

学科通融 · 之六十

临界回落

学习为何在抵达之后被记为从未发生，以及承接结构为什么不是帮助完成、而是阻止关闭

教育学 × 化学 × 管理学

作者 王德生 + Claude

栏目 学科通融 · 站外碰撞

篇幅 约3.4万字 · 45页

发表 2026年8月3日

摘要

一个学生正在往前想。他已经离开了原来那个安稳的不懂，还没有到达懂，他处在一个自己维持不住的位置上。然后铃响了，或者老师说“我来讲一下”，或者他自己觉得说出来会显得很蠢。三十秒后他回到原处。**在这堂课留下的全部记录里，这三十秒不存在**——他与后排那个从头到尾在想中午吃什么的同学，在所有的账上是同一笔。本文把这一类事件从不存在变成一个可以被清点的对象，并说明为什么它至今没有一行：三条互不相容的现行解释——取向说、路径说、计量说——对“那几十秒该由谁动”给出三个不能同真的答案（教师动／制度动／不需要任何人动），**而每一家的处方恰好是另一家诊断出的病因**；三家却共同假定了一条谁也没说出口的东西：在一段有界的过程里，改变从哪个位置发起是由角色事先定死的。推翻它的材料来自三家之一自己——在反应条件下，起作用的活性相不是投料时投进去的那一相，而是被反应本身长出来的，**发起权在同一个封闭体系内部换过手**。由此得到一张两轴辨别格（抵达 × 承接）、一句不含情态词的问话，以及两个必须分开的读数：**回落率**是三领域可通约的上界，**关闭率**才是承重的那个子集。把两个读数与机制放在一起还能直接推出一条**容量折线**：净学习在抵达数等于承接容量处取最大值，此后按固定斜率下降，**而拐点的位置只由课时长度、班额、允许的沉默时长这些结构量决定，可在开课之前算出来**；并在一份公开的多机构数据（三十一门课／二十八所机构／两千八百五十五名学生）上对这条折线做了实测——**控制承接容量之后，抵达事件的偏效应为零，该预言未获支持**；而无功效的原因本身是一个结果：现实中凡是提高提问密度的做法同时也提高了小组时间，**本模型所要的那个「容量不变而抵达增加」的对照在野外不存在，必须新造**。本文最锋利的一步在这里：承接的功能不是帮助学习者完成那次转化，而是在那段不可自持的时间里**阻止一次关闭性归因**——这解释了为什么及时给出答案在短期指标上是改善、在关闭上是加速，为什么同伴的作用常大于教师且在同伴水平更低时依然如此（同伴不具备关闭的权威），以及为什么大规模的思维干预效应很小（它改的是关闭的措辞，不是是否关闭）。最后一句本文不打算缓和：**近三十年真实有效的教学改良提高了抵达率，而承接容量没有同步改变，因此它们在提高完成数的同时更快地提高了回落数，而后者不出现在任何一份报告里**。

这一篇由哪三个学科的理论体系撞成

把三家摆到同一个问题上——一个学习者已经进入了他自己维持不住的推断，接下来的那几十秒该由谁动——三个答案不能同时为真。第一家答由教师动：取向不是学生能自己选的，它由感知到的评价要求诱发，指望他在这几十秒里自己翻转它等于指望结果去改原因；第二家答由制度动且绝不能由现场的个体判断来动，因为不被计量的动作在多任务代理下必然被从可测的那一侧抽走，而可审计只能审到可见动作那一层；第三家答不需要任何人动，因为转化不发生只由路径决定，在有多条竞争通道的体系里普遍加压永远是选择性的敌人。三条互相取消，而且各自的处方正是另一家诊断出的病因：教师及时进入正是第二家所说的普遍加压，把它做成可考核正是第一家所说的评价条件诱发浅层取向，零干预只修路径正是第三家所说的不可测者被判为不存在。三家均为站外的公开文献，链接直达原始出处，可自行核对。

教育学 · 深浅学习取向，与它下游那条更硬的证据

[Marton & Smljouml;, On Qualitative Differences in Learning: I—Outcome and Process \(<i>British Journal of Educational Psychology</i> 46:1, 1976\); 下游见 Kirschner, Sweller & Clark, Why Minimal Guidance During Instruction Does Not Work \(<i>Educational Psychologist</i> 41:2, 2006\)](#)

主张同一堂课同一教师同一份材料，取向不同学得完全不同，故教学法不决定学习、取向决定学习，而取向由学生感知到的评价要求诱发；而这一支自己确立的另一件事——取向不是一个人的稳定属性，同一学生在同一天的两门课里可以是两种取向——等于承认它是在每一次具体互动里被重新定住的，只是它从不问定在哪一秒。

<https://doi.org/10.1111/j.2044-8279.1976.tb02980.x>

化学 · 过渡态理论的单因主张，与反应条件下的活性相重构

[Tao, Grass, Zhang, Butcher, Renzas, Liu, Chung, Mun, Salmeron & Somorjai, Reaction-Driven Restructuring of Rh-Pd and Pt-Pd Core-Shell Nanoparticles \(<i>Science</i> 322:5903, 2008\); 另见 Halpern, Mechanism and Stereoselectivity of Asymmetric Hydrogenation \(<i>Science</i> 217:4558, 1982\); 理论侧见 Truhlar, Garrett & Klippenstein, Current Status of Transition-State Theory \(<i>J. Phys. Chem.</i> 100:31, 1996\)](#)

主张热力学允许的转化绝大多数不发生，不是因为没能量，而是因为没有可走的路，故唯有路径是决定项、普遍加压是选择性的敌人；而这一支自己在近常压条件下反复观测到的一件事——活性相是被反应本身长出来的，不是投料时投进去的那一相，且气氛一变可以换回去——说明发起权在同一个封闭体系内部换过手，只是化学把它记在“反应前的表征为什么与实际活性对不上”名下。

<https://doi.org/10.1126/science.1164170>

管理学 · 计量结构说，与它自己关于脱耦换代的观察

[Espeland & Sauder, Rankings and Reactivity: How Public Measures Recreate Social Worlds \(<i>American Journal of Sociology</i> 113:1, 2007\); 另见 Kerr, On the Folly of Rewarding A, While Hoping for B \(<i>AMJ</i> 18:4, 1975\); Holmstrm & Milgrom, Multitask Principal–Agent Analyses \(<i>JLEO</i> 7, 1991\); Bromley & Powell, From Smoke and Mirrors to Walking the Talk \(<i>Academy of Management Annals</i> 6:1, 2012\)](#)

主张只有什么被算作数是决定性的：当一项任务可测而另一项不可测时，对可测任务的任何强化激励都会把努力从不可测那一侧抽走，而被测者会重新组织整个条件场以适应测量；而这一支自己写明的脱耦换代——新的那一种是组织真的照做了、做得越来越认真、而这些动作与本来要服务的目的之间的联系已经断了——它用来解释合法性演化，没有用来问那些真的做出来的动作去了哪里。

<https://doi.org/10.1086/517897>

目 录

摘要6

一、引言：一个每天发生而没有名字的现象7

二、三条现行解释路线，以及它们各自最硬的那一条8

三、三条路线共同站着的那块地11

四、来自催化化学自己的两件旧证据12

五、临界回落：一类不在任何账本上的事件13

六、辨别格：抵达 × 承接14

七、一句不含“应当”的问话16

八、两个读数：可通约的那一个，与承重的那一个16

九、承接不是帮助完成，是阻止关闭18

十、三条路线为什么各自到不了这里23

十一、与最近的几种说法的分界24

十二、划完之后：一块所有人都站着的地基32

十三、三处反过来加在本文身上的约束34

十四、轮次：谁先动，谁跟上，谁反过来改了谁35

十五、证伪条件36

十六、本文可以被分层引用38

十七、反噬：这个读数一旦进入考核39

十八、不适用的边界39

十九、伦理与证据边界40

二十、本文自己落在哪一格40

二十一、一条写死日期的赌注	41
二十二、结语	41

摘要

对课堂上真正的学习很少发生这一现象，目前有三条互不相容的单因解释路线：学习取向说（教育学）、路径缺失说（催化化学，作为一个持有立场的学科而非比喻来源）、以及计量结构说（管理学与制度理论）。把三者摆到同一个问题上——一个学习者进入了他自己维持不住的推断之后，那几十秒该由谁动——三家给出三个不能同真的答案：教师动、制度动、不需要任何人动；而**每一家的处方恰好是另一家诊断出的病因**。本文不在三者之间裁判。本文指出三者共同假定而从未说出的一条：在一段有界的过程里，改变从哪个位置发起，是由角色事先定死的；并用催化化学自己长期掌握的一件旧事实推翻它——在反应条件下，起作用的活性相不是投料时投进去的那一相，而是被反应本身长出来的，发起权在同一个体系内部换过手。这条前提一旦松动，就出现一类在所有现行账本上都没有字段的事件：一次转化已经离开原态、抵达了不可维持的临界区，又因为它无法自持的那一小段时间里没有任何东西接住它而退回原处；这类事件在全部现行记录中与从未开始记为同一笔。本文提出一张两轴辨别格（抵达 × 承接）、一句不含情态词的诊断问话，以及两个必须分开的读数——**回落率**（三领域量纲不同而比率相同，是上界）与**关闭率**（承重的那个子集，只在会改变下一次概率的体系里成立）。本文论证承接结构的功能不是帮助学习者完成，而是阻止关闭性归因；给出轮次形式的正向预测与机制层面的证伪条件；并把与它最近的三位——僵局驱动的学习、有生产力的挣扎、习得性无助的归因重构——逐一交出可判决的分离线，其中最近的那一位与本文只隔一个词：**意识到**。把两个读数与该机制放在一起，可直接推出一条容量折线：净学习在抵达数等于承接容量处取最大值，此后按固定斜率下降，**而拐点的位置只由结构量决定、可在开课前算出**；据此对既有主动学习元分析中「效应在小班最大」这一结果给出相反的读法——**班额是承接容量的代理而非互动量的代理**，并对这条折线的区分性预言做了一次实测——在一份公开的多机构数据（三十一门课／二十八所机构）上，控制承接容量之后抵达事件的偏效应为零，**该预言未获支持**；无功效的原因本身是一个结果：现实中提高提问密度的做法同时提高了小组时间，「容量不变而抵达增加」的对照在野外不存在，必须新造。本文进一步指出，合意困难那一支已不能同时持有「可事先指导设计」与「困难须在学习者能力范围内」这两条，除非补上本文的变量或放弃前一条。一句话结论：近三十年真实有效的教学改良提高了抵达率，而承接容量没有同步改变，因此它们在提高完成数的同时更快地提高了回落数，而后者不出现在任何一份报告里。

关键词 临界回落；承接窗口；回落率与关闭率；关闭性归因；痕迹非负假定；抵达而未完成

Abstract. Three research programmes explain the scarcity of genuine classroom learning by three mutually exclusive single drivers: learners' approach orientation (education), the absence of a traversable pathway over the barrier (chemistry of catalysis, taken as a position rather than a metaphor), and the structure of what counts and is counted (management and institutional theory). Placed against a single question — who is to move during the seconds after a learner has entered an inference he cannot sustain — the three give answers that cannot all be true, and each one's prescription is precisely the pathology another one diagnoses. This paper does not adjudicate among them. It identifies what all three take for

granted — that in a bounded episode the position from which change originates is fixed by role — and overturns it with a fact long held inside catalysis itself: under reaction conditions the active phase is generated, not supplied, and initiative demonstrably changes hands within one system. Released from that assumption, a class of events becomes visible for which no existing ledger has a field: a transformation that has left its origin, reached the unstable region, and returned, because nothing held it during the window in which it could not sustain itself. In every current record such an event is indistinguishable from never having begun. The paper proposes a two-axis discriminant grid (arrival $\times$ uptake), a modality-free diagnostic question, and two readings that must be kept apart — the recrossing rate, commensurable across the three fields but only an upper bound, and the closure rate, the load-bearing subset. It argues that the function of uptake is not to help the learner finish but to prevent closural attribution, derives a round-structured prediction, specifies mechanism-level falsification conditions, and separates itself from its three nearest neighbours — impasse-driven learning, productive struggle, and the attributional reformulation of learned helplessness — the nearest of which differs from it by a single word: noticing. Combining the two readings with the mechanism yields a capacity kink: net learning peaks where arrivals equal uptake capacity and declines at a fixed slope thereafter, with the kink located by structural quantities alone and computable before a course begins. On that basis the paper re-reads the class-size finding in the existing active-learning meta-analysis — class size proxies uptake capacity, not interaction volume — and then tests the kink's discriminating prediction on an open multi-institutional dataset (31 courses, 28 institutions): controlling for uptake capacity, the partial effect of arrival events is null, so **the prediction is not supported**. The reason it has no power is itself a result: in the wild, every practice that raises question density also raises group time, so the contrast the model needs — capacity held constant while arrivals rise — does not exist and must be built. It further shows that the desirable-difficulties programme cannot hold both of its commitments, that difficulty can be prescribed in advance and that difficulty is desirable only within the learner's capability, without either adopting this paper's variable or demoting the first.

Keywords: critical recrossing; uptake window; recrossing rate; closure rate; closural attribution; nonnegative-trace assumption; arrival without completion

一、引言：一个每天发生而没有名字的现象

有一件事，几乎每个上过课的人都经历过，而它至今没有一个可以写进研究报告变量表的名字。

一个学生正在往前想。他还没有想清楚，但他已经离开了原来那个“不懂”的位置——他察觉到了一个不对劲的地方，他的表述开始变得不完整：“是不是因为……”“那如果反过来……”“可是刚才那个不是

说……”。这是一个特殊的状态：他既回不到之前那种安稳的不懂，也还没有到达懂。他处在一个无法长时间维持的位置上。

然后，其中一件事发生了：铃响了；老师说“好，我来讲一下”；另一个同学抢先给出了正确答案；或者他自己觉得说出来会显得很蠢，把嘴闭上了。

三十秒之后，他回到了原处。

到这里，问题出现了：**在这堂课留下的全部记录里，这三十秒不存在。** 课堂观察量表上，他没有举手，记零；当堂检测上，他这道题不会，记零；参与度评分上，他没有发言，记零。他与教室后排那个从头到尾在想中午吃什么的同学，在所有的账上是同一笔。

这不是“记录得不够细”。这是一件更根本的事：**现行的每一套课堂记录系统，都没有为“抵达过而未完成”设置字段。** 不是这个字段填得潦草，是没有这个字段。你可以把课堂录像看一百遍，可以把每一次转写编码到话轮层，你依然读不出这个数——因为没有人被要求判定“这个学生已经进入了一个他自己维持不住的推断”。判定这件事的编码规则不存在。

本文的全部工作，是把这一类事件从“不存在”变成“一个可以被清点的对象”，并说明：一旦它被清点，关于“为什么课堂上真正的学习很少发生”的三条现行解释，各自都要让出一大块地方。

需要先说明本文为什么把催化化学放进来，以及放进来的是什么。本文不是拿化学作比喻。比喻是“像”，它可以随时被替换成另一个比喻而论证不受影响。本文取的是催化化学作为一个学科所持有的一条实质立场——**决定一个转化发不发生的，唯有路径**——以及为支撑这条立场而积累的一批经验事实。这条立场与教育学的立场、与管理学的立场彼此不相容，正因为不相容，才有本文要做的事。若把化学换成任何一个不持有这条立场的学科，本文第四节的论证会立刻塌掉。这是本文可以被检验的一处：**去掉催化化学的那两件材料，第五节的构念就没有来源。**

二、三条现行解释路线，以及它们各自最硬的那一条

要看清一个共同的盲区，先要把三条路线各自最强的形式摆出来。摆的时候必须按它们自己最强的说法摆，不能摆成便于反驳的样子。三条都不是稻草人，三条在自己的问题里都是对的。

2.1 学习取向说：只有学生把这件事组织进哪条序列是决定性的

这条路线最经典的形式来自马顿与萨尔约 1976 年那篇研究。他们让学生读同一篇文章，然后发现，学生之间最大的差别不在能力，也不在教师，而在于他们在阅读时到底在做什么：一部分人在追那篇文章要说的意思，另一部分人在记那篇文章里可能被考到的东西。前者被称为深层取向，后者被称为浅层取向。此后四十年的大量研究反复确认了同一件事：**同一堂课、同一个教师、同一份材料，取向不同，学得完全不同。**

这条路线的单因主张是硬的：教学法不决定学习，取向决定学习；而取向由学生感知到的评价要求诱发。它给出的动力式是——**学生自己要拿到的那个东西，与他所处的那套评价条件长期不容，逼得他的学习做出另一副样子。**从理解型的样子，变成应试型的样子。

它的时序读数也是清楚的：考纲先动，取向在一个学期内跟上，教师的教法在两三个学期后跟上。这条路线因此对“改教法”始终不乐观——教法是被驱动的，不是驱动者。

这条路线自己知道、却不往那个方向用的一件事是：**取向不是一个人的稳定属性。**同一个学生在同一天的两门课里可以是两种取向；同一门课在期中之后可以整体翻转。既然它是逐次发生的，那么它就不是“学生带进教室的东西”，而是在每一次具体的教学互动里被重新生成的东西。这一点这条路线并不否认，但它没有把它当作一个需要解释的对象——因为它的问题是“什么使一个学生倾向于深层取向”，而不是“取向在一次互动的哪一秒被定住”。

2.2 路径决定说：只有能垒上有没有路是决定性的

催化化学在过去一百年里守着一极其强硬的单因主张：**热力学允许的转化，绝大多数不发生；它们不发生不是因为体系没有能量，而是因为没有任何可走的路。**

这条主张的精确形式来自过渡态理论。一个转化能否发生，由起点与那个最高点之间的高度差决定，而不由起点与终点之间的高度差决定。改变前者的手段有两种：一种是把整个体系加热——让所有分子都获得更多能量；另一种是修一条新的路——让转化沿一条最高点更低的路径进行。**前者粗暴且破坏选择性：加热在把目标反应推快的同时，把所有副反应也推快了。后者不消耗任何额外能量，什么都不必加，只是把路修通，转化就发生了。**这就是催化。

这条路线因此对“施加更多压力”这件事有一种结构性的怀疑：**在一个有多条竞争通道的体系里，普遍加压总是选择性的敌人。**它的动力式是——**已经存在的那些物种，与它们所处的条件场之间的不容，逼出一条新的路径。**

这条路线自己知道、却不往这个方向用的事实有两件，第四节会详细展开。这里先记住它们的名字：一是反应驱动的重构，二是“看得见的多数不是有效的那一条”。

2.3 计量结构说：只有什么被算作数是决定性的

第三条路线来自管理学与制度理论。它最锋利的表述属于科尔 1975 年那篇文章：奖励甲，指望乙。此后的发展把它推到了更硬的位置。霍姆斯特罗姆与米尔格罗姆 1991 年的多任务代理模型给出了一个结构性的结论：**当一项任务可测、另一项任务不可测时，对可测任务的任何强化激励，都会把努力从不可测任务里抽走。**这不是激励设计得不好，这是激励只要生效就必然发生的事。

埃斯佩兰与索德 2007 年关于法学院排名的研究把这条推得更远：排名不是在测量法学院，排名重造了法学院。被测者会重新组织自己所处的整个条件场以适应测量——改招生策略、改经费分配、改人事、改课程，最后连“什么是一所好法学院”这个概念本身都被改掉了。

这条路线的动力式是——**实际做出来的东西，与名义上的目标序列长期不容，逼得组织去改造它自己置身其中的那套条件**：改口径、改什么算数、改谁在看。它主张组织不会改变行为，组织只会改变它计量的东西。

这条路线自己知道、却不往这个方向用的事实是布罗姆利与鲍威尔 2012 年的那个观察：脱耦有两种，而它们正在换代。旧的那种是政策与实践的脱耦——说一套做一套，仪式性地挂着规章，实际上照旧。新的那种是手段与目的的脱耦——**组织真的照做了，做得越来越认真、越来越有技术含量，而这些动作与它们本来要服务的那个目的之间的联系已经断了**。评估越密，前者转成后者。这一点它写得很清楚，但它用这个观察去解释组织合法性的演化，没有用它去解释一件更具体的事：**当所有人都真的在做、做得越来越好，而目的没有到达时，那些真的做出来的动作去了哪里**。

2.4 三条路线为什么不能靠“综合”处理

有一种处理方式是：说这三条都对，它们分别是三个层面的因素，共同作用。这个处理是无效的，理由不在于它不温和，而在于它取消了这三条路线各自最有价值的东西。

三条路线的价值恰恰在于它们各自的排他性。取向说之所以有力，是因为它说教法不决定学习；如果它退成“教法也是因素之一”，它就不再能解释为什么四十年的教法改良没有改变深浅取向的分布。路径说之所以有力，是因为它说加压无效；如果它退成“能量也重要”，它就不再能解释为什么催化能做到加热做不到的事。计量说之所以有力，是因为它说行为不会因倡导而改变；如果它退成“文化建设也有帮助”，它就退回了它自己批判的那个位置。

三条路线不能被综合，只能被同时超越。 而要同时超越三条互不相容的解释，唯一的办法是找到它们共同踩着的那块地。

2.5 把三条摆到同一个问题上：那几十秒该由谁动

三条路线不能综合，还有一个比上一小节更硬的理由：把它们摆到同一个具体问题上，三个答案不能同时为真。

问题是这一个：一个学习者已经离开了原来那个安稳的不懂、进入了一个他自己维持不住的推断，在接下来的那几十秒里，该由谁动？

取向说答：由教师动。 取向不是学生能自己选的——它由学生感知到的评价要求诱发。学生在那几十秒里能调动的全部，恰恰是他已经带进来的那套取向；指望他在这几十秒里自己把它翻转过来，等于指望结果去改原因。这条路线的下游还有一条更硬的证据指着同一个方向：把探索完全交给新手的教学法，在几十年的对照研究里稳定地差于有指导的教学（基尔施纳、斯韦勒与克拉克 2006）。所以那几十秒里正确的动作是教师及时进入。

计量说答：由制度动，而且绝不能由现场的个体判断来动。 一个不被计量的动作，在多任务代理的结构下必然被从可测的那一侧抽走——这不是激励设计得不好，这是激励只要生效就必然发生的事。要那几十

秒里发生正确的事，唯一的办法是先让它可测、可审计、可进考核；而可审计只能审到可见动作那一层。所以那几十秒里正确的动作是把它变成一条可被检查的程序。

路径说答：不需要任何人动。 一个转化能不能发生由路径决定，不由外部推动力决定。在一个有多条竞争通道的体系里，普遍加压总是选择性的敌人——它把目标反应推快的同时把所有副反应也推快了。所以那几十秒里正确的动作是**零动作**：只要路径已经修通，转化自己会翻过去；此时任何外加的推动都只会把它推向别的通道。

三个答案穷尽了“谁动”的全部可能——教师、制度、没有人——而且三条互相取消：

- 教师及时进入，正是路径说诊断的那件事：**在有竞争通道的体系里普遍加压**。按路径说，及时讲解不是把学生推向理解，是把他推向那条阻力更小的通道——记住结论。
- 让它可测可审计，正是取向说诊断的那件事：**评价条件诱发取向**。按取向说，一旦“那几十秒里做了什么”进入考核，学生感知到的评价要求随之改变，而取向会在一两个学期内跟上；跟成哪一种，取向说四十年的证据说是浅层那一种。
- 零动作、只修路径，正是计量说诊断的那件事：**不可测者被判为不存在**。按计量说，一件不产生可见动作、不被记录、不进任何报表的事，在制度里等于没有发生；把它交给“路径自己会走通”，等于把它交给一个没有人负责、因而也没有人会去做的位置。

三家各自的处方，恰好是另一家诊断出的病因。 这是本文认为这三条不是三个层面而是一场活争论的全部理由：它们不在补充彼此，它们在同一个位置上互判对方为错。任何一个“三者共同作用”的说法，都要先说明为什么可以同时执行三条互相取消的处方。

还有一件事记在这里，因为它是第九节那条推论的伏笔：三个答案穷尽了“由谁动”，**却都假定那几十秒里需要有一个动作者**。教师、制度、或者明确的“没有人”——三者是对同一个空格的三种填法。本文最后给出的答案不在这个格子里：动的不是任何一方，**要被撑住的是那几十秒本身**。

三、三条路线共同站着的那块地

把三条路线的动力式并排放好，会看到一件它们自己不会指出来的事。

- 取向说：评价条件先动，学生做出来的样子跟着变。**学生是被驱动的一方**。
- 路径说：条件场先动，路径被逼出来。**路径是被驱动的一方**。
- 计量说：实际产出与名义目标的不容先在，条件场被改造。**条件场是被驱动的一方**。

三条路线各自指认了一个“被驱动者”，而且三个被驱动者互不相同。这看起来像是三种不同的立场。但它们有一个共同点，而这个共同点从来没有被写出来：

在一段有界的过程里，谁是发起者、谁只能跟着动，是事先给定的。

取向说假定学生永远处在被评价条件推动的位置上——它从不设想学生的某一次推断反过来重设了这堂课的评价方式。路径说假定条件场是外加的、给定的——催化剂是投进去的，温度是设定的，路径是在这些给定条件下被逼出来的结果。计量说假定组织的调节永远落在条件场那一侧——它从不设想一次具体的实际产出反过来把“什么算数”这件事重设了。

把这条前提念给三条路线中的任何一条听，都会得到同一个反应：这还用说吗。教师是教的，学生是学的，制度是定规则的；催化剂是催化剂，反应物是反应物。**正因为三条都会说“这还用说吗”，它才是那条共有的前提。**会引起争论的东西不是共有前提，是第四种立场。

这条前提有一个直接后果，而这个后果正是本文要处理的东西：

如果发起位置是固定的，那么“一次转化中途折返”这件事就没有独立的本体地位。它只是“这一轮没有被驱动成功”，也就是一个否定性的结果，而否定性的结果与“根本没有被驱动”在账上是同一件事。三条路线不是漏记了它，是**它们的结构使得记它成为不必要的**——如果发起者始终是那一个，那么只要看发起者做了什么、被驱动者最终变成了什么，账就记平了。中间那一段属于谁，不需要问。

要拆掉这条前提，不能从外面搬一个理由。从外面搬理由得到的是第四种立场，是加入争论，不是取消争论。要取消它，必须用三条路线之一自己已经掌握、已经反复引用、只是没有往这个方向用的事实。

这件材料在催化化学手里，而且已经放了十几年。

四、来自催化化学自己的两件旧证据

4.1 活性相不是投进去的那一相

2008 年，陶等人在 Science 上发表了一项在近常压条件下用光电子能谱观察双金属纳米颗粒的工作。他们把铈与钡以核壳结构做成纳米颗粒，然后在真实的反应气氛下观察它。结果是：在氧化条件下，铈原子偏析到壳层；在还原气氛下，换成钡原子扩散到壳层。组成与化学态发生剧烈的、可逆的改变。

在催化化学内部，这项工作回答的是一个非常具体的问题：**为什么反应前的表征结果与实际活性对不上。**长期以来，人们在反应之外测量催化剂的结构，然后用这个结构去解释反应中的活性。这项工作以及同一时期整批同类工作说明：不能这样做。**真正的活性相是在反应条件下被生成的，它不是你投进去的那一相。**由此发展出的一整套原位与工况表征方法，如今是催化研究的标准配置。

现在把这件事从它原来回答的问题里取出来，问一个催化化学从不问的问题：**在这个体系里，谁是条件，谁是被条件逼出来的东西？**

反应开始之前，答案是清楚的：催化剂表面是给定的条件，反应路径是待定的、要被这些条件逼出来的结果。反应开始之后，答案倒了过来：被逼出来的那条路径改造了催化剂表面，于是下一段时间里真正起作用的那个“给定条件”，是上一段时间的产物。**在同一个封闭体系里，不加入任何外部干预，发起权换了手。**而且陶等人的观察还说明它是可逆的——气氛一变，它可以换回来。

这就是那件推翻材料。它满足最严的一条要求：**它不是从外部搬来的理由，它是三条路线之一自己长期掌握、反复引用、写进教科书的常识；它之所以能推翻那条共有前提，正是因为它在它自己的领域里它回答的是另一个问题。**

一旦承认发起位置不是固定的，“中途折返”就立刻获得了独立地位。因为如果发起权可以换手，那么“这一段时间里本该由某个位置接手而没有接”就是一个真实的事件，而不是一个否定性的结果。它有确定的时间边界，有确定的位置，有可以被指认的后果。

4.2 看得见的多数不是有效的那一条

第二件材料更老。1982 年，霍尔本在 Science 上给出了铑-手性膦催化不对称氢化的机理结论：立体选择不是由底物与手性催化剂之间那个占优势的初始结合模式决定的，而是由那个不占优势的次要非对映体高得多的反应性决定的。

用非化学的话说：在溶液里，你用核磁看得见的那个中间体，占绝大多数，稳定，好测；而产物几乎不从它那里来。产物来自另一个中间体，它浓度极低、常常低到测不出来，但它翻过去的速度快得多。

在化学内部，这项工作回答的问题是“为什么对映选择性与结合的热力学偏好相反”，它终结了当时流行的“锁钥”式想象。它从来没有被用作一条关于观测的一般命题。但它本身就是一条关于观测的一般命题：

在一个有多条并行通道的体系里，最容易被观测到的那一条，恰恰可能是通不过去的那一条；而真正通得过去的那一条，可能低到检测不出。

这两件材料合起来，给出了本文要做的事的全部依据。第一件说明：中途折返是一个有独立地位的事件。第二件说明：这类事件恰恰是最不容易被现行观测抓到的一类——因为**现行的课堂观测全部建立在“看得见什么”的基础上**：他举手了吗，他发言了吗，他写了吗，他答对了吗。而抵达临界并折返的那一段，全程可以是完全无声的。

五、临界回落：一类不在任何账本上的事件

现在可以把那一类事件定义出来。

临界回落：一次转化已经离开原态，进入了一个它自己无法长时间维持的中间状态；在它无法自持的那一段时间里，没有任何外部结构承接它；于是它回到原态。这一次往返，在所有现行记录中与“从未离开原态”记为同一笔。

三个要件，缺一不可：**离开过原态、进入了不可自持的状态、在那段时间内未被承接**。第三个要件里的“承接”是什么，第九节会给出一个与直觉相反的答案。

先说清楚它为什么不是“记录得少”，而是“不算一样东西”。

在教育学的账本上，字段是这些：出勤、参与度、发言次数、正确率、作业完成度、测验分数、学习时长。这七个字段没有一个能容纳临界回落。参与度和发言次数测的是“有没有做出可见动作”，而临界回落全程可以无动作。正确率与测验分数测的是终态，而临界回落的终态与从未开始的终态完全一致。学习时长测的是暴露量，而两个暴露时长相同的学生，一个抵达过五次全部回落，一个从未抵达，读数一样。

在管理学的账本上，字段是这些：交付、完成率、周期时间、在制品、返工、缺陷。约束理论与精益的整套工具建立在一个假定上——**损失会堆积，所以看瓶颈前的堆积就能找到损失。**而临界回落的关键性质恰恰是：**它不排队。**一个未被接住的推断不会积压在任何地方等着被处理，它消失了。这就是为什么一整套以“看积压”为基础的管理工具从来没有发现过它。

在化学的账本上，字段是这些：转化率、选择性、收率、周转数、周转频率。有意思的是，化学是三者中唯一给这件事留了一个位置的：过渡态理论假定凡是越过分界面的轨迹都到达产物，而实际上一部分轨迹会折返，因此真实速率要乘上一个小于一的透射系数。但化学**正确地**把它降级成了一个修正因子而不是一个对象——理由第十三节会讲，它同时也是对本文的一处重要约束。

于是有了本文的第一条判断，它是三条路线中任何一条都不会同意、而三条路线的证据合起来只能得出的：

课堂上真正的学习很少发生，不是因为学生没有进入学习，不是因为难度没有匹配，也不是因为激励指向了别处；而是因为学生每天大量地进入学习又退回，而“进入过又退回”与“从未进入”在所有现行记录里是同一笔账——于是三条路线各自的补救，全都以这笔账为输入，因而全都在提高进入率的同时更快地提高回落率。

最后半句是本文最容易被检验、也最容易被误读的一句，需要立刻说清楚。它不是说主动学习无效。弗里曼等人 2014 年在 PNAS 上那份覆盖两百多项研究的元分析给出的正效应是稳定的，本文不挑战它。本文说的是另一件事：**那份元分析的结果变量是考试分数与挂科率，也就是“完成转化”那一格的读数；“抵达而未完成”那一格的读数，在全部两百多项研究中没有一项报告过，因为没有一项设过这个字段。**一个提高抵达率而承接容量不变的改革，会同时提高完成数和回落数；前者被报告，后者不被报告。这两件事完全可以同时为真。

这也解释了一个真实存在而长期被归为“教师的主观印象”的经验难题：为什么大量一线教师在采用主动学习之后，普遍报告“课堂比以前活跃多了，但真正看懂的人好像没多几个”。按本文的判断，这个印象不是错觉，它是对回落数上升的直接感知——只是它没有可以落笔的地方，于是被写成了一句抱怨。

六、辨别格：抵达 × 承接

一个名字不解决问题。要让它成为一个能分开两样看起来一样的东西的读数，它必须是两轴的。

第一根轴：这一次，系统有没有真的离开原态、进入不可自持的中间状态。第二根轴：在它处于该状态的那段时间里，有没有东西接住它。

未被承接	被承接
抵达临界 II 临界回落：往返一次，记为零。短期指标改善，失效无信号，越正规化越严重。 未抵达临界 IV 显性失败：任何人都看得见，不需要本文。 四格必须逐格给出读数方式，否则它只是一张画。 格I 转化完成。 读数：延迟测验上出现该概念的迁移性使用，且学生能复述自己曾经卡在哪里。第二个条件不是修辞——完成的转化留下路径记忆，未完成的不留。 格II 临界回落（靶格）。 读数：课堂转写中出现一个未完成的推断（以试探性连接词开头、句法未闭合、包含一个自我提出的疑问），在此后若干个话轮内没有任何人接续该推断所指的那个疑问，话题转走。这一格是本文全部工作的目标，它必须满足三条，否则第二根轴选错了： · 它在短期指标上表现为改善。老师讲解之后当堂正确率上升，课堂气氛比沉默时活跃，参与度评分更高。三项都是改善。 · 它的失效不产生任何信号。回落的学生不抗议、不缺勤、不表现异常；他与从未抵达的同学在所有记录上一致。 · 它自我加固。教师观察到“我一讲解，正确率就上去”，于是下一次讲解得更早；讲解得越早，学生停留在临界状态的时间越短，回落越多；而回落越多，当堂正确率对讲解的依赖越强。越是把教学做得规范、及时、反应快，这一格越大。 三条全部成立，靶格成立。 格III 空承接。 读数：承接动作（提示、追问、同伴讨论、支架材料）密集出现，而转写中未完成推断的出现频次接近零。这是当代精细化教学设计的系统性归宿：每一个可能的困难都被提前预设、被拆解成可管理的小步，于是学生一次也没有进入过不可自持的状态。这一格的所有指标是四格中最好看的——流畅、无卡顿、当堂达成率高、学生评价满意。它的问题不是学生没被帮助，是没有任何东西需要被帮助。 格IV 显性失败。 读数：既无未完成推断，也无承接动作。这一格不需要任何框架就能识别，它也不是本文的对象。 这张表最重要的用途是把两组容易被混为一谈的东西分开： · 格II 与格IV：现行的一切课堂记录把它们记为同一笔。本文的第一项主张是它们必须分开。 · 格III 与格I：现行的一切课堂评价把它们记为同一笔，而且格III的读数通常还更好看。本文的第二项主张是它们必须分开。	I 转化完成：真正意义上的学习发生。 III 空承接：承接结构在空转。教学设计精美，流程完整，指标最好看。

七、一句不含“应当”的问话

一个构念要成为工具，需要一句可以拿去问记录、而不是拿去问人的判断的问话。它不能包含“应当”“有意义”“真正”“充分”“恰当”这类词——凡是需要评判者作价值判断的问话，都会退化成评判者的偏好。

本文的那一句是：

从这个学生说出一个他自己没有说完的推断，到有人给出结论，中间隔了几个话轮？在这几个话轮里，有没有人朝他那个未完成的疑问再走一步？

它可以拿去问课堂录像，可以拿去问转写文本，可以拿去问在线讨论的时间戳。它不问任何人“这堂课好不好”。

把它放到五个互不相同的具体场景里跑一遍：

场景一，中学数学习题课。 学生说“如果这个点在圆外的话是不是就……”，教师说“对，很好，我们看下一步”。答案：零个话轮，承接动作为零。教师说了“很好”，那是评价，不是朝那个疑问走一步。

场景二，研究生讨论课。 学生提出一个半成型的反例，另一位学生立刻指出这个反例不成立。答案：一个话轮，承接动作为负——反例被证伪不等于那个疑问被推进，而且它给出了一个明确的关闭信号。

场景三，手术示教。 住院医师的手停住了，他不确定下一层是什么。带教说“你先说说你在犹豫什么”。答案：一个话轮，承接动作成立。请注意这里承接的内容不是答案，是**要求他把那个不可自持的状态说出来**——这个动作本身把状态延长了，使它不必立刻塌回去。

场景四，企业新人跟项目。 新人在会上问了半个问题就被打断，会后没有人回到那个问题。答案：全程零，且窗口在会议结束时永久关闭。

场景五，独自阅读。 读者读到一个不通的地方，停住，翻回去两页。答案：承接者是他自己，承接动作是回读。这一场景说明承接者不必是另一个人。

五个答案互不相同。其中至少两个不能从构念本身推出来，是查出来的：场景二的“承接动作为负”——反例被驳倒会加速回落，这不是定义的推论，是需要观测的事实；场景三的“承接的内容不是答案而是要求外化”——这条在第九节要单独论证，它不能从“承接”这个词推出来。

八、两个读数：可通约的那一个，与承重的那一个

一个构念如果只在一个领域里可测，它就不是跨领域的判断，是一个换了行头的本领域概念。检验方式只有一个：给出一个三领域量纲完全不同、而比率含义相同的读数。

回落率 = 最近 N 次抵达临界的事件中，未被承接而退回原态的次数所占的比例。

在化学里：透射系数的补数。轨迹越过分界区后返回反应物一侧的比例。可以从分子动力学轨迹文件里逐条清点。

在课堂里：转写文本中，学生说出的未完成推断在其后若干话轮内未被任何人接续、话题转走的次数，占未完成推断总数的比例。可以从课堂录像里逐条清点。

在组织里：已启动并进入不可自持状态（已投入资源、已产生对它的依赖、尚未交付）而后被搁置的工作项，占已启动工作项的比例。可以从工单系统与版本库的时间戳里逐条清点。

三处量纲完全不同：一处是轨迹条数，一处是话轮，一处是工作项。三处的比率含义相同。

这个读数要成立，必须过四条检查，其中第四条最容易漏而最要命。

第一，无量纲。 三领域可以并排比较。✓

第二，可事后清点。 三处所需的原始材料在多数机构的既有记录里已经存在——轨迹文件、课堂录像、工单日志。不需要新建任何采集系统。✓

第三，它把一个印象变成一个数。 “这堂课挺热闹但好像没学到什么”，这是一个印象；“这堂课出现十七次未完成推断，其中十四次未被接续”，这是一个数。✓

第四，它必须与既有的主要指标正交。 这一条的检验方式是：举出一个“既有指标最好看、而这个数最难看”的真实例子。举不出来，说明这个数会被立刻读成既有指标的一个代理，白造。

三个例子：

- **化学：**一个收率百分之九十五的反应可以有很高的折返比例。只要给足时间与推动力，折返的分子会再爬一次。反过来，一个折返比例接近零的反应可以收率极低——路修得笔直，但通向的是副产物。**收率与折返比例可以任意组合。**
- **课堂：**一堂当堂正确率百分之百、参与度评分满分、学生满意度最高的公开课，回落率可以是一。每一个半成品念头都被教师接过去替他说完了；每一次接过去都是一次评价性收口。**这正是格II的定义。**
- **组织：**一个按时交付率百分之百的团队可以有很高的回落率。所有超出既定范围的探索都被及时砍掉了，砍得越及时，交付率越好看。

三个例子都成立，正交性通过。

需要立刻说明的一点：**分母比分子难。** “抵达临界”的判定需要一套编码规则，而这套规则目前不存在。本文给出的操作定义（试探性连接词开头、句法未闭合、包含自我提出的疑问）是一个起点，不是一个成熟量表；它的编码者一致性需要被单独检验，而这正是本文最容易被推翻的地方之一——**如果两个受过训练的编码者对“这个学生是否已经抵达”无法达成可接受的一致性，那么整个读数就不成立。** 这一点写在这里，不写在注释里。

8.1 一处必须做的拆分：这个比率是上界，不是那个承重的数

上面那个数还不能直接用，因为它的分子里混着两类符号相反的事件。

第九节将论证：回落之所以是损失，只在它伴随一次关闭性归因时才成立；不伴随的那一类留下的痕迹仍是正的（第十三节第三处约束会把这条写死）。既然如此，一个把两类一起数进去的比率，就不是本文机制所预测的那个量。**这不是精度问题，是分子不干净。**

所以本文交出两个数，不是一个：

- **回落率 R** = 最近 N 次抵达中，未被承接而退回原态的次数所占的比例。只需一套转写编码，成本低，且三个领域量纲不同而比率含义相同。
- **关闭率 C** = 其中伴随一次关闭性归因的次数所占的比例。需要多编一项：那次退回之后的短时间内，学习者是否作出了一次关掉自己、关掉对象或关掉路径的表述。

两者恒有 $R \geq C$ 。差值 $R - C$ 是本文解释不了的那一类回落，第十三节第三处会承认它在数量上可能占多数。

本文的机制预测的是 C ，不是 R 。 R 仍然要报，有两个理由，而且只有这两个：其一，它是 C 的上界，因此 **R 为零可以直接排除本文的机制**，这是一条极便宜的否定通道；其二，它比 C 便宜一个量级，是唯一能大规模回溯清点的那个数。

还有一件事，它把本文的适用范围钉在了读数这一层：**只有 R 是三领域可通约的， C 不是。** 化学里没有关闭性归因的对应物——分子不作归因，一个折返的分子下一次爬得一样快。这不是本文没把化学那一侧写完，这恰恰是第十三节第一处约束在读数层的形式：**在折返可重试的体系里 R 就是全部；只有在一次折返会改变下一次折返概率的体系里，承重的才是 C 。** 两个数在哪里合一、在哪里分开，本身标出了本文说得上话的范围。

此处拆分必须能被推翻，否则它只是把一个数写成两个：若在多个班级中 R 与 C 的相关系数稳定高于 0.8，则拆分没有做任何工作， C 只是 R 的一个代理，本文应当退回单读数的版本，并连同交出第九节那条机制。这一条进第十五节的清单。

九、承接不是帮助完成，是阻止关闭

到这里，最自然的推论是：既然回落是因为没被接住，那就多接。而“接住”最自然的实现方式是——在他卡住的时候给他提示，帮他把那一步走完。

这个推论是错的，而它错在哪里，是本文最锋利的一处。

先看一个反常的经验事实。生成效应的研究，尤其是科内尔、海斯与比约克 2009 年的那组实验，给出了一个反直觉的结果：**让学习者先尝试回忆一个他根本还不知道的答案，哪怕这次尝试完全失败，之后再看到正确答案，学习效果也优于直接看答案。** 失败的提取本身有正效应。

这与本文的判断似乎冲突：如果失败的尝试有正效应，回落怎么会是损失？

它不冲突，而这个不冲突之处正好把本文的机制逼精确。在科内尔那组实验里，**尝试失败之后，正确答案随即到来了**。也就是说，那个不可自持的状态被终结了，而终结它的方式是给出结论。按本文的格子，那是格 I，不是格 II。

那么格 II 与格 I 的差别到底在哪一秒发生？本文的回答是：

在于那次未完成的抵达之后，学习者是否作出了一次关闭性归因。

关闭性归因是一个具体的、可以被听见的动作，它有三种典型措辞：“**我不行**”（关掉自己）、“**这个东西没意义**”（关掉对象）、“**这条路走不通**”（关掉路径）。三种都有同一个后果：**下一次遇到同一类不对劲时，不再离开原态**。它不是情绪，是一次路径的封闭。

这就是为什么回落不是回到原点。**回落把系统送进了一个比原点更深的位置**。在催化里这件事有精确的对应：一个中间体可以跑出催化循环，掉进一个更稳定的热力学阱里，把催化剂锁住。它比反应开始前更难动。

于是承接的功能被重新定义了：

承接结构的功能不是帮助学习者完成那次转化，而是在那段不可自持的时间里阻止关闭性归因的发生。

这条推论一旦成立，几件长期矛盾的观察立刻并到一条线上：

第一，为什么“及时给出答案”是最强的关闭动作。 教师给出答案，同时给出了两个信号：这个疑问已经被处理了；处理它的人不是你。第二个信号正是“这条路我走不通”的最有效确认。**给答案在短期指标上是改善，在关闭性归因上是加速**。这两件事从来没有被放在一起考虑，因为后者没有读数。

第二，为什么“你先说说你在犹豫什么”这类动作有效，而且效果与它是否解决问题无关。 它把一个内部的、不可自持的状态外化了。外化之后，那个状态获得了一个外部支撑物，它可以在不被解决的情况下继续存在。**它买到的不是答案，是时间**。

第三，为什么同伴的作用常常大于教师，而且在同伴水平更低时依然如此。 同伴不具备关闭的权威。一个同样不确定的人接住你的疑问，他无法给出结论，因此他不会关闭；他只能陪你在那个状态里多待一会儿。**从本文的角度看，这是承接的纯粹形式**。

第四，为什么成长型思维的干预在大规模复制中效应很小。 耶格尔等人 2019 年那项覆盖全美的实验给出的效应量相当小，且集中在特定学生群体。按本文的判断，这是意料之中的：成长型思维干预改变的是关闭性归因的措辞（从“我不行”改成“我还不行”），而不是**是否发生关闭**。一个持成长型信念的学生在无人承接的窗口里，会把归因从关掉自己改成关掉路径——“这条路走不通，我换一条”——路径同样被关闭，而且这一版的关闭听起来更健康、更不容易被察觉。**在承接结构不变的条件下，本文预测思维干预对回落率没有影响**。

这一段是本文最容易被误读为“反对教师讲解”的地方，所以要写死一句：**本文不主张教师不给答案。**

本文主张的是一件更窄也更可查的事——**给答案的时点，与承接窗口的宽度，是两个独立的变量，而现行的一切教学建议只调了前一个。**

9.1 容量约束：从两个读数直接推出来的一条折线

第八节交出两个数，第九节交出一条机制。把两者放在一起，会得到一个正文此前没有说过的东西——它不是一个新主张，是那两样东西的直接后果。

记一节课中抵达的次数为 N ；记这节课的**承接容量** K ——在当前的课时结构下，最多有多少次抵达能被接住。 K 完全由结构量构成：课时长度、班额、允许的沉默时长、场内有多少人可以充当承接者。它不含任何关于学生能力或内容难度的量。

于是被承接数为 $\min(N, K)$ ，未被承接数为 $\max(0, N-K)$ 。记 p 为未被承接的抵达中伴随关闭性归因的比例，则第八节那两个数就是 $R = \max(0, N-K)/N$ 与 $C = p \cdot R$ 。再记 λ 为一次关闭性归因造成的未来抵达损失——第九节已经论证它大于零：关闭关掉的是下一次离开原态的可能。

本节的净学习量因此是：

$$L(N) = \min(N, K) - \lambda \cdot p \cdot \max(0, N-K)$$

分成两段看： $N \leq K$ 时 $L = N$ ，每多一次抵达净增一份； $N > K$ 时 $L = K - \lambda p(N-K)$ ，每多一次抵达净减 λp 。 L 在 $N = K$ 处取最大值，此后单调下降，而下降的斜率与 N 无关。四条推论，都不是本文此前说过的。

推论一：拐点的位置由承接容量决定，与学生能力、内容难度、教师水平全都无关。 一切关于“参与度要适度”的说法，长期停在“适度”两个字上，因为没有人说得那个度在哪里。这里说得出：它在 $N = K$ 。而 K 全部由结构量构成，因此它可以在开课之前算出来，不需要等任何一次测验。

推论二：收益曲线是折线，不是钟形。 这一条比“倒 U 形”更强，也更容易被推翻。倒 U 预言两侧都是曲线、拐点平滑；折线预言**左侧斜率恒为一、右侧斜率恒为 $-\lambda p$ 、拐点是尖的**。二者可分辨：在同一门课程上以固定步长逐级提高抵达率并记录延迟迁移，形状就出来了。若实测是光滑的倒 U，本节的容量模型为伪——那说明承接容量不是一个硬上限，而是随负荷渐减的量，本节应当重写。

推论三：既有元分析里那条关于班额的结果，可以用本模型重新读一遍。 弗里曼等人那份覆盖两百多项研究的分析报告：主动学习在各种班额下都有效，而**效应在不超过五十人的小班里最大**。标准解释是小班里互动更多——按本文的记法，那是说小班的 N 更高。本文给出相反的读法：**班额是 K 的代理，不是 N 的代理**。小班里可充当承接者的人数更多、教师能分给每一个未完成推断的时间更长、允许的沉默更长，因此 N/K 更低，而不是 N 更高。

写到这里必须停一下，因为本文原本打算把这一步说成一条便宜的检验，而查证之后不能这么说了。那份元分析已经按干预类型做过一次异质性分析，结果是零：八个类别之间效应量无显著差异。若只看这个结果，它对本文不利。

但它对本文并不构成反证，理由本模型自己给得出：那八个类别分的是学生在做什么——应答器、学案、问题导向、工作室、案例、演示——而不是结论在什么时候到达。它们整体落在 N 那一侧。一个与 K 正交的分类法，必然给出零异质性；这正是本模型的预言，而不是它的反例。

更值得记下来的是那份分析自己在同一处写下的话。作者说这个零结果不可当真：样本极不平衡；一大批实验只能被笼统编码为“学案”，因为原作者用的是“课堂上的合作活动”“小组解题”“课内导引”这类说法；目前学界对各种主动学习类型在实践中究竟指什么并无共识；而且原作者很少交代活动的强度、时长、提问层次、分组方式与任务内容，也很少说明练习是否计分、按参与还是按正确计分。作者由此明确提出一项要求：这个领域需要一套客观、可重复、并且有理论根据的干预分类法，否则效应量的异质性无从理解。

本文提供的正是这样一套，而且它只有一个问题、只有两个取值：这项干预的定义里，有没有一条“学生交出之前教师不给出结论”的规则？有则记 K，无则记 N。它不问活动形式，不问强度，不问技术，因此与那八个类别正交。

代价也要如实交出。本文对该元分析的文献表做了一次试编码，只按方法名判读：一百八十六条里，三十一条可判为 K（集中在同伴教学、同伴引导小组、引导式探究与问题导向这四族），三十七条可判为 N（集中在应答器与测验、技术增量、演示与节奏调整这三族），其余一百一十八条仅凭方法名无法判定，占六成三。也就是说，这条检验不需要新采集数据，但需要回到约一百五十份原始文献的方法部分逐份判读；而按该元分析自己的说法，其中相当一部分原文并没有写清判读所需的那些细节。本文此前把这一步称作最便宜的一条检验，那是说错了，现在收回。它是可做的，但它是一个需要数周文献工作的项目，不是一次案头计算。

这一条的证伪形式因此要写得更硬：若按上述二值编码重新分析之后，K 与 N 之间仍无显著异质性，而班额仍是独立的强调节变量，则本文对班额那条结果的重读为伪，容量模型失去它唯一一处可在既有文献上兑现的支点。

推论四：那份元分析自己留在讨论部分的一个未决问题，本模型有一个明确的答案。该文写明：那些研究里实施主动学习的教师都是自愿者，若要求全体教师实施，效应是否还有这么大，是一个开放的问题。按本模型，答案是应当显著衰减，而且原因是可以说出来的——自愿者是那些愿意等的人，他们的实践本身就是高 K 的；而行政推行能提高的是 N 不是 K（课时长度、班额、允许的沉默时长在推行前后一律不变）。衰减幅度应当随 N/K 上升。这个洞不是本文去找的，是那份元分析自己交出来的；本文只是把它填上，并给了一个可查的形式。

最后一条，它同时是本文适用范围的形式表达。令 $\lambda \rightarrow 0$ ，全式退化为 $L = \min(N, K)$ ：回落只是没有收益，不构成损失，本文整篇不成立。 $\lambda > 0$ 正是第十三节第一处那约束——抵达机会稀缺、且每次抵达改变体系本身——的形式版本。那一节交出去的那句话，与本节的这个极限，是同一件事的两种说法；全文因此只有一套说法，没有两套。

9.2 这条预言已经被检验过一次，结果是否定的

上一小节把话说得很满，所以这一小节必须先交结果：本文对 9.1 那条区分性预言做了一次实测，它没有通过。

数据。 桑德斯特伦等人二〇二五年发表了一份多机构研究，比较四种成熟主动学习方法（ISLE、同伴教学、教学导引、SCALE-UP）对概念理解的相对作用：三十一门课、二十八所机构、两千八百五十五名学生，每门课有前后测概念量表、三次完整课堂录像的逐两分钟行为编码，**全部数据公开**。这正是弗里曼那份元分析所呼吁的第二代研究，而且它顺带回答了上一小节留下的第一个问题：**当方法被精确命名并被实地观察时，方法之间的异质性是显著的**（ $Q=8.11$, $df=3$, $p=0.044$ ）。弗里曼那个零结果确实是编码粗糙的产物，这一点本文猜对了。

操作化，在看任何结果之前写死。 用该研究自己的行为编码汇总：**抵达事件 N** = 出现「教师提问」或「教师提应答器问题」的时段占比；**承接容量 K** = 出现「应答器组内讨论」或「学案小组」或「其他小组活动」的时段占比。两条互斥预言由此分岔：**标准解释**说学生中心的时间越多越好，因此在 g 对 K 与 N 的回归里 **N 的系数应当不为负**；**容量模型**说同样的容量里塞进更多抵达会增加回落，因此 **N 的系数应当为负**。二十六门课同时具备两侧数据。

结果。

方法	课数	学习增益 g	N	K	$\rho=N/K$	平均班额
ISLE	5	0.60	0.80	0.36	2.52	66
同伴教学	9	0.62	0.79	0.36	2.67	177
SCALE-UP	5	1.25	0.82	0.66	1.26	69
教学导引	7	0.93	0.85	0.72	1.22	51

第一眼看上去像是支持： ρ 与 g 的相关是 -0.314 ，置换检验单尾 $p=0.037$ 。但它不成立，而这一条是本文自己去查出来的：K 的倒数单独与 g 的相关是 -0.393 ，比 ρ 更强。 ρ 那个负相关全部来自分母，不构成对本模型的任何独立支持。

决定性的那一条：控制 K 之后，N 的系数是 -0.046 ，置换检验 $p=0.449$ ，自助百分之九十五区间从 -1.280 到 $+0.692$ 。零结果。点估计的方向是对的，但与零无法区分。单变量上，N 与 g 的相关是 $+0.220$ ，是正的。

结论写死：9.1 那条把本模型与标准解释分开的预言，在目前唯一能跑它的数据上没有得到支持。

为什么没有功效——这一点本身是个结果。 四种方法的 N 几乎相同（0.79 到 0.85），而 K 相差一倍（0.36 对 0.66 与 0.72）。逐课看，K 的变异系数是 N 的二点二倍，而且 N 与 K 之间的相关是 $+0.473$ 。也就是说：在现实中，凡是提高提问密度的做法，同时也提高了小组时间。没有人把这两样分开过，因为没有人有理由把它们分开。本模型所要的那个对照——容量不变而抵达增加——在野外不存在。

这给出一个明确的后果，也是本文在此处唯一诚实的位置：这条预言不能靠回看既有数据来判，那个实验必须被造出来。设计现成且便宜：同一门课的两组，小组时间与教师巡视时间严格匹配到分钟，只改一件事——在那段时间里抛出的未闭合问题的个数（例如三个对九个）。标准解释预言两组无差、或问得多的那组略好；本模型预言问得多的那组延迟迁移更差。若这个实验做出来仍无负效应，第 9.1 节的容量模型应当被放弃，而不是被解释。

最后交出一条对本文不利、而这份数据给得很干净的观察。该研究还发现：四种方法之间学生同伴网络的发育没有差别，差别在课堂时间如何分配。而本文第九节把同伴说成承接的纯粹形式。如果同伴关系的网络结构与成效无关，那么“谁在承接”这个问题的答案可能比本文说的更简单——要紧的也许只是有没有那一段时间，而不是那一段时间里站着谁。本文没有能力在这份数据上把这两种读法分开，把它记在这里。

十、三条路线为什么各自到不了这里

一个领域看不见什么，通常不是它的缺陷，是它的问题结构的影子。三条路线各自不是因为疏忽才停在原地。

取向说到不了，因为它的测量单位比这件事大。它的问题是“什么使一个学生倾向于深层取向”，因此它的全部读数建立在“学习者”这个持续存在的单位上——一次访谈、一份问卷、一个学期的成绩轨迹。而临界回落是一个事件级的对象，它的时间尺度是十几秒到几分钟，它落在这条路线所有测量单位的下面。要它看见这件事，等于要它换一个问题。

路径说到不了，因为在它的体系里，折返可以重试。化学处理的是每毫升 10^{21} 量级的分子。一个折返的分子会再爬一次，再一次，直到它翻过去。所以折返的比例只影响速率，不影响最终能不能到达。在这样的体系里，把折返降级为一个修正因子是完全正确的处理。它看不见的是那一类抵达机会稀缺、且每次抵达都改变体系本身的场合——而这恰好是课堂。

计量说到不了，因为它的对象是工作项，而这件事是中间态。它的问题是“如何使实际活动与目标对齐”，因此它的账本上只有已完成与未完成两种状态。“抵达而未被承接”在工单系统里没有状态码，在预算表里没有科目，在绩效表里没有行。它不是被记成了零，是没有可以记零的那一行。更关键的是第五节说过的那一条：它不排队。而整套管理工具的探测原理是看堆积。一种不排队的损失，对以看堆积为原理的工具是不可见的。

三处“到不了”有一个共同的形状：不是他们没看，是他们各自的问题使得看这件事成为不必要的。

十一、与最近的几种说法的分界

这一节决定本文是不是一个既有概念的改名。判定标准是硬的：**能不能说出一个具体的观察，看到它就说明本文对而那个说法错，看到另一个就说明反过来。**说不出这样的观察，就等于承认被那个说法覆盖。“本文更强调”“本文更深入”“层次更根本”“视角不同”这类措辞不构成分界，它们不可判定。

11.1 承接（尼斯特兰德与加莫兰，教育学，1997）

这是最近的一位。尼斯特兰德在对话式教学的研究中提出，教师是否把学生的回答编入下一个问题，是课堂对话性的核心指标。他称之为承接，并证明高承接的课堂学习成效更好。

它说到哪一步：承接是教师的一种话语行为，测的是教师是否让学生的贡献影响了后续的教学进程。

分离线：尼斯特兰德的承接是**教师层面的话语纳入**，本文的承接是在**临界窗口内阻止关闭性归因**。三处不同：承接者不必是教师（第九节场景五，可以是他自己）；承接不必是话语（一次回读、一次让他动手试的沉默，都是承接）；最关键的是，**一次符合尼斯特兰德定义的承接，可以同时是一次关闭动作**——教师把学生的回答纳入之后立刻给出评价性收口，这在编码上计为高承接，在本文的读数上计为回落。

判定形式：在同一批课堂中同时编码尼斯特兰德式承接率与评价性收口率。若观察到高承接且高收口的课堂，其延迟测验成绩不高于低承接课堂，则本文对而该概念错；若高承接组无论收口率高低都优于低承接组，则反之。

必须连带点名的一处：尼斯特兰德的承接概念出自课堂话语分析的传统，而那个传统里早有一个词占着本文所说的评价性收口这个位置——发起一回应一评价这个三步结构里的第三步（米汉 1979）。这一步被批评了四十多年，批评的内容与第九节第一条几乎重合：教师给出评价，学生的思考随之终止。**本文与它的分界只有一处，但这一处是全部：**那个传统把第三步当作**教师的一种话步选择**来批评，因此它的处方是换一种话步——把评价换成复述、追问、或把发言权转给别人；本文把它当作**关闭性归因的触发条件**来处理，因此本文预测**换话步不够**。**判定形式：**把评价性收口换成非评价的复述或追问，而仍在窗口内给出结论。若关闭率不下降，则本文对而换话步的处方错；若关闭率随话步类型而降、与是否给出结论无关，则反之。

11.2 等待时间（罗，教育学，1974/1986）

罗发现教师提问后的平均等待时间不足一秒，把它延长到三秒左右，学生回答的长度、推理性回答的比例、学生之间的相互回应都显著上升。这是教育学中最经济、最常被引用的干预之一。

它说到哪一步：思考是一个可以被时间延长的连续过程，给更多时间就有更多思考。

分离线：等待时间是一个关于**时长**的量，本文的承接窗口是一个关于**结构**的量。二者可以背离：把等待时间从一秒延到五秒而不改变谁来收口，结果是更多学生进入了未完成推断的状态，然后仍然由教师收口——**抵达数上升，承接结构不变，回落数按比例上升。**

判定形式：若观察到延长等待时间之后，未完成推断数上升而未被接续比例不变或上升，且延迟测验成绩未同比例改善，则本文对而等待时间的剂量—反应逻辑错；若延迟成绩随等待时长单调改善且与接续比例无关，则反之。

11.3 有效失败（卡普尔，教育学，2008/2016）

卡普尔的设计是：先让学生在没有任何教学的情况下尝试解决超出他们能力的问题，失败，然后再教。这样比先教后练效果更好。

它说到哪一步：失败本身具有生产性，探索阶段的生成活动是效果的来源。

分离线：卡普尔的设计里，**巩固阶段是必要组成部分**，但在他的理论中它被处理为“把探索所得整理起来”，即一个收尾环节。本文认为它是承重的，而且它承重的方式与卡普尔说的不同——它的作用不是整理，是在探索制造的大批不可自持状态集体塌回去之前接住它们。

判定形式：若观察到去掉巩固阶段的有效失败设计，其效果不是减半而是低于直接教学组，且劣势幅度随探索时长增加，则本文对而该理论错——因为它把一批学生批量送上临界点然后集体回落；若去掉巩固后效果仍优于直接教学，则反之。这一条在现有文献中已有部分数据可供检索。

11.4 参与模式框架（齐与怀利，教育学，2014）

该框架按学生的外显动作把学习活动分为四级：互动优于建构，建构优于主动，主动优于被动。它的力量在于给了教学设计一个可操作的排序。

它说到哪一步：学习成效由参与模式决定，而参与模式可以从外部可见的动作读出来。

分离线：本文认为**同一个动作可以是抵达，也可以是空承接**。一个学生在小组里说了三句话，按该框架计为互动级；他可能正处在一个未完成的推断里，也可能在复述已经确定的东西。二者的外显动作相同。

判定形式：取同为建构级的两次活动，一次在生成后立即给出标准答案，一次在生成后设一个同伴质询窗口。若该框架预测两次学习量相同而实际第二次显著更高、第一次接近零，则本文对；若两次相同，则反之。

11.5 表现与学习的分离、合意困难（比约克；索德斯特罗姆与比约克，心理学，1994/2015）

这一脉给出的是当代学习科学中最重要的一条区分：当下的表现不等于长期的学习，某些会降低当下表现的困难反而提升长期保持。

它说到哪一步：同一个学习者的两个读数不一致——学习是潜在的，表现是当下的。

分离线：这一脉说的是**一个人的两个时间点**，本文说的是**两个人的同一笔账**。抵达并回落的学生与从未抵达的学生，在当下表现和长期保持上都可能一致，而他们经历的不是同一件事。更关键的一处：这一脉隐含地假定每一次学习经历留下的痕迹非负——差别只在于它何时显现。本文主张伴随关闭性归因的回落留下的痕迹为负。

判定形式：若观察到未被承接且伴随关闭性归因的抵达者，在延迟测验上不高于从未抵达者，甚至更低，则本文对而痕迹非负的假定错；若他们凭借生成效应仍高于从未抵达者，则反之。**这是本文最重要的一条可判定分离线，也是最可能推翻本文的一条。**

11.6 认知负荷理论（斯韦勒，教育心理学，1988）

该理论主张工作记忆容量有限，教学设计应降低外在负荷、管理内在负荷，困惑是需要被减少的东西。

它说到哪一步：负荷是一个标量，越低越有利于图式建构（专长逆转效应给出了对高水平者的例外）。

分离线：本文认为同样水平的负荷可以是抵达，也可以是淹没，二者的区别不在量而在是否存在承接。

判定形式：若观察到把内在负荷降到学生从不进入不可自持状态的水平时，主观指标全部最优而延迟迁移测验接近零，则本文对而“负荷越低越有利”错；若延迟迁移随负荷降低单调改善，则反之。这一格正是本文的格III。

11.7 支架与最近发展区（伍德、布鲁纳与罗斯，1976；维果茨基）

支架理论主张成人提供恰好使儿童能完成任务的支撑，并随能力增长而撤除。

它说到哪一步：支撑应当贴合，撤除应当渐进。

分离线：本文认为在临界窗口内提供内容性支撑（提示、部分答案）与提供结构性支撑（要求他把犹豫说出来、保持沉默、把疑问交给同伴）是方向相反的两件事。前者提高当堂表现、加速关闭；后者降低当堂表现、延长窗口。支架理论把二者合为一类。

判定形式：若观察到内容性支撑组的当堂表现最高而两周后保持最低、结构性支撑组反之，则本文对而“支撑越贴合越好”错；若两组的当堂与延迟成绩同向，则反之。

11.8 目标置换（默顿，1940；塞尔兹尼克，1949，组织社会学）

把本文的判断剥掉全部教育学名词，只留下形式：一个系统提高了进入不稳定态的频率，而承接容量未变，于是未完成的转化按比例上升，且未完成与未尝试在账上同形。剥到这一层之后必须去问：这个形式在上游学科里叫什么。

最接近的两个是目标置换与手段一目的脱耦。目标置换说的是手段吞噬目的——为维持程序而把程序要服务的目的架空。这是一个有八十多年历史的命名，任何主张“越规范越失效”的判断都必须先与它划清。

分离线：目标置换的机制是手段被当成了目的，其观察签名是组织成员坚持程序而无视结果。本文的机制不是替换而是同形记账：所有人仍然以结果为目的，程序也确实指向结果，问题在于两种结果无法被区分。在目标置换里，人们知道自己在维持程序；在临界回落里，人们不知道有一样东西正在流失。

判定形式：若观察到教师完全以学生理解为目标、并主动违反程序去追求它，而回落率不下降，则本文对而目标置换的解释不适用；若回落率随“以程序为目的”的程度单调变化，则反之。

11.9 成功陷阱与能力刚性（列文塔尔与马奇，1993；莱昂纳德-巴顿，1992）

凡是剥出来接近“越成功越失败”这一模板的判断，都必须点名这个模板早已有的名字。成功陷阱说的是：组织在利用中获得的即时回报挤出了探索，最终锁死在旧能力里。

分离线：成功陷阱预测**探索活动减少**。本文预测的恰恰相反——**探索活动增加**（课堂比以前活跃、活动比以前多、生成比以前频繁），而完成比例下降。两条预测在同一个观察上分岔。

判定形式：若观察到教学改良之后，学生的生成性活动频次上升而完成比例下降，则本文对而成功陷阱不适用；若生成频次本身下降，则本文的解释在该案例上失效。

11.10 多任务代理与计量固化（霍姆斯特罗姆与米尔格罗姆，1991；埃斯佩兰与索德，2007；比万与胡德，2006）

这一脉解释了大量教育系统的病理：可测的挤出不可测的，被排名者重造自己。

分离线：这一脉的机制要求存在一个**被计量的对象**，努力从不可测的那一侧被抽到可测的那一侧。本文的对象**根本没有被计量，也没有可以被抽去的对侧**——临界回落不是被挤走的努力，它是已经付出的努力在一个无人记账的位置上消失。**它甚至不需要有一个竞争性的指标存在。**

判定形式：若观察到一个完全没有外部评价压力的学习场景（自选课程、无学分、无排名）中回落率同样高，则本文对而该机制不适用；若回落率随外部计量压力单调变化，则反之。这一条是低成本可实做的，第十五节会再提。

11.11 防御性推理与单环学习（阿吉里斯，管理学，1991）

阿吉里斯观察到，越是聪明的专业人士，在面对自己的失败时越倾向于防御性推理，因而越不会双环学习。这是关闭性归因最近的一位邻居。

分离线：阿吉里斯说的是一种**跨情境稳定的倾向**，接近人格或组织文化。本文说的是一个**窗口内的事件**：同一个人在被承接时不产生关闭性归因，在未被承接时产生。

判定形式：若观察到同一个人在同一天两节课中关闭性归因的出现与否随承接结构变化、跨情境相关系数低，则本文对而“稳定倾向”的读法错；若跨情境高度一致，则反之。

11.12 两位方法学占位者

内容上的邻居容易找，方法上的邻居常被整批漏掉。凡是带着一个操作的判断，都必须问这个操作本身有没有一个成熟的现成范式。本文带的操作是“逐条清点未完成而未被接续的事件”，与之最接近的两个成熟范式是：

消融范式（机器学习）：逐项移除一个成分，看结果掉多少。若本文的回落率读数与“移除承接动作后的成绩下降幅度”高度相关，那么本文的读数就应当被消融设计取代——它更便宜。**分离线：**消融测的是移除某成分后的净效应，它无法分辨“没有承接因此回落”与“根本没有抵达因此无所谓承接”。**判定形**

式：若在抵达率极低的班级中移除承接动作，成绩不下降，而消融设计据此判定承接无效，则本文对——因为本文预测的是承接的效应量应当与抵达率成正比，而不是一个常数。

传递链范式（巴特利特，1932；梅苏迪与惠滕，2008）：把材料依次交给下一个人，看还剩什么。它与本文的相似之处是都在追踪“某样东西在传递中的流失”。**分离线**：传递链追踪的是**内容的流失**，本文追踪的是**未完成状态的流失**，后者从来没有进入过内容。**判定形式**：若在一条传递链中，每一环的复述内容完整而每一环的未完成推断全部消失，则两者测的是不同的东西。

11.13 僵局驱动的学习（范莱恩等，认知科学与智能教学系统，2003）

这是全部之中最近的一位，比尼斯特兰德更近。

它说到哪一步：范莱恩等人分析了约一百二十五小时的物理学专家辅导对话，逐段比较学习确实产生的片段与没有产生的片段，得到一个极强的结论：**学习几乎只发生在学生已经抵达僵局的片段里；学生不在僵局上时，无论辅导者给出何种解释，学习都罕见。** 这条结论与第五节那条判断在方向上完全一致，而且它有数据，本文没有。

分离线：范莱恩自己给僵局的说明（他标明这是非操作性的）是：**学生意识到自己在某一处缺乏完整的理解。**这个“意识到”，是本文与它的全部差别所在。

- 僵局要求学习者**察觉并停住**。本文的抵达不要求。第一节那个学生正在往前想，他没有停，他没有卡住的感觉，他甚至可能觉得自己推进得很顺利——“是不是因为……那如果反过来……”是一个向前的句子，不是一个停下的句子。
- 因此**本文的靶格恰好是僵局框架结构性排除的那一类**：没有僵局的外观，所以没有人来救；而没有僵局的外观，正是它至今不被记账的原因。一个体系若只在有求助信号时才响应，它的响应能力再强，也覆盖不到不发信号的那一类。
- 反过来，一个已经抵达僵局并明确说出“我卡住了”的学生，在本文的账上属于**已被外化**的状态——第七节场景三说过，把维持不住的状态说出来本身就是一次承接。**僵局是本文靶格的对立面，不是它的同义词。**

判定形式：在同一批课堂转写中同时编码两类事件——僵局事件（明确表示不理解、停顿超过阈值、或提出求助）与本文的抵达事件（未完成推断，且无停顿、无求助、无可见错误）。若在**无僵局签名**的抵达事件上，承接密度与两周后保持量仍呈剂量一反应关系，则本文覆盖了僵局框架结构性覆盖不到的一格；若两类事件高度重合，则本文可被它吸收。

这里必须补一句对本文不利的话。 范莱恩那条结论如果成立，它对第九节既是支持也是威胁。支持在于辅导者的解释在学生不处于关键状态时无效；威胁在于，若学习只发生在被意识到的僵局上，那么本文这一整类无声的、学习者自己都未必察觉的抵达，可能根本不产生学习，因而它的回落也就不构成损失。**这条威胁本文没有数据可以排除**，它被写进第十五节，并且是那份清单里最致命的一条。

11.14 有生产力的挣扎与教师回应的四种类型（沃绍尔，数学教育，2015）

它说到哪一步：沃绍尔逐条编码了中学数学课堂里一百八十六个学生挣扎片段，把学生的挣扎分为四类——开不了头、走不下去、说不出理由、说出了错的东西；把教师的回应也分为四类，排在一条从降低认知需求到维持认知需求的连续谱上：直接告知、导向式引导、探询式引导、留出余地。实测分布是：直接告知约百分之二十七，导向式引导约百分之三十五，探询式引导约百分之二十八，留出余地约百分之十一——**前两类合计约百分之六十二，都落在不支持挣扎的那一侧。**

这一位在方法上离本文最近：它已经在做第八节想做的事——逐条清点、按回应类型分类、给出比例，而且它的编码体系比本文成熟得多。**本文若只是把这件事重做一遍，就应当被它取代。**

分离线：沃绍尔的四类学生挣扎，**每一类都有外显签名**——开不了头是可见的停滞，走不下去是可见的中断，说不出理由与说出错的东西都是可听见的表述。这不是她的疏忽，这是编码得以进行的前提：**没有签名就没有可编码的片段。**而本文的靶格恰恰定义为无签名的那一类：学生正在往前说，句法未闭合，没有错，没有停，没有求助。**那一百八十六个片段里，本文的靶格一个也进不去。**

第二条分离线在因变量：她按**教师回应的类型**编码，本文按**是否发生了关闭性归因**编码。二者可以背离。第七节场景三那种“你先说说你在犹豫什么”，在她的谱系里属于探询式引导，在本文的读数上是承接；而一次探询式引导若以评价性收口结束——“对，很好，我们看下一步”——在她的编码上仍是支持性回应，在本文的读数上是一次关闭。

判定形式：在同一批转写上同时跑两套编码。若关闭率与她那四型的分布之间相关系数稳定高于 0.6，则本文的读数是她那套分类的一个代理，应当被并入而不是另立；若在**四型分布相同**的两组课堂中关闭率显著不同，且延迟保持随关闭率变化而不随四型分布变化，则本文对而那套分类不够。**这一条所需的原始材料在她的公开研究里已经存在，可直接做重分析，不必新采数据。**

11.15 习得性无助的归因重构（阿布拉姆森、塞利格曼与蒂斯代尔，心理学，1978）

第九节那三种关闭性归因的措辞，必须与这一位划清，因为其中至少一种几乎是它的字面形式。

它说到哪一步：归因重构把一次失败之后的解释拆成三个维度——内部还是外部、稳定还是不稳定、笼统还是特定。把失败归为内部的、稳定的、笼统的原因的人，会在此后的类似情境里不再尝试。这是心理学过去半个世纪最扎实的结论之一。

分离线：这一位测的是**跨情境稳定的解释风格**，接近人格；本文说的是**窗口内的单次事件**。这条分离线与 11.11 对阿吉里斯用的是同一条，判定方式也相同：同一个人在同一天的两节课中，关闭性归因的出现与否随承接结构变化，跨情境相关系数低。

但必须承认一处不对称。三种措辞里，**关掉自己那一种确实可被归因重构一比一覆盖，本文在这一条上没有增量。**本文的增量在另两种，尤其是**关掉对象那一种**。归因理论处理的是对**失败原因**的解释，而“这个东西没意义”不是对失败的解释——它是对**对象价值**的重新裁定，可以在学习者完全不认为自己失败的情况下作出。作出之后，他不会表现出任何无助的行为签名，他只是不再走这条路，而且他有一套听起来完全合理的理由。**本文认为这一种在课堂里最常见，也最不可能被现行的任何量表捕捉到。**

判定形式：若在未被承接的抵达者中，三种措辞对延迟成绩的区分作用完全一致，则本文的三分是多余的，应退回因重构的单一维度；若关掉对象那一种的负效应显著大于关掉自己那一种，且它的持有者在无助量表上得分正常，则本文对而归因重构漏掉了一类。

11.16 同一栏目此前六篇的划界

本文与本栏此前六篇同处一个问题域，用的是同样三个学科。前五篇可以用一条判据一次划清，第六篇要单独处理。这是本文最需要自查的地方，因此把五篇一并摆出来，用一条判据一次划清，再给出双向判决。

那五篇各自的承重读数是：**首次选择的可逆性**（一次已经做出的选择能否被撤回）、**结转率**（上一轮结束时留下的未完成处有没有被下一轮接上）、**节拍比**（验收间隔与更换一次做法所需时长之比）、**溯改率**（本轮的实际产物能否改写下一轮的判据）、**解绑代价**（同一份知识换一个条件场后重新达到同等水平的重做量）。

一条判据把它们与本文一次分开：**那五个读数的分母，全部要求被读的那样东西活过本轮的结束。** 一次选择要已经做出，才谈得上可否撤回；一个未完成处要在轮末还在，才谈得上有没有被结转；一次改变要已经发生，才谈得上节拍是否压掉了它；一份产物要已经产出，才谈得上能否用它改判据；一份知识要已经绑定，才谈得上解绑的代价。

而本文的对象在本轮内部就终止了。 一个未完成的推断在三十秒内塌回原处，到这节课结束时它已经不存在，它没有留下任何可以被撤回、被结转、被压掉、被用来改判据或被解绑的东西。**因此它不进入上述五个分母中的任何一个。** 这条分离线在分母那一层，不在措辞那一层：六个读数的分母两两相交，唯独本文这一个与其余五个的交集为空。

双向判决形式：若在同一批班级中观察到那五个读数中的三项以上同时改善，而课堂转写显示未被接续的比例不变或上升，且延迟迁移未改善，则本文测的是那五篇测不到的东西；反过来，若本文的回落率与那五个读数中任何一个的相关系数稳定高于 0.6，则本文是它的一个代理，应当并入那一篇而不是另立。

互补关系，以及本文在这一串里的确切位置。 把六个读数按学习者手里那样东西的去向排一次序，会得到一条链：**抵达**（它有没有在那几十秒里活下来）→ **首次选择的可逆性**（活下来之后做出的第一个选择能不能改）→ **轮末的结转**（本轮没走完的部分能不能进下一轮）→ **验收的节拍**（下一轮的间隔容不容得下它）→ **判据的溯改**（它能不能反过来改写标准）→ **跨条件场的重绑**（它能不能换个场合还立得住）。

此前五篇覆盖的是这条链的第二到第六个位置。**它们全部预设了链首已经有东西抵达。** 本文处理的是链首本身：一次抵达能不能存活到被处置。由此得到一句六篇合起来才说得出的话，它也是本文对这一串的实际贡献——**如果链首为零，后面五个读数全部做到最好也没有输入。** 可撤回的选择、被结转的中间体、对得上的节拍、开放的改判通道、跨场的重绑，处理的都是已经产生出来的东西；而一个在三十秒内被关闭掉的推断，从来没有到达它们中的任何一个。

第六篇不在这条链上，它记的是另一本账。 本栏最新的一篇处理的是**局部处置能力对差异信号的清除**：一个位置就地解决问题的能力越强，同一个差异到达下一位置的概率越低，而这两件事由同一个动作完

成，无法分开调节；它的读数是差异的**透过率**与系统一切正常的**诱导期长度**，并且明写不看学生内部，只从教案版本与题库改动记录算出。

两篇最容易被看成一篇，因为它们指着同一个可见动作——教师在学生说到一半时接过去讲完。三处把两本账分开：

- **分母不同。** 那一篇的分母是**差异信号**：某处出了错、有人掉了队、有一个可被记录的不对劲。本文的分母是**未完成的推断**：没有错，没有卡住，没有求助，学生正在往前说。**没有差异，就没有东西可供清除。**这与 11.13、11.14 用的是同一条线：本文的靶格定义为无签名的那一类，而无签名的东西进不了任何一个以“有东西可被清除”为前提的分母。
- **丢的是谁的东西不同。** 那一篇记的是**系统没有收到信号**——下一轮与本轮无关，制度学不到东西。本文记的是**学习者自己关掉了一条路**——损失落在他身上，且下一次遇到同一类不对劲时他不再离开原态。一个是信息流的账，一个是学习者状态的账。
- **两个读数可以同时走向相反方向，这是判决点。** 一间教室的透过率可以很高而关闭率同时为一：教师把这个困难记了下来、改了下一道题，差异确实传到了下一位置，而他是在给出结论的同时记下来的，关闭同时发生。反过来，透过率可以为零而关闭率也为零：一个同样不确定的同伴在窗口里陪着他，没有任何东西需要向上传递，也没有任何东西被关闭。

双向判决形式：若透过率与关闭率的相关系数稳定高于 0.6，则两篇测的是同一件事，本文应当并入；若观察到高透过率而高关闭率的课堂——教师记录充分、改进及时、延迟迁移仍差——则两本账必须分开记。

还有一处必须诚实说明。那一篇的三家互否轴是“处置能力该不该加强”，本文的是“那几十秒该由谁动”，**两条轴不同，因而两篇不是同一次碰撞的两种写法**；但两篇都得到了“越把课上好，这一格越大”这一形状的结论。**这种形状上的重合本身值得警惕**——它可能说明两篇各自摸到了同一样更深的东西的两个侧面，也可能只说明这个形状太容易被写出来。本文没有能力判断是哪一种，把这一句留在这里，不写进注释。

11.17 最近的那一位

在以上全部之中，最近的不是尼斯特兰德，是范莱恩（11.13）。

尼斯特兰德与本文共用一个词、一个观察位置、一部分编码方法，但两边会对同一份材料给出相反评价（11.1），分界是硬的。范莱恩不一样：他与本文共用**同一条实质命题**——学习只发生在学习者已经进入那个特殊状态的片段里，此外的教学动作近乎无效。本文与他之间只隔一个词：**意识到**。

这一个词是否足以支撑一篇独立的论文，本文认为是，理由写在 11.13 的第二条：不被意识到、因而不产生任何求助信号的抵达，恰恰是不会有人来接的那一类，也因此是全部损失的所在。但**这个理由是论证，不是数据**。若那条最便宜的检验做出来，无僵局签名的抵达事件与从未抵达者在延迟测验上无差异，本文

就应当被并入范莱恩，而本文此前的全部划界都救不了它。**这是本文最脆弱的一处，写在正文里，不写在注释里。**

十二、划完之后：一块所有人都站着的地基

把上面十七位邻居逐一问一句“它把什么当作给定，因此结构性地看不见什么”，会出现一件值得注意的事。

- 等待时间把“思考是可延长的连续过程”当作给定，因此看不见“存在一个有窗口而非有时长的不可自持状态”。一旦承认后者，它的剂量—反应逻辑就塌了。
- 有效失败把“失败是一个已完成的事件”当作给定，因此看不见“失败可以是未完成的”。一旦承认，它的探索阶段就不再同质。
- 参与模式框架把“参与可以从外显动作读出”当作给定，因此看不见“同一动作可以是抵达也可以是空转”。一旦承认，它的分级依据就失效。
- 表现与学习的分离把“每一次经历留下的痕迹非负”当作给定，因此看不见“关闭性回落留下负痕迹”。一旦承认，合意困难的“在范围内越难越好”就不成立。
- 多任务代理把“未被计量的努力仍然发生并留下价值”当作给定，因此看不见“未被计量的抵达会因未被承接而变为负值”。一旦承认，它关于“努力被转移”的整个记账方式就要重写。

这五行——以及其余各行——背后是不是同一样东西？是。

凡是发生过的事，都留下了非负的痕迹；差别只在于我们测不测得到。

本文称之为**痕迹非负假定**。它不是任何一派的主张，它是所有派别（包括互相批判的各派）共同站着的那块地。正因为大家都站在它上面，谁也没有回头看它。而它一旦被写出来，就会发现它并不显然：**一次未被承接的抵达，可能不只是没有留下痕迹，而是关掉了一条以后不会再被走的路。**

这里必须把话收窄一格，否则本文就说过头了。本文说过这一类事件在所有现行记录中与从未开始记为同一笔，这条仍然成立；但如果把它读成“这一类事件从来没有被任何学科辨识”，那就不实——11.13 与 11.14 已经把不实的地方指了出来：**范莱恩在事件层辨识过它，沃绍尔在事件层逐条编码过它的可见亲属。**准确的说法只有一句：

它没有被任何学科当作一个记账对象辨识。

被辨识与被记账是两件事。范莱恩把它当作**学习发生的条件**来研究——它出现，则解释有效；它不出现，则解释无效。在这个用法里它是一个筛子，一个决定别的东西有没有有效的条件；它自己的数量、它的流失、一堂课里它被浪费掉多少次，都不构成问题。沃绍尔把它当作**教师应当回应的情境**来编码——回应类

型是因变量，挣扎片段是背景。两处都不需要去问：这一堂课总共有多少次抵达，其中多少次没有活下来。

于是“为什么它至今没有一行”这个问题的答案是：不是没有人看，是**看的人都站在痕迹非负假定上面**。所有以改进测量为出路的研究纲领，其前提都是那件东西存在、只是测不到；而本文说的这一类事件，在这个读数被造出来之前不作为一个可被清点的量存在——不是难测，是没有这样一个对象。删掉这个读数，它不是变得难测，是不存在。

必须诚实地补一句：本文只能证明这五行的交集是同一样东西，不能证明这块地基是错的。**本文的主张是它是可疑的，而不是它已经被推翻**。推翻它需要的是第十一节 11.5 那条判定形式的实际结果。

12.1 强制不一致：合意困难那一支不能同时持有的两条

上一节说这块地基“可疑而未被推翻”。这是诚实的，但还可以再走一步：**不必推翻它，只要指出站在它上面的某一支已经不能同时持有它自己的两条承诺**。

被点名的是 11.5 那一支——表现与学习的分离，以及由它发展出来的合意困难。选它不是因为它最弱，恰恰因为它最强，而且 11.5 已经写明它握着最可能推翻本文的那条判定。

它同时持有两条：

- **甲：合意困难是一条可以指导设计的原则。** 某些困难可以被**事先**识别为有益——间隔、交错、提取练习，这是被反复给出的标准清单，而清单之所以有用，正因为它不必等结果就能用。
- **乙：困难只有在学习者能够成功应对时才是合意的；超出的部分有害。** 这是该支自己给出的边界条件，不是外人加的。

两条要同真，需要一个**事先可用的判据**来回答“会不会超出”。该支给的答案是难度与学习者能力的匹配。而这个答案不够用：**同一难度、同一先备水平的两组学生，结果可以相反**。一旦承认这一点，只剩两条路。

退到“只有事后由结果判定”，那么“合意”就从一条指导设计的原则退化成一个事后贴的标签——成功了就叫合意，失败了就叫不合意。这取消的正是甲的全部实用价值，也就是这条原则被提出来的理由。

或者保留甲，承认判据不完整，那就必须补上第三个量。本文提供的正是它：**合意不是困难的属性，是困难与承接结构的联合属性**。同一个困难，落在有承接的窗口里是合意的，落在无承接的窗口里不是，而这与学习者的能力无关——第九节场景三那位住院医师，能力一分未变，变的只是有没有人要求他把犹豫说出来。

此处不一致已经在它现行的处方里发作。合意困难的教学建议一向同时给出两样东西：一份**可以事先照做的难度操作清单**，和一句**注意不要超出学习者能力**的现场提醒。两样并排放在同一份材料里，而**从来没有给出“如何事先知道会不会超出”的判据**。此处沉默不是疏忽——要给出这个判据，就得承认一个该支变量表上没有的量。

退路已经替它写好，因为它确实有一条。该支可以说：所谓“能力”，本来就指在具体情境中能否应对，因此承接结构早已被“情境”二字吸收，不需要另立新量。这条退路成立，代价也是明确的——“合意”从此不可事先判定，因为情境要到发生时才知道；而那正是甲被提出来时要避免的处境。本文接受这条退路是一个真的选项，并把它列进第十五节：若该支明确采纳这一版，并相应地把困难清单从设计指南降为事后描述，则本节的不一致解除，本文在这一处的增量归零。

十三、三处反过来加在本文身上的约束

一个判断如果只从别处取材料而从不被别处削弱，它多半没有真的进入那些领域。以下三处，每一处都削掉了本文原本打算主张的东西。

13.1 化学削掉了本文的适用范围

化学之所以正确地把折返降级为修正因子，是因为在它的体系里折返可以重试：分子数是天文量级，每个分子可以反复尝试，折返只影响速率不影响到达。

这迫使本文交出一句原本会写下的更强的话。如果没有这一处，本文原本会写：“回落率是一切学习系统的决定性读数。”现在只能写：

只有在抵达机会稀缺、且每一次抵达都改变体系本身的系统里，回落率才是决定性的。

课堂恰好是这样的系统：一个核心概念在一个学期里可能只出现两三次真正的抵达机会；而每一次未被承接的抵达都通过关闭性归因改变了下一次抵达的概率。但这是一个需要论证的经验主张，而不是一个可以假定的性质。在抵达机会密集的场所——比如一个人独自反复琢磨同一个问题、比如高频次的刻意练习——本文的读数应当迅速失去解释力。这是本文适用范围的边界，也是它的一条证伪条件。

13.2 管理学毁掉了本文最自然的处方

本文最自然的处方是：给回落率记账，把它做成一个指标。

管理学自己的证据直接毁掉了这条路。埃斯佩兰与索德关于反应性的证据说明：一个读数一旦被用于评价，被评价者会重组自己以适应它。布罗姆利与鲍威尔的手段一目的脱耦说明：评估越密，“真的照做而与目的失联”越普遍。把这两条合起来预测本文的处方：回落率一旦进入教师评价，最省事的达标方式是不再制造临界。教师会把课设计成学生从不进入不可自持状态——分母趋近于零，读数完美，而那正是格Ⅲ。

这削掉的不是一个补充条件，是本文的价值排序本身。本文原本会主张“这个读数应当被制度采用”，现在必须主张相反的东西：

这个读数必须由制造临界的人自己持有，且不得外报。一旦它成为外部评价的输入，它就自动失效，且失效的方式不产生任何信号。

这是一个不舒服的结论，因为它意味着本文提出的东西在结构上无法规模化。本文接受这个结论。

13.3 教育学削掉了本文关于负痕迹的强主张

第九节提到的生成效应证据——失败的提取尝试之后仍有正效应——迫使本文把“负痕迹”从一条普遍主张降为一条条件主张。

如果没有这一处，本文原本会写：“未被承接的抵达留下负痕迹。”现在只能写：

只有当回落伴随一次关闭性归因时，痕迹才为负；单纯的未完成生成留下的仍是正痕迹。

这一削弱的代价是本文的核心机制从“承接”缩到了“阻止关闭”，因而它更窄；它的收益是本文因此获得了第九节那一整段推论——承接不是帮助完成——而那一段是本文最有解释力的部分。这是一处削弱同时是一处收获，但削弱是真的：它意味着本文无法解释那些不伴随关闭性归因的回落，而这类回落数量上可能占多数。

十四、轮次：谁先动，谁跟上，谁反过来改了谁

第三节说过，三条路线共同假定发起位置固定；第四节用催化化学自己的证据说明它不固定。这条推翻如果只用来立一个构念，就浪费了它最值钱的一段——发起权换手是一个有时序结构的东西，因此本文的正向预测应当写成轮次，而不是一次先后。

一个完整的轮次有四步：这一轮先动的是谁 → 另两样多久跟上 → 有没有反向改写 → 下一轮的发起处是不是换了人。

第一轮（评价先动）。 考试方式或升学口径改变，属于条件场先动。学生做出来的样子在一到两个学期内跟上——取向随之改变，这一段有四十年的证据支持。教师的教法在两到三个学期后跟上。反向改写发生在这里：新的教法改变了课堂的时间结构（单位时间内的话轮密度上升、教师讲述时长下降），而这个时间结构反过来重设了下一轮的条件场。

第二轮（发起处换人：教法先动）。 这一轮的发起者不再是评价，而是教学路径本身——学校采用了小组探究、翻转课堂、生成式活动。抵达临界的次数在数周内上升，这是最快跟上的一样。而承接结构跟不上：一节课的长度没变，班额没变，教师人数没变，允许的沉默时长没变，评分表上仍然没有“未完成推断”这一栏。于是回落率上升。

这就是本文的核心正向预测，它可以被写成一句可查的话：

在推行生成式教学改革学校，“未完成推断的出现频次”的上升，应当早于承接结构的任何变化 2 到 6 个学期出现；且在这段时间里，未被接续比例应当上升而非下降。

如果观察到承接结构的改变与抵达频次同时发生，或早于它，本文的轮次结构为伪。

第三轮（发起处再换：学习者先动）。 当回落积累到一定程度，学生做出来的样子会自己先动——出现一种新的课堂形态：**熟练的参与**。学生学会了生产符合活动要求的可见动作（说出恰好足够的话、在小组里承担恰当角色、写出符合评分表的反思），而这些动作与是否抵达临界脱钩。这正是手段一目的脱耦在课堂尺度上的形式。反向改写发生在这里：教师据此调整活动难度与评分标准，条件场被重设。

顺序预测比相关预测难被巧合满足得多。以上三轮各自给出的是“哪一样先动、另两样多久跟上”，它们都可以在既有的学校档案里被回溯清点：教研记录有教法变更的时间戳，课表有课时长度，评分表有版本，课堂录像有日期。**这一条不需要新建任何采集系统。**

十五、证伪条件

一个不会被触发的条件等于没有。以下各条都写成可以被触发的形式。

先写出最强的竞争解释，不挑好打的。 对本文而言，最难反驳的那句“其实不就是……吗”是：

其实不就是练习量不够吗？ 学生没学会是因为一次不够；需要的是重复、间隔、提取练习。所谓“未被承接”，不过是“这次没练够”的另一种说法。

这个解释朴素、有大量证据、且能覆盖本文的大部分观察。本文的证伪条件必须能排除它。

主证伪条件（机制层）：

控制总练习量、间隔安排与初始知识水平后，若“临界窗口内的承接事件密度”与“两周后保持量”之间不存在剂量—反应关系，则本判断为伪——届时应把差异归因于练习总量而非承接结构。

这一条必须用 C，不能用 R。 第 8.1 节已经说明 R 的分子里混着两类符号相反的事件；用 R 跑这条检验，两类会在同一个系数里互相抵消，届时做不出结果也无法判断是机制错了还是分子脏了——一个分不清这两种失败的检验，不构成一条证伪条件。

样本设计（这一条是硬约束）： 要谈剂量—反应关系，样本至少六个单元以上，且在关键混淆变量上同质。**不得用两个国家或两个地区的对比充数**——那类比较在文化、规模、历史、语言上处处不同，所谓“控制掉某变量”在方法上站不住。合格的设计是：同一所学校、同一学科、同一份教案的十二个平行班，随机分配承接结构（生成后立即给标准答案 / 生成后设三十秒同伴质询窗口 / 生成后要求外化犹豫），课堂录像编码回落率，两周后延迟测验。**所需数据在多数学校的既有记录里已经存在，这一条低成本可实做。**

其余证伪条件：

- **编码可靠性。** 若两位受过训练的编码者对“该学生是否已抵达未完成推断”的一致性系数低于可接受水平，则本文的分母不可测，整个读数不成立。**这是最先应当做、也最可能失败的一条。**
- **拆分的必要性（对应 8.1）。** 若在多个班级中 R 与 C 的相关系数稳定高于 0.8，则两个数没有分开任何东西，本文应退回单读数版本并交出第九节的机制。
- **无签名的抵达是否产生学习（对应 11.13）。** 若那些没有任何僵局外观——不停顿、不求助、不出错——的抵达事件，在延迟测验上与从未抵达者无差异，则本文的靶格根本不产生学习，因而它的回落也就不构成损失，本文的全部主张随之归零。**这是整份清单里最致命的一条，也是最便宜的一条，应当排在编码可靠性之后立刻做。**
- **折线还是倒 U（对应 9.1 推论二）。** 若在同一门课程上逐级提高抵达率而延迟迁移呈光滑的倒 U、拐点不尖，则容量不是硬上限，9.1 的模型为伪。
- **容量折线（对应 9.1 与 9.2）。** 已在桑德斯特伦等人的公开数据上跑过一次：控制承接容量之后，抵达事件的偏效应为零（见 9.2）。**这一条尚未被支持。** 判决要靠一个必须新造的对照：小组时间与巡视时间匹配到分钟、只改抛出的未闭合问题个数。**若该实验仍无负效应，容量折线应当被放弃。** 至于那份元分析的重编码：不需新采集，但需回读约一百五十份原始文献的方法部分，本文一次只按方法名的试编码显示其中约六成三无法从方法名判定。
- **合意困难的退路（对应 12.1）。** 若该支文明确采纳“能力已含情境”那一版，并把困难清单从设计指南降为事后描述，则 12.1 的不一致解除，本文在那一处的增量归零。
- **正交性。** 若在多个班级中，回落率与当堂正确率、参与度评分之间出现稳定的强相关，则本读数是既有指标的代理，白造。
- **关闭性归因的中介作用。** 若在未被承接的抵达者中，关闭性归因的出现与否对延迟成绩没有区分作用，则第九节的机制为伪，本文退回一个更弱的版本。
- **抵达机会密度的调节作用（对应 13.1）。** 若在抵达机会密集的场所（高频刻意练习、自主反复琢磨），回落率仍具有同等解释力，则本文对自己适用范围的限定是错的——这对本文是坏消息还是好消息尚不确定，但它证伪了 13.1 那条约束。
- **外部计量压力的作用（对应 11.10）。** 若在完全无外部评价压力的学习场景中回落率显著低于有压力场景，则计量说足以解释本文的现象，本文的独立性不成立。
- **轮次结构（对应第十四节）。** 若承接结构的改变与抵达频次同时或更早发生，本文的轮次结构为伪。
- **负痕迹（对应 11.5）。** 若未被承接且伴随关闭性归因的抵达者，在延迟测验上稳定高于从未抵达者，则痕迹非负假定成立，本文第十二节的整段论证倒塌。

如果本文被证伪，我们会学到什么？ 若第 1 条失败（编码不可靠），学到的是这一类事件在原则上不可外部判定，那么正确的方向不是改进编码，而是把读数交给学习者自己持有——这指向一个完全不同的干预设计。若第 7 条失败（痕迹确实非负），学到的是回落只是慢，不是损失，那么正确的干预是提高抵达频次而完全不必管承接——这与本文的处方相反，而且更便宜。**两种证伪各自指向一个具体且不同的行动方向，这是本文认为它值得被检验的理由。**

两个本文解释不了的洞：

其一，本文完全没有处理**为什么有些学生在完全无人承接的情况下不产生关闭性归因**。这类学生存在，而且他们身上可能藏着本文全部处方的替代方案。本文对此没有机制。

其二，本文无法解释**承接窗口的宽度是否有个体差异，以及它是否可以被训练**。如果它可训练，本文关于“结构不可规模化”的悲观结论就要被修正；如果不可，本文的处方只能停留在教师的现场判断上。本文倾向于前者，但没有证据。

十六、本文可以被分层引用

本文的主张分三层，各层的稳固程度不同，可以被分开引用与分开推翻。写清这一点是诚实，也是保护。

第一层（最强，最易倒）： 临界回落是一类具有独立本体地位的事件，且现行记录把它与“从未开始”记为同一笔。倒在编码可靠性、无签名的抵达、痕迹非负这三条证伪中的任何一条上。

第二层（中）： 抵达 × 承接的两轴辨别格作为分析工具。它不依赖第一层的本体主张——即使临界回落不被承认为一类独立事件，这张格子至少把格Ⅲ（空承接）从格Ⅰ（转化完成）里分了出来，而这一分是独立成立的。

9.1 的容量折线不在这一层，它现在单独站着，而且站得比本文原先说的低。 9.2 已经报告：它那条区分性预言在唯一能跑的公开数据上是零结果，而且无功效的原因是野外根本不存在「容量不变而抵达增加」的对照。**在那个实验被造出来之前，容量折线只是一个有明确证伪条件的待检模型，不是本文的一项结论。** 引用它的人应当连同 9.2 一起引。即使第一层的本体主张被削弱，这张格子仍可作为课堂观察的分类框架使用——它至少把格Ⅲ（空承接）从格Ⅰ（转化完成）里分了出来，而这一分是独立成立的。

第三层（最稳，也最便宜）： 一条不依赖前两层的经验主张——**生成式教学改革提高了未完成推断的出现频次，而课时长度、班额、评分口径在同期未发生相应改变。** 这一条完全可以用学校档案回溯验证，不需要任何本文的概念。它若成立，本文的实践含义已经大半成立；它若不成立，本文的其余部分都失去落点。

十七、反噬：这个读数一旦进入考核

任何一个读数一旦被制度采用，最省事的实现方式会摧毁它本来要保护的东西。本文必须写出自己的那一条。

回落率一旦进入教师评价，最省事的达标法有两条，都不需要任何弄虚作假：

第一，**把分母做小**。不再制造临界。把每一个可能的困惑提前拆解掉，课上得流畅、精确、无卡顿。回落率的分子和分母同时趋近于零，读数完美。这正是格III，而格III的所有其他指标也是最好看的——它没有任何代价。

第二，**把“承接”的定义放宽**。在转写编码时，把教师的任何一次回应都计为承接。这不需要任何人说谎，只需要编码规则里的一句话稍微松一点。而编码规则通常由被评价方或其上级提供。

更根本的一处：**回落率的分母不可被外部审计**。“这个学生是否已经抵达”这一判定，需要对具体内容与具体学生有足够了解，而具备这个了解的人正是被评价的那个教师。任何外部审计都只能审到可见动作那一层，而本文全部的论证正是可见动作那一层不够。

我看不到出口。看得到出口的反噬不是反噬，是待办事项。本文能做的只有第 13.2 节那条：主张这个读数不得被用于评价，并承认这条主张在制度上几乎不可能被遵守。

十八、不适用的边界

本文不适用于以下几类场合，逐条写明理由：

其一，抵达机会密集且可重试的场合。独自反复琢磨同一个问题、高频次的刻意练习、大量同质题目的重复训练。理由见 13.1：折返可重试时，折返比例只影响速率不影响到达。

其二，转化本身不经过不可自持状态的学习。大量的事实性记忆、程序性动作的初期建立、纯粹的信息获取。这些学习不需要经过一个学习者自己维持不住的中间状态，因此第一根轴的读数恒为“未抵达”，整张格子退化为一条线。这类学习占学校教学的相当大比例，本文对它们无话可说。

其三，安全关键的训练场合，在承接结构未到位之前。见下一节。

其四，回落被承接但承接者是评价者本人的场合。当承接者同时握有评价权时，本文第九节的机制会与评价的关闭效应叠加，二者无法在观测上分离。这不是本文不适用，是本文在这类场合给不出可判定的读数。

十九、伦理与证据边界

本文提出的是一个比率，而比率几乎必然会被拿去考核人。因此以下三条是发表条件，与本文的其余主张无关：

第一，这个读数指的是位置，不是人。 回落率读的是某一段课时、某一个交接口的承接结构，不是某位教师的能力或态度。同一位教师在两种课时结构下会得到完全不同的读数。**任何把回落率写进教师个人评价表的做法，都违背本文提出它的全部理由。**

第二，不得为了把这个数做好看而在安全关键场合减少真实困惑的暴露。 在医学操作、实验安全、司法训练、机械操作这类场合，学习者必须在受控条件下真实地遭遇他维持不住的状态——因为不遭遇的代价会转移到将来的真实现场。在这些场合，**降低回落率的正确方式只能是增加承接容量（增加带教密度、延长窗口），绝不能是减少抵达。**本文明确禁止后一种实现方式。

第三，已经依据这个读数作出不利安排的机构，必须公开编码规则并提供复核通道。 编码规则是这个读数的全部要害——“什么算抵达”“什么算承接”这两个定义可以把同一堂课的读数从零改到一。不公开编码规则的读数不构成证据。

此外一条关于研究伦理：本文暗示的干预中包含“延迟给出答案”“在学习者困惑时保持沉默”。这类操作在成年学习者的自愿研究中可以实施；**在未成年学习者身上实施时，不得延长到学习者出现明显痛苦的程度，且必须在同一课时内给出闭合。** 本文不主张让任何学生带着未解决的困惑离开教室——第九节的机制是阻止关闭，不是制造悬置。

二十、本文自己落在哪一格

用本文自己的格子给本文定位，得到的答案不好看。

本文把三条研究路线各自送到了它们维持不住的位置上——每一条都被逼到承认自己长期掌握的某个事实指向了它自己不愿意去的方向。而本文**不提供任何承接结构**。它是一篇论文；读者读完之后产生的每一个半成品念头，没有任何机制去接。按本文自己的判据，**本文的读者回落率应当接近一。**

这个定位有具体后果，不是一句谦辞：

其一，本文最可能被引用的部分，是第六节那张两轴格子——因为它是可复述的。而本文的承重主张是“抵达与从未抵达在账上同形”，它不可复述，只能在一次具体的观察中被认出。**本文因此传不了它自己主张的那一格。** 这是可以预测的：本文若被引用超过一定次数，引用中提及格子的比例应当显著高于提及同形记账的比例。

其二，本文属于“提出”而不是“核验”。它从三个领域取材料，而它在这三个领域中的任何一个里都不是内行。这一点在写作过程中已经兑现过一次：本文初稿漏掉了与它最近的三位邻居（11.13 至 11.15），其中一位与本文只隔一个词。**一篇声称某件事从来没有被辨识过的论文，它自己没有把该领域**

看全，这是同一种病的两次发作。到本稿这一版，同样的事又发生了第二次：9.1 原本把那条重编码称作最便宜的检验，而查证之后发现该元分析早已做过一次类型分析、结果为零，且它自己写明了为什么那个零不可当真。两处都不是被人指出来的，是自己去查的时候撞上的；这既不值得表功，也说明一件事——凡是本文说“某处还没有人做过”的地方，都应当先去查，而不是先去写。第三次紧接着来了：去查之后真找到了能跑的数据，跑出来是零结果（9.2）。三次里有两次是本文说错，一次是本文预言落空；三次都写在正文里，没有一次挪进注释。这是本文能对自己做的全部，也是它唯一能示范的东西。第 11.12 节那两位方法学邻居随时可能把本文的读数收编——如果消融设计能以更低的成本给出相关的结果，本文的读数就应当被取代。本文对此没有防御，也不认为需要有。

其三，本文写作过程本身就落在格III的风险里：每一处可能的困难都被提前处理成了一个小节，读者一次也不必进入不可自持的状态。本文读上去越流畅，它越可能是自己所批评的那个东西。唯一能避免这一点的写法是在关键处留白不作解释，而那与学术论文的形式要求冲突。本文选择了形式要求，因此承担这个代价。

二十一、一条写死日期的赌注

到 2031 年 8 月 31 日之前，将至少出现一份经同行评议的课堂研究，在其变量表中报告一个以“抵达而未完成”为分子的比率（名称不限），并将它作为独立于参与度与正确率的结果变量。

另有一条只关于容量折线，期限更近：到 2030 年 12 月 31 日之前，将至少出现一项把小组时间严格匹配、只改抛出问题个数的对照实验并报告延迟迁移结果。若到期没有这样的实验，说明这条预言连一次像样的检验都没能招来，本文对它的主张应当自行降为猜想；若实验出现而未见负效应，按 9.2 的写法，容量折线应当被放弃。

回到主赌注。什么不算命中，写死三条：

- 只在讨论部分提及“有些学生想到一半没说出来”，不算。必须进变量表。
- 作为过程性描述报告，不算。必须作为结果变量，且与至少一个既有指标同时报告。
- 以“未回答问题数”“沉默时长”“教师等待时长”代替，不算。分子必须要求编码者作出一项判定：该学习者已经进入一个未完成的推断。不含这项判定的任何替代读数一律不算。

若到期未出现，本文所主张的“这一类事件具有独立本体地位”这条判断，在学术共同体层面就没有得到承认，本文应当被读作一次未被接住的抵达。这句话不是修辞：按本文自己的机制，那意味着一次关闭性归因将出现在本文的方向上，而下一个会更晚才走上这条路。

二十二、结语

把全文压成一句：

课堂上真正的学习很少发生，不是因为学生没有进入学习，而是因为他们每天进入又退回，而“进入过又退回”与“从未进入”在所有账本上是同一笔；正因为是同一笔，所有旨在提高进入率的改良都在同时提高退回率，而后者不出现在任何一份报告里。

由此得到的实践含义只有一条，而且它与几乎所有现行的教学建议相反：

不要问怎样让学生更多地进入。先问在他进入之后的那几十秒里，谁在那里，那个人做了什么。如果那个人做的事是给出结论，那么进入得越多，损失越大。

而承接的内容，不是答案。是别让他在那几十秒里对自己、对这个东西、或这条路作出一次判决。

注释

〔1〕本文第二节对三条研究路线的表述，一律按各自最强的排他形式给出，而非按“多因素共同作用”的温和形式。这是有意的：三条路线的解释力恰恰来自它们的排他性，弱化它们等于取消本文要处理的问题。

〔2〕第四节所引的两项化学工作，在其原领域内回答的都不是本文的问题。这是本文选择它们的理由，而不是本文的疏忽：一件材料之所以能推翻一条被普遍默认的前提，通常正是因为它在它自己的领域里它回答的是另一个问题，因而从未被当作关于该前提的证据来读。

〔3〕第八节给出的“抵达”操作定义（试探性连接词开头、句法未闭合、包含自我提出的疑问）是一个起点，不是一个经过验证的量表。它的编码者一致性未经检验，这是本文最先应当被检验、也最可能失败的一处。

〔4〕第十四节的轮次预测所需的全部数据，均可从学校既有档案中回溯获得：教研记录、课表版本、评分表版本、课堂录像。本文未作此项回溯。

〔5〕关于“承接”一词与尼斯特兰德同名概念的关系，见 11.1。本文保留这个词而非另造新词，是因为另造新词会掩盖二者在观察位置上的重叠；重叠必须被看见，才谈得上分开。

〔6〕本文交出两个读数而非一个（8.1）。这不是为了精确，是因为第十三节第三处约束逼出的一个后果：两类回落的痕迹符号相反，把它们数进同一个分子，得到的就不是任何一条机制所预测的量。凡是接受那处约束的读法，都必须做这一次拆分。

〔7〕11.13 至 11.15 三位邻居在本文初稿中缺席，是初稿最大的一处漏。补上之后，第十二节原有的一句“从来没有被任何学科辨识”不再成立，已按 11.13 与 11.14 的实际情况收窄为“没有被任何学科当作一个记账对象辨识”。这一处修改削弱了本文的措辞，没有削弱本文的主张：被辨识与被记账是两件事。

参考文献

Abramson, L. Y., Seligman, M. E. P., & Teasdale, J. D. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology*, 87(1), 49–74.

Argyris, C. (1991). Teaching smart people how to learn. *Harvard Business Review*, 69(3), 99–109.

- Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press.
- Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. *Public Administration*, 84(3), 517–538.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about Knowing* (pp. 185–205). MIT Press.
- Bromley, P., & Powell, W. W. (2012). From smoke and mirrors to walking the talk: Decoupling in the contemporary world. *Academy of Management Annals*, 6(1), 483–530.
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243.
- Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, 113(1), 1–40.
- Eyring, H. (1935). The activated complex in chemical reactions. *The Journal of Chemical Physics*, 3(2), 107–115.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415.
- Halpern, J. (1982). Mechanism and stereoselectivity of asymmetric hydrogenation. *Science*, 217(4558), 401–407.
- Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 371–404). Information Age.
- Holmström, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, 7(Special Issue), 24–52.
- Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26(3), 379–424.
- Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist*, 51(2), 289–299.
- Kerr, S. (1975). On the folly of rewarding A, while hoping for B. *Academy of Management Journal*, 18(4), 769–783.
- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(4), 989–998.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86.

- Leonard-Barton, D. (1992). Core capabilities and core rigidities: A paradox in managing new product development. *Strategic Management Journal*, 13(S1), 111–125.
- Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal*, 14(S2), 95–112.
- Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—Outcome and process. *British Journal of Educational Psychology*, 46(1), 4–11.
- Mehan, H. (1979). *Learning Lessons: Social Organization in the Classroom*. Harvard University Press.
- Merton, R. K. (1940). Bureaucratic structure and personality. *Social Forces*, 18(4), 560–568.
- Mesoudi, A., & Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society B*, 363(1509), 3489–3501.
- Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83(2), 340–363.
- Nystrand, M. (1997). *Opening Dialogue: Understanding the Dynamics of Language and Learning in the English Classroom*. Teachers College Press.
- Rowe, M. B. (1974). Wait-time and rewards as instructional variables, their influence on language, logic, and fate control: Part one—wait-time. *Journal of Research in Science Teaching*, 11(2), 81–94.
- Rowe, M. B. (1986). Wait time: Slowing down may be a way of speeding up. *Journal of Teacher Education*, 37(1), 43–50.
- Selznick, P. (1949). *TVA and the Grass Roots*. University of California Press.
- Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, 10(2), 176–199.
- Sundstrom, M., Gambrell, J., Green, C., Traxler, A. L., & Brewe, E. (2025). Relative benefits of different active learning methods to conceptual physics learning. arXiv:2505.04577. Data: github.com/msundstrom33/ComparingActiveLearningMethods
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
- Tao, F., Grass, M. E., Zhang, Y., Butcher, D. R., Renzas, J. R., Liu, Z., Chung, J. Y., Mun, B. S., Salmeron, M., & Somorjai, G. A. (2008). Reaction-driven restructuring of Rh-Pd and Pt-Pd core-shell nanoparticles. *Science*, 322(5903), 932–934.
- Truhlar, D. G., Garrett, B. C., & Klippenstein, S. J. (1996). Current status of transition-state theory. *The Journal of Physical Chemistry*, 100(31), 12771–12800.
- VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction*, 21(3), 209–249.

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

Warshauer, H. K. (2015). Productive struggle in middle school mathematics classrooms. *Journal of Mathematics Teacher Education*, 18(4), 375–400.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100.

Yeager, D. S., Hanselman, P., Walton, G. M., et al. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, 573, 364–369.

人机分工声明 本文的问题、三条解释路线的选定、以及跨领域取材的方向由作者确定；材料检索、论证结构的组织与初稿撰写由作者与语言模型协作完成；第四节所引两项化学工作的书目信息经在线核验；全部判断、证伪条件与伦理条款由作者负责。本文未使用任何自评分数，其学术价值应由独立评审判定。

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