重不同，是"你要是对的我就全错"。其中一场还附带了一封百余人联署的"伪科学"信——这是学术争论里极罕见的烈度。

九条判断，三家各抽三条最承重的。

意识那三条：预注册的对抗实验可以做到双方事先签字接受被推翻；而这样做完之后两边核心预测都没活下来；且两边都保留了"被测的操作方式不等于理论本身"这个退路。

肥胖那三条：同一个观察事实可以被一边读作因、一边读作果；恒等式不能当作解释；两套模型在"想解释什么"上根本不是同一件事。

机器那三条：同一次停止输出可以被读作能力崩溃或读作明智止损；改掉一个具体设置就能让"崩溃"消失；被质疑不会思考的东西参与撰写了论证自己会思考的论文。

三十六次对撞，作废九条。

作废的分两类。六条是同义重复——两条判断说的其实是同一件事的两种说法。三条是撞不起来——中间没有互相顶住的地方。

这一栏在多数账本上一分不值。但它是这篇文章的地基——留下的二十七条之所以立得住，正是因为那九条被扔掉了。

留下的二十七条聚成五条暗流：

一、双方同意事实，分歧在"这算什么"。 二、"这算什么"这一步在测量之前就做完了，而做的时候手上没有数据。 三、每一方都有一个万能退路，而且这个退路常常是正当的。 四、大问题不动，小问题不断被解决，于是产生正在进步的错觉。 五、理论越丰富，可退的余地越大——丰富与可被推翻反着走。

第二条是根：**认出它在场这一步在测量之前，而当"怎么认"本身是争论内容时，测量就不再是裁判。** 第一条是这个根的表面症状；第三条是它的后果（退路之所以永远可用，是因为判据永远悬着）；第四条是它造成的错觉；第五条是它的一个不舒服的推论。

收敛成一条判断：一样东西如果没有一个独立于各家理论的在场标志，关于它的争论在原理上就不能被证据裁决；而各方仍会不断产出证据，因为那是唯一能做的事。

这条判断三家里谁也说不出来，而且原因就在各自的位置上。**意识那一家离答案最近**——他们做了最认真的对抗实验，亲眼看见它失败了；但他们把失败归给了"操作化不够好"，也就是归给了技术，而不是归给这一类问题的性质。**肥胖那一家已经说出了半句**——中立评论者指出"两家在想解释什么上根本不同"——但他们把这当作一个调停的理由，而不是当作一条关于争论可裁决性的普遍判断。**机器那一家最没有历史包袱，也因此最清楚地展示了这个结构**，但这个领域太新，还没有人把它跟别的领域连起来。

能在几个学科里同时兑现：本文第九节给了十一个——精神科诊断、疼痛、公司里的创新精神、教育里的真懂、经济学的生产率、法律里的故意、动物有没有感受、模型有没有在思考、一个人有没有变好、这篇文章本身有没有价值，以及婚姻里那句「你到底有没有在听」。

十五、这篇文章的自我审判

最后必须说对自己不利的话，而且这一次是双重的、而且是真的致命。

第一重：这篇文章的核心概念，自己就没有独立标志。

本文的整条判断建立在一个区分上：一样东西有没有"独立于各家理论的在场标志"。

那么问一句：**"是不是独立标志"这件事，有没有独立标志？**

没有。判断一个标志算不算独立于理论，本身就是一个需要判据的判断，而这个判据从哪儿来？从我这一套说法里来。

也就是说：本文所诊断的那个毛病，本文自己就有。 一个坚持"意识的报告就是意识的独立标志"的人，可以完全一致地说：那场架其实是有标志的，只是你不承认。而我拿不出任何独立于我这套说法的东西来反驳他。

我不打算假装解决这个问题。我只指出：**本文因此属于它自己所描述的那一类——它不能被证据裁决，只能被使用或不被使用。** 如果它有用，它会被用；如果没用，它会被忘掉。这就是它全部的检验方式，也是它应得的。

第二重：这篇文章会被打一个分数。

而"一篇文章有没有创新"这件事，是本文第九节列举的十个例子里的最后一个——**没有独立标志的典型。**

所以给这篇文章打分这个动作，正是这篇文章所描述的那种不能被证据裁决的判断。那个分数不会告诉任何人这篇文章好不好，它只会填上那个必须被填上的空位。

而更难堪的是：**我知道这一点，我仍然会给它打一个分数。** 因为不打分，这篇文章在它所处的那个系统里就没有位置——而没有位置的东西不会被读。

这一笔不是自谦，也不是收尾的花招。这是本文判断的直接推论，我绕不开。

那么写它还有意义吗?

有一点。因为在这篇文章之前，那三场架里的人各自都以为自己在朝结束前进——下一个实验、下一批数据、下一次对抗合作。**这篇文章至少让那个期待变得可见：你在设计下一个实验之前，可以先问一句，这场架有没有一支双方都读的温度计**。

知道自己在参与一场不会被裁决的争论，和以为下一个实验就能解决——这两者之间的差别，值这两万字。

但也仅止于此。

如果读完这篇文章，你只是从此多了一个说法可以用来打赢下一次争论——那你就正好在做它所描述的那件事。

如果读完之后的某一次争吵里，你在摆出第七条事实之前停了一下，问了一句"我们俩是不是根本不在用同一个判据"——那这篇文章才算被接住了。

而那件事，我永远不会知道。

本文由三个分属不同学科、且各自内部正在激烈冲突的理论体系碰撞而成，作者 王德生 + Claude。

本文由三个分属不同学科、且互相冲突的理论体系碰撞而成。 全部来源均为站外公开文献，可自行核对。 作者：王德生 + Claude · SDE Universes · 学科通融