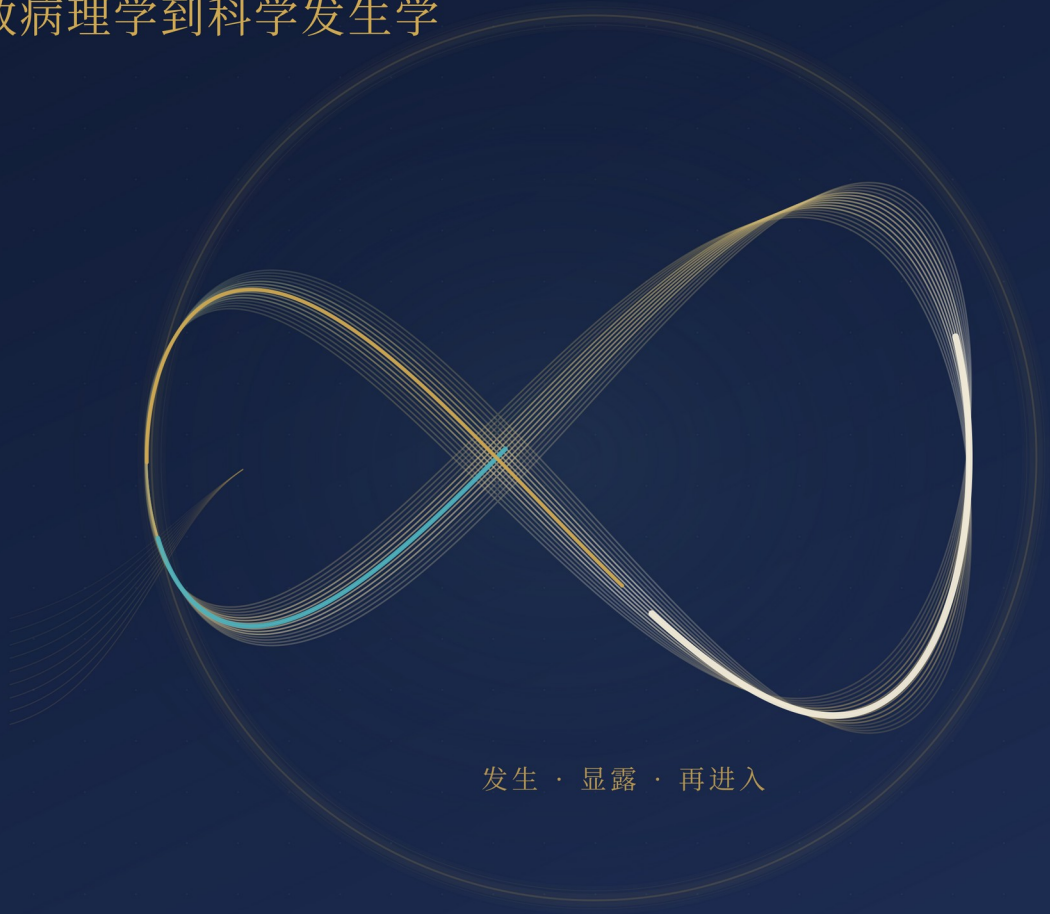
第379卷

属性纠缠与科研新典范

从不收敛病理学到科学发生学



王德生　著

德麦国际出版社
Demai International Press

属性纠缠与科研新典范

从不收敛病理学到科学发生学

王德生 著

德麦国际专著 第 379 卷

全书统稿版 v1.1

德麦国际出版社

Demai International Press

2026 年 10 月

本书研究属性的稳定性如何发生，测量怎样参与发生，以及读数、模型与制度如何改变下一轮发生条件。以 SDE 为研究框架，将三号位的属性侧重、科学测量、生命过程、经济制度和人机科研置于同一问题体系中展开论证。

全书八编四十章，另设导论与结语。概念论证、数学结果、已有经验研究与待检验方案分别呈现；文中构造例用于说明机制，研究方案不作为已经执行的实验结果。

全书八编四十章 · 正文约 32.3 万字

出版信息

属性纠缠与科研新典范

从不收敛病理学到科学发生学

王德生 著

出版 德麦国际出版社 Demai International Press

德麦国际专著 第 379 卷

ISBN 979-8-90690-485-0

定价 US$35

2026 年 10 月 全书统稿版 v1.1

开本 190 mm × 250 mm

阅读 学习包 智能体

https://sdeuniverses.com/books/m/379/

本书正文、章节学习任务与智能体研生共同组成出版单元。学习任务为配套教学设计；模型建议须由读者核对原文与证据。

阅读说明

本书从属性的归属问题进入，先建立发生分析的概念与方法，再以物理学检验解释边界，以生命和社会过程展开回写机制，最终讨论 AI 与科研制度怎样组织新的知识发生。读者可顺序阅读，也可由目录进入具体领域，再返回第一编核对共同概念。

S 表示显露，D 表示差异序列，E 表示特征纠缠；客体、互动与主体表示属性的号位侧重。两套区分承担不同任务。三方程、六路径与三原理共同组织研究，不以三个名称替代具体机制。本书的特征纠缠属于发生分析概念，物理学的量子纠缠另有严格的态结构定义，两者在具体问题中的联系需要另行论证。

正文中的方括号数字指向书末参考资料。各领域保留自己的对象、方法与证据标准，跨领域比较围绕属性怎样取得稳定性、稳定何时失配以及发生条件怎样改变展开。概念解释、数学条件结论、经验材料、待检验假说与规范理由各自承担相应的论证责任。

一些问题会在不同编章中再次出现，所承担的工作随之改变。一杯甜饮料首先帮助划定发生单位，随后成为读数记录、感知过程与区分实验的不同研究对象；一次评价也从能力的证据推进到机会的分配与修复。共同问题把全书连在一起，具体机制和证明方式则由各章展开。

目录

导论 从属性之问到科研新典范.....1

第一编 属性发生的本体论基础

第 1 章 属性之问与发生单位.....13
第 2 章 三号位的稳定性侧重发生.....22
第 3 章 三方程六路径三原理的联合方法.....31
第 4 章 混沌 分化与成熟.....40
第 5 章 稳定 退化与再生的机制.....49

第二编 物理学中的属性显露

第 6 章 经典物理的客体化条件.....58
第 7 章 相对论中的测量路径与不变量.....67
第 8 章 量子测量中的装置与结果.....76
第 9 章 量子场论与客体身份的形成.....85
第 10 章 物理学史的发生学重读.....94

第三编 测量 读数事件与不收敛

第 11 章 从属性值到读数事件.....103
第 12 章 测量装置与判断的封装.....112
第 13 章 不收敛的分类与数学边界.....121
第 14 章 回写机制与因果辨识.....131
第 15 章 发生模型的比较与迁移.....140

第四编 生命 神经与感知的发生

第 16 章 味觉中的三属性发生.....149
第 17 章 细胞状态与历史调控.....158
第 18 章 神经活动 感受野与学习.....167
第 19 章 生命稳态 变异与恢复.....176
第 20 章 多尺度回写与生命研究设计.....185

第五编 经济指标与制度回写

第 21 章 经济参数的条件性.....194
第 22 章 市场价格 预期与交易事件.....203

第 23 章 股市信息与公开预测的回写.....212
第 24 章 GDP 的核算边界与制度作用.....221
第 25 章 经济制度如何改变发生路径.....230

第六编 伦理 教育与意义发生

第 26 章 爱 包容与伦理事件的同一性.....239
第 27 章 规范判断与责任的发生.....248
第 28 章 教育评价与能力生成.....257
第 29 章 意义的多元发生与解释.....266
第 30 章 教育和伦理中的制度修复.....276

第七编 AI 与人机科研系统

第 31 章 模型 智能体与人机研究单位.....285
第 32 章 记忆 目标与后果的跨轮承接.....294
第 33 章 未决问题与创造的条件组织.....303
第 34 章 人机科研中的判断与知识发生.....312
第 35 章 AI 研究系统的对照实验.....321

第八编 科研新典范的建立与检验

第 36 章 科研问题如何发生.....331
第 37 章 对象建构与跨学科协作.....340
第 38 章 实验设计 证据与模型修订.....349
第 39 章 发表 复现与知识基础设施.....358
第 40 章 新典范的研究纲领与开放边界.....367
结语 让科学研究承担发生的历史.....377
参考资料.....389

导论 从属性之问到科研新典范

0.1 一个数怎样成为知识

研究者把一项测量结果写进表格，接下来的工作似乎只是比较、估计和解释。然而，在那个数获得意义以前，许多事情已经发生：某个对象被选出来，某种差异被认作值得测量，装置或任务建立了接触，观察者采用一种规则读出结果，记录又把复杂过程压成可以传递的形式。多数时候，这些安排可靠地工作，因此不必每次重新展开。也正因为它们经常成功，研究者容易忘记数值的可用性依赖一整套发生条件。

一个数进入知识，并非仅因它被准确写下。它需要有对象、有比较范围、有误差说明，还需要能够承接后续判断。某个甜度评分用于比较同一程序下的样本，与用于预测不同人群在日常饮用中的体验，承担的任务并不相同。某个考试成绩用于评价独立作答，与用于评价人机协作产出，也不能直接互换。数字可以相同，知识用途已经变化；数字可以不同，转换后的共同规律仍然成立。科学的严格性不仅在于把数算得准确，还在于说明数究竟在说明什么。

本书从属性之问进入这一问题。甜、重、能力、风险、信任，这些词常被当作某种对象已经拥有的性质。研究随后寻找它们的数值，解释数值为何变化。这样的工作方式在许多场景中有效，因为对象身份和操作条件已经获得足够稳定的规定。然而，一旦发生条件改变，原来能够压缩的过程重新变得相关，单纯调整数值可能无法解决困难。问题可能已经从某个量是多少，转为我们是否仍在测量同一类发生。

设想一个并未实际执行的比较研究。同一组学生先独立完成任务，再使用智能工具完成相似任务，成绩普遍提高。如果只保留总分，容易把变化归为个人能力增加；如果只强调工具参与，又容易把提高全部判为外来代劳。两种判断都过快。新的任务产出可能真实改善，个人核验能力可能有所成长，也可能没有迁移；协作过程还可能改变学生以后如何界定问题。应当分别记录这些后果，并说明它们怎样在同一学习发生中相互规定。新单位使研究可以保留原先被单一判断压平的差异。

这种重新提问不是对测量的拒绝。它要求测量更清楚地承担自己的条件，也让测量获得新对象。原先被当作附属信息的路径、基准和后效，可以成为解释的一部分；原先被当作异常的稳定失败，可以推动分类或目标重建。科学发生学由此提出一个积极任务：研究

对象的规定性如何形成，研究活动怎样进入这种形成，以及知识如何使后续发生获得新的可能。

属性纠缠论为这一任务提供本体论与方法论的共同起点。它把属性理解为同一 SDE 发生中具有号位侧重的稳定性。客体性、互动性和主体性并非三类先在实体各自携带的东西，再在某一时刻相加。它们是同一发生中不同规定获得突出与延续的方式。这个起点既保留属性能够被认识的信念，也要求解释认识所依赖的对象、过程和判断位置怎样共同成立。

本书所说的新典范，因而是一项需要论证和建设的研究纲领。它的资格不来自宣布旧知识全部失效，而来自能够提出新问题、形成新对象、组织新证据，并在与已有方法的比较中显示增量。它必须能够解释稳定，也必须能够分析失配；必须保留科学成功的条件，也必须让失败改变研究。一个数怎样成为知识，最终联系着知识怎样参与世界的下一次发生。

还可以从一项更普通的判断看到这种责任。两份报告给出相近的平均结果，其中一份来自同一程序的重复，另一份来自不同条件的混合。平均值虽然相近，它们支持的知识却不同。第一份可能帮助估计程序内的稳定，第二份可能把条件差异压成总体描述。若把两者都当成一个固定属性的直接测量，就会在后续使用中混淆对象。数值相似不是知识相同，数值不同也不是知识冲突。研究需要把数字带回能够说明它的事件结构。

这也解释了为什么科学发生学与档案工作密切相关。档案保存的不是结果旁边可有可无的故事，而是使结果能够被重新理解的条件。一个没有来源和程序的精确数字，可能无法进入可靠比较；一段记录清楚但精度有限的材料，却可能帮助识别关键机制。精度与可解释性不是同一尺度，成熟研究需要把两者组织起来。这样的组织使后来的研究者既能利用成果，也能在新问题出现时回到其形成过程。

0.2 发现与发生的相互支持

发现的语言强调，研究者遇到某种不能任凭意愿改变的对象，通过观察、实验与推理获得知识。这一立场具有重要力量。它使结论承担证据约束，使不同研究者可以围绕共同材料争论，也使任何个人权威都不能自动取代检验。发生的语言则追问，能够被指认和检验的对象怎样取得当前规定性，操作与分类怎样参与这一过程。两种语言可以互相补充。发现需要稳定对象，发生研究解释这种稳定如何建立和改变。

把二者写成绝对对立，会遮蔽真实科学的复杂性。科学并非一直只寻找无条件常数，也并非从未研究历史、反馈和测量安排。许多成熟方法本来就处理条件依赖、过程变化和模型修订。本书的任务不是把这些已有工作从科学史中删去，再宣布发生第一次出现。它提出的是更集中而系统的组织：把属性发生的说明责任贯穿对象建构、测量、模型、制度使用与知识积累，使局部认识能够在共同框架中互相检验。

发现与发生的区别可以通过一个思想实验说明。研究者发现一种材料在规定条件下具有稳定响应，这是发现的成果；他继续研究制备路径怎样形成这一响应，哪些变化使材料进入另一种状态，这是发生的研究。后者可能解释前者，也可能改变前者的对象边界。若新的制备程序形成一种不同组织，研究者需要重新判断它是否仍属于原类别。发现提供可追踪结果，发生研究提供结果成立的过程，两者在科学活动中彼此推进。

在社会与教育领域，研究活动的参与性更容易看见。一个评价指标被公布以后，参与者可能据此调整策略；一个预测成为公共信息以后，行动可能改变预测所针对的过程。这里也不能简单说发现已无可能。研究仍然可以发现反馈规律、条件稳定和制度差异，只是对象需要包含信息怎样进入行动的通道。发生学的要求是把这一通道纳入解释，而不是因对象会回应就放弃客观研究。

客观性因此不必以研究条件完全退出为前提。它可以通过公开对象定义、操作规则、转换依据和失败条件建立。一个结果依赖观察安排，并不意味着任何安排都同样有效；不同安排经比较后可能支持某些共同关系，也可能揭示不能合并的差异。客观性在这种比较中获得具体内容。它不是从世界之外俯瞰全部发生，而是让处在世界之中的判断能够经受公共检查。

发生学也需要防止另一种极端：把所有知识都说成制造的结果，因而不再区分可靠发现与任意命名。对象能够抵抗我们的分类，操作能够产生不符合预期的后果，竞争模型能够击败偏好的解释。构成不等于任意。研究者可以参与形成对象的某些条件，却不能凭参与取消证据责任。恰恰相反，参与越多，越需要说明自己的操作产生了什么，以及哪些结果不受自己的愿望支配。

本书保留发现的证据纪律，并将其推进到发生条件本身。我们既问结论是否可靠，也问可靠性依赖哪些安排；既问模型是否符合资料，也问资料如何被组织成模型对象；既问制度是否有效，也问指标进入制度以后怎样改变效果。这样的推进不是对已有科学的外部审判，而是使科学能够更完整地理解自己正在做的事情。

0.3 属性没有孤立的主人

通常说法把属性交给某个主人：甜属于糖，能力属于个人，风险属于资产，意义属于句子。这些表达在日常任务中可以简洁有效，但若被当作最后解释，就容易遮蔽属性成立所需的发生。糖的某些结构可以被客观研究，却不能自动包办某次甜味体验；个人的知识与技能重要，却不能单独说明在不同工具和任务中的全部表现；资产特征可以进入风险分析，风险判断仍需要时间、信息与用途。属性的归属需要以构成说明支撑。

所谓没有孤立的主人，不是说对象不存在，也不是说任何性质都可以无条件转移。它指属性的规定性不能仅靠一个先在载体的名称完成解释。对象自身也是在测量、行动与分类中取得可追踪身份的发生单位。一个杯子、一个细胞、一个市场指标和一个研究团队，可以具有不同尺度的连续性；研究者需要说明哪些关系支持这种连续性，以及何种变化要求重新识别。

三号位提供对这一整体的不同追问。客体性侧重追问什么显露为可指认、可比较的对象，互动性侧重追问差异怎样接续并形成过程，主体性侧重追问判断从什么位置发生并怎样承接后果。它们不对应三个独立零件，也不与 S、D、E 一一配对。每一种侧重仍需自己的显露、差异序列与特征纠缠。这样，三号位既能够区分，又不会失去同一发生的整体性。

研究底稿从甜的问题出发，经历了一个有意义的推进：最初问甜属于哪一方，随后要求客体甜、互动甜和主体甜分别得到说明，最后要求在同一 SDE 发生中理解它们如何共同长成。[1] 本书沿着这条推进展开，但不把任何一个操作指标直接等同某个号位。浓度、时间曲线和期待报告只是有限代理，它们怎样支持属性判断，需要在具体研究中继续论证。

号位侧重还包含稳定这一要求。若某种规定完全不能再次识别，属性词就难以支持知识积累；可是，再次识别需要保持哪些条件，仍然是一个开放问题。名称可以延续而构成已经改变，数值可以变化而对象仍有共同身份。第 1 至 5 章依次处理单位、侧重、方法、成熟和维持机制，目的正是把这些不同意义拆开，再说明它们如何接续。导论只将属性规定为需要解释的稳定性，不提前替具体领域决定稳定来自何处。

这种属性观还提出三层成熟之间的关系问题。活动本身形成了怎样的组织，是构成层问题；研究者能否理解这种组织，是认识层问题；过程、条件、结果与判断能否形成来源清楚、关系可追踪的复核材料，是记录层问题。记录可以先于充分解释而建立，并为后来

认识提供入口。三层可能互相促进，也可能不同时发展。第 4 章将分别建立判据。现在需要先避免一种短路：记录空白不能直接说明发生缺少相应号位，记录增加也不能独自证明能力已经长成。成熟所追问的是相互规定如何取得更可延续的形式。

属性没有孤立的主人，最终也改变了责任的理解。若结果由多种条件共同形成，责任不能只凭最显眼的载体分配；但共同形成也不能成为无人负责的理由。研究应追踪谁能够改变哪项条件，哪些决定如何进入后果，哪些环节需要共同维护。整体性不是取消具体职责，而是让职责获得更准确的发生位置。

0.4 读数何时需要成为事件

研究若要追查属性的发生，首先会遇到一个实际难题：有限记录究竟应当保留什么。保存一个结果，便于比较，却可能丢掉解释它的线索；保存全部过程，似乎更忠实，却可能使任何结果都被无穷背景淹没。两种极端都没有回答材料如何获得相关性。本书提出读数事件，首先是为了把这个选择本身变成可以论证的问题。

读数事件将数值与其所指对象、取得过程、判断规则和后续用途联系起来。它与发生单位的关系需要先作说明：发生单位界定我们正在研究的连续过程；读数事件则是其中一次读出及其可追踪关联的研究表达。一次发生可以包含多个读数事件，一个读数也可能连接前后不同的发生单位。研究者必须指出连接所依据的材料，不能因为使用事件一词，就假定数值已经完整保存了发生。

设想一间作坊为陶杯检验漏水。甲记录注水结束时的液面，乙记录一段静置后的变化，丙只记录最后能否交付。三份记录各有用途。若管理者仅想剔除明显破损，简短结论或许足够；若要判断裂纹处理的效果，就需要知道检查等待了多久；若要追踪售后出现的问题，还要问检验与使用是否处于可比较条件。这个设想没有预先决定哪份档案最优。它把记录要求与推断用途相接，使我们能够问：在哪一步，原先足够的信息开始不够。

困难还在于记录会移动问题的位置。补录检验时间，可能只是使分析者看见过去已经存在的差异；要求检验者在固定时点反复查看，则改变了工作的安排。前者增加可获得信息，后者可能改变过程。它们有时共同发生，却需要不同证据。若记录后的判断更好，究竟是因为遗漏条件被找回，还是因为检验行为本身更有秩序？这个分岔将引向第 11 章的读数事件和第 14 章的因果通道，本节先把问题保留下来。

事件的边界也不必与文件的边界重合。一张表可以汇总许多次检验，一次检验也可以散见于交接单、修补记录和最终报告。研究者若直接把表格的一行当作一次独立发生，便

可能把共同经历拆成互不相关的样本。相反，若因为资料来自同一作坊，就把所有记录视为一个不能拆开的整体，又会失去比较机会。合适的事件划分应使共同条件与可比较差异同时可见，而划分是否合适，要由它能支持的推断来检验。

研究底稿从误差项进入这些问题，试图把尚未解释的差异与发生条件联系起来。[2] 在陶杯设想中，某批次结果反复偏离预期，可以成为重新核对等待时间和处理程序的理由；但偏离自身没有说明答案。资料也许记错了，也许杯子分类过粗，也许原有机制假设不适用。事件记录所承诺的，是让候选解释能够落到共同材料，并使下一步补充有所针对，而非把误差重新命名后便获得理论成果。

还有一种更难判断的情况：不同历史在当前任务中给出相同后果。此时是否仍有必要保留历史？如果未来用途不会触及这些差异，删除细节可能合理；如果研究正要检验迁移，过早删除便可能关闭重要入口。我们需要同时考虑眼前任务和可合理预见的下一问题，却又不能用任何未来都有可能为无限保存辩护。事件单位由此包含知识的经济性问题：哪一种压缩能够留下必要的重查路径，哪一种压缩会使错误无法再被定位。

这些问题使读数事件具有明确的研究身份。它既涉及对象怎样被表述，也涉及资料怎样被使用，还涉及记录劳动由谁承担。第 11 至 15 章将分别处理事件表达、装置封装、数学边界、因果辨识和模型比较。把它们分开展开，是因为一份更丰富的记录并不能同时回答所有问题；把它们重新接起，是因为每一种推断最终都要回到它实际研究的发生。

0.5 不收敛为何需要分开追问

不收敛是这项研究的一个起点。某些参数在条件转换后不再可靠，某些估计随着资料增加仍然摇摆，某些过程持续变化，促使我们怀疑现成属性是否得到充分说明。然而，一个词能够聚集疑问，并不意味着它已经统一了疑问的结构。研究首先要确定，究竟是什么在趋向什么，又是谁根据何种材料判断它没有趋近。

可以设想一个团队连续校验同一装置。每轮都得到分散的读数，平均值却逐渐稳定；随后更换材料，平均值又出现差异；再后来装置经过维修，原先的换算方式不再适用。这三段经历都可能被口头称作数值不稳定，但需要的行动不同。第一段要问散布与估计精度的关系，第二段要问条件改变是否进入被测量定义，第三段要问程序版本与目标关系怎样转换。把它们混为一谈，会让研究者用增加重复次数回应一个本应重新校准的问题。

本书据此区分统计收敛、动力学收敛和跨条件迁移。统计问题关心估计量怎样随资料增加接近规定目标；动力学问题关心状态怎样沿时间演化；迁移问题关心已获关系能否用

于改变后的条件。这里只给出问题的入口，具体假设与数学边界在第 13 章建立。读者需要保留的是三个不同的追问方向，而不是预先接受它们都将导向某种失败。

进一步的困难是，观测稳定并不自动使内部解释唯一。两个候选过程可以产生相近结果，研究者可能缺少区分它们的材料；观测变化也不自动使研究无从推进，因为变化自身可以有可靠规律。稳定性与可辨识性由此成为需要分别考察的维度。怎样组织新观测，才能使相似输出背后的解释分开，将在模型与辨识的讨论中取得具体形式。这里尚不能从输出通道的数量直接决定答案。

回写又引入另一个问题。一次结果被保存以后，后来的过程可能据此调整，也可能完全没有使用这份记录。即便调整确实发生，它可能纠正偏差，也可能改变目标，还可能重新组织可选路径。我们要追问保存、调用和改变之间的实际联系。三方程所表达的相互规定，为这个追问安排了位置；某项回写是否存在、方向如何，则仍需时序、操作和机制证据。这一区别使本体论命题能够引导研究，同时不越过经验识别所需的工作。

跨条件失准还带来一个有实际锋芒的问题：原模型在什么地方超出了自己的权限。一个公式可以在原范围内很好地工作，使用者却把它带入已经改变的任务；一个参数可以是有用摘要，制度却要求它承担所有个体判断。此时应当检查的既有模型，也有外推行为。第 13 章将给出条件预测的限制，第 15 章讨论比较与迁移。它们要回答的并不是所有简化是否应该废弃，而是哪些条件足以支持继续使用，哪些差异要求改变使用方式。

病理学的对象也需在这里初步限定。本书所说的病理，指模型或制度与发生条件之间持续错配，以及这种错配如何被维持。一个异常结果不足以建立判断，某种稳定也不天然可疑。需要继续查明：同类困难是否反复出现，合理修订是否受阻，代价怎样进入后续发生。这个问题把认识缺口与公共后果联系起来，但具体评价仍要分别承担机制与规范的理由。

不收敛因而具有双重作用。它能够打断过早结束的问题，让被压在误差项里的差异重新成为材料；它也要求研究者约束自己的概括，避免用一个响亮名称遮住不同困难。全书从这一入口出发，最终要面对更宽的问题：当旧有判断不足时，新的对象、关系和稳定怎样被建立。回答之前，先让每一种困难保有可以被证明或推翻的明确形式。

0.6 跨学科转换怎样获得依据

这些问题将在不同领域反复出现，但相同的问句可能面对不同结构。材料检验中的稳定，学习中的能力保持，制度中的规则延续，都涉及时间，却不共享同一个计时单位；都

可以记录结果，却不具有同样的误差含义。跨学科研究的第一项困难，是辨认共享之处究竟在哪里。若共享的是提问方式，就不应提前借用共享机制的确定性；若已经有形式联系，还需说明各个符号在不同操作中指什么。

设想三个团队共同研究某种流程的可靠性。工程团队记录装置在扰动后恢复所需时间，教育团队关注新手能否在指导撤除后继续完成任务，组织团队关注关键成员离开以后工作能否接续。三者都使用恢复和保持的词语，却可能是在追踪数值、能力与分工。若直接把结果压成一个总分，词语上的一致会掩盖对象差异。若三者完全各说各话，又会错失共同问题：一次成功由哪些条件维持，条件替换以后何种关系仍然有效。

可行的转换从一项有限对应开始。三个团队都可以先说明原任务、改变了的条件和继续成功的判据，然后考察各自资料是否保留这三个环节。此时共享的是比较结构。要进一步说它们具有相同的反馈机制，就必须证明在各自系统中，结果实际通过某条通道影响后续行动。工程量的传输、学习者对提示的理解、组织中的规则修订，分别需要自己的证据。结构相似可以组织合作，却不能替代这一步证明。

转换也可能停在中途。一个领域能够操纵关键条件，另一个领域只能利用既有变化；一个领域的对象允许多次近似复现，另一个领域的每次操作都会留下历史。这样的不对称无需被掩饰。它会改变什么强度的结论可以成立，也会改变什么样的材料值得共同保存。相同问题若在不同领域只能获得不同层次的回答，跨学科工作应当呈现这一差别，让较弱结论保留自身价值，而不是向较强领域借来确定语气。

另一种不对称来自规范目标。装置恢复得快，可以相对于明确任务评价；一个人迅速回到原行为方式，却可能只是再次承受不合理要求。这里的问题已经从过程是否恢复，转到恢复的目标是否值得维持。事实材料能够帮助说明后果，但价值判断还要比较理由、选择与代价。把规范层次加入研究，并不意味着任意放弃经验约束，而是让经验说明它实际能够支持什么，并给尚需论证的要求留下位置。

跨学科的统一因此需要保留差异账目。当一个概念进入新领域，研究者应说明它保留了哪种关系，改变了哪个含义，放弃了哪些旧假设。以对象同一性为例，材料连续性、功能延续与制度承认可能承担不同角色。若一种连续性不足以支持新任务，就要补充转换依据，而非只重复对象仍然相同。第 37 章将把这项工作放进协作过程；这里先确立其方向：共同语言应使分歧能够被精确表达，而不是让分歧从文字中消失。

领域材料还必须能够反向限制共同框架。假如一个严格建立的不变量已经组织了大量条件差异，发生学应当追问这些条件怎样支持不变，不能因为期待变化就将稳定宣布为假象。假如某个社会案例显示公开信息并未改变后续行为，也需要缩小关于回写的主张。只有各领域有权给出不利回答，统一框架才具有实际内容。否则，跨学科广度不过是一个概念被允许进入多少段叙述。

本书按领域次序推进，是为了逐步增加这些转换责任。物理学要求我们尊重精确关系与操作条件；生命问题迫使我们处理物质历史和多尺度承接；经济中的公共信息使描述与行动发生联系；伦理和教育要求公开规范目标与发展条件；AI 进一步使记录、判断和行动的分工成为可设计对象。这些是进入问题的理由，具体答案留待各编论证。后文之间可能互相支持，也可能迫使先前表述改写。

从跨学科研究中能够取得的成果，因而不限于一条普遍定律。明确某种机制不可迁移，可以避免错误类比；建立一种可转换的记录形式，可以使不同团队共享材料；发现一个反复被忽略的对象边界，可以改变研究设计。这些成果层次不同，却都必须让接续者知道还缺哪一步。跨学科知识的统一要求每一答案都能在共同问题中占有准确位置，并说明它与其他答案之间可以转换到何种程度。

0.7 创造从未决问题的组织开始

科研创造常被描述为一个新想法突然出现。新想法当然重要，但它能够成为知识，还需要未决问题、材料、工具、判断和共同体形成适当联系。一个看似大胆的概念若不能改变研究动作，可能只增加叙述；一个看似微小的记录改造，却可能使长期混在一起的机制终于能够区分。创造的发生条件值得研究，因为它关系到新对象和新问题怎样获得可持续生命。

未决问题不是尚未填写答案的空白。它可以有明确结构：两个结果不能在当前模型中同时成立，一项分类无法处理反复出现的边界事件，一个成功实践无法迁移到新条件，一种工具改善产出却没有改善独立能力。把这些结构保存下来，研究者才知道新的尝试应该回应什么。若每次探索都从一个宏大口号出发，而不承接具体未决，工作容易在新颖措辞之间循环。

创造的组织还需要回答从哪里开始。材料尚杂乱时，研究者可能先追查一个异常，也可能先制作一个能暴露困难的原型；已有可靠成果时，则可能从迁移失败重新打开问题。选择入口会改变最先取得的证据，也会影响哪些解释暂时被保留。第 3 章将集中说明三方

程、六路径与三原理怎样联合承担这些工作。这里先提出它们需要满足的要求：每次转向都有实际困难作为理由，改变之后能够返回原问题接受检查。

原型具有特殊作用。研究者在尚未完全理解机制时，做出一个可以被试验的安排，使问题获得更清楚形态。原型可以是任务、仪器、分类、档案或协作流程。它的价值不只在于是否立即成功，还在于使哪些条件变得可见。一个失败原型若准确暴露了隐藏依赖，可以为下一轮理论提供材料；一个成功原型若无法说明范围，也仍需要继续研究。创造因此包含试作、比较、失败与修订的全过程。

创造还需要保护不同贡献的可见性。提出概念、整理异常、改造装置、建立资料、识别错误、完成复核，都可能改变问题的发生。若制度只奖励最后一个显眼结论，维持知识所需的劳动容易被压到幕后。发生学不是要求把所有贡献视为相同，而是要求说明它们怎样进入成果。公开这一过程，有助于公平分工，也有助于后来者知道如何继承和改进研究。

共同体对创造的支持不应等同于降低证据要求。新想法可以先作为猜想获得讨论，但其身份必须清楚；探索结果可以暂时不具普遍性，但不能被包装成完成的定律。更合适的安排是给不同阶段以不同责任：概念阶段要求一致性与问题价值，模型阶段要求明确假设与推导，经验阶段要求可追踪材料与竞争检验，制度应用阶段还需承担实际后果。这样既为新生事物留下空间，也使成熟判断不被过早借用。

创造的评价还要允许时间差。某种概念重构可能先改善对象识别，稍后才形成新的预测；某个工具可能立即提高效率，却长期没有改变知识。短期收益与长期贡献需要分别记录，不能互相冒领。研究纲领可以为较慢形成的成果保留空间，同时要求每个阶段交出可检查产物。这样，耐心与责任能够共同存在，探索既不会被即时指标完全支配，也不会成为无限延期的承诺。

本书自身也处于这一创造过程之中。它提出一个跨学科研究纲领，建立概念关系、形式边界和具体研究方向，并不把所有设想写成已经执行的实验。书中的虚构案例用于检验概念，数学模型用于证明条件结论，领域资料用于限制经验判断。不同材料各自承担责任，才能使一个哲学纲领既保持想象力，又逐步获得能够被他人继续研究的形式。

创造还包含判断何时不再增加新概念。已有语言若能准确组织问题，应当继续使用；新术语只有在承担新的区分、解释或行动任务时才值得保留。概念节制使真正的新问题获得更清楚的位置，也使跨学科交流不被不断增殖的名称阻断。

0.8 本书的论证道路与开放承诺

全书八编形成从本体论基础到科研制度的连续推进。第一编建立发生单位、三号位、联合方法、成熟与稳定机制，使后续讨论有共同语言。第二编以物理学检验客体化、测量路径、装置与不变量，明确哲学重读与新增物理主张的不同责任。第三编把这些问题转成读数事件、判断封装、不收敛分类、因果辨识和模型迁移，使概念能够进入具体研究设计。

第四编研究味觉、细胞、神经、生命稳态与多尺度回写，追踪历史怎样进入状态，稳定怎样由持续过程维持。第五编转向经济参数、市场价格、公开预测、GDP 和制度路径，分析读数怎样从描述进入行动。第六编讨论爱、规范、教育、意义与制度修复，在共同发生分析中保留规范判断的独立责任。第七编以 AI 和人机科研系统为对象，区分模型、智能体、记忆、目标和后果承接，并把创造与知识发生带入可比较任务。

第八编返回科研本身。问题如何形成，对象怎样在跨学科合作中建立，实验怎样支持修订，发表与复现如何保存知识，新典范又应以什么研究纲领继续推进。这一返回不是把所有案例强行收束为一个答案，而是检查共同方法经历不同领域以后还剩下哪些可靠内容。某些初始设想可能得到加强，某些需要缩小，某些则应明确放弃。理论的统一以能够承接这些结果为条件。

全书承担三个相互联系的主张。第一，属性的稳定性需要发生说明，客体性、互动性与主体性应在同一 SDE 过程内理解。第二，测量与知识使用可能参与后续发生，研究单位应在必要时保留这种承接。第三，科研创造能够通过问题、证据、方法与制度条件的重组获得公共支持。三个主张并不具有相同的证明方式。第一主要承担概念与本体论论证，第二需要形式模型和经验检验，第三还涉及研究组织与规范选择。

这种分工使本书能够接受局部失败。某个回写机制不成立，不必自动推翻属性发生的全部立场；但本体论宽泛性也不能保护每个失败模型。某种记录方式没有增益，就应简化；某种现成理论已经充分解释案例，就应承认其成果；某个领域需要不同的单位，就应重建转换。理论的力量不在从来没有失败，而在失败能够准确落到应当修订的层次。

与此同时，本书不把谨慎当作最后目标。说明边界，是为了让更强的主张能够在合适位置成立。稳定机制的分析可以帮助设计更可靠的系统，事件记录可以使新条件进入知识，成熟研究可以揭示能力如何形成，联合方法可以使未决问题转成可执行研究。哲学的

创造力在于重新组织可能性，而学术责任在于让这些可能性接受材料与后果的约束。二者相互需要。

读者也可以依自己的问题进入本书。关注哲学基础者，可以先沿第一编理解属性和发生单位，再用后续领域检验其解释力；关注实证设计者，可以从第三编进入，把概念转换为辨识与迁移问题；关注制度和人机研究者，则可以从具体案例返回共同方法。不同阅读路径不改变论证责任。每一领域都需要保留自己的证据，每一概念也需要接受领域反向限制。全书希望形成的不是只能顺次接受的封闭体系，而是一组可以互相校正的研究通道。

这种开放结构还给后续研究留下任务。本书提供初步机制、条件命题和研究方案，未来应通过更细的案例、正式实验与共同体实践检验它们。检验不能只寻找成功材料，也应公开无增益、不可辨识和迁移失败的结果。若新框架能够使这些结果更容易定位和继承，即使部分具体猜想被撤回，科研也可能因此获得进步。理论承担历史，不是保证自己永远延续，而是使自己的改变能够成为公共知识的一部分。

从不收敛病理学到科学发生学，首先是一种问题范围的扩展。不收敛使某些旧判断无法继续沿用，发生学则把新的对象、方法与稳定如何形成提到前景。这个转向能否成立，要看后文是否能够说明原问题遗漏了什么，新单位又增加了何种可受检查的规定。我们先从最小的属性判断进入：一次甜的发生，为什么难以仅由一个既定主人说明；沿着这一困难，发生单位的边界将逐步变得清楚。

第一编 属性发生的本体论基础

第 1 章 属性之问与发生单位

1.1 从一口甜开始的归属困难

桌上有一杯糖水。有人说它甜，有人说它不够甜，第三个人在吃过一块更甜的点心以后，觉得这杯水几乎没有味道。这个日常场景很容易被整理成两个问题：糖水究竟含多少糖，人的感觉为什么不同。第一个交给化学分析，第二个交给感知研究，各自都能获得知识。然而，当问题是这一次甜究竟怎样发生时，两个答案尚未接上。浓度记录并不是那一口甜的完整记录，感觉报告也没有把产生报告的过程全部交出来。杯中液体、取样方式、口腔过程和比较经验并非分别结束工作之后，才在最后一刻合成一个结果。它们的作用随着发生推进而获得具体含义。

设想两个人饮用同样配方的糖水。一人迅速吞下，一人缓慢含在口中。这里不需要先断言哪一人感觉更甜。真正值得注意的是，同样的配方没有保证同样的呈递过程；即使最后评分恰好相同，也不能据此断言两次发生完全一样。相同评分可能来自相似的过程，也可能来自不同变化的补偿。反过来，不同评分也没有立即证明配方发生了变化。研究若只保留末端数值，就会把这些不同可能压进同一张表。表格并没有做错事情，但它完成的任务比研究者后来赋予它的解释任务窄。

属性归属的问题由此发生了移动。通常问法是，甜究竟属于糖、属于人，还是属于两者的互动。这种提问容易使三个候选者在问题提出以前就被安排妥当：糖是对象，人是主体，互动是连接。随后无论把甜分给哪一方，乃至把它平均分给三方，都没有触及划分本身如何成立。属性纠缠论要求向前再走一步：在这一次发生中，什么被指认为同一对象，什么差异被读成甜，什么过程让这种指认能够延续。这三个问题在回答中相互约束，不能把其中两个永远留作不言自明的背景。

这一步并不取消糖、水、身体和器具的日常存在。它改变的是说明责任。将糖水作为实验样本时，研究者已经选择了配方、容器、温度范围和取样规则；将一个人作为评分者时，已经规定了任务、报告尺度和比较方式。这些规定不必全都改变结果，却决定了哪些变化可以被当作同一被测量的变化。发生单位因而比日常名词的边界更细，也更有组织性。它不是把世界无限放大，而是把支撑当前判断的必要关联纳入同一个可追踪过程。

研究底稿中关于甜的连续追问，重要之处正在于拒绝让主体感觉垄断显露，同时也拒绝把客体甜重新放回一个孤立容器。客体性、互动性与主体性是同一甜味发生的不同侧重；每一种侧重都应继续说明自己的显露、差异序列与特征纠缠。[1] 这一原则既给研究增加任务，也帮助研究停止无效争论。争论不再围绕谁最终拥有甜，而转向不同侧重在什么条件下能够得到明确说明，以及这些说明如何彼此支持。

1.2 属性判断首先组织差异

把一块布称为柔软，看起来是在陈述布的性质。这个陈述却已经组织了一组差异：与什么相比柔软，在怎样的接触下柔软，弯曲、压缩还是贴肤的体验承担了判断。不同任务可以合理地采用不同组织方式。包装材料需要的柔软与衣领需要的柔软，可能分享一些机械条件，也可能在具体使用中发生分歧。因此，属性词不是贴在对象上的透明标签。它把某些差异保留下来，使它们可以重复指认，同时让另一些差异暂时退出当前任务。

这种组织并非随意命名。若一种布无论怎样接触都割伤皮肤，仅靠改变措辞不能使它获得适合贴肤的柔软。世界通过后果限制我们的分类。可是，后果的限制也需要发生过程才能显露：切割、拉伸、穿着和清洗可能揭示不同的耐受边界。对象对分类的抵抗不是一个脱离操作的终极裁判，而是经由可重复过程进入判断的约束。由此可以同时拒绝两种省事的解释：属性既非任凭名称创造，也非只需等待名称忠实抄录。

考虑一个明确的思想实验。两种织物在某项压缩测试中读数相同，其中一种在反复弯折后仍保持表面平整，另一种出现局部硬结。若任务是一次性填充，两者可能仍属于同一可用类别；若任务是长期贴身穿着，原有等价就需要拆开。这里出现的不是同一个永恒属性被偶然情境污染，而是属性判断的用途使原先被压缩的历史差异变得相关。新测量的价值在于改变了可采取的行动，而不只在于获得更多小数位。

差异组织也包含尺度。一个杯口从远处看是平滑的，触摸时可能察觉缺口，显微观察又会给出更多起伏。不能据此说杯口同时有无限多种互相否定的性质。每个判断需要声明观察尺度、比较标准与任务。对饮用安全而言，有些微小起伏并不重要，某个肉眼不明显的锋利缺口却可能重要。相关尺度并不总由几何大小决定，还取决于后果怎样传播。发生研究因此关心差异被哪一过程放大、被哪一过程抑制，以及这种选择能否在新条件中维持。

属性是稳定性的侧重，意味着差异必须形成某种可延续的规定性，才能支持重复判断。若每次说柔软都任意改变标准，这个词不能协调行动；若标准永不允许改变，又会把

已经失效的分类当作对象的错误。稳定与修订在这里构成同一任务的两个方面。稳定使经验能够积累，修订使积累能够面对新的发生。二者之间没有一个脱离用途的最佳比例，但研究可以给出具体的比较：某种分类在何种任务中减少误判，又在何种迁移中持续失配。

1.3 对象身份在发生中延续

一只陶杯掉下一小块釉面，日常语言通常仍把它认作原来的杯子。若杯底裂开并持续漏水，它还是不是同一只杯子，会随问题而变。修复记录关心材料与历史连续性，饮用安排关心容纳液体的能力，商品交易还关心承诺的品质。不同判断并不必然彼此矛盾，因为同一性总要回答某一项具体的追踪任务。困难发生在一种任务的同一性标准被悄悄用于另一任务，而使用者仍以为自己没有转换问题。

发生单位必须把这种转换放进研究。它包含某个对象如何被带入任务、哪些变化被允许、哪些变化会要求重新识别。这里的对象不是无边界的世界整体。研究可以合理地把杯子从房间中区分出来，因为许多房间变化不影响当前判断；但当环境温度使杯壁出现新的裂纹时，被忽略的关联就需要重新进入。边界的有效性由其承担的任务与经受的变化来检验。边界越明确，越能说明何时应当扩展，而不是越能保证永不改变。

对象身份的历史性也解释了为什么简单复现有时并不简单。两次看起来相同的注水，可以作用于不同的杯壁状态：前次清洗、干燥或修补可能留下后果。要判断它们是否是同一种检验，研究者需要说明当前任务容许哪些历史差异。如果任务只检验此刻能否容水，过去的细节可以被当前表现部分替代；如果任务是比较某种修补方式的耐久性，修补前后的连续记录便成为单位的一部分。历史由此获得的是相对于问题的必要性，而非因为发生过就一律进入解释的资格。

不过，历史相关并不意味着每个过去事件都必须永久保存。一个过程可能存在足以承接后果的状态摘要。假如某项任务只需要知道材料当前含水率，早先每一滴水何时进入未必提供额外帮助；若干燥路径改变了内部结构，含水率又可能不够。研究的关键在于检验摘要何时充分。可以比较保留完整历史与只保留摘要的预测，观察两者在新操作后的差异。若摘要持续有效，压缩便构成知识成果；若失效集中于特定路径，发生单位需要补入那些路径。

由此出现一种比固定实体清单更严格的对象观。严格不在于把对象规定得永不变化，而在于说明跨变化追踪它的理由。材料连续性、功能连续性、组织连续性和社会承认可能分别承担理由，不能用其中一种自动替代其他种。一次修复可以保留材料却改变功能，也

可以更换材料而保持某项制度身份。属性纠缠论应把这些差异写清楚，研究它们如何共同支持对象识别，不能把所有复杂情况一概归为模糊。

1.4 发生单位的四个环节

一个可工作的发生单位，至少需要呈现进入条件、展开过程、可辨认结果和后果承接。进入条件说明当前发生继承了什么，包括材料状态、操作规则和可利用的历史。展开过程说明哪些差异被依次组织，哪些选择改变了后续可能。可辨认结果使这一次发生能够被记录、比较和讨论。后果承接则追踪哪些结果成为下一轮条件。这四个环节是分析安排，并非四个相互隔绝的时间盒子；在连续活动中，某一结果可能同时成为另一路过程的进入条件。

这四个环节同时约束单位的大小。若问题是杯壁处理能否减少渗漏，研究应追踪处理前后可比较的杯壁、施加工序、检验结果以及下一次使用；杯子所在房间的全部历史则无需自动纳入。若后来发现检验者依据先前标签改变等待时间，标签与等待程序才获得进入单位的具体理由。单位扩展须指出一条新增联系改变了哪个推断。这样的准则使边界既可打开又可收束：重要的不是纳入条件最多，而是现有边界能够支持当前结论，潜在扩展能够说明预期增加什么知识。

以共同撰写一份实验报告为思想实验。研究者甲留下初稿，乙发现某项变量定义前后不一致，提出修改；修改使原先能够合并的数据需要分组，分组又改变结论。若发生单位只取最后发表的文字，就难以说明问题为什么被发现。若只取甲或乙的个人能力，又会漏掉文档、记录和批注所承担的承接。有效单位应围绕该结论的形成组织材料：谁在什么版本上发现差异，什么证据迫使哪一步改变，改变后如何重新约束论证。个人并未因此消失，他们的贡献获得了更具体的界定。

进入条件也包含未决之处。一个尚未解决的问题可能长期留在研究中，影响材料选择与方法改造。若只记录已获答案，研究的方向性会被误写成一串偶然成功。未决问题并非缺少内容的空格，它可以有明确结构：两个测量不能同时满足某个假设，某个模型无法区分两种机制，某项解释没有交代跨条件差异。把这种结构保存下来，下一轮工作才知道自己应当改变什么。发生单位由此包含问题的连续性，而不仅是结果的连续性。

后果承接不必总靠有意识的记忆。装置校准、文件版本、材料变形和制度规定都可以保存前次过程的影响。主体性侧重也不能因此被简化成存储位置。记住一条信息、把它当作理由、据此改变目标，是不同的承接方式。在人机合作中，系统保存历史消息只保证资

料有机会再次进入处理，不保证它理解了理由，更不保证后续行动承担了相应责任。发生分析需要说明保存、调用、评价与行动之间实际存在的联系。

选择单位还涉及成本。过窄的单位容易遗漏机制，过宽的单位可能让任何差异都获得事后解释。一个良好单位应当具有可缩减性：研究者能说明删去某项记录会失去什么，也能说明哪些关联目前可以忽略。这样，整体性就不会成为要求记录一切的口号。整体指当前属性成立所需的相互规定，而不是宇宙全部事实的总和。单位的界定应随证据改进，却必须留下版本，使不同研究者知道他们是否仍在回答同一个问题。

1.5 观察者如何进入发生

把观察安排写入研究，不等于主张观察者的意识创造了所有对象。读温度计的人可以影响记录方式，也可能通过接触或移动装置改变测量过程；这两种影响需要分别说明。前者涉及信息如何进入公共判断，后者涉及被测过程如何受操作改变。它们有时同时存在，有时可以有效分离。哲学上的主体性侧重若要进入具体研究，必须沿这些可追踪途径展开，不能让一个含混的观察者概念承担所有因果工作。

观察之前的选择同样重要。研究者选择了什么量、用什么单位、把哪段时间作为窗口，这些选择决定了哪些差异会显得稳定。选择并不自动使知识失去客观性。相反，公开选择理由和转换规则，能够使不同观察者检查同一判断。若一个结果只在某种窗口下成立，说明窗口比声称无条件客观更有信息。若不同窗口经过合理转换后给出相容结果，又可以据此识别更深的稳定性。客观性因此可以被理解为经受公开比较的能力，而不必依赖观察条件完全退出世界。

设想研究一段会议讨论。只计发言次数，可能发现甲比乙活跃；记录提议如何被修订，可能发现乙的一次质疑改变了全部方案；记录会后谁承担实施，又得到新的贡献结构。这里没有一个单一指标可以自动代表全部参与。三种记录分别服务不同问题，任何一种都可能有合理用途。问题在于将发言次数当作贡献的全部，再根据它配置下一轮机会。这时读数开始回写发生：被评价者可能改变发言策略，原指标的含义也随之变化。研究必须追踪这一转换，而不能只增加一次统计控制。

观察反应性也可以成为积极资源。公开某项安全检查可能促使参与者在下一轮更认真核对，教师反馈可以改变学生的解题路径。不能因为测量改变对象，就宣布测量失败。应当先判断任务是描述原有状态，还是支持状态改善；前者需要控制或估计反应，后者需要

评估改变是否达成目标。相同的介入在两个任务中承担不同职责。发生单位使这种差别可见，也使研究者能够为自己造成的变化负责。

对于不具有理解能力的系统，观察与后果之间仍需按具体机制分析。反馈控制可以使某个物理量保持稳定，仪器读取也可以在一定精度要求下近似不影响目标。既然忽略观察影响有时有效，发生学就应说明这一近似如何成立，而不是把成功重新命名为压制。近似是从复杂发生中提取可靠结构的方式。它的边界越清楚，越能在失效时引导下一项研究。主体性进入说明，不是为了削弱所有测量，而是为了使测量的成立条件能够被认识。

1.6 从关系描述走向构成说明

说一个属性依赖关系，尚未说明它怎样形成。几乎任何结果都能列出许多相关因素；把清单画成网络，也不保证获得发生解释。构成说明需要回答这些因素为何在当前任务中获得特定角色，哪些联系使对象可以被重复识别，哪些联系一旦改变便要求重建判断。关系的存在、关系的作用和关系的组织方式是三个层次。属性纠缠论的工作集中在第三层，同时要求前两层提供可检查的支持。

以两个人共同搬运长桌为思想实验。最终搬到的位置相同，不代表合作方式相同。一种方式是双方事先约定节奏，途中各自调整；另一种方式是一方不断发令，另一方被动执行。两者都可能成功完成短距离搬运，但面对狭窄转角时，适应方式可能不同。若研究只以终点成功评价协作，就会丢失转向能力的差异；若只以发令次数评价支配，又可能忽略特定场景中集中指挥的作用。构成说明应追踪重量、空间、动作反馈和判断位置怎样共同组织出可行路径。

这一例子还表明，属性并非每一轮都从毫无结构的零点开始。动作习惯、约定和工具都是过去发生留下的资源。它们在当前过程中被调用，也可能被修改。强调发生不等于把所有现成结构看作虚假；现成结构正是发生能够继续的条件。所谓纠缠，是这些条件在当前活动中彼此规定，而不是任何部分都必须持续剧烈变化。一个被可靠调用的约定，可以让复杂协作显得简洁；其简洁是积累的结果，值得解释和保存。

构成说明具有反事实要求。若声称某项反馈是合作能力的必要条件，应当说明取消或替换它会发生什么。这个反事实不必总以现实干预实施，也可以先用模型、历史对照或不同情境比较来展开，但证据强度需要相应限定。如果取消某一通道后功能不变，可能存在替代通道，原先的必要性判断就需要修改。若只有多条通道同时受损才出现问题，说明能力可能由冗余结构维持。整体性不排斥局部分析，反而要求局部改变在整体中得到定位。

对同一现象可以存在多个有效的构成层次。搬运任务可以研究肌肉与关节，也可以研究共同注意，还可以研究劳动组织。层次不同不意味着其中只有一个真实，其余都是表面。关键是解释各自回应什么差异，允许什么操作，产生什么后果。跨层研究的困难在于建立转换，而非宣布一个最高层概念已经包办所有层次。SDE 提供的是组织转换的框架；具体机制仍需由相应尺度的材料来承担。

1.7 发生单位与科学创造

新的科研问题常从单位的改变开始。原先被当作同类的对象被重新区分，原先分开的现象被发现共享一个过程，原先只记为误差的差异进入主要问题。这些改变不保证产生正确理论，却改变了研究可以看见什么。发生单位因此既是分析工具，也是创造条件。它让研究者能够在不随意扩大因果解释的前提下，重新组织观察对象，并检验新组织是否带来更强的说明。

设想一项关于手工制作的研究，最初只比较完成速度，后来发现相同速度的作品在返修率上差别很大。增加返修率后，又发现差异主要集中在材料交接处。此时发生单位从个人完成一次动作，扩展为材料识别、交接、制作和复核的连续过程。新问题不再只是怎样让工人更快，而是哪些交接信息使后续动作能够可靠完成。这个重构并没有否认速度的重要性，而是说明速度与质量如何在共同发生中获得可比较意义。

这种创造可以沿三个方向展开。第一是发现遗漏的承接：前一轮操作留下了哪些状态，后续分析却将其当作无历史的输入。第二是识别隐藏的转换：一个物理读数如何变成分类，一项分类如何变成行动理由。第三是寻找可以重组的条件：既然某种能力由多个环节共同维持，是否存在成本更低或迁移更好的安排。三者都要求把新问题带回具体材料，防止概念重构停留在语言的新颖。

新的发生单位必须承受竞争。若原有单位加上一个普通条件变量就能同样解释现象，复杂重构未必必要；若新单位只改善事后叙述，却不能帮助预测、干预或复核，其贡献应限于解释组织。另一方面，不能只用原有单位认可的结果评价新单位。一个改变问题的框架，可能首先让此前无法描述的失败得到清楚界定。对此应比较是否形成了新的可检验后果，是否减少了判断冲突，是否让不同研究者能够围绕同一材料继续工作。

手工制作的设想显示，单位重建的成果可以很具体：它把返修从个人速度的附属损失，变成检验交接是否充分的材料。新的单位因此需要承诺一种可观察改进。若改进交接以后，返修仍不随相关信息变化，便应返回材料差异或其他工序寻找解释；若新单位只让

事后故事更完整，却无法改变任何比较，其作用仍停留在叙述层。创造单位与约束单位属于同一项工作。接下来必须说明，研究者怎样把这种有理由的边界交给他人，而不是让新单位继续依赖提出者的直觉。

1.8 边界的公开与单位的修订

发生单位一旦进入共同研究，就不能只存在于研究者的直觉中。它需要一个可供他人使用的边界说明。这个说明不必列尽全部条件，但必须交代纳入与排除的理由。以陶杯的研究为例，若当前问题是清洗后的容积稳定，杯壁颜色可以暂时退出；若问题是颜色变化是否提示表面损伤，原先被排除的记录便重新相关。同一材料可以承担不同研究单位，而每个单位都需要保持自己的问题连续性。否则，研究者可能在结果不利时偷偷更换目标，却把新结果写成对旧判断的证实。

单位修订应当区分三种情形。第一种是信息增加，原来被假定相同的条件现在能够被测量，但研究问题没有改变。第二种是对象重分，资料表明原有类别包含具有不同后果的过程，需要建立子类。第三种是任务改变，研究已经从描述结果转向设计干预，或者从评价个人转向评价协作系统。这三种变化都可能合理，却具有不同的连续性要求。第一种可以更新原有估计，第二种需要重新解释总体结论，第三种则必须说明新旧结论各自回答的问题。

假设陶杯检验最初按单只杯子登记，后来发现同一修补批次的杯子容易一起出现问题。研究者可以增加批次标识，保持单杯问题不变；也可以重新区分修补工序，使原来一个类别变成若干过程类别；还可以转而研究批次交付的可靠性。三种处理对应不同单位变更。第一种仍回答单杯发生，第二种改变类别的可比范围，第三种把共同批次的后果作为目标。资料可能来自同一批杯子，却不能把三种结果无缝拼成同一结论。边界说明应保留这一选择，使后来者知道新证据是在补充旧问题、改写对象，还是开启了另一项任务。

边界公开还可以减少跨学科合作中的误会。研究者常使用同一个词，却在追踪不同对象。工程人员说稳定，可能指允许误差内的持续输出；组织研究者说稳定，可能指人员和规则的可预期性；伦理讨论说稳定，可能指承诺值得信赖。三者可以关联，却不能直接互换。合作需要先让每一种稳定说明自己的事件、时间范围与后果，再寻找实际相接的环节。真正的统一从转换关系中产生，不从术语表面的一致中产生。

一个单位能否被共同使用，还取决于它怎样处理未知。未知可以位于进入条件、过程通道或后果承接，也可能涉及对象身份本身。将未知定位，比笼统说系统复杂更有用。若

尚不能测得内部状态，但可以可靠记录输入和输出，研究仍可建立有限预测；若连哪些结果属于同一类事件都未确定，优先任务应是比较和分类。不同未知要求不同的下一步行动，不能用同一套增加样本的办法解决。发生单位由此也组织了研究的次序。

合理的单位修订应保留导致改变的材料，说明新单位解决了什么旧困难，并指出新增风险。杯子与批次的比较只有在保留共同条件时才有意义，若每次失败都靠无限扩展单位获得事后解释，研究就无法区分有效边界与补救叙事。新单位应当在此前未使用的材料上给出更清楚的区分，或使另一位研究者重现同样的判断。本章由此取得一个最低要求：属性研究的单位必须能够被指认、能够承接差异，并能公开修订。尚待回答的是，这一个单位内部为什么可以突出客体性、互动性与主体性，而又不被拆成三个先在实体。

第 2 章 三号位的稳定性侧重发生

2.1 侧重如何产生可辨认的属性

在同一场合里，一张桌子可以被看作待移动的物件、协作搬运的支点，也可以成为讨论中谁能够发言的位置条件。桌子的材料没有因为提问改变就任意变化，但它在不同任务中显露的规定性确有差异。若只把差异归给观察者的兴趣，便会遗漏任务实际改变了什么；若把每种用途都说成一件新实体，又会失去同一发生的连续性。号位侧重提供另一种理解：某些相互规定在当前过程中得到突出，成为稳定指认和行动的依据，其他规定仍参与其成立。

侧重不是注意力的一束光，照到哪里就创造哪里的属性。注意力可以参与侧重形成，却不能独自保证属性成立。搬运者把一处桌沿当作着力点，桌沿是否足以承受力量，仍受材料、形状与动作方式约束；讨论者把桌首当作主持位置，这一位置是否实际获得组织能力，也取决于规则、共同承认和发言过程。属性侧重由一组可延续的配合形成。它既有选择，也有承受选择的条件，并经后果检验其稳定程度。

客体性侧重使某物能够作为可指认、可比较的对象进入判断；互动性侧重使接续、转换与相互调整形成可追踪的过程；主体性侧重使某个判断或行动具有位置、取向与后果承接。三者是分析同一发生的三个号位，不是三种材料。不能把桌子全部放进客体格、搬运动作全部放进互动格、搬运者全部放进主体格，然后宣布分类完成。同一个搬运者的身体状态也可以进入客体性侧重，他的动作参与互动性侧重，他对危险和目标的判断参与主体性侧重；但这些并非三个互不相干的人。

属性作为侧重稳定性，还要求区分一时突出与持续规定。一次偶然发言不一定使某人获得主持位置，一次偶然承重也不一定使桌沿成为可靠的着力点。稳定不是必须永远不变，而是在一定范围内能够支持再次识别。这个范围包含时间、扰动和用途。若主持位置只在特定议题中成立，就不宜把它扩展为不受限制的权威；若着力点只适合两人缓慢搬运，就不宜用于单侧猛拉。侧重的范围越清楚，属性越能成为公共知识。

三号位同时参与，也不意味着每个时刻都同等明显。一个熟练动作可能使互动过程退居意识之外，可靠装置可能让操作者不必逐次反思判断，已经形成的分类可能使客体显得自然清楚。隐去并不等于没有构成作用。对于具体操作者、装置或反馈通道，发生研究要区分实际未参与、参与但未被认识、已被认识却未被记录三种情况。这里可能未参与的是

某个具体承担者或通道，而不是某一号位的构成作用。否则，研究者容易把自己的记录缺口当作世界缺少某一维度，再以补充记录冒充新对象的诞生。

2.2 客体性侧重与共同指认

客体性首先涉及共同指认。两位研究者要比较结果，必须有理由相信他们所谈的是同类量、同类事件或可转换的对象。这种共同性并非只有名称一致。两个文件都写温度，不保证采用同一测量位置；两份报告都写满意度，不保证题目与对象相同。客体性侧重要求建立可追踪的指认路径，让不同位置上的研究者能够回到足够相容的对象。编号、单位、取样、校准和分类因此不是附属手续，而是客体化工作的组成部分。

客体化具有积极的认识价值。它让个别经验成为能够比较、传递和修订的材料。搬运者甲说桌子容易倾斜，乙说途中没有困难；若不知道两人指的是哪一段路径、在哪一边着力，这两项经验难以比较。将位置、路线和动作阶段对应起来，才可能判断他们是否描述同一过程。共同指认不要求两人报告相同感觉，它要求分歧能够回到可共同查验的环节。由此，客体性并非从互动之外取得对象，而是在发生中建立可以相互纠正的参照。

客体性并不与主体性零和竞争。研究者公开自己的判断规则，往往使对象更容易被别人识别。一个分类边界若只有创立者凭经验才能使用，客体化程度有限；当他能够说明典型实例、边界实例和错误情形，分类便获得更强的公共性。主体位置没有消失，而是把自身的判断转换为可检查条件。这种转换也会反过来改变判断者：他可能发现原来认为明显的区别并不稳定，需要修订标准。客体化因此可以是主体性承担责任的结果。

考虑一个思想实验：研究团队收集破损陶杯，准备分析漏水。最初按照裂纹长度分类，后来发现相同长度在不同位置产生不同后果。团队于是把位置、深度与杯壁结构纳入。这个过程不是承认所有对象都无法客观分类，而是使分类更贴近发生。旧指标仍可用于某些比较，但不能独占解释。客体性在这里通过差异的细化获得发展，也通过新的失败条件获得边界。若新增分类在后续样本上无助于判断，又应删减不必要项目。

共同指认还需要允许转换。不同装置可能用不同方式接近同一目标，只要转换依据和不确定范围清楚，就不必要求原始读数完全相同。把读数一致当作客体性唯一标准，会把可换算差异误判为对象分裂。相反，只要给出转换公式就宣布全部相容，也可能掩盖定义变化。客体性侧重需要同时检查对象身份与转换有效性：我们是否仍在测同一个目标，转换是否在当前条件内获得支持。两个问题缺一不可。

客体化的限度也应通过公共程序呈现。若一种分类在边界案例上经常产生争议，研究者可以保留争议类别，而不强迫所有材料立即归入已有格子。暂缓分类并非放弃认识，它能防止不确定判断在后续统计和制度使用中被当作确定事实。与此同时，争议必须组织成可继续工作的差异：分歧来自材料不足、标准不同，还是对象本身正在变化。这样，共同指认不再意味着所有人必须立即一致，而意味着人们能够准确说明一致到哪里、分歧从哪里开始。

2.3 互动性侧重与过程的规定力

互动性侧重让过程本身进入属性判断。一个动作的后果不仅取决于起点与终点，也可能取决于接触顺序、持续时间、等待间隔和中途反馈。这里的可能性必须逐案检验，不能把所有系统都宣布为有记忆。但一旦路径影响后续结果，研究就不能继续把它仅作为输送现成性质的管道。过程不仅传递条件，也组织条件，使某些差异能够发生而另一些差异暂时不能发生。

把两种粉末混合，先后顺序是否重要，要由具体材料决定；在讨论中先给结论再给理由，和先呈现疑问再比较证据，可能改变参与者如何理解任务；在学习中先尝试再接受提示，和直接接受答案，也可能形成不同的后续能力。这些例子不能互相证明，但共同提示一种研究入口：终点相近时，哪些过程差异仍然在下一轮留下后果。互动性侧重尤其重视这类被末端结果遮住的承接。

过程的规定力还体现在节奏。太快的反馈可能使参与者来不及形成自己的判断，太慢的反馈可能无法与原动作联系；适合哪一种节奏，取决于任务与承接条件。不能据此提出一个普遍最佳速度。更有价值的做法，是记录行动、反馈和修改之间的时间关系，比较它们如何影响下一步。节奏不是附加在既定互动上的装饰，它可以决定反馈是否成为可利用的信息，进而改变整个过程的组织方式。

互动性并非只有人际关系才具有意义。装置与样本的接触、程序模块之间的数据传递、材料加工中的先后顺序，都可以形成可研究的接续结构。使用互动一词不应暗示这些系统具有意图，也不能把人的协商与物理传递视为同一机制。它们共享的是过程需要被说明，而非已经共享同一个动力方程。概念的统一应留下领域差异，使具体机制有位置展开。

以两人搬运长桌为例，双方之间的调整可以通过语言、动作感觉或事先约定完成。某条通道暂时失效时，其他通道可能补偿。若只统计说话次数，无法完整判断协调是否存

在；若完全不记录语言，又可能错过转角处关键指令。互动性侧重应寻找实际承担接续的通道，而非预先指定某种沟通方式为全部过程。一个成熟协作可能话很少，却拥有清楚的调整能力；一个高频交流的协作也可能反复误解。密度不是有效性的替代指标。

2.4 主体性侧重与判断位置

主体性侧重涉及谁从何处判断、以什么为目标、怎样承接后果。这里的谁不必被缩为孤立心灵，也不能被扩张成任何有状态的系统。一个研究团队能够共同形成目标和判断规则，但责任仍需落实到可说明的分工；一个计算程序可以根据历史状态改变输出，却未因此自动获得人的经验与责任。主体性的分析应区分取向、判断、经验和责任等不同问题，不以单个功能指标包办全部含义。

主体位置会随对任务的理解而变化。搬运者最初把尽快到达当作主要目标，经过狭窄转角时却发现需要优先保护桌沿，并重新分配停顿与发令。此时即使终点与用时没有改变，判断的组织已经不同。研究者应分别记录目标如何被解释、谁有权提出暂停，以及暂停是否实际改变后续动作。同一目标下动作更熟练，与目标本身被重新规定，承担不同的说明。主体性侧重正在这种有位置的取舍中获得内容，而非只由参与者拥有某种内在偏好来定义。

主体性也通过坚持而显露。某人面对新的压力仍维护原先边界，并不必然说明他僵化；可能是边界仍有充分理由。相反，频繁改变立场也不自动表明开放，可能只是追随即时奖惩。主体性侧重关心判断怎样与理由、历史和后果相接。稳定与变化都要接受这一检查。只有把理由和承接过程保留下来，才能分辨原则的持续、习惯的惯性、策略的调整和被迫的顺从。

在共同研究中，主体位置还影响什么被看成问题。实验执行者可能首先察觉操作困难，统计分析者可能首先察觉分布结构，实际使用者可能首先察觉迁移失败。这些位置提供不同入口，却没有任何一个位置天然拥有全部真相。合作的价值在于使入口能够互相改变，而非把多份意见机械汇总。执行困难可能迫使分析者重定义变量，使用失败可能迫使设计者修改目标，模型判断也可能纠正使用者对偶然结果的归因。

承接后果使主体性具有时间厚度。一个判断若不必面对后来发生的事情，可以轻易获得表面的确定；当判断者需要追踪错误、解释修订并改变行动，取向才进入持续过程。科学共同体的主体性不应只由研究者自述的意图说明，还应看其纠错安排如何运作。错误能

否返回原结论，失败能否改变下一轮设计，受影响者能否提出有作用的异议，都是判断位置如何形成的具体问题。主体性因此与公共知识的组织紧密相连。

2.5 每个号位都有自己的 SDE

三号位与 S、D、E 不能一一对应。把客体等同显露、互动等同差异、主体等同纠缠，虽然看起来整齐，却使每一号位失去自己的发生过程。客体性需要显露为可指认对象，需要差异序列建立分类和比较，需要特征纠缠使对象在特定条件中保持身份。互动性也需要可辨认的过程显露，需要动作与反馈的差异接续，需要维持接续的组织条件。主体性同样具有判断的显露、目标和基准的变动，以及使位置能够承接后果的纠缠。

这一交叉可以沿搬运长桌的同一过程展开。客体性侧重下，桌沿被指认为能够着力的位置；其显露是某处可被共同找到的支点，差异序列包括转动前后位置的改变，特征纠缠则涉及形状、承重和路线怎样共同限定可用位置。互动性侧重下，抬起、转向、暂停形成可追踪的接续；其显露可以是一次协调完成，差异序列是彼此调整的次序，纠缠是动作如何通过桌体和提示相互限制。主体性侧重下，暂停成为有理由的判断；其显露、理由变化与责任安排，也需要分别说明。九个说明位置围绕同一发生展开，并没有生成九件桌子或九次搬运。

这些描述可以共享同一条记录。一次转角处的暂停，既可帮助定位桌体位置，也可标识动作接续，还可成为判断取向的材料。每次使用所支持的推断却不同：位置记录可供几何比较，动作次序可供过程比较，暂停理由则需有关判断的证据。把记录分别放进三格，并没有增加三份独立信息。交叉说明的作用是明确同一材料承担哪些责任，以及还欠缺哪些材料；它不要求每一格都设一个独立仪表，更不以表格填满作为解释完成的标准。

同样，特征纠缠不能被缩成外在环境清单。温度、规则和记忆若只是同时列在纸上，还没有说明它们怎样共同规定结果。纠缠的说明需要展示联系的具体形式：某一条件是否改变另一条件的作用范围，某一结果是否成为后续选择的前提，某一分类是否使某种差异进入或退出判断。E 不是装剩余因素的篮子。凡是尚未知道怎样关联的内容，应明确保留为未知，而非通过命名纠缠获得解释完成的外观。

逐格追问使新增观察有了顺序。若路线位置已经清楚，却不知道暂停如何由一方传到另一方，可以先追踪提示与动作的接续；若接续清楚，却无法解释参与者为何接受危险安排，可以追问目标与责任的理解；若各项记录充分而结果仍相反，则需要比较替代机制。

缺口位于哪里，决定下一项材料应从哪里取得。这样的组织保留同一发生，同时允许研究在当前任务中只深入最有区分力的联系。

2.6 侧重转换与属性迁移

侧重转换发生在原先承担辅助作用的规定进入主要判断时。一个原本只被当作支撑物的桌子，在协作设计中成为动作协调的关键；一个原本只用于统计的分数，在奖惩制度中成为参与者调整策略的目标。转换不是词语换位，因为它改变了哪些差异获得实际作用。研究应追踪转换之前和之后的任务、规则与后果，而不能仅凭解释者换了一套术语，就宣布发生已经改变。

某种属性能够迁移，意味着它在新的发生中仍然支持相关判断。可迁移性有程度和范围。一个人在熟悉工具下形成的检查能力，可能迁移到相似工具，却未必迁移到全新任务；一种材料分类可以适用于固定温度范围，超过范围后需要重建。迁移失败并不自动证明原属性虚假。它可能说明原有稳定性依赖的条件没有被一同带过去。属性研究因此需要区分已有能力、支持能力的条件以及迁移所需的转换。

设想一个团队在长期合作中形成高效的无言配合，新成员加入后效率下降。若把高效协作当作每个老成员独自拥有的属性，便难以解释为什么更换一个人会改变全部过程；若把失败完全归为新人能力不足，又可能忽略原团队从未将关键约定公开。发生分析会检查哪些判断寄存在历史互动中，哪些提示只有熟悉者才能理解，哪些环节可以通过重新设计转为公共资源。迁移因此成为揭示属性构成的机会。

这个例子显示，实践成功与解释充分之间可以存在距离。团队能够完成任务，却尚未将关键约定变成新成员可用的资源；迁移失败于是使既有能力的构成变得可追问。本书在这里区分两项工作：先确认某种协调在原条件中确实存在，再解释支持它的条件怎样进入新的发生。第 4 章将进一步区分构成、认识与记录三层成熟；本节只需保留这一结论：迁移是检验侧重稳定性如何获得支持的机会，既可能扩大其范围，也可能使原先过宽的属性归属得到收缩。

侧重转换还可能暴露权力与成本分配。一个机构把复杂劳动压成容易比较的指标，管理获得便利，执行者可能承担更多未被记录的协调工作。不能仅凭压缩就判其不正当，因为共同指标也可能带来可预期性。需要问哪些工作被省略、谁承担后果、申诉能否修订规则。客体性、互动性和主体性的分析在此成为同一制度事件的不同入口，规范判断仍需明确价值理由，不能从三号位齐全直接推出制度公正。

迁移研究还有一个容易忽略的方向：新的发生可能改变原来的属性理解。团队为培训新人而公开隐含约定之后，老成员也可能更容易发现旧流程中的冲突。迁移并非把已经完成的能力搬到别处，它可以反过来揭示能力原先的构成。若这种回流被记录下来，属性研究便获得双向材料：哪些条件支持复制，复制中的失败又如何促使原系统重新组织。由此形成的知识比一份只记录成功经验的操作手册更有生命力，因为它保留了能够继续学习的结构。

2.7 可分性与不可分性的不同含义

属性纠缠并不要求所有变量在任何意义上都不可分。研究至少要区分概念区分、操作分离、统计独立和发生上的不可孤立归属。概念区分使我们能够分别追问不同侧重；操作分离意味着某项条件可以在保持其他条件近似不变时被改变；统计独立是特定分布关系；不可孤立归属则涉及属性的构成说明。四者有关联，却不能互相替代。一个操作能够局部隔离某种影响，不代表属性从此拥有完全独立的主人。

同样，统计相关也不自动证明构成纠缠。两项记录可以因共同原因一起变化，也可以因尺度选择出现关联。若要从相关进入发生说明，需要增加时序、干预或结构约束，展示哪一种联系承担了实际作用。显著交互项只说明在给定模型和尺度下，一项条件的响应随另一项条件变化；它没有自动证明双向反馈，更没有证明三号位本体论是唯一解释。保持这些层次清楚，可以让经验研究支持具体命题，而不被迫一次证明全部框架。

可分离实验仍然有重要地位。它通过安排对照，使某些差异获得更明确的解释。发生学不应要求研究者放弃控制，而应要求说明控制做了什么。固定外部配方可能减少材料差异，却未必固定全部内部状态；统一报告程序可能提高可比性，也可能改变参与者注意。实验的有效性取决于这些安排与目标之间的关系。控制可以帮助揭示构成，也可以在不恰当外推时遮蔽构成，二者不能预先一概判定。

如果三种侧重的读数始终同步，理论是否失败，也需要看具体命题。同步可能来自真实的共同约束，可能来自测量指标实际重复，也可能来自观察范围过窄。它首先要求检验操作化，而不自动取消概念区分。若经过充分设计，某种区分始终不改变问题、判断或干预，便应承认该案例不需要这项复杂分析。框架的价值不以每个案例都得到最大展开为条件。能够指出何时可以有效简化，本身就是发生研究的成果。

不可分性最有力的表达，应是带条件的说明：在当前任务和允许的操作下，若去掉某类关联，便无法维持对象身份、解释后续结果或承担相应判断。这个表达允许检验，也允

许发现替代结构。它不把整体神秘化，因为每项必要性都要指出失败方式；它也不把整体碎片化，因为局部作用始终通过共同结果获得意义。三号位理论由此从一种分类语言，进入可以指导具体研究的构成框架。

2.8 三种侧重之间的解释接续

区分三号位以后，仍有一项工作尚未完成：三种说明怎样能够接成关于同一发生的知识。客体性记录清楚，未必自动说明动作如何接续；过程记录完整，也未必交代参与者为什么改变目标。若只是把三份材料并排放置，整体仍然留在研究者的口头宣称中。解释接续要求指出一份材料怎样限制另一份判断，以及这种限制通过何种共同对象或过程发生。

回到搬运长桌的设想。若甲指出某个桌沿不能作为着力点，乙需要知道这一判断将改变哪一个动作；若乙提出先转向再抬高，甲需要知道新的次序是否仍容许可靠着力；若双方决定在某处暂停，理由又会改变他们怎样评价速度和风险。这里共同发生的内容并非三份单独答案相加，而是每一个答案都改变了其余问题可以接受的回答。研究者应把这种限制保存下来，才能解释为什么换一个动作会使原有支点失去意义，或者换一种目标会使原本可用的路线不再可接受。

解释接续可以失败在不同位置。一个位置名称未经共同确认，后续指令便可能指向不同地点，这是指认失配；双方知道地点，却不知道何时转换动作，是接续失配；动作可以完成，但一方把危险当作另一方已经同意的代价，是判断与责任失配。三种失败可能相互影响，研究应沿实际材料查明传播。它们分别显示三号位为何值得区分，也说明区分之后还需要建立转换关系。

转换的证据不能只依赖同一词语反复出现。如果两份记录都写合作良好，却没有共同时间点、动作或理由可供核对，它们未必互相支持。反过来，一份记录写转角受阻，另一份写按要求暂停，表面冲突可能只是评价目标不同。研究可以先重建两者所指的动作，再比较各自标准。共同性建立在可追踪指认上，冲突则应保留到足以判定其来源的层次，而非立即用多数意见压平。

这种接续也使三号位获得反向约束。若某项主体解释声称参与者主动选择暂停，过程材料却表明通道完全堵塞，解释需要调整；若客体说明声称桌沿可以可靠承重，动作记录反复显示同处失效，就需要重新检查可用条件。反向约束并不意味着一项记录天然高于其他记录。它意味着每种说明都必须面对能够使自己改变的材料，因此同一发生可以承受不同入口的公共检查。

三号位的构成作用、研究者对它们的认识、实验对它们的记录，在这些检查中保持不同身份。新增一个传感器可能增加记录，却没有新造出互动性；共同明确暂停理由可能改善认识，也可能进一步改变实际动作。何时只是表达变清楚，何时发生本身得到改造，需要逐项说明。本体论所要求的共同构成，为研究留下完整问题；实际的因果改变，还要由通道、时序和操作承担证明。

因此，本章得到的不是把对象分进三格的分类法，而是三种不可相互代替、又须彼此接续的说明职责。其最低成果，是让人能够指出当前判断依赖哪一种侧重，以及另外两种侧重如何限制它。接着的问题在于，资料通常并不齐备，研究该从哪个入口开始，什么困难足以要求转向，又如何使新解释返回既有材料。第 3 章的联合方法将为这一具体工作建立研究次序。

第 3 章 三方程六路径三原理的联合方法

3.1 方法必须使问题发生变化

一组研究者拿到同一批材料，可以写出很不相同的报告。有人强调平均结果，有人强调个体差异，有人追踪过程，有人讨论制度条件。观点增多并不自动使研究进步。若各种报告只是把既有材料装进不同词汇，却没有改变下一项应该观察什么、哪些解释值得比较、什么结果会促使修订，那么方法的增量仍然有限。科研方法应当把理解转化为可继续行动的组织，使问题经过研究以后取得更明确的形态。

SDE 方法的统一不来自把所有学科写成同一类方程，而来自三项互相限制的工作。三方程要求显露、差异序列与特征纠缠相互说明；六路径组织研究从不同材料状态进入问题；三原理追踪某两维之间的张力如何推动第三维改变，并继续回写前两维。三者若被拆开，容易分别退化成整体性宣言、路径标签和变化故事。联合使用则要求每一次研究推进都回答：从什么材料进入，哪一项规定发生了困难，通过什么改变获得新的说明。

导论以工具参与的任务提出评价对象问题，这里以一个虚构案例进一步分析研究怎样选择证据与转向。某团队研究学生如何利用工具检查推理错误。第一次任务中，多数学生能发现明显错误，第二次面对表面形式改变的题目，表现分化。团队已有答案、时间记录和部分操作日志，但还不能确定差异来自知识、工具使用、任务理解还是评分方式。SDE 分析不应立即把某个原因填进三个号位。它先组织已有显露、差异与纠缠之间的关系，寻找可以区分解释的下一步。

方法的约束首先体现为不抢答。现有评分是 S 的一个入口，却不是全部显露；两次任务间的差异提供 D 的材料，却未必已经构成完整序列；学生、工具、任务与评价规则共同参与 E，但列出四个名词还没有说明纠缠。研究需要在这些不完整材料之间建立可检查的联系。若评分标准变化已经足以解释分化，就应先修正比较；若比较相容而差异仍在，才进一步追踪学习与工具依赖。

不抢答并非消极等待。它使研究能够提出更有针对性的动作。重新评分、改变提示、撤除工具、追踪修改理由，各自能区分不同解释。联合方法的成果应当表现为这种差别：原来只能说能力不稳定，现在能说明哪一种能力、在什么条件下、经哪条路径发生改变。概念创造在这里与实验创造相接，新的问题单位使新的证据安排成为可能。

3.2 三方程的相互规定

三方程写为：

$$S=F(D,E),D=G(S,E),E=H(S,D).$$

在概念层，它们表达三维相互规定。显露依赖差异怎样展开以及特征怎样纠缠；差异序列又受既成显露与纠缠条件限制；特征纠缠也在显露和差异的展开中得到维持或重组。这种表达不能直接充当某个经验系统的动力方程。函数的变量域、时间、测量规则与参数，仍需在具体模型中另行规定。概念关系给出说明责任，经验模型承担预测责任，两者通过操作化相接。

这里的中心工作超出向固定对象的回归式增加条件变量。在学习设想中，若评分只把最终答案当作目标，加入工具使用时间仍可能维持原对象；若日志表明同一正确答案分别来自直接采纳、独立核验和发现错误后的修订，研究便须判断它们是否仍应被称为同一种能力。新的显露分类又会反过来规定哪些步骤值得记录、哪些工具差异能够进入条件解释。对象、差异与纠缠描述由此共同取得含义。这个重建属于构成与研究单位的工作；何种步骤实际造成后续能力变化，仍须另行识别因果通道。

在学习案例中，某个答案被评为正确，是在任务、评分规则和作答路径中形成的显露。它又会影响后续差异：教师可能据此增加难度，学生可能据此改变信心或检查策略。新的差异反过来改变原先的条件组织，例如团队发现工具提示让学生跳过某个关键判断，于是重新安排任务。这样，S 不只是末端结果，D 不只是外部时间，E 也不只是固定环境。每一维都能在具体过程里参与下一步规定。

相互规定并不要求任何两项都具有同时发生的双向因果。研究者可以在一个时间尺度上把某些条件近似固定，再分析局部变化；在更长尺度上，固定条件本身可能被更新。三方程提醒我们检查这种固定的适用范围，但不替代对因果方向的识别。若某次评分只改变报告文件，没有改变学生或后续安排，就不能凭三方程说它已经回写学习。回写必须有实际承接通道。

同时，相互规定也不等于循环论证。循环论证用待证结论证明自身，不能提供独立约束；构成说明则可以让不同材料相互限制。例如任务记录限制我们怎样解释分数，作答过程限制我们怎样理解任务，后续迁移限制我们怎样评价过程。三项材料若来自不同观察与操作，就有机会排除某些解释。若三项只是同一句判断的改写，则没有增加约束。三方程的研究使用必须追求前一种结构。

一个有用做法，是为每一维写出可能使当前解释失败的材料。若声称某个分数代表自主检查能力，可以观察撤除工具后的判断；若声称某条路径形成迁移，可以改变题目表面而保留推理结构；若声称某种条件组合是必要的，可以比较替代安排。三维各有反向约束，整体就不再是一幅无法触碰的图。它能通过局部失败获得更清楚的规定，也能通过条件稳定保留已经成功的部分。

概念层与模型层之间还需要一层操作约定。设团队把一次正确检查记为结果，把检查步骤记为序列，把任务与工具安排记为条件，这样的对应只是起点。还要说明正确由谁判断，步骤怎样切分，哪些条件被记录为固定，哪些允许改变。不同切分可能改变模型结果，因此操作约定需要用边界案例检查。例如学生先接受工具建议，随后独立发现其中错误，究竟算工具依赖还是自主核验，不能由研究者事后挑选有利标签。应当把过程按时间保存，让不同判断能够回到共同材料。

这种操作约定使三方程获得有限而明确的落点。研究者无需声称已把显露、差异与纠缠全部量化，却可以说明当前代理覆盖了哪些关系。遗漏部分不应被悄悄假定为零，也不应无限扩张为所有结果的解释。若某项未知足以改变关键判断，就优先设计补证；若它在当前误差与任务范围内影响有限，可以保留近似。联合方法因此支持有理由的取舍，而不是要求每项研究都达到全知。

3.3 从纠缠进入的沉淀与锚定

六路径中的前两条从 E 进入。E—D—S 称为沉淀型，研究从已有条件与关联开始，追踪它们怎样经过差异展开形成可辨认结果。E—S—D 称为锚定型，研究先在复杂纠缠中确认一个有用显露，再以它为参照组织后续差异。两者都从 E 开始，却承担不同任务。前者重在解释结果如何形成，后者重在建立可以继续比较的抓手。

在学习案例中，沉淀型可以从教学安排开始。团队检查学生接触过哪些提示，工具在哪些环节提供答案，教师怎样反馈错误，再追踪这些安排经过练习与修改如何形成不同的检查习惯。研究不能从安排名称直接推出结果。例如同样叫提示，可能是指出错误位置，也可能是给出完整推理。只有记录差异如何进入行动，才能说明哪些条件沉淀成了可迁移能力，哪些只形成当前任务中的便利。

沉淀不是单向积累。某种习惯形成后，会改变学生如何使用下一次提示，也会促使教师调整安排。三方程要求研究在得到 S 以后回看 E 与 D：显露是否证实原先的条件解释，还是暴露了遗漏的过程。三原理则帮助定位张力。如果已有安排无法解释某组学生的稳定

迁移，就需要改变对纠缠的理解；如果过程材料显示学生一直绕过关键判断，所谓能力显露也需要重新命名。路径结束于新的问题，而非结束于一个标签。

锚定型适合条件复杂但已有某种结果值得追踪的情形。团队发现一些学生能在陌生题目中明确指出工具建议的错误理由，这个显露可以作为锚点。接下来比较他们在不同任务中的检查次序、使用工具方式和接受反馈的过程。锚点不是预先宣布存在一种独立能力，而是把某个可复核结果固定下来，让差异研究有共同参照。若后续比较发现这些学生仅在熟悉错误类型上表现良好，锚点的解释范围就需要缩小。

锚定也有危险：一个鲜明结果容易吸引研究者只寻找支持它的材料。联合方法要求让锚点接受反向改变。可以设置同样成功但过程不同的案例，也可以设置过程相似却结果失败的案例。两者迫使研究者检查哪些条件真正相关。E—S—D 中的 S 因此是暂时的研究支点，不是不可动摇的中心。路径价值在于使复杂条件获得可比较入口，同时保留入口被新差异重建的可能。

3.4 从差异进入的倒逼与原型

D—E—S 称为倒逼型。它从不能被当前解释安置的差异进入，追查条件组织，并据此形成新的显露或对象识别。D—S—E 称为原型型。它从差异出发，先做出一个可检验的初步形式，再用这一形式探索其成立条件。前者适合解释异常，后者适合通过试作生成问题。两条路径都承认差异能够推动研究，但不把差异自动归为新理论的胜利。

学习案例中的倒逼，从两次任务的分化开始。团队先确认评分、难度和参与者是否可比，再追踪差异集中在哪里。如果只有依赖某种提示的学生在新题目中失败，工具使用历史便成为条件候选；如果分化主要由任务阅读误解解释，就应重新考虑被评估的能力。E 的重建最终可能形成新的 S：原先名为推理能力的总分，被拆成错误识别、理由说明和迁移检查三个可区分结果。新对象不是由词汇增殖产生，而由差异迫使分类承担更细的责任。

倒逼型容易出现一种事后万能解释：任何差异都能在条件清单中找到对应说法。防止这一点，需要让新条件预告尚未使用的结果。若提出工具提示造成某种依赖，应当说明撤除或改变提示后哪些行为会改变；若提出任务理解是关键，应设计能够区分理解困难与推理困难的材料。差异推动解释，解释再返回差异，这个回路使方法获得约束。没有返回检验，倒逼只剩重新叙述。

原型型则允许研究者在尚未完全理解时先做出一个可操作安排。例如团队据已有差异设计一张检查卡，要求学生先复述问题、再判断建议、最后给出独立理由。这张卡是原型，不是理论已经成立的证明。它把模糊的学习困难转成可观察步骤，使研究者能够发现哪些步骤有用、哪些徒增负担、哪些改变了任务本身。试作在这里产生知识，因为它让原先不可见的条件依赖显露出来。

原型成功也不能结束研究。若检查卡改善当前任务，却没有改善后续迁移，研究需要说明收益的边界；若少数学生因步骤过多反而受阻，应检查任务与资源差异。由 S 返回 E 的工作，正是识别原型依赖什么条件。进一步的修改可能使原型更简洁，也可能使其分成多个版本。科研创造不必总从完整理论到应用，有时通过有约束的试作，理论对象才逐渐获得明确形式。

3.5 从显露进入的进入与再创

S—E—D 称为进入型。它从已经显露的结果出发，进入其条件组织，再沿新理解安排差异研究。S—D—E 称为再创型。它先让既成显露在新的比较、操作或转换中发生差异，再据此重组特征纠缠。两者都尊重已有成果，却不把成果当作完成的终点。已有结果可以成为通向发生条件的入口，也可以成为重新创造条件的资源。

进入型适合面对一个稳定成功但机制不明的实践。某班级持续表现良好，研究者不必先怀疑成绩造假，也不应立即归功于某种教学理念。可以从可复核成绩进入，检查任务构成、教师反馈、同伴协作和工具使用，再选择能够区分解释的比较。例如在保持目标相近时改变反馈次序，或在另一任务中保留相同协作方式。S 提供值得解释的现象，E 使候选条件变得明确，D 则把条件解释转成下一轮研究。

进入型也可以处理稳定失败。某类错误反复出现，研究者从错误作品进入，不急于给学生贴能力标签，而是查看他们怎样理解题目、怎样调用工具、在哪一步接受了错误建议。失败的显露由此成为重建发生的入口。若发现错误在多个学生中由同一界面提示诱发，干预对象可能从个人训练转向工具设计。属性判断的单位随之改变，研究不再只是比较谁做得差，而是说明什么安排持续生成了同类困难。

再创型更强调对既成结果进行新的差异操作。团队已有一种成功的检查程序，接下来改变任务结构、合作人数或工具条件，看其哪些部分保持、哪些部分失效。通过这些变化，原先隐含的条件逐渐显露，程序可以被重新组织。再创不是把旧方案改头换面，而是

用新的差异迫使其构成发生变化。成功可能表现为适用范围扩大，也可能表现为明确缩小范围而提高可靠性。

六路径之间没有永久等级。一项研究可以从显露进入，遇到异常后转入倒逼，再以原型试作探索新条件。路径转换的理由需要记录：什么材料使原路径不足，新的入口将解决哪项困难。也不能把每次转换都夸大为范式变革。许多转换只是正常科研推进。SDE 方法的贡献在于使这些推进形成可复走结构，让研究者能够解释自己的选择，并让后来者在相同材料上判断是否存在更好的路径。

六路径表示研究进入的次序，不应被解释成全部真实事件的时间顺序。研究者可以先看到 S，再返回较早发生的 D 与 E；也可以从制度条件进入，追踪它们跨多个时间尺度产生的结果。研究次序与对象时序一旦混淆，就容易把先被研究者发现的因素误当作先发生的原因。日志应分别记录事件何时发生、资料何时获得、解释何时提出。三种时间的区分，使回溯研究能够保持诚实，也使预测检验能够明确哪些信息当时已经可用。

同一条路径还可能在不同尺度上重复。一次任务内部可以研究提示怎样改变检查步骤，一个学期内可以研究反复反馈怎样形成习惯，制度层面可以研究评分规则怎样改变教学安排。这些研究共享入口结构，却不共享同一个时间步长。不能把短期过程的发现直接外推为长期机制，也不能用长期相关替代每一轮承接的证据。路径提供研究组织，尺度决定具体问题，两者需要共同声明。

3.6 三原理与张力的可检查形式

三原理可以写成三个张力相位：S 与 D 之间的张力推动 E 改变，D 与 E 之间的张力推动 S 改变，S 与 E 之间的张力推动 D 改变。若用乘号表示张力，它只是关系记号，不是数值乘法。第三维改变以后还需返回前两维，检验原张力是否得到解除、转移或重新组织。没有这个返回，三原理容易成为一个随时可以套用的三步故事。

S 与 D 的张力，在学习案例中并非稳定成绩与变化过程天然不相容，而是二者同某个具体解释不能同时成立。若团队将持续正确归为每次都进行了自主核验，日志却显示关键判断被直接跳过，原解释就遇到困难。重建 E 需要说明工具、任务与评分怎样共同产生正确显露，再据此调整 S 的解释和 D 的记录安排。若撤除工具后的新任务表明学生仍能独立核验，则跳过步骤也可能来自已掌握判断而作的压缩。张力要求比较这两种说明，不预先把工具依赖写成结果。

D 与 E 的张力，可以在既定条件解释无法容纳实际分化时出现。近似相同的记录下出现不同序列，首先可能是模型容许的随机差异，也可能是记录遗漏；只有原解释预告的范围被持续突破，才需要重组说明。改变 S 是一条可能路径：原先的一类能力是否混合不同任务，原先的一类错误是否混合不同过程。新分类必须对尚未用于分类的材料给出更清楚判断。若每个异常都被另命名一类，条件解释的失败只是被转移到了分类之中。

S 与 E 的张力也可能位于证据而非对象过程。团队把既成成绩称为可迁移能力，现有条件却只包含重复练习，尚无材料支持这个较宽命名。此时要求 D 改变，是增加真正不同的任务，让原命名接受检验。若新差异支持迁移，显露的解释范围获得扩展；若不支持，就收缩结果名称或重建条件说明。这里发生的是研究中的证据张力，不能据此断言学习者内部已经发生了某种动力冲突。对张力所在层次的定位，使三原理能够推动工作，又不借方法名义预先决定因果。

三原理中的推动也不意味着历史必然性。研究者可能忽略张力，也可能用不恰当办法压平张力；制度还可能使某种问题长期无法进入研究。三原理描述可供组织研究的转换关系，不保证实际科学总沿最合理路径前进。正因为推进不是自动的，方法才需要日志、比较和共同批评。张力能否变成知识，取决于它是否被准确提出，是否获得足够材料，以及改变能否承担新的检验。

3.7 联合运行的完整研究过程

把三方程、六路径与三原理放在一起，研究可以形成一个连续过程。团队首先记录已有显露和差异，说明当前对象与任务；随后依据材料状态选择入口，而非先选喜欢的路径名称；在推进中定位张力，提出改变哪一维能够形成可区分后果；最后用新材料返回检查原说明。这个过程不是一次填完的表格，而是有版本的研究活动。每次修改都应保留此前为何合理、后来为何不足的证据。

在虚构学习案例中，可以先走 D—E—S。两次任务的差异迫使团队检查条件，发现原总分混合了直接答案与理由核验，因而重建结果类别。接着走 D—S—E，试作新的检查任务，使理由核验获得可观察形式。原型产生一些稳定成功，也产生新的失败；团队再走 S—E—D，追查成功依赖的条件，并安排陌生题目和工具变化。三条路径不是三次独立验证，而是同一问题逐步取得更强规定性的不同阶段。

三方程在全过程中承担相互约束。每次结果分类改变，都要重新检查差异是否仍可比较，条件解释是否仍适用；每次操作改变，都要说明它是否改变了目标对象；每次条件重

组，都要指出哪些显露与过程将随之变化。三原理则帮助团队解释为何在某一节点转向。这样的方法日志不只记录做过什么，还记录为什么此时需要这样做，以及什么后果会说明选择失败。

强竞争解释应当在联合运行中保有真实位置。普通练习效应、任务难度差异、既有知识和评分误差，可能已经解释大部分结果。三属性分析不能先把这些解释压缩成薄弱的单因子模型，再宣布自己胜出。公平比较应尽可能使用相同资料和相近调参机会，明确新增记录带来的收益。如果一个成熟的条件模型已经给出同样预测，SDE 方法仍可能在对象重建或研究组织上有贡献，但经验机制的原创性需要另行证明。

研究结束也可以有不同成果。有时得到新的机制，有时得到可靠的条件关系，有时只确认原有简化在某范围内有效。还可能发现当前材料无法区分两种解释，从而形成明确的下一项测量。不能只有支持理论才算完成，否则研究会在结果不利时无限延长。联合方法应使失败能够定位，使成功能够说明范围，使未决能够组织下一步。其严格性来自这些不同出口都能够被他人复核。

一个完整记录还应保存未被采用的合理路径。团队可能因为资源有限先研究任务理解，暂缓研究工具界面；这种选择不意味着后者已经被排除。把资源约束与证据判断分开，有助于后来者决定从哪里接续。研究史中的许多空白并非理论明确否定某种可能，而是当时没有条件继续。公共日志若只留下最终路线，容易使偶然选择看起来像唯一合理道路。

研究过程也需要适时停止。停止并不总因问题彻底解决，可能因为当前任务已达到足够精度，也可能因为新增记录成本超过预期收益。停止理由应与结论范围配套。如果为一个低风险任务保留简单预测，就不应将其提升为普遍本体判断；如果为了探索新机制接受较高成本，就应说明探索成果尚未转为成熟应用。联合方法所追求的完整性，是推断责任的完整，而不是任何情况下材料数量的最大。

3.8 方法怎样被另一位研究者接续

联合方法的公共性可以从一个具体问题判断：另一位研究者面对同一批材料，能否理解原团队为何在某处改变研究方向。接续者不必选择完全相同的路径，但应当能够辨认各条路径正在解决什么困难。因而需要交出的，不只是最终类别和六个名称，还包括转向之前的候选解释、促成改变的材料，以及新方案预期增加的区分。方法的一致性由这些责任

保持。进入不同领域以后，资料形式可以改变，转向仍须说明由什么困难触发、准备增添哪一种约束。

设想学习案例中的团队甲从成绩分化进入，优先追查任务理解；团队乙认为最急迫的是评分目标不清，先重新核对正确答案与理由核验的关系。两种入口都可能合理。判断优劣时，不能数谁走过更多路径，而要比较目前哪项不确定性最影响后续推断。若评分口径尚不相容，甲即使收集更多操作日志，也可能仍在比较不同对象；若评分已经清楚，乙继续修订标签则可能耽误对真实过程的研究。

路径比较因此有一个具体的停止条件：当某项补充已使当前竞争解释具有相同的预期后果，再沿同一路径增加相似材料的收益可能有限。团队需要重新寻找能够拉开解释的操作，或承认当前问题暂时不能辨识。这并不否定原路径，而是明确它已经完成了哪一段工作。方法日志若能保存这个位置，接续者就不必把全部研究再做一遍，可以从仍未解决的差异继续。

原型失败也应按这一方式交接。检查卡没有改善表现，可能是学生没有按卡操作，也可能是卡片步骤完成而假定机制没有出现，还可能是当前任务根本不需要这一步。三种情况对后续路径有不同含义。第一种要求检查执行条件，第二种限制机制解释，第三种提示重新选择有区分力的任务。失败记录的价值在于保留这些可改变下一动作的区别，而不是把失败本身装饰成创造的必经阶段。

方法训练可以据此采用同一材料的分歧分析。研究者分别提出最先补充的一项证据，再公开它将排除什么解释、保留什么未知。若两种建议都不改变任何判断，应返回问题定义；若两者各能处理一部分困难，可以说明先后依赖与资源理由。这种训练使三方程的约束、路径选择和张力转向同时进入实际推理，六路径也因而成为可以比较的工作方式。

联合方法仍需与熟练的常规研究比较。若常规方案已经清楚定位上述问题，采用新术语未必增加知识；若联合使用使一个原被忽略的对象变化得到确认，则可以把增量报告在对象层；若进一步产生独特的机制预测，再由相应材料检验。方法不需要在每一层同时获胜，但它必须说明贡献实际发生在哪里。这样的公共接续将问题带入下一层：当分类、过程与判断逐步取得协调时，究竟什么发生了成熟，什么只是被记录得更清楚。

第 4 章 混沌 分化与成熟

4.1 尚未分明并不等于没有发生

一项新研究开始时，人们常常已经察觉问题，却说不清问题究竟在哪里。样本有差异，记录有矛盾，操作时有成功时有失败；研究者能够指认某些异常片段，却还不能把它们组织成共同对象。把这一状态称为混沌，首先是描述规定性尚未分明。这里的混沌不是数学动力系统中已经定义的混沌行为，也不意味着世界本身毫无结构。可能是结构尚未形成，也可能是结构已经运作而我们尚未辨认，两种情况必须区分。

设想几位教师共同设计一种新的实践课程。他们希望学生能够独立判断资料是否可信，却发现各人理解的独立判断不同：有人重视找出事实错误，有人重视说明来源，有人重视在证据不足时暂缓结论。学生作品中也出现多种表现，有的引用很多却不能比较冲突，有的判断敏锐却不愿留下依据。这一阶段并不是没有能力发生，而是目标、任务和评价尚未取得相互支持的规定性。直接给出统一分数，可能过早抹平真正值得研究的差异。

混沌状态中的材料仍然可以积累。教师可以保留典型作品、记录讨论分歧、比较同一学生在不同任务中的判断，而不急于把全部表现放进固定量表。这样做不是拒绝概括，而是为概括准备能够限制它的材料。若每一次新情况都被当作独一无二，研究无法推进；若过早把它们归入已有类别，新的问题又可能失去显露机会。混沌阶段的任务，是在保持差异的同时寻找足以组织差异的初步规定。

三号位在这一过程中并非尚不存在。作品已经作为可指认材料进入客体性侧重，教学与判断已经形成互动过程，教师与学生已经从各自位置理解任务。尚未成熟的是这些侧重之间的关系，以及它们被认识和记录的程度。不能因为三者还未清楚分工，就说三方程只在未来成熟时才成立。三方程表达相互规定，成熟描述相互规定如何分化、协调并获得更强的承接能力。两个命题处在不同层次。

这个区分使发生学能够研究真正的成长。若把成熟以前的状态一概当作缺陷，就只能从最终标准回看过去；若把所有早期差异都赞美为创造，又无法解释为何某些探索必须被放弃。成长需要形成规定，也需要保留改变规定的可能。混沌的认识价值在于尚有多个可发展方向，困难则在于方向之间缺少足够约束。研究必须让探索逐渐产生可比较后果，使分化不只是分出更多名称。

4.2 分化形成新的差异尺度

分化不是从一个现成整体中切出早已等候的部件。它是某些差异开始取得稳定作用，使先前无法分别判断的情况能够被区分。课程案例中，教师可能逐步区分资料检索、证据评价和结论修订。这个区分是否有价值，不由三个词是否整齐决定，而由它们是否帮助解释作品、设计反馈和预测后续表现。若同一组学生在三个任务中始终表现相同，区分仍可能有教学用途，却需要说明其实际收益；若差异能够指导不同干预，分化便获得更强支持。

新的差异尺度往往通过比较失败形成。一个学生能找到来源，却无法判断来源之间的冲突；另一个学生能察觉冲突，却无法说明何种证据会改变自己的结论。原先的总评价无法指明两人分别需要什么帮助，分化于是改变了教学行动。关键不在于把学生分得更细，而在于使差异能够进入适当回应。没有行动或判断上的后果，细分可能只是增加管理负担。

分化还会改变对象本身。教师公开新的评价标准后，学生可能开始主动记录结论修订过程，作品由此包含以前没有的材料。新尺度既发现差异，也参与差异的形成。发生研究应同时记录这两种作用：在旧作品中重新分类得到什么，在新规则下产生的新作品又显示什么。若把后者全部当作原有能力被准确测出，就会遗漏评价怎样组织学习；若把前者全部说成任意建构，又会忽视材料对分类的限制。

分化通常是不均衡的。某一环节先形成可靠指标，另一些环节仍依赖经验判断。不能因为可测部分更清楚，就把它当作全部成熟度。课程可能很容易统计来源数量，却难以评价证据使用；若制度只奖励前者，学生的学习路径会被引向容易计数的活动。分化因此需要反思哪些规定性被工具放大，哪些尚未获得合适表达。工具的便利不应自动决定属性的重要性。

分化也包含重新合并。若两类差异经过研究发现依赖同一操作，分别记录没有额外收益，可以建立更简洁的共同概念。成熟不是类别数量单调增加，而是类别与任务的联系逐步可靠。某些区分在探索中有用，进入实践时可以压缩；某些压缩在日常使用中有效，遇到迁移时需要重新展开。分化与整合构成持续过程，使研究既能保留细节，又能获得可传递的知识。

分化还需要能够承受边界实例。若资料检索与证据评价在某项活动中紧密相连，不能为了保持分类整齐而强行切开。可以记录一个步骤同时承担两种职责，并进一步检查这种

重叠是否稳定。边界实例往往最能说明分类的性质：它让研究者看到差异并非来自名词本身，而来自活动怎样组织后果。若重叠持续存在且有解释价值，理论可以保留交叉类别；若重叠只是记录粗糙造成，则应改进观察。不同处理需要不同证据。

差异尺度的形成也依赖共同学习。教师最初可能对作品分歧很大，经过讨论具体案例，逐步说明哪些理由支持哪种判断。这个过程若只以提高评分一致率为目标，可能把真实分歧压成服从；若完全不追求一致，又难以形成公共评价。更有价值的是记录一致怎样达成，保留仍有理由的分歧，并检查新的标准是否帮助学生改进。分化因此同时发生于对象、判断和共同语言之中。

4.3 协调使分化成为能力

各部分分清以后，整体并不会自动成熟。一个课程可以准确区分检索、评价和修订，却仍然让学生把三项任务分开完成，无法在真实问题中接续。协调涉及一种更高的要求：前一环节的结果能够成为后一环节合适的条件，后一环节的困难也能够返回前一环节。找不到支持材料时，问题可能需要缩小；证据相互冲突时，检索策略可能需要改变；结论修订后，原先的评价标准也可能需要说明。

协调通过转换规则实现。转换规则不必都是明文程序，也可以是熟练实践中的判断，但要在研究中尽可能追踪。学生怎样从多个来源形成证据比较，怎样从比较形成暂时结论，怎样从新反例返回修改，这些转折比单个终点更能显示能力组织。一个人每一步单独测试都表现良好，仍可能在转折处失败。能力的发生因此不能只用各项测验分数相加解释，还需要研究接续怎样成立。

协调也需要适度冗余。不同环节可能保留部分重叠，使一个错误有机会在后续被发现。若为了效率把每一步严格隔离，只允许它传递一个最终结果，后续环节可能失去检查依据。相反，保留全部细节又可能使信息负担过重。成熟的协调需要选择哪些过程理由必须随结果一起传递，哪些细节可以在需要时再展开。这种选择体现了知识的组织，而不只是通信容量的大小。

考虑研究团队的资料交接。执行者只交一张数值表，分析者很难判断异常是否来自操作变化；执行者交出全部原始日志而没有索引，分析者也可能无法有效使用。合适的交接应包含关键条件、版本变化和异常说明，并保留回到原始材料的路径。这样的安排使客体性记录、互动性接续和主体性判断相互支持。协调的成熟并不是所有人知道所有事，而是关键判断能够获得所需材料，错误能够沿适当路径返回。

协调还会形成新的稳定性。多个环节各自有波动，整体却可能通过互相调整保持可靠结果；单个环节表现更精确，也可能因接续方式不当而使整体更脆弱。不能只从局部性能推断整体成熟。研究需要比较扰动怎样传播，哪些环节能够吸收差异，哪些转换会放大差异。由此，成熟成为可以用恢复、迁移与修订表现来研究的过程，而不是一个抽象赞词。

4.4 构成 认识与记录的三层成熟

一个技艺熟练的人可能做得很好，却难以说清自己的判断；一个理论训练充分的人可能解释得清楚，却尚不能稳定完成操作。前者说明实践构成可以先于明确认识，后者说明认识表达也可能先于实践协调。发生学需要区分构成成熟、认识成熟与记录成熟。构成成熟涉及活动本身能否形成稳定而可调整的组织，认识成熟涉及这种组织能否被准确理解，记录成熟涉及过程、条件、结果与判断能否形成来源清楚、关系可追踪的复核材料。记录的组织可以先于机制理解而成熟：材料能够支持后来者复查发生了什么，却仍不足以说明为何如此。三层判据都须对应具体任务，不能以材料数量代替记录是否保住了任务所需的区别。

三种成熟度并非彼此独立。记录可以帮助认识发现规律，认识可以指导实践改进，实践失败也可以暴露记录的缺口。但它们之间没有自动等价。文件齐全的项目未必实际运作良好，实际成功的项目也未必已经拥有可靠解释。将三者区分，可以避免用形式化程度给所有实践排序，也可以避免把无法解释的成功浪漫化为不可言说的智慧。

以修复陶器为思想实验。熟练者根据触感和声音判断裂纹处理时机，作品经使用保持可靠。研究者最初只有成功记录，无法确定哪些线索发挥作用。他可以通过观察不同情境、比较判断理由、记录失误与修订，逐步形成认识。若随后设计训练材料，使新手能够识别关键差异，认识开始转为可传递记录。但训练材料能否支持实际操作，还要通过新情境中的表现检验。任何一阶段的成功都不能替代另外两阶段。

三层差异可以通过交接作进一步判断。若同一熟练者始终成功，而新手依据说明仍然失败，至少表明记录尚未承接原实践；但还不能立即判定说明错误，也可能是新手缺少训练条件。若训练充分后仍在同一环节失败，就要追问熟练者调用了哪些没有表达的线索。反过来，若新手能按记录完成，却无法应对边界变化，记录可能已经支持复制，认识却尚未解释转换。成熟评价需要沿失败发生的位置推进，才能把需要补材料、补解释和补实践的任务分开。

这一区分也保护了三号位理论的严谨性。研究者未记录主体判断，不意味着主体性侧重没有参与；实践者不能解释互动过程，不意味着过程尚未形成；某项条件被写进模型，也不意味着它已经在对象中获得新的因果地位。对认识变化与对象变化分别记账，才能判断某种成熟究竟发生在哪里。否则，哲学会把自己概念的出现误当作世界结构的出现。

认识成熟还包含对自身限度的明确说明。一个模型能够预测某范围内的结果，却无法识别内部机制，这是一种有限但真实的知识；一套解释能组织案例，却尚不能产生独立预测，也应准确说明其层次。成熟不要求全知，而要求不同能力之间的关系清楚。研究者知道自己能够说明什么，不能说明什么，以及下一项材料可能改变哪部分判断，知识才具备继续成长的条件。

4.5 成熟的多维判据

成熟若要成为研究概念，就需要从赞许性语言转为可比较判断。可以首先考察规定性是否稳定：不同使用者在明确条件下能否识别同一对象，重复活动能否形成相容结果。其次考察协调是否可靠：局部变化能否被适当传递，关键错误能否返回。再次考察迁移是否有界：系统能否在某些新情境中继续工作，失效时能否说明范围。最后考察修订是否可能：新证据能否改变规则而不使全部知识失去连续性。

这些判据不需要合成一个普遍分数。一个体系可以识别稳定而迁移有限，可以迁移广泛却解释不足，也可以纠错积极却维护成本过高。不同任务对各维度的要求不同。用于低风险重复操作的工具，可以优先追求简单可靠；用于开放研究的系统，则需要保留更多探索与修订能力。成熟评价应公开任务权重，不能把某种场景下的优势扩成所有场景的优越。

成熟也不是最大灵活性。若一个分类每天改变，参与者无法形成合理预期；若一套程序对每个新案例都重新设计，已有经验就难以积累。有效修订需要触发条件和过渡安排。系统应当知道哪些变化可以在原规则内处理，哪些变化需要重新校准，哪些变化要求重建对象。这种分层回应使稳定与变化获得组织。成熟的力量常表现在不必为每个扰动重做全部工作，却能在真正失配时及时改变。

可以用一个虚构的检测流程说明。最初流程只能在一种材料上稳定工作，后来研究者发现它对温度变化敏感，于是增加条件记录；再后来通过校准，能够在一定范围内转换读数。此时成熟度提高，不是因为温度影响消失，而是因为系统能够说明并处理影响。若进

入全新材料后转换失效，成熟的回应是限制使用并重新研究，而非把失效掩盖为偶然误差。边界意识本身构成可靠性的一部分。

成熟的代价也应进入判断。一个制度可能通过大量人工核对维持稳定输出，另一个制度可能以较少成本获得足够可靠的结果。比较时不能只看表面精度，还要看维护劳动、时间与错误后果如何分配。高成本并不必然不合理，关键是代价与任务价值相称，并且被公开承担。忽略代价的成熟叙述，容易把少数人的隐形劳动写成系统自然具有的能力。

成熟判据还应包含对意外的响应。一个体系在预定任务中表现良好，遇到未列出的情况却只能给出武断答案，说明其边界识别尚不充分。相反，能够准确暂停、请求补充信息或转交适当环节，也是一种成熟表现。暂停本身不是失败，但需要说明触发理由，避免把难题随意推出责任范围。成熟的停止能力与成熟的行动能力相互支持：前者保护结论边界，后者使知识能够真正进入实践。

可以进一步比较恢复方式。扰动以后迅速返回原状态，有时表明协调可靠，有时只是系统拒绝吸收新信息；缓慢恢复，有时意味着资源不足，有时则是在形成更合适的新结构。恢复时间不能脱离目标解释。研究需要区分返回旧状态、形成新的稳定状态和持续未决三种后果，观察每一种后果对任务、成本和未来适应的意义。成熟研究由此避免把某个单一动态指标升格为普遍价值尺度。

4.6 不同步与局部成熟

发生的不同方面常有不同速度。对象分类可能较快形成，实践路径需要长期练习，主体判断的承接又可能依赖更复杂的共同体安排。不能假定三号位会同时达到相同成熟程度。不同步可以造成困难，也可以成为发展资源。某项清晰记录可以帮助整理尚未稳定的过程，一种成熟实践可以反过来暴露旧分类的不足。研究应追踪这种互相带动，而不是把所有不均衡都判作缺陷。

局部成熟尤其重要。一个团队在处理常见任务时协调良好，面对新类型问题时却无法分工；一个学生能在熟悉题目中检查理由，换到陌生领域后仍依赖权威提示。这些情况不必被描述为原有能力虚假。能力真实存在于特定发生条件中，迁移尚未形成。把两者分开，既能保留已获知识，也能明确新的学习任务。若一旦迁移失败就否定此前全部成功，成长便失去可承接的基础。

局部成熟也可能掩盖整体困难。每个部门都能完成自己的指标，交接处却反复出现错误；每个研究环节都有技术规范，结论与实际用途却持续错配。这里需要研究局部稳定怎

样在整体中组合。某一环节的优化可能把负担转移给另一环节，使局部成绩提高而整体收益下降。发生单位若只围绕单个部门，便看不到这种传播。适当扩大单位，是为了使后果进入评价，而不是取消局部责任。

时间尺度的不匹配同样会产生表面矛盾。一个反馈安排可以改善短期表现，却削弱长期自主判断；一种缓慢训练可能暂时降低速度，却提高后续迁移。成熟评价若只保留一个终点，会偏向某一尺度。研究应至少说明当前观察窗能够支持什么，并在理论涉及长期后果时安排相应追踪。没有长期材料时，可以提出假说，但不能把短期曲线直接解释为完整成长轨迹。

不同步还意味着发展路线可能多样。某些系统先通过试作形成实践，再逐步建立理论；另一些系统先获得形式工具，再寻找适合的对象。SDE 六路径允许这种入口差异，成熟研究则追踪不同入口如何取得协调。不存在一条必须由混沌到分化再到统一的单线程序。某些阶段可以反复出现，某些结构可以在局部成熟后重新开放。成长的连续性来自问题与后果的承接，不来自阶段标签的固定次序。

4.7 成熟判断为何会发生错位

局部成熟与整体成熟分开以后，评价仍可能发生错位。一个团队产出可靠，评价者便推定其说明充分；一份操作文本完整，管理者便推定任何接续者都能使用；一种分类在熟悉案例上没有争议，研究者便推定新对象也会落入相同边界。这些推断把某一层的成功借给了另一层。问题不在成功虚假，而在成功被要求证明尚未接受检验的事情。

设想修复陶器的团队准备扩大工作。原先由熟练者完成关键判断，后来将流程写成手册，并把手册完整性当作可以交给新团队的依据。此时至少有三项成果需要分别确认：原实践是否可靠，手册是否准确说明关键关系，新团队是否获得使用说明所需的训练。每一项都可以独立进步。若仅以手册页数判断成熟，新增记录可能掩盖尚未解决的实践缺口；若只看原熟练者的成绩，又无法知道能力能否接续。

评价错位还可能来自时间。检验一次修复，需要的观察范围与判断长期维护不同。短期成功可以支持有限结论，却不能替未来表现作保。成熟度报告应当保留观察窗，尤其说明哪些能力尚未得到追踪。这样的保留不是把成果降为无意义，而是使成果能够在恰当位置继续使用。下一次研究可以据此直接检验仍然开放的部分，而不必先纠正一个被夸大的总评价。

当评价结果反过来组织工作，错位便有了实际后果。若奖励只依据手册完整，团队可能优先补充文字；若只依据短期产出，训练和边界比较可能被推迟。这些后果仍需具体材料确认，但在概念上已经可以看到：成熟的指标本身参与了成熟条件的分配。评价者因此需要说明指标究竟代表哪一层能力，以及其他层次由什么材料补充，而不把一个总分当作整个发生的替身。

修订这种错位，首先是把成功归还给它实际支持的层次。保留已经可靠的实践，同时补充尚欠缺的解释；保留能够复核的记录，同时检验它能否被新成员使用；保留有限范围的分类，同时开放边界材料。第 5 章将进一步分析能力何以维持、何时脆弱及怎样再生。本章取得的判据先行划清了评价对象，使后面的机制分析不会把记录增长直接写成能力增长，也不会把一处失效写成整体衰退。

4.8 再分化与开放成熟

成熟的系统面对新条件时，可能需要重新进入某种不分明状态。已有分类暂时不足，原有路径需要试作，判断标准尚待协调。这个过程不是简单回到最初，因为它继承了过去的材料、错误和方法。再分化的任务，是让这些积累在新条件中重新获得作用。过去经验既提供资源，也可能形成盲点；研究应让两者同时进入比较。

课程案例可以继续说明这一点。假设原任务只使用来源清楚的短文，后来加入材料不全、相互引用且结论冲突的文献，原先按照来源数量和单项准确性形成的评价便可能不足。教师仍可保留已成熟的检索与核对技能，却要重新研究学生如何表达未知、追查引用关系和暂缓总体判断。新任务没有让旧能力一夜消失，它改变了旧能力彼此接续的要求。再分化的成果应当能指出哪些旧判断仍然可用，哪些转换需要新的练习，而不是仅以新材料更复杂为由宣布必须更换全部课程。

再分化需要保护探索空间。若所有试作都必须立即达到成熟程序的效率，新的对象和方法难以出现；若探索永远不承担结果，又会失去公共支持。可以把探索与应用设为相互联系但责任不同的阶段：探索阶段允许较多候选，要求清楚记录；应用阶段要求可靠范围与可执行规则；两阶段之间通过证据门槛转换。这样既不把未成熟想法过早制度化，也不让成熟标准永久封闭创造。

开放成熟还包括能够重新理解自己的历史。一个团队可能原先以为成功来自某个明星成员，后来通过记录发现关键在于一套不显眼的交接安排。新的理解不会改变过去发生的事实，却会改变未来怎样继承成功。历史重读因此不是纯粹事后叙述，它可以重组公共记

忆和资源分配。其正当性仍取决于材料与竞争解释，不能为了建立新身份而改写不利证据。

开放成熟也要求把修订权分配给能够接触后果的人。若只有中心能够修改规则，而一线参与者只能提交无法得到回应的异常，系统的学习会受到限制。反过来，若每个人随意改变规则而没有共同版本，局部改进又可能破坏整体协调。需要建立能够连接局部发现与共同修订的过程：异常先被定位，修改理由接受比较，变更范围明确，后果随后复查。这样的组织不是附加在知识之外的行政工作，它决定哪些差异能够进入下一轮规定。

再分化的成果应当保留旧结构的可理解性。后来者不仅需要知道新规则是什么，还需要知道旧规则曾解决什么问题、为何在新条件下不足。这样才能防止每次理论更新都抹去前人的有效经验，也能避免旧规则在不知情的情况下继续被误用。知识的成熟因而包含历史可追踪性：变化有理由，保留有范围，失败有位置，新的尝试能够从已有材料继续。

本章以资料判断课程和陶器修复区分了几种常被混在一起的改变：活动可以形成新组织，认识可以找到原有组织的线索，记录可以使这些线索被接续；三者的进展并不同步。成熟评价据此取得相对任务、时间和接续能力的判据，既可以承认局部成功，也能定位尚未完成的转换。仍需解释的是，已经形成的规定性靠什么维持。只知道一个体系能够工作，还不知道它依赖固定约束、持续纠偏还是他处补偿。这一机制差别决定了同样成熟的表面为何会在扰动后走向不同结果。

第 5 章 稳定 退化与再生的机制

5.1 稳定是需要解释的成果

两只杯子都能稳稳放在桌上，一只是杯底平整，另一只是在桌面凹处找到了恰好适合的位置。短时间观察可能得到相同结果，移动位置以后却会出现差异。这个简单思想实验提示，稳定的表面形式不足以说明稳定的发生机制。稳定可以由对象结构、条件约束、持续调整、不同变化的补偿，或者观察尺度的选择形成。研究若只把不变当作解释终点，就会错过这些机制之间的差别。

稳定首先要说明稳定什么。一个读数保持不变，不代表内部状态不变；一个分布保持相近，也不代表每次结果相同；一项功能可靠完成，可能依靠不同部件交替承担任务。稳定还需要说明在哪个范围内成立。范围可以由时间、误差、扰动类型和任务定义构成。没有这些条件，稳定只能是一种印象。规定条件不是削弱稳定，而是让稳定能够被别人使用，并在变化以后接受检查。

第 4 章区分了能力形成、能力被理解和能力被记录。本章进一步追问，已形成的规定性如何在扰动中延续。一次可靠表现需要支持再次识别的组织，而这种组织可能持续调整自身。发生因而包括稳定的形成、维持、失效以及失效后的重建。要把成熟判据转成解释，就必须比较维持方式：同样一项功能可以依靠强约束，也可以依靠持续修正；两者的日常结果相近，却可能在更换条件后显出不同边界。稳定本身在这里成为待解释的成果。

稳定机制的研究可以从扰动开始。改变位置、延长时间、更换参与者、撤去某项支持，观察结果怎样响应。扰动不必很大，也不总需要破坏性干预。一个小范围的条件变化，往往足以区分被动保持与主动纠偏。若稳定依赖某个隐藏支持，撤去支持后便可能显露；若稳定来自冗余组织，局部改变未必影响功能。无效扰动也有信息，它说明当前改变尚未触及关键条件，或系统存在替代能力。

研究稳定的目的不是寻找永不变化的最后核心，而是建立可以解释保持与改变的模型。模型需要同时面对两个问题：为什么这段范围内没有出现重要变化，为什么越过某个条件后变化变得显著。只解释失稳而不能解释此前稳定，或只解释稳定而把后续失配都称为偶然，都不够完整。稳定的发生学要求将维持过程与变化边界接在一起，使已有成果成为下一轮研究的资源。

5.2 约束与重复的稳定

一种稳定来自约束。容器限制液体的运动范围，程序限制操作顺序，制度规定可以采用的判断标准。约束使某些变化无法发生，因而提高结果的可预期性。这种稳定不必被贬为人为压制。没有适当约束，许多任务根本无法完成。需要研究的是约束限制了哪些自由，保留了哪些能力，以及约束的维护成本与任务收益怎样对应。

实验中的标准化便包含这种机制。相同材料、统一程序和明确时间窗，可以减少与当前问题无关的差异，使目标关系更清楚。标准化获得的知识是真实的，但其外推需要说明条件。若目标是理解标准程序下的表现，实验已完成重要工作；若目标是预测开放场景，研究者还需检验被固定条件改变以后结果如何。问题不在于标准化本身，而在于把条件知识无声地扩成无条件断言。

重复可以巩固约束，也可以暴露约束不足。一个操作在相同安排下反复成功，说明当前组合具有一定可靠性；但重复次数增加并不自动扩展适用范围。始终沿同一路径重复，可能更清楚地认识这一路径，却不能代替其他路径的比较。稳定研究应区分内部重复与条件迁移，两者回答不同问题。前者估计既定安排的表现，后者检验安排依赖哪些构成。

设想某团队用固定模板记录实验异常。模板使遗漏减少，分类更加一致，这是约束带来的收益。后来出现模板没有栏目容纳的新问题，执行者只能把它写进备注。如果备注反复出现同类内容，模板就需要修改。成熟约束应当有发现自身遗漏的通道。若所有未列入栏目的情况都被拒收，稳定记录可能只是记录规则封闭的结果；若模板随每个个案任意变化，又会损害比较。约束的有效性在共同标准与例外返回之间取得平衡。

约束也有层次。某些限制来自材料性质，某些来自装置设计，某些来自共同规则。不同层次改变的成本和后果不同。研究不能把它们一概称为环境。材料限制可能需要替代方案，规则限制可能可以通过协商修订，装置限制可能通过重新设计改变。将约束具体化，才能把稳定解释转成行动知识。稳定不再只是现象的标签，而成为能够说明如何维持、如何扩大或何时放弃的成果。

5.3 反馈怎样维持稳定

另一种稳定来自持续纠偏。系统并非没有偏离，而是偏离会触发某种调整，使后续状态回到允许范围。为了说明这个机制，可以构造一个最简单的离散模型。设某个状态在每一步保留一部分自身，同时向目标值调整：

$$x_{t+1}=(1-\lambda)x_t+\lambda r.$$

其中 r 是固定目标，调整比例 λ 在零与一之间。将两边减去 r，可以得到偏差每一步都乘以一减 λ，因此逐步缩小。这个模型展示了反馈与收敛可以共存。它不宣称真实系统只有一个状态，也不把目标的来源解释完毕，但足以说明回写并不必然制造不收敛。

这个模型还可以说明收敛与有界保持为何需要区分。设每一步额外受到一个绝对值不超过 δ 的扰动，目标与调整比例仍保持不变，则下一步偏差的绝对值至多是本步偏差乘以一减 λ，再加 δ。如果本步偏差不超过 δ 除以 λ，下一步仍不超过这一范围。因此，持续扰动未必使状态趋向同一个点，却可以容许一个被反馈维持的偏差范围。这里取得的是模型内的条件结论。现实中是否具有相应更新规则，需要另行证据；但一旦规则得到支持，研究就可比较调整强度、扰动尺度与允许误差，而不是只看曲线是否完全变平。

在实际研究中，目标本身可能变化，反馈可能延迟，测量可能有误差，调整也可能受到限制。加入这些条件以后，需要重新分析稳定性。过度调整可以造成来回摆动，过迟反馈可能无法处理当前变化，错误目标则可能使系统稳定地朝不合适方向工作。反馈是否存在，与反馈是否有效，是两个问题；有效维持某个指标，与实现整体任务，也不是同一个问题。发生研究必须追踪目标、读数、调整与后果之间的完整联系。

以共同搬运为思想实验，一方察觉倾斜，提醒另一方调整；对方调整过大，又会引起反向修正。熟练配合可能通过适当幅度和节奏保持稳定。这里的稳定来自持续互动，不能解释为互动被取消。若研究只在终点记录桌子没有掉落，就看不到维持过程；若只记录调整频率，又可能把频繁补救误判为高效协调。需要同时观察偏差大小、调整方式和任务后果。

反馈还可以改变目标。一个学习者发现原来的速度目标导致大量错误，重新选择更重视核验的目标；一个制度发现原指标诱发不利行为，调整评价规则。这种变化超出了围绕固定目标纠偏的模型，需要把目标更新纳入发生单位。目标更新不必每次发生，也不能被写成任意适应一切结果的装置。研究应说明何种证据触发更新，谁有权更新，以及更新后怎样检验效果。

反馈稳定性的公共意义在于，它把可靠性理解为可维护能力。一个系统能够发现偏差、组织修正并保存经验，往往比一个偶然输出正确结果的系统更值得依赖。可靠性由此不只属于当下数值，还属于后续承接结构。这个结构可以支持稳定，也可以在规则失配时支持改变。稳定与学习不是必然对立，它们可以通过合适的反馈层次相互促进。

反馈的方向不能仅凭曲线外形确认。一段回到原水平的序列，可能来自主动纠偏，也可能来自外部扰动自行消退；一段持续漂移的序列，可能来自目标更新，也可能来自测量偏差。要识别反馈，需要追踪读数怎样进入调整，或通过改变反馈通道比较后果。若系统在停止报告以后仍以同样方式变化，报告行为可能不是关键机制；若改变调整规则使恢复速度按预期改变，机制解释才获得更具体支持。稳定曲线提供线索，因果说明还需要承接材料。

目标由谁规定，也影响反馈稳定的意义。一个装置可以围绕设定值可靠工作，但设定值是否适合实际用途仍由更大的判断过程决定。机构可以高效提高某个指标，却未必改善参与者真正关心的结果。研究应把局部控制目标与整体任务并列，而不是以控制成功自动证明目标合理。这样，反馈分析既保留技术判断的精确，也为规范讨论提供清楚对象，两者不必互相替代。

5.4 补偿 冗余与聚合

多个变化可以相互补偿，使总体结果保持稳定。一个团队中某人暂时速度下降，其他成员改变分工，最终任务仍按时完成。此时总时间稳定，内部贡献却发生变化。若研究只看总时间，可能误以为每个人都保持原样；若只看个体波动，又可能低估整体协调。稳定的单位与变化的单位需要同时说明，才能理解补偿发生在哪里。

补偿并不总是可持续的。一位成员长期承担额外工作，短期结果仍稳定，长期可能形成负担累积。稳定机制的分析因此需要记录代价如何分配，以及代价是否进入后续条件。若代价没有返回正式评价，系统可能在输出良好时逐渐失去恢复能力。这里有理由讨论脆弱性，但不能仅凭存在补偿就断言退化。分工调整也可能是公平而有效的协作，需要结合任务、资源和参与者的判断。

冗余提供另一种稳定。多个通道能够承担相近功能，某条通道暂时失效时，其他通道继续工作。冗余不是简单重复，它可以有不同条件优势。一个团队既有书面记录，也有口头交接；两者各有局限，合适配合可以降低遗漏。若两个通道共享同一错误来源，表面上的双重检查却未必增加可靠性。因此研究冗余需要检查通道的实际独立程度，以及错误怎样共同传播。

聚合也能形成稳定。个体结果有波动，总体平均可能较为平稳。这种稳定可以服务某些决策，却未必适合个体判断。若总体均值稳定来自不同群体变化的相互抵消，制度使用

者就需要知道谁在变化。聚合并非天然遮蔽，它是一种有用途的压缩；问题在于压缩后的量被用来承担什么任务。用于总体规划可能合适，用于判断每个参与者则可能失配。

补偿、冗余与聚合的区别，可以用各自的失败方式检验。补偿依靠其他部分增加贡献，因而需要追踪负担；冗余依靠替代通道，在一条通道受阻时应检查功能由谁接续；聚合则可能只改变观察层次，分开记录各部分便会暴露原先被抵消的变化。三者可以同时存在，名称本身不能决定比例。它们共同提出一个要求：稳定的量、承担稳定的部分以及稳定的代价需要分别记录，才能把相同输出背后的组织差别转成下一步可比较的问题。

5.5 压缩与有效简化

科学和日常行动都依赖压缩。地图不保留每块石头，摘要不保留每句话，指标不保留全部过程。若要求所有判断都重新展开完整历史，知识将难以使用。有效简化的任务，是在给定用途和容许误差下保留足够信息。属性纠缠论不应取消这个任务，而应使简化的条件和代价更清楚。简化成功，是发生研究需要解释并保存的成果。

一个简化可以通过任务比较获得支持。若只保留当前状态与保留完整历史在新情境中表现相近，历史摘要可能已经足够；若完整历史明显改善某类判断，就需要查明哪些部分具有价值。不是记录越多越好，而是新增信息能否改变可靠行动。复杂模型可能因估计困难而表现更差，简单规则也可能在有限任务中更稳健。模型选择应把信息价值、估计误差与使用成本共同考虑。

设想一种器具借用制度，最初只记录借出与归还时间，足以支持日常管理。后来器具经常归还后立即出现问题，管理者增加使用状态与维修记录。新增字段若帮助定位损坏发生与维护安排，就有明确价值；若要求记录每次使用者的所有活动，却无法改善判断，则属于无效负担。压缩与展开应围绕具体失配进行，使档案服务任务，而不是让任务服务档案。

有效简化还需要可展开性。一个指标可以在日常使用中保持简洁，但在出现异常时允许回到原始条件与过程。若只有最终分数，研究者无法解释变化；若分数关联必要的事件记录，就可以在需要时重新分析。可展开性使简化不必永久遗忘其来源。它让知识在稳定使用与深入研究之间转换，而不必每次从零重建。

简化的失效通常具有条件结构。某个模型在多数场景表现良好，只在特定人群、路径或时间范围内失准。此时可以先做局部校准，也可能需要重建对象，不能一概要求全面复

杂化。关键是保留失效分布，避免平均表现掩盖重要后果。一个总体准确的规则若持续伤害某个关键群体，规范判断还需明确如何权衡；统计性能不能独自决定其使用正当性。

简化还可能在不同使用者之间产生不对称。熟悉原始过程的人能够从一个简短指标恢复很多隐含信息，后来接手的人却只能按字面理解。同样的摘要对前者充分，对后者不足。知识传递因此不能只问保留多少信息，还要问使用者已有何种条件。一个面向内部团队有效的简表，公开传播时可能需要补充定义、范围和例外。压缩的有效性属于信息与使用者共同构成的任务关系，而不是文件本身独有的性质。

这一点解释了为什么复制规则未必复制能力。规则原先依靠共同经验补足含义，迁移到新团队后，隐含条件没有一同迁移。适当的解决可能是增加训练案例，也可能是改造规则，使关键判断更明确；不是简单要求新团队照做得更严格。发生学通过检查压缩的前提，让知识转移从口号进入可操作的条件组织。

5.6 从脆弱性到病理判断

脆弱性指某种稳定容易在特定变化下失效。它可以在系统正常工作时存在，也可能直到扰动才显露。病理判断则更进一步，涉及模型或制度与发生条件之间持续错配，以及错配为何没有得到适当修订。两者不能混为一谈。某项工具适用范围很窄，只要边界明确并被正确使用，不必称为病理；一种复杂制度即使不断调整，也可能因为拒绝关键证据而持续错配。

病理需要具体对象。可以研究一个估计模型为何在制度改变后仍被外推，一个评价指标为何把策略迎合当作能力增长，一套程序为何让反复出现的异常无法进入修订。不能把波动的人群或不符合预期的个体直接称为病。研究的重点是判断与使用方式怎样失去响应，使本来可以被识别的差异持续产生可避免后果。这样的病理概念保留批判力量，又不把差异本身污名化。

持续错配也需要时间证据。一次预测错误可能来自随机变化，一次操作失败可能来自执行偏差，不能立即升级为制度性病理。若错误反复集中于相同条件，既有解释无法消除，修订通道又持续阻断，就有更强理由研究结构问题。证据应同时包含结果、条件与回应：发生了什么，哪些解释被检验，机构或研究者如何处理失败。只有结果曲线，没有回应过程，难以判断持续错配怎样被维持。

病理机制可以通过反馈扭曲形成。一个指标进入奖惩以后，参与者改变行为，原先指标与目标的关系随之变化；管理者仍按旧关系解释，便可能强化产生失配的行为。这种循

环不必导致数值发散，它甚至可以产生非常稳定的指标。问题在于稳定围绕了什么目标，以及真实后果是否被排除。发生学在这里关注反馈的方向和对象，不以收敛或不收敛直接代替价值判断。

病理判断还要容纳修复成本。并非发现任何差异都应立即改制，改变规则也可能损害合理预期、增加执行负担或带来新的不公平。合理回应可以是维持并监测、局部校准、扩大记录、替换模型或暂停使用。研究应说明为何某种回应与当前证据相称。批判若只要求不断变化，就无法区分负责的稳定与顽固的维持；修复必须比较实际后果，才能成为知识而非姿态。

5.7 再生改变承接结构

再生意味着系统在失配以后形成新的维持与修订能力。它不必恢复原来的全部状态，也不必彻底抛弃旧结构。一个团队失去关键成员后，通过重新分工、公开隐含知识和改造工具，可能获得不同但有效的协作。再生的对象是能够承接任务与后果的组织，而不是过去形式的机械复制。研究应比较哪些功能被保留，哪些条件被改变，新的代价由谁承担。

再生往往从一个局部入口开始。反复出现的异常被重新记录，一项原先不被承认的经验获得公共表达，一个工具被修改以保留关键过程。这些变化可以沿三号位传播：新的客体记录使互动问题可见，新的互动安排改变参与者判断位置，新的主体性承接又促使记录标准修订。传播不是必然的，每一步都需要实际通道。若异常记录没有进入决策，资料增加并不会自动带来再生。

以器具借用制度为例，维修记录揭示损坏并非集中于某个使用者，而是集中于一种交接方式。机构可以改变检查与说明程序，给使用者提供及时反馈，并追踪新安排是否减少后续问题。若效果只在试点中出现，扩大后失效，就需要检查资源和执行条件。再生不是一次宣布改革，而是从新安排到新后果再到修订的连续过程。成功的标志是系统能够更可靠地学习，而不仅是短期指标好看。

再生也可能形成新的对象身份。某个研究工具原先用于给个人排序，后来改为定位学习困难；同一分数不再承担原来的制度功能，记录与解释随之变化。不能仅因为名称和界面相同，就把新旧系统视为同一对象。发生研究应说明目标、规则和后果怎样重组。历史连续性仍然存在，但连续不等于不变。新的身份可以通过公开版本与转换理由获得共同理解。

再生需要保留失败的可见性。改革者容易只展示新方案的成功，把未改善部分解释为旧制度残留；这种做法会使再生再次陷入不能修订的结构。更可靠的安排是预先说明哪些结果要求修改或撤回方案，并让受影响者能够报告新的问题。再生的成熟不在于证明改革者始终正确，而在于形成比以前更好的纠错与承接条件。稳定由此重新获得基础，变化也有了继续发生的空间。

再生与恢复之间也有重要差异。恢复可以是在扰动以后回到原有工作范围，再生则可能改变支撑该范围的组织。一个团队通过加班暂时恢复产出，尚未解决知识交接问题；通过训练和流程改造形成替代能力，才可能获得新的维持条件。两者都可能必要，但不能把短期恢复当作长期修复已经完成。研究需要分别观察即时输出、负担分配和后续独立运行，才能判断变化发生在何种层次。

再生还涉及放弃。某些旧任务已经不再适合，继续维护原结构的成本超过其价值；某些指标因用途改变失去意义，应当停止使用。放弃不是把失败解释为成功，而是明确承认边界，并把已获得知识转移到仍有价值的问题。一个能够合理停止无效活动的系统，可能比持续维持全部旧功能的系统更有发展能力。停止需要公开理由和后果安排，尤其应关注依赖旧结构的人如何获得新的承接。

5.8 稳定机制的比较研究

稳定、退化与再生可以被组织成同一项比较研究。先明确任务和目标量，再观察稳定的表面范围，提出约束、反馈、补偿、冗余、聚合等候选机制，最后选择能够区分它们的扰动或资料。机制比较应预告不同后果。例如撤去某项支持，约束型稳定可能立即失效，冗余型稳定可能保持功能但增加其他部分负担；改变观察尺度，聚合型稳定可能显露内部差异。具体预测仍需依对象模型建立，不能把这些示意当作普遍定律。

比较还需要同时追踪维持成本与恢复能力。两个系统在常规条件下输出相近，一个依赖持续高强度补救，另一个具有较大余量；两者面对新任务的表现可能不同。成本不只是货币或时间，也可以包括参与者无法表达的负担、维护知识的集中程度和纠错机会的减少。哪些成本相关，需要由研究任务与规范理由共同决定。不能把所有难以计量的后果排除，也不能用含混价值词替代实际记录。

按照第 3 章的联合方法，比较应先让稳定显露、扰动序列和维持条件互相限制，而不预选一个喜欢的机制名称。假设两个团队日常完成时间相同，研究可以先分别撤去一种辅助提示。若时间仍相同，却有成员工作量明显增加，补偿解释获得新线索；若替代通道自

动接续且负担变化有限，冗余解释更值得追踪；若总体相同仅因两类任务一快一慢相抵，就应重新分组。每种线索都须返回原结论：稳定到底属于整个功能、某项指标，还是一种聚合方式。这样使用联合方法，所重建的是稳定对象及其支持关系，具体因果仍由实际比较承担。

比较研究还应保留尺度之间的联系。短期稳定可能以长期消耗为代价，局部变化也可能维护整体功能。研究者不必一次观察所有尺度，但应说明自己的观察窗位于何处，以及哪些未观察后果足以改变结论。当某项长期机制尚无资料时，可以先提出明确预测，为后续研究留下入口。这样的有限结论比把短期结果直接推广到整个系统更容易积累，也比因复杂而拒绝判断更有行动价值。

稳定知识的最终形态，应包含可使用的范围和可启动的研究入口。使用者知道什么时候可以依赖简化结果，也知道什么迹象要求展开过程、补充条件或暂停判断。这种知识比无条件承诺更有实际力量，因为它将成功与失效放在同一说明中。边界没有削弱成果，反而使成果能够迁移而不掩盖代价，能够保持而不拒绝变化。

第一编至此为属性发生提出了可分工的说明：发生单位界定追踪什么，号位侧重组织同一事件的不同规定，联合方法约束研究怎样转向，三层成熟区分实际组织与认识记录，维持机制则解释规定性如何延续。这些说明只有进入具体材料才会获得更强约束。物理学将首先提出严峻问题：当转换关系与不变量已经得到精确规定时，发生分析还能增加哪一项必要说明？答案必须从具体对象、操作与数学关系中取得，不能仅靠把稳定重新命名。

第二编 物理学中的属性显露

第 6 章 经典物理的客体化条件

6.1 一根金属棒何以具有长度

把一根金属棒放到桌上，问它有多长，通常会得到一个数。这个数可以写进图纸，交给另一家工厂，也可以成为争议中的验收依据。交接能够成功，是因为长度在大量实际行动中具有可靠的稳定性。假如每一次换人、搬动或者重新放置都会毫无规律地改变结果，图纸和零件便难以配合。然而，这种可靠性并不意味着测量已经与放置方式、温度和端点规定脱离关系。恰恰因为这些关系得到适当处理，一个数才有资格代表一根棒，而不只是代表某次匆忙读出的刻度。

这里出现了经典物理最有生产力的操作：把一段受到条件约束的发生，组织成可以归属于对象的性质。客体化使人们能够暂时离开完整现场，携带长度、质量、弹性系数等量进入推算。它不是凭空制造一个对象，也不是简单揭开对象外面的遮布。它把若干差别证明为无关，把另一些差别写进修正关系，再把余下的稳定内容交给共同使用。所谓客体性，由此包含一种跨场景承接能力：条件按约定改变，性质仍能通过明确关系被重新找到。

设想同一根棒在两个房间被测量，两个读数略有差别。这个思想实验至少允许四种解释：对象的状态变了，仪器的校准变了，端点判断变了，或者差值仍在约定不确定度以内。仅凭两个数，不能决定哪一种解释成立。为了形成关于棒的知识，必须把差别重新放回发生过程。接触测头是否施加同样压力，温度是否充分平衡，表面附着物是否被当作边界，都可能改变“同一根棒的长度”这个表达所实际指向的内容。

属性纠缠论在此提出的主要问题，不是取消长度的客观性，而是追问它怎样取得客观性。客体侧重表现为棒及其边界可被重复指认；互动侧重表现为支撑、接触、热交换和读数路径；主体侧重表现为目标长度的选择、端点规则和接受精度。三个侧重并不对应三件各自完整的东西。同一测头既是物质对象，也参与接触路径，还承载已经确定的判断规则。把测头划入某一栏，不会消除它在另外两种分析中的作用。

经典物理的成功常被概括为排除了主体。更准确的说法是，它发展了许多办法，使某些主体差别不再影响所声明的结果。一个人喜欢这根棒，另一个人讨厌它，只要这种偏好没有经操作进入测量，两人的情绪就不属于长度模型的必要变量。这个排除是研究成果，

而非关于世界的先验禁令。一旦人的预期改变了施压、取点或筛选规则，相关差别便需重新进入分析。排除是否合理，应由作用路径和误差范围决定。

因此，客体化的首要条件不是现场无人，而是现场中哪些差别可以被忽略已经获得说明。它同时创造自由与责任：研究者不必每次重写全部现场，却要知道省略的条件何时需要恢复。长度是这个问题的朴素入口。从它出发，经典物理呈现为一种构造可承接稳定性的实践，属性发生学则追踪这种构造的成立范围和重新展开方式。

6.2 任意观察者与可交换的操作

两个观察者给出相同结果，可能有很不相同的原因。他们可能独立校准后测得一致，也可能照抄同一张表；可能采用不同方法而经转换一致，也可能共同受到一个没有发现的偏差影响。结果一致本身不能充分证明客观性。真正有力的一致，来自差异受到考验之后仍然成立：换人没有改变操作的有效含义，换仪器没有改变所指目标，换表达没有改变可检验关系。

观察者的可交换性首先是一种有条件的对称。它允许更换承担操作的人，同时维持任务、训练、程序和信息条件。这个要求不同于让任何人在任何状态下都得到相同数字。一个人用尺读端点，另一个人用手估计，虽然都面对同一根棒，却没有实施同一种操作。将他们的差别统称为主体性，既不能说明为什么不同，也不能指出怎样改善。更有用的做法，是寻找差别进入读数的具体位置。

可交换性还不同于所有坐标数值相同。即使在经典运动学里，同一运动在不同运动状态的坐标系中也可以具有不同速度。共同规律依靠的是表达之间的转换，而非数字逐项不动。把客观性理解成任意观察位置下都得到同一数值，会把经典理论本身已经具备的关系结构误写成例外。属性稳定性有时表现为一个量不变，有时表现为一组量按照固定关系一起改变。两者都可以支持跨观察者的知识。

设想一辆车在笔直道路上匀速运动，车内记录者和路边记录者分别描述一个落下的小球。若只比较各自写下的水平速度，两份记录可能显得冲突；若同时写明坐标系，两份记录便可指向同一事件。这一例子不需要把人的意识纳入动力学。主体侧重在这里主要表现为描述位置和坐标规则。它已经进入知识的组织，却不因此成为推动小球的附加作用力。主体的构成作用与主体的物理干预需要分开。

这个区分关系到本书所说的号位显题化。某个条件在发生中起作用，某个理论明确描述它，某份实验记录保存它，是三个不同层次。观察位置即使没有被逐项写入简短报告，

也可能通过固定实验室布置得到控制；理论即使把位置参数写出，也未必完整记录每次实际安排。由报告中没有“主体”二字，推断主体号位不存在，是从文字缺席跳到了构成缺席。反过来，由报告提到观察者，也不能推断理论承认一种独立的心灵力量。

经典物理提供的建设性经验，是把可交换性做成能够维护的制度。校准、共同训练、误差检查和公开操作让不同人接续同一问题。由此产生的稳定性既不属于某个人，也不悬浮在实践之外。它通过一群人能够替换彼此而取得共同效力。属性发生学应当继承这种能力，并在交换失败时展开条件，而不是预先宣布一切交换都压抑了主体。

6.3 路径怎样被保留又怎样被压缩

同样条件下能够复现，不意味着改变任何路径都必须得到相同结果。复现所要求的是某些与结果有关的条件得到保持或按模型转换。将一个小球沿不同轨道释放，即使起点高度相同，也不能只凭“经典物理”四个字要求全部运动细节一致。轨道形状、摩擦、转动以及约束方式可能改变时间过程。只有在明确的假设下，某些末态量才与具体路径无关。路径无关性是需要证明的性质，不能被当作全部经典科学的定义。

为了看清这点，可以构造一个简化思想模型。两个装置都把同一个物块从高处送到低处，其中一个允许近似无耗散运动，另一个通过可调阻力使物块缓慢下降。最终高度相同，但向环境转移能量的方式不同。若任务只关心位置，两个过程可能被视作等效；若任务关心发热、到达时间或可提取功，它们就不能合并。对象是否“相同”，在此还取决于研究保留的是哪一类后果。

经典物理之所以强大，正因为它能够选择适当的压缩。对于某些问题，用起点和终点描述即可完成推算；对于另一些问题，需要轨迹；再进一步，还要记录约束和环境交换。压缩没有统一的善恶。遗漏对任务无关的信息，能使规律更清楚；遗漏下一步需要的信息，则会使规律在迁移时失效。发生学批评应当落在遗漏与用途的关系上，而不是把所有结果表达都判为过程消失。

互动侧重的显题化，就是使这些选择变成可讨论的问题。一个力学模型把物体视为质点，意味着在当前精度和问题中，内部形变不参与必要解释；当接触或破裂成为研究对象，内部结构需要重新展开。不是模型先前完全错误，而是它授权的省略已不能承担新任务。对象的边界随问题重建，仍受原有成功经验约束。重建不能任意取消已经解释良好的部分。

这也改变了“历史”的含义。历史并非凡是过去发生的一切，而是那些仍通过状态、约束或记录影响当前过程的先前变化。一个弹性模型可以把适用范围内的过去操作压缩进当前状态；一旦材料表现出加载历史的重要性，状态说明就可能需要扩展。对于既定被估量的预测模型，是否增加历史变量，应由可区分的预测或干预后果检验。若历史信息首先改变被估量、事件同一性或比较资格，则需先论证这种对象重建，再在共同目标下比较预测。新增叙述应说明自己改变的是对象规定还是机制解释，才能取得与其作用相称的位置。

在 SDE 语言中，D 作为差异序列不能被缩减为钟表时间上的点列。它包含哪些变化被连接为有意义的路径，哪些差异能够推动显露改变。E 作为特征纠缠也不只是轨道外面的环境；约束、材料性质、测量尺度和可实现的操作彼此规定，使某些路径成为可能。S 则是这个关系中可被承接的显露。路径压缩是否成立，要同时检查它保留了哪些显露、依赖什么纠缠结构，以及怎样改变下一轮能够追问的问题。

6.4 仪器并非对象之外的透明通道

一台弹簧测力装置可以把作用力转换为可见位移。使用者看到的是指针位置，想知道的是力。两者之间需要一个经过检验的关系。指针必须能够运动，又不能把接触过程扰乱到关系不再成立；装置必须足够敏感，又不能对无关扰动同样敏感。仪器的设计因而不是把影响降到绝对零，而是把影响组织为可识别、可校准和可控制的部分。

国际计量学词汇把被测量作为拟测的量来界定，并要求相关条件足以明确其含义。测量不确定度又涉及归属于被测量之量值的分散程度。这一语言提醒人们，装置输出、被测量和报告结果之间存在需要说明的联系。它没有要求研究者把整个宇宙写进记录，而要求写明足以确定目标和评价结果的条件。[3] 本书从这里进一步提出：仪器是属性得以稳定承接的一种组织，不只是既成属性经过的通道。

设想测量一片非常柔软的薄片厚度。接触压力增加时，读数可能改变。若任务是比较规定压力下的厚度，压力是被测量定义的一部分；若任务是推断无载厚度，压力则进入测量模型与外推过程。这两种研究都可以合理，却不能把各自得到的值直接混在一起，以差异证明厚度“没有真值”。先明确要回答哪一个问题，才能判断差异属于对象响应、模型不足还是操作偏差。

仪器也保存着先前判断。量程选择规定哪些差异能被看见，滤波规则规定哪些变化被当作有效信号，读数时间规定一个过程被合并到多长窗口。说仪器保存判断，不等于说仪

器随意创造事实。材料和相互作用会抵抗不合适的设计。一个被选中的频段可能确实含有目标信号，也可能使关键变化消失；答案要由独立检验给出。判断的物质化使知识能够重复，也使错误能够被稳定复制。

可以把一次校准理解为有限的交换实验：用已知或可溯源的输入检查输出关系，再评价这段关系能够覆盖怎样的工作范围。若装置对时间和温度敏感，校准就需要相应维度。这里已有一种回写：本轮检查改变下一轮使用的转换关系。但是，校准回写不是被测对象必然发生的回写。研究者修正换算表，和装置使薄片变形，发生在不同位置。把两者都叫反馈之后，仍然必须分清各自的因果路径。

经典测量给属性发生学的一个重要约束正在这里。认识活动能够改变关于对象的表述，物理操作能够改变对象状态，两者还可能通过控制系统连接起来。它们不必处处同时发生。一个充分的发生分析允许某条路径在当前任务中为零，也允许一项控制保持目标稳定。只有保留这些差别，三号位分析才有辨认能力；若凡是变化都归入同一“纠缠”，概念反而失去对研究的帮助。

6.5 从微观差异到宏观状态

一杯水的温度能够成为生活和实验中的可靠量，并不要求人们逐个追踪全部微观运动。宏观描述通过选择状态量，把大量差异组织成少量可使用的信息。这样的选择经常被理解为删除了真实过程。但如果没有这种压缩，人们连“这杯水与另一杯水是否达到同样热状态”都难以形成可执行的问题。宏观属性的价值，在于它让特定层次的比较取得闭合，而不是冒充包含了每一条微观轨迹。

为了说明层次之间的关系，可以使用一个明确的思想实验。设想两个容器在某项粗略观测下给出同样平均读数，其中一个内部接近均匀，另一个保留明显空间差异。如果下一步只需要判断是否超过一个宽松阈值，这个平均值可能足够；如果需要预测局部材料反应，两者就不能等同。同一压缩对于前一任务是有效知识，对于后一任务则遗漏了必要结构。问题不在于平均本身，而在于平均被授权代表什么。

宏观状态是否充分，实际上取决于它能否支持后续演化的判断。给定同样的宏观描述，若系统在相同控制下仍产生系统性不同的后果，就应寻找遗漏的状态、历史或分布信息。这个判断允许逐步扩展模型。它不要求一开始就采用最大数量的变量，因为变量越多并不意味着可辨识性越强。无法独立测量的细节可能增加解释自由，却没有增加可检验的约束。

主体侧重在宏观建模中表现为选择尺度和用途，互动侧重表现为交换与演化路径，客体侧重表现为可重复确认的状态。三种侧重在同一层次描述中交织，却不因此使宏观性质成为任意的心理产物。一个选择若不能稳定支持任务，会在新的条件下暴露不足。理论的自由受到对象响应和实践后果的双重限制。选择可以更换，约束不能靠重新命名消失。

这也使“条件场”获得比环境更厚的含义。容器、边界、交换方式、观测分辨率以及允许使用的状态量共同形成一种可区分结构。E 所指的特征纠缠，是这些规定相互嵌入之后，使某些差异能够显露、另一些差异暂时不可见的组织。把它只写成外部温度或背景噪声，会漏掉尺度和操作的构成作用；把它无限扩大成所有关联，又会使研究无法落地。E 需要围绕具体任务被界定和检验。

由此得到的成熟性标准，不是所有差异都被消灭，也不是全部细节都被记录。一个成熟的描述能够说明哪些差异被压缩，为什么可以压缩，以及何种结果要求恢复被压缩的维度。它既有稳定的内核，也有可定位的边界。经典物理的客体化正是在这类工作中取得力量：它使对象能够作为宏观状态持续进入行动，同时把进一步展开的可能性留在适用条件里。

6.6 标准如何使数值离开现场

一个读数离开原实验室后仍然有意义，需要的不只是数字和单位。接收者还须知道数值所指对象、实现办法和不确定度。共同单位使表达可以交换，溯源链使不同实现能够比较，测量模型使条件差异能够转换。标准化因此是稳定性的公共组织。它把许多本来需要临场解释的问题预先处理，让分散行动能够协调，也让偏差可以沿共同链条追查。

国际单位制以七个定义常数的固定数值规定单位，其数值在这一制度内没有不确定度；实际实现这些单位仍有不确定度。[4] 这两种陈述必须同时保留。定义数值的精确性不保证每台装置实现得同样精确，实验实现的差异也不说明定义常数正在漂移。把约定、实现和自然过程混为一体，会使关于常数的讨论一开始就错置对象。

定义并不是从世界撤退到语言。一个单位之所以能有广泛效力，还需要稳定可复制的实现和相互比较。约定解决共同如何表达，实验检验表达怎样落实。两者之间既有分工，也有反馈。当实现手段改进，共同标准能够以更低不确定度被使用；当某项比较出现张力，研究者会检查实现模型、输入数据与相关性。共同定义没有消除研究，而是把研究焦点转移到可定位的问题上。

设想两家工厂使用同一种长度单位，却分别在不同温度条件下验收零件。它们拥有共同单位，并不自动拥有共同被测量。若验收规定没有温度、支撑和测量方向，争议仍然可能出现。将规定补充清楚之后，一部分差异可以通过转换解决，另一部分可能需要改变工艺。共同标准的价值就在于它使这种区别能够公开表达，而非由权威宣布其中一个数字天然正确。

标准化也有成本。某些为规模协作而设定的容许范围，可能不适合研究极微小变化；某种普遍程序可能对特殊材料产生系统偏差。因此，标准不应被理解为永远禁止例外。提出例外的人需要给出明确理由和可比较的替代程序；维护标准的人也需要说明它当前支持的用途。发生学对制度的贡献，应是帮助界定何时继续使用共同规则，何时建立有说明的分支。

在这里，主体侧重既没有被消灭，也没有被无限放大。它通过定义、协商和判断门槛进入公共形式，随后接受跨实验室的约束。个人偏好让位于共同程序，并不必然是主体性受压；它也可以是主体能够相互接续的条件。客体性由这种接续得到增强。经典客体化的成熟形式，因而不是把所有人排除，而是使人的不同位置可以在共同证据中获得协调。

6.7 一项摆实验中的控制与恢复

设想一个研究小组用摆的周期估计某个稳定参数。这是为澄清方法而构造的案例。第一轮，他们记录摆长和多次往返的总时间，依一个简化模型作出估计；第二轮，不同成员按相同说明操作，结果出现系统差异。问题由此落在一项具体工作上：找出哪些控制使两轮操作可以交换，哪些被省略条件需要恢复。一个合适的模型应当既解释已有差别，又说明怎样操作可以使差别按预期改变。

被估量的定义先决定控制的范围。摆长从何处量起，周期从哪一相位计起，摆幅和支点是否落在模型的适用范围，都影响实际过程与简化描述的对应。可以先把释放、计时和几何测量分给不同成员，交叉交换各项任务。如果差异跟随计时岗位而转移，便把注意力集中到计时；如果差异跟随某一套摆装置而转移，则进一步检查几何与约束。这种交换保存了总体任务，同时有计划地移动局部条件，因而比笼统比较两组结果更能定位问题。

计时差异还可以继续分解。设想总时间含有一个近似固定的起止偏差，将其除以所计周期数之后，平均每周期的偏差会随周期数增加而缩小；若差异来自每一周期都存在的动力学效应，单纯多计周期通常不会按同一方式消失。这是关于两个简化误差模型的推演。

它提示团队同时改变所计周期数和计时方式，并检查偏差的缩放关系。若观察不符合这两种形状，就应修改候选解释，不能仅凭自动计时比人工计时更稳定便结束分析。

同样，检验摆幅近似需要让幅度变化与计时安排尽量分开。可以在同一装置和计时规则下改变释放条件，再交换装置重复；支点则作为另一项受控条件。这样的分解建立一种局部恢复次序：先恢复最可能改变预测的维度，再判断是否值得增加新维度。每增加一个变量，都应减少某种已明确的解释歧义。若新变量只让每组数据各有一个自由修正值，却不能预言下一组结果，它尚未增加模型的约束力。

这一实验中的 S、D、E 因而具有可以逐项追踪的内容。周期图样是显露，释放、摆动、计时与交换构成差异序列，几何、约束和计时规则共同规定哪些图样具有可比意义。调整计时路径之后，原来的周期估计需要重算；调整摆长定义之后，原来的几何记录也需要重新解释。三方程所要求的相互规定，在这里表现为一个局部修订会改变其余关系的使用条件，而不是在数据之外再添加三个独立因素。

研究可以从异常周期进入，也可以从一套新计时装置开始。前者先寻找差别所依赖的条件，再安排交换；后者先建立装置的工作范围，再看它能支持哪些周期区分。六路径的价值，是使这两种工作有机会会合。新的显露与既有路径不相容，推动条件说明改变；改变的条件又要求重做相应比较。三原理因而落实为有次序的实验修订，每一步都能指出被解决的张力以及仍然保留的限制。

最后，团队可以得到几种不同而有内容的结论。计时修正后结果相容，说明操作的可交换性得到增强；保持计时后仍有随摆幅变化的规律，说明动力学近似需要展开；不同控制组合给出相同预测，则允许继续压缩记录。经典客体化在这里没有把过程赶出知识。它通过控制、交换和恢复，逐渐取得一种经过验证的省略权，使一个参数可以离开这次现场，进入下一次推算。

6.8 客体化的成就与重新展开的权利

经典物理的客体化使知识获得了一种重要的可携带性。人们能够用少量量和关系跨越现场，把不同地点的行动接起来。零件装配、轨道推算、能量核算和仪器校准，都需要这种能力。发生学如果把可携带性一概描述为对丰富世界的压迫，就会失去解释科学成功的任务。真正困难的问题是：在怎样的条件下，压缩既足够简洁又保持对下一步行动负责。

这个问题可以用两种失配来说明。第一种失配发生在对象或用途改变之后，模型仍沿用先前的省略。原本可以忽略的接触、历史或尺度因素开始影响结果，旧参数却被继续当

作无条件保证。第二种失配发生在研究者过度展开之后，模型容纳一切故事，却不能提供比简化描述更明确的判断。前者把条件冻结，后者让条件无限增生。两者都削弱知识的公共可检验性。

从摆实验可以提炼出客体化的两项配套能力：日常使用时，把已经控制的条件压缩进参数与适用范围；发生偏离时，沿保留的交换关系恢复必要环节。交付一个参数，也就是交付一种有边界的使用资格。接收者应能判断何时可以直接使用，何时须重新校准，何时已经改变被估量。知识基础设施保存这些分界，便使简洁表述与详细发生能够互相接通，而无须要求每位使用者每次重做全部实验。

主体侧重在这项工作中承担选择恢复入口的责任。研究者要在有限资源下判断，优先交换计时者、扩大摆幅范围，还是重新检查支点。选择可以包含经验和直觉，随后却要转化为他人能够接续的操作理由。当一次交换没有改变差异，原先偏爱的解释就需要退让；对于目标已经确定的物理机制模型，新条件应通过可区分后果取得增加机制项的资格；若它先改变目标或比较范围，则应公开这种对象重建，并重新确定检验任务。主体的创造性由此表现为提出可受纠正的区分，并与经典研究的可重复性共同成长。

从这一位置重读经典物理，可以明确本书关于“客体侧重”的判断：经典理论经常把可承接的对象性质放在显著位置，使主体与互动的许多规定通过边界条件、理想化和程序得到处理。但这是一种分析侧重，不是说经典世界只存在一种属性，也不是说所有经典理论都把另外两种侧重设为零。一个含有装置耦合或反馈控制的经典模型，完全可以明确处理复杂的相互作用。

经典物理留下的遗产，是一套使差异能够在共同关系中被理解的办法。接下来的问题更进一步：当测量长度和时间所使用的共同框架本身需要重建，客体化是否还能维持？相对论给出的回答，不是放弃共同知识，而是在不同参考系之间建立更严格的联系。属性发生学由此获得下一条线索：稳定未必寄居在一个不变的读数里，它也可以存在于读数之间可以公开转换的结构中。

第 7 章 相对论中的测量路径与不变量

7.1 两个长度怎样同时成立

一根匀速运动的杆可以被赋予不同长度，而这些长度不必互相否定。这个陈述容易被误读为测量最终失去了确定答案。相对论给出的实际要求恰好更加严格：先说明在哪一个参考系中，同时记录杆两端的什么事件，再比较两份结果。差别不再被当作模糊的观察偏好，而被写入确定的变换关系。读数发生变化，规律的约束并没有因此减弱。

爱因斯坦一九〇五年的论述区分了与杆共动的测量和在另一惯性系中同时确定两端位置的测量。[5] 这一分法的重要性，在于“量同一根杆”并不能独自确定正在比较的是哪两个事件。杆的身份可以保持，而构成长度的端点事件组合会随同时性规定改变。对象同一和事件选择相同，是两个需要分别确认的问题。

设想一根杆沿自身长轴方向运动。杆的前端和后端各自在时空中形成一串事件。某一时刻的长度不是从全部事件中任取两个位置相减，而是取该参考系判为同时的两个端点事件。换到另一参考系，原来同时的那一对事件一般不再同时。另一参考系要量自己的长度，就须重新选择事件对。两组长度关系因而包含了端点世界线与同时性切面的共同规定。

这使本书提出的“互动活过来”获得一个可以严格展开的意义。互动侧重在这里不是额外增加了一种推动杆变短的神秘作用，而是测量所依赖的信号、同步与事件比较不再能够被当作无条件透明的程序。所谓长度具有关系性，首先意味着其定义包含参考系和事件选择。它并不取消杆的物质结构，也不意味着任意改变仪器就能任意得到合法长度。

主体侧重同样需要精确限定。一个观察者选择坐标系，和一个观察者移动到某处进行物理操作，是不同的行为。前者可能只是对既有事件重新表达，后者可能改变能实际接收的信息与实验安排。相对论的参考系不是人的私人感受。两个研究者完全可以在同一参考系中合作，也可以由一个研究者同时计算多个参考系中的描述。主体位置进入理论，不等于意识成为物理变量。

从发生学看，这里发生了对客体化条件的深化。第 6 章讨论过，经典描述把足够稳定的结果交给公共使用；相对论进一步要求把时空比较的条件写进公共表达。一个长度的可靠性不只来自尺子的精确，还来自事件选择的可说明性。能够在条件变化时重新找到对象，构成比“任何情况下都写同一个数”更强的客观性。

7.2 同时性是一种比较结构

两件相隔遥远的事是否同时，不能只靠站在某一个位置的人“同时看见”来回答。光从不同地点抵达观察者需要时间，接收时刻相同并不自动表示发出时刻相同。一个研究共同体若要给远处事件分配时间，就必须建立信号传递与时钟比较的程序。时间坐标由此不仅是一列刻度，还包含将分散事件组织到同一比较框架中的规则。

可以用一个思想实验说明这一点。设一个站点向远处反射器发出信号，信号返回后，站点保存发出与接收时刻。按照所采用的同步规定，可以给远处反射事件分配一个坐标时刻。关键不在于把人从过程里抹去，而在于所有使用这一规定的人都清楚它如何工作、在什么条件下工作。信号路径、时钟状态和传播模型共同参与时间的可比较性。

一旦这些条件进入定义，“同时”就不再像日常语言那样只是一个无需解释的词。但它也没有退化为任意约定。实际信号必须满足物理规律，时钟比较必须接受实验约束，不同规定必须与其他可观测关系协调。研究者能够选择表达方式，却不能借选择消除局部相遇、信号实际收发以及具有因果关联的事件先后。类空间隔事件的坐标先后仍须按参考系区分。操作规则与物理约束在这里形成相互限制。

从不同参考系选取不同同时性面，并不是对同一事件记录进行随意修改。记录中的局部相遇、信号到达和装置响应可以保持，改变的是如何把相隔的事件归为同一坐标时刻。这个区别很重要，因为它把“描述条件改变”与“对象被干预”分开。若将两者混在一起，就容易说每次换参考系都实际改造了杆，或者说观测者的理解决定了杆的材料状态。

SDE 分析可以在这里追踪一种相互规定。可公开指认的事件是显露的基础；信号往返与比较程序构成差异序列；时钟、传播规律、参考系和同步关系所形成的特征纠缠，使这些事件能够进入一个共同时间结构。新信号与既有比较关系不相容时，研究者可能检查时钟，也可能重建传播模型。哪一种改变成立，需要由进一步证据决定。

同时性的显题化给跨学科研究提供的并非一个可以随处套用的结论，而是一种问题意识。许多领域使用“同一时点”“同一阶段”“同时发生”，却没有说明信息何时产生、何时被接收、何时进入判断。相对论不能直接证明经济或教育中的某种因果关系，但它提醒研究者：比较所使用的时间结构本身值得审查。这个方法论启发必须保持与物理公式的距离。

7.3 从变换关系到共同对象

在最简单的惯性运动情形中，沿运动方向的杆长可以写成一个清楚的关系。设杆在自身静止系中的长度为 L_0，在相对速度为 v 的另一惯性系中按该系同时性测得的长度为 L，则有

$$L = \frac{L_0}{\gamma}, \gamma = \frac{1}{\sqrt{1 - v^2/c^2}}.$$

这条关系给出的是条件化长度之间的转换。它同时保留了对象、相对速度和测量定义。给定这些内容，理论并没有留下“长度可以任意变化”的自由。相反，偏离关系的结果需要检查条件、误差或理论适用性。数值不同不是检验的终点，而是进入共同关系检验的起点。

一个简短推演能够展示条件的重要性。在与杆共动的坐标中，前后端的位置差为 L_0。用洛伦兹变换联系另一惯性系中的端点事件，并要求另一系的两个事件具有相同坐标时间，便会得到上述长度关系。要求同时的步骤不能省略。若先选共动系中同时的端点事件，再把同一对事件直接转换过去，得到的位置差一般不是另一系定义的杆长，因为那两个事件在另一系并不同时。

因此，两个合法读数的差别既不等于仪器误差，也不等于统计不收敛。设想在固定参考系、固定速度和固定测量程序下重复估计 L，只要相应统计条件成立，估计仍然可以越来越精确。把不同速度下的 L 全部混成一个样本，再问它们为何不集中到同一个值，是改变了估计目标。合理的分析可以先按速度分层，或根据模型转换到共同目标，再检验剩余差异。

这个思想对常数讨论尤其有约束。相对论说明某些量随参考系改变，并不说明任何物理常数都必然随实验者改变。一个理论可以同时包含依赖参考系的量和不依赖参考系的结构。若从长度的参考系依赖直接推出所有常数没有稳定意义，就越过了理论实际提供的证据。发生学的建设任务，是辨认哪一层关系保持、哪一层读数改变，而不是把“改变”扩展成没有层次的总判断。

固有长度在这个关系中具有明确地位。它是与杆共动的参考系中规定的长度，不是一个不受任何状态条件影响的神秘本质，也不是相对论已经取消的伪概念。对一根会受温度或应力影响的杆，固有长度仍可能随物质状态改变；这属于状态问题，与在同一状态下更换惯性系所得到的长度变化不同。两种变化可以同时存在，却要分别建模。

可以把这个关系变成一项完全说明条件的计算示例。假定相对速度为光速的五分之三，则洛伦兹因子为一点二五，运动系长度为固有长度的五分之四。这是公式下的演算，不是新增实验。若两组分别报告一和零点八两个归一化长度，把它们当作同一目标的重复估计求平均，得到的零点九没有自动取得特殊地位。它既不是前一组的固有长度，也不是后一组规定的运动系长度。对不同目标求平均，不会因为算术正确便成为物理上适当的综合。

真正的综合需要共同目标。如果选择固有长度，就把第二组读数按已知相对速度转换，并传播速度及长度估计的不确定度。若速度本身来自共享装置，还要考虑共同输入带来的相关性。转换之后仍有差别，才进入关于装置、材料状态或模型的进一步讨论。这个例子说明，发生档案并不只是附加说明文字。它能直接改变哪些数据可以合并、怎样加权，以及合并结果所指何物。保存条件是形成数值判断的组成部分。

共同对象正是在这种区分中得以维持。人们可以承认多种合法表达，并通过变换把它们联系起来。对象同一不需要每个数字同一，而需要不同记录之间存在受到约束的协调关系。属性发生学因而可以把稳定性理解为可转换的持存：在某些差异出现时，仍能通过明确规则认出正在讨论的同一过程。

7.4 不变量为什么比单一读数更有力量

如果每个观察者都只能坚持自己的数字，知识便会分裂为无法沟通的记录。相对论并未停留在多样性上。它以不变量和协变关系组织不同描述，使差异具有共同的约束。时空间隔是其中一个基本例子。在平直时空和所选符号约定下，两事件之间的间隔可以写为

$$s^2 = c^2\Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2.$$

时间差和空间差可以在不同惯性系之间变化，这个组合却保持。这里的稳定性不在某个孤立分量中，而在分量之间的结构中。把不变量称为结构，并不意味着它比实验更抽象而因此更真实；它的力量来自能够协调不同实验表达，减少彼此独立的任意选择。一个框架若允许每个读数单独调整，便很难失败；不变量反而增加了它失败的可能。

为了理解这种约束，可以构造一个记录核查的思想实验。两个小组分别报告同一对事件的时空坐标差，所用惯性系不同。如果按各自坐标计算的间隔不一致，就不能只说双方观察位置不同而结束争论。至少有一处需要检查：他们是否指认了同一对事件，是否使用

相容单位，是否正确转换，或者是否超出了平直时空近似。多视角不是免于核查的理由，而是核查得以展开的材料。

不变量也不意味着一切条件都可以不记。要计算和比较它，仍需知道记录对应什么事件、使用什么模型。它是一种经过组织的独立性：某些表达差别被证明不影响结果，其他差别仍然重要。因而，发现不变量和恢复发生条件并不冲突。前者给出能够跨变化承接的内容，后者说明这种承接如何成立。两者共同支持一种有边界的普遍性。

本书把相对论理解为互动侧重的显露，可以在这里获得更坚实的形式。互动的进入并没有把世界化为一串互不相通的关系，而是促使共同性从单一读数转移到关系约束。变化和稳定被放进同一个表达。若把相对论只当作“读数不同，所以不收敛”，便恰好漏掉了它最深的组织成就：它使不同读数比以往受到更紧密的共同限制。

这一点也为 SDE 提供自我要求。三方程若只意味着所有维度都互相有关，尚不足以形成有力研究。研究必须进一步说明，哪些改变可以一起发生，哪些组合被排除，怎样的转换保持对象同一。在没有明确函数或可检验限制之前，三方程是相互规定的研究要求，而不是能够替代洛伦兹变换的物理方程。它们应推动寻找约束，而不能因写成方程样式就被当作约束已经找到。

7.5 几何进入动力过程

狭义相对论要求重建不同惯性系之间的时空比较，广义相对论进一步使时空几何进入引力的动力学描述。几何不再只是用来放置物质运动的外部网格。关于物质、能量与几何的关系，需要在一个相互制约的体系中求解。这个转变对发生学具有吸引力，因为它使通常被当作条件的结构，也成为需要解释的内容。

这使条件研究出现两层不同任务。第一层由具体物理理论确定度规及其动力学，给出运动、信号传播和测量之间的可计算关系；第二层研究这些关系怎样成为可操作的共同描述。度规属于第一层的明确数学对象，特征纠缠 E 属于本书组织发生分析的概念。两层的连接点在具体实验：几何假设怎样进入信号预测，装置怎样实现比较，哪些偏离会迫使研究者修改对应关系。这样的映射限定了哲学解释可以借助物理证据的范围。

这里可以采用一种有限而有力的重读：广义相对论展示了“研究条件本身如何成为研究对象”的可能性。原来用于描述运动的几何结构，其确定也需要物质分布和边界条件；反过来，运动又受几何约束。这种相互规定不要求人们先寻找一个绝对不被影响的出发点。它要求给出共同成立的方程、适当初始或边界资料，以及可以与观测对接的结果。

引力波提供了一个具体支点。LIGO 与 Virgo 合作组在二〇一六年发表的首个直接探测报告，记载了二〇一五年九月的瞬态信号，并将其与广义相对论中双黑洞并合的波形相比较。[6] 这个例子支持动力几何能够产生可测后果。它把“条件可以变化”推进到更具体的层次：候选变化需要具有可以同记录比较的波形结构。波形在时间中的组合关系限制了任意解释，也使理论与装置共同接受检验。由此获得的是明确物理模型的证据，跨领域的解释仍需建立各自的对应关系。

发生学从这一事实继续提出的是研究组织问题。测得信号、波形模型、装置响应和背景估计如何相互约束？某个候选波形与数据相近时，研究者还须检查哪些替代解释？新事件又怎样改变参数范围和下一轮搜索？这些追问并不减损引力波探测的客观性，而是说明客观结果怎样经过多层检验取得资格。把结论重新接回发生过程，可以增加理解，而无须把它降格为人为制造。

更深一层，动力条件不必意味着失稳。一个相互制约的系统可以支持稳定解、周期过程或复杂变化。是否稳定取决于具体方程和条件。用“背景活起来”概括广义相对论时，应保留这种多样性。活起来所表达的是原先省略的依存关系获得理论地位，不是自然界从此不能保持任何结构。稳定本身也可以由相互制约产生。

7.6 因果结构与观察能力的边界

当观察路径受到限制时，人们容易把无法获取信息解释为因果本身失效。两者并不相同。一个事件的信号暂时无法到达某个观察者，说明该观察者的可知范围受到约束；要进一步说事件没有确定因果关系，需要另外的理论论证。时空描述之所以重要，恰在于它能够把信息可达性、坐标表达和物理关系分开，而不是把它们合成一个笼统的“不可知”。

黑洞视界尤其容易被用作跨领域想象的捷径。它是需要严肃处理的因果边界，却不能因其特殊性便被当作一切局部顺序消失的地点。对常规广义相对论情形，视界、曲率奇点和可能需要量子引力的新尺度是不同问题。要把这些现象理解为号位难以分离，必须分别给出明确模型，并将它们展开为可以回答的物理与解释问题。

这样的限制并不消除本书提出的问题：当原来可以分开的尺度和描述发生相互制约，研究应怎样重新组织？它使问题变得更精确。哪一个近似失效，哪一类观测仍可定义，哪些变量必须共同求解，现有理论在哪些条件下彼此不相容，都可以形成具体任务。仅说“因果秩序被撤销”，却不说明什么观测能区分不同机制，会把研究过早关闭在宏大的结论里。

同样，时间尺度接近并不必然取消可分解分析。两个过程有相近时间常数，仍可能通过不同观测通道辨认；一个过程远快于采样，也可能只留下有效平均作用。可分与不可分取决于耦合、观测设计、目标精度和模型，而非一个单独的速度比较。后续生命编讨论“同轮”时，将继承这个约束：同轮必须围绕实际任务定义，不能成为抹去所有内部差异的口号。

主体侧重在因果边界问题中主要表现为研究者处在何种信息位置。处于不同位置的人可以拥有不同资料，这不等于他们分别生活在任意的私人世界。共同理论需要解释信息差别如何形成，以及在什么条件下可以比较或合并。一个研究共同体若忽略资料可达性的差别，可能把后来获得的信息偷偷放进先前判断；若夸大差别，又可能失去共同验证的可能。

相对论在这里提供了一种克制的雄心：承认位置和路径的重要性，同时坚持位置之间存在公共约束。发生学应把这种雄心带入跨领域问题。所谓主体回到研究，并不是让每种观点免受约束，而是让每种观点说明自己怎样取得信息、承担何种限制，以及如何与其他位置形成可检验的联系。

7.7 从变化波善到可承担的解释

本书用“变化、波、善”表达互动侧重在相对论中的显露，这个对应可以作为哲学解释的入口，却需要展开其不同层次。“变化”可以指读数随参考系而变，也可以指物理状态随过程演化；两者不能直接互换。“波”在引力理论中有具体的动力学意义，在更一般的发生分析中则可能只是传播与关联变化的类比。“善”涉及规范评价，不能由坐标变换公式直接推出。

把各惯性系在物理定律上的平等理解为反对特权视角的启发，是一种可讨论的哲学选择。它提示研究者不要把某个方便的位置误当作唯一可能的位置。但是，从物理理论中的对称性到社会成员应获得怎样的权利，中间仍需要关于人的地位、伤害、责任和制度目的的论证。两个领域可以互相启发，不能互相代交证明。

为了保留这一对应的建设性，可以把它写成三个研究动作。面对变化，先寻找合法转换与剩余差异；面对波，研究影响如何传播、传播是否改变原有条件；面对善，询问研究规则是否允许不同位置提出证据并纠正结论。前两个动作可能在具体物理模型中定量化，第三个动作属于知识共同体的规范组织。它们共同支持研究，却承担不同的理由责任。

这里的“互动活过来”也不应被理解为经典物理完全没有互动，到了相对论才突然出现。前一章已经说明，力、约束和测量路径可以被经典模型明确处理。相对论的特殊贡献，是对某些原先默认的时空比较条件进行重建，并以新的关系协调描述。发生学可以把这称为某种互动侧重的显题化，但不能把显题化写成互动在历史上第一次存在。

若要检验这种解释是否有价值，不能再用相对论已经解释的效应作为独占证据。更合适的评价是：它是否帮助读者发现原先混淆的问题，是否形成更清楚的研究记录，是否使跨理论比较减少偷换，是否提出可由新材料区分的历史判断。解释性增益也是真实贡献，只是它与新物理预测不同。把这一区别说明白，反而能使发生学避免承担它尚未完成的任务。

三方程、六路径和三原理在这里共同约束重读。研究可以从稳定的变换关系进入，追问它组织了怎样的测量序列；也可以从新观测进入，追问条件结构怎样需要修改。但无论从哪里进入，都须返回另外两维，检查新的显露、路径和特征纠缠是否相容。这样，哲学对应能够成为连续工作，而不是在“变化”“波”“善”三个词之间快速滑动。

7.8 转换闭合怎样保证共同判断

不同参考系之间建立联系，最终要使同一材料经不同转换路线仍能得到相容判断。可以设想三个惯性系中的研究者，共享一组带有明确标识的事件记录。甲把记录直接转换到丙的表达，乙则先将甲的记录转为自己的表达，再转到丙。只要各步变换及其相对运动关系彼此相容，两条路线应当描述同一组事件。这种闭合提供了比两组数值看起来相近更强的检查：某一中间环节若引入了不相容的坐标约定，差异会在路线会合时暴露。

这里必须保留第 7.3 节已经建立的区分。转换同一对端点事件，与在新参考系重新选择同时的端点事件，是两种操作。前者保持事件身份，后者构成该系的长度测量。若甲传递的是事件坐标，乙却把其中的位置差直接作为自己参考系中的杆长，路线从这一步开始就改变了目标。后续公式即使计算无误，也只能传递这个错置。共同事件标识的作用，正是让读者能够定位究竟在哪一步发生了重新选择。

一个完整比较因此先保存事件层，再说明量的构造层。事件层包含端点相遇、信号发出与到达等所采用记录；构造层说明如何选择事件对、赋予坐标并计算长度或间隔。这样，当同步约定或输入参数修订时，可以重新构造被比较的量，而无须把全部原始记录改写。对象的共同性由可追溯的转换关系维持，研究者的不同位置则被保存在构造条件中。

不确定度也要沿同一关系传播。若甲到乙、乙到丙的计算共同使用某个速度估计，两步误差一般不能被当作互不相关的新噪声。共同输入可能使两条路线同时偏离，也可能在差值中部分抵消。计量学的不确定度指南已经提供处理相关输入的原则，并提示时间相关观测需要适当方法。[7] 本章的特定任务，是把这些原则落实到转换链：区分独立获得的新约束与对同一输入的重复使用。

可以进一步设想，直接路线与中转路线相容，但二者都与一项独立记录不相容。路线闭合此时只支持内部一致，尚未保证外部准确。研究者需要检查共同校准、共同事件选择和共享模型前提。相反，两条路线不闭合，也不立即意味着时空理论失效；它首先要求寻找单位、同步、相对速度关系或事件配对中出现分歧的位置。这种分层定位使争议能够逐步缩小，而不是在“全部正确”与“新物理出现”之间跳跃。

真正的新预测要对这种已经严格组织的关系增加约束。如果一个候选修正预言某种条件下仍有额外差异，应说明它改变哪一段转换、为什么不会被共同输入解释，以及在哪一项独立比较中留下后果。若研究的成果是使原先混淆的事件重新可比，其价值属于知识组织和测量设计；若产生额外而可复核的经验限制，才进一步进入物理机制的竞争。两种推进可以接续，各有明确完成条件。

由此，相对论给本书提供了第二种稳定性结构。第 6 章通过控制与交换取得局部省略权，本章则通过变换闭合与不变量组织条件差别。稳定性可以在多种合法读数之间成立，而不要求先消除这些读数的差别。量子测量将提出另一层问题：即使制备与装置安排已经固定，单次结果仍可不同；共同判断因此还需要在事件、分布与测量关系之间重新定位。

第 8 章 量子测量中的装置与结果

8.1 从一次点击到一种可重复关系

探测器出现一次点击，屏幕留下一个斑点，计数器增加一个单位，这些事件看起来比抽象的波函数更确定。研究者首先拥有的是发生过的记录，而不是一份已经解释完毕的微观世界图景。但是，单次记录也不能独自说明它为何出现。背景事件、装置响应、制备过程和选择的可观测量，都参与决定一个记录能够支持怎样的判断。量子研究从未只靠点击本身完成，它依靠的是点击怎样在一套受控制的关系中反复出现。

这里的稳定性经常表现为分布的可重复，而不是每一次结果都相同。两次按相同程序制备和测量，可能得到不同单次结果；大量记录所形成的频率结构，却可以受到理论的严格约束。将单次变化直接叫作“不收敛”，会混淆对象层次。需要先问：被要求收敛的是某个单次结果，结果分布的估计，还是跨实验装置的共同参数？这些问题有不同的数学含义。

以一个两结果的理想测量模型为例。每次实验记录“是”或“否”，若条件固定且重复满足适当独立性，某一结果的经验频率可以趋近稳定概率。即使稳定概率是二分之一，单次记录也不会逐渐变成半次点击。概率估计的收敛和单次事件的随机性能够并存。这个简单区别，足以阻止把量子不确定性直接翻译成所有测量都无法稳定的结论。

装置在这一关系中具有实际构成作用。它决定哪些相互作用被转换为记录，怎样的信号被接受为事件，以及哪些结果可以被区分。制备同样重要，因为不同制备对应不同状态。研究者选择可观测量，必须通过装置安排实现。主体侧重因而可以通过目标选择、实验程序和记录规则进入分析，但这些作用并不要求意识在最后一刻向系统施加一种未写入理论的力量。

本书所说的“主体活过来”，可以理解为测量安排和判断条件不再适合被看作完全透明的通道。这个理解保留了思想的锋利处：结果的意义不能只向一个孤立对象索取。同时，它需要承认量子理论已经发展出处理制备、演化和测量的丰富结构。属性发生学的工作，是把这些结构中的发生关系讲清楚，并提出可以继续研究的问题；它不能只通过改名宣布已有理论失去了全部解释力。

从点击到统计关系，再从统计关系到对象判断，存在连续而不同的证明责任。哪一种记录属于目标事件，频率估计是否可靠，模型是否能够在新条件下预测，最后又支持怎样

的存在论解释，不能在一段话里一并结算。量子测量之所以成为发生学的重要边界案例，就在于它迫使这些平时容易折叠的层次重新展开。

8.2 装置选择怎样进入问题

同一个制备过程可以接受不同测量。这种差别不只是同一把尺换了颜色，而可能改变被询问的可观测量。为了避免把抽象讨论变成关于人心的猜想，可以采用一个理想两态系统。沿一种选择测量，结果有两类；改用另一种不相容选择，得到的统计关系可能改变。这里的改变由状态与测量算符共同规定，不需要假定操作者心中对答案的期待直接成为物理原因。

装置选择具有两种相连但不相同的意义。它在物理上实施某种耦合，使系统信息进入记录；它在研究上确定结果被解释为何种量。一个方向旋钮既可以改变实际测量安排，也代表研究者对问题的选择。把这两个层次分开，才能说明主体如何参与而不夸大其作用。研究者决定转动旋钮，并不意味着转动后的结果可以由意愿指定。

还应区分“选择哪一种测量”和“在看到结果后选择保存哪些记录”。后一种选择可能改变最终分析的条件样本。若研究报告只展示通过筛选的事件，读者需要知道筛选规则；否则，有条件分布容易被误当成全部实验的分布。量子实验中的后选择具有明确用途，但它并不免除样本条件的说明。任何领域中，结果选择与过程干预混在一起，都会使因果判断变得困难。

可以构造一个方法学例子。两个小组使用同一制备和同一探测器，其中一组保存全部有效记录，另一组只保留与辅助信号同时出现的记录。两组结果不同，并不能立刻证明装置产生了不同物理规律。它们可能是在比较不同条件下的分布。要判断是否存在新增效应，需要把筛选条件、事件配对和未选择记录重新放回分析。这一工作属于测量对象的明确化，不是对异常的敷衍。

装置中还沉淀着过去的判断。哪些自由度被视为无关，哪些事件被合并，什么时间间隔算作符合，怎样判断探测成功，通常已经进入硬件或软件。它们让实验能够高速重复，也使研究者可能忘记自己正在使用一种特定安排。发生学将其展开，目的不是让每次读数重新接受无限争论，而是保留在结果异常或用途变化时追溯这些判断的能力。

因此，装置可以被看作一种条件化的提问形式。它把理论允许的区分转成物理上可执行的区分，同时接受材料、噪声、效率和耦合强度的限制。一个好问题需要一个能承受其

证明责任的装置。主体号位的建设性作用正在这里：让问题、实现和解释相互校准，使世界的回应能够纠正原来的提问。

8.3 不确定性与估计精度的分层

量子不确定关系最容易被日常语言带偏。“测不准”似乎意味着仪器不够精密，或者重复多少次都不能改善知识。一般的态标准差关系并不是这个意思。对同一量子态中的两个可观测量，它约束相应概率分布的宽度。Robertson 的一般关系以两算符对易子的期望为右端；位置和动量的常见形式则具有熟悉的下限。[8]

$$\sigma_x \sigma_p \geq \frac{\hbar}{2}.$$

这里的标准差首先属于态所规定的结果分布，不是研究者对分布参数估计的标准误，也不是某台仪器的刻度误差。假如独立重复满足通常统计条件，一个分布均值的估计可以随样本量增加而改善，尽管单次结果的分布仍然具有非零宽度。分布不变窄与对分布认识更精确，可以同时成立。把这两者区别开，量子论与统计收敛之间的表面冲突就会消失。

前述两结果模型可以把这一区分进一步量化。若结果以零和一编码，概率各为二分之一，单次结果的标准差为二分之一；在独立同分布的理想条件下，一百次样本均值的标准误为零点零五，一万次则为零点零零五。增加样本所改善的是概率估计，而非把单次结果的分布压窄。这里的计算只使用假定的统计模型。若资料存在相关、制备漂移或事件筛选改变，标准误就须按实际条件重算。两批频率是否相近回答分布稳定性问题，系统是否预先具有全部确定属性则属于另一层解释，数值精度本身没有跨越两层的证明力。

更进一步，样本量也不能无条件补救目标漂移。如果前半段测量一种制备，后半段测量另一种制备，合并频率可能趋于一个稳定混合值，却不能代表任一制备的概率。稳定估计仍可能回答了错误的问题。这个反例使发生学与统计学形成正面联系：记录发生条件既帮助解释变化，也帮助识别一种看似成功的收敛是否具有所声称的意义。

还需要区分测量误差和测量扰动。一个装置对某量的测量偏差，与它对另一量后续状态造成的改变，不是态标准差的同一个概念。它们可以建立关系，却必须依明确的测量模型定义。用同一个“Δ”表示全部不确定性，容易把不同命题互相替换。两项结果的分布不能同时变得任意窄，与两个分布参数的估计不能同时改善，是不同命题。只有把它们分别写出，才能知道一次实验究竟检验了哪一项限制。

弱测量实验使这一接缝更加可见。Rozema 等人的实验讨论了朴素测量误差与扰动乘积关系的失效，并与修正关系比较；它没有推翻态标准差的 Robertson 关系。[9] 新的测量方案可以改变信息获取和扰动的分配，但若比较的量已经改变，就不能把新的数值代入旧命题之后宣布原理被否定。严谨区分并不削弱实验的意义，反而说明它究竟推进了哪一个问题。

这一分层对发生学具有直接价值。S 可以指单次记录，也可以指稳定分布或参数估计，但一次分析必须说清楚当前指的是哪一层。D 可以包含重复制备、测量和估计程序，E 则包含使这些重复可比较的特征纠缠。若每一轮制备状态已经改变，经验频率的变化可能反映条件漂移；若条件固定，频率波动可能仍符合统计模型。两个情形不能仅凭“都在变化”而归为同一机制。

因此，量子不确定关系不是“不收敛病理学”的现成证明。它提供的是一种更精确的启发：某些稳定性不能通过把全部单次属性同时压成确定数值来获得，研究应寻找分布、关系和受约束的可测结构。这个启发能够推动新的提问，却不授权把已有数学命题改写成它没有说过的内容。

8.4 测量相互作用与记录的形成

一个理想化的测量模型，可以把系统与装置从起初不相关的状态，演化为结果与指针相互关联的状态。若系统有两个可区分的状态，装置能够对它们分别留下不同记录，那么装置便取得了区分能力。这个模型的重点不在于宣称真实仪器只有两个自由度，而在于展示：结果可被记录，需要系统与记录载体之间建立受控制的相关关系。

$$(a|0\rangle+b|1\rangle)|M\rangle \to a|0\rangle|M_0\rangle+b|1\rangle|M_1\rangle.$$

这里还必须区分物理学的量子纠缠与本书的特征纠缠 E。对复合纯态，量子纠缠具有不能分解为各子系统状态直积的数学含义。在上面的理想模型中，两个系数均非零、两种装置记录彼此正交而可区分时，右端不能写成一个系统纯态与一个装置纯态的简单直积。[10] 特征纠缠 E 则是更宽的发生分析概念，关注使显露与差异序列相互规定的组织。同名不意味着 E 自动具有这种量子态性质，也不意味着量子纠缠的实验证据可以直接证明 E 在其他领域的主张。二者只有经过明确映射，才可能在一个具体研究中形成有限联系。

这行理想模型同时展示了测量的成就与难题。右端把不同系统状态与不同装置记录联系起来，因此后续读取装置可以获得信息。但仅仅写出这种关联，并不自动回答为什么一

次实验经验中出现一个确定结果。对复合系统的演化描述、对记录概率的计算，以及对单一结果的解释，仍是不同问题。它们彼此关联，却不能由一句“装置已经参与”便全部解决。

从属性发生学看，记录并不是客体属性外面的一张复印件。记录需要材料载体、放大、保持和读取；它也需要能够区分有效事件与背景的程序。微观相互作用只有在适当安排下，才能成为研究共同体可以共享的证据。这个过程具有客体、互动和主体三种侧重，但它们仍然属于同一次发生的不同分析入口。记录载体不是纯粹客体，读取规则也不是脱离物质实现的纯粹主体。

可以设想一个不可靠记录器，它在不同微观结果之后总显示同一个符号。这样的装置可能稳定，却没有完成当前所需区分。也可以设想一个记录器能够区分结果，却无法保存到读取时刻。它具有瞬时响应，却没有形成可承接证据。由此可见，稳定性不能脱离用途评价。科学需要的是与目标差异相联系的稳定记录，而不是任何不变状态。

记录还可以反过来进入控制。在一种预先规定的实验中，第一轮结果决定下一轮制备或设置，这时形成了实际的跨轮反馈。与此不同，若研究者只在实验结束后重新整理记录，改变的是认识和分析过程，不一定改变此前已经发生的系统演化。两种回写都值得研究，但需要不同的时间记录和因果证据。量子测量的全部意义不能被压缩成“每一次看都会改变下一次”。

三原理可以指导寻找尚未协调的环节。可见记录与目标区分不匹配，可能要求改变记录条件；操作路径与装置能力不匹配，可能推动新的显露方式；既有记录与条件假设冲突，可能推动新的制备和检验序列。这些方向没有预先指定哪一种量子解释正确。它们要求把哲学问题转为能够推进的研究动作，让每次修正都能接受独立检查。

8.5 退相干提供了什么又留下什么

真实装置处于复杂环境中，系统与装置所形成的相关关系还会扩展到更多自由度。对部分自由度不再逐一追踪时，某些相干效应在所研究的子系统描述中受到抑制。退相干研究给出了理解经典记录稳定性的一条重要路径。它使人们能够讨论哪些状态更能维持可记录性，以及环境怎样参与这种选择。[10]

这是一项建设性进展，因为它不再把环境仅仅视为需要删除的噪声来源。环境可以破坏某些相干关系，也可以使某些记录形式更容易持存。发生学从中得到的启发，是稳定性与可见性有时由相互作用共同形成。隔离并非唯一的稳定办法，某种有选择的耦合也可能

产生稳定的公共线索。不过，这一启发需要落实到具体动力学，不能被推广为所有耦合都能生成有效记录。

退相干也不能被草率写成测量问题已经彻底结束。对子系统约化状态的解释、单一结果如何理解、概率如何获得意义，仍会进入不同理论解释之间的讨论。指出这些问题，并不等于退相干没有解释贡献。科学研究经常通过解决一个有边界的子问题推进整体。一个机制能够说明某类干涉为何不可见，就已经提供了约束；它不需要同时承担关于一切存在的最后答案。

记录的时间结构由几种不同关系共同决定：相干效应受到抑制需要多久，装置记录保持多久，以及研究者在什么时间分辨率上读取。退相干速度依系统、耦合、状态差异与环境条件而变，在某些模型中可以极快。[10] 一个适合读取的窗口，应使目标记录已经形成，又尚未因后续变化失去区分能力。窗口是否存在，需要具体模型与装置共同说明。因而，跨领域比较应逐项比较形成、保持和读取的时间，而不能用“物理慢、生命快”替代这些关系。

从方法上看，环境信息不易恢复和信息原则上不存在，也需要区分。研究者可用的通道有限，导致某些相干关系对当前任务不可见；这一事实支持一种有效描述。若以后出现更精细的操控，描述层次可能改变。发生学所说的显露，正可以用来追踪这种有条件的可见性，但不能把暂时不可见直接说成从未存在，也不能把数学上可写出的关系都当作当前可测对象。

“场”在本书的视觉 SDE 解释中，可以引导研究者关注记录能够成立的关联范围。但光学相干长度、环境耦合范围、探测器时间分辨率与主体可获得的信息范围，是不同概念。它们可能共同影响实验，却没有一条普遍等式把它们合成一个场宽度。保留差别，能够使这一解释从诗性联想转向细致分析：究竟哪一种范围被改变，改变如何传到记录，又如何影响下一轮判断。

8.6 计算共同性与解释分歧

量子理论的计算成功与其解释争论可以同时存在。不同解释可能在广泛条件下给出相同观测预测，也可能在扩展模型中产生可检验差异。讨论解释时，首先应问比较的是计算规则、本体承诺还是新增动力学。若这些层次不断移动，一个解释在经验上获得的成功，就容易被错误转移给另一套尚未产生独立预测的哲学语言。

Bell 的工作为这一问题提供了重要约束。它针对具有特定局域性条件的隐变量构造建立限制，并不排除所有可能的隐变量理论。[11] 因此，解释比较须把所用局域性条件和其余前提分别写出，再判断各方案承担了什么不同代价。一个排除结论的价值在于精确缩小允许范围；尚被允许的方案则继续接受一致性与经验约束。这样，Bell 的结果成为组织分歧的共同条件，而不是把开放问题一次清空的宣告。

同样，把某种解释简单归为“主体独裁”“客体独裁”或“互动独裁”，可以作为阅读线索，却不能代替对其最好版本的重建。解释可能已经包含观察者、环境、记录和相关结构，只是赋予它们不同地位。若仅凭它没有采用本书的词汇，就认定它忽略了某个号位，争论会成为命名权竞争。更有力的问题是：在什么明确任务中，它遗漏了必要的约束或产生了与证据不符的结果？

属性纠缠论因此需要承认一种可能：在某些物理问题上，它提供新的组织方式，却没有提供与已有量子理论不同的预测。这个结果并非毫无价值。新的组织方式可以帮助区分被混淆的层次，改善记录或使跨领域迁移更谨慎。但它不能据此声称已经解决坍缩、概率或量子引力。贡献的边界被说明后，后续研究才能知道应该增加什么，而不是一再宣告已经完成。

一个公平的比较可以围绕同一实验任务展开。先写明各解释共同接受的制备、装置与记录，再列出它们各自增加的假设。如果新增假设只改变叙述，没有改变任何可观察限制，比较主要属于解释的一致性、简洁性和问题生成能力；如果新增动力学改变统计，就需要独立实验检验。两类评价都应公开，不能用哲学偏好代替数据，也不能用暂时观测等价宣布所有解释永远同样好。

这种竞争纪律同样适用于本书。三方程要求相互规定，六路径帮助进入问题，三原理追踪张力与回写；只有当它们进一步约束了可出现的关系，或改善了研究组织的可检验结果，才形成明确增益。量子解释的开放性是一项需要承担的工作，不是任何宏大框架都可以自由填入的空白。

8.7 分布场美的建设性重读

本书以“分布、场、美”描述主体侧重，并把观察装置作为这一侧重的物质界面。[12] 这个思想最有价值的部分，是将主体性从孤立的内心活动扩展到选择如何被实现、可能性如何被保留、判断如何取得后果。若仅把主体理解为某个人是否睁眼，许多真正进入实验的决定反而会被遗漏。装置设计和研究程序保存的判断，值得成为哲学分析的对象。

“分布”可以首先追问判断权与信息分配。谁决定记录哪些结果，谁能改变设置，谁知道筛选规则，谁有资格把某个信号解释为有效事件？这些问题属于研究系统的组织，并不等同于量子态的概率分布。两种分布可以通过操作发生联系，例如某种决策改变了实际测量设置，但联系需要被说明。相同的词不能自动建立因果桥梁。

“场”可以追问一个判断能够作用的条件范围。一个研究者拥有的装置、时序控制和可获得信息，决定其可以实施哪些区分。这个范围不是单纯的外部环境，也包括已经建立的技术与概念关系。把这些关系纳入 E，有助于解释为什么同一理论问题在不同研究条件下会形成不同实验。与此同时，E 不能被简化为仪器清单；它需要表现这些条件怎样相互规定问题的可能性。

“美”则应当谨慎地保留其关于未完成可能性的哲学含义。一次研究在作出判断之前，可能同时保存若干候选解释、实验路径和问题方向。它们尚未被结算为最后答案，却可能对下一轮研究有价值。将这种开放性称为美，可以形成关于科研创造的规范性论证。但它不等于物理势能，也不意味着量子测量普遍把势能转换成真与善。物理能量与可能性的价值属于不同概念层次。

这种重读能够导向实际设计。一个研究档案不仅保存获胜解释，也保存被排除解释为何失败、哪些分支因资料不足暂缓、哪些条件变化可能使它们重新相关。这样，未决可能性能够进入未来行动，而不是只作为浪漫的无限开放存在。主体侧重的成熟性，表现为既能作出必要判断，也能保存判断尚未覆盖的边界。拒绝永远结算与过早结算，同样可能妨碍研究。

三属性并不因此分别占有分布、场和美。每一种主体侧重仍有自己的显露、差异序列和特征纠缠；客体与互动侧重也同样如此。一个保存备选解释的档案具有可见内容、更新路径和维持条件。将这三个维度写清楚，才能讨论它怎样影响判断。只在装置旁贴上三个词，既没有解释量子结果，也没有形成新的科研方法。

8.8 从结果分布到可靠记录链

量子测量中的稳定性至少需要两段关系同时成立：制备与测量安排规定结果分布，结果与装置记录之间保持可说明的对应。第 8.3 节讨论前一段中的分布宽度和估计精度，第 8.4 节讨论相关记录如何形成。把两段连接起来，才能解释为什么一种稳定的仪器输出有时可靠，有时却稳定地抹去了目标差异。这也使本章的发生学分析取得具体落点：追踪区分能力怎样在记录链中被获得、保存或丢失。

可以在前面的理想两结果模型之后，再附加一个明确假设的读出模型。系统结果已按所采用量子测量模型给出，记录器随后有一定概率把零记成一，也有一定概率把一记成零。这里只分析记录错误，不以这个经典误差模型解释量子结果为何发生。若两种错误概率能够通过独立校准确定，观测到的一的频率便可以用来约束原结果概率；校准不确定度也必须随之进入估计。这比笼统声称装置参与测量，多给出了一段可以实际检查的转换关系。

这个模型还能显示信息丢失的条件。设想记录器无论接收零还是一，都各以一半概率输出两个符号。那么输出频率会稳定趋近二分之一，却完全不能区分输入概率。增加记录次数只会让这种无区分能力的输出更加精确。若记录器对两类输入仍有不同响应，才可能恢复一定区分力；响应越接近，反演通常越容易受读出误差影响。稳定性、精度和辨识能力由此被放在不同位置，各自需要相应证据。

另一种丢失发生在筛选环节。若只保存满足辅助条件的点击，最终频率针对的是被选择的事件集合。即使读出没有错误，这个条件频率也不能直接替代全部制备事件的概率。研究者应保存总尝试次数、有效记录数和筛选规则，才能判断差别发生在系统统计、记录效率还是选择过程。这里要保留的条件由具体转换决定，无须把与当前推断无关的一切现场细节都收入模型。

这条链也为装置改进确定了清楚目标。降低错误概率，可以改善已有量的读出；扩大可区分设置，可以使新的测量任务成为可做；把先前结果送回下一轮设置，则改变跨轮实验设计。三种工作增加的能力不同，报告应说明改善落在哪一段。仪器增益可以非常重要，其物理意义仍由它支持的预测与检验决定。新增统计关系若由已知测量理论和读出修正充分说明，就保留这项明确成就；若仍有额外关系，则另行建立竞争机制。

主体侧重在链条中获得的是选择与解释的可定位责任。研究者决定采用什么制备、允许什么设置、保留什么记录，再让输出反过来约束这些决定。E 体现各项安排使哪些区分可实现，D 体现区分经过怎样的过程进入记录，S 体现当前可以承接的结果。它们共同组织一项测量，却不因使用了同一套语言便解决量子解释的全部问题。发生学的积极任务，是使一项判断所需的关系保持连续。

于是，经典条件控制可以在量子实验中继续发挥作用，但稳定性的承载处已经进一步分层：单次事件留下记录，重复试验约束分布，装置校准保证两者之间的联系。在此基础上，第 9 章将处理另一项工作：多个不同过程的记录，怎样共同支持一个不能直接摆在面前的理论对象。

第 9 章 量子场论与客体身份的形成

9.1 看见记录怎样成为认识对象

一个在探测器中短暂出现的事件，怎样支持关于某种粒子的判断？日常经验习惯于先有一件持续存在的东西，再观察它留下的痕迹；高能物理中的许多对象却不能以这种方式摆在桌面上。研究者需要从产生条件、衰变产物、运动关系和统计结构中识别它们。对象的知识身份通过一组相互约束的证据形成，而不是由单张漂亮的事件图独自保证。

这并不意味着粒子只是研究者发明的名称。一个名称能否成为科学对象，要看它能否把不同记录联系起来，能否限制尚未观察的结果，能否在装置和分析改变后仍然得到支持。任意命名容易，建立这种承接关系困难。发生学对客体身份的追问，应当落在困难所在：不同证据如何共同指向同一种对象，什么结果又会迫使人们拆开原来的同一性。

可以构造一个简单例子。某项实验的分布中出现一段超出预期的聚集。它可能提示新的物理过程，也可能来自背景模型不足、选择条件或装置响应。仅仅把聚集命名为一个新对象，并没有增加解释力。研究者需要寻找独立关系：改变选择后聚集怎样变化，其他通道中是否出现相容结构，相关运动量是否满足候选模型限制。对象的稳定身份来自这种多方向约束。

同一性在这里具有层次。某一个碰撞事件是否被重复计数，是记录层面的同一；一组事件是否属于同一产生过程，是模型层面的分类；不同过程是否反映同一种粒子，是理论对象的同一。三者可能相互支持，却不能互相替代。一个事件标识清楚，并不自动证明其物理分类正确；一种分类统计稳定，也未必意味着关于微观实体的全部解释已经确定。

主体侧重进入对象形成的方式，首先是提出候选分类和安排检验。互动侧重表现为产生、传播、衰变和探测路径，客体侧重表现为能够跨这些路径识别的稳定关系。它们不是三个预先完整的实体相加。所谓粒子的可识别性质，恰在不同过程与不同记录之间取得公共意义。研究者参与了识别条件的建立，却不能凭参与取消外部约束。

因此，“客体作为发生物返回物理学”应当被理解为一个需要展开的研究命题。它可以指物理对象在过程中产生或衰变，也可以指研究对象的身份通过证据逐渐成立，还可以指理论把对象重新放进更深层结构。三种发生相关而不相同。把它们区分开，才能既承认知识建构的工作，又不把知识形成误写成意识创造粒子。

9.2 从孤立小球到场的描述

量子场论为粒子与相互作用提供了不同于一群经典小球的理论语言。在适当情形下，粒子可以作为场的量子激发得到描述，产生与湮灭也能够进入同一框架。这个转变的重要性，不在于把小球换成另一种看不见的物质，而在于对象身份与动力结构、对称关系和可能过程更加紧密地联系起来。对象是什么，与它能够怎样出现和相互作用，不再容易分开叙述。

然而，场也不能被当作已经理解的一切之母。“粒子是场的激发”是一种有条件的理论概括，不能自动消除态、观测和相互作用中的困难。特别在强相互作用、复杂背景或不同近似下，粒子描述的适当性需要具体讨论。本书不以一句场的口号替代这些技术问题，而从中提取一个有限的哲学方向：客体身份可以依赖关系结构，同时保持可检验的稳定性。

想象一段振动能够在某种介质中持续传播，人们可以在不同地点识别同一种模式。这个经典类比有助于理解“稳定形式与承载结构不完全相同”，但不能直接当作量子场论的证明。量子场不必等于某种机械介质，粒子也不是水面上的小波包。类比的作用是打开问题：一种可辨认对象能否以过程方式存在？答案仍需由各自理论给出。

过程式身份与稳定并不冲突。一种激发若完全没有可重复的性质，就很难成为对象；一种对象若只能在特定条件下稳定，也不因此失去科学意义。关键是稳定范围及其转换关系能够被说明。由此看，发生并非稳定的反义词。某些稳定性正是由特定组织持续维持或由特定对称关系限制而成立。属性发生学应研究这种成立方式，而非把一切固定性质视为旧范式残余。

场的语言还促使人们重新思考边界。日常物体的边界常由表面给出，理论对象的边界却可能由可区分的自由度、守恒关系或可操作范围给出。这些边界不必对应一个可触摸外壳。研究者需要说明哪些变化仍属于同一对象状态，哪些变化改变对象数目或类别，以及哪些描述差别只是表达方式不同。边界由此成为理论与实验共同承担的工作。

在 SDE 框架中，不能因为 E 被称为场，就把量子场直接等同于特征纠缠。量子场是物理理论中的特定对象；E 则关注使显露与差异序列相互规定的组织。某个场的状态、装置的耦合和理论的可观察限制，都可能参与具体 E 的分析。它们怎样共同作用需要逐项说明。把两个词合并，会同时失去物理的精确性和发生学的分析幅度。

9.3 多通道证据怎样加固对象身份

一个对象若只能依靠单一观测通道出现，关于它的判断往往更依赖该通道的模型。不同通道提供相容证据，可以增强对象身份，但这种增强不是简单把数量相加。不同通道是否真正独立，是否共享校准或理论输入，各自对候选对象有什么不同敏感性，都影响综合的意义。多样证据的价值，在于它们可能以不同方式约束同一主张。

ATLAS 合作组二〇一二年的新粒子报告提供了一个具体例子。研究合并多个产生与衰变相关搜索，给出新中性玻色子的证据，并说明观察结果与标准模型 Higgs 粒子的产生和衰变相容。[13] 这里需要保留“观察证据”“模型相容”与“全部性质已确定”的区别。发现对象与继续测定对象性质，是相连而未必同时完成的研究阶段。

发生学可以从这一结构提出更一般的问题。一个新对象的身份往往并非在第一次异常出现时一次决定，而是在不同测量持续排除替代解释的过程中变得更稳定。最初的名称可以只是候选，后来得到更多限制，某些早期表述还可能需要调整。成熟对象并不因此摆脱发生史；它的成熟恰恰包含知道哪些判断已经牢固，哪些仍属于开放属性。

设想一种候选对象能够解释某个峰，却对另一通道给出过多事件的预测，而后者没有出现。这个反例会迫使研究者缩小或改变候选解释。若框架为了保留对象而随意为每个通道增加独立参数，它也许仍能拟合数据，却失去了跨通道约束。对象理论的力量不在于每处都能补洞，而在于同一组关系能够承担多处检验。失败通道应进入对象形成史，而不是被从叙事中删除。

还有一种不同情形：多个通道都出现相容结构，但共享某个关键校准，因此共同偏差仍可能存在。这要求研究者寻找新的实现路径或独立检查。所谓独立不是团队名字不同，也不是仪器外观不同，而是导致错误的关键机制不完全相同。跨通道一致性只有在误差结构得到说明时，才能获得应有的证据重量。

对象识别还会遭遇一个循环风险：研究者根据候选对象设计筛选，再用筛选后的资料证明候选对象存在。这个循环并不必然无效，因为理论引导观察是研究的正常部分；关键是筛选规则是否产生了能够独立检验的后果。可以在看见目标结果之前固定主要规则，用背景区域检查装置与模型，再在独立资料中复核。这里提出的是方法设计，而不是声称所有实际发现都遵循同一程序。它把对象期待与对象证据之间的关系公开，使期待能够指导工作，也能够被工作纠正。

此外，事件图的直观性和证据的强度需要分别评价。一幅图可以清楚展示某次碰撞的记录结构，却不能单独告诉我们这种结构在背景中有多常见。统计证据需要比较发生概率，物理解释需要比较可产生这些记录的过程。将两者接起来，对象才从一个引人注目的形象成为承受反例的主张。发生学关注显露，不能只偏爱最容易看见的显露；它还须研究那些通过分布、比较和缺失结果才能显现的约束。

本书由此将客体身份理解为一种不断加固的约束中心。不同路径围绕它形成共同预测，不同装置让它承受不同风险，新的观测使其属性边界更清楚。这个中心不是任意投票形成，也不必被想象成脱离一切过程的微小实体。它是一种可以在反对证据面前承担后果的公共对象。对象建构与对象抵抗在这里共同发生。

9.4 重整化与能标依赖

同一种相互作用在不同能标下可以使用不同的有效耦合值描述，这一事实常被读成常数本身失去意义。更准确的理解，是参数的定义与所采用尺度需要共同说明。能标改变，理论把不同贡献重新组织到参数和显式计算中；可观察预测必须保持相容。数值的尺度依赖因而不是对任意变化的许可，而是有规律的联系。

精细结构常数的跑动是一个重要例子。粒子数据组的电弱综述将低能值、能标依赖及相关量子修正置于明确框架中讨论。[14] 这里的能标变化不同于宇宙随时间演化导致常数改变，也不同于实验室之间对共同目标的估计存在张力。三种问题都可能使用“α 变化”的表述，但它们需要不同数据、不同模型和不同否定条件。

重整化在哲学上值得研究，首先因为它使参数的定义与物理预测的关系更明确。这里需要分开实际过程的能量条件与计算中采用的重整化尺度。前者改变时，过程的可观察结果本来就可能改变；后者改变时，参数与其余计算贡献应按理论相互配合，以描述同一物理过程。因而，“保持相容”要求保持预测关系的协调，并非要求所有能量下的观测数值相等。这一区分使能标依赖既能够被研究，也不会变成随意调整参数的自由。

可以用一个抽象的计算思想模型说明这种协调责任。设同一观测量的预测由一个按尺度定义的参数和一项显式修正共同给出。改变参数采用的尺度时，若只更新参数而保留原修正，便混合了两套计算规定；若两部分按同一规则一起更新，才有资格检验它们是否仍描述同一观测量。这个模型只展示一致性要求，不替代具体场论计算。它把属性讨论的重心从“哪一个参数数字最真实”，移到“哪些配套关系共同承担可观察预测”。

对于属性发生学，真正值得保留的是参数与实现关系的历史。报告某个数时，应明确采用的定义、尺度和方案；比较两项结果时，应先确认是否已转换到共同目标。若忽略这一点，理论已经预言的差异会被误当作异常；若把所有差异都归给定义，又可能掩盖真正张力。有效分析必须同时避免过度惊讶和过度吸收。

尺度依赖还要求重新思考“根属性”的语言。一项参数不是脱离理论便能自行说话的数字，但它也不因此成为纯粹约定。它受到可观察关系的限制。两套不同参数化若给出相同预测，研究者可以讨论各自便利；若预测不同，实验具有裁决作用。发生学应把注意力放在这条从表示到约束的链上，使客体身份在不同描述中保持可核查联系。

9.5 常数张力怎样转成研究问题

为什么精细结构常数取某个值，是关于理论解释的深问题；不同实验对同一目标是否相容，是计量与模型检验问题；它是否随时间变化，则是另一个经验假说。三者可以相互启发，但没有一条简单推理把其中一个问题的未解决自动转为另一个问题的肯定答案。理论尚不能解释某个数值，不意味着该值已被证明随主体而变。

CODATA 的常数调整以多项实验与理论关系组织共同估计，其中一些输入存在需要处理的张力。[15] 这说明最精密的研究仍会遭遇不一致，也说明共同估计本身是一项持续工作。它不能单独决定张力来自未识别系统效应、理论输入、数据处理还是新物理。把这些可能性保持开放，并为它们设计不同检验，比立即宣布一个总体本体论结论更能推进问题。

一项发生学再分析可以从资料身份开始。不同方法测量的是同一参数，还是通过不同观测方程反演参数？它们共享哪些理论输入，使用哪些单位定义，哪些不确定度来自共同来源？把这些联系写清楚，可以发现表面独立的数据其实部分相关，也可能发现一种真正有价值的独立比较。三号位记录的作用，是让这些结构可见，而不是预先规定结果一定分裂。

原子钟比较又提供另一种任务：利用频率比的历时资料约束可能的常数时间变化。Lange 等人的研究属于这一类检验。[16] 它与改变能标考察跑动不同。即使两种研究都涉及 α，也不能把所有数值按发表年份排成一列，便称其为宇宙时间演化。发表时间、观测时间、理论版本和目标定义各自具有不同地位。

可以构造一个再分析思想实验。某方法在新一代装置投入后出现中心值移动，同时研究者也更新了输入理论。若只看年份，会得到一条方向性变化；若同时记录装置与理论版

本，就会发现时间与方法完全纠缠。此时，资料还不能区分真实时间变化和方法改变。需要交叠时期的不同方法、独立校准，或者对理论输入的专门检查。记录更多字段的收益，是揭示哪些结论暂时不可识别。

若经过共同目标转换、完整不确定度分析和独立重复，仍有稳定张力，它便成为值得继续研究的对象。发生学可以提出有关路径和条件的新候选机制，但这些机制必须承担额外预测。若某个参数只是补回遗漏校准量，成果属于测量改进；若它在独立资料中揭示新的物理关系，才可能产生更深理论意义。两者都值得承认，不能以追求宏大结论压低具体改进的价值。

9.6 有效理论与有边界的真实

一种理论能够在某个范围内极其成功，同时不被宣称为全部尺度的终极描述。有效场论把这类思想发展为有技术内容的研究方式。Weinberg 关于有效场论的回顾，讨论了标准模型与广义相对论作为更一般描述中低阶结构的可能地位。[17] 对本书而言，关键不是从中宣布某一终极图景，而是理解有限适用范围怎样仍然能够产生严格知识。

“有效”容易被误解为只是勉强有用。实际上，一个有明确范围和误差控制的描述，可能比声称解释一切却没有可检验限制的体系更可靠。它说明保留哪些贡献、忽略哪些贡献，以及精度提高时需要怎样扩展。边界不是事后失败的借口，而可以进入研究的事前承诺。知道何时不该使用一个模型，本身是模型成熟的一部分。

设想对某过程的近似描述，只在一个小参数足够小时保留前几项。研究者在扩大范围之前，应先检查遗漏项可能达到的量级。若已经超出控制范围，不能以原模型预测失效证明自然界突然改变了法则。相反，若在承诺范围内出现超过误差预算的系统偏离，模型就需要承担压力。这个思想模型说明，评价失败必须同时知道模型曾经承诺什么。

属性发生学可以从这里重建“条件恒常”的意义。一个稳定参数可能属于某个有效层次，在该层次内支持广泛预测。更深层描述改变它的解释，并不必然取消其有效性。例如，一个宏观量可以由更复杂过程组织而成，但在适当尺度仍是最有用的变量。由构成复杂性直接否定宏观稳定，会破坏不同层次之间的知识承接。

同时，有效理论也要求防止范围无限后退。若每次失败都通过缩小适用域而保留理论，最后可能只剩对已知资料的描述。一个有生命力的研究纲领，应说明缩域之后仍能完成什么新任务，并提出如何重新扩展。条件化不是永不失败的保护壳，它需要与预测能力、独立检验和研究收益一起评价。

跨尺度承接也必须保留误差责任。一个模型在粗尺度上把多种内部过程合并为同一状态，换到细尺度时，原先等价的状态可能分开。细尺度发现差异，并不自动否定粗尺度模型在原任务中的成功；粗尺度模型成功，也不授权它判断所有细节。两层模型需要说明相互转换后哪些量应当一致，哪些信息已经在压缩中不可恢复。这样的关系比把两种理论简单排成真假等级更能组织后续研究。

这给三方程的成熟性研究提供了一条清楚道路。相互规定并不意味着任何问题都必须把全部维度同时精细求解。某些任务可以合法地固定或近似处理部分关系，只要保留近似的理由和退出条件。成熟模型具有受控制的省略，新的发生分析在省略失效时帮助重新展开。如此，统一框架与局部有效理论可以相互支持，而不必互相取消。

9.7 真空质量与生成问题的限度

量子场论中的真空不是日常语言所说的绝对空无。它在具体理论与状态条件中具有确定意义，不能只凭“空”这个字来判断应否具有结构。另一方面，把真空称为有结构，也不等于已经解决真空能与引力之间的全部问题。理论量的定义、重整化处理、引力作用和观测推断之间需要逐项连接。一个极大的数值差异如果没有说明估计前提，就不能独自承担整体哲学结论。

本书从真空问题提出的发生学追问，是“被当作无背景的条件，实际包含哪些结构”。这一追问可以帮助识别默认设定，却不能代替具体计算。研究者若认为某种状态结构参与客体性质的形成，必须说明可观察后果及与已有模型的区别。没有这一步，“万物从场中发生”仍然只是宽泛语言，尚不足以组织独立的研究计划。

质量的讨论同样需要层次。某个理论机制参与解释某类基本粒子的质量，并不意味着日常物体的全部质量都可用一句相同故事交代。复合系统、相互作用能和基本参数之间存在不同关系。本书不把任何单一机制升格为全部属性生成的通用答案，而把它看作一个范例：原先直接赋予对象的性质，可以在更深理论中获得部分构成解释。

这种解释会改变我们关于客体的想象。一个量曾经只是模型输入，后来可能与其他结构建立关系；一个对象曾经按稳定标签分类，后来可能显示与状态或对称结构的联系。科学因此不仅增加数字精度，也会改变“什么需要解释”的边界。发生学的任务，是记录这类边界移动，并分析它怎样打开新的问题，而不是在每次移动后宣布旧对象从未真实。

可以区分三个生成问题。第一，某个过程怎样产生一个可记录事件；第二，一种性质怎样由理论中的关系获得说明；第三，一个研究共同体怎样建立关于该性质的可信知识。

第一是过程机制，第二是理论构成，第三是知识发生。它们可能在同一研究中交织，却没有一个可以直接替代另外两个。探测器制造了记录，不意味着它制造了关于粒子的全部性质；理论解释了性质关系，也不意味着实验实现已无需研究。

三号位分析在这里获得一种正面任务：把三个生成问题连接起来，同时防止它们互相吞没。客体侧重要求性质之间有稳定限制，互动侧重要求产生与识别路径可说明，主体侧重要求问题和证据规则能够反思与修改。每种侧重都需要自己的 SDE 展开。只有这样，客体的生成性才能成为可研究的内容，而不是对一切实体词语的普遍怀疑。

9.8 对象身份怎样约束整条证据链

第 8 章考察结果怎样经过装置成为记录，本章进一步追问记录怎样共同指向对象。原始事件经过校准与选择形成可分析量，可分析量约束理论参数，参数又必须在不同产生和衰变过程中承担联合预测。对象身份的力量，恰在于它使这些环节不再能够彼此任意调整。一项解释若只在每条通道中分别拟合良好，却无法说明通道之间的联系，仍未完成同一对象的构造。

可以设想两种衰变通道都出现候选结构。研究者先用第一通道约束某些对象参数，再根据同一模型推导第二通道应有的分布或范围。第二通道中的背景和装置响应仍需分别处理，但不能为了保住对象而任意放弃共同参数所带来的限制。联合分析的额外价值，来自一条通道中的资料缩小了另一条通道允许的结果。这种约束传递，是共同对象超出两个相似峰形的地方。

反过来，一条通道中的缺失也可能有内容。如果模型在已声明条件下要求第二通道出现可识别结构，而相应资料排除了该范围，研究者就需检查候选对象、产生与衰变关系或实验响应。若模型本来允许这一通道极弱，缺失便不能承担同样否定力。负结果的作用因此取决于对象模型预先承诺了什么。对象身份逐渐加固，也就是相关承诺越来越明确，能够承担的反对证据越来越具体。

证据链还存在另一种关键区分：对象参数与装置参数各自能在多大范围内被调整。设想两条通道共享某个能量校准，二者的相容可能部分来自共同输入；再加入一种具有不同响应路径的观测，才能对这个共同来源施加新限制。多通道并不天然独立，独立性要沿误差产生的路径判断。把共享输入显题化，使对象的统一获得更严格的检验，而不是因发现共同条件便宣布所有证据无效。

尺度变化把同一责任扩展到理论表示。参数按新尺度重写之后，比较应在共同观测预测处会合；若实验过程的能量本身改变，则应检验模型预言的条件响应。前一种工作维护表示的一致性，后一种工作考察对象在新过程中的表现。区分它们，才能让精细结构常数的跑动、实验估计张力与时间变化各自保有清楚问题。一个统一名词只有在这些不同任务之间保存转换规则，才增加知识。

在这样的研究中，六路径可以表现为不同的约束取得次序。已有候选对象时，从其允许过程安排搜索；已有异常结构时，先检查能够维持它的条件，再寻找其他过程；已有新装置时，先确定它能区分哪些关系，再决定值得建立的对象。入口可以不同，最后都要把局部结果放回联合限制。三原理的作用则是在一处不相容出现时，定位需要重建的是可见结构、搜索路径还是使它们可比的条件，并检查修订怎样改变其余承诺。

由此得到的客体身份具有可以积累的厚度。一次观测可能只支持某个结构存在，后续资料再约束它的性质与作用范围；部分参数修订时，其他已经获得支持的关系仍可保留。这样的成长允许理论对象既真实地受到外部限制，又拥有尚未完成的认识历史。对象建构的正面成果，是形成越来越难以随意改写、同时仍可接受新证据的约束中心。

本编至此展现了四种彼此交叠的稳定性工作：经典研究通过控制取得有根据的省略，相对论通过变换组织不同描述，量子测量通过制备与记录约束统计关系，场论则以跨过程和跨尺度的限制支持对象身份。第 10 章所需新增的工作，是解释这些认识形式怎样进入历史叙述：什么材料足以表明一个条件被显题化，哪些变化只是表达方式更新，以及旧知识如何在新问题中继续承担责任。

第 10 章 物理学史的发生学重读

10.1 历史能够支持怎样的哲学主张

一部科学史可以按照理论名称排列，也可以按照问题如何改变来书写。前一种写法告诉我们哪些理论先后出现，后一种写法追问：原先可以不问的条件，为何后来成为不得不问的对象？本书重读物理学史，采取的是后一种方向。它关心属性怎样获得稳定的科学地位，测量怎样参与这一过程，以及理论如何在新的困难面前重新安排对象、路径与判断。

由此提出的核心解释是：物理学的发展使属性发生中的不同侧重逐渐显题化。经典客体化突出可承接的对象性质，相对论重建时空比较条件，量子测量使制备和装置安排更明确地进入结果解释，量子场论又加深了对象身份与理论结构的联系。这是一条有组织力的阅读主线，但它不是宣称四种理论分别只拥有一个维度，也不是宣称历史必然朝本书的词汇前进。

历史解释要承担两种不同责任。它需要忠实于过去研究者面对的问题、能够使用的工具和实际提出的主张；它也可以用今天的概念发现跨时期的关系。两者之间需要一条清楚界线。当本书说某个理论使互动侧重显题化，这是对该理论工作的重读，不是替它的创造者补写一份他们从未发表的宣言。重读的价值，要由它能否解释更多细节、产生更好的历史问题来证明。

一个反面的写法，是先决定三个号位必须按某种顺序苏醒，再把每一段历史填进相应格子。这样总能找到例子，却很难发现反例。另一种写法则从实际问题出发，考察哪些条件被固定、哪些变化造成压力、哪些新关系被建立，再问三号位分析是否有助于解释这些工作。后者允许一段历史同时显露多种侧重，也允许某个变化与预设顺序不同。

科学的成熟并不总以公开危机出现。很多概念改变来自长期校准、仪器改进、计算方法和跨实验比较。若只寻找戏剧性的革命，就会把维持可靠知识的工作排除在发生之外。属性稳定性既需要新理论，也需要反复维护。一个共同常数值、一种可重复制备或一条稳定校准链，都有自己的形成史。发生学应当能够书写这些不那么壮观的成就。

因此，物理史不是本书已经获胜的证书，而是一块能够承受细致检验的材料。它让“属性是发生的稳定性”面对真实知识形成过程，要求这一命题解释成功、失败、延续和转向。只有四者都能进入叙事，属性发生学才有可能成为科学史的分析方法，而非给著名理论重新贴上哲学标签。

10.2 被省略的条件怎样进入历史证据

第 6 章说明了经典研究怎样取得有根据的省略权。历史研究需要继续问：一项省略何时已经成熟到可以不再逐次解释，又在何种情形下被重新打开？这要求把理论文本与实践材料放在不同证据位置。一篇论文省去了某项操作，可能因为作者没有认识它，也可能因为同行已将它作为共同惯例，还可能因为出版形式只容纳最终推导。仅凭文字长短无法在这些解释之间作出判断。

可以构造一个档案分析的思想案例。第一份操作说明详细交代支点处理，第二份简报只列参数，第三份争议记录又反复讨论支点。如果只看三份材料的篇幅，容易写成条件先显现、后消失、再复活的故事。若第二份简报默认采用第一份程序，其省略却可能表明操作已被共同承接。第三份记录重新展开支点，则可能源于用途扩大或实现条件改变。这里需要查明的不是某个词出现多少次，而是省略依靠什么关系维持，以及后来哪一项关系不再足够。

条件的隐含地位因此至少有三种。它可以尚未被注意，可以已被理解并稳定控制，也可以因组织分工而由别处承担。三种隐含都可能在最终文本中表现为沉默，却具有不同历史意义。第一种适合研究发现，第二种适合研究成熟与迁移，第三种适合研究知识劳动的分配。发生学把它们分开，便能够解释同样简洁的公式为什么可能依赖很不相同的实践基础。

重建这种基础，需要沿知识的使用关系寻找材料。例如，参数报告援引哪一版操作说明，接手者是否需要额外训练，出现争议时首先追查哪个环节，以及修订之后哪些用途恢复。它们共同表明一项省略是否具有稳定的恢复通道。如果没有这些材料，历史解释可以把某种机制列为候选，却应保留其他可能。证据的不完整在这里形成研究问题，而不应被填成理论所期待的完整故事。

这也给主体侧重的历史位置提供了更细的说明。判断可能从某位实验者的临场选择，转入共享程序，再转入装置和校准关系。随着判断形式改变，个人在现场需要解释的内容减少，而维护共同程序所需的劳动可能增加。这样的变化是判断的组织方式改变。研究可以追踪决定在哪里作出、由谁复核、何时需要重新讨论，从而避免把操作自动化直接写成主体性消失。

所谓经典客体化的历史成熟，便表现为对象能够携带已被验证的适用条件进入新的工作，而无需在每处从头建立。它的失败也有具体形状：被携带的条件与实际用途脱节，或

者恢复通道失去可用材料。本章在第 6 章之上增加的，是区分这些历史状态的证据问题。对象稳定的历史不仅包括参数如何被发现，也包括参数如何被传递、何时值得信任，以及何时需要恢复其背后的工作。

10.3 问题单位的改变如何被辨认

在爱因斯坦一九〇五年的论述中，时间与同时性的操作说明进入理论展开的起点。[5] 第 7 章已经分析其物理内容，此处关注的是这一文本安排承担的论证工作：它使原先看来可以直接使用的时间规定，本身成为需要说明的对象。这个判断来自可阅读的论证结构；关于作者最初怎样想到问题、哪些经历最关键，则需要另外的历史材料。两类提问相连，却不能仅凭成文顺序相互推出。

问题单位改变，可以通过三种相互支持的文本迹象辨认。首先，旧提问中的某个默认词取得新定义；其次，新定义要求读者改变可以合法比较的记录；最后，推导后果依赖这项改变，而非仅在开头增添哲学说明。以同时性为例，其意义不只在于出现了观察者和时钟，而在于事件的选择方式进入长度与时间的理论关系。历史解释因而可以追踪一项定义怎样改变后续问题的可成立形式。

这种追踪不同于寻找一句最像今天术语的话。假如早期文本已经提到观察位置，不能仅凭名称相近就认定它完成了相同重建；假如后来文本使用不同词汇，也可能承担相似论证任务。需要比较的是定义、推理和实践之间的功能关系。概念史中的连续性由这种关系支持，词语的连续只提供入口。发生学的显题化解释应在此接受材料的约束。

可以设想两种历史解释竞争。第一种把转变归于测量条件在理论中的重新定位，第二种强调理论结构之间的一致性要求。它们可能指向同一文本的不同侧面，未必互相排斥。研究应进一步问：哪一种解释能说明为何恰好这些条件进入论证，而其他条件仍被省略？哪一种能说明定义变化带来了哪些后续限制？若两者需要共同作用，就应把这种联合组织写清楚，而不是为了突出号位显题化删去数学与物理问题自身的压力。

问题单位的变化还应与传播范围区分。一篇论文建立了新的表达，并不自动意味着实验操作、教学或工程计算同时改变。要研究这些后续变化，需要考察新表达怎样被转写为训练、计算规则和仪器比较。它们可能拥有不同节奏，旧表达也可能在限定范围继续有效。因此，理论显题化、实践采用和公共常识之间可以存在时间差，而这种时间差正是知识发生的材料。

由此，互动侧重的显题化具有一种可以争论的历史含义：原先承接结果所依赖的某段比较关系，获得了不可省略的理论地位。这个解释的力量取决于它能否说明论证任务的变化，而非能否把全部相对论效应重新命名。它为后面的量子概念分析留下一个新的问题：有时并非旧问题的默认条件改变，而是一个旧名称被分化为若干需要分别研究的问题。

10.4 概念分化怎样改变研究的方向

科学史常以一个理论替代另一个理论来组织转折，但概念分化提供了另一种成长形式。一个长期使用的词可能同时覆盖几种不同限制；当这些限制获得各自的数学对象和实验任务，旧知识的一部分被保留，新问题则在其旁边形成。第 8 章关于态展宽、测量误差、扰动和估计精度的区分，为这种历史分析提供了一个具体入口。

Robertson 的态标准差关系与 Rozema 等人研究的误差—扰动关系，具有不同的比较对象。[8][9] 两项工作之间的关系不能仅用“后者推翻前者”表达。更有解释力的追问是：同一个“不确定性”名称在不同论证中指向了什么，装置如何参与各自量的定义，以及何种实验结果能够限制哪一项命题。这样的重建既保留已有结果，也显示新研究究竟打开了什么空间。

这里已有的材料足以支持命题对象的区别，却不足以独自复原两项工作之间的完整思想史。若进一步研究概念怎样分化，就需要比较原始定义、后来重述以及相应实验的任务说明。特别要避免用今天已经清楚的区分，推断所有早期研究者都以同样方式理解问题。一个概念在后来取得精确表达，既可能澄清早期意图，也可能改变早期问题本身。两者应通过具体文本关系判断。

概念分化对历史叙事有一个重要后果：学科的增长不必表现为原理论整体退出。态标准差关系可以继续约束分布，关于装置误差与扰动的研究可以发展自己的模型，统计推断又继续处理有限样本。它们共同增加研究能力，却没有一个简单的总尺度能把三者排成越来越接近最后答案的序列。成长表现为不同问题获得更明确的承接方式，以及原来混合在一起的证据责任得到分开。

这种分化也会改变争论的资源分配。若争论双方实际上回答不同问题，继续增加同类数据未必能结束争论；明确量的定义之后，实验才能被分派到相应命题。若双方回答相同问题，却采用不同装置模型，才需要围绕模型差异设计比较。历史研究可以据此追踪争论何时由解释口号转为任务分工，以及这种转变是否产生新的可操作区分。这是一项关于研究方向怎样形成的判断，而不仅是术语整理。

主体侧重在这里表现为提问形式被重新组织，互动侧重表现为实验安排取得更细的作用位置，客体侧重表现为每项命题能够稳定指向自己的对象。三种侧重共同参与概念成长，不能把某一实验仅因包含装置便划为互动胜利，也不能把概率讨论全部归给主体。发生学的历史贡献，是追踪区分怎样产生、怎样被使用以及怎样改变后续研究，而让每一项物理命题继续承担自身的证明责任。

10.5 新理论怎样继承旧知识

理论更新之后，旧知识如何继续有效，是物理史必须解释的另一条主线。第 9 章讨论有效理论的有限适用范围，此处需要进一步区分几种继承关系：旧结论在限定条件下仍可计算，旧测量可以在新模型中重新解释，旧对象名称继续使用却改变了构成说明。这些关系可能同时存在，但各自需要不同证据。仅仅说新理论“包含”旧理论，尚未说明实际继承了什么。

Weinberg 关于有效场论的讨论表明，一种描述可以在明确尺度中保持严格意义，同时容许更一般的理论结构。[17] 对历史解释而言，这提供了研究连续性的一个方向：科学成长可以增加旧结果的成立理由，也可以重画其适用边界。更深的说明与更精密的数值并非同一种进展。某项新理论在当时尚未改善预测精度，也可能改变哪些关系被视为需要解释；这类改变应按其实际贡献进入历史叙述。

可以用一个理论比较的思想模型把继承关系写清楚。设旧模型在某个控制范围内给出一组成功预测，新模型在更广范围给出预测。如果二者在原范围内按已说明的近似相容，旧计算便有继续使用的根据。若新模型只与旧资料拟合相近，却没有交代近似如何成立，继承关系仍需研究。数值接近是线索，参数对应、误差范围和对象定义共同说明这种接近为何值得信任。

历史材料还需要区分理论继承与工具继承。即使对象的解释改变，某种仪器校准、数据处理或实验技能仍可能被继续使用；反过来，理论语言变化不大，实际装置却可能已全面重建。两种轨迹若被合并，科学史就容易只剩概念名称的更替。发生学可以沿具体工作追问：新研究从过去带走了哪些可用约束，又在哪个环节必须重新取得资格。

对象身份同样有多种延续方式。一个名称可能保持而性质范围改变，也可能名称改变而部分观测关系保持。判断是否仍在讨论“同一对象”，要看理论与实验怎样连接这些阶段。把名称连续直接当作实体连续，或把名称改变直接当作旧对象消失，都会跳过真正困

难的工作。历史中的对象由不断调整的证据联系承接，这种承接既允许真实修订，也保留已经经受检验的内容。

由此可将成熟性理解为可说明的继承能力。新理论能够指出旧结果何以有效、何处需要修改；旧方法能够在限定任务中继续工作，并向更广研究提供材料。三属性成长在这里不表现为越来越多名词进入同一段说明，而表现为对象、路径与判断之间的联系更加可追踪。这个尺度允许历史同时书写突破和延续，也使新典范必须解释自己准备继承什么，而不仅是宣布自己打算超越什么。

10.6 六路径怎样进入历史研究

六路径是研究进入次序，不能直接当作历史必然次序。一个理论在成书时可能先给定义再推结果，在真实研究中却可能先遇到异常，再反复修改概念。论文叙述、实验操作和问题形成具有不同时间结构。若仅按发表文本的章节顺序推断真实发生，就会把经过整理的论证误当作发现过程的完整记录。

以从显露进入条件的路径为例，研究者可能先注意到某个无法稳定复现的结果，再检查装置、样本和计算规则。这个过程具有 S—E—D 的进入型特征，因为新路径是在条件检查之后形成。但另一小组也可能先建立新的实验条件，进而发现以前不可见的结果，表现出不同入口。两种路径可能围绕同一对象相遇，并共同改变理论。历史不是只允许一条合法道路。

从条件出发的沉淀型研究尤其值得关注。长期建立的技术、计算和共同标准，会让某些问题逐渐变得可处理。新的显露不是突然从无处出现，而是沿已有条件中的差异序列被组织出来。这样的研究往往不如革命故事显眼，却对科学成熟不可缺少。发生学若只讨论异常如何倒逼理论，会低估准备条件的创造劳动。

原型型与再创型又展示另一类关系。一个试探性操作可能先形成可见样本，随后研究者才认识到需要怎样的稳定条件；一个已经确立的对象，也可能经新路径获得新的组织方式。历史中的技术与理论并不总是谁先谁后的固定关系。它们可以在不同阶段互相提供问题、工具和限制。六路径的联合使用正是为了防止把某一入口绝对化。

开展具体历史研究时，可以为一个问题保存几种时间记录：异常首次被注意，候选解释首次写出，关键装置形成，独立检验完成，概念进入共同使用。它们不一定重合。研究者还应保存失败路径，因为失败常常解释为什么后来选择某一种方案。只保留成功结果，会使发生过程看起来比实际更直、更必然。

三原理则用于分析转折中的压力。现有显露与操作路径不相容，可能推动条件重建；路径与条件冲突，可能催生新的显露；显露与条件判断冲突，可能要求新路径。这样的分析需要指出具体不相容是什么，改变怎样发生，以及改变如何回写前两维。若无法给出这些内容，就不应把某次理论更替直接称为原理的实现。框架应帮助发现联系，而非代替联系的证据。

10.7 显题化有三种不同的历史时间

前面几节已经呈现三种需要分别追踪的时间：某种关系在实际过程中成立的时间，研究者认识并表述它的时间，以及它进入稳定记录与公共使用的时间。它们可以相互接近，也可能相隔很久。发生学若把三者排成一条“属性逐渐出生”的历史，就会把知识增长误写成世界等待理论命名。更适当的做法，是研究三种时间怎样交错，并指出每一层可以获得什么证据。

以记录稳定性为例，退相干理论为某些相干效应受抑制以及记录形式的选择提供了动力学研究。[10] 物理过程的时间尺度属于这种模型与实验所约束的问题；相应解释什么时候进入共同认识，则属于知识史；研究团队何时开始在某类记录中保存相关条件，又是实践史。即使第一层过程极快，也不能据此判断后两层同步完成。物理速度与知识成熟速度需要各自的度量与材料。

一个仪器改进的思想案例可以展示这种交错。某种效应原来已经发生，却被粗分辨率平均掉；新装置使它能够被区分，随后新模型才说明其关系。在另一种可能历史里，理论先给出可区分预测，研究者再建设装置取得记录。两种路径都可能使相关条件显题化，但其创造发生在不同环节。比较它们能够帮助解释科学为何有时由新观察推动，有时由新计算或新技术可能性推动。

这也是从物理案例进入生命、神经与 AI 研究时值得携带的区分。后续各编可能研究系统自身形成新组织能力，也可能研究人们取得新观测通道，还可能研究已有材料被新的问题重新解释。这三类增长都可以有价值，却不应被同一个“主体成长”指标混合。判断对象究竟变了什么，需要先定位变化属于哪一层，再考察它怎样改变另外两层的任务。

历史比较还应追踪未发生的承接。一项理论认识可能长期没有可用记录，一种新装置可能产生大量数据却尚无适当问题，一个稳定操作也可能缺少完整解释。这些不同时并非一概属于失败。它们可能构成后续研究的条件储备，也可能暴露某个方向始终没有取得可行动关系。显题化假说需要解释这些停留与错位，才有能力超出只挑选成功结局的叙事。

因此，本编与后续各编的联系是一组可追问的关系，而不是物理先于生命、生命高于物理的等级安排。物理案例提供了区分控制、转换、统计与身份的严格材料；其他领域将检验这些区分在新对象上需要怎样改变。历史解释本身也由此保持开放：它研究知识与能力如何相互承接，同时允许某项认识已经成熟、某项操作仍待形成，以及某个宏大对应尚未获得实质内容。

10.8 一部开放的显题化历史

属性发生学对物理史的主张，可以被表达为一种开放的历史假说：科学在处理持续困难时，常会把原来被稳定封装的条件重新展开，使客体、互动或主体侧重获得新的理论与实验地位。这个假说不要求每个领域按同一顺序发展，也不保证展开之后必然获得更好理论。它预期的是某种可考察的工作结构，而非一条预先写好的胜利路线。

要检验这个假说，可以选择相互差异足够大的历史案例，比较问题出现前后哪些条件进入论证核心。研究者应同时寻找不符合预期的案例：困难是否有时由简单校准解决，而没有发生概念重建？某个理论是否在没有明显异常压力时由数学或技术可能性推动？某种新语言是否增加解释吸引力，却没有改变实验工作？反例能够帮助缩小主张，使显题化不成为对所有变化的通用命名。

这种研究可以先从一个规模有限的比较开始。对每个历史案例，分别整理问题出现时的原始表述、后来解决方案的论证结构，以及新方法进入实践之后的操作变化。三组资料不应被后来获得的结论一并改写。早期研究者认为困难来自哪里，后来研究者怎样重述这个困难，实际操作又改变了什么，可能出现不一致。正是这些不一致，能够显示理论认识、实验记录与构成条件不是同步成长的。

研究还可以邀请熟悉不同领域的读者分别标记关键转折，再比较分歧。这里需要的不是把复杂历史强行编码成一个总分，而是定位解释在哪些段落上发生争议。一位读者认为新装置只是提高精度，另一位认为它改变了被测量定义，双方就必须指出各自根据的操作材料。这样的比较让 SDE 解释接受领域知识的压力，也避免解释者只挑选最符合自己语言的历史片段。

如果比较发现某些理论变化主要来自表示简化，却没有新增经验约束，这一结果仍然重要。表示简化可能降低计算成本，允许新的推导，也可能让多个原本分散的问题进入同一框架。概念组织收益需要按这些后果评价，而不必伪装成新机制。若某次变化进一步导

致了独立实验预测，则可以另行登记经验增量。把不同贡献分别承认，历史叙事就不必把每一项成就都写成一次全面革命。

历史解释还需要公平处理稳定时期。一个长期有效的标准、一套成熟模型或一种可以被不同团队接续的操作，不应只在后来失效时才进入叙事。它们怎样维持、怎样积累条件知识、怎样培养新的问题，同样构成发生史。稳定不是历史暂停，而是某些关系持续被维护。维护中的细小修正，也可能为后来变化提供关键材料。

本书由此提出一种成熟性观：成熟框架既能够稳定回答已有问题，也能够明确保留自身的开放边界。它不会因发现条件依赖就放弃全部共同知识，也不会因已有成功就禁止问题单位改变。客体性、互动性和主体性在这样的研究中不是竞争王位，而是共同承担发生的不同责任。任何一维被概念上省略，都需要说明省略的范围和恢复通道。

这一历史重读也给本书自身设定要求。三属性、三方程、六路径和三原理不能因为能够重新讲述过去，就被宣布已形成新的科学典范。它们还需要在当代研究中帮助产生问题、明确对象、设计检验，并允许结果改变框架。历史解释是支持研究纲领的一种理由，不能替代研究纲领的实际工作。真正的推进将发生在下一次具体判断中。

这种写法还要求版本意识。同一个问题在早期札记、正式论文和后来的教材中可能具有不同边界。引用哪一个版本，必须服务于当前所问的历史问题。用晚期的清楚表达解释早期的模糊处可以帮助理解，却不能反过来证明研究者一开始已经拥有全部后来知识。

本编的历史解释由此获得一个明确结论：属性的科学稳定性并非一个无需说明的起点，它有条件、有路径、有公开承接的历史。理解这段历史，会改变我们对不收敛、误差和异常的提问方式。第 3 编将从理论史返回测量现场，把这里获得的区分转成读数事件、装置分析、数学判据和模型比较，使发生学进入能够被执行、被纠正的科研程序。

第三编 测量 读数事件与不收敛

第 11 章 从属性值到读数事件

11.1 一个数字究竟说出了什么

一张表上写着某杯溶液的甜度为七。这个数可以用于安排配方，也可以用于比较受试者，还可以被解释成产品在市场中的接受程度。三个用途看似共享同一数据，实际却提出了不同问题。配方比较关心在规定程序下改变浓度会怎样改变评分；受试者比较关心同样安排下不同人的反应；接受程度还涉及价格、饮用场合和选择行为。如果数字离开了生成它的问题，越精确的记载越容易制造错误的确定感。小数位不能替代用途，统一单位也不能保证判断的是同一种东西。

测量并非先取得一个完全中性的数，再由研究者任意附加意义。操作开始以前，研究者已经选择了样品、时间、参照和响应方式。读数出现以后，又要经过校正、汇总与解释，才成为可以进入论证的结果。将这些步骤统称为背景，会把影响结果身份的条件与暂时可以忽略的细节混在一起。发生学的入口在于重新区分：哪些条件规定我们在测什么，哪些条件只是提高测量的精度，哪些条件会被测量结果反过来改变。

设想两名研究者分别报告同一水槽的温度。一人将探头放在表面，另一人放在底部；一人在加热刚停止时读取，另一人在充分搅拌后读取。数值不同不必意味着至少一人出错。若目标是整个水槽的均匀平衡温度，两次程序可能尚未实现目标；若目标是空间温度分布，它们可能都是有效的局部记录。问题不能仅靠求平均解决，因为平均以前就需要判断这些读数究竟属于一个总体、多个位置，还是不同过程阶段。这里的目标不是由数值自行宣布的，而由对象、安排与提问共同规定。

国际计量词汇把被测量理解为拟测量的量，并区分测量结果、不确定度、计量可比性与兼容性。这个区分已经说明，现代计量学不是把裸数当作全部知识。[3] 本书由此继续推进的地方，是把条件从一次报告的说明项延伸为跨轮发生的组成部分：探头可能改变局部状态，读取方式可能改变行动，公布结果可能改变下一次安排。测量不只是给已有状态加一个名称，也可能参与状态怎样继续形成。

这种推进并不要求每个读数都附带世界的完整历史。一杯水的温度记录无须收集研究者全部生活经验，甜度评价也不必保存所有心理活动。所需记录取决于具体争议和任务。若两种测量程序的差异已被可靠换算，它们可以共同支持一个目标量；若某段历史在允许

误差内不改变判断，就有理由将其压缩。读数事件的要求是有根据地决定什么可以省略，而不是把任何省略都宣布为压抑。

“七”因此不是失去价值的符号。它是在某组条件下形成、承担有限任务并可能进入下一轮行动的结果。只有弄清这些关系，研究者才知道什么时候可以把它与另一个七放在一起，什么时候必须保留差别。数字的力量来自可追踪的关系，而非来自脱离关系后的孤立。测量的对象性也正是在这样的追踪中获得公共检验的形式。

11.2 事件的边界与对象的同一性

把读数放回事件，首先遇到的困难不是字段太少，而是事件从哪里开始、在哪里结束。以一次品尝为例，端起杯子、接触溶液、吞咽或吐出、选择评分，都可以看作动作边界。但前一次刺激可能仍在影响当前反应，期待可能在看到包装时形成，评分又可能成为下一轮比较的锚点。若把事件只截取为按键的一刻，便会丢失解释当前输出所必需的过程。若把事件无限扩展到人的全部历史，分析又无法开始。

事件边界应当由待解决的差异决定。研究配方的即时效应，可以先把当前刺激呈现到评分结束作为核心窗口；研究适应则需要连接相邻品尝；研究长期训练则要保留多次会话之间的更新。边界不是随意的叙事选择，而是一个可受检验的假设：窗口之外的信息是否仍能系统改变我们对窗口内结果的解释或预测。如果增加前史显著改变归因，原边界就过窄；如果不断增加历史并不带来任务收益，扩张便没有充分理由。

对象的同一性同样需要分层。配方相同可以支持“同一名义刺激”，却不保证温度、口腔状态和呈现过程相同；同一个人可以支持身份连续，却不保证状态恒定；同一套量表可以支持形式一致，却不保证使用者对刻度的解释不变。我们不需要因此放弃“同一对象”，而要说明同一性落实在哪一层。它可能是材料同一、制度身份同一、功能同一或经过转换后的结构同一。不同层次之间不能未经论证地相互代替。

考虑一种明确虚构的教育场景：学生两次完成同一道推理题，第二次答对了。若研究问题是题目得分稳定性，第二次见到相同题目是一种历史污染；若问题是错误反馈怎样促进理解，这段历史正是研究对象。相同的操作在不同任务中具有不同地位。事件化记录不预先决定它应被控制还是保留，而是让这种决定进入可公开审查的论证。研究者必须说明为何记忆在此处属于干扰、在彼处属于机制。

事件划分还影响样本单位。一个人连续给出十次评分，并不自动等于十个彼此独立的人。十次响应可以提高对个人轨迹的了解，却不能以同样方式增加对人群差异的了解。将

过程中的多个点当作相互独立的样本，会夸大资料对某些问题的支持；反过来，仅把整个过程压成一个平均数，也可能浪费轨迹中的信息。恰当做法是同时保留事件内部结构与事件之间的组织方式，在分析时让它们承担各自能够承担的推论。

由此可见，读数事件不是一个把数字装进去的容器。它是对象身份、时间边界、允许比较和推论范围共同构成的单位。边界可以修订，同一性可以分层，比较协议也可以改变，但修订必须留下理由。前一版本不能被新版本悄悄覆盖，否则历史差异会在档案中消失，研究者只能看到最后一次安排所认可的对象。这种保留不是文书负担，而是判断知识怎样形成的必要材料。

11.3 三种侧重怎样进入同一次记录

客体性、互动性与主体性侧重不能被当作三类互不相干的变量盒子。浓度看似属于客体，实际必须通过取样、校准与单位约定才能成为可比较的浓度；时间曲线看似属于互动，却依赖什么响应被记录以及谁能够持续报告；期待看似属于主体，也须借助问题措辞、表达方式和既有经验才显现。每个候选指标都在一次共同发生中取得自己的位置，没有哪个指标可以凭名称免除对发生条件的说明。

这并不妨碍做分析区分。研究者可以先问，哪些稳定性使材料能够被反复指认，哪些路径把材料与响应接起来，哪些判断位置决定什么被注意和报告。三种提问各有侧重，却持续相互约束。例如一个评分装置把作答时间限制为两秒，既改变互动节奏，也改变主体能够使用的判断策略，并影响最终被称为能力的客体化结果。将时限单独放进“情境变量”并不能自动消除它对目标量身份的影响。

第 1 至 5 章建立的概念区分，在记录层面提出一个具体要求：同一字段可以参与不同侧重的说明，但须保存它实际怎样取得。例如“反应时间”既可能用于描述装置延迟，也可能用于分析判断策略；前一用途需要设备时钟和触发记录，后一用途还需要题目要求与作答规则。字段名称相同，不足以保证证据可以互换。记录应把直接指示、计算变换和理论用途连起来，使后来者能够从解释退回取得程序，再判断是否接受这一用途。

仍以评分装置为例，从显露进入可以先发现两秒时限下的答错集中在哪些题型，再追问造成这些差异的组织条件。从差异序列进入，可以比较先解释再选择与先选择再解释的不同路径。从特征纠缠进入，则可研究题目表达、工具能力与评分制度怎样共同规定可采取的策略。三条进入路线最后应在同一资料和机制上相遇；如果每条路线都只得到一组独立标签，分析尚未说明各侧重怎样共同形成结果。

主体侧重尤其需要有限而明确的代理。预期报告、参照点选择、题意复述与纠错行为，都可能帮助分析某种判断位置，却不能直接充当主体性的总量，也不能成为意识的直接读数。一个人拒绝报告期待，不意味着没有期待；一次报告与行为不一致，也不必意味着报告虚假，可能是取得时点不同或问题诱发了新的判断。代理的有效性必须在具体用途上建立，不能从术语的宏大程度中借来。

记录要求本身也应成为事件的一部分。询问受试者预计会有多甜，与品尝结束后要求回忆预期，得到的是不同位置的证据；要求学生解释错误，则同时增加了一项学习活动。档案至少应保存询问措辞、出现时点、是否向参与者回显，以及谁能够看到回答。本章先保留这些操作事实，暂不把观察到的后续差异归因于询问。第 14 章再辨识记录动作和信息展示的作用。如此分工，档案既能支持因果研究，也能供不接受该因果解释的人使用。

11.4 从概念单位到可以交接的档案

读数事件可以用紧凑形式表示为读数、测量安排、判断位置与发生条件的组合。这个表示有助于提醒研究者，数字只是事件的一个投影。但四项过于概括，直接用于协作时容易产生另一种混乱：一名研究者把仪器版本写在安排中，另一名写在条件中；有人把关于机制的解释写成原始事实，有人则只保留数值。因此概念表达需要展开成工作档案，而工作档案不能反过来被当作存在本身的穷尽分类。

本书采用十项相互连接的记录：被估量、对象状态、测量安排、时间与历史、主体侧重代理、互动侧重代理、特征纠缠说明、读数与误差、回写与后效、治理信息。它们不要求每项都有同样精度，也不要求每个项目全部采集。被估量说明当前问题和单位，对象状态保存材料与制度身份，测量安排交代工具和操作，时间与历史连接事件发生与资料取得。主体与互动代理分别承担有限证据职责，避免直接把理论名称当成观察。

其余字段使记录能够进入后续判断。特征纠缠说明保存被认为相关的组织关系，同时明确它是观察摘要、推断还是待检验假说；读数与误差保留原始结果、转换结果及其不确定性；回写与后效记录结果是否被展示、采取了何种行动、随后发生什么；治理信息保存事件标识、版本、权限、用途和修订理由。事件标识属于档案管理，不必再制造另一套与十字段竞争的概念结构。

一个虚构的品尝事件可以说明这种展开的作用。研究者不只写“甜度七”，还写当前目标是比较某配方在指定温度和报告程序下的短时评分；保留溶液批次、呈现次序与评分时点；把品尝前的期待报告和品尝后的回顾分开；标记受试者是否看见上一次评分。若同

一事件随后被用于产品接受度分析，原记录不应被改名为购买意愿，而须另建用途说明，指出缺少选择行为等证据。十字段由此约束的是结果怎样被再次使用：新增解释可以连接到旧事件，却不能追溯性地改变当时究竟问了什么。

此处尤其重要的是区分原始事实和解释。按键时间可以由系统直接记录，“受试者因失望而降低评分”则是关于原因的解释。即便解释十分合理，也不能与按键记录拥有同一证据身份。可以把它写入档案，但须允许后来根据访谈、对照或新模型修订。如果研究者把每一个波动都预先标成回写，档案将成为理论的自证装置：模型在资料进入以前就决定了资料能够说什么。

可交接性还要求他人知道一个空白意味着什么。未采集、不可取得、因隐私限制不公开、设备故障、对当前任务无关，都是不同状态。将它们全部编码为零，会把没有信息改造成一个确定读数；将它们全部删除，又可能改变样本构成。说明缺失来源，比强求表格满格更有研究价值。档案的成熟不表现为没有空白，而表现为空白的意义可理解，结论能够根据这些限制适当收缩。

11.5 时间不只是一列日期

一个经济数据发生于某个季度，却可能在以后多次发布和修订；一次实验发生在上午，分析规则可能在下午改变；一份学习评价生成于今天，却可能使用昨天才获得的辅助信息。若档案只有一个日期，这些差别很容易被抹平。发生时间回答世界何时出现某件事，取得时间回答研究者何时能够知道，使用时间回答结果何时进入行动，版本时间回答记录何时被改写。对于跨轮研究，这几条时间线各有因果意义。

设想一个研究者使用最终修订的数据重建过去的预测。他可能得到极为清晰的关系，但当时的决策者没有这些资料。最终数据适合研究后来认定的历史过程，却未必适合评估当时是否能够作出某项判断。将两种任务混在一起，就会把后来知道的事情偷偷放进过去的行动条件。这里的问题不是资料真假，而是知识的位置发生了变化。事件研究必须把“事情已经发生”与“行动者已经知道”分开。

时间还以路径的形式进入结果。两个过程可能拥有同样的起点和终点，却经过不同中间状态。若中间状态影响疲劳、学习、库存或承诺，终点的同值便不能证明过程等价。反过来，并非每一段路径都必须保存。有些系统在足够长的时间后会忘记初始差异，某些任务只关心最终稳态，那么一个经过证明的摘要就能代替全部轨迹。判断历史能否压缩，应基于未来任务对历史的依赖，而不能只凭研究者是否喜欢讲过程故事。

一个有用的历史记录应当能够改变问题，而不只是增加篇幅。比如在连续答题中，保存“曾经答错”这一事实可能不足以区分知识不足与错误反馈造成的偏移；保存错误类型、反馈内容和后续修订，才可能说明不同路径。对于研究者而言，历史的粒度应随竞争解释而定。如果所有可考虑的模型都只需要上一轮状态，那么记录全部毫秒日志没有显著收益；若存在长时滞机制，仅保留上一轮又可能系统漏掉影响。

历史档案也需要承认不可恢复的部分。现在询问一个人半年前的期待，得到的是当前对过去的回忆，不能无条件补成当时的期待。旧实验没有记录某个状态，后来构造的代理可以帮助分析，却不能抹去推断的性质。事件化研究的诚实并不表现为为每个过去补齐故事，而表现为分开当时记录、现在报告和模型反演。三者可以共同提供知识，只是各自的不确定性和误差来源不同。

时间维度最终改变了复现的含义。重复一次操作可能不能让一个已经学习的系统回到过去状态，但仍可以检验相同更新关系是否在新的序列中出现。重复的对象可以是固定条件下的分布，也可以是某种干预引发的路径变化。两者都必须明确规定，不能在同值复现失败后随意改成形状复现，又在形状不符时改成“存在历史性”便算成功。时间增加了研究对象，也增加了研究者对事前定义的责任。

11.6 比较怎样从事件之间建立起来

事件各有历史，并不意味着它们彼此不可比较。比较依赖的不是抹去全部差异，而是规定哪些差异可以转换，哪些差异需要分层，哪些差异直接改变目标。两支温度计经过校准后可以比较，同一量表在不同任务中的分数则可能不能直接相减。比较协议应先于合并：被估量是否一致，单位与参照是否对应，样本进入规则是否相容，操作是否改变了所判断的属性。这些问题明确以后，数值关系才获得解释。

对于很多研究，比较的单位也不必是单个均值。可以比较整个条件分布、响应随条件改变的方向、冲击后的恢复过程，或同一规则在不同场景中的预测表现。例如两种教学安排的平均分数相同，却可能在错误类型和延迟迁移上不同；一种安排帮助原本困难的学生，另一种只提高已掌握者的速度。平均数并未说谎，它只是不能承担尚未被赋予的分布判断。事件记录的价值在于提供提出这些额外问题的根据。

比较协议还应说明通过比较以后允许哪一种合并。若只证明两个程序的排序接近，可以共享排序判断，却仍须保留各自单位；若在目标范围内建立了稳定换算，可以转换数值，但换算的不确定性也要传递；若仅确认总体均值接近，则不能据此把个体轨迹互换。

这里的区别直接影响档案交接：接收者应当取得一条已获支持的转换关系，而非一个含混的“可比”标记。分布、排序和数值各自能够流通多远，由不同的验证结果决定。

比较还涉及反事实问题。观察到甲安排下分布较宽、乙安排下分布较窄，不能直接断定乙压缩了发生。两组对象可能不同，测量精度可能不同，样本选择也可能改变。若要判断安排的作用，需要把这些解释纳入设计，或明确依赖哪些不可检验假设。事件之间的描述性差异与改变安排能够造成的差异，是两个有关联但不同的命题。后者需要第 14 章所讨论的因果辨识。

一次有价值的比较可能得到“可以合并”的结论。若不同路径经过校准后在预设容许范围内相容，研究者便获得了一种经过检验的稳定性。此时压缩没有遮蔽问题，而是研究成果的一部分。保留事件信息使这种压缩可以接受未来复核：当用途改变或新条件出现，研究者知道原合并根据什么成立。稳定因此不是发生学的对立面，而是需要说明如何建立、在哪些范围内维持的成就。

公共知识正是由这些局部可比较性逐渐连接起来的。没有必要要求所有领域共享同一种量尺，才能承认共同的问题结构；也不能因为都使用分布、历史和回写等词语，就宣称跨领域机制已经相同。共同协议可以规定怎样描述目标、怎样处理不确定性、怎样登记失败，具体模型仍须尊重对象。跨领域统一首先是一种证明责任的统一，然后才可能在充分证据下成为机制的统一。

11.7 不完整记录中的知识责任

理想档案与实际研究之间始终存在资源差距。完整轨迹需要存储和维护，细致报告占用参与者时间，某些状态涉及不宜收集的个人信息。若把完整性设成任何知识的入场条件，事件化方法就会排除大量历史资料，也会使资源较少的研究者失去参与资格。更合理的要求是让记录强度与任务风险、预期收益和取得成本相称，并明确剩余不确定性，而不是用全知理想惩罚有限研究。

对于一项低风险的产品比较，只要程序标准化、目标明确、样本适当，有限记录就可能足够。若结果将被用于解释个人能力、分配重要机会或改变制度规则，则需要更充分的用途论证和复核通道。记录数量并不直接等于责任程度：收集大量与决定无关的资料，既增加负担，也可能给人一种已经理解个体的错觉。责任体现为能够解释哪些资料支撑了决定、哪些差异尚未被覆盖，以及结论怎样被纠正。

缺失资料可以通过多种方式处理。研究者可以缩小目标，只讨论有记录的条件；可以引入合理范围，检查不同缺失假设是否改变结论；也可以安排小规模补充研究，判断缺失信息是否值得大规模取得。这些选择各有成本。重要的是不要在没有依据时把缺失状态当成典型状态，也不要让模型填补的数值看起来像亲眼观察。补全方法只能在其假设成立的范围内提供知识。

一个具体的例子是没有保存答题过程的旧考试。它仍可能支持关于该考试分数分布和某些后续表现的研究，但很难独自回答某个学生究竟依靠什么策略得分。研究者可以利用新的同类任务补充策略证据，却须说明新样本与旧样本之间的迁移条件。不能因为事件字段不齐就宣布旧分数毫无意义，也不能因为旧分数预测某个结果便断定已经测得全部能力。有限资料对应有限但真实的知识。

权限设计也属于认识条件。为保护隐私而限制原始资料，不必意味着研究无法复核；可以开放脱敏后的摘要、分析规则和独立审计结果，同时说明哪些问题不能由公开材料检查。反过来，全部公开也不自动保证可理解。没有变量字典、版本说明和变换过程的海量日志，往往比一份结构清晰的有限资料更难审查。知识的可继承性需要合适的组织，而不是单纯的可下载性。

研究者对这些限制承担的责任，最终要落实到结论措辞和行动范围。一个推论若依赖未验证的历史代理，应让读者知道替代解释仍然开放；一个制度若使用不完整事件作决定，应提供补充和修订的渠道。限制可以推动新的研究，也可以支持停止继续采集。只要理由公开且能接受检验，简化与扩充都是知识工作的一部分。

11.8 事件化客观性的形成

把安排、判断位置和历史写进测量，似乎会削弱客观性，因为每个结果都显得有条件。然而无条件的外观并不等于更强的客观性。一个脱离来源的数无法说明谁能复核、怎样比较、什么结果会使它失效；一个明确条件的结果则允许不同研究者按照公开关系重新取得或检验它。事件化客观性建立在这种可进入、可转换和可修订的结构上，而不依赖观察者被文字抹去。

这种客观性有三个相互支持的方面。其一是对象约束：解释不能任意改写材料和结果，失败必须留下痕迹。其二是程序约束：资料怎样取得、变换和比较应能被他人理解。其三是后果约束：模型一旦进入行动，新的结果有权迫使模型和规则修订。三者共同防止两种失误，一种把程序稳定误认为对象永恒，另一种把条件多样误认为任何说法都成立。

事件记录本身可以通过交接任务检验。设想原分析者只留下“七”与一份简明事件档案，两名后来者分别判断该结果能否与另一批记录合并。有效的档案应使他们定位到同一目标、程序版本与缺失来源；意见仍有分歧时，也应能说清分歧在单位换算、样本边界还是用途。若增加字段只延长阅读而没有减少这类误认，就应调整结构或删除无用项。档案的收益首先表现为判断依据可恢复，而不是让每项后续统计必然更优。

由属性值到读数事件的转变，因而不是禁止数字独立流通。公共生活需要简明结果，科学也离不开有效摘要。转变在于让摘要能够返回自己的成立条件，让接受者知道它解决什么问题，并在条件改变时有途径恢复被压缩的信息。一个数可以继续出现在图表、标准和制度中，但不再被赋予它没有证明过的无限外推权。

到这里，“甜度七”的身份已经得到限定：它能够支持规定程序下的评分比较，进入其他用途时需要补充证据；它的历史位置、转换关系和缺失也能够被接收者识别。这是读数事件相对于裸数取得的具体成果。不过，档案中写着“依照标准程序测量”，仍可能把大量选择留在一个黑箱里。滤波窗口、量表阈值和版本更新如何进入输出，须沿装置内部的转换链继续追踪。第 12 章由这一尚未展开的入口进入判断的封装。

第 12 章 测量装置与判断的封装

12.1 输出之前已经作出的选择

按下按钮，屏幕给出一个数，这种简洁是测量技术的重要成就。使用者不必重新理解每一根导线、每一步校准和每一种算法，便能完成日常工作。但装置的易用性也会造成一种认识错觉：结果似乎直接来自对象，判断只发生在读数之后。实际上，在屏幕出现数字以前，什么变化可以进入系统、什么变化将被过滤、什么时间段被汇总、什么响应算作异常，往往已经被决定。装置把这些判断转化为能够反复执行的操作。

这里的装置不限于物理仪器。问卷将可说的话变成选项，考试将可以展示的能力变成规定时段内的作答，经济核算将复杂活动变成能够入账的类别，软件将原始数据变成可读取的特征。它们都在某种意义上规定了结果的入口。进入入口并不意味着结果虚构，而意味着结果的可靠性依赖入口怎样与任务相连。判断的封装既是知识积累的形式，也是某些假设不再被日常看见的原因。

设想一台用于生产过程的温度监测器。为了避免短时尖峰不断触发警报，系统对过去若干秒的读数取平均。若目标是控制缓慢变化的整体温度，这种处理可能非常有效；若目标是发现会损伤材料的短时过热，平均却可能隐藏关键事件。系统在两种任务中都忠实执行同一算法，差别来自算法所服务的判断。称它为“更稳定的仪器”并不足以决定它是否更适合当前用途。

这个例子揭示封装的双重结构。第一层是材料与信号的转换，某个物理变化成为电信号或数字；第二层是信号与判断的转换，数字被解释为温度、异常、合格或风险。两层之间还可能存在滤波、校正和模型反演。每一步都可以增加准确性，也可以改变能够看见的变化类型。打开封装不是要求所有使用者同时掌握全部工程细节，而是使重要选择在需要时能够被追溯。

封装还会改变劳动分工。制造者设计响应规则，校准者建立对应关系，使用者决定放置位置，分析者选择汇总方法，管理者规定行动阈值。最终读数却常常以一个机构或一个人的判断出现。若发生错误，只追问最后按键者是否谨慎，就可能遗漏已经进入系统的前提。认识责任需要沿着转换链分配，而不能被屏幕的单一出口集中到一个虚构的全知主体身上。

这并不意味着每次测量都需要一场全系统审计。成熟装置的价值，恰在于大量条件已经得到验证并可以重复利用。关键是区分能够继承的验证与必须重新检验的用途变化。当对象、速度、材料范围或决定后果改变时，过去的可靠性是否仍然适用，需要新的论证。封装的边界应当与其适用证据一起保存，使工具既能稳定运行，也能在越界时暴露自己的有限性。

12.2 校准建立的是哪一种对应

校准通常被理解为把不准的仪器调准，但从认识上看，它首先建立指示与参考之间的对应，并说明这种对应的不确定性。调节仪器可以随后发生，却不是对应关系本身。一个未经校准的指示可能非常稳定，只是持续偏离参考；一个经过校准的系统也可能在特定范围之外失去适用性。稳定与准确因此不能被同一个词吞没，重复出现同一数字不等于测到了所需目标。

考虑两个传感器都读取同一变化中的环境。一个输出随真实量线性改变但始终多出一个偏移，另一个在低范围准确、在高范围逐渐饱和。用单一参考点调整二者，可能使它们在该点完全一致，却不能证明整个范围上的一致。不同偏差结构需要不同校准设计。多个参考点、变化方向、等待时间和重复过程，分别提供不同约束；选择哪些约束取决于预期用途，而不是参考点越多就必然越好。

校准也不是获得了不受条件影响的绝对值。参考标准有其实现方式和不确定性，环境与连接过程会引入变化，被校准对象也可能随时间改变。可靠的结果来自这条对应链被说明并适当控制。测量不确定度指南明确处理输入之间的相关性，并不将一切误差都看成独立白噪声。[7] 对发生学而言，这意味着不能靠批评一个过于简单的误差模型来宣布既有计量学失效；必须说明某种新的历史或回写关系在现有预算中为何没有得到充分表达。

一个容易被忽视的情形是共享参考造成的相关误差。两台仪器若使用同一个偏移了的参考源，它们可能彼此高度一致，却共同偏离目标。将二者平均并不会像平均独立误差那样大幅降低不确定性。若把一致当作独立验证，便会高估证据强度。这不是说共同参考不可用，而是说共同来源需要进入误差结构。所谓多通道验证，只有在通道提供新的约束时才真正增加认识。

这个关系可以用一个简单模型说明。两次读数分别等于目标值、共享偏差和各自噪声之和。平均以后，各自噪声可能部分抵消，共享偏差却原样保留。若只观察两次读数之差，共享偏差还会完全消失，使差值看起来非常小。因此“差值接近零”回答的是两次指

示是否相容，不能独自回答二者是否接近目标。发生条件决定哪个问题被数据约束，不能由日常语言中的“互相证实”替代。

把校准放进读数事件，并不会使每次使用都重新发明标准。它要求记录对应关系来自何处、何时取得、在哪个范围内得到检验，以及哪些变化会触发复核。校准因此具有历史性，却不因此失去客观效力。恰恰因为关系能够被继承和重建，来自不同时间与机构的读数才有可能连接。客观性在这里表现为有根据的共同尺度，而不是没有来源的数字一致。

12.3 算法把什么留下又把什么丢掉

任何实用装置都进行压缩。连续信号被离散采样，大量观测被汇总成特征，多个特征又可能被转化为一个分数。压缩并不天然损害知识，因为任务本身通常只需要世界的某些方面。若要计算一批产品的平均重量，保存每件产品的装饰图案可能无关紧要；若要研究装饰工艺与重量偏差的关系，同一信息就变得必要。压缩是否恰当，必须相对于任务和允许误差判断。

设想两组数列拥有相同平均数，一组每个点都接近平均，另一组交替出现很高和很低的值。只保存平均数，足以回答总体平均水平，却不能回答是否存在危险峰值。这里没有一种摘要能够在不附加条件的情况下支配所有任务。若研究者事后换了问题，旧摘要就可能不够。这种不足不能追溯性地证明原摘要错误，但它能够说明外推的边界：为一种任务验证的压缩，不能自动继承另一种任务的合法性。

可以把这种关系理解为事件之间的任务等价。若两个完整事件对某项行动给出同样判断，并且在允许误差内具有同样后果，就可以在该任务下用同一摘要代表。等价并不宣称事件本身相同，只宣称当前问题不需要区分它们。当任务由均值估计改为极端事件识别，原来的等价类可能必须拆开。这一思想让发生学获得建设性方向：不是拒绝归约，而是研究怎样建立可以说明的归约。

压缩还存在顺序效应。先平均再做非线性转换，通常不等于先转换每个值再平均。例如某项损失随偏差平方增加，正负偏差在均值中可以相互抵消，实际损失却不会抵消。如果装置先把偏差压成零，再据此判断损失很小，它已经丢掉了任务需要的信息。研究者需要追踪每个变换的顺序，尤其在阈值、比例、对数和分类规则介入时，不能把整个处理链视为无害的记账。

路径信息也会在压缩中改变意义。一条先升后降的曲线和一条持续平缓的曲线可能有相同终点；若系统记住峰值，终点并不足以预测下一轮。相反，若已经证明某个状态摘要

包含未来所需的信息，保存全部轨迹就可能没有额外价值。状态摘要的合法性因此是一个经验与形式共同承担的问题：形式模型说明在什么条件下摘要充分，实际资料检验条件是否近似成立。

对于复杂软件，压缩往往分散在多层处理之中。原始数据清理、异常值删除、特征选择、模型训练和阈值设置，都可能改变哪些差异最终可见。一个在训练资料上表现良好的评分系统，也可能依赖某些仅在训练环境中成立的关系。使用者只看输出时，很难知道这种依赖已经改变。因此至少应保存关键变换版本和适用范围，使失败能够回到具体环节，而不是将所有问题归咎于“模型不理解现实”。

12.4 判断怎样变成标准操作

当一项研究程序被反复使用，它常会从有争议的选择变成默认操作。最初选择某个阈值的人可能清楚它是成本与误报之间的折中，后来使用者却可能把阈值当作对象自身的边界。最初为方便协作规定的分类，也可能逐渐被理解为自然界已有的分区。封装不仅保存技术，还保存某些规范选择；随着其来源被遗忘，规范便获得了类似事实的外观。

这种过程在质量检验中尤其容易看清。产品某项指标超过阈值被判不合格，并不意味着阈值两边存在一道绝对自然断裂。阈值可能来自可接受风险、生产成本、客户用途和测量不确定性的综合判断。它可以非常合理，也可以具有公共约束力，但合理性需要依据。若用途改变，或者测量误差大到足以改变临界附近的分类，制度就应重新考虑如何处理边界个案，而不能仅重复“标准就是标准”。

阈值还会反过来影响被测对象。生产者知道检验规则以后，可能把资源集中在刚好越过合格线的指标上；学生知道评分方法以后，可能改变学习策略；机构知道考核指标以后，可能重组记录和服务。由此产生的行为不是外加的小噪声，它可能改变标准最初依据的资料分布。测量规则作为信息进入系统以后，规则与对象便共同演化。能否持续有效，需要在新的行为环境中检验。

不过，行为响应并不总是规避或损害。公开标准也能协调努力，减少不必要的不确定性，并使改进有明确方向。一项指标若推动了所重视的能力与服务，就实现了制度目的。问题在于指标所奖励的行为与目的是否仍然相连。不能因为存在策略性调整便宣布评价无效，也不能因为平均指标提高便认定目的已经实现。要追踪的是从标准到行为、从行为到结果的具体路径。

这要求把测量判断与行动判断适当分开。前者询问结果在给定程序下是否可信，后者询问依据结果采取某项行动是否合理。一个分数可以测得可靠，却不足以决定一个人的全部机会；一个存在不确定性的测量，也可能在比较行动代价后足以支持谨慎决定。两种判断互相联系，但各有证据责任。装置把它们合并成一个红灯或绿灯时，使用者尤其需要知道合并依赖什么前提。

封装的透明也不等于把每个参数同时展示给所有人。过量技术细节可能使真正重要的选择更难辨认。面向不同使用者，应提供与其责任相匹配的说明：操作者需要知道有效范围和异常处理，研究者需要知道变换与误差，决定者需要知道用途限制和行动后果。透明的目标是使相关判断可以被理解与质疑，而不是让文档数量成为无需回答问题的挡箭牌。

12.5 打开封装需要改变哪一层

前述压缩和阈值经由装置稳定执行以后，打开封装需要确定从哪一层开始。对一个监测器，可以改变输入信号、处理规则或行动阈值，三种改动分别检查传感、汇总和决策。若三者同时更新，警报改善便很难归到具体环节。装置分析的工作是保留可逐项比较的接口，让一处改动先与其余部分连接，再判断旧验证能否继续使用。这里检验的是判断如何被封装，不必立即扩展为整个对象系统的机制实验。

打开哪一层，取决于争议发生在哪里。若两个装置面对同一参考输入仍不同，应先查传感与校准；若保存的输入相同而摘要不同，应检查窗口和转换；若摘要相同却启动不同动作，则需要回到阈值与用途。材料是否真正变热是另一项判断，不能由软件版本差异代答。把争议定位到接口，可以使成熟部分继续使用，让修订集中于尚未合格的环节，也避免每次出现异常都推倒整套测量安排。

在温度监测器的构造例子中，可以先把同一段已保存的输入分别送入两个滤波版本。由于输入完全相同，输出差异能够定位到处理链，尚不涉及材料变化。随后把两个版本并行用于新输入，检验原先定位的差异是否在现场仍存在。最后才让警报进入控制行动，检查动作是否反过来改变输入。三个阶段逐步打开封装，分别保留离线变换、现场测量和投入行动的证据，避免把软件差异与对象响应混写成一次“升级效果”。

改变输入时也应选择能够触及假设的形状。若旧滤波窗口只为缓慢变化验证过，适当的检查应包含持续时间不同而峰值相近的信号；若担心饱和，则需要覆盖高范围而非只增加低范围重复。各项输入须处在设备允许范围内，并保留发生器或参考记录。输出未随短

时尖峰改变，可以符合平均任务的设计，却同时表明该版本不足以担任峰值警报。相同结果因此能批准一项用途并限制另一项用途，封装的边界由此成为可执行的使用条件。

逐层打开也会暴露记录的分辨率限制。若旧版只保存每分钟均值，维护者便无法据此检查秒级峰值是否丢失；需要另行取得更细输入，再判断值得保留何种摘要。反过来，如果任务只是缓慢调节，永久保存全部高频输入可能徒增维护负担。可以在验证窗口保留较细轨迹，确认摘要合格后恢复常规压缩，并规定何种越界重新开启细记录。封装由此具备可恢复的层次，而不必在全量记录和不可追踪之间二选一。

装置检查的结果还应转成使用者能够执行的规则。某个窗口适合慢变控制，就说明有效变化速度；某个阈值附近分类不稳，就保留复测或待定状态；某个版本在高范围饱和，就要求越界提示，避免继续显示貌似精确的结果。这样的规则把试验所得接回日常操作。封装的成熟表现为已验证选择可以被放心继承，而在未验证位置及时暴露判断需求；使用者无需重做全部研究，却能知道何时应暂停自动结论。

12.6 版本更换为何改变了测量身份

装置会升级，软件会更新，题库会调整，核算规则也会修订。版本更换常被描述为精度提升，仿佛它只把同一个对象看得更清楚。然而更新可能改变采样频率、异常处理、分类边界和目标定义。两个版本给出不同结果时，首先需要判断变化发生在哪一层：是减少旧偏差，还是改变了被测量，或者只是重新表达同一信息。三者的历史意义完全不同。

设想一个评分软件旧版将遗漏答案记为错误，新版将其中一类技术中断记为缺失。新版平均分提高，不能直接解释为学生能力提高。改变的是记录与汇总规则。若研究者没有保存版本，跨时期比较会把软件变化混入对象变化。反过来，如果新规则纠正了确实存在的误记，也不能为了保持数字连续而拒绝改进。需要做的是建立旧版与新版之间的桥接分析，并说明哪些历史比较仍然成立。

桥接并不总能通过一个固定换算式完成。若版本差异对不同群体或条件有不同影响，单一校正可能掩盖新的不公平或失配。可以在重叠样本上同时运行两个版本，观察差异怎样随条件改变；也可以保留一段并行记录，以区分对象真实变化与工具变化。若无法并行，则应明确不可恢复的部分，而不能用平滑曲线制造看似完整的历史。

版本还包含分析选择。原始信号不变，改变滤波参数或分类算法，也可能得到不同的属性描述。研究者应保存从原始资料到结果的变换，使他人能够判断差异来自资料还是处理。只保留最终表格，会使方法更新看起来像对象自行改写了过去。一个可继承的档案应

能重现旧结果，也能说明为什么新结果取代旧结果；两者的并存是认识演进的记录，不是必须清除的混乱。

对持续学习系统，版本问题更加复杂。系统可能在运行中更新参数，今天与明天不再是同一测量安排。若更新依赖昨日输出和使用者反馈，模型本身就进入回写环。此时只报告软件名称已经不够，需要说明更新何时发生、使用了哪些信息、哪些任务受到影响。无法保存全部内部状态时，也应保留足以支持关键比较的检查点和更新日志，使变化具有可追踪的边界。

版本治理的目的不是阻止修订，而是使修订成为公共知识。一个系统能够快速纠错是优点，但如果纠错抹掉了旧错误的条件，其他使用者便无法学习。记录失败版本及其撤回原因，可以帮助后来研究者判断相似情形，避免把同一种失配不断重新发现。测量历史由此从一个仅供追责的附属物，转变为理解工具怎样获得和失去适用性的研究材料。

12.7 封装能否保留恢复历史的入口

所有资料都长期完整保留并不现实。存储、整理、权限维护和解释成本会不断增长，某些原始记录还包含不宜保留的敏感内容。因此，封装的理想不是永不丢弃信息，而是在可能重要的地方保留恢复入口。这个入口可以是原始数据的有限样本、校准记录、版本说明、关键状态摘要，或能够重新执行的程序。它使未来问题出现时不必完全从零开始。

入口的设计应由失配风险决定。若一个摘要用于高频日常控制，可以只保留稳定运行的汇总，而在异常触发时保存较细轨迹；若研究关注罕见事件，就应警惕只保存“异常”的规则是否已经以旧模型筛掉未知结构。两种安排各有盲区，需要通过抽样保存正常时期和不确定事件来补充。否则，档案只能证明系统如何处理它已经知道的异常，无法支持发现新异常。

历史恢复还需要解释资源。即使原始信号全部存在，如果没有单位、设备状态和处理说明，未来使用者也可能无法重建结果。一段数字序列并不会自动告诉人们哪次重启改变了基线，哪段缺口来自设备停机。元资料因此不是低于数据的附属品，而是数据能够进入推论的条件。但元资料同样应避免无边界扩张，优先保存能够改变目标身份、误差解释或后续行动的信息。

一个可操作的判断是比较恢复代价与重新取得资料的代价。对可以廉价重复的普通测量，有限保存可能足够；对无法重现的历史状态、一次性事件或长期训练过程，提前保留关键轨迹更有价值。这个判断也提醒我们，不同领域不宜强行使用同一档案强度。一次实

验中的可逆操作与一项制度实施后的社会变化，恢复条件差异很大，记录责任应随之改变。

恢复入口还要经得起实际重建。设想维护者从某个警报回查，至少应能取得当时窗口中的输入摘要、所用滤波版本、阈值以及触发时点。如果为了降低存储只保留了均值，便不能再声称可以恢复已经丢弃的峰值；此时入口只能支持均值核验，其他问题需要重新测量。恢复能力应按可重建的判断明确声明。若后来修改异常标签，则保存旧标签及其产生规则，让使用者分清重新解释历史与重新取得观测。

封装与开放于是形成一种持续张力。完全开放会使工作难以运行，完全封闭又会使失败难以解释。成熟方法在两者之间建立分层：日常使用保持简明，重要争议能够下钻，必要修订能够重新部署。它既不把技术便利等同于认识充分，也不把复杂透明当作天然优越，而是在具体任务中比较成本、可靠性和修复能力。

12.8 从装置可靠性到发生条件的可靠性

一台装置经过多次重复仍给出近似结果，可以证明它在某些条件下可靠，却不能独自证明这些条件代表所有用途。可靠性必须具有对象、范围和时间。一个在实验室内表现稳定的系统进入现场以后失效，可能是对象范围改变，也可能是使用者策略改变，还可能是装置通过自身输出改变了过程。只有将这些解释区分，改进才有明确方向。

如果问题来自已知偏差，重新校准可能足够；如果问题来自任务变化，应调整目标或摘要；如果问题来自跨轮反馈，则需要重新理解装置怎样参与系统。三种修复的层次不同，不能一律靠增加测量精度解决。更高精度可以使一个不适当目标被更精确地读取，却不能使目标自动变得适当。另一方面，也不能在精度不足时急于诉诸宏大理论，把本可通过工程改进解决的问题写成存在论困境。

发生条件的可靠性，指向工具与用途之间的关系能够在变化中被说明、监测和修订。它不要求系统永远不变，而要求变化能够触发相应判断。一个能够报告适用范围、识别越界并保留恢复入口的装置，即使偶尔需要停用，也可能比一个始终输出确定数值的装置更可靠。确定性不是可靠性的唯一外观，适时暴露不确定性同样是功能的一部分。

回到监测器，发现短时过热被平均掩盖以后，可能的修订包括缩短窗口、增加独立峰值通道，或承认现装置仅用于慢变控制。这些方案改变不同关系，也有不同代价：缩短窗口可能增加误报，增加通道需要维护，缩小用途则需要其他工具补位。选择必须围绕材料

保护的任务展开。SDE 所要求的相互规定，在这里落实为输出、处理路径与使用规则的联合修订，并由后续警报表现检验；无需另加一层抽象标签来证明修订已经发生。

本章得到的封装结构包括三项可继承内容：指示与参考的校准关系、信号到摘要的任务条件、摘要到行动的判断规则。版本更换必须说明修改了哪一项，恢复入口则保证争议出现时能够返回相应依据。由此，屏幕上稳定的数字既有工程来源，也有用途边界。剩下的困难是，打开封装后发现的散布、不可辨识或预测失配究竟限制什么；第 13 章需要为这些不同失败规定数学对象，才能决定应提高精度、增加约束，还是改变模型目标。

第 13 章 不收敛的分类与数学边界

13.1 同一个词遮住的四个问题

一串读数没有逐次靠近某个固定值，一种估计在增加样本后仍然不稳定，一组参数不能由资料唯一恢复，一个模型换了环境便失去准确性，日常语言都可能称之为不收敛。然而这些现象指向不同数学对象，具有不同失败条件，也需要不同修复。若不先作区分，研究者可能用更多样本解决不了结构辨识问题，也可能把本来正常的随机波动误认为估计方法失效。

第一个问题是读数本身怎样随时间变化。室温围绕一个范围波动，市场价格连续改变，评分受到前史影响，都属于观测过程的行为。第二个问题是估计量怎样随资料增加变化。即使单次读数一直有随机性，样本平均仍可能越来越接近一个稳定目标。第三个问题是辨识：允许取得的资料是否能够区分不同机制或参数。第四个问题是迁移：在原条件中有效的模型，能否在新的条件、任务或干预下继续成立。四者能够相互影响，但不能相互定义。

一个简单例子足以揭示第一种混淆。反复独立掷一枚理想硬币，单次结果不会随着次数增加而变成某个中间值；它仍然在两种结果中出现。但正面频率可以趋近其概率。单次输出的持续分散与估计量的一致性完全可以共存。若把输出必须逐次接近常数设为测量成功标准，就会排除大量正常的概率知识。科学可以精确描述分布，而不要求每个事件给出同一个数。

辨识则涉及另一种困难。若只能看到两个未知贡献之和，那么总和可能十分稳定，两个贡献却未必能分别恢复。更多同类观测能够提高总和的精度，却不能自动增加关于分解的信息。此时问题不在噪声过大，而在资料与目标之间缺少区分约束。改变观察或实验输入可能有用，继续重复同一个操作则可能只让一个不完整答案变得更精确。

迁移失败也有不同逻辑。一个模型在固定人群、固定规则下预测良好，换到新规则后误差增大，并不说明原模型在原条件下没有知识。它说明原知识的适用范围受到限制。研究者需要进一步判断限制来自参数变化、目标改变、信息不足还是模型结构遗漏。把任何迁移失败都称为常数不存在，会丢失最有建设性的部分：哪些关系仍然稳定，哪些条件必须重新进入模型。

本书所说的病理因此应落实为某种持续错配。例如一个制度明知目标已经改变，仍强迫所有结果服从旧摘要；一个模型反复产生有方向的错误，却没有相应修订规则；一个研究把无法辨识的成分当作已经单独测得。这些是具体的认识和制度问题。随机分散、条件稳定甚至长期波动本身，都不能未经任务说明便被诊断为病态。

13.2 稳定分布为什么容许持续散布

在一个最简单的模型中，观测等于固定目标与随机误差之和。若误差相互独立、均值为零并具有相同的有限方差，样本平均的期望等于目标，方差随样本数增加而下降。这里的数学关系说明，单次误差不会因为测量重复而消失，但汇总对共同目标的估计会改善。为了清楚呈现论证，以下模型均为本书构造的形式例子，不对应某项已经执行的实验。公式中的期望算子表示数学期望，与 SDE 中的特征纠缠 E 不同。

$$Y_i = \mu + \varepsilon_i,\quad E(\varepsilon_i) = 0,\quad Var(\varepsilon_i) = \sigma^2.$$

$$E(\overline{Y}_n) = \mu,\quad Var(\overline{Y}_n) = \frac{\sigma^2}{n}.$$

由切比雪夫不等式，对任意正的容许误差，平均数偏离目标超过该误差的概率不大于方差除以容许误差的平方。随着样本增加，这个上界趋近零。这个结论依赖所写前提，并不宣称所有真实序列都适用；它证明的是持续散布并不妨碍一致估计，因而足以否决从“每次读数不同”直接推出“无法得到稳定知识”的逻辑。

独立性也不是稳定估计的唯一道路。对于有时间依赖的序列，平均数方差还包含不同时点之间的协方差。若相关持续很久，同样数量的观测可能只提供较少独立信息；若相关以适当方式衰减，平均仍可能集中。关键不是是否看见任何自相关，而是依赖怎样影响目标估计。一个固定的自相关阈值无法承担所有样本长度和任务中的判断。

$$Var(\overline{Y}_n) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov(Y_i, Y_j).$$

这个公式还有一个重要含义：收集频率提高不等于有效信息按相同比例增加。若连续读数几乎复制同一状态，增加采样点主要改善时间分辨率，未必同样改善长期平均的精度。反过来，不能为了追求独立而任意稀疏采样，因为快速机制可能恰好需要密集轨迹。设计应分别考虑过程辨认与目标估计，而不是用一个样本数概括全部信息。

双峰分布也不必破坏这种稳定性。设某种结果来自两个稳定成分，每次独立地按固定概率选择其中一个。总体可以有两个峰，但仍拥有稳定均值和方差，样本平均也可以收敛。双峰说明单一中心不能完整表达分布，却不意味着概率结构不存在，更不意味着观测不独立。只有当混合比例、成分位置或产生机制随条件改变，才需要进一步研究相应的非平稳性与迁移问题。

这个协方差表达式也给出不能靠加密观测消除的不确定性。设想每次读数含有同一个随机校准偏移，另加相互独立、方差相同且有限的零均值噪声；共享偏移与这些噪声独立且方差有限。平均以后，独立噪声的方差随次数下降，共享偏移的方差却保留。因此样本均值可以越来越稳定地指向带偏移的值，却不能仅靠这批重复消除相对于原目标的不确定性。增加一个独立校准条件与增加一千次相同读数，提供的是不同信息。统计判据的建设性作用正在这里：从协方差怎样进入误差，推导下一步需要更长观测、独立批次还是新的参考约束。

13.3 不可辨识究竟阻止什么

辨识是一个相对于目标和设计的概念。设模型参数为一组未知量，允许的观察安排已经规定，每个参数值都诱导一种可观测资料的概率分布。若两个不同参数始终产生完全相同的可观测分布，研究者便无法仅凭这些资料判断哪一个成立。但这还不足以说明所有目标都无法估计：如果所关心的目标在这两个参数下相同，它仍可能被识别。辨识限制必须指向具体目标，而不能只对模型整体作模糊宣布。

$$P_{\theta,d}=P_{\theta',d},\psi(\theta)\neq\psi(\theta').$$

这里的设计符号代表输入、观察通道、时段和取得程序；目标函数表示研究真正想知道的量。若只在一个固定设计下得到分布相同，结论也只限于该设计；要对整套允许安排宣称不可辨识，必须证明其中每种安排都不能区分这两个目标。结构可辨识研究正是围绕输入输出关系与内部结构的区分展开，不能由输出通道数量替代。[18] 参数模型的辨识理论还有各自的正则条件和局部、全局区分，本书不把某个局部判准推广为全部模型的定理。[19]

可以严格证明一个有限但重要的结论。假定两种参数在每一种可得样本长度上都给出相同资料分布，而目标值不同。取两目标距离的三分之一为正数。若某个估计程序在第一种参数下依概率趋近第一个目标，那么它进入该目标小邻域的概率应趋近一；若同一程序在第二种参数下也一致，就应以趋近一的概率进入第二个目标的小邻域。两个邻域不相

交，而资料分布完全相同，两个事件的概率不可能同时趋近一。于是不存在对两种参数都正确的一致估计程序。

$$\delta=\frac{\left|\psi(\theta)-\psi(\theta')\right|}{3}>0.$$

$$A_n=\{\left|T_n-\psi(\theta)\right|<\delta\},B_n=\{\left|T_n-\psi(\theta')\right|<\delta\},A_n\cap B_n=\varnothing.$$

证明约束的是“对不同真目标同时一致恢复”的可能性。它没有说估计程序不会收敛；程序完全可以稳定收敛到一个约定值，只是不能保证在两种模型下都正确。它也没有说原始输出一定波动。一个不可辨识模型可以持续产生恒定输出。这种区分尤其重要，因为把不可辨识描述成数值永远散开，会把认识限制误写为对象的动态性质。

例如只观察两个未知常数之和，所有总和为一的参数对都给出恒为一的输出。若目标是总和，已经完全知道；若目标是第一个贡献，则现有资料不足。增加对总和的重复观察不能改变这个结构，除非新观察或干预提供了独立约束。这里的正确修复可能是改变输入、单独校准一个成分，或者承认只报告可辨识组合。强行报告两个精确成分，才是认识上的错配。

13.4 单一通道可以包含多个约束

“输入多路而输出单一”听起来像必然的信息不足，但单一通道并不等于只有一个数。一个通道在不同时间、不同输入下可以产生多个独立约束。若已知输出由两个不同衰减模式组成，观察其时间轨迹就可能分别识别两个系数。关键在于模式是否可区分、观察时段是否足够、模型形式是否正确，以及噪声是否允许实际恢复。

$$y(t)=ae^{-t}+be^{-2t}.$$

在理想无噪声条件下，初始值给出两系数之和，初始斜率给出它们的另一个线性组合。由这两个关系可以直接解出系数。也可以取两个适当时点的读数求解，不必真的对带噪资料求导。这个例子不证明所有单通道系统都可辨识，却证明通道数不能单独充当不可辨识定理的前提。

$$y(0)=a+b,y'(0)=-a-2b.$$

$$a=2y(0)+y'(0),b=-y(0)-y'(0).$$

回路同样不能自动破坏可观测性。考虑三个连续状态构成的循环系统，每个状态受到自身衰减和下一状态影响，只观察第一个状态。以下系数固定且已知，目的在于研究状态能否由轨迹恢复，而非估计未知系数。

$$\dot{x}_1=-2x_1+x_2,\dot{x}_2=-2x_2+x_3,\dot{x}_3=-2x_3+x_1,y=x_1.$$

从第一式可得第二状态等于输出导数加上两倍输出，再从第二式得到第三状态。因而理想轨迹包含全部三个状态的信息。系统有闭合循环，却不妨碍这种恢复。动态可观测性需要研究状态与观测轨迹的对应，相关理论也不是只数瞬时出口。[20]

$$x_2=\dot{y}+2y,x_3=\ddot{y}+4\dot{y}+4y.$$

这个系统还会稳定趋近零。取三个状态平方和的一半作为非负函数，对时间求导，得到负的平方和项与相邻状态乘积之和。利用每个乘积不大于对应平方和的一半，可以得到该函数的导数不大于其负二倍，故它指数衰减。输出和全部状态随之趋近零。这里同时存在多状态、循环回写、单一输出、可观测性和稳定收敛，足以否决由前三者直接推出必不收敛的充分性声称。

$$V=\frac{1}{2}\left(x_1^2+x_2^2+x_3^2\right),\dot{V}=-2\sum_{i=1}^{3}x_i^2+x_1x_2+x_2x_3+x_3x_1\le -2V.$$

形式反例不等于关于真实 SDE 发生的经验裁决。它所做的是限定推理：如果某个概念命题声称仅凭这些结构条件便推出数学结论，那么一个满足条件却违反结论的模型已经足以否决该推导。接下来可以增加关于非线性、更新、时滞或观察映射的具体条件，但新增条件必须独立说明，不能事后把所有稳定系统重新命名为没有真正回写。

13.5 条件改变何时使固定预测必然失败

虽然纠缠不能自动推出不收敛，但错误恒定化仍然可以受到严格批评。假定两种情境下目标均值相差超过两倍允许误差，而一个模型对两种情境必须给出同一个预测，那么它不可能同时在两边达标。这是一个简单而明确的条件结论。它不需要先证明世界不存在常数，也不需要把参数辨识与动态波动混在一起。

$$\left|\mu_1-\mu_2\right|>2\varepsilon.$$

若同一预测值在两种情境中的绝对误差都不超过容许值，三角不等式就会推出两个真实均值之间的距离不超过两倍容许值，与前提矛盾。因此至少一个情境的预测必须失败。

这一证明说明，把实际重要的情境差异排除在模型输入之外，会产生无法通过继续拟合同一个常数消除的损失。

$$\left|\mu_1-k\right)\le\varepsilon,\left|\mu_2-k\right)\le\varepsilon\Rightarrow\left|\mu_1-\mu_2\right)\le2\varepsilon.$$

这个结论的力量来自限制清楚。它针对同值预测模型，不针对所有输出一个标量的模型。一个标量预测函数完全可以读取不同情境并给出不同数值，也可以通过状态摘要反映历史。不能因为最终输出仍是一个数，就将它归为已经被证明失败的模型类。发生学真正要批评的是错误地删除决定性条件，而不是数字形式本身。

实际研究还需要承担均值未知的问题。研究者看到的是有限样本，不是真实均值；必须考虑估计不确定性，才能判断差异是否确实超过用途所允许的范围。即使差异在统计上可检出，也可能太小而不影响行动；反之，一个可能造成重大损失的差异，即便暂时证据不足，也值得改进设计。容许误差应根据任务确定，不能在观察差异以后再选择恰好支持理论的门槛。

同值预测的失败也不直接指向某个新机制。条件差异可能来自材料、样本、已知适应或制度规则，SDE 框架需要提出额外的区分性问题，才能获得独立经验内容。它可以帮助定位哪些条件在原模型中被省略，组织回写假说和新观察，但不能把证明固定常数失效当作证明自己的唯一性。竞争模型若已经读取这些条件并成功预测，也应得到承认。

这条条件结论仍有重要建设性意义。它把“需要保留发生条件”转化为一个可讨论的误差问题：哪些条件差异大到使当前摘要不能同时服务多个情境，增加哪项信息可以恢复任务表现，恢复需要多少代价。答案可能是扩充模型，也可能是缩小用途，或者分别保留两个条件常数。不同修复各有合理性，不必强迫所有案例以同一种理论姿态收场。

13.6 四类分布变化及其解释距离

当研究者确认差异值得继续分析，可以先从四类可描述变化入手：中心漂移、尺度变动、响应方向翻转和分布分裂。这四类不是四种必然机制，而是对数据问题的组织。它们可以同时出现，也可以随观察尺度改变。分类的价值在于明确接下来需要解释什么，而不在于给每一种形状预先分配一个 SDE 原因。

中心漂移表示分布位置随时间或条件改变。它可能来自对象状态变化，也可能来自仪器基线、量表使用、样本构成或分析版本。首先需要检查漂移发生在哪一层：原始指示已经改变，还是校正以后才改变；同一对象内部改变，还是进入样本的人不同。只有把这些

来源分开，研究者才能提出某种跨轮更新如何推动中心移动。否则，一条平滑趋势很容易承载比资料多得多的故事。

尺度变动表示散布大小改变，即使均值可能保持不变。它可以反映响应异质性增大，也可以来自测量精度变化或条件混合。去趋势残差的均值通常接近零，因此用其标准差除以均值构造变异系数，可能无定义或极不稳定。应根据量表性质使用原单位尺度、可靠参考尺度或明确模型中的方差参数。标准化本身也是选择，必须说明分母代表什么，不能让一个看似无量纲的指标隐藏新的不稳定。

响应方向翻转关心某种条件变化的作用方向。它比一般散布更接近机制问题，但仍需区分整体与分层。不同群体比例改变，可能使总体方向与组内方向不同；基线不同、上限效应或量表变换，也可能改变表面关系。研究者应先确定比较是否在相同目标和条件下进行，再判断是否存在真实的机制转换。方向性预测特别有价值，因为它较难被任意幅度调整吸收，但失败必须按事前规定记录。

分布分裂则指单一中心不足以描述资料，可能出现多个集中区域。Hartigan 的 dip 检验针对单峰性，拒绝单峰并不确定恰有两个峰，也不识别峰背后的状态机制。[21] 若要讨论双稳态，还需证明系统具有不同稳定状态及相应的转移和恢复特征，不能仅凭一张直方图。固定人群混合、阈值化响应和非线性观察都可能产生类似形状，需要通过轨迹与干预继续区分。

自相关、分组差、交互项和形状检验因此都适合作为寻找机制的入口。NIST 关于自相关图的说明强调时间依赖与模型检查，并以样本量相关的置信带表达不确定性。[22] 这些工具不直接告诉研究者回写发生在哪条通道。形状与机制之间的距离不是方法缺陷，而是研究需要承担的工作：从描述进入可区分解释，再用新的资料缩小解释集合。

13.7 理想可解与实际可测之间

形式上可辨识不等于实际资料足以精确估计。两个参数可能在理论上产生不同分布，但差异小到现有样本和测量精度难以分辨；两个时间模式可能足够独立，却在实际观察窗口内几乎重合。此时增加资料、改善精度或改变设计可能有帮助。它与结构不可辨识不同：后者在规定资料集合中根本没有区分信息，前者有信息却不足以支持所需精度。

前述通过输出导数恢复状态的例子尤其说明这一点。理想轨迹上的求导关系严格成立，带噪观测的直接差分却可能放大高频误差。采用平滑、状态估计或更合适的观测设计，可以改善数值表现，但每种方法引入新的假设与偏差。这种实践困难不应被用来抹去

形式反例，也不能因形式反例成立便轻视实际困难。两层问题需要分别报告，才能找到正确修复。

模型形式已知也是强前提。若输出中的衰减率同样未知，或者存在未记录的外部输入，原先对系数的简单恢复可能不再成立。辨识不是对象永久拥有或缺乏的一种性质，而是目标、模型类和设计共同决定的关系。研究者每次改变允许模型，都应重新审查辨识。一个过于狭窄模型中的唯一答案，可能只是排除了真实竞争解释；一个无限宽泛模型中的不可辨识，也可能不能指导任何具体行动。

前述两个衰减模式能够说明实际精度怎样限制理想可解。若两个读数时点非常接近，两种模式在该小窗口内产生的变化也接近，解方程时便需要用一个很小的差异区分系数；读数中很小的扰动可能被放大为很大的系数变化。把第二个时点适当移远，可以增加模式差异；移得过远又可能使信号衰减到噪声附近。设计因此需要在可区分变化与可测信号之间选取窗口。单纯增加相邻读数，可能远不如补充一个有区分力的时点。

负结果同样需要精度说明。未检出某个效应，不等于效应为零；但也不能永远以样本不足保护理论。研究者可以报告当前资料能够排除多大范围的效应，以及剩余范围是否足以改变用途。若多次适当设计都把可能增益压到实践上无关的范围，继续维持强机制声称便缺少理由。退出不是等待绝对证明不存在，而是在明确任务和证据标准下接受足够有约束力的失败。

实际可测性的判断应把目标精度带回设计。若用途只要求预测输出，两个系数各自区间很宽未必妨碍任务；若用途要求单独改变某个内部成分，同样精度便可能不足。研究者可以先考察与现有资料相容的参数集合，再检查它们对目标操作的预测是否接近。预测仍接近时，可以报告内部机制未定而任务判断可用；预测分散时，则应把下一轮观测放到这些模型产生最大分歧且能够可靠测量的位置。形式可解因而被转换为有限资源下的取得顺序。

辨识还可以随干预集合改变。设两种机制在自然观察中总是给出同一结果，但它们对某个可实行输入的反应不同，那么观察等价不再意味着干预等价。研究者此时有理由把资源用于改变输入，而不是无限延长原序列。相反，如果所有允许干预下两种机制仍然等价，那么当前研究只能报告共同可知的部分。这种结论并不要求在两种机制之间强选一个，也不阻止依据它们共同支持的预测采取行动。

一个明确构造的例子是两个成分始终随同一个温度一起变化，只观察总响应时难以分别估计。若存在只改变其中一个成分的独立操作，新增实验就可能拆开原有关系；若所谓独立操作同时改变两个成分，研究者必须把这种作用写入模型，不能仅凭操作名称宣布辨识完成。实验之所以增加知识，在于它改变了可观察关系，而非因为它比观察研究在形式上更像实验。控制是否成功，本身需要材料支持。

模型之间的等价也未必是研究失败。在某些工程任务中，多种内部解释对全部相关操作给出相同预测，当前目的可能只要求选择一个便于维护的表达。此时应承认机制未定，同时保留任务层面的可靠性。只有当用途改变、需要判断新的干预或责任归属时，原先可以搁置的内部差异才重新重要。辨识目标随任务而变，正说明不能把一个笼统的可辨识标签贴在系统上永久使用。

还应区分状态可观测与参数可辨识。第 13.4 节循环系统的系数已经给定，输出导数恢复的是当前三个状态；若把衰减系数和耦合系数也都当作未知量，原推导不会自动证明这些系数可唯一估计。此时需要重新列出不同参数与初始状态可能产生的轨迹，并检查新增输入是否拆开这种等价。研究报告应明确哪些量经过外部校准，哪些由当前数据估计。否则，把已知系数的状态恢复结果称为未知机制已被识别，就会在数学结论与实际声称之间漏掉一层证明。

13.8 让边界成为理论的生产条件

经过这些区分，不收敛不再是一个可以覆盖所有异常的总称。原始输出可能不稳定而分布稳定，目标可能可估计而内部成分不可辨识，机制可能有回路而整体收敛，局部模型可能准确而迁移失败。每种组合都提出不同研究任务。数学边界的作用，是让这些任务不被一个宏大判断提前终结。

对具体研究，可以沿目标建立一条诊断次序。先确认讨论的是观测轨迹、估计程序、内部目标还是跨条件预测，再规定允许资料与误差范围；随后检查当前失败是否由共同偏移、观察等价或条件删除造成。这一次序把问题送向不同修复：共同偏移需要独立参考，观察等价需要新增约束，条件删除需要扩充输入或缩小用途。稳定性的发生由此进入可计算关系，而不是只用“复杂”概括困难。

这些修复之间仍有连接。新增观察通道可能打破原有等价，也可能带来额外误差；扩充条件可以降低同值预测的必然损失，却会增加有限样本估计的负担。因而数学诊断不能在宣布“有解”时结束，还应说明改善哪项限制、引入什么新限制。例如给两个未知贡献

分别加传感器，形式上增加了约束；若两传感器共享未知校准源，原来的和问题可能只是转成了新的偏差分解问题。每次改动都需要重做目标与观察关系，而非沿用上一次可辨识结论。

真正有力量的理论愿意说明自己禁止什么。若允许任何分布、任何方向、任何稳定性都被解释成同一种纠缠，那么它只能提供语言组织，不能独自承担经验裁决。语言组织可以有哲学价值，但若进一步声称发现了机制，就必须写出哪些结果不应出现、哪些干预会改变路径、哪些条件下应退出。数学与经验在这里共同把研究从解释的无限可能引向有限而可积累的知识。

本章取得的结果是有条件的：独立零均值且具有相同有限方差的误差下，均值可以一致估计；观察等价可以阻止特定目标恢复；单通道循环系统仍可稳定且可观测；删除足够大的条件差异则使同值预测必然失配。这些结果共同限制了“不收敛”的使用。它们还没有回答过去的读数是否造成未来变化：历史依赖可能只是状态延续，也可能经过信息和行动形成通道。第 14 章由此进入因果辨识，将能够描述的时间结构进一步转成可干预的关系。

第 14 章 回写机制与因果辨识

14.1 过去在场不等于过去的读数产生作用

一次评分偏高以后，下一次评分也偏高，这种连续性可以由多种机制产生。受试者可能记住前一次评分并以它为锚点，也可能只是保持了某个持续状态；仪器可能带有缓慢漂移，实验安排也可能使相邻刺激彼此相似。只有第一种解释直接涉及读数作为信息进入后续判断。若将所有时间依赖都直接称为读数回写，就无法区分信息使用与其他承接，也无法告诉研究者应该改变什么。

全书的一般回写概念包含此前过程改变后续发生条件的各种承接，既可以经过物质状态，也可以经过记忆、行动或制度规则。细胞内部的调节历史并不需要先变成公开数字，才具有回写意义。本章集中辨识其中较窄的一类：已经取得的读数或形成的信息，怎样经由展示、理解和使用改变后续过程。这个范围限定让研究者能够提出“遮蔽读数以后会怎样”的明确问题，同时保留物质历史等其他通道作为竞争解释或共同机制。

对这一较窄问题，需要区分状态延续与读数介入。前者由某个持续状态同时影响相邻结果，读数主要是它的指示；后者使已取得的结果通过被看见、解释或使用而改变系统。持续状态的形成自身可能包含一般回写，只是不能从其延续直接推出读数的因果作用。若遮蔽结果不改变未来，受检验的信息通道受到限制，物质承接仍可能存在；若随机改变展示内容能够改变未来，则取得了超出状态连续性的通道证据。

这种区分也防止把“观测改变对象”说得过于宽泛。探头接触改变局部温度，是测量操作的物理作用；公布评分改变人的策略，是结果作为信息的作用；软件将评分写入下一轮参数，是制度化更新。三者都值得研究，但对应不同干预与证据。把它们合称为一个神秘的观察者效应，会削弱而不是加强发生学，因为具体通道恰恰是可以积累知识的地方。

一个虚构的连续判断任务可以作为贯穿例子。参与者每轮估计一段线条的长度，系统有时显示其上一轮答案，有时不显示。下一轮表现改变，可能来自显示内容，也可能来自显示动作吸引注意，或额外等待改变疲劳。要识别回写，不能只比较“有反馈”和“无反馈”两个标签，必须说明反馈究竟包含什么信息，哪些操作保持相同。因果问题开始于可区分的操作，而不是开始于对两组曲线的命名。

回写研究的目标也不应预设它必然有害。反馈可以纠正偏差、协调行动、形成稳定参照，也可以放大偏差和造成锁定。某个系统的稳定可能正是持续回写的成果。因而研究首

先要说明作用方向、强度、时滞和恢复条件，再判断它在具体任务中的价值。历史进入下一轮是一种结构，健康与病理则需要目标与后果的进一步判断。

14.2 把闭环展开为可观察的过程

闭环在图形上很容易画成一个圆，但一个圆无法直接用于辨识。研究需要沿时间把它展开，分别说明状态如何产生结果，结果如何被呈现，行动如何响应呈现，行动又怎样改变状态。这样，同一系统在不同时点的关系就可以被具体观察。没有时间展开的“相互影响”，常常把原因和结果同时放进一个解释，导致任何方向都显得可能。

以下是一个用于分析的简化模型。状态代表当前材料、知识或资源条件，观测是状态在测量安排下的显露，展示信息可以是原始观测或某种加工结果，行动根据展示和既有历史形成。下一轮状态受当前状态、行动与外部输入共同影响。模型不是三号位各对应一个独立变量，而是为某个回写假说建立有限操作表达。

$$y_t = h\left(x_t, q_t\right) + \varepsilon_t, m_t = g\left(y_t, z_t\right), a_t = \pi\left(m_t, H_t\right).$$

$$x_{t+1} = f\left(x_t, a_t, u_t\right) + \eta_t.$$

其中测量安排与展示规则都可以成为研究者改变的对象，历史代表行动者在当时能够使用的信息。若展示规则被随机改变，研究者可以考察信息通道；若行动受到限制，可以考察从信息到状态之间的中介；若测量安排改变，则可能同时影响指示与对象。每个操作需要单独分析，因为切断一条边往往会改变环中其他关系。

这个展开有助于区分“结果具有预测力”和“结果具有因果作用”。一个读数可能准确预示下一轮状态，只因它反映了当前状态；把读数隐藏起来，未来仍然相同。相反，某个公开指标原本只是粗糙代理，却因为被用于分配资源而开始改变未来。预测价值与因果力量不必同强同弱。发生学应同时研究两者，而不能用一个相关系数代替它们之间的差异。

闭环还可能有多个时间尺度。即时展示改变下一次动作，累计表现改变长期目标，制度评价改变未来可获得的资源。这些更新可以层层嵌套。研究者不必一开始建立全部尺度模型，而应根据当前争议选择最小足够结构，同时检查被省略的慢变量是否改变结论。一个短期有效的纠错机制，可能在长期改变参与者策略，原模型的适用窗口因此需要明确。

时间展开还规定了每项资料应在哪个位置取得。展示前状态用于说明起点，实际显示日志用于证明操作到达，随后的行动记录用于定位承接，下一轮结果用于判断后果。若行动只能在结果之后回忆，便需要把回忆误差另列；若展示分配了但未被接收，则分配效果与接收效果也应分开。三原理所关心的更新由此有了可追踪的时点。一个缺少中间记录的闭环可以仍是候选解释，但其缺口不能靠箭头方向自动补齐。

14.3 从信息开放到记录反应性

增加记录有两种完全不同的效果。它可以让分析者知道更多已经发生的事情，也可以通过要求报告、提供反馈或改变界面，影响事情怎样发生。两种效果都可能有价值，却需要不同研究设计。若将它们混合，研究者可能把记录造成的变化当作原本潜藏的结构，又可能把信息带来的预测改善误认为测量改变了对象。

第一种问题可以在同一批事件上研究。先以尽可能低干扰的方式取得完整资料，再让不同分析程序分别使用有限字段与较完整字段。响应变量、采样频率和测试资料保持一致，比较信息增加是否改善预测、解释或错误定位。因为事件本身相同，原始边际分布不会因分析者后来看到标签而改变。改变的是对条件结构的识别能力，而非资料已经发生的历史。

这一问题的结果应限定为资料使用的收益。设同一批历史日志中，一组分析者取得实际展示值，另一组只能取得原始评分；前者预测更好，说明展示值提供了额外条件信息，仍未证明展示造成变化。因为实际展示可能按受试者状态选择，两者可能共同反映未记录条件。第 15 章负责信息与模型资源的公平比较；这里需要保留的因果判断是：分析者看见什么与参与者经历什么，是两个不同的干预对象，不能用前者的改善替后者作证。

第二种问题则需要改变记录行为。参与者可以被分配到主动报告、低干扰记录和常规流程等安排，但主要结果的测量必须保持可比。若主动报告组反复作答，而常规组只在结尾作答，两组差异同时包含采样密度、注意分配、练习和报告行为，不能单独归于记录主体状态。设计的关键是把可能作用拆成可以比较的操作，而不是仅使表格中的字段数量不同。

一个适当的分步实验可以先比较是否询问预期，再比较是否把预期展示回来，最后比较展示内容怎样影响后续判断。询问可能形成新的期待，展示可能巩固已有期待，内容变化则可能提供方向性证据。这些关系并非一定出现，实验的意义就在于允许它们分别失败。若只有展示动作有作用而具体内容没有作用，注意或任务切换可能是更好的解释。

还应考虑参与者对研究安排的理解。人们可能把额外记录视为研究者希望他们关注某个方面的提示，由此改变策略。这个反应不必被简单称为污染，它本身可以成为研究对象；但若目标是自然状态下的机制，就必须把安排诱发的作用与自然过程区分。测量的反身性不是取消因果研究的理由，而是要求研究把自己的介入写进因果问题。

14.4 随机展示与回写通道的检验

若研究者能够在不改变原始状态的前提下随机改变展示信息，就有机会识别某条信息回写通道。核心思想是让展示内容与原先影响结果的状态相互独立，从而比较不同展示造成的后续差异。这里的独立来自设计，而不是事后看到两组平均值接近就可以宣布。随机分配也不能保证所有实施问题消失，仍需检查参与者是否实际接收信息以及后续资料是否完整。

在前述长度判断的虚构任务中，可以保持前一轮真实表现与呈现时长一致，只随机展示不同形式的历史摘要。若下一轮判断随展示方向系统改变，信息内容的因果作用比单纯自相关得到更强支持。进一步还可比较真实历史、重新排列的历史与不含方向信息的占位显示，以区分具体历史、一般反馈和界面注意。所有操纵都须符合参与者知情与研究目的的要求，不能将研究便利当作任意误导的理由。

即使随机展示改变了下一轮，也不能立即宣布完整机制已经识别。展示可能通过期待、注意、动机或策略等多个中介起作用。要主张其中一个通道，需要额外测量或干预，并考虑中介本身受到哪些共同原因影响。单独把某个自我报告放入回归后，效应变小，并不足以证明该报告就是唯一中介，因为报告可能有误差，也可能被处理本身改变。

回写阻断实验提供另一类证据。若某个结果通常进入下一轮资源分配，可以暂时保持分配不随结果变化，比较后续状态。若差异消失，支持行动通道的重要性；但阻断也可能改变参与者预期，使其采用新策略。研究者应说明阻断是否只影响目标通道，还是同时改变了整个任务的意义。机制检验需要对干预的附带作用保持具体，而不是把“阻断”当成已经纯净的操作。

假反馈与真实反馈之间的比较同样有局限。真实反馈往往与当前状态一致，假反馈可能引起怀疑，从而同时改变可信度。若没有记录可信度或安排相应对照，差异可能来自信息内容，也可能来自参与者是否相信装置。一个设计良好的研究会把可信度作为可讨论的条件，而不在结果不符时临时宣称参与者没有真正进入实验。

因果辨识的强度因此来自多种相互约束的证据。随机展示说明信息有作用，通道干预说明作用经过哪些环节，恢复实验说明改变是否持续，外部任务说明机制是否只适用于特定界面。这些证据并非必须一次齐全，但每增加一个环节，都应缩小竞争解释，而不是仅增加一项支持标签。回写的可信性是在这种逐步排除中建立的。

14.5 时滞与稳定性不能由形状代替

回写经过时间，时滞因此是机制的一部分，但长时滞不自动导致失稳，短时滞也不保证安全。作用方向、增益、饱和、外部驱动和其他反馈共同决定结果。一个缓慢的负反馈可以稳定系统，一个迅速而强烈的正反馈则可能放大偏差；当系统具有边界或饱和时，正反馈也未必无限增长。分类必须进入具体更新关系，不能凭快慢给系统分配命运。

考虑一个离散时间的简化模型。当前状态的一部分保留到下一轮，行动根据当前读数形成，行动再改变下一轮状态。在忽略噪声、读数等于状态且各关系线性的条件下，闭环更新系数是状态保留与反馈作用的合成。只要其绝对值小于一，状态就趋近零；大于一时一般会放大，等于边界值则需另行分析。这个例子说明，回写是否存在与回写是否稳定是两个问题。

$$x_{t+1}=r\,x_t+b\,a_t,\quad a_t=k\,x_t,\quad x_{t+1}=\left(r+b\,k\right)x_t.$$

$$\left|r+b\,k\right|<1\Rightarrow x_t\to 0.$$

若反馈基于上一轮而非当前状态，系统就需要同时保留两个时间点，稳定性取决于相应特征根。时滞可能改变振荡与恢复方式，但并不单独决定结果。研究者可以通过受控扰动观察响应曲线，比较其峰值、衰减与延迟，而不是把自相关长尾直接翻译为某种不可逆结构。一个缓慢外部驱动也能产生长尾，却不一定包含目标回写。

恢复过程尤其有区分价值。短暂改变输入以后，系统是返回原工作点、停留在新状态，还是因规则更新而沿新路径运行，分别对应不同假说。若撤回干预后仍有差异，说明存在持久后效，但持久并不自动等于不可逆。观察窗口不足可能使缓慢恢复看起来永久，未记录的持续输入也可能维持差异。不可逆性需要明确时间尺度和允许的反向操作。

同样，两次扰动大小相同而恢复不同，可能揭示路径依赖或状态依赖，但也可能来自初始状态不一致。研究者需要说明比较是在何种状态匹配下进行。所谓迟滞若只是输入上升和下降时记录延迟造成的图形，也不能直接证明系统有不同稳定分支。数据的时钟、装置响应和真实状态变化必须先被分开，才有资格讨论更深的动力学结构。

对 SDE 而言，时滞与增益可以帮助把“相互规定”转成更具体的问题：哪类张力能够推动更新，更新需要积累到什么程度，更新后怎样改变原有关系。它们不必在所有领域采用同一种数值单位，却应保留相同的解释责任。一个机制图只有在这些细节上能被资料约束，才超越“所有因素互相影响”的一般描述。

14.6 无法随机化时能够知道什么

许多重要回写发生在无法随意操纵的系统中。研究者不能为辨识方便任意改变宏观政策、家庭处境或长期制度，也未必能够随机决定某个公开指标是否发布。观察研究仍然可以积累机制证据，但必须更清楚地说明哪些关系来自资料，哪些关系依赖假设。不能因为没有随机实验便放弃研究，也不能因为问题重要便降低因果声称的门槛。

时间顺序首先应成为排除具体解释的工具。设想某实验室的反馈界面在不同工作站分批启用，启用后连续判断偏差减小；若同时更换显示器或重新训练操作者，前后差异便混合了多种变化。可以追查更新日志，比较尚未启用工作站在同一时段的变化，并检验启用前趋势是否已经分开。如果工作站原本按困难程度分配对象，比较还需说明这种选择。这里得到的约束来自实施历史，不来自“前后比较”这个方法名称。

历史版本在这里尤其重要。行动者依据当时可见的资料决策，后来修订后的资料不能无条件替代当时信息。实时数据研究指出，使用实时资料与事后修订资料可能产生不同政策判断。[23] 对回写研究而言，这提供一条明确方法要求：重建每个行动时点能够获得什么，随后哪些消息进入解释与行动，再判断这些行动是否改变了未来资料。否则，研究者可能用后来更准确的知识构造一个从未真实发生过的闭环。

比较组也需要机制上的可比性。未受某项信息影响的群体可能在资源、规则和趋势上与受影响群体不同；把它们放在一起不能自动形成反事实。研究者可以利用制度边界、发布时间差异或不同暴露程度寻找额外约束，但每种策略都需要说明为何其他差异不会产生同样结果。方法名称不能替代这段实质论证，尤其不能用复杂估计掩盖明显不可比的对象。

负对照有助于检查某些替代解释。例如一个被认为只应影响某类任务的信息，若同时改变了理论上不相关的结果，便提示存在一般注意、共同冲击或选择机制。负对照并不自动证明主结果错误，却能迫使模型解释更广的资料。类似地，虚构的提前发布时间若同样出现“效应”，说明原时间关联可能反映趋势，而不是目标发布的作用。对照的价值在于给理论制造可以失败的地方。

最终，观察资料可以支持不同强度的结论。它可以确认时间结构，建立条件关联，排除若干竞争机制，也可能在强而透明的识别假设下支持因果判断。每一层都有价值，不能为了得到统一的“已证实回写”而抹平差别。如果资料只能支持机制相容，就应保留这个位置，同时指出哪种新观察或制度变化能够进一步区分解释。诚实的未决比虚假的闭环更有生产力。

还要注意不同研究单位之间的信息传播。一个人接受反馈以后可能把策略告诉同伴，一个机构调整规则可能改变周边机构的选择。若把这些单位当作互不影响的对象，处理组与比较组之间的差异可能被传播缩小，也可能被竞争放大。研究者可以按实际传播边界组织分配，记录接触关系，或把传播本身纳入目标。重要的是不能同时依赖系统自反性解释主结果，又在估计时默认单位之间完全隔离。

退出研究的行为也可能进入回写。某种反馈使困难者更早放弃，留下者的平均表现便会改善，即使原有个体并未进步。只分析完成全部轮次的人，会把选择后果误认为学习后果。因此，事件档案应保留何时退出及当时可见信息，并分别讨论参与持续性与完成者表现。若无法取得退出后的结果，应报告这种缺失可能怎样改变结论，而不是将没有读数的人从发生过程里抹去。

14.7 主体相关指标与因果通道的区别

研究者常通过期待评分、参照点选择或自我核验记录描述主体相关状态。这些是认识上的代理指标，不是代替主体行动的代理人，也不构成主体性的总量。指标在本章承担两个需要分别验证的任务：表征某种判断状态，以及帮助说明该状态是否传递了展示的作用。一个期待评分可能同时反映熟悉度、动机和此前经验；即便它预测后续表现，也需要额外约束才能成为通道证据。

代理的取得时点决定其证据身份。任务前报告可以作为当时状态的有限证据，任务后回忆可能被结果重新组织。若研究者用后者解释前者，就有可能把结果造成的叙事当作结果的原因。两者都可以研究，只是问题不同：一个关心期待怎样影响显露，另一个关心显露怎样改写对期待的理解。读数事件需要保留这条方向，而不是把同名问卷分数放进同一列。

多个代理之间不一致，也不应立即归结为主体不可测。它们可能对应不同层面：口头预期体现可表达判断，选择行为体现任务中的权衡，纠错轨迹体现后续更新。三者未必应该完全一致。研究者需要先说明各指标承担什么解释职责，再判断不一致是否构成效度问

题。若所有指标都被要求证明同一个笼统主体性，任何差异都会显得神秘，具体机制反而无法建立。

对代理进行干预还可能改变构念本身。要求参与者公开承诺一个预期，可能使其后来维护这个预期；让其反复解释策略，可能促成原本不存在的策略意识。因此干预所识别的是某种操作造成的后果，不必等于自然状态中同名变量的效应。研究可以将这种创造作用纳入发生模型，但必须明确：这里研究的是预期如何被组织出来，而不只是读取一个原先已经完整存在的内部数值。

设想展示改变了期待报告，也改变了下一轮判断。至少存在两种结构：展示经由期待改变判断；或者展示分别改变报告方式和判断策略，两项结果彼此相关却不构成中介。要缩小分歧，可以安排尽量不改变信息内容的报告措辞变化，检查报告数值改变时任务效果是否保留；也可以用行为选择与事前报告共同约束期待状态。每个新操作仍有附带作用，需要说明其范围。这比直接把报告分数的回归系数称为主体的因果力量，更能积累可被修订的机制知识。

指标失配还可以指导下一步取得资料。如果口头预期随展示改变，而选择与纠错均未改变，模型应检查表达或量表使用；如果选择改变但口头报告不变，可以研究未被该问卷捕捉的策略。两种情形都保留主体侧重，却对当前指标作出不同判断。一个代理有效只支持规定任务中的操作关系，意识及完整主体性的论证仍有独立要求。主体相关研究的进步，体现为能够区分这些失败位置，而不是不断添加同义问卷来追求一个全面总分。

14.8 回写解释何时值得保留

一个回写模型值得保留，不仅因为它能够重述已经发生的轨迹，还因为它比竞争解释更有效地组织新的观察和干预。它应当指出，在何种条件下前一轮结果会改变下一轮，改变方向是什么，作用经过哪些环节，以及阻断或替换环节以后会发生什么。若这些预测与一个更简单的状态延续模型完全相同，回写语言可能有组织价值，但经验增量尚未建立。

比较必须允许旧有动态模型以强版本进入。适应、学习、滞后响应、异质性和制度反馈并非等待 SDE 才出现的研究对象。新的框架需要说明它增加了什么：也许是过去分散记录的环节被连接，也许是发现了被忽略的更新规则，也许是改变了辨识目标与实验入口。贡献可以是方法组织、机制发现或概念澄清，但三者不能相互代领证据。

贯穿本章的长度判断任务可以形成一项具体结论：若展示内容的随机差异改变后续响应，且呈现时长和界面注意不足以解释这种差异，就可以确认受检验的显示安排具有作

用；若进一步限制某种调整行动后效应减弱，则为行动承接提供额外证据。结论仍应保存未解决部分，例如期待与策略是否能够分开。这种写法把已经辨识的边与仍未辨识的边同时留下，使后来研究者知道应复核哪项操作、增加哪项观察。

失败也应针对具体层次处理。若结构不能重现，首先检查测量、抽样和探索选择；若结构稳定但干预不改变它，应重新考虑通道；若通道成立却没有额外任务收益，则需要重新评估模型复杂度与用途。不能每次都以“纠缠太复杂”解释失败，因为这种解释没有给下一次研究提供不同操作。能使研究改变方向的失败，才真正进入理论的学习过程。

回写模型还有一种容易忽略的责任：它可能因为被采用而改变自身适用条件。当机构按照模型调节反馈，原先估计的更新关系可能随新策略发生变化。成功干预未必维持原来的统计规律，有时恰好会消除原先的异常。研究者应事前说明成功意味着保持哪种关系、改变哪种关系，并在部署后重新检查。模型进入过程以后，不能继续把自己当作过程外的旁观者。

本章因此把过去在场细分为持续状态、实际可见信息、行动承接与后续更新，并为显示、阻断和恢复操作规定了各自的证明范围。一个可用的回写解释，应能据此指出改变哪一环会产生什么后果，同时允许剩余通道未决。因果辨识解决的是这种操作关系；是否值得长期采用该模型，还涉及它与其他表达在相同任务上的精度、成本和迁移。第 15 章接过这一评价问题，使已被识别的通道进入公平竞争，而非仅凭机制名称取得使用资格。

第 15 章 发生模型的比较与迁移

15.1 解释丰富并不自动成为解释优势

两个模型可以用完全不同的语言解释同一组资料。一个使用条件变量和动态参数，另一个使用发生条件和回写路径；如果它们在当前任务上给出相同结果，仅凭语言的丰富程度无法判断哪一个提供更多知识。理论可以帮助提出此前没有被问过的问题，但一旦声称改善了解释、预测或行动，就需要规定改善发生在哪个层次，并让竞争者面对同样的任务。

模型比较首先是任务比较。预测下一轮评分、解释某次评分的形成、估计干预效果和选择制度规则，需要不同评价。一个模型可以善于预测，却无法识别因果通道；另一个模型可能揭示机制，却因资料成本太高而不适合日常部署。若把所有优点合成一个模糊的“解释力”，研究者便容易在某项结果失败时转向另一项长处，使比较失去共同尺度。

任务还规定需要预测什么。点预测关心一个代表值，分布预测关心可能结果及其概率，排序关心对象之间的相对位置，决策则关心采取行动后的损失。均值预测准确的模型未必善于识别极端情况，排序稳定的指标也未必支持绝对差距判断。发生模型应说明自己增加的条件信息服务哪种任务，而不是把输出从一个数改为一张复杂图就宣布优越。

第 12 章的温度监测器可以把任务差异继续推进到模型选择。设模型甲预测下一时段的平均温度，模型乙利用近期轨迹预测超过阈值的持续时间。在缓慢调节任务中，甲可能以更低成本达到要求；在材料保护任务中，乙若能减少漏报，就可能更有价值。但两个任务应各自固定损失和检验资料，不能因乙在平均误差上落后就否认其警报用途，也不能因其识别过一次峰值就宣布全面优越。同一装置可以保留两个经过分别验证的模型，而不必选出一个适合所有用途的冠军。

公平比较也意味着重建竞争解释的强版本。异质性、适应、状态空间模型和制度反馈已经能够处理许多条件变化。把旧模型写成永远坚持一个无条件常数，再让任何具有情境变量的新模型取胜，并不能建立 SDE 的独立贡献。本书可以从最简基线开始，但最终对手应包括最有理由成功的既有模型。理论的增量在强比较中可能缩小，却也会因此变得真实而可继承。

发生模型的增量须在三个层次分别提出。对象重建改变被估量、事件同一性或比较资格，例如把混在一起的个人作答与人机协作重新划定为两个评价对象；设计辨别力使原先

等价的解释在可执行安排下给出不同预期；经验机制则要求取得资料，支持某条具体通道及其条件。前两种贡献可以先于预测胜出成立，却不能代替第三种证据。对象重建后还要交代新旧目标如何对应：若仍声称更好预测同一结果，必须与同目标、同资源的强基线比较；若目标已经改变，则分别评价新任务的理由和表现，不能拿目标变化本身计算预测增益。

15.2 从基线到回写模型的竞争阶梯

一个有解释力的模型比较不应只设置两个极端：一边是没有条件的常数，另一边是包含所有可用信息的复杂系统。这样的胜负无法说明究竟哪种新增结构有用。更有建设性的做法是建立逐步扩展的模型阶梯，让每一步对应一个明确问题。阶梯不是规定真理必然位于最高层，而是帮助研究者定位改善来自哪里。

最简层可以是固定目标加测量误差。它说明在不引入情境差异时资料能够被解释到什么程度。下一层加入有根据的对象异质性和条件变化，检验组间差异是否已足够解释结构。再下一层加入既有适应、状态延续或动态更新，检验时间依赖是否需要目标回写。最后才加入明确的新回写通道，并要求它产生独立预测。每层应拥有适当复杂度，而不是故意让低层模型无法学习明显关系。

例如某种评分随重复次数下降。固定模型失败以后，加入个体差异仍不能解释同一人内部的下降；加入随时间变化的疲劳状态可能已经充分预测。此时再加入“上一轮评分改变下一轮期待”的通道，只有在信息展示等设计中产生额外区分，才获得独立支持。若两种机制在自然轨迹上等价，比较应进入可区分干预，而不是继续在同一数据上争论哪种语言更深刻。

阶梯还应分离信息与模型形式。一个模型使用期待、路径和历史，另一个只使用当前刺激，前者获胜可能只是因为信息更多。要判断新增结构的价值，可以让两类模型使用同一信息，比较不同组织方式；要判断新增记录的价值，可以固定模型类，逐步开放字段。两种比较共同说明，收益究竟来自记录、函数灵活性、更新规则还是其组合。

模型选择也受到调参预算影响。复杂模型若获得大量尝试机会，而基线只运行一次，最终差异包含搜索资源的不平等。研究应尽量保持可比的调参规则，或至少公开各自使用的资源。预处理、变量筛选和早停规则都是模型的一部分，不能把它们隐藏在“数据准备”中，使新模型额外享有看过测试资料的机会。公平并不要求所有模型形式相同，而要求比较不通过未披露的优势预先决定。

阶梯比较最终可能停在中间层。若疲劳状态已经解释重复评分的下降，下一项有价值的比较是：增加展示历史以后，是否改善新展示安排下的预测，而非继续降低旧轨迹误差。若没有改善，可以保留动态模型，并把信息回写列为当前未获支持的候选；若只有新增记录起作用而函数形式没有优势，则把贡献归到信息。这样的中间结论具体规定了应继承什么。阶梯是一套贡献分解，层次高低本身不表示理论先进程度。

15.3 新增信息的理想收益与实际代价

读数事件主张保留更多与任务有关的条件，但“更多信息一定更好”需要精确定义。在一个理想化的平方损失预测问题中，如果可以知道真实条件分布，并且所有变量具有所需有限矩，那么增加信息不会提高最优预测风险。这个结论说明信息可能具有价值，却不能直接保证有限资料中的具体模型一定受益。将理想界限与实际估计分开，才能正确评价事件档案。

设结果为 Y，原信息为 X，新增记录为 Z。只使用 X 时，平方损失下的最佳预测是 Y 在 X 条件下的数学期望；使用 X 与 Z 时，则是更细条件下的期望。这里的期望算子表示数学期望，应与 SDE 中的特征纠缠 E 区分。两个最优风险的差，等于两种条件预测差异平方的平均。

$$m_X = E\left[Y \mid X\right], m_{XZ} = E\left[Y \mid X, Z\right].$$

$$E\left[\left(Y - m_X\right)^2\right] - E\left[\left(Y - m_{XZ}\right)^2\right] = E\left[\left(m_{XZ} - m_X\right)^2\right] \geq 0.$$

证明来自一个直接分解。把只使用原信息的预测误差写为完整信息下的误差，加上两种预测之间的差。展开平方以后，交叉项的期望为零，因为在完整信息条件下，结果减去其条件期望的平均为零，而两种预测之差是这些信息的函数。于是剩下非负平方项。该证明没有假定新增记录来自主体、互动或任何特定理论，因而信息增益本身不能证明 SDE 的独占性。

$$Y - m_X = \left(Y - m_{XZ}\right) + \left(m_{XZ} - m_X\right).$$

如果新增记录不改变条件期望，右侧风险差便为零。它仍可能改善分布尾部或其他任务，但对当前平方损失下的均值预测没有理想增益。这个结论有助于约束无限记录冲动：变量具有哲学意义，不意味着它对每项预测都有必要；字段可以在某个任务中无用，在另一个任务中重要。任务与信息的关系需要实际检验。

现实模型并不知道真实条件期望。增加变量可能引入估计误差、过拟合和缺失处理负担，记录本身还可能改变过程。即使理想风险降低，取得信息的成本也可能超过收益。因此应区分不可约的不确定性、模型近似误差、有限样本估计误差和采集代价。一个档案在理论上包含更多信息，却因样本不足无法有效使用，并不矛盾。

这条分解还需要经过有限样本学习才能转成实际选择。可以固定训练资料规模与评估任务，分别记录少字段和多字段模型的测试损失，再观察新增记录的取得成本。如果多字段只在罕见高损失情境中有价值，就须单独报告该条件，不能让总体均值抹去收益；若改善随训练样本减少而消失，则其使用范围应包含最低资料条件。按需采集还需在结果出现前定义触发条件，否则会以未来结果决定是否取得当前信息。理想增益提供方向，实际增益及其取得时点才决定字段能否进入长期系统。

15.4 拟合 数据外检验与机制判断

模型在训练资料上表现良好，说明它能够表达已经看到的关系，却不能自动说明这些关系可以重复。数据被用于选择变量、调整形式和决定阈值以后，最终拟合已经包含选择过程。继续用同一资料评价模型，容易把偶然结构当作能力。独立检验的作用，是让模型面对尚未参与自身形成的结果，从而约束这种选择自由。

资料分割必须与研究问题匹配。若同一个人的相邻记录同时进入训练与测试，模型可能利用个人身份和近邻状态，得到很好的测试表现，却未必能够迁移到新个体。若未来资料的信息进入了特征构造，时间预测也会被人为简化。按个体、会话、机构或时间分割，各自对应不同外推目标，不能把一次随机拆分的成功说成所有迁移都已通过。

模型复杂度的评价与外部预测也应分开。Schwarz 提出的准则在相应模型与抽样前提下对似然拟合加入参数维度惩罚。[24] 常用写法由最大似然项与样本量相关的复杂度项组成。它是在特定框架内组织模型比较的工具，不是独立测试集损失，也不是某个因果机制成立的证书。把准则改善两点写成新本体论被证实，远远超过它承担的任务。

$$B I C = -2 \log \hat{L} + k \log n .$$

外部检验同样不能只看一个平均误差。一个模型可能总体准确，却在某类条件下持续失准；也可能平均误差略大，却对重大错误更保守。评价应包括与用途相连的分层表现和不确定性校准。若模型宣称某个范围具有较高覆盖概率，就应检查实际结果是否与宣称相称。准确性与诚实表达不确定性共同决定模型可被怎样使用。

机制判断需要更进一步。一个灵活预测器可以通过多种替代关联获得成功，而不揭示目标通道。若要证明某个回写机制，仍需第 14 章所讨论的时间、干预和竞争解释。预测增益可以支持模型值得保留，却不能替代因果辨识；反过来，一个因果通道得到识别，也不保证在每种预测任务上都有显著优势。两种成果应分别报告，以免彼此借用权威。

当外部检验失败，研究者应保存失败的具体形态。是所有条件都偏离，还是特定群体失准；是均值错误，还是不确定性过窄；是模型形式问题，还是输入资料质量下降。这样的失败记录能够指导修订。只报告一个总体评分，会把有方向的失配重新压成无解释的残差。模型比较本身也需要事件化，让评估结果能够返回自己的产生条件。

15.5 历史样本中的迁移与信息时点

第 11 章已经区分发生时间与可见时间；历史模型比较要进一步把这一区分写成检验协议。在每个预测起点，训练集、变量定义和可用版本都应固定下来，随后只评价当时实际能够生成的预测。若某个输入要到结果出现后才修订，就不能把修订值无标记地放回原起点。原版本预测与修订资料下的重算可以并列：前者回答当时能做什么，后者帮助定位资料修订造成多少差异。两项结果分别保存，回测才不会把后见之明登记为模型能力。

实时档案需要同时保存对象发生的时期和资料公开的时期。某项经济活动属于第一季度，初值可能在第二季度发布，后续又多次修订。评估某个历史决定时，应尽量使用决定当时能够取得的版本。ALFRED 的官方说明提供了历史版本与当时可用信息的组织方式，可以作为这种设计的数据基础；拥有该渠道不等于已经完成有效回测。[25]

滚动检验是一种自然安排。模型在一个时间点只使用此前资料训练，对后续窗口预测；随后窗口向前移动，记录表现怎样变化。每一轮允许的更新规则应预先说明，不能在看到失败以后临时决定哪次重训属于模型、哪次重训不算。若系统本来就具有持续学习，更新能力可以成为被评价的部分，但必须把更新所需资料与时间成本计入。

历史迁移还需要区分资料修订与世界改变。一个参数在不同版本中变化，可能是对过去认识的修订；一个参数在不同年代变化，则可能反映对象或制度改变。两者可以同时出现，却应沿不同时间轴分析。把版本变化直接当作对象漂移，会夸大世界的不稳定；把真实时期差异全部视为统计误差，又会错过发生条件的变化。

滚动窗口还应处理更新和评分之间的时间差。设想一个模型每轮取得上轮已公布结果后更新，而另一个按固定参数运行；比较时须让二者面对相同的结果公布节奏，记录更新完成前的等待成本。若新结果延迟，更新模型就应依照事先规则使用旧状态或暂缓输出，

而不是在回测中假定新标签即时存在。窗口中的缺口与不可见资料同样属于任务。只有把这些现实限制纳入协议，才能判断持续学习的收益来自更新规则，还是来自理想化的信息便利。

历史研究最有价值的结果，往往不是找到一个永远正确的参数，而是说明哪些关系在条件改变后仍然保留，哪些需要更新，以及更新依据在当时是否可见。这样的结果直接服务于模型治理：什么时候继续使用，什么时候降低信任，什么时候需要重新定义目标。发生模型的历史厚度，应体现在这些可操作区分上，而不是体现在事后叙事包含多少背景。

15.6 多分布研究中的稳定结构

迁移可以发生在人群之间、机构之间、时间之间，也可以发生在不同任务和工具之间。把所有变化统称为环境变化，会遮住模型究竟需要适应什么。某些迁移只是输入构成改变，给定条件下的结果关系仍然稳定；另一些迁移改变了结果生成关系；还有一些直接改变被估量和评价目的。不同变化需要不同策略，不能用一种重新加权或重训操作包办。

设想一个评分模型在两个机构使用。第二机构的对象构成不同，但同样条件下的响应关系保持一致，适当调整构成可能恢复总体预测。若第二机构的操作规则改变了条件与结果之间的关系，仅调整构成就不够。若第二机构用分数决定完全不同的事项，问题甚至超出预测关系，涉及用途效度。事件记录能够帮助区分这三层，但只有资料和设计支持时才能真正识别。

多分布比较不应只追求所有场景表现相同。某些差异正是模型正确反映条件变化的结果。一个能够预测不同条件下分布怎样改变的模型，比一个输出始终稳定却持续错误的模型更有知识。但稳定结构仍然重要：它可能是更新规律、条件关系、转换方式或误差范围。发生学要寻找的是恰当层次的不变量，而不是把一切稳定都看作对象被压缩后的残余。

可以在多个场景中同时考察中心、尺度、响应方向和分布分裂，观察哪些变化由同一条件关系组织。若模型只有在每个场景单独重新拟合后才成功，它提供的迁移内容可能很弱；若少量可测条件能够预测新场景的变化，则具有更强增量。这里的“少量”不只是参数个数，还包括取得条件的成本和是否需要事后知道结果。一个依赖未来标签的适应过程不能被称为即时迁移。

验证稳定结构可以采用逐个留出场景的比较。研究者在若干机构上确定模型结构，再把一个未参与结构选择的机构作为迁移目标，分别评价完全沿用、只更新指定参数和重新

估计。三种安排使用目标资料的数量须明确。若局部更新以很少新资料达到重估水平，便取得了可继承结构的证据；若仍需大幅改变函数或重新定义结果，原迁移声称应收缩。这个比较并不要求各机构分布相同，它检验的是事先指定哪些关系值得保留。

多分布研究也提供反例资源。一个在原场景成立的机制，在另一场景缺失，可能揭示必要条件；一个原以为关键的字段在多个场景都没有增益，可能应从常规记录中删除。迁移失败不只是理论的损失，也可以成为识别机制边界的材料。前提是失败被保留，并且新解释在新的资料中再次接受检验，不能无限使用同一批例外来修饰理论。

迁移还可以是局部的，而不必在全盘成功与完全失败之间二选一。一个模型的测量部分可能保持稳定，行动规则却随制度改变；某些个体差异可以共享估计，另一些必须在新场景重新校准。把模型拆成有明确职责的部分，可以研究哪些部分值得继承。这样的拆分需要机制根据，不能仅在结果出来以后把成功部分称为核心、失败部分称为外围。继承范围应在新资料到来前尽可能说明。

新场景资料较少时，完全沿用旧模型与完全重新估计之间也存在中间选择。研究者可以允许一部分参数调整，并用原场景知识约束其合理范围。但约束越强，越可能把旧偏差带入新场景；约束越弱，则越依赖有限新样本。比较这种权衡能够使“历史如何进入现在”成为实际建模问题，而不只是哲学描述。历史可以提供资源，也可能形成惯性，两种作用都须接受新任务的检验。

迁移失败的程度也应与行动联系。短期误差略增可能仍在用途允许范围内，某种罕见但方向一致的错误却可能要求暂停使用。平均表现不能替代严重失败的分析，单次严重失败也不必推翻全部用途。研究者应同时报告总体表现、关键条件和不可接受后果，使模型的保留、缩域与停用具有可讨论的依据。

15.7 干预后的预测比事后拟合要求更多

干预改变了系统的发生条件，因此对干预后结果的预测，是检验模型机制内容的重要机会。一个依靠稳定关联的模型，在干预后可能失效；一个正确描述某些更新通道的模型，可能保留更多预测能力。但这不是保证，因果模型也可能遗漏其他通道，干预还可能改变目标或实施方式。所谓机制模型应在具体干预集合中接受比较，而不能凭名称享有迁移特权。

设想教育系统依据错误类型提供不同反馈。原数据中，反馈形式与学生状态同时变化，预测器可能借助这种关联表现良好。一旦研究者改变反馈分配，原关联就不再成立。

若模型描述了反馈怎样影响后续策略，便可能预测新安排下的结果；若只是把反馈标签当作学生类型的替代，它就可能失败。这种干预提供了区分解释的机会，前提是实施和结果记录足够清楚。

干预后的预测需要保存三个不能相互替代的成绩。第一项是在看见新结果前冻结模型所作的预测，检验旧知识能直接承担多少改变；第二项是在预定少量新资料下完成指定更新的预测，检验调整规则是否有效；第三项是充分取得新资料后重新建模的结果，说明当前任务能够达到的参照水平。三者应使用各自独立的评价资料。若只展示第三项，就无法知道原模型究竟继承了什么；若只允许第一项，又会排除本来以适应为功能的系统。

反事实问题还要求说明干预是否可实行。把一个变量数值改掉而保持其他关系不变，未必对应现实操作。例如直接指定一个人的期待，同时假设表达、经验和行动都不变，可能没有明确实现方式。模型可以使用这种抽象研究逻辑后果，但现实结论需要连接到具体操作。发生条件相互规定，正意味着干预常常不是孤立地旋转一个旋钮。

此外，干预成功可能使原来用于辨识的关系消失。若某个错误反馈环被纠正，后续残差结构不再出现，这可能是机制解释得到支持的结果，而非理论失去对象。研究者需要事前说明这种消失的预期，并检查是否存在其他导致消失的变化。反过来，若干预只改变了记录而没有改变实际结果，漂亮的新指标也不能被当作修复成功。

干预后预测使模型不仅说明已经发生的结果，还要在改变条件时承担可能出错的判断。这种承担并不要求所有研究都立即部署重大干预。可以从低风险、可逆的小范围操作开始，逐步扩大证据。每一步都比较收益、损失和恢复条件，使理论的创造性通过可检验的安排进入世界，而不是停留在对既有资料的重新叙述。

15.8 模型选择也是一种可修订的制度

实际使用中，模型不是一次比较后永久获胜。资料会变化，任务会改变，维护资源也会减少。一个原先复杂而有效的模型可能变得难以更新，一个简单模型可能在新条件下重新足够。模型治理因此需要把适用证据、更新触发与退出规则一起保存。选择模型不是给理论授予终身地位，而是在当前任务和条件下作出可解释的决定。

成本应进入这种决定，但不能只按运行速度计算。采集敏感资料的代价、解释结果的难度、错误发生后的修复能力和对使用者的负担，都影响模型价值。一个略微提高平均精度却显著增加不可复核决策的系统，未必更适合公共用途；一个精度相当但能定位失败原因的模型，也可能具有重要优势。价值判断需要明确，不能藏在技术评分背后。

这种选择可以落实为可追踪的决定。对监测器，研究者先说明在规定输入范围内，两个模型分别漏掉什么、误报什么及需要多少资料；使用方再依据材料损伤与停机代价选择用途。如果提高灵敏度需要更多复测，复测负担也应进入同一决定。几个月后用途改变，可以回到原权衡重新选择，无须宣布旧模型从未有效。这里保存的不只是最佳分数，还包括为何这一误差结构在当时足够，以及哪些条件变化会使原选择失去理由。

发生模型本身应接受同样治理。若十字段档案中某些记录长期没有收益，可以简化；若新增回写通道不能改善干预判断，应缩限机制声称；若一个成熟既有模型已经充分处理任务，应保留它。另一方面，当制度不断遭遇可重复的条件失配，却拒绝增加必要信息或修改用途，简洁也会变成逃避。方法的经济性与认识的充分性需要一起检验。

第 11 至 15 章由此形成分工明确的工作链：事件档案给出结果身份，装置分析解释稳定操作如何形成，数学判据区分不同限制，因果辨识确定可改变的通道，模型比较决定哪些表达值得继承。每一环都向后一环交付具体条件，同时保留返回修订的入口。一个模型可以在预测任务上被采用而机制仍未定，也可以提供可靠机制却暂不适合部署；把这两种位置写清楚，是本编相对于笼统“解释力”判断取得的实质成果。

生命材料会使这套分工面对更强的历史要求：同一评分可能来自不同感知过程，同一细胞状态可能保留不同承接能力，同一稳定输出也可能依赖不断更新的内部组织。味觉、细胞、神经与恢复研究各有成熟方法，需要按其观察条件建立具体模型。进入这些领域时可继承的是已明确的判据与比较责任；哪些状态、通道和稳定关系实际成立，仍须让材料决定。

第四编 生命 神经与感知的发生

第 16 章 味觉中的三属性发生

16.1 一杯甜饮料中的事件边界

一杯饮料在配方表上只有一组糖酸浓度，在品尝者的叙述中却经历入口、展开、转折和余味。第 1 章借甜味划定发生单位，第 11 章说明读数必须带上用途与记录条件；这里进一步追问，哪些时间差异参与了甜感的形成。总体评分相同的两次品尝，可能在达到峰值的时刻、强度保持的时段和余味的衰减上不同。本章以相同评分下的不同时间进程为起点，把材料接触、感知更新与报告动作之间的时间关系作为独立解释任务。

设想一个构造案例：两位参与者先后品尝同批饮料，给出相近的总体甜度。第一位在入口时觉得明显甜，随后迅速减弱；第二位开始感觉平淡，含留之后才觉得甜味突出。这里没有假定任何真实实验结果，只是在提出相同评分下存在不同时间进程的可能性。如果研究目标是一次性购买选择，总体评分或许已经够用；如果目标是改进饮用过程中的体验，两个相同评分便掩盖了需要分别处理的过程。属性描述的充分性须与用途共同说明。

对上述差异，首先需要确定共同的时间起点。从举杯开始计时，保存的是整个操作过程；从溶液接触口腔开始计时，保存的是接触后的展开；从参与者报告察觉开始计时，则已经把察觉之前的一段时间排除。三个起点各有用途。如果两位参与者的操作速度不同，以举杯为零点得到的峰值错位，可能在接触对齐后缩小；如果接触对齐后仍存在差异，才需要继续分析感知和表达过程。这样，事件边界便具体地进入了时间机制，而不只是附在曲线旁边的一条说明。

所谓客体甜，是在某种制备、接触和判断制度中能够归还给样品的稳定差异。它依靠配方可重复、批次可比较，也依靠品尝任务具有足够明确的共同规则。互动甜侧重入口方式、含留过程和报告动作如何使这种差异展开。主体甜则侧重既往比较、期待与辨别能力怎样使某一体验获得确定意义。这三种侧重都涉及显露、差异序列和特征纠缠，不能把样品直接叫作显露，把接触直接叫作差异，把人直接叫作纠缠。

这一界定改变了通常的归属争论。即使参与者报告稳定，主体号位也没有消失；训练、共同参照和明确任务可能正是稳定的生成条件。即使不同参与者报告分散，也不能立刻认为客体性已经瓦解；材料差异可能仍在每个人内部保持一致的排序。研究需要把组间

差异、个体内部排序和重复试次变化分开。三号位的意义在于追问稳定如何取得，并非预先要求所有研究都出现不稳定。

事件边界还决定谁与谁能够比较。一杯饮料的甜味强度、它的可口程度和人愿意再次饮用的倾向并不相同。有人可以认为饮料很甜却不喜欢，也可以觉得甜度适中但不愿购买。把这些判断压入一个甜度分数，会让对配方的物理比较悄悄承担价值判断的任务。发生分析因此先保持问题的区分，再考察它们如何在一次品尝中互相影响；这比把全部差异归入主观性更能发现新的研究对象。

16.2 从配方到报告的多层约束

浓度、受体活动、神经反应和甜感报告可以被置于同一研究中，却不因此成为同一种量的四个名字。浓度属于材料描述，细胞和神经记录属于特定生理测量，报告则包含感受与判断的表达。它们之间需要经验桥梁。某一层改变而另一层保持稳定，可能来自补偿、测量灵敏度差异或记录任务不同。不能只因两个量相关，就认定其中一个已经完整代表另一个；也不能因关系并非一一对应，就放弃建立条件明确的映射。

把映射写清楚，能够为 SDE 提供具体工作。假设研究者选取一段连续甜感报告作为显露，差异序列包括刺激呈递、口腔接触和判断更新，特征纠缠则指这些特征以何种依赖方式共同规定可辨别的甜感。这里的纠缠不是装有所有遗漏因素的容器。例如，同一个标签只在某种呈递顺序下改变判断，便需要研究标签和顺序的条件关系，而不能分别加一个标签变量和一个顺序变量后宣告问题解决。

尺度桥梁可以先从三个时点建立：刺激实际接触、参与者作出可见报告、记录设备保存响应。只有最后两个时点被准确记录时，仍需估计接触的不确定性。例如，假设报告晚了一段时间，既可以由甜感形成较晚引起，也可以由发现甜感之后的按键较慢引起。增加一个独立的动作校准任务，能够限制后一种解释，却不能保证校准中的反应速度完全适用于品尝。模型应保留这段残余不确定性，使时间机制的解释精度与实际证据相称。

材料控制的价值也由此得到保留。对同批样品进行编号、记录准备时间和呈递条件，能够降低无需诉诸新机制的解释空间。若未记录批次差异，却以参与者的历史说明所有读数变化，主体分析只是替材料管理的缺口承担了不恰当的责任。客体性的建立要求可追溯的制备程序，主体性的研究也同样依赖它。把二者理解为共同发生，意味着研究一种侧重时必须给其他侧重提供足够可靠的条件说明。

量表是一项尤其容易被忽视的约束。端点文字、示范样品和评分次数共同规定数字如何使用。假设某参与者把最高分留给日常经验中极甜的食物，另一位把最高分留给本次实验中最甜的样品。同样的数字就有不同的参照范围。训练可以减少这种差异，但训练本身也在建立研究所需要的判断方式。量表校准取得的不是脱离历史的纯粹内心读数，而是能够服务明确比较的报告制度。

这并不意味着所有报告都只是语言游戏。若浓度变化在多种报告方法下产生有方向的一致差异，材料约束就获得了更强支持。若某标签效应只出现于一种措辞而不能迁移到辨别任务，则它的解释范围应相应收窄。多层证据的作用在于建立相互限制的关系：报告可以提示生理测量未捕捉的问题，生理测量也可以限制对报告变化的过度解释。没有哪一层因更靠近仪器便自动成为最终裁判。

16.3 混合效应中的加和与重组

酸甜混合提供了一个清楚的起点：混合物中的甜感不必等于单独糖溶液甜感的直接搬入。Green 等研究了单一味刺激及其多种混合，用同一评价框架区分具体味质和总体强度。研究中的混合抑制与混合内部感知强度的加和可以并存。这提示，加和是否成立取决于加和的对象究竟是混合前的单项刺激，还是混合后已经形成的感知分量。[26]

这一差别对属性发生论尤为关键。若把混合前的甜感和酸感当作保持原样的零件，混合之后偏离其总和便显得神秘。若先承认混合改变了各分量的成立条件，那么混合内部仍可呈现某种稳定关系。发生分析并不需要排斥加法。它要确定哪些量已经通过共同过程生成，然后才讨论这些量是否可加。经过发生而得到的可加性，与未经检验便假定各要素独立，是两个不同的主张。

但抑制这个名称仍不足以提供机制。甜感评分下降，可以由感知层面的改变、注意分配、报告策略或它们的组合造成。若研究只测一个最终评分，便无法从结果的方向区分这些可能。SDE 所要求的增量，是找到能够使竞争解释分开的条件，而不是在下降曲线上加注三个号位。例如，在同一材料下改变问题措辞，与在同一措辞下改变入口顺序，分别提供不同的证据；两种变化一致也仍需进一步分析共同原因。

Schifferstein 对果糖与柠檬酸混合的情境研究，把内部表征和反应选择作为需要区分的环节。其问题设置表明，情境对混合刺激和单一刺激的影响是否相同，本身就是经验问题，不能预设。[27] 本书由此提出一个方法判断：既然抑制量常由两种评分的差得到，

研究必须说明这两种评分是否经过相同的报告映射。若映射随情境不同而改变，差值可能同时包含感知变化和量表使用变化。

考虑另一个构造例子。研究者在甲情境中只呈递低甜度范围，在乙情境中加入高甜度样品。某混合物的评分下降，而它与相邻参照样品的辨别正确率保持不变。这一结果首先要求检查报告尺度是否被重新展开。若辨别表现也改变，则单纯尺度解释面临压力。例子的作用不在于预言真实结果，而在于说明同一个评分变化可以推动不同的后续问题。理论成熟的标志是能够安排这样的分歧检验。

酸甜混合还揭示关系的非对称性。即便某种配方下酸对甜有抑制，也不能据此推断甜对酸的影响等大，更不能把不同浓度区间的方向一概固定。研究需要分别确定目标味质、刺激区间和比较基准。E 作为特征纠缠，正是在这些不对称、条件化和可能改变的依赖中获得内容。它不是一句一切互相影响，而是可以被局部识别、可能被不同模型解释的具体组织。

16.4 期待怎样进入当下的甜

期待既可能改变体验，也可能改变人如何表达体验。Woods 等在特定甜饮料任务中发现，较高甜度期待对应的条件同时改变了报告和味觉皮层相关活动。这个结果支持在其任务范围内进一步研究期待参与味觉加工的机制，却不足以把所有报告差异都还原成期待。[28] 对发生分析而言，其价值在于提示，材料到报告之间存在可以被研究者操纵并需要独立检验的历史条件。

期待不能被简单当作从外面加到甜上的一个数。一个人听到甜这个词时，已经在使用既往形成的比较序列；当下的饮料又可能修订这种比较。标签所产生的期待和体验所更新的期待并不必然相同。若前者只改变本次描述，后者却改变下一次选择，两者便承担不同的因果角色。主体号位的研究由此转为考察参照、辨别和承接后果的能力，而不是寻找一个完全自足的内心原因。

期待的承接可以通过线索的可信度变化进一步分析。假设某标签在训练初期可靠地预示较甜的样品，随后却反复与实际体验冲突。如果参与者逐渐减少对标签的依赖，研究需要辨认被更新的是标签判断、甜度类别，还是作答时避免受提示影响的策略。三者在无标签的新样品上应有不同后果：标签判断的修订可以只影响带标签试次，类别重组则可能改变无标签比较，作答策略还可能随着提问方式而改变。期待因而成为有特定作用对象和迁移边界的历史状态，而非每次差异都能调用的总原因。

期待的可操纵性也带来一个研究悖论。为了知道参与者在品尝前期待什么，研究者会询问预期甜度；但询问可能使参与者把注意集中到甜度，并生成本来并不突出的期待。于是期待测量既是信息来源，也可能是条件干预。解决办法不是停止询问，而是区分提问的目的。若目的是预测，提问形成的任务可以作为研究对象；若目的是估计未被询问情境中的体验，就需要比较主动提问和较少打扰的程序。

标签效应消失同样有理论价值。它可能说明材料线索足够强，期待在该任务中没有额外作用；也可能说明标签并未形成预期，或参与者经过重复后学会不信任它。几种解释对后续试次的预言不同。研究者应检查操纵是否到达其目标，而不是把统计不显著解释成主体缺席。反过来，显著标签效应也不能被直接写成主体构成体验的全部证据，因为需求特征与策略性作答仍可能产生类似结果。

期待还具有社会来源。商品名称、熟悉的容器和他人的评价都可以成为有待研究的线索，但这些线索不应被一次性塞进同一个心理背景变量。容器可能提供视觉线索，评价可能建立可信度，名称可能激活类别。把来源拆开，才能确定哪种历史进入了哪一段差异序列。当这些线索相互冲突时，发生理论的问题更具体：主体怎样分配可信度，哪种冲突改变了后续比较规则，以及这种改变能否跨材料保留。

16.5 时间曲线与跨轮历史

连续时间记录的意义在于保存过程，不在于它天然比单点评分更真实。要求参与者持续移动评分器，可能增加注意负担，改变含留节奏；在品尝结束后回忆整段曲线，则又引入记忆重构。两者是不同的测量安排，都有明确用途。研究可以比较它们能否预测后来选择或辨别表现，但不能仅因曲线包含更多数据点，就宣告它已经接近未经干预的体验本身。

过程资料需要区分至少两种时间。一种是单次品尝内部的展开时间，另一种是连续试次之间的历史时间。前者可能帮助研究峰值和余味，后者涉及适应、学习、疲劳和参照更新。如果把两者混成一条长序列，模型可能用试次之间的缓慢变化去解释单次内部的转折，也可能用单次曲线的形状掩盖长期的评分漂移。时间戳的精细并不能自动解决这种层次混淆，事件划分依然需要理论理由。

一个适于推演的构造方案是，让两组经历不同的训练序列，随后接受相同且不再带标签的测试材料。如果两组仍出现可重复差异，过去的安排就可能通过某种状态进入现在。然而差异也可能来自对实验者意图的猜测或量表习惯。进一步的辨别任务和改变报告方法

的测试，能够对这些解释施加约束。历史效应的发现只是入口，历史以何种机制继续发挥作用才是发生研究的中心问题。

回写要求指出被改变的是什么。一次体验可能更新对标签可靠性的判断，也可能改变后续选择的材料范围，或者使参与者改变报告速度。这些路径的后果不同。若材料范围改变，下一轮研究面对的是新的输入分布；若只改变报告速度，则观测映射改变；若辨别能力改变，则可得经验的组织方式改变。把三者统称为反馈，会丢失对干预位置最有价值的信息。

历史的作用也可能表现为稳定。经过参照训练后，同一参与者能够在不同日期维持相近的排序，这说明历史塑造了可保持的判断。此时不应把收敛当作发生停止，而应研究训练如何筛选差异、如何形成校准，以及校准在哪些新条件下仍有效。若新材料带来排序反转，既可以检查旧参照的适用范围，也可以研究新的区分怎样形成。稳定与再学习由同一段历史连接，并不互相排斥。

对曲线进行压缩，需要保留与后效有关的时间结构。用一个纯粹构造的算术例子说明：把一次品尝分成三个等长时段，甲的报告依次为零、四、二，乙依次为二、四、零。两者的峰值相同，各时段强度乘时长后的总和也相同，但末段状态不同。若下一次判断只依赖离开本次品尝时的状态，两个摘要都不足以预测后效；若后续结果只与全程累积有关，这种压缩反而可以成功。峰值、面积和末段值因此对应不同机制承诺。选择摘要量的依据，是它能够保存哪种承接关系，而非它比原曲线更方便或更复杂。

16.6 甜感过程的定位与重组

时间资料把三方程的相互规定转成一个定位问题：当前报告的转折发生在材料接触、体验组织还是表达环节？三者可以同时变化，但研究必须说明每一项变化对其他环节施加了什么约束。若甜感减弱后参与者加快吞咽，新的接触时长又进入后续体验，报告便同时是先前过程的显露和下一段差异的条件。把这段联系纳入单位，可以提出具体后效预测，而无需先断言哪一号位独自拥有甜。

一个构造模型能够揭示定位的困难。假定记录值在每个时点都是体验强度与报告增益的乘积，那么同一条下降曲线既可以由强度下降、增益保持产生，也可以由强度保持、增益下降产生。曲线记录得再密，两种分解仍可能完全相同。要区分它们，需要获得对某个环节的额外约束，例如在可比时点安排辨别任务，或者改变报告装置而保留接触程序。信息不足在这里表现为机制不可辨识，增加采样点并不必然解决。

反过来，若某种过程差异跨报告方法保持，而改变接触时长可以稳定地移动该差异，在操作条件可比时，就有理由优先检查接触过程及其感知后效。这一结果缩小了解释范围，但两种报告方法共有的判断策略仍须纳入比较。若差异主要随量表参照改变，材料操作的作用很小，则应优先研究判断映射。两种结果都推进了味觉研究，因为它们指出下一次值得改变的环节。定位逐步形成，所需的是不同操作对候选机制施加的联合限制，而非单项变化与一个内部环节的一次对应。

六路径在这里用于安排实际入口。已有可信的材料与量表校准时，可以沿沉淀型入口改变接触序列，考察时间形状；已有一种反复出现的余味异常时，可以沿进入型入口检查相容的材料、注意和表达条件，再安排区分试次。前者利用已知条件制造差异，后者利用可靠显露寻找条件。入口不同，对资料的要求也不同：前者尤其需要操作记录，后者尤其需要确认异常不是一次偶然报告。

当参与者能够辨认余味差异却持续给出相同总体甜度时，显露与现有区分方式之间出现张力。此时可以增加一个余味维度，使特征组织从单一强度轴转为包含时间位置的描述。新维度只有能够支持稳定比较才值得保留；若只是为每个样品增加一个事后标签，便没有提高解释力。三原理所追踪的改变由此落实为一个可以回到资料的动作：新的区分是否使原先互相矛盾的报告获得一致解释，并改善下一次比较。

进一步的反例是，训练后出现了更细的语言，却没有出现更好的辨别或迁移。它支持表达能力改变，却不足以支持感知分辨率提高。反向情形也可能成立：辨别改善而描述词汇不变，提示原量表未能保存新差异。把语言、辨别与后效分开，可以辨认主体号位的不同承接方式。主体的参与因此在具体能力上取得内容，而不是由报告是否复杂来估算。

本章得到的研究对象，是在规定时间窗口中能够被材料操作、比较任务和表达程序共同限制的甜感过程。其可重复性可以表现为曲线形状，也可以表现为时间摘要与后效之间的稳定关系。含序列更新的心理物理模型为这种研究提供直接基线；SDE 增加的工作，是检查基线遗漏了哪一段承接，并说明打开这一段以后能够得到什么新的定位结果。跨领域竞争检验将在第 38 章展开，此处首先把味觉问题本身建立清楚。

16.7 从解释进入可执行的研究

具体研究可以从时间差异的重现开始。假设要解释两组饮料在总体甜度相近时为何有不同余味，先导阶段应分别确认接触持续时间、末段报告和下一次判断的关系。酸甜组成与期待条件可以作为有理由的候选因素，但初始目标是确定差异位于哪个窗口。若差异只

在入口对齐之前出现，应先改善呈递；若在接触对齐后稳定存在，才值得扩大到感知条件的操纵。每次增加比较，都应使一个已经明确的时间问题得到解答。

执行程序应使材料制备者、呈递者和数据分析者各自知道完成任务所需的信息，同时减少无意的提示。重要的不是追求一种抽象的绝对盲法，而是记录哪些信息在什么时候可见。若参与者察觉了配方线索，应保留这一事实；若呈递者无法遮蔽某条件，也应说明其可能影响。发生档案保存的是实际发生的实验，不能只保存计划书上理想化的实验。

顺序安排需要同时处理比较与承接。随机化可以降低某些系统性偏差，平衡顺序可以减少条件与位置的混淆，但它们不会自动消除记忆。若关注跨轮回写，历史正是研究对象，不能先用统计处理把它抹去再声称没有历史效应。研究者应分别规定哪些序列作为干预，哪些序列作为需要平衡的条件，并在解释中保持这种角色的区别。

采样方案也需要覆盖机制可能分开的时段。若两个解释只在接触结束后的短窗口产生不同预言，把多数采样放在入口阶段便无法裁决它们；若关注的是跨试次承接，就应保留下一次刺激之前的间隔和当时状态。可以先用较稀疏记录定位转折，再在新的样本中针对该窗口增加观察，避免在同一资料中反复寻找最有利时点。持续报告与分时报告具有不同负担，选择依据应是它们是否保留关键窗口，并是否改变所要解释的饮用进程。

主要终点也需要提前确定。如果预先关心的是取消标签后的后效，就不能在后效不明显时改用当轮峰值作为同一假说的胜利。探索可以发现新问题，但它应当产生下一项独立检验。个体和时间块的划分尤其重要：同一人的相邻试次若同时进入模型训练与检验，模型可能只是识别熟悉的报告习惯。是否迁移到新人、新日期或新材料，应根据研究声称的用途逐项评估。

一项研究的有效产物不必是三属性机制获胜。它可能发现，已知适应与量表模型已经足以解释现有差异；也可能发现，过程记录提供增益，但没有支持特定回写机制；还可能发现，新的历史操纵产生了原模型未能预测的结果。这三种产物都能够改进研究。只有把概念收益、信息收益和机制收益分别说明，才不会把整个纲领的命运押在某一个显著性结果上。

16.8 可迁移的甜与不可省略的历史

味觉案例的深层贡献，是使属性稳定性获得可追问的生成条件。对某一产品配方，研究者可能最终建立一个相当简洁的预测式；这个结果无需被复杂性叙述削弱。它的可靠性来自材料、群体、报告制度和用途之间的匹配。若这些条件在新场景中仍成立，简洁模型

可以继续使用。发生分析所提供的，是对何时应保留模型以及何时需要重新打开过程的判断依据。

跨文化或跨群体比较尤其需要这种判断。若两个群体使用不同的日常甜度参照，直接比较平均分可能混合感受、分类与量表使用差异。增加共同训练可能提高可比性，却也把原来的生活情境改造成实验情境。两种设计都可以有价值，但回答的是不同问题：一种研究各群体日常判断如何成立，另一种研究在共同参照下如何比较。不能把后者产生的一致性当作前者差异不真实的证据。

从实验室走向实际饮用，还会改变事件边界。实验中单独评价甜度，日常消费中却可能同时涉及温度、气味、饱足感和社交场景。并非所有这些条件都必须进入每个模型；只有当它们改变目标判断或限制外推时，才应扩展记录。一个能够明确说明自身适用范围的局部模型，比一个声称包含全部生活却不能区分预测的总模型更接近发生研究的要求。

对于创造性任务，SDE 还提供了另一种用途。已有配方没有满足体验目标时，可以从被忽略的过程差异寻找新设计，而不只沿糖浓度轴增加或减少。这里的创造并不意味着必然得到更好的产品，而是扩大可检验问题的空间：某种余味是否值得独立评价，某种呈递方式是否改变比较基准，某种标签是否使体验目标发生偏移。每一个新问题都必须重新进入材料和行为的约束，不能停留在命名的兴奋中。

味觉中的主体性同时提醒研究者，个人报告应受到尊重，但并非所有报告都回答了研究者原先的问题。参与者可能在描述喜欢程度，研究者却把它记为强度；参与者可能在比较前一杯，研究者却按绝对尺度解释。研究的改进首先是使表达与问题互相校准。主体号位进入科学，不意味着任何解释都获得同等证据地位，而意味着判断的成立条件应当能够被当事人表达、被程序记录并接受共同检验。

从一杯饮料出发，发生单位已经包含材料的可重复性、感觉过程的展开和历史条件的承接。它仍然保留清楚的局部边界，并没有变成无所不包的世界模型。下一步转向细胞时，不能把人的期待直接移植为细胞的心理，也不能把甜感报告移植为表达谱。可以迁移的是一个更精确的问题：当前状态通过何种物质过程取得稳定，又怎样改变下一轮响应能够发生的条件。细胞研究将使这个问题获得不同于自我报告的证据形式。

第 17 章 细胞状态与历史调控

17.1 一个状态为什么不能只是一张清单

两份细胞测量可能在所记录的表达指标上十分相近，随后却对同一刺激产生不同响应。这个构造情形提出了一个状态定义问题：表达清单是否包含预测下一步所需要的信息？如果没有，缺少的可能是尚未测量的分子条件，也可能是空间位置、先前刺激或细胞周期等因素。不能仅凭预测失败就宣称细胞超越物理规律，但也不能把不断出现的失配永久留在误差项中。状态的充分性需要相对于明确的响应任务接受检验。

通常将状态理解为某一时点的变量取值。这种理解并不排斥历史，因为历史可以通过当前的物质结构继续起作用。问题在于研究者是否选取了能够保存这种作用的变量。如果某个调控条件维持较久，短时表达水平便可能不足以代表它；如果先前刺激只留下暂时的分子存量，则未必需要诉诸稳定的表观遗传记忆。发生分析首先要求把历史作用的承载方式区别开来，再决定需要多长的记录窗口。

细胞也不是一个任意划定的小盒子。研究可能关心单个细胞的响应、同一谱系中的承接，或者细胞群体的分布。在第一种问题中，相邻细胞主要表现为输入条件；在第二种问题中，分裂使个体身份发生变化，承接关系成为核心；在第三种问题中，增殖和淘汰会改变组成。三种问题能够相互连接，但不能用群体平均的稳定替代单个细胞的稳定，也不能用单个细胞的轨迹直接说明群体比例为何改变。

所谓客体侧重在这里体现为可追踪的细胞身份及其条件稳定性。互动侧重关注接触、交换和刺激怎样进入响应。主体侧重若要使用，则只能指系统对自身维持具有选择性后果的调控组织，不能由这个词推断细胞具有人的反思体验。这三种侧重均是同一发生的不同观察重点。若研究只列出基因、信号与调控网络三个名词，仍未说明显露怎样依赖差异序列和特征纠缠，也未完成任何发生模型。

细胞状态可以被看作一组对未来条件具有约束力的可能性。这里的可能性不是脱离材料的潜能，而是对于给定刺激范围能够作出怎样响应的可检验陈述。相同的当前表达水平若对应不同的响应集合，研究者就有理由补充状态描述。若一个简单状态模型在声明范围内已经预测准确，则没有理由仅为符合理论图式增加无法识别的隐藏变量。SDE 的解释力取决于它能否改善状态的选择，而非变量数量的多少。

由此，细胞历史研究面对两个相连的问题。第一，过去怎样成为现在的物质条件；第二，现在的条件又怎样选择哪些未来差异能够被放大、缓冲或忽略。前者防止把记忆神秘化，后者防止把细胞写成被动存储器。只有两者结合，当前显露才能被理解为历史约束下的响应，同时又是下一段历史的起点。状态不再是一张静态清单，却仍然是可以通过测量与干预逐步限定的研究对象。

17.2 变异的来源与解释层次

同源细胞之间的差异不能一概归为测量失误，也不能一概归为有意义的历史回写。Elowitz 等构造了能够区分基因表达变异来源的细胞系统，显示生化表达过程自身的随机性和其他细胞成分的波动都可贡献总体差异。[29] 这一研究的重要启发是，变异可以被分解成具有不同机制含义的部分。承认随机性，并不意味着停止解释；研究随机性的生成与调节，本来就是机制研究的一部分。

对 SDE 而言，随机性与发生性并不冲突。差异序列可以包含随机事件，特征纠缠可以限制随机差异的传播方式，显露则可能是稳定分布而非稳定单值。若某种表达波动围绕相近分布长期重复，统计估计仍可能收敛。此时需要研究的是分布为何保持以及何时迁移，而不是把每一个分散的单细胞读数都命名为不收敛。随机结果的存在没有取消参数估计的可能性。

群体分布还可能隐藏组成效应。假设某刺激之后高表达细胞的比例上升。这既可能由原有低表达细胞发生状态转变造成，也可能由高表达细胞增殖较快或存活较多造成。仅有刺激前后两张分布图，不能区分两种机制。若将比例变化直接解释成每个细胞都被重新编程，便把群体层面的选择偷换成了个体层面的转变。谱系追踪与个体重复观测因此承担具体的辨识任务。

同样的区分也适用于反馈。两种蛋白水平在群体中相关，可能因为共同受到一个上游因素影响；某指标先变而另一指标后变，可能来自不同的测量延迟。回写需要更强的证据：改变某段响应能否改变随后输入的有效条件，或者改变下一次响应的规则。时间先后帮助排除某些解释，却不能独自建立闭环。发生分析的价值，是把需要证据的环节展开，而不是把相关网络自动升级为调控网络。

噪声一词的用途也应随问题而变。对一个要求准确报告平均水平的仪器来说，细胞间差异可能增加估计困难；对研究群体组成的项目来说，同样的差异正是对象。不能由某研究把变异纳入误差模型，推断它在本体论上否认变异存在。真正值得批判的是，在变异持

续决定干预效果时仍不检查其结构。发生学重读应当把这种条件性的忽略指认出来，而不把整门分子生物学描述成只认可均值。

因此，变异研究需要保留分层的解释秩序。先判断测量技术与样品处理是否造成差异，再分析个体状态和群体组成，随后比较随机动力学、隐藏条件与反馈机制。这个秩序不是要求所有项目都完成无穷排除，而是要求最接近数据的已知解释先获得合理机会。若新增的历史变量确实提供独立预测增益，便可以继续追查它的承载机制；若增益消失，框架也应接受在当前尺度上收缩。

17.3 记忆的承载与主动维持

细胞记忆可以首先被操作地界定为：在当前外部条件相近时，先前经历对后续响应仍具有可重复的影响。这个界定允许多种实现方式，并不要求所有记忆都经历染色质修改。持续存在的分子、反馈维持的表达状态、空间组织或群体组成都可能形成历史依赖，但它们的持续时间和干预敏感性不同。只有明确历史依赖由什么保存，记忆才从一个形容词成为能够比较的机制概念。

Acar 等对酵母半乳糖调控的研究表明，不同反馈环对稳定表达状态和先前条件的保持具有不同作用；在其研究系统中，并不是所有正反馈环都等价地生成记忆。[30] 这使发生分析获得一个重要约束：有回路不等于有某种特定的历史承接。回路位置、作用方式和随机转变都需要进入解释。稳定可以是持续调控的产物，不能因为输出保持不变便断言系统不再发生。

为说明承载方式的区别，可以构造两个具有相似后效的模型。模型甲以过去刺激留下的分子存量作为状态，撤去刺激后仅有衰减；模型乙在刺激后维持持续补充该状态的调控。短窗口内两者都可呈现平坦输出，延长等待后才可能分开。若甲中的存量按固定比例减少，经历相同累计衰减后就应接近相同响应；若乙包含维持阈值，短暂扰动可能被补回，越过阈值的扰动却可能使状态切换。比较的关键因此是扰动后的去向，而非后效持续了多久。上述预言属于构造模型，具体细胞是否满足它们须由相应测量判定。

历史承接也不等于完整保存。一次经历可能只留下对某类输入敏感性的改变，而没有保留其全部细节。这样的选择性压缩使后续响应获得方向，也使系统能够忽略无关扰动。若把记忆理解成过去的逐项复制，就难以解释为什么不同经历可以导向相近状态。SDE 所强调的稳定性，恰好可以研究哪些差异被沉淀为条件，哪些差异未能进入下一轮，以及这种筛选怎样改变未来可出现的显露。

对承接的评价还需要区分功能与适应价值。某历史状态能够保持，只说明它具有持续机制；是否有利于当前环境，是另外的问题。在构造的交替环境中，长期保持可能减少反复调整的成本，也可能使系统对突然变化反应迟缓。没有一种记忆长度在所有任务下都占优。研究应当比较保持、切换和重新建立状态的代价，而不是将更长记忆自动理解为更成熟的主体性。

细胞记忆与人的记忆由此形成有限的比较。两者都可以被追问过去如何影响下一步，但人的记忆还涉及叙述、理由、身份与意识等问题。细胞实验能够提供历史承接的物质模型，不能直接证明人的体验由同一种调控结构充分解释。发生框架的统一性在于允许提出相似的承接问题，其学术纪律则在于要求每个尺度重新建立证据桥梁。概念相似不能代替机制同一。

17.4 快慢过程怎样共同构成响应

细胞内部包含多种时间尺度。把它们概括为比观察更快，无法解释为何研究者有时可以追踪反馈，有时只能得到平均效应。一个过程的快慢须相对于另一个过程、测量间隔和研究问题说明。快速过程可能成为慢过程模型中的有效参数，慢过程也可能决定快速响应的边界。只要这种近似在明确范围内有效，时间尺度的区分就是知识，而不是对生命整体性的背叛。

假设一个细胞对短刺激迅速产生响应，同时某种较慢的调控改变随后刺激的敏感性。若实验只观察刺激后很短的时段，它主要识别即时响应；若在较长间隔后重复刺激，便可能识别历史依赖。若观察间隔过密导致测量本身干扰细胞，获得更多时间点也未必改善机制识别。时间分辨率与观察负担应被共同设计，不能把更快拍摄当作唯一的认识进步。

快过程成为慢状态的条件，可以用一个构造比较说清楚。两段刺激具有相同总量，一段密集到达，另一段分散到达。若中间状态在刺激间隔内基本不流失，总量可能足以预测最终响应；若中间状态持续衰减，两段安排就可能积累出不同末态。此时需要保留的是到达时间与流失速度的相对关系。增加全部历史未必必要，记录一个经过检验的累积状态便可能足够。若更换间隔后该状态失去预测力，才应重新打开更细过程，检查是否存在饱和、阈值或另外的承载环节。

跨尺度模型必须说明哪些过程被显式保留，哪些被平均化。若快速变化在慢窗口内呈现稳定统计关系，可以先用有效关系处理；当干预改变了快速过程的分布，原来的有效关系可能失效，此时才需要打开更细尺度。反过来，也可能存在对微观细节不敏感的总体规

律。发生研究应当解释这种稳健性怎样形成，而不要求每一种宏观预测都回到全部分子事件。

历史窗口的选择也需要可修订。开始时研究者可能只保留前一次刺激，随后发现更早的序列仍能改善预测。增加历史长度应该依据独立资料中的增益，而非模型拟合能够持续变好这一事实。过长的历史表示可能把批次、时间漂移和实验者操作习惯一起记住，造成虚假的记忆机制。只有当旧历史对新干预提出可检验的差异，研究者才有理由认为它捕捉到了状态的承接。

快慢分析最终不是为了在物理和生命之间建立一道绝对边界。反馈系统在不同领域都可能表现出尺度分离，也都可能需要联合处理。生命研究的特殊困难在于，研究对象的维持、繁殖与环境交换常常使事件边界本身成为问题。SDE 可以帮助追踪这些边界如何被维持和重建，却不能由这种困难推断所有现有分步实验都失效。有效的分解与必要的联合，应在同一个研究程序中互相修正。

17.5 空间组织与细胞身份

细胞响应并非只取决于分子种类和总量，空间安排也可能改变能够建立的接触和信号关系。Warmflash 等在体外微图案化培养中，通过控制细胞群落几何并给予特定信号，研究了有序空间分化图式。[31] 该工作提供的是一种受控模型系统，不能等同于完整胚胎发育，却说明几何约束能够成为可操纵的发生条件，而非只能作为不可处理的背景存在。

这一事实对 E 的理解有直接启发，但不能把 E 简单等同于空间环境。相同几何可能在不同细胞状态下产生不同结果，相同状态也可能因接触关系改变而进入不同序列。特征纠缠指的是位置、状态、信号与相互作用如何共同规定可出现的图式。几何只提供其中一个可控入口。若研究者把全部空间变量称为 E，再把所有细胞输出称为 S，仍然需要说明两者之间实际通过哪些差异过程连接。

假设研究者在两种排列中得到相似的群体平均表达，却观察到不同的空间分区。平均值相同并未使两个系统等同，因为后续刺激可能首先作用于不同边界位置，接着经不同邻接关系传播。此时状态的充分描述需要包含一定的空间结构。反过来，如果重新排列并不影响研究所关心的输出，空间细节在该用途下可以被压缩。保留结构与舍弃结构都应由任务预测来检验。

细胞身份也因此有多种判据。按某些标志物分类，可以得到便于沟通的类别；按响应潜力分类，可能得到另一种划分；按谱系关系分类，又回答不同问题。类别之间不一致

时，不必立即宣布其中一种虚假。应当检查各自的目的：标志物类别适合描述当前分布，响应类别适合预测干预，谱系类别适合研究承接。发生理论所要求的是解释这些分类之间的联系及其不能互换的条件。

还可以让三种身份判据在同一构造案例中接受检验。某群体的标志物分布回到处理前范围，原有谱系却减少，而剩余谱系面对新刺激的响应发生改变。按照当前标志物，系统可以被归回旧类别；按照谱系，发生了成员更替；按照响应潜力，则出现了新的功能划分。三项记录各自正确，却对应不同承接关系。只有保留它们之间的连接，才能追问是旧成员恢复了状态，还是新组成重新实现了旧外观。若类别变化仅来自分析阈值修订，证据只支持分类历史改变；对象转变必须另有记录。

空间研究由此扩展了细胞的发生单位。某些问题能够以单细胞为单位，另一些问题需要细胞及其邻接关系，还有一些问题需要培养边界与外部供给一起描述。单位扩展不是越大越好，而是要足以解释被研究的稳定性。若新增边界信息改善了迁移预测，便说明旧单位遗漏了相关条件；若没有增益，则可以保留更简洁的描述。对象的形成与模型的节制，在这里是同一项科学责任。

17.6 观察怎样进入细胞过程

活细胞观测的一个直接问题是，获取信号的操作可能影响被观察系统。Kiepas 等对荧光成像的原始方法研究指出，样品被照明而相机未实际采集信号的照明开销，也可能增加光毒性与漂白。[32] 这一结果并不意味所有荧光记录都不可用，而是说明观察剂量、采集时序和系统状态需要被共同记录。测量的质量不能只由图像是否清晰决定。

这里能够清楚地区分两种回写。第一种是观察动作通过物理过程改变细胞，因而影响后续状态。第二种是研究者根据图像改变采集计划，例如决定下一时段拍得更密。后一种回写由实验控制程序实现，可能进一步改变前一种作用。若不保存控制决策，最后得到的时间序列就会混合细胞自身变化和观察程序的响应。读数事件档案需要同时记录采集事实与采集规则的版本。

假设自动系统发现某细胞形态变化后增加照明频率，而这些细胞后来表现出更多异常。观察到的关联既可能来自早期变化确实预示后续异常，也可能被增量照明加强。简单地比较高频组与低频组无法区分两者，因为高频组本来就被选择了。要识别观察的作用，可以在适当条件下对采集策略进行受控比较，或使用不依赖当前异常的安排；具体操作仍须符合样品和设备的限制。

破坏性测量提出另一种困难。如果每次测量消耗不同细胞，时间序列实际来自群体切片，而不是同一细胞的生命史。将这些切片排列成一条轨迹，需要假设样本之间具有可比性，并说明排序依据。这样的重建可以产生有用假说，却不能直接冒充逐细胞观察到的转变。发生研究应保留这种证据身份，使后续谱系资料能够检验重建，而不是让重建结果反过来定义所有可能的轨迹。

单细胞追踪也并非天然没有选择偏差。能够持续成像、能够稳定分割和能够保持标记的细胞，可能只是样本中的一部分。丢失的轨迹如果与细胞状态有关，剩余资料便会改变对稳定性和转变率的判断。研究者需要记录为何退出追踪，并比较不同处理方式对结论的影响。把不可观察的部分承认为资料边界，比将幸存轨迹视为全体历史更符合发生单位的要求。

观察参与发生之后，客观性并未消失，而是转化为可检查的条件透明。其他团队能够知道采集规则、材料处理和分析阈值，便有可能重建差异的来源。绝对不介入通常不是可实现的目标，控制和说明介入才是实际工作。细胞研究的科学性由此建立在双重可追溯性上：既追踪细胞的状态历史，也追踪研究者怎样取得和处理这段历史。两条历史相互关联，但仍然承担不同的解释责任。

17.7 让历史假说接受干预

要检验历史调控，首先需要让不同历史在当前比较点尽可能可比。构造一个两阶段方案：第一阶段给予不同的刺激序列，第二阶段在共同条件下施加相同探测刺激。若第二阶段响应不同，便得到历史依赖的初步证据。下一步不是马上归因于某条反馈环，而是检查第一阶段是否改变了群体组成、当前残留物或基本生理状态。不同解释对等待时间和后续操纵具有不同的敏感性。

对回路进行干预时，也要区分形成与维持。某调控因素可能只在历史状态形成时必要，形成之后由其他过程维持；也可能持续参与维持，却不决定最初选择。若只在整个实验期间统一改变它，就难以分辨两个角色。将干预安排在不同阶段，可以更细致地定位历史怎样进入未来。SDE 的三原理在这里应落实为具体的阶段性预测，而不是把所有差异都解释成三维同时变化。

恢复实验还能够加强解释。如果改变候选承载机制之后历史差异减弱，而恢复相应条件能够恢复部分差异，便为该机制提供更有力的支持。但恢复程度不足也有多种可能：干预并不充分，其他机制共同承载，或者先前的解释本来错误。发生模型应预先说明哪些结

果支持必要性、哪些支持充分性、哪些只提供相关线索。单次成功不能替代对这些逻辑关系的区分。

在这些阶段性操作之后，模型比较才有明确对象。存量模型应预测延长间隔怎样削弱历史效应，群体选择模型应预测效应怎样随成员比例变化，持续调控模型则应预测形成期与维持期干预为何不同。让三个模型只竞争同一条终点曲线，往往保留过大的解释自由；让它们共同预测间隔、比例与干预时段三个改变，才可能暴露各自缺失的承载条件。比较应使用同样可获得的状态信息；某模型若额外知道谱系，首先应计入谱系信息的贡献，再判断其更新规则是否提供新的收益。

预测检验应当越过模型形成时的条件。对于声称解释历史保持的模型，新的刺激顺序比相同顺序的重复更有辨识力；对于声称解释空间组织的模型，新的几何条件比同一图案的更多图像更能检验外推。新的条件仍需处于合理的实验范围，不能故意选择任何模型都无法处理的极端而宣称旧知识失效。有效的压力测试应当针对模型已经作出的明确承诺。

研究结果还应回写状态定义。若某组历史变量在多种条件下没有增益，可以把它们从常规档案中移出，保留为特殊情境的补充；若某个旧类别内部出现稳定的响应分化，则需要修订分类。发生理论的建设性在于支持这种有理由的增删。它不要求档案无限增厚，而要求模型能够把失败转化为具体修订，并使修订后的对象在下一轮研究中接受检验。

17.8 从细胞历史到组织能力

细胞状态的历史性并不使每个细胞都不可比较。相反，某些不同历史可以汇入相近的响应结构，某些共同历史也可以分化为多个状态。研究需要找到能够说明这种汇聚与分化的条件。稳定类别因此可以保留，但其意义从永远固定的本质转向可重复建立的发生结构。一个类别若能预测新条件下的响应，就具有科学价值；若只能整理现有数据，它仍有描述价值，但不能承担更强的因果任务。

组织能力也不能由单细胞能力直接相加。假设每个细胞单独都能对刺激恢复，而群体中相互竞争的供给使总体无法恢复；或者单个细胞的响应差异很大，群体却通过分工保持稳定输出。这些可能性要求在细胞与群体之间建立新的测量映射。E 在组织尺度上将包含新的特征依赖，其内容不能由单细胞尺度的变量简单复制。尺度改变意味着研究对象的重新限定。

这也说明涌现不应被当作无法解释的剩余。若组织层面的稳定需要特定空间安排、交换关系与更新规则，那么这些条件就是可研究的机制入口。研究可以先建立局部模型，再

检验它们怎样组合以及在哪些条件下组合失效。SDE 并不要求从整体直接跳到整体，而是要求局部知识在组合时保留其成立条件。所谓整体性，只有能改变实验问题和预测安排时，才具有超出口号的学术内容。

承接关系还需要跨分裂说明。假设某状态在母细胞中保持，而两个子细胞的响应不同。研究可以追问承载材料是否分配不均，分裂后是否受到不同输入，或者旧状态是否需要重新建立。母细胞的一项终点指标不足以决定这三个解释；谱系连接需要同时保存分裂时点、承接对象和子细胞后续条件。由此可见，延续的单位有时是一个细胞，有时是一条谱系中的状态传递，二者不必重合。最低功能意义上的主体号位关注这种选择性承接及其维持后果，并不据此赋予细胞现象意识或道德地位。

同样，稳定状态与疾病之间没有单一对应。某些稳定性可能是正常功能的条件，某些持续失配也可能稳定存在。仅凭表达分布集中或波动加大，不能作出临床判断。发生研究在基础层面可以考察响应范围、恢复路径和历史代价，但这些结果若要进入医学，还需要独立的病理、生理和临床证据。概念上的病理学在本书中指向模型与条件的持续错配，不承担未经验证的诊断功能。

细胞历史所形成的正面结果，是把过去的作用定位到仍可改变的物质条件：存量怎样衰减，调控怎样补充，成员怎样更替，空间关系怎样继续限制响应。一个足以预测新刺激的状态，可以省去大量过去；当这种压缩失败时，谱系和阶段记录帮助找到需要重新打开的承接。神经学习在此基础上增加了一个问题：任务改变以后，已有活动组织怎样被重新用于区别和行动。第 18 章由此转向感受野与群体活动，追踪物质状态进入任务能力的具体方式。

第 18 章 神经活动 感受野与学习

18.1 从响应位置进入功能问题

一个神经元在某种视觉刺激出现时增加放电，研究者据此描绘它的感受野。这个描述已经包含刺激方式、动物状态、记录位置和响应判据。感受野不是从神经元内部取出的地图，而是在这些条件下建立的响应关系。它可以具有高度重复性，也可能在任务改变后需要重新估计。发生分析要解释的正是这种条件稳定性：一个响应关系怎样被建立，哪些变化会使它继续有效，哪些变化会迫使研究者重新定义它。

神经活动的功能不能仅凭一张响应图确定。某神经元对刺激敏感，不等于它在当前任务中决定了行为；某活动能预测选择，也不等于它直接生成选择。预测、相关和因果贡献需要不同证据。SDE 在神经研究中的入口因此不应是给每种活动指定一种哲学属性，而是把活动显露、刺激与行动的差异序列、以及限定可用关系的特征组织放在同一分析单位内。

这个单位仍需受任务限制。研究视觉辨别时，身体运动可能主要作为需要控制或记录的条件；研究自然行动时，运动本身则改变后续输入，成为核心机制。若在前一种任务中获得稳定感受野，不能直接断言自然行为中的神经功能同样固定；若在后一种任务中看到复杂变化，也不能据此否定受控实验的知识价值。两种研究保存了不同的关系，重要的是建立它们之间的迁移条件。

假设同一视觉图形在任务甲中是目标，在任务乙中是干扰物。某神经记录随任务而变，可以说明刺激到活动的关系受到任务条件调节，但还不能说明动物如何理解目标。需要比较行为表现、任务规则学习和具体干预。神经活动在这里提供了一种显露，而不是对主体状态的完整替代。主体号位必须通过系统如何区分目标、承接反馈及改变行动来研究，不能被压缩为一个脑区或单一指标。

任务变化使号位侧重的区分取得了更具体的内容。同一神经元身份的持续辨认，使两次记录能够比较；刺激、动作与反馈的安排，使其活动获得不同的任务关联；动物或参与者对规则的学习，则决定哪些信息可以用于选择。三者各自都有形成和更新的过程。若只把任务名称加为一个标签，就仍未说明哪一步改变了响应。研究应指出新规则改变的是目标位置、奖励条件、可用动作还是既有策略，并检查这种改变怎样进入记录与行为之间的关系。

神经科学由此成为对发生理论的严格检验。它提供丰富的时间记录，同时也容易诱使研究者把读数密度当作解释深度。更多神经元、更多采样点和更复杂模型，并不自动回答功能如何形成。真正的推进在于，记录能够区分哪些竞争机制，干预能够改变哪些后果，以及模型能否预测新的任务安排。发生单位只有在这些要求下才从宏观哲学语言进入神经研究。

18.2 感受野的稳定与可变

Womelsdorf 等在猕猴视觉皮层 MT 区域的研究中，观察到空间注意相关的感受野位移，并伴有较小程度的收缩。[33] 这一结果支持在其任务条件中把注意安排纳入感受野描述。它并不意味着所有感受野都随时任意改变，也不意味着神经元缺乏可识别的结构。相反，只有先建立足够可靠的记录和比较，研究者才有可能把有方向的变化与估计误差区分开来。

发生分析从这里提出的第一个问题，是保持了什么才使变化能够被称为同一感受野的改变。记录的神经元身份、刺激空间和响应判据需要足够可比，否则两次图形的差异可能只是对象或测量方式改变。第二个问题是变化怎样影响任务中的可用信息。某位置的响应增强，不一定提高全部辨别；对目标有利的调整也可能降低对其他位置的敏感性。功能判断必须明确其服务的任务。

一个简化模型可以帮助区分位置变化与响应放大。假定感受野曲线的形状保持，只把所有位置的响应乘以同一正数；在完整采样且估计方法允许幅度独立变化时，曲线的峰位应当保持。若以是否越过固定阈值决定响应区域，或者只采集曲线的一侧，同样的放大却可能改变被纳入估计的位置。于是表观位移既需要与纯幅度模型比较，也需要检查覆盖范围和阈值。只有这些条件得到控制，剩余的位置变化才更有力地要求空间重组解释。这是对估计逻辑的构造推演，并非对所引实验结果的替代解释。

感受野的估计尤其依赖刺激集合。若只在局部区域呈递刺激，模型对未采样区域的形状可能主要来自预设；若任务条件改变了有效采样范围，两次估计便不完全可比。发生档案应保存刺激实际出现的位置、时间和任务关联，而不是只保存拟合后的中心与宽度。这样，后来的研究者能够判断变化发生在数据、拟合规则还是两者之间的关系上。

稳定感受野同样能够支持发生论证。假如某响应关系在不同注意安排下保持不变，这可能说明该层处理具有较强的不变性，也可能说明所用测量不足以捕捉任务相关变化。两种可能不能凭理论偏好裁定，需要通过测量敏感性和任务操纵来判断。若证据支持稳定，

就应当把稳定作为结果，继续研究系统怎样在其他层面完成任务调整。发生框架必须容纳分工，而不能要求每个局部都承担全部可塑性。

由此可以区分参数变化与对象失效。感受野参数随条件改变，通常意味着描述需要包含条件，而非神经元这个研究对象已经瓦解。只有当原有单位无法组织关键依赖，例如单神经元模型持续遗漏群体协调，才需要扩展单位。对象的重新建构应针对实际解释困难，不应由变化这一事实直接推导。对稳定性的精确说明，比宣布一切都流动更能推动新的神经实验。

18.3 群体活动与学习的可达范围

学习要求系统改变可产生的活动与行动关系，但这种改变并非完全自由。Sadtler 等利用脑机接口任务研究了既有神经群体活动结构对学习的约束，发现与既有结构关系不同的任务变化，在所考察学习窗口中具有不同的可学习性。[34] 这里的约束是具体任务和时间范围中的结果，不能被解释为个体具有永远不可改变的能力上限。

这一研究问题适合 SDE 的一个核心判断：新的显露需要从已有的特征组织中发生。若群体活动已经形成某些协同关系，新的任务可以重用这些关系，也可能要求建立新的组织。两种改变在训练量、时间和反馈需求上可以不同。发生理论不必预先决定哪些改变更容易，而应把既有组织与新任务之间的相容程度转化为可比较的研究条件。

假设一个控制任务原先要求几种活动组合对应不同移动方向。新任务甲只是重新分配这些组合的行动意义，新任务乙则要求产生过去很少出现的组合。即使两种任务的目标距离相同，它们对系统提出的改变也不同。若只用总体成功率描述难度，就可能遗漏这个差别。分析单位需要包含活动结构、映射规则和反馈历史，才能说明行为改善究竟来自何处。

两种任务若获得相近成功率，仍可以通过后续探测区分学习内容。沿用上述构造案例，暂时恢复旧映射：若原有活动组合仍能迅速支持旧行动，改善可能主要来自新旧映射之间的选择；若旧映射也受持续影响，则需要考察活动组织是否改变。再安排未训练过的中间映射，可以检验系统是否只记住两个固定解法。恢复、保持和新映射迁移分别约束不同层次，任何一个成功率都无法独自概括它们。学习的对象因此是活动组合与任务规则之间可重新建立的关系，而不只是分数的提高。

E 在这种分析中不是一个静态的神经网络图。它包含任务相关活动之间的约束，以及这些约束与反馈、身体和工具之间的关系。连接结构可能限制活动，活动历史也可能改变

可用连接的功能表达。仅凭一张相关矩阵不能决定这些关系的方向。研究需要把观测到的协同、模型中假定的约束和干预支持的机制分别标明，再讨论哪种组织真正承担学习的限制。

学习的困难也可以生成创造性问题。若某种新行为长期不能获得，研究者不必只增加重复次数，还可以改变反馈的形式、任务分解方式或初始映射，考察是否存在能够进入新组织的路径。这里对应的六路径是研究进入次序：从现有活动结构出发安排差异，或者从失败的行动显露倒推需要改变的条件。路径是否有价值，取决于它能否产生新的可检验进展，而不是其名称是否完整。

18.4 行动反馈与表征的共同形成

在主动任务中，动作改变下一刻能够获得的刺激。眼睛转向某处，身体接近某物，控制器移动到某位置，都使输入依赖刚刚发生的输出。这种闭环区别于把刺激作为完全外部序列的安排，但两种安排可以被有目的地比较。若同样的刺激记录在主动与回放条件下产生不同响应，需要进一步分析行动预测、控制感、注意和身体状态等候选机制，而不能将差异一次性归给主体性。

构造一个低风险的视觉控制例子。参与者通过手柄移动光点，在条件甲中光点运动由其动作决定，在条件乙中观看先前记录的相同轨迹。两种条件在视觉输入上可以相近，但动作与后果的关系不同。若研究目标是行动回写，就应记录动作、反馈延迟和误差处理，而不只比较平均脑活动。这个例子没有声称任何结果，其作用是说明怎样把关系差异转为实际可比较的任务。

反馈还具有解释内容。同样的失败提示可能告诉参与者方向错了，也可能只告诉其得分不足。前者提供具体修改线索，后者可能促使其探索其他策略。神经活动变化因此需要与可获得信息一起解释。若不记录反馈内容，研究者可能把任务信息差异误判成个体能力差异。发生档案应保存提示的版本和可见时间，使学习过程能够被重建，而非只留下训练前后的两次分数。

反馈的时间还影响后果被归给哪一次行动。设想参与者连续作出两次选择，评分提示在第二次选择之后才出现。如果提示未标明所对应的动作，规则学习的困难可能来自归属含混，而非缺乏产生正确动作的能力。可以在保持奖励数量不变时增加事件对应标记，观察策略是否随之改变；也可以固定标记而调整反馈间隔，研究时间承接的作用。行动显露

与既有区分之间的张力，由此推动任务组织的修订；修订再改变下一轮可获得的差异。这使三原理落在后果归属的具体机制上。

成功的反馈也可能稳定原有组织。当某策略持续满足任务，系统可能减少探索并保持一致表现。这种收敛可以是学习的成就。若任务后来改变，原有稳定策略才可能成为限制。稳定与灵活的评价必须相对于条件和目标进行，不能以行为变异大小直接判断系统是否有生命力。一个在熟练任务中低变异、在新任务中能够重组的系统，与一个在所有任务中都随机波动的系统，具有完全不同的能力。

主动任务同时把研究者纳入闭环。实验者可能根据参与者表现调整难度，算法可能改变奖励频率，观察者可能给予额外提示。若这些调整没有记录，所谓学习轨迹就是对象变化与任务变化的混合。研究者不必消除所有自适应安排，但必须让它们成为模型可见的条件。研究单位因而包含一个共同演化的任务制度，而不只是接受固定刺激的孤立大脑。

18.5 同步 差异与因果的尺度

神经信号的同步可以由多种机制产生，异步也可以由多种机制产生。共同输入、相互耦合、任务节律和记录处理都可能影响时间关系。因而不能把三种指标不同步直接当作三号位同台的证据，也不能把同步当作某一号位压制其他号位的标志。时间结构的科学意义，需要通过其与任务、干预和模型的关系来确定。

假设两组神经活动在某频段相干增强，同时行为改善。至少可以提出三种解释：协调本身有助于任务，共同的任务状态同时增强协调与表现，或者分析方法在活动强度改变时产生相干估计的变化。仅有相关结果不足以选择其中一种。研究应当检查统计估计的条件，再寻找能够改变协调关系而尽量保留其他条件的干预。即使这样的干预可行，其作用范围仍需谨慎限定。

时间先后也不能直接等同于功能层级。某区域先出现可测响应，可能因为其信号更容易被当前技术检测；另一区域的早期变化也可能未进入记录。不同测量方式具有不同的延迟和时间平滑。若把未经校正的时间差当成信息传递顺序，便会把仪器性质写进神经机制。发生分析要求读数事件包含观察映射，使时间证据不被过度解释。

另一方面，时间关系可以帮助构造更强假说。如果某种干预只在学习前期有效，而在熟练阶段作用减弱，就可以进一步研究它是否参与组织形成而非维持。如果改变反馈延迟产生可重复的策略变化，则延迟可能是差异序列中的重要变量。这里的结论来自有方向的条件比较，而不是来自对波形的哲学命名。SDE 应帮助提出这些比较，不能替代它们。

多尺度时间模型还需要处理同时性。细胞膜过程、群体活动、一次动作和长期训练不必共享同一离散步长。把它们强行排列在同一轮中，可能制造看似闭合但实际上跨越多个机制的回路。研究可以采用嵌套窗口，分别定义快速状态、任务状态和学习状态，并说明它们如何交换信息。所谓同一发生，不要求所有变化发生在同一个时间格子，而要求它们在所解释事件中具有可说明的依赖。

时间窗口之外，任务坐标也决定哪些关系可以保持。假设同一目标在屏幕上移动，而参与者的任务始终是判断目标左侧的图形。按屏幕位置汇总与按目标相对位置汇总，可能显示不同的稳定关系。研究应预先说明其表征主张对应哪一种坐标，并检验更换坐标是否只是改进统计对齐，还是确实对应行为所利用的信息。如果行为依赖目标相对关系，而模型只记住屏幕位置，原任务内的准确预测就可能在目标移动后失效。任务意义在这里转化为可检验的坐标选择。

18.6 主体号位与意识问题的距离

神经研究中的主体号位首先可以围绕目标区分、后果承接和行动调整来操作化。系统是否把某刺激作为任务目标，是否利用反馈修订策略，是否在新条件中保持学习所得，都能够形成可比较问题。这种操作化仍然没有解决意识是什么。一个系统表现出记忆或自我调整，并不足以证明它具有某种主观体验；相反，无法从现有记录读出体验，也不意味着体验不存在。

人的自我报告为意识研究提供重要材料，却不是所有神经任务都具备的材料。动物任务中的选择、延迟保持和行为灵活性，可以支持功能推断，但其解释需要明确边界。若用人的语言直接描述动物或细胞的内心，便可能把研究者的叙述投射成对象事实。SDE 的主体号位应允许不同层次的证据，而不能利用同一个词在功能、体验和道德地位之间自由滑动。

假设一个受试者能够准确区分刺激，却报告没有明确看见。这个情形提出的是表现与报告之间的关系问题，而不是自动证明某个脑区拥有或缺少主体性。研究需要检查任务要求、报告标准、置信程度以及重复条件。主体的表达同样发生在一种判断制度中，既不能被神经记录完全替代，也不能不经分析便作为对机制的最终裁决。

真、善、美若被用于解释主体活动，只能承担哲学上的对应工作。某种神经分布不能直接被测为美，某种相位协调也不能直接证明善。经验研究可以考察人怎样作审美判断、怎样评价行为，并研究相关神经过程；但价值判断的正当性仍需要理由。把物理或神经术

语当作伦理词的单位，会掩盖两个领域各自的证明责任，也使本来可以开展的具体研究失去清楚目标。

主体号位的建设性作用，是防止任务意义在模型中消失。两个系统可以输出相同动作，却通过不同规则、不同训练和不同后果关系取得结果。若研究只保留动作标签，就无法说明能力的形成。把主体号位纳入研究，意味着保留系统能够获得什么信息、怎样选择目标和怎样承担后效的证据。这样的工作可以在不先解决全部意识哲学的情况下进行，同时为更深问题提供有约束的材料。

这也为跨人机比较建立了边界。神经学习与人工系统都可能表现出状态更新和任务适应，但其实现方式、训练范围和后果承接不同。不能由共同使用反馈一词，就推断两者具有相同主体性。有效比较应选取明确任务，分别描述记忆、目标更新和环境耦合，再检验哪些功能关系可以迁移。第七编将继续这一问题，此处只确立一个原则：功能相似性是一项研究结果，不是词语相同所保证的事实。

18.7 一个关于表征更新的研究方案

可以把本章的问题组织成一个尚未实施的研究方案：考察任务规则改变之后，感受野估计、群体活动组织和行为策略怎样共同更新。方案首先需要固定所研究的物种、脑区和任务类型，再确定技术条件；不能从抽象框架直接推出统一的实验参数。研究目标也应有限，例如区分行为改善主要依赖已有活动的重新利用，还是需要出现新的活动组织，而不是一次实验解释全部学习。

资料应保留训练前、规则改变期和稳定表现期的连续关系。仅比较前后两个平均状态，会遗漏中间的探索与过渡；只记录过渡期，又难以判断最终结构是否保持。各阶段的刺激、反馈和动作需要同步记录，并标明任何任务难度调整。对于神经记录，需保存信号质量与单元识别的变化，以免把记录对象变化误认成可塑性。不同阶段的可比性是后续机制分析的基础。

干预安排可以围绕差异预测展开。若模型认为反馈只改变动作选择而不改变感受表征，那么在保持视觉输入相近时改变反馈规则，应主要影响行为映射；若模型认为反馈参与表征更新，则可能对后续无反馈的感知任务提出不同预测。此处的预测方向需要具体模型决定，SDE 不能提前替所有模型填写答案。框架的工作是要求后效能够被独立观察，而不是只解释训练中的成功。

在该方案中，可以把三个候选变化分别留下探测条件。重新使用旧活动的模型，应预测映射恢复时的表现；表征改变的模型，应预测新映射撤去以后对相同刺激的响应；依赖反馈规则的策略模型，应预测没有反馈时以及反馈延迟改变时的选择。三类模型都应获得任务状态和可用历史，避免把信息缺失制造成结构劣势。真正值得比较的是它们对这些探测条件的联合限制：一个模型若靠分别调整每个条件的自由参数才拟合成功，就尚未说明同一学习机制如何贯穿任务变化。

验证应包含新规则或新刺激的有限迁移。若模型能解释训练记录却不能预测相邻任务，可能是学到了实验特有的统计关系。若在新条件中仍有效，则可以更有信心地讨论其所捕捉的组织。迁移失败也应被定位：是感受层映射改变、行动策略改变，还是任务意义改变。将失败拆开，能够为下一轮研究选择更合适的入口，而不是把所有失败归于系统太复杂。

研究的判定还需要处理不一致的结果。若所记录的神经响应在相应灵敏度与任务范围内保持，而行动映射改变，结果与已有表征重新使用的解释相容，仍需检查未测变化和替代策略；若响应变化但行为迁移没有改善，应检查这种变化是否承担了所声称的功能；若行为和神经记录都改变，而其先后在现有分辨率下不可分，则只能支持联合更新的描述。将这些结果预先列为不同出口，能够让研究在局部层次形成明确知识。后续的干预据此选择对象：对映射变化检查反馈归属，对表征变化检查持续性，对联合变化补充能够分离环节的任务条件。

18.8 学习作为稳定性的再生产

学习并非只有从不稳定走向稳定一种方向。它可以建立新的稳定行为，也可以使系统学会在条件改变时选择不同策略。一个稳定的总分可能掩盖策略重组，一个波动的分数也可能只是随机误差。研究应当把表现、策略和可迁移能力分开追踪。发生论的贡献，是将学习理解为稳定性条件的更新：系统不仅产生新输出，还可能改变下一次输出能够如何产生。

这种理解也保留了遗忘和退出的功能位置。某种旧规则在新任务中不再适用，减少其影响可能有助于形成新表现；但遗忘也可能损失仍有用的知识。不能把保持或改变任一方向设为普遍价值。基础研究需要比较不同历史对当前任务的具体后果，教育和伦理则还需说明哪些目标值得追求。神经机制能够限制可行性，却不能独自决定培养什么能力。

对个体差异的解释同样需要历史视角。相同训练后表现不同，可能来自初始组织、可获得反馈、采用策略和练习过程的差异。若研究只把最终差异归为固定能力，便可能遗漏可改变条件；若把全部差异归为训练制度，又可能忽略实际的生物与经验差异。SDE 不预先选择单一归属，而是要求找到能够支持不同解释的资料，并保留当事人的活动与研究安排之间的关系。

神经研究还有一种更深的客观性要求：模型应当使不一致可以被公开定位。不同实验室得到不同感受野或学习结果，可能因为任务、记录、分析或样本条件不同。可共享的事件档案能够帮助重建这些差别。其目的不是让所有实验最终只得到同一个数，而是建立条件之间可说明的转换；若无法转换，则明确指出资料缺口和模型边界。这样的客观性比掩盖条件的表面一致更有力量。

生命系统需要维持功能，学习又改变功能实现的方式，两者之间存在持续张力。过强保持可能限制新任务，过度更新也可能破坏已获得的协调。神经活动的发生因此不能只用可塑性赞歌来描述。研究应当同时追踪改变的收益、维持的成本和恢复的范围。稳定不是缺乏历史，灵活也不是没有约束；二者都是组织在具体条件下取得的能力。

感受野变化、活动组合的重新使用与反馈归属，分别显示任务能力在何处取得稳定。这样的稳定服务特定辨别与行动目标，其评价依据与内部变量的生理维持有所区别。第 19 章转向稳态时，需要进一步考察维持的资源、补偿的代价和扰动后的恢复范围。神经学习留下的问题因此是：系统通过改变实现方式保持功能以后，还剩下多大的调节余地，又会在何种新扰动下暴露原先不可见的限制。

第 19 章 生命稳态 变异与恢复

19.1 稳态是一项需要解释的成就

生命研究面对的一个基本事实形式是，外部条件变化之后，某些内部变量仍能保持在有限范围内。这里的保持并不要求每一时刻完全相同，也不要求所有组成部分静止。若只把稳定看成没有事情发生，就无法理解维持稳定所需的交换、调节与修复。发生分析首先把稳态重新放回问题中心：系统靠哪些差异过程，使什么样的显露在何种扰动范围内维持可重复的关系？

稳态必须有明确的观察对象。某个细胞的输出保持、群体平均保持和组织功能保持并非同一个结论。群体平均不变可能伴随成员更替；某个指标不变可能伴随维持成本增加；功能暂时保持也可能掩盖可恢复范围缩小。因此，稳态描述至少需要说明被维持的变量、观察窗口、允许波动以及系统边界。缺少这些限定，稳定只是对曲线外观的印象。

假设两套细胞培养系统在一段时间内具有相近的平均输出。甲系统通过持续调整保持输出，乙系统依靠较大的初始储备暂时抵消损耗。两者在当前曲线上相似，却可能在延长观察或再次扰动时分开。发生研究的任务是识别维持机制，而非仅凭当前稳定给两者贴上相同标签。稳定性作为属性，包含对后续条件的有限承诺，不只是过去数据的描述。

SDE 的客体侧重在这里表现为可持续辨认的状态或功能，互动侧重表现为扰动与补偿的过程，主体侧重则在最低功能意义上关注系统如何组织对自身维持有后果的差异。三者并不意味着系统拥有一个内在的小管理者。维持可以由局部规则实现，也可以不依赖任何中央判断。发生论需要解释选择性调节如何形成，不能把调节能力重新神秘化为一个无须说明的主体。

稳态也不是单一终点。某些系统可能围绕固定范围波动，某些系统可能沿周期运行，还有些系统在不同条件之间切换。只要这些模式与所研究功能之间具有可靠关系，都可以成为稳定性研究的对象。是否存在数学收敛，要看具体状态和模型；是否具有生物意义，则要看维持了什么。将两者直接等同，会使周期性过程被误判为失败，也会使无功能的恒定被误判为成功。

从这一点出发，本书关于不收敛的讨论必须与稳态研究相互约束。如果反馈常常能够生成稳定，便不能把反馈存在当作不收敛的普遍原因。更有价值的问题是，哪些结构使扰

动得到吸收，哪些条件使调整越过有效范围，以及模型如何区分暂态、稳定波动与持续漂移。发生理论的力量应体现在这些区分上，而不是给所有曲线预先规定相同的哲学含义。

19.2 反馈为什么能够稳定

Yi 等对细菌趋化模型的分析，将鲁棒适应与积分反馈联系起来，并明确讨论了相关结构和假设。[35] 这一理论工作提供了一个重要范例：输出回到某个水平，可以依靠系统内部持续更新的调控，而不是因为外部刺激已经消失。其结论具有模型条件，不能推广为所有生命稳态的唯一机制，但足以说明稳定与回写可以同时成立。

用一个完全构造的离散模型，可以更直接地展示这种关系。令 x 表示相对于目标的偏差，a 表示系统原有变化的保留程度，b 表示调节输入的作用，u 表示本轮调节。假设每轮变化满足下式。这只是用于说明逻辑的数学例子，并未对应某种具体生理变量。

$$x_{t+1} = a\,x_t + b\,u_t$$

若调节规则为 u 等于负 k 乘以 x，下一轮偏差便等于 a 减 bk 之后再乘本轮偏差。因此，在 a 减 bk 的绝对值小于一时，偏差趋向零。这里可以选择 a 为一点二、b 为一、k 为零点九，得到每轮偏差乘以零点三。无反馈时非零偏差逐轮放大，加入上述反馈后则逐轮衰减；反馈在这个例子中使原先不稳定的系统收敛。这个形式结论只依赖所写假设，却足以反驳有回写就必不收敛的普遍命题。

同一模型也说明反馈并非越强越好。保持 a 与 b 不变，将 k 改为二点四，偏差每轮乘以负一点二，便出现交替放大的轨迹。稳定与失稳的区别来自具体增益条件，而不是来自有没有反馈。若再加入延迟、饱和或随机扰动，稳定条件会改变，需要重新分析。不能把上述简单阈值直接移植到细胞或人体，但可以把它作为模型写作的纪律：每个稳定性主张都应说明结构与参数条件。

生物调节中的目标也不必以一个显式数字存放在某处。某些网络关系可以使输出在一定范围内回到相近状态，即使系统没有符号化的目标表示。研究者使用目标值，是为了描述动态关系，不意味着对象内部存在与工程图完全相同的控制器。SDE 的主体侧重同样需要注意这一点：功能上的选择性可以被分析，却不能从拟人化词语直接推出心理主体。

线性例子的边界也能产生新的研究问题。它假定调节输入可以随偏差无限增大；若实际可用输入存在上限，小偏差可能衰减，大偏差却可能超出可恢复范围。例如，假设两个系统在小扰动下具有同一衰减倍数，输入上限不同，在较大扰动下便会分开。因此，估计

局部反馈强度之后，还需检查可用范围及其恢复。显露偏差与调节能力之间的张力，在这里推动对特征组织的修订：模型必须把资源上限纳入条件，再重新预测允许的差异序列。三原理所组织的是这一步有依据的重建。

19.3 稳定输出背后的可塑性

神经系统提供了稳定与可塑性共同存在的研究材料。Turrigiano 等在培养新皮层神经元中研究了活动相关的突触强度调整，显示系统能够依据活动条件改变突触作用。[36] 这一结果支持具体培养条件下的调节机制研究，不应被扩展为对所有神经功能或临床状态的解释。它提示的是，局部性质的改变可以参与总体活动的维持。

对于发生论，稳定输出并不能告诉我们内部是否改变。一个系统可能通过改变组成维持功能，也可能在组成近似不变时改变运行方式。研究如果只记录最终输出，就可能把这两种过程混同。反过来，内部指标改变也不意味着功能已经改变，因为不同部分可以互相补偿。属性的稳定性需要在所研究尺度上定义，不能把每个层次的变化都推成整体失稳。

假设一套网络中的某条作用减弱，另一条作用增强，最终输出仍保持相近。若研究者只分别测量两条作用，可能认为系统发生了重大改变；若只测输出，又可能认为没有变化。完整解释需要回答补偿是否有代价，以及补偿后的系统在新扰动下是否仍能维持。当前表现相同，并不保证未来可达范围相同；这正是恢复实验比静态比较更有信息的原因。

补偿也可能具有选择性。系统可以保留某一功能，同时牺牲另一功能，或者在短时间内维持输出却增加长期成本。不能因为主要指标稳定，就把全部后果称为成功。基础研究需要明确被测功能和未测功能，规范评价则还需说明哪些功能更值得保护。发生分析把这些问题连接起来，但不允许用一个动态模型替代价值取舍的论证。

可塑性因而应被理解为改变关系的能力，而不只是改变某个数值。若系统能够在新条件下重新建立有效协调，它具有某种适应能力；若只是所有指标都变得更易波动，则不能据此判断更灵活。研究应观察变化与任务需求之间的关系，以及改变后是否形成可持续的工作方式。将变化本身当作进步，会使发生理论失去辨别好的调整与无效扰动的能力。

补偿后的余地可以通过一个构造模型来理解。设两条路径共同维持总输出，甲系统两条路径各有剩余容量，乙系统则主要依赖已经接近上限的一条路径。当前总输出可以完全相同，进一步削弱主要路径时，乙系统更容易失去维持能力。于是余地的估计需要第二次、方向不同的扰动，或者对路径容量的独立测量。若第一次补偿同时扩大了备用路径的

容量，结果还可能相反。可塑性的意义应由这种后续能力说明，稳定与改变之间也因此呈现具体的结构关系。

19.4 变异何时成为资源

变异可以为系统提供不同响应的可能性，也可以破坏已经建立的协调。它是否有价值，取决于变异的结构与环境要求是否匹配。一个输出在允许范围内探索，和一个输出无方向地偏离目标，不能仅按方差大小归为同类。发生研究需要说明变异发生在哪些维度、持续多长时间、能否被选择和是否改变后续状态，而不是把波动总量当作生命活力的直接测量。

构造一个群体模型：不同成员采用不同响应方式，环境偶尔改变，有些方式在当前条件下更有效。如果成员差异使群体在条件切换后仍保留可用响应，它可能提高群体层面的持续性。但这并不保证每个成员都获益，也不说明任何更大差异都更好。若差异扩大到失去基本协调，群体功能可能反而降低。对变异的评价须同时说明分析单位与代价承担者。

同样的方差可以来自不同结构。周期性波动、随机扰动、缓慢漂移和两个状态之间的切换，可能具有相近的总体离散程度，却对后续预测提出不同要求。若只用一个变异系数概括它们，便无法判断系统在受到新扰动时怎样响应。时间关系和条件依赖因此不是附加装饰，而是区分机制的必要信息。SDE 对差异序列的强调，在这里落实为对波动组织的研究。

变异还可能属于测量而非对象。信号质量下降、采样频率变化和分析阈值调整，都可以改变记录的离散程度。如果研究者预先相信变异增加意味着恢复能力增强，就容易忽略这些技术来源。反过来，预先相信变异只是噪声，也可能漏掉状态转变。更可靠的程序是先确认读数结构，再比较对象机制；哲学解释应当接受这种证据顺序，而不能凭价值偏好挑选曲线。

探索与利用的区分可以帮助提出具体问题，但不能直接套用为全部生命机制。例如，假设一个系统在熟悉条件中保持低变异，在不确定条件中扩大尝试范围，可能比始终保持同一变异程度更有效。要检验这种可能，必须说明系统如何感知条件改变，以及变异调整怎样影响后续结果。没有这些环节，只看到不确定时波动增加，还无法确定它是功能性探索还是调节失效。

变异的组织可以比其大小更重要。构造两个各含两名成员的系统，每名成员都围绕基线作幅度相同的上下波动。若两名成员总朝相反方向变化，总输出可以保持；若总朝同一

方向变化，总输出就会放大波动。个体方差相同，群体后果却不同，因为成员之间的协调关系不同。若协调被打断，原先有用的变异也可能转成成本。把变异视为资源，因而需要同时找到差异的来源、使差异相互补偿或被选择的关系，以及这种组织对后续功能的收益。

19.5 恢复是一条路径

恢复常被描述为指标回到原值，但原值相同并不保证系统回到原来的组织。一次扰动之后，系统可能通过新的结构达到相近功能，也可能只在外部支持下暂时维持旧水平。研究因此需要区分数值恢复、功能恢复和组织恢复。三者可以重合，也可以分离。若不说明目标，恢复就会成为一个含混的肯定词，掩盖不同后果之间的差异。

假设甲系统受扰动后迅速回到基线，却在下一次扰动时出现更大偏离；乙系统回到基线较慢，但后来能够承受更广的变化。仅按首次恢复速度排序，会忽略后续能力。反过来，乙系统的缓慢恢复也可能造成不可接受的即时成本。评价需要同时记录恢复速度、持续代价和后续可工作范围，不能只选择最符合理论叙事的那一项。

恢复路径可能具有历史依赖。曾经经历某种扰动的系统，面对同类扰动时可能反应不同；但这种差异既可能来自学习或预适应，也可能来自未恢复的残留。区分二者需要检查当前状态和新条件中的表现。若差异只在尚未恢复时出现，不能把它自动称为长期记忆；若在共同条件下持续并可被特定干预改变，才有更充分理由研究其承载机制。

恢复也不必回到旧目标。环境长期改变时，重新建立一个可工作的范围，可能比返回旧值更符合当前要求。这种调整在描述层面可以被研究，但其价值不能自动由系统能够做到推出。对于人类相关研究，还要考虑当事人的目标、负担和选择权。基础生命机制提供可行性信息，伦理判断则决定如何理解和采用这种信息。两者的联系需要明确的理由桥梁。

恢复过程因此宜用成对探测，而不只用一次回到基线来表征。第一次扰动后，可以在不同等待时点施加同一探测，区分恢复进度；也可以在相同等待时点改变探测方向，区分恢复范围。若输出已经回归而对某方向仍异常敏感，就需要保留未恢复的条件；若所有已检验方向都与原先相近，才可以在该范围内采用恢复的判断。这里得到的是带探测范围的功能等价，它允许内部组织已经改变，也为 SDE 追踪新组织怎样限制下一轮响应提供了具体入口。

恢复研究最有价值的资料，往往包括没有恢复的情形。若只分析成功轨迹，就无法知道机制的适用边界。失败可能来自扰动超出范围、资源不足、调节延迟或目标冲突。研究者应保留失败条件和退出记录，使模型能够解释何时不应期待恢复。一个只在成功之后讲通故事的发生模型，不能支持可靠干预，也不能成为科研新典范的示范。

19.6 病理判断与模型失配

本书使用病理一词时，首先指向一种认识与制度问题：模型持续把已经改变的发生条件压回旧描述，并据此产生反复失配。这个概念不能直接替代医学中的疾病判断。一个细胞指标稳定或波动，都不足以单独证明健康或病态；需要结合具体生物机制与相关证据。发生理论如果借用临床语言扩大自己的确定性，反而会违背其要求条件透明的基本主张。

模型失配可以在没有疾病的情况下发生。假设某实验系统改变培养条件后，原来用于预测输出的固定参数不再有效。问题可能只是参数需要随条件更新，也可能是研究单位遗漏了新的相互作用。如果团队仍把每次偏差当作偶然误差，便形成持续失配。改进方法是定位失效条件、比较候选机制并修订适用范围，而不是立即给对象贴上退化的标签。

反过来，模型可以准确预测一个不理想状态。预测准确并不意味着目标值得维持，也不意味着系统具有良好恢复能力。由此需要区分知识的正确性与状态的价值。发生分析能够帮助揭示不理想状态怎样被再生产，但是否应当改变、如何改变以及由谁承担代价，需要额外论证。对于人类相关场景，这些问题尤其不能被一个稳定性指标替代。

所谓僵化也应被具体化。若系统面对条件变化仍只能重复同一响应，并因此持续失败，可以研究其可达范围是否受限；但在明确边界内坚持某种反应，也可能是可靠功能。没有任务条件和后果证据，不能把反应重复直接解释为僵化。类似地，灵活变化只有在能够响应相关差异时才具有功能意义。对稳定的贬低与对变化的赞美，都需要让位于更精确的描述。

制度也会制造对稳定的偏好。若研究评价只奖励低方差和整齐曲线，研究者可能倾向于删除难以归类的轨迹；若评价反过来只奖励新奇变化，也可能放大偶然波动。两种偏好都会损害知识。发生档案应保存预先规定的排除理由、原始分布和不利结果，使稳定与变化都在相同证据标准下接受解释。批判的对象是选择性处理，而不是某一种曲线形状。

持续失配的修订可以分三步落实。先保留原模型曾经有效的输入范围，确定失败从哪种条件开始；再把观测到的偏离与容量、时延、基线和组成等候选变化连接；最后选择能够改变后续判断的最小修订。例如，加入输入上限即可预测大扰动失败时，无须把全部反

馈结构推翻；若新增上限仍解释不了方向不对称，则应检查不同路径。病理判断的知识价值就在于这种定位：它使重复失败转成可比较的机制问题，并保留旧模型已经取得的局部成果。

19.7 稳态与恢复的比较设计

一项关于稳态的研究，应先规定扰动和维持对象。若目标是考察某输出在输入改变后是否恢复，就要记录输入实际变化、输出的暂态和足够长的后续过程。若目标是考察组织是否保持功能，则还需独立的功能指标。不同目标可能需要不同采样频率和观察时长，不能用同一套默认参数处理所有生命过程。设计的依据应当来自候选机制，而不是数据越多越好的直觉。

比较中应包含持续刺激与撤去刺激两类问题。输出在刺激持续时恢复，可以支持某类适应机制；输出只在刺激撤去后恢复，则回答不同问题。若把两种恢复混写，模型可能获得虚假的解释力。还应考虑重复扰动：第一次响应帮助识别即时调节，后续响应帮助检验历史承接。重复本身可能改变系统，所以它既是测量安排，也是需要明确的干预。

观察窗口需要与失败风险一起设计。窗口太短，缓慢恢复会被误判为永久漂移；窗口太长，又可能把无关的培养变化纳入同一事件。研究者可以预先规定多个机制相关窗口，并说明每个窗口回答什么问题。若结果提示原窗口不足，可以扩展研究，但应保留这种修订的时间和理由。后见之明可以推动新设计，不能改写原先假说已经承担的预测责任。

模型比较需要允许不同稳定形式。固定点模型、周期模型、随机稳态模型和具有状态切换的模型，各自适用于不同结构。先检查资料是否足以区分，再讨论哪个机制更合理。对一个短序列拟合复杂状态切换模型，可能只是把噪声分成多个类别；对明显的阶段变化坚持固定均值，则可能遗漏真实结构。发生理论应帮助寻找适当分辨率，而不是预先指定最复杂模型。

评价标准应包括预测与解释的不同收益。一个简单模型可以很好预测平均输出，却不能预测恢复时间；一个机制模型可以解释某项干预，却未必提高全部预测。研究者应根据原先用途报告这些差异，避免把任何一项改善写成全面胜利。模型需要承担的任务越多，证据要求也越高。尤其当研究结果准备跨尺度使用时，必须重新检验原有变量与新任务之间的映射。

基线的选择还会影响恢复的判定。如果基线本身随着实验进程缓慢变化，将所有后续值都与最早时点相比，可能把正常的条件迁移算作无法恢复；如果每次都用最近值重新定

义基线，又可能把持续偏移吸收掉。研究应当同时保存固定参照与局部参照，说明两者各自的用途。参照更新必须有独立依据，不能仅为了让曲线看起来恢复而移动目标。

多次扰动的顺序也可以提供额外辨识。先经历甲扰动再经历乙扰动，与相反顺序产生不同结果时，说明两个过程不能在所考察窗口内简单交换。但顺序效应仍可能来自未洗脱的残留，而非新的调控结构。增加等待条件、改变扰动间隔和记录当前状态，能够帮助区分这些解释。研究的推进由此从存在顺序差异走向差异承载机制，而不是停留在历史重要这一宽泛结论。

最终的产物应是一份有边界的稳定性说明：在什么输入范围、时间尺度和资源条件下，系统维持了哪些变量；哪些后果没有被测量；哪些改变会要求重新估计。这样的说明既可以支持工程应用，也可以为后续机制研究提供清楚入口。它不需要把稳定性神秘化为生命本质，却能够揭示稳定作为持续生成的结果所具有的结构与代价。

19.8 从维持到开放的能力

生命稳态的讨论使发生理论获得一个必要修正：稳定不是发生的例外，而是发生的一种重要成果。系统通过持续交换和调节建立可重复关系，科学通过控制和校准识别这些关系。两者都可以在历史中改变，但改变并不取消先前的有效性。真正需要避免的是把有限范围中的稳定扩大为无条件本质，或者把条件变化误写成规律彻底失效。

开放能力也应被限定为对相关新条件的响应范围。它可以表现为恢复、切换、重新学习或建立新的协同，但每一种表现都有成本。一个系统不可能在不受任何约束的情况下响应所有未来。发生研究应比较可达范围怎样被历史塑造，以及维持当前功能需要放弃哪些可能。这样，未显露的可能性就不再只是诗性语言，而成为可以通过适当探测逐步限定的能力描述。

这种能力描述必须保留未知。未尝试的扰动不等于已经能够处理，未观察到失败也不等于无限稳健。研究者可以根据模型提出外推，但应明确外推依赖的条件。若新条件破坏了这些条件，失败可能提示模型范围，而非单个参数误差。SDE 所要求的历史承担，包含保存这些外推承诺，使后来的修订能够追溯到具体假设，而不是悄悄改变理论所说的对象。

从细胞到组织，稳定性还可能存在层次冲突。局部维持可能妨碍整体调整，整体稳定也可能依赖局部更新。这样的冲突不能通过选择一个最高层级就自动解决，需要研究具体

交换关系和功能目标。对于涉及人的应用，还应当说明谁享有收益、谁承受代价。基础机制与规范选择在此相遇，但并不合并为同一方程。

发生论因此可以把稳态研究与创造性研究连接起来。发现原有恢复范围不足时，研究者可以寻找新的调节路径、重新组织边界或改变任务条件，而不只是追求更强反馈。新的方案是否有效，仍需与原有方案在同一目标和成本条件下比较。创造来自对约束的重新组织，并不意味着取消约束；新稳定性只有经过独立检验，才成为可供他人使用的知识。

稳态研究由此取得了三个相互连接的对象：持续维持的输出、实现维持的补偿组织，以及在新扰动下仍可使用的恢复余地。三者需要不同资料，当前输出的稳定尤其不能代替后两项。把这种区别带入第 20 章，多尺度设计便有了明确任务：在局部调节与整体功能之间建立能够接受探测的映射，并保存哪一项证据支持哪一种跨尺度结论。恢复成为桥梁的条件，正是其范围与实现路径能够被分开说明。

第 20 章 多尺度回写与生命研究设计

20.1 先确定跨越了什么尺度

多尺度研究的困难，并不只是资料来自不同仪器。细胞表达、神经群体活动和行为选择分别使用不同单位、不同时间窗口与不同解释目标。将它们放进同一数据库，并未自动建立尺度之间的关系。如果一种变量在细胞层面表示分子数量，在组织层面表示成员比例，在行为层面又被当作能力强弱，那么相同名称甚至可能掩盖三个不同对象。跨尺度研究必须先说明每次转换保留了什么、舍弃了什么。

一种转换是聚合：把多个细胞记录合成群体平均。另一种是功能映射：用某类神经活动预测行为。还有一种是机制连接：说明细胞过程怎样通过特定组织关系改变总体功能。三者的证据要求不同。聚合可以由统计定义完成，功能映射需要预测检验，机制连接则需要额外的干预或结构证据。不能因为聚合后的变量与行为相关，就认定已经完成从分子到行为的机制解释。

尺度转换是否可靠，可以通过两个次序的比较来检查。一个次序是先用局部模型预测各成员的响应，再聚合为总体；另一个次序是先把当前局部状态聚合，再用总体模型预测。如果两个次序在新扰动下给出相容结果，所选摘要便可能保存了与该任务相关的关系；若结果持续不同，就应寻找聚合遗漏的差异。成员之间的补偿、边界位置与状态比例都可以成为候选。SDE 的特征纠缠在这里体现为哪些组织关系必须穿过尺度转换而被保留，不能由一个均值或关系清单预先替代。

构造一个跨尺度问题：一项细胞处理之后，群体输出恢复，但成员之间的差异增加。若研究目标是解释总体维持，成员差异可能是补偿机制的一部分；若目标是预测下一轮扰动，差异分布可能比当前均值更重要。研究者不应先决定保留所有单细胞细节，而应比较哪些摘要能够承担目标任务。多尺度方法的质量，体现在能够找到有理由的压缩，而不是拒绝任何压缩。

研究单位也会随问题改变。若干预对整个培养容器实施，容器通常承担分配单位；容器内成千上万个细胞并不会自动提供同样数量的独立干预重复。若关注细胞间作用，成员之间的依赖恰是对象，又不能为了统计便利假定完全独立。设计需要同时规定干预单位、观察单位和推断单位，使样本量与结论层次对应。否则再精细的记录也可能产生过强的确定性。

跨尺度不是单向地从小到大。宏观任务条件可以改变个体能够获得的输入，空间安排可以改变局部接触，研究制度也可以改变采样策略。承认这些作用，并不意味着宏观概念无需物质实现；它要求研究者说明宏观安排通过哪些局部路径产生后果。发生理论所强调的相互规定，应在这种可追踪的实现关系中展开，避免以宏观决定一切或微观解释一切取代具体工作。

20.2 时间对齐不等于因果对齐

在需要比较动态先后或重建实时控制的联合研究中，可靠的时间对应是必要条件，却不是因果关系的充分证据。设备可能记录不同处理阶段，有些信号经过积分或平滑，有些指标只在采样结束后获得。若直接比较峰值先后，可能把测量延迟解释成机制顺序。研究需要保存时钟来源、触发方式和信号处理步骤，并区分事件发生时间、采集时间与结果可获得时间。

这种区分对自适应实验尤其重要。控制程序只能依据当时已经获得的结果调整下一步，而不能依据后来完成分析的指标。若模型训练时使用了延后才可获得的信息，却被报告为实时预测，就产生时间泄漏。发生档案应重建每一决策点的信息状态，使研究者能够检查当时的控制是否确实可行。历史不是按最后版本整理好的全知叙述，而是有限信息逐步进入行动的过程。

观察窗口也会改变可见结构。一个快过程在低频采样下可能表现为平稳平均，一个缓慢变化在短窗口中可能看似固定。更高频并不总能解决问题，因为它可能增加观测负担、数据相关性和技术误差。合理做法是根据候选机制设置不同分辨率，并在重叠窗口中检验它们能否相互解释。尺度之间的对齐应当有模型支持，而不能只靠插值让所有曲线拥有相同长度。

假设细胞层指标每隔一段时间取得一次，群体输出则连续记录。研究者发现某次细胞指标变化后群体输出改变，但细胞变化的实际时间只知道落在两个采样点之间。如果这一区间与群体变化重叠，资料便不足以确定先后。诚实的结论是时间分辨率不支持该顺序推断。下一项研究可以缩小关键窗口，而不是让模型拟合出的精确时点替代尚未观察到的事实。

回写周期也不能由自相关峰直接读取。周期性可以来自外部驱动、内部振荡或分析过程；闭环则要求输出通过某条路径改变未来条件。研究可以先利用时间结构寻找候选环节，再通过干预、延迟操纵或机制模型检验。若把任何有规律的时间序列都称为回写，概

念就失去区分能力。时间证据有价值，正因为它能够限制候选机制，而不是自动给出唯一机制。

时间映射可以包含部分顺序。若局部变化发生的区间与总体变化重叠，研究先保留二者在该窗口内共同变化的结论；若通过更细观察缩小区间后顺序仍不确定，就应选择能够检验方向的干预。对于只能在结束后测量的局部指标，则可以把它用作过程后的状态证据，避免将其放入当时的实时控制模型。如此处理，时间对齐成为一组可检查的对应关系：哪些次序由记录确定，哪些由模型推断，哪些仍待区分。技术分辨率的提高可以改变这些关系的确定程度。

20.3 从依赖图走向可干预关系

多组学和多通道记录容易产生密集的依赖图，但图中的每条边都需要解释。它可能表示统计相关、条件预测、物理接触或因果作用。不同类型的边不能因为画在同一图上就获得相同地位。SDE 中的特征纠缠可以借助这些关系获得操作内容，却不能被等同于任何一张复杂网络。研究首先应说明边的证据来源，再讨论它们如何共同限定显露。

回写机制尤其需要跨时间的方向证据。若当前输出改变了下一轮输入的有效性，就应当能找到某种中介或约束变化。例如，假设前一次刺激改变了后续响应，并不必然需要反向作用于外部环境；它也可能通过内部状态承接。因此，闭环的边界可以包含对象内部的更新，也可以包含对象与环境的相互改变。研究者需要明确所称的回写属于哪一种，避免用同一符号混合不同机制。

干预对象应尽量贴近候选解释。若假说指向某段历史的保持，单纯改变当前刺激强度可能无法检验它；若假说指向空间邻接，改变总体细胞数量也可能同时改变多个条件。理想的干预不是绝对只改变一个物理变量，而是在可行范围内使不同解释产生不同预测，并记录不可避免的伴随变化。控制的价值在于缩小解释空间，而不在于宣称获得完全孤立的因果箭头。

还需要区分切断路径与改变节点水平。将一个调控因素降低，可能影响多条路径；观察到后效消失，未必说明它只通过假定的那条边起作用。可以通过阶段性干预、补充测量和竞争模型逐步约束解释。若技术条件不足以独立改变某条路径，应当如实报告这种辨识限制。发生模型的丰富性不能成为越过技术限制的理由，反而要求更明确地标出哪些关系尚未被单独检验。

因果效应也可能依赖历史。某干预在初始状态有效，在训练或适应后作用减弱，可能说明系统改用了其他路径，也可能只是目标变量进入不同范围。研究应将干预与状态条件共同报告，而不是把平均效应当作无条件常数。多尺度设计因此需要在关键状态中保留足够比较，使效应的迁移范围能够被检查。否则一次总体平均可能同时掩盖不同方向的局部作用。

对 SDE 而言，因果图的作用是把共同发生变成可争论的具体结构。它不需要表示世界中所有关系，而应包含足以回答当前问题的候选路径。三方程提醒研究者检查双向规定，六路径帮助选择研究入口，三原理提出可能的更新方向；真正的因果主张仍由实验与模型条件支持。框架越有解释雄心，就越需要在每条边上保持这种证据节制。

20.4 让过程档案成为研究工具

过程档案的第一项任务是保存对象身份与状态。对于细胞研究，这包括样品来源、培养批次、处理记录和追踪方式；对于神经任务，还包括训练阶段与任务版本。档案不必记录所有能够想到的字段，而应优先保存影响比较和重建的条件。一个字段若无法解释、无法可靠取得或从未进入任何研究问题，就可能只是增加负担。完整性应当相对于可追溯目标定义。

第二项任务是保存差异序列。输入实际发生的时间、持续方式、顺序以及任何偏离计划的情况，都比事后简化的条件标签更有价值。若一次刺激未按计划到达，或操作者临时改变了采集间隔，这些事实需要进入记录。发生研究面对的是实际事件，不能只保留理想协议。对偏离的诚实记录，能够使其他团队判断结果差异来自执行还是机制。

第三项任务是保存读数的形成方式。原始信号、处理参数、校准版本和排除理由需要建立可追溯关系。不同分析版本可以产生不同的分类和指标，这种变化应被记录为知识处理历史。它不等于对象本身发生改变，却可能影响下一轮实验的选择。把对象历史与分析历史放在同一档案中并明确区分，能够研究知识怎样回写研究，同时避免两种回写被混为一谈。

第四项任务是保存后效。许多研究只记录主要终点，却没有保留下一轮能够检验历史承接的资料。若项目声称存在回写，档案至少需要说明后续观察的窗口、条件和缺失原因。没有后效记录的研究仍可能提供重要当轮证据，但不能承担跨轮机制的全部责任。研究设计可以将后效追踪作为分阶段扩展，而不必要求每个项目一次完成所有层次。

档案也应有权限与用途边界。涉及人的资料不应因发生理论重视历史，就被无限收集。研究伦理需要尊重参与者、自主决定、风险收益和公平等要求；《贝尔蒙报告》为这些问题提供了明确的原则讨论。[37] 这里的引用不是用规范文本证明某个经验机制，而是说明方法选择本身承担价值责任。资料越能够重建个人历史，越需要清楚说明收集目的与使用范围。

好的档案应支持删减。若先导研究发现某些字段不可靠或没有增益，可以减少常规采集，并保留决定依据。若新增机制问题出现，又可以有目的地补充。这样，档案成为随问题更新的研究工具，而不是不断膨胀的仓库。SDE 所要求的历史厚度不是资料数量，而是关键后果能够被承接、关键决策能够被复查，以及新的研究能够从旧失败中得到具体约束。

20.5 一个从局部恢复到群体输出的设计

第 17 章已经建立了区分细胞历史承载的两阶段思路，第 19 章则说明数值恢复与恢复余地可以分离。这里构造一个尚未实施的研究，把两项问题连接起来：群体平均输出恢复之后，下一轮扰动的结果主要取决于平均状态、成员组成，还是成员之间的空间组织？研究同时保留单细胞层与群体层的测量，但其目标是识别跨尺度所需的最小表示，而非把所有尺度一次测全。

第一阶段建立一组相近总体输出的条件。它们可以来自不同处理历史，但在比较时应说明培养环境、群体数量和测量窗口的匹配程度。匹配只保证已经记录的方面可比，不预先保证微观状态相同。需要独立估计总体输出与单细胞记录之间的关系：总体量是否确实可由所采成员加总，是否包含未进入成像区域的贡献，以及是否存在额外的非线性转换。只有这条观察映射建立以后，局部记录才有资格解释总体差异。

第二阶段保留三种逐步增加信息的摘要。最简摘要只有平均状态；第二种增加成员比例或分布形状；第三种再增加与当前机制有关的邻接或边界信息。使用相同记录形成这些摘要，可以比较信息压缩的损失。假如平均值已能预测后续探测，复杂表示应暂缓；假如组成改善预测而空间信息没有增益，就把后续工作集中在成员更替及状态转变。每一项新增表示都需要承担具体任务，以免多尺度研究变成无限增加变量。

第三阶段施加能够区分这些表示的后续扰动。若平均模型认为历史已经被当前均值充分吸收，它应预测均值相近条件中的响应相近；若组成模型认为成员比例仍重要，它应预测同均值异组成的差异；若空间模型认为传递依赖邻接关系，它还应对同组成异排列提出

额外预测。实际操作往往同时改变多个条件，因此需要记录伴随变化，并检查这些变化是否已经足以解释结果。此处列出的是模型之间的分歧，方向与幅度须由具体系统确定。

第四阶段区分局部机制与整体承接。某项细胞层干预改变了成员响应，只有当这种变化通过已验证的尺度映射预测总体后果时，才能支持更强的跨尺度解释。若局部变化发生而整体输出保持，应检查补偿或观察范围；若整体改变而被测局部指标不变，应检查遗漏路径。成像程序也可能参与这些差异，已有方法研究使照明过程成为应控制的候选条件。[32] 其影响需要在本设计中估计，而不能直接搬用其他实验的数值。

最后把已经建立的映射带到新的历史或排列条件。若局部机制仍有效而总体预测失败，修订的重点可以是组合方式；若连局部探测也失效，则需先回到状态定义。这样的失败定位，使团队能够交接明确的成果：哪些局部关系已经成立，哪些聚合在何种范围内有效，哪一种新组织要求重新估计。分阶段设计的价值，是让机制、组合和用途逐步获得各自的证据，而不是把一次局部成功扩写成从细胞到整体的全部解释。

20.6 模型竞争与有限不变性

多尺度模型容易因参数众多而获得更好拟合，因此比较必须对称。各模型应尽可能使用相同的训练资料、相近的调参机会和一致的验证窗口。若一个模型拥有完整历史，另一个只能看到当前值，比较首先证明的是历史信息的价值。要检验更新机制，则还需要在相同信息条件下比较不同结构。把信息优势与机制优势分开，能够更准确地说明研究真正取得了什么。

跨尺度比较尤其需要建立一个直接预测的基线：只用群体当前状态预测下一轮输出，再与通过单细胞状态和组合规则得到的预测相比。若后者更好，应继续检查增益来自更多信息，还是来自可迁移的组合约束。可以保留相同局部信息而替换组合方式，或者固定组合方式而减少局部字段。两类比较的结果分别决定下一步投资：改善采集还是改善连接。第 15 章的一般模型判据在这里落实为尺度接口的检验，使模型优劣对应能够承担的具体预测。

评价也需要多种任务。短期预测检验局部动态，长期预测检验状态承接，干预预测检验机制，跨条件预测检验迁移。一个模型可能在其中一项较强、另一项较弱。报告应保留这种结构，而不是只公布最有利的总分。研究者还应考虑错误代价：对平均输出的微小改善，未必值得失去对关键失败情境的识别。模型选择需要与研究目的共同论证。

有限不变性是多尺度研究的重要成果。如果某关系在一组条件变化下保持，就可以作为连接不同实验的桥梁。但不变性需要明确所经受的变化类型，不能从几次重复推出普遍有效。某模型可能对输入强度稳健，却对顺序敏感；可能在同一批次内有效，却无法迁移到新群体。记录这些边界，使常数和稳定关系重新获得可使用的地位，而不是成为发生理论必须排斥的对象。

失败定位可以沿尺度转换的两个次序展开。若局部模型在新条件下已经失准，总体失准首先来自状态或局部机制；若局部预测准确而聚合后失准，需检查成员依赖、未观察部分和加总规则；若两条预测都准确，却无法支持原定行动，则需要修订用途。另有一种情形是两条次序碰巧得到同一错误均值，这时平均误差会掩盖方向相反的局部偏差。保留分项结果和关键异常，使这种抵消能够被发现。尺度一致性由此成为可以失败的要求，而不只是图上连通的形式。

SDE 的增量应在这套竞争中被具体衡量。它可能帮助找到先前遗漏的事件边界，提出能够区分模型的干预，或者组织跨尺度档案以便重建失败。只有当这些贡献进一步产生可检验机制时，才可以作更强经验主张。概念贡献并不卑微，经验贡献也不能凭概念雄心预支。把贡献层次分开，反而使框架能够在不同成熟阶段获得清楚而持续的进展。

20.7 协作中的解释权与代价

跨学科项目常常由不同团队掌握不同环节：有人理解分子机制，有人负责成像，有人构建模型，有人解释行为。若每个团队只收到最后的摘要量，接口就可能丢失关键条件。合作需要共同定义事件和变量，说明哪些处理步骤会影响其他团队的解释。所谓同一发生，在研究组织上首先意味着没有任何一方可以把自己无法处理的条件默认为他人已经解决。

共同语言也不等于统一术语。一个团队说状态，可能指瞬时表达，另一个团队说状态，可能指可持续的行为策略。项目应记录这些词的操作含义，并通过具体案例检验是否指向同一对象。若不能统一，应保留并列描述和映射关系。强行使用同一词汇可能产生表面一致，却让后续争论无法定位。解释的协作需要允许有理由的差异。

采集成本必须进入方法评价。增加连续记录可能提高信息量，也可能消耗样品、延长任务或增加参与者负担。一个理论如果要求所有条件全部入账，却不考虑成本，就无法成为实际科研程序。研究应选择最能区分候选机制的观测，而不是最能展示复杂性的观测。

对人类参与者而言，额外负担还需要伦理理由；对基础实验而言，样品和资源成本同样决定研究能否重复。

解释权也与责任相连。模型团队可以提出某关系可能是反馈，却不应替实验团队宣称已完成选择性干预；实验团队可以确认操作成功，却不应仅凭操作名称决定模型因果结构。各方需要明确主张的证据来源和复核者。跨学科研究真正的整合，是让不同证据互相限制，而不是让权威较强的一方替所有尺度下结论。

发表安排应保留不利发现和协议修订。若先导研究发现某记录方式改变对象，应该把它写成方法结果，而不是因不符合最初目标而删除。若模型比较显示简单模型已足够，也应保存适用条件。这样的成果能够减少其他团队重复错误，并为未来扩展提供基线。科研新典范的建设不只在论文中的理论名称，也在共同体如何保存这些通常不够戏剧性的知识。

团队交接还需要一个最小的复核动作：接收方能够凭记录重建一个真实事件，并指出哪些判断来自原始资料、哪些来自处理规则。如果只能运行代码而无法解释变量，交接尚未完成；如果能够口头解释却无法定位证据，交接也不充分。选取少量具有代表性且包含异常的事件共同复核，比仅签署统一术语表更能暴露尺度接口中的缺口。

协作的成熟还体现在允许一个问题暂时停留在有限层次。某项目可能只能证明历史信息改善预测，不能定位分子机制；另一项目可能识别了局部机制，却尚无跨尺度功能证据。两者都可以成为可靠积木。只有当连接关系得到支持，才应合成为更强解释。SDE 的共同框架应帮助这些成果对接，而不是要求每篇研究都声称完成从分子到主体的全部链条。

20.8 生命研究的创造性进入

多尺度研究最富创造性的时刻，常常出现在旧分类无法安放一个可靠现象的时候。味觉曲线提示最终评分遗漏过程，细胞历史提示当前状态遗漏承接，神经学习提示任务结果遗漏可达组织。此时研究者需要决定从哪里进入新的问题。六路径提供的不是自动答案，而是多个有纪律的入口：从稳定条件寻找序列，从异常显露重建关系，或从新原型反推其成立条件。

例如，一个构造项目反复发现两类细胞具有相同终点指标却不同后效。可以从显露进入，寻找与后效共同变化的特征；也可以从差异序列进入，改变两类细胞形成之前的经历；还可以从特征组织进入，操纵空间或调控条件并观察是否生成相同后效。三种入口可

能互相补充。选择应根据可获得证据和可区分预测，而不是根据哪种路径听起来更接近整体。

研究入口的价值还要由接口能否承接来检验。一个新增细胞类别若只在一种仪器中可辨认，其他尺度便暂时只能把它作为局部发现；若类别可以稳定映射到后续群体响应，它才开始承担跨尺度条件。三方程要求这项映射同时照顾显露、形成序列与特征依赖，三原理则追问接口失配迫使哪一项重建。例如，局部类别相同而整体后效不同，可以推动加入排列条件；若排列已经保留仍无法预测，则应回查类别是否省略了相关历史。每一步改变都有明确对象，也留下下一轮检验的入口。

创造也包括发现一个问题没有必要扩大。如果现有状态模型在相关任务中持续有效，最好的研究可能是确定其边界和降低采集成本，而不是增加新的哲学变量。若一个新指标只重复已有信息，删除它可以提高可解释性。发生理论不应把复杂化当成唯一前进方向。它的目标是使研究单位与问题更加匹配，这种匹配有时通过扩展实现，有时通过明确压缩实现。

生命部分由此形成一项有限而积极的主张：属性的稳定和变化可以通过事件、历史与条件组织共同研究，测量程序也能够进入可检验的解释。味觉、细胞、神经与稳态并不因此共享一个普遍动力方程。它们共享的是对生成条件、承接机制和独立检验的要求。若这些要求能够持续产生更好的问题和更可靠的比较，SDE 就获得了方法上的实质贡献。

这些尺度接口仍主要连接物质状态、任务活动与观察程序。进入第 21 章后，还需增加另一类承接：被研究者能够理解公开判断，并据此修改计划、承诺与资源安排。局部状态到总体指标的映射纪律可以继续使用，但人的理由和制度规则需要新的说明。生命研究在此留下的成果，是让每一次跨越都带着已经验证的关系及其范围，避免将细胞中的调节机制直接当作经济行动的解释。

第五编 经济指标与制度回写

第 21 章 经济参数的条件性

21.1 一个数承担了什么承诺

同样一笔额外收入，可能被一个家庭用于偿还到期账单，被另一个家庭留在存款账户，被第三个家庭转为一件早已计划购买的耐用品。若把三个家庭的反应平均起来，就能得到一个有明确统计意义的数。这个数能够帮助判断一次转移支付在特定时期对消费的影响，却未必能够直接说明另一批家庭在另一种支付安排下会怎样行动。困难并不在于平均数虚假，而在于从一个有条件的描述走向另一个场景的预测时，哪些条件被默默带走了，哪些条件已经改变。

经济参数经常处在这样的转接位置。它把许多行为和制度关系压缩为可以进入计算的表达，同时向使用者许诺某种可重复性。这个许诺有时只适用于一个样本，有时指向一个人群，有时进一步承担政策干预后的预测。三个层次并不相同。一项统计量可以被精确计算，却不能直接外推；一个结构参数可以具有清楚的理论定义，却在现有数据中估计得很不精确；一个政策关系也可能在观察期内稳定，而在政策制度改变后发生变化。只有先弄清参数承担哪一种承诺，所谓不稳定才有可以追问的对象。

本章把经济参数放回具体发生之中，所针对的不是数学表达本身，而是表达与使用之间的脱节。边际消费倾向、货币乘数、自然失业率或非加速通胀失业率、资本产出比与全要素生产率、风险厌恶系数，常被并列为经济学试图寻找固定量的例子。这样的排列容易掩盖根本区别：其中有行为响应、有资产负债表比率、有模型中的潜在状态、有核算残差，也有偏好模型的参数。它们的变动原因不同，证据要求也不同。将它们统称为普遍常数，然后宣布常数寻找失败，并不能推进发生研究。

属性发生的观点首先改变提问方式。一个经济参数显露为什么样的稳定性，要看哪些差异被纳入比较，哪些差异被平均，哪些特征通过制度和行为彼此规定。这里的特征纠缠并非给已有回归式加上一个叫作环境的变量。它指向更具体的构成：收入何时可用，承诺是否可信，债务何时到期，未来能否借款，行动者怎样理解当前事件，这些特征共同决定什么才算一次可比较的收入变化。发生单位没有确定，参数的数值就可能是若干不同问题的混合回答。

这一取向与现代经济学的条件分析有广泛交集。异质主体模型、动态决策理论和制度比较早已把许多条件写进研究。发生学应从这些成果继续推进，不能靠删除它们来获得原创性。其独立责任在于指出：某个被默认不变的对象身份、响应窗口或更新规则，是否实际上由研究过程和制度实践共同维持；一旦它改变，研究者是否应重新定义参数，或者改变记录与干预的方法。可成立的贡献必须落在这些具体变化上。

21.2 消费响应中的收入事件

边际消费倾向描述收入变化所引起的消费变化，但“收入变化”并不是一种只有金额的事件。一项预先知道的工资增长、一笔突然而至的退款、一项附带资格限制的补助，以及一次必须偿还的借款，可能在到账时增加相近的可用资金，却具有不同的未来含义。若研究者把这些变化放在同一列中，再以一个系数概括响应，系数的跨研究差异就可能首先来自事件定义，而不必来自人性的飘忽。

Kaplan 与 Violante 对异质主体模型的讨论把资产结构、收入过程和家庭差异放在解释消费响应的中心。拥有财富并不意味着当下可以方便地支出；非流动资产与流动资产的配置会改变家庭对收入变化的反应。[38] 这意味着一个强的比较基线已经存在。发生研究不能满足于再次证明富裕家庭和流动性紧张家庭有所不同，而应追问相同可观测资产状况之下，响应差异是否还由尚未记录的事件历史或制度安排产生。

设想两个家庭在月底拥有相同的存款、债务和收入水平。其中一个家庭刚经历过两次补助延迟，另一个家庭始终按时收到款项。这是一个分析性情境，并非已完成的调查。若两者对下一笔补助的支出不同，至少有三种解释：前者形成了不同的到账信念，前者还承担了未被记录的费用，或者研究者把消费决策窗口选得太短。只有区分这些解释，才能把历史进入当前行动的方式说清楚。简单增加一个“过去经历”标签，会把不同机制再次揉在一起。

可以先取得支付宣布时间、实际到账时间、最早可用时间和家庭支出时间的连续记录。宣布使未来资源进入预期，到账改变资产负债表，可用性决定能否完成某些交易，消费发生则表现为不同类别的支出。四个时点可能重合，也可能分离。一次干预如果改变的是支付可预期性，它的机制便不能用到账后的余额变化完全代表。若改变的是支付渠道，手续费和取款条件也可能进入响应。这样的记录使“同样的钱”成为可以检查的判断，而不是研究者的预设。

这套记录还要区分支出时点移动与累计消费增加。假设延期家庭原本计划在通知之后购买必需品，却因无法借款而推迟到到账之后；按到账后一周计算，它可能显示较高消费响应，按通知前后较长窗口计算，差别却主要是支出的重新排序。另一家庭可能先借款消费、到账后还款，于是在同样的一周窗口内显示较低响应。这里的两种低高排列可以完全来自同一生活计划受到不同支付条件约束。研究应同时保存消费类别、债务变化和预先规定的累计窗口，才能判断新增资源究竟改变消费总额、消费时间还是融资方式。

特征纠缠在此有明确对象：到期责任改变了可用收入的含义，支付可信度改变了通知能够支持的计划，借款机会又改变到账与消费的先后关系。这些特征共同规定一次收入事件能与哪些事件比较。显露的消费增量和差异序列因此应一起界定；若窗口只是事后按最显著的结果挑选，研究者得到的便不是可迁移的条件参数。

预测检验应留出完整的新支付安排，而不只是从同一批记录中随机抽取若干行。若一个模型能够解释已有家庭，却不能预测支付时间改变后的反应，它可能遗漏了可用性机制。若加入历史变量的模型只提高原样本拟合，而在新安排中没有收益，历史解释就尚未获得支持。反过来，如果既有资产和预期模型已经充分解释响应，新术语便没有新增任务。发生学的价值不在于保留更多历史，而在于找出何种历史对于当前问题不可压缩。

21.3 货币乘数与资产负债表的发生顺序

货币乘数的讨论特别容易把一个事后比率误写成一条自动运转的因果规律。广义货币与基础货币之间可以计算比值，但这个比值存在，并不等于银行在每个时期都按某个固定倍数，把先收到的准备金转化为贷款和存款。定义上的比例、特定制度下的模型关系和实际业务的发生顺序，需要分别处理。

英格兰银行关于现代货币创造的说明指出，商业银行发放贷款通常同时创造存款，银行并非只把已有储蓄转贷出去，也不是简单按照固定倍数扩张中央银行货币。这一过程仍受到盈利、风险、监管、结算以及货币政策等条件约束。[39] 因而，批评固定倍数的机械图式可以成立，但不能进一步推成银行不受约束，或者所有货币总量都是任意构造。发生顺序的修正必须连同约束一起保留。

可以用一项简化业务说明。某企业提出贷款申请，银行评估其未来收入和担保，签订合同并记入贷款资产，同时在企业账户形成存款负债。企业把存款支付给另一家银行的客户，才出现银行之间的结算需求。这只是便于分析的简化过程，现实业务还包含授信额

度、抵押、流动性管理和监管要求。它的意义在于显示，贷款决策、存款形成和准备金结算虽相互关联，却不能被置换为一条无条件的单向倍增链。

如果研究者观察到准备金增加而信贷没有按比例扩张，至少需要检查新增准备金的来源、银行是否面临资本约束、借款需求是否改变、贷款风险如何评价，以及持有准备金的收益和替代资产回报。不同答案指向不同机制。仅用“乘数失效”概括，可能把一个本来就不承担稳定因果含义的比率当作失败理论；仅用“制度变了”解释，又可能没有说明制度究竟改变了哪项行为。

还可以把比率变化分成两个不同问题。一个货币比率上升，在算术上可能来自分子增加，也可能来自分母减少；即使变化由同一端引起，所对应的资产交易仍可能不同。因而，只把前一期比率代入下一期总量预测，已经额外承诺了分子与分母之间的响应关系。这个承诺需要业务机制支持，不能由比率的定义取得。相反，若研究任务只是描述某一时点两类货币存量的相对规模，该比率完全可以成立。错误发生在从存量关系跳到信贷生成规则的那一步，而不在计算比率本身。

更有辨别力的研究可以比较两种制度改变。第一种只改变准备金数量，第二种改变贷款损失承担方式。即使两者导致同样的货币比率变化，其后续信贷配置也可能不同。应记录业务发生顺序，观察风险判断何时变化、哪些申请获得批准、结算需求如何调整，再判断哪条机制解释了响应。若既有银行模型能够给出同样预测，发生学的增量应转向事件记录和跨制度比较，而不应把同一解释重新命名。

21.4 潜在失业状态与实时判断

失业率是按照特定调查与分类规则形成的统计量，自然失业率和非加速通胀失业率则属于另外一层对象。后两者与理论结构、劳动市场机制和通胀关系有关，通常不能像调查中的就业状态那样直接清点。它们也不能不加说明地互换。若研究者把不同概念、不同估计方法和不同数据版本列成一串数，再以数值变化证明同一个常数不存在，论证从起点就失去了可比性。

Staiger、Stock 与 Watson 的研究在多种设定中考察了非加速通胀失业率估计的精度，显示有关判断可能具有相当大的不确定性。[40] 精度不足不等于概念没有意义，更不等于结构不可辨识或动力学不收敛。一个参数可以在理论上可辨识，却因为样本长度、噪声和模型选择而难以精确估计；也可能估计得很精确，但模型中的对象已不适合新的制度条件。两种困难需要不同的改进。

政策制定还有一个独立问题：当时能够知道什么。研究者事后使用修订后的就业、价格和产出数据，可以得到比当时更连贯的历史图像；决策者却只能使用当时已经发布的版本。若把事后数据回填到历史政策模型，就可能把后来才获得的信息当成过去已经拥有的知识。发生研究因此需要两条时间轴，一条记录经济活动及其后果，另一条记录信息何时被获得、修订和采用。

设想某次政策评估认为失业已经低于一个估计的可持续水平。随后工资、价格和参与率的变化与预测不符。研究者可以从三个方向修订：估计原本过于不精确，劳动市场关系已经改变，或者政策自身改变了求职、招聘和议价行为。第三种解释不能只靠政策发生在前、结果变化在后来成立。政策可能是在回应同一个冲击，而工资与价格也受到其他因素影响。需要追踪政策规则、预期更新和实际行为之间的连接。

在这种情形下，估计值具有双重地位。它一方面是对潜在状态的判断，另一方面可能进入政策规则，成为下一轮行动的条件。两种地位可以同时存在，却不能相互替代。估计影响了政策，不证明估计准确；估计不准确，也不意味着它没有制度效力。若一个被误估的状态导致了真实干预，研究者必须同时研究认识误差与行为后果，而不能只在修订后的统计表中寻找答案。

可以把修订拆成三次受控比较：保持模型不变只换数据版本，保持数据版本不变只换估计设定，再考察不同政策阶段中关系是否持续偏离。第一步定位资料更新，第二步显示模型依赖，第三步才涉及结构变化的候选解释。三步不能保证完全分解，因为资料与模型也可能同时更新，但它们至少阻止把全部修订都归于劳动市场本身。最终报告的状态判断应与当时信息和估计区间绑定。如此得到的并非永不变化的失业门槛，而是政策在某一知识状态下可以承担的判断范围。

21.5 资本比率与生产率残差

一座工厂的设备价值与它的年度产出之间可以形成比率，一个国家的财富总额与收入之间也可以形成比率。两者都涉及存量与流量，却不是同一对象。设备、土地、住房、金融资产和其他财富项目的估值方式不同，纳入的经济关系也不同。把财富收入比的历史变化直接当作生产资本产出关系的反例，会把分类差异误当成理论冲突。

Piketty 与 Zucman 对富裕国家财富收入比的历史研究为长期变化提供了材料，但其财富口径不能未经处理就等同于生产设备资本。[41] 对发生研究来说，关键不是借一条长曲线宣布资本不稳定，而是分解曲线为何变化。资产数量增加、资产价格上涨、收入暂时

下降或统计边界扩展，都可能提高比率；这些变化对生产能力、分配关系和未来投资的意义并不相同。

假设两个地区的资本产出比相同，一个地区的设备保持闲置，另一个地区的设备持续运行，但统计中的设备价值较低。仅凭相同比率无法判断两者的生产过程是否相同。相反，若两个地区比率不同，也可能因为估价时间或行业结构不同，而不意味着技术关系完全改变。因此，比率不是一张自动透明的窗口，它把多种过程压进了分子和分母。研究者需要拆开这种压缩，确定什么差异与当前问题有关。

全要素生产率又是另一种对象。在增长核算中，产出变化中未由所计量投入解释的部分，可以被表达为生产率变化，但残差的名称不能保证它只包含纯粹技术。投入质量、利用率、测量误差和模型设定都可能影响它。Fernald 的季度利用率调整研究已经把劳动努力和资本使用强度纳入处理。[42] 因此，“残差混合了多种机制”是应当继承的研究问题，而不是发生学首次揭露的秘密。

一个具体研究可以从同一工厂的生产调整开始。订单增加后，管理者先延长机器运行时间，再调整工序，最后购置设备。若只按季度记录设备存量和劳动人数，前两步可能被压入残差。增加运行时间和工序记录，有机会把某些变化从未知项转为可解释过程。然而，记录更细并不自动带来更好的跨工厂预测。某个工序指标可能只适用于一种设备，增加它还可能引入不一致的报告方式。模型应同时接受解释收益和采集成本的检验。

这里值得关注的发生问题，是生产能力本身怎样通过组织学习、维护和协调形成。相同设备并不保证相同的可用能力，因为故障处理、知识分布和操作安排会参与能力的实现。但这不要求把所有差异都称为不可还原。若经过适当调整的投入模型已经解释了变化，就应承认这种压缩成功；若无法解释，则要提出可以观察的更新机制。残差的开放性提供研究入口，却不替任何特定本体论作证。

21.6 风险选择与参考点的形成

一个人在小额选择中避免风险，在另一次选择中接受较大风险，并不直接证明其风险厌恶系数已经改变。选择还取决于他怎样理解概率、把什么当作得失起点、是否相信承诺能够兑现，以及是否还有其他未被题面写出的责任。模型中的效用曲率是一种解释资源，但观察到的选择是多种条件共同形成的结果。二者之间需要识别，而不是凭直觉对应。

Kahneman 与 Tversky 的前景理论把收益与损失的区分、参考点和决策权重纳入风险选择的解释。[43] 这为本书提供一个强近邻：情境依赖并不等于只能诉诸任意变化。研

究者可以建立有结构的条件模型，并在不同任务中检验。发生学若要继续推进，尤其应追问参考点怎样形成、怎样被保留，以及一次选择的后果如何成为下一次选择的条件。

考虑一个明确假设的情境。参与者连续处理若干具有相同客观概率的任务，第一次的结果被标记为奖励，第二次的同额结果被标记为恢复先前损失。行为差异可能来自参考点，也可能来自对实验者意图的猜测、对概率的误解或疲劳。若只把前后差异归入主体性，就没有完成解释。应分别记录任务理解、结果反馈和后续信念，并设计能够改变其中一个环节而尽量保持其他条件的比较。

更深的困难在于，要求参与者报告参考点本身可能改变其注意。一个人原本并未明确计算自己“应该拥有多少”，被询问后却开始围绕这一数值组织选择。此时，测量并不是一扇不留痕迹的窗口。可以比较无提示、事后回顾和事前询问三种程序，但不能把它们的结果混为同一构念的三个无差别读数。记录程序需要进入事件描述，研究者才能说明所发现的是原有过程，还是被询问后形成的新过程。

这种分析不会使任何选择都免于预测。相反，它提出更具体的预测责任：若参考点确实由某种反馈形成，改变反馈方式应当在预定条件下改变后续选择；若反馈实际不变，仅改变叙述就复制了原先归因于反馈的选择差异，则需进一步比较叙述、任务理解与参考点更新等机制；叙述改变而选择保持不变，本身并不反驳反馈机制。稳定的个人差异仍然可能存在，并可在适当模型中发挥作用。发生学要解释的是个人差异与任务形成之间怎样结合，而不是用情境的丰富性取消对人的持续认识。

风险厌恶、损失厌恶与对概率的判断还应在估计中分开。给定一种效用形式，风险厌恶系数描述的是该形式中的曲率；以参考点区分得失，则改变了结果被评价的位置。一次拒绝同样的赌局，可以由不同参数组合解释。若资料只包含单一金额和一种表述，复杂模型可能在拟合上彼此接近，参数值却取决于未被检验的限制。更有辨别力的任务应分别改变结果规模、得失参照与概率说明，并说明哪些变化足以区分候选模型。这种任务扩展不保证识别成功，却能指出新观察应落在哪里。对于仍不可区分的部分，应保留参数集合或预测范围；用一个精确的点估计替代这种未决，会给迁移造成虚假的确定性。

21.7 政策规则如何进入参数

在某一政策制度下估计出的行为关系，未必适合评价另一制度。Lucas 对政策评价的批评已经明确提出这一困难：决策规则与预期可能随政策安排改变，历史关系的拟合能力不能保证政策反事实的有效性。[44] 本书据此不再把“政策改变参数”作为新发现，而把

问题推进到更细的层面：政策改变通过哪些被记录或未被记录的过程进入了参数，研究怎样识别这些过程，何时仍能维持一个较简单的表示。

例如，一项按固定规则反复实施的补助，与一次被明确宣布不会重复的临时补助，可能具有相近的当期金额。家庭对未来收入的判断却未必相同。若研究者用前者的平均消费响应预测后者，预测失配并不需要假设偏好突然变化；承诺的持续性已经足以构成候选机制。相反，如果家庭根本不相信持续性承诺，两种政策也可能产生接近的响应。政策文本、政策可信度和行动者实际采用的判断，不能被当成同一个变量。

这要求研究把规则的宣布、执行和体验分开。政府宣布一种资格规则，执行部门可能按不同节奏处理，家庭则通过个人经历形成对规则的理解。三个层次之间的差异，可能在总量数据中只表现为一个不稳定系数。发生分析的工作，是从系数差异倒推值得追踪的环节，再用新的记录或干预加以区分，而不是在看到不稳定后立即宣布整体制度不可分析。

政策还可能把参数本身变成目标。若一个部门根据估计的信贷响应安排资源，银行的报告和业务选择就可能适应这一标准。此时，参数变化既可能反映真实机制，也可能反映被评价方式改变后的行为。研究者需要追踪指标何时进入考核、哪些行动因之受奖惩，以及指标改善是否对应目标结果。这样的分析连接认识与制度，却不把任何适应都视为欺骗。行动者可能合理地回应新的激励，也可能通过学习改善过程。

政策反身性需要比系数变动更强的证据。以补助为例，可以先比较政策宣布是否改变未来收入判断，再比较这一判断是否进入当前消费计划，最后检验计划变化是否在实际可用资金的约束下实现。若宣布之后支出增加，却没有信念资料，现金提前到账和收入预期改善仍是竞争解释；若信念变化但支出未变，流动性约束或既有承诺可能阻断行为通道。发生研究应据此决定重估何种参数：改变的可能是信念分布、约束状态或条件响应，而不必是模型中最深层的偏好。把需要更新的位置找准，才能使政策解释具有超出“制度不同所以结果不同”的内容。

把政策纳入参数发生，并不要求所有参数随时重估。持续估计同样有成本，也可能把短期噪声误当结构变化。更合理的做法是预先规定更新触发条件：数据定义变化、政策规则变化、响应持续超出误差范围，或某个关键中介关系不再成立。稳定时期可以保留简洁模型，变化时期再开启更细的研究。这样形成的稳定不是盲信，而是一种有监测、有修订入口的知识安排。

21.8 条件参数的研究资格

五类案例说明，经济参数的条件性不是一个足以解释一切的结论。它更像一项研究要求：每个数都应说明自己怎样形成，能够回答什么问题，以及在何种变化下需要重新解释。对事后比率，应优先检查口径与组成；对行为响应，应界定事件和窗口；对潜在状态，应区分概念与估计精度；对核算残差，应追踪尚未分离的来源；对偏好参数，应检验它与信念、参考点及选择集合能否被区分。

一个参数要获得跨情境使用的资格，需要比原样本中的显著性更多的证据。首先，目标人群应与原人群具有足够的共同支持，否则重新加权也无法创造从未观察过的情况。其次，相关条件下的响应关系需要具有可迁移性，不能因为观察变量相同就假定机制相同。再次，信息时点必须一致，不能用后来修订的数据模拟当时预测。最后，改变参数用途可能改变行动者行为，这种制度后果需要被纳入使用范围。

迁移失败以后，处理方式也应随参数身份改变。消费响应在新支付安排下失准，应重新比较事件与家庭构成；货币比率变化，先分解两端存量及业务来源；潜在失业状态修订，先区分资料与模型；生产率残差扩大，先检查投入和产出的计量；风险选择反转，则要排查参照、信念与选择集合。五项诊断不能互换。只有在相应替代解释受到约束后，才有理由提出更深的结构修订。这样，条件性从一句普遍断言转成了选择下一项观察的规则，也允许不同参数得到不同结局：保留、重估、重新定义或暂缓使用。

发生学在经济研究中的第一项责任因此是辨别，而不是宣判。它需要辨别固定关系的真实成功与被强行延长的适用范围，辨别资料修订与机制变化，辨别行为异质性与制度产生的分类差异。它还要解释为什么某些稳定关系可以跨越相当长的时间：也许约束持续存在，也许行动者不断校准，也许不同变化相互补偿。稳定是需要解释的正面事实，不是必须清除的旧典范遗迹。

本章的五类参数，大多是研究者从事件中估计或计算的表达。市场成交价还多一层构成：参与者在交易中就要使用它，并以其为条件作出承诺。即使研究者完全不估计价格系数，成交仍要发生或失败。第 22 章据此把问题从“一个数能否迁移”推进为“一个数怎样取得可履约的身份”，使参数的条件分析进入价格形成的具体过程。

第 22 章 市场价格 预期与交易事件

22.1 从一次成交辨认价格

一件物品标着一百元，买方报出九十元，卖方暂时没有接受；几轮商议后，双方以九十五元成交。这里的三个数具有不同身份：标价提供公开条件，报价提出一项尚待接受的承诺，成交价记录双方已经接受的交换条件。九十元甚至未必等于买方真正愿意支付的最高金额，它可能只是一次试探。若研究者不区分这些身份，就会把展示、策略和结果混成同一种价格。确定数值究竟记录什么事件，是解释价格大小的前提。

成交也并不总在付款的一刻完全结束。商品是否交付、质量是否达到约定、支付能否撤回、责任如何承担，都会影响这笔交易最终意味着什么。一个附有退货权的九十五元，与一个不可退货的九十五元，名义数值相同，实际承担的条件却不同。这些条件并非附着在纯粹价格外部的装饰。对行动者而言，它们共同决定接受这个数意味着作出怎样的承诺。

本章采用的发生单位，是在明确规则下形成并被参与者承接的一次交易事件。它可以包括此前的搜寻、谈判与预期，也可以包括之后的结算与履约，但范围应由研究问题限定。若问题是竞价规则如何影响成交，就需要保留报价和撮合次序；若问题是价格如何影响使用，就还需要履约后的观察。无限扩大事件不会自动提高解释力，恰当的边界应让候选机制有机会相互区分。

价格具有客观可核查的一面。只要记录可靠，研究者可以确认某人在某时按照某种条件支付了多少。价格也具有主体性和互动性的侧重，因为这个数来自参与者的判断、承诺和操作。两种认识没有冲突。发生于交易，并不使价格成为任意意见；能够被记录，也不使它成为与交易安排无关的天然属性。问题在于，一种稳定的公共记录怎样从分散而有差异的行动中形成，又怎样成为进一步行动的共同条件。

在这里，显露不是一个已经完成的数值被动地等待解释。标价转为报价、报价进入匹配、匹配经过履约，价格的身份会随着事件推进而变化。差异序列也不是任意的时间列表，它包括那些使意愿能够转化为承诺的关键区别。特征纠缠表现为数量、质量、权利、时间和信用的相互规定：改变其中一项，其他项的经济意义可能随之改变。只有追踪这些规定关系，发生学才超出给交易贴上三个标签的层次。

22.2 愿付价格从何而来

一个人愿意支付多少，不能仅由他昨天见过的价格推出。既有价格可以成为参照，但用途、收入、替代品、信息和等待能力也会参与判断。若把所有愿付价格都解释为过去价格的回写，研究便形成了一个自我封闭的循环：当前价格来自过去价格，过去价格再来自更早价格，而生产条件与生活需要反而无处进入。承认历史效应，不应取消其他差异的作用。

设想两个人购买同样一件工具。一人需要立刻完成一项已经承诺的工作，另一人只是为以后可能发生的任务做准备。他们拥有相同收入，也知道同样的市场报价，最高愿付价格却可能不同。这是由用途和时间约束产生的差异。若工具可以借用，差异又可能缩小；若借用需要额外担保，判断再次改变。价格形成之中，替代路径并不是抽象选择集合上的一个符号，而是行动者是否真正能够采用的安排。

过去价格的作用可以在这一结构中被精确定位。它可能帮助行动者判断什么价格通常能够成交，可能成为公平判断的参照，也可能被用来预测未来转售价值。这三种作用不同。预期成交价影响出价策略，公平参照影响接受与拒绝，转售预期影响持有决定。若改变过去价格的呈现方式，三种机制未必产生相同反应。研究需要提出能够区分它们的操作，而不是把所有历史依赖统称为锚定。

同样，卖方的保留价格也有多个来源。继续持有的收益、库存成本、生产安排和现金需求，可能共同影响最低接受条件。买方与卖方之间的差异不只是心理刻度不同，还包括他们承担不同后果。若一次拒绝对一方意味着等待，对另一方意味着无法履行迫近的义务，那么相同的议价程序可能产生不同的实际机会。这一事实能够进入因果分析，同时还会引出规范评价；两者应相互联系，却不能彼此代替。

成交记录对愿付价格的支持通常有限。在这个简化交易中，接受九十五元至多表明买方在当时条件下接受了该支出，不能单凭成交知道他在一百元时是否也会接受，更不能倒推出全部效用结构。卖方同意九十五元，同样没有公开其最低可接受条件。若进一步询问最高或最低价格，回答又可能受交易策略影响。于是，主体判断与成交记录之间存在需要研究的间隔：有时可以通过不同报价下的接受和拒绝缩小范围，有时只能在明确行为假设下估计。把范围当成范围，能够防止研究者把自己推算的尺度冒充行动者已经表露的尺度。

这样理解愿付价格，能够说明差异尺度的形成：人并不是带着一把完全独立的价值尺进入交易，判断的尺度会在交易经验中形成。但尺度也并非只由旧成交价复制出来。它在需要、资源、规则与经验之间获得形状，并通过接受、拒绝和履约后果继续调整。研究的任务是找到尺度改变的具体通道，并说明这种改变何时有助于形成稳定判断，何时造成持续失配。

22.3 交易规则怎样规定可见的差异

同样一组参与者，如果只能公开报出一个不可撤回的价格，与可以反复修改报价时，面对的行动问题不同。这一点不需要预先假定哪种安排更有效。规则首先改变的是哪些差异能够进入比较、哪些行动可以在看到他人行动后发生，以及承诺何时变得不可撤回。规则由此参与价格的形成，而不是在价格已经产生以后才负责记录。

可以构造两种分析性市场。第一种市场要求买卖双方同时提交密封报价，再按公开规则确定是否交易；第二种市场允许双方依次公开调整报价。两者中的参与者、物品和初始信息保持相同，只改变行动次序。第一种安排下，对他人意愿的判断主要通过事前预期形成；第二种安排下，报价本身还成为可观察信号。于是同一数值的报价可能在前者表示最终意愿，在后者表示试探或策略性承诺。若只比较平均成交价，就会遗漏行为身份的变化。

两种市场的比较还必须给未成交留下位置。假设可反复报价的安排让一部分时间充裕者继续谈判，而必须立即离开者退出，最后成交价即使更集中，也可能只反映留下的人更相似。研究不能把没有成交的人随意赋予零价格，再与成交均价相加；没有成交本身是另一类结果。应先比较进入率、等待与退出，再在各自成交者中比较价格及履约条件。若规则改变了成交者构成，成交均价的差异便同时包含选择与交易方式的作用。只有给出共同可比的人群范围，才有条件判断规则怎样改变了同类交易。

规则还决定谁有资格进入。设想一个交易系统要求参与者先缴纳保证金，另一个允许在较低保证金下参与但设置更严格的成交额度。两种制度可能形成相近的成交价，却吸引不同的人进入，排除不同的人离开。仅从价格稳定判断制度没有变化，就可能忽略参与机会的重组。相反，价格发生变化也不必意味着规则失效，它可能反映先前被排除的需求获得表达。

在三方程的语言中，新的显露迫使研究者检查差异怎样被排序，排序方式又依赖特征之间的规定关系。保证金、撤单权与履约责任不是三个可以独立添加的控制变量；改变撤

单权可能改变保证金需求，改变履约责任也可能改变报价的可信度。三方程提示这种相互规定，但不能替代具体的制度模型。研究者仍需写明行动集合、信息可见性和后果分配，才能从概念关系走向可比较的解释。

这一分析也使制度讨论摆脱单一结果评价。一个规则可能让成交更快，却使某些参与者承担更高的不确定性；也可能提高价格可比性，却降低特定需求的表达空间。评价哪种制度更好，需要明确目标和受影响者，而不能由价格是否接近某个平均值自动决定。发生研究在这里承担解释任务：它说明不同规则怎样产生不同可见结果，为规范选择提供更清楚的后果图景。

22.4 价格的信息作用与反向影响

价格能够传递信息，也能够改变行动。这两个方向经常在同一事件中相遇。一个买方看到价格上升，可能推测需求增加，调整采购；他的采购又可能参与推动价格。另一个买方则可能因价格上升而减少购买，形成相反作用。信息解释和约束作用未必同向，也不必具有相同时间尺度。把这种过程简单称为自反性，只说明存在研究入口，还没有说明反馈的方向与强度。

Grossman 与 Stiglitz 的模型已经将价格的信息含量与信息获取行为一起放入均衡分析。知情者的交易影响价格，不知情者从价格学习，获取信息的人数又取决于信息收益与成本。[45] 这一强近邻表明，市场理论并非只能把信息看作完全外来的输入。发生学若要提出额外问题，需要检查信息的身份如何在传播和行动中变化，或者既有模型的行动与更新规则是否遗漏了制度形成过程。

设想某种商品的价格突然上升。它可能反映供应减少，可能反映购买者集中行动，也可能来自交易规则变化后记录到的成交构成改变。观察者若把三者都解释为需求增长，其后续行动就可能进一步改变市场。这里有两个需要分开的因果问题：最初变化由什么产生，解释这一变化的行动又造成什么后果。后一个问题真实存在，并不使前一个问题失去意义。

同一条信息进入不同人的决策，也可能获得不同含义。对现有持有者，价格上涨意味着出售机会；对潜在进入者，它可能意味着更高成本；对依靠资产担保的人，它还可能改变可借款范围。这些机制不是心理分歧的简单加总，因为制度权利和资产位置决定价格如何进入行动。研究若只记录“看涨”或“看跌”，就可能遗漏具有相同预期的人为何采取相反行动。

价格反馈因此应被画成带有中介的路径：公开价格被何人看到，被解释为何种信息，通过哪项资源或权利改变行动，行动又如何进入下一次成交。每个箭头都有可能失败。价格可能早已被预期，参与者可能没有行动能力，行动也可能被相反方向的交易抵消。只有把这些可能性保留在模型中，才能解释为什么有时反馈迅速放大，有时只产生短暂调整，有时甚至没有可辨认后果。

这种研究也需要区分价格水平与价格变化的含义。一个长期稳定的高价可能已经进入生产和消费安排，而一次短期跳升主要触发等待或替代。把两者合并为同样的“价格增加”，会混淆适应与冲击。历史不是作为无限背景附在模型外面，而是通过已形成的合同、计划和参照进入当前事件。寻找这些承接关系，比笼统强调价格会影响价格更接近可检验的机制。

22.5 自反过程中的均衡与稳定

均衡并不等于世界停止变化。一个模型中的均衡可以表示行动彼此相容、预期与结果在特定意义上协调，也可以包含持续的随机变化。即使存在某种静态均衡，它是否会被实际过程达到，仍是另一个问题。价格形成、均衡存在、均衡唯一和动态稳定是不同判断，不能从其中一个词直接跳到其他结论。

一个最简单的反馈模型足以显示这种区分。假设下一轮价格由固定部分和当前价格的线性影响构成，反馈系数的绝对值小于一。在这个人为构造的模型中，当前价格确实进入下一轮价格，但偏离会逐轮缩小，过程趋向一个固定值。若系数改变，稳定性可能随之改变。这个例子不是对实际市场的证明；它只说明从存在回写直接推出必不收敛，在逻辑上缺少条件。

现实研究需要寻找这些条件的经济含义。价格上升可能诱发增加供应的行动，也可能因担保价值提高而扩大购买能力。前者在某些条件下抑制继续上升，后者在某些条件下放大上升。两种机制可以并存，其相对速度、容量限制和触发阈值会影响结果。研究不能只数有多少回路，还必须辨认每条回路怎样承担后果，以及它们能否在所观察的时间尺度内发挥作用。

预期也可以支持稳定。若参与者相信某种公开规则会被执行，并据此安排可相容的行动，这种相互承接可能减少不必要的反复调整。公开信息不是必然破坏规律的力量，它也可能成为协调条件。相反，若参与者对规则含义理解不同，或者执行持续偏离承诺，同样

的公开形式可能引起不稳定。发生学需要解释稳定预期怎样形成与维护，而不是只在预期落空时才进入现场。

稳定的价格序列也不必证明内部过程简单。不同方向的作用可能相互抵消，持续校准可能维持同一显露，交易规则也可能过滤某些变化。要区分这些机制，可以观察扰动后的恢复、某条路径暂时受限后的变化，以及相同价格下交易组成是否改变。若内部机制不同却暂时同价，新的干预可能使它们分离。若各种可行干预下仍不可区分，研究者就应承认当前证据只能支持较粗的描述。

由此至少需要区分三个时间窗口。撮合窗口内没有进一步成交，可能只意味着当前订单无法匹配；履约窗口内价格维持，仍可能伴随库存和未完成合同累积；更长窗口中的恢复，则取决于供给、替代与融资是否能够调整。研究若用第一个窗口的平静证明第三个窗口的稳定，便跳过了承接条件。反过来，短期价格波动也不能直接推翻长期协调。均衡模型在明确时间范围内组织相容条件，发生分析则检查这些条件如何被维持，以及在哪个窗口开始失配。二者由此具有可以合作也可以冲突的具体位置。

22.6 时间压力与交易之外的路径

交易中的等待能力常常影响行动。一个必须立即支付费用的人，与一个可以推迟交易的人，面对相同报价时可能具有不同选择。把这种差异记录为时间压力，可以提出可检验的问题；把它立即提升为所有市场价格的唯一根源，则会遮蔽其他机制。生产用途、偏好、产权和替代机会仍然需要具体分析。

可以比较两种明确的分析性安排。第一种安排允许参与者获得一段可靠的支付延期，第二种只提供关于价格走势的信息。前者改变可等待的时间，后者改变对未来的判断。若两者都改变报价，机制仍可能不同。延期待遇可能降低迫切成交的需要，信息则可能改变预期收益。通过分别改变时间约束和信息，可以检查原来被统称为焦虑的行为究竟由什么支持。

交易之外的替代路径也有多种形式。公共供给、共同使用、借用、家庭内部安排和赠与，能够改变某些市场交易的必要性，但各自形成不同责任与限制。共同使用需要协调时间，借用需要处理损坏责任，公共供给需要决定资格和资源分配。它们不是无条件自由的同义词，也不天然比市场路径更公平。发生分析应比较实际可用性，而不是只把是否标价当作优劣边界。

若一项公共服务扩大了可获得性，市场中的愿付价格可能改变，某些交易可能不再发生，另一些交易的质量要求也可能上升。即使平均价格没有变化，交易人数和退出选择也可能已经改变。研究因此应观察可行行动集合，而不只盯着一个价格。制度的意义可能体现在人能够拒绝什么、能够等待多久，以及能否从一种不利安排转向另一种安排。

另一方面，价格管制、补贴和支付规则也不能先验地被划为只改变显露。它们可能改变收益、资源约束和预期，进而重组交易过程。某项措施只在表面上留下短暂效果，还是实际改变了发生路径，需要由后果判断。理论若规定只有某几类干预才配称结构改变，就会在研究之前封闭发现空间。发生学更适合提出比较：哪些措施只是把原有压力转移到其他环节，哪些措施真正改变了参与者能够承担的选择。

这些问题把因果解释推进到规范评价的门口。让一个人拥有更多退出机会通常具有重要意义，但具体政策是否值得采用，还要考虑资源成本、他人权利与长期可持续性。描述价格由不平等机会形成，不能单独决定采取何种分配原则；说明某个制度可以改变价格，也不证明它符合所选择的价值。后续伦理章节需要承担这些规范论证，而经济发生分析应提供足够清楚的机制和后果。

22.7 价格客观性的实践基础

如果价格由参与者的判断和交易规则形成，研究者还能否对它作出客观判断？这个问题的答案取决于客观性被要求做什么。如果要求价格必须完全独立于一切人的行动，那许多社会对象从定义上就被排除；如果要求记录可核查、概念可比较、推断能够接受共同检验，那么发生过程反而提供了建设客观性的材料。

一项可靠的价格记录需要确定商品身份、数量单位、时间、支付方式和履约条件。不同市场之间比较时，还需检查费用、质量、税费和权利是否一致。这些工作并没有把价格变成主观判断，相反，它们使比较能够承受追问。客观性在这里表现为对差异的公开处理，而不是假装差异不存在。一个没有标明条件的精确数字，可能比一个范围明确的区间更难被正确使用。

回到九十五元的交易，可以进一步比较当场付款与约定以后付款。只要付款时间不同，卖方承担的等待、违约可能和资金安排就不同。若后付款伴随额外担保，担保又改变了同一付款期限的含义。比较者可以为研究目的把这些条件尽量标准化，也可以保留它们来研究信用怎样进入价格；两种做法分别回答不同问题。重要的不是把所有条件折算成一

个自称绝对真实的价格，而是说明哪些条件已被调整、哪些仍保留在比较中。同一金额因而可以是真实记录，同时尚不足以构成可比价格。

研究者自身的分类也可能回写市场。若一个公开分类被消费者或生产者采用，商品设计和购买行为可能开始围绕它组织。此时，分类既帮助观察，也成为行动条件。要判断这一作用，必须说明分类何时公开、谁实际采用、通过何种后果发生影响。仅仅发现分类与行为相关，还不能排除分类原本就在追踪已有变化。认识与对象之间的双向关系需要更多辨识工作，而不是更少。

价格客观性的另一面是可追溯责任。当某个交易数值被用于合同、税收或赔偿，参与者需要知道谁有权修改记录、错误如何纠正、争议如何处理。这些制度安排并非外在于价格的纯粹价值表达。它们决定记录能否在跨时间行动中被信赖。稳定的公共价格，往往需要持续维护记录与执行，而不只是一次即时撮合。

这使客体性侧重获得一种具体解释：价格可以在后续行动中作为相对独立的对象被引用，因为形成它的关键条件已经被封装、保存并获得制度承认。这种独立性是有限的，却是真实有效的。当封装中的条件发生变化，研究者需要重新打开它；当条件保持相容，就不必每次交易都重述全部发生史。发生学因此能够解释对象化为什么必要，以及对象化什么时候需要被修订。

22.8 从成交记录到价格研究

以发生为单位研究价格，并不是要求每一个项目都建立一个包罗万象的社会模型。一个可执行的研究应从具体失配或新问题出发。例如，相同报价在两种规则下产生了不同成交率，或者相同成交价下的退出行为发生改变。问题足够明确，才可能选择需要保留的事件环节，并控制记录的负担。

研究可以先重建一次交易的最短链条：行动者能够看到什么，能够采取什么行动，何时必须承诺，结果如何被执行。随后提出至少两种不同机制。若差异来自信息解释，改变信息内容或可见性应有相应后果；若差异来自时间约束，单纯增加信息未必有效；若差异来自履约风险，调整责任安排可能比改变报价程序更关键。每种解释都应对某项可行变化作出不同判断。

例如，若两套规则中的成交均价都是九十五元，一套安排只让少数需求急迫者成交，另一套允许更多人以较短等待完成交易，研究结论应首先指向成交机会与时间成本的区别。若两套规则的参与者和成交比例接近，但一套包含更可靠的退货承诺，则需要转向质

量风险与权利的区别。价格数字没有变化，并未使两种机制等同；但若研究目的仅是核算当日支付金额，这些区别又可能无需全部纳入。研究入口由用途选择，解释的终点则由事件中真正改变的承诺来约束。

一个重要的成功结果可能是证明简单模型足够。若经过质量、权利和时间调整后，价格差异被稳定解释，研究就获得了可继承的压缩方法。另一个成功结果可能是发现原来被混称为同一种价格的记录，其实来自不同交易事件，需要重新分类。还有一种结果是识别制度反馈：某个公开价格通过信用、计划或退出机会改变后续行动。这三种成果价值不同，不应被压成是否证成整个发生本体论的单一考核。

上述比较也给出明确的否定条件。若所谓履约差异在合同和实际执行中均未出现，就不能用它解释价格；若等待缩短只来自商品库存增加，排序规则便不是已获支持的原因；只有在参与者范围、成交率及成交者在价格相关特征上的分布均保持可比时，选择构成解释才受到相应限制；未成交者性质相同本身不足以排除选择效应。否定这些具体通道以后，研究仍可以保留真实的成交差异，但应另找机制。事件分析的进展，正表现为把一个含糊的“制度影响价格”收紧为某种承诺经何种操作发生改变。

跨市场比较还需要防止一种过早的同一化。食品、住房和生产设备即使都用货币报价，交易频率、检验质量的可能性与退出成本也不相同。一项机制在短期标准化交易中成立，并不保证在长期合同中保持同样作用。比较的共同单位应是被明确界定的行动问题，而不是抽象的价格数字。能够指出哪些条件可保留、哪些条件必须重建，才构成真正的迁移知识。

成交价把已经接受的交换条件固定下来，公开预测则在成交以前提出关于未来的判断。两者都可能影响行动，但预测还要面对一种特殊责任：它所引发的行动是否改变了自己原来声称的未来。第 23 章从这一差别出发，将价格作为既成交易记录的作用，与预测作为公共陈述的作用分开，再检查它们怎样在市场中相接。

第 23 章 股市信息与公开预测的回写

23.1 预测公布以后发生了什么

一份报告预测某类资产的需求将增加。有人据此买入，有人怀疑报告发布者的动机而卖出，有人虽然相信判断，却因资金限制无法行动。价格随后发生变化，研究者不能仅凭方向一致就宣布预测得到验证，也不能仅凭方向相反就宣布预测因公开而自我消灭。报告、解释、行动和结果之间存在若干需要分别检查的环节。

公开预测至少承担两种角色。作为认识陈述，它对未来事件提出可以追究的判断；作为公共行动，它向潜在参与者提供信息，并可能改变他们彼此之间的预期。一份报告还可能成为合同、风险限制或政策讨论的依据，获得超出最初传播范围的制度效力。认识角色与行动角色并不互相排斥，但评价它们需要不同证据。预测改变了结果，不等于其原始判断准确；预测准确，也不意味着公开它没有产生作用。

必须先确定预测内容。预测某公司未来收入增长，与预测其股票价格上涨，不是同一命题；预测短期价格变化，与预测长期收益分布，也不是同一任务。若报告只给出条件性情景，使用者却把它当成无条件承诺，后续失配可能首先来自解释转换。发生研究应保存原始命题、适用条件与传播后的版本，才能区分预测本身的失败和命题在采用过程中的变化。

同样需要确定被公开的是什么。报告可能包含一个新数据、一套分析方法、对已有材料的新解释，或对某种行动的建议。新数据的价格反应与行动建议的采用，不应被混为一种信息效应。前者可能改变对现金流的估计，后者可能改变持仓和交易安排；方法公开则可能让更多人采取相似的判断程序。三者都能进入市场，但形成的反馈链条不同。

公开预测进入现实，需要区分两条经常重叠的路径。一条从新证据出发：报告使投资者修正对企业原有业务的判断，交易据此调整，价格重新表达前景。另一条从采用出发：报告改变交易或融资条件，这些变化进一步改变企业能够实施的项目，未来业务本身随之变化。两条路径都可能产生同方向的股价反应；仅凭发布后上涨，无法判定是哪一种。这里的新增研究对象是从价格反应走向项目和产出改变的承接，而不是重复第 22 章关于价格形成的一般分析。

23.2 可预测收益与无风险机会的区别

市场效率的讨论常被压缩为一句话：价格应当随机游走，因此凡是可预测的模式都是理论失败。这种表述省略了资产定价模型的作用。一个有风险的资产可以具有可预测的预期收益，收益也可能随经济条件变化；这与存在无需承担相应风险和成本的确定获利机会不同。研究者若不区分两者，就可能把补偿风险的规律误当成任何人都可以无成本利用的漏洞。

Fama 对有效市场研究的总结明确讨论了预期收益的时变性与可预测性，也强调效率检验与所采用的均衡定价模型相联系。[46] 因而，发生学不能把一个最强对手替换成“任何可预测变化都必须立刻消失”的简化命题。更有实质的问题是：当预测被采用后，承担风险的人、信息成本和交易约束是否改变，原来估计的关系是否仍适用于新的行动结构。

设想一种分析性情境。某项策略在一段时期内具有正的平均收益，同时承担偶发而较大的损失。策略公开后，更多人采用它，平均收益降低。这可以来自竞争，也可以来自风险被重新分配；不能仅凭收益下降就断言市场已经消灭了全部规律。若仍有一部分预期收益补偿承担的风险，预测可能继续有效，但其可实现收益和容量已经改变。认识命题与实施方案需要分别更新。

相反，一项模式在公开后增强，也不能直接证明一种特殊本体论。参与者的相同行动可能短期推高价格，随后又因为持仓集中而增加反向调整的压力。若研究只选择增强阶段，就会把暂时反馈写成永久机制。观察窗口应与所声称的回写过程匹配，既记录进入，也记录退出、撤销和后果承接。持续性需要证据，而不能由最显眼的一段曲线代替。

交易成本和执行条件也不是可以在理论上忽略、在失败后才补上的借口。公开预测涉及实际采用，采用就必须经过资金、价格和成交机会。一个纸面上可以预测的变化，不保证报告发布后仍能按原价完成交易。研究应区分发布前已经持有、发布时试图进入和发布后才能行动的人。不同位置获得不同后果，正是公开预测形成新差异的一个来源。

这一讨论不会把市场效率变成不可反驳的万能解释。相反，它要求明确比较对象：风险如何度量，成本如何计算，信息何时可得，策略如何执行，以及预测是否在未参与发现的数据上成立。发生学可以在这些条件之外提出新的更新机制，但必须承担同样具体的检验责任。把不能立即解释的价格变化全部归入发生，和把它们全部归入噪声一样，都可能停止研究。

23.3 公开预测的三种后果

公开预测可能产生抵消、强化和协调三类后果，但这只是候选机制的区分，并非所有市场必然落入其中一种纯粹类型。抵消意味着人们针对预测采取行动，使原先预期的差异缩小；强化意味着采用过程扩大了原先预期的变化；协调意味着公开陈述帮助行动彼此相容，使某种可持续状态得以形成。三者还可能在同一事件的不同时段交替出现。

抵消可以通过一个简单的假设例子说明。如果人们预期某种商品下期短缺，生产者提前增加供给，消费者提前寻找替代品，最终短缺程度可能小于原先预测。原预测因行动而未完全实现，但这种未实现并不自动意味着预测毫无认识价值。它可能准确指出了在没有响应时会发生的风险。要评价它，需要明确原命题是对无干预路径的预测，还是对实际响应后结果的预测。二者不可事后任意切换。

强化也需要具体路径。一个关于价格上涨的报告若导致更多购买，而购买又改变可用数量或融资条件，价格可能进一步上涨。可是购买行为并非唯一反应：已有持有者可能出售，生产者可能扩大供给，风险管理者可能收紧敞口。强化是否占优势，取决于这些作用的相对速度和约束。研究不应把一个方向的箭头画完后，便假定其他路径不存在。

协调的可能性尤其重要。设想一份明确标注条件的产能报告：若一定数量的购买者提前确认订单，供应者便能按报告所列成本组织生产，并在约定时间交付。购买者此前可能因为担心无法交付而不愿承诺，供应者也可能因为订单不足而不愿投入；当双方理解并接受同一条件安排，报告所述交付路径才可能实现。这个构造例子中的公开判断没有自动消灭自身，而是帮助相容行动形成。不过，协调证据必须包括实际订单、投入与交付，不能只看各方表达了信心。它说明采用可以促成一种条件预测的实现，不说明任意乐观预言都能创造资源。

三类机制必须通过中介和反事实区分。仅观察结果方向不够，因为同样的上涨既可能来自新信息，也可能来自采用反馈，还可能来自共同冲击。可以观察报告是否改变信念、信念是否改变订单、订单是否在所研究的制度下影响价格。如果某个环节没有发生，就应缩小回写主张。若多个机制同时存在，则应报告无法完全分离的范围，而不是强行分配为独立份额。

这一框架也保护预测评价的诚实性。预测者不能在结果相符时称自己准确，在结果不符时自动称自己成功阻止了事件。后一主张同样需要响应链条的证据。应在发布时说明预

测是否考虑采用效应，以及哪些观察可以区分准确预测、行动改变和普通错误。公开预测的责任由此不仅包括说出未来，还包括说明自己的话以何种方式可能进入未来。

23.4 价格怎样改变它所反映的对象

股价通常被用来理解企业前景，但股价也可能影响企业获得资源、面对的收购压力和决策机会。这里出现一个比价格影响下一笔交易更深的回路：市场判断可能改变被判断对象后来的活动。若价格能够影响企业行为，就不能把所有未来现金流都预先当成与市场过程完全无关的外部给定量。然而，这也不意味着现金流分析失去意义，而是要求说明现金流在什么机制下形成。

Edmans、Goldstein 与 Jiang 对价格和收购的研究，区分了市场价格影响收购活动的触发作用，以及收购预期反过来影响价格的作用，并使用工具变量方法尝试识别前者。[47] 这项研究的重要性不在于替整个市场提供一个统一答案，而在于展示研究价格反馈时必须处理的方向问题。价格与企业行动共同变化，既可能是价格造成行动，也可能是市场提前反映行动，或两者受到共同因素影响。

发生学可以沿着这一问题进一步追踪对象身份。一个企业在报告发布前后可能仍有同样名称、设备和法律主体，但它面对的融资条件、管理目标和可行投资已经改变。如果研究者只把企业看成一个固定容器，再把市场价格放在外部，就可能遗漏市场判断怎样进入企业后续能力。不过，改变能力的具体中介仍需证明。企业名称保持不变不能证明机制不变，价格变化也不能自动证明企业能力已经改变。

可以构造两个发布时股价反应相同的企业情境。甲企业的报告只是揭示已签订合同，投资者据此提高估值，生产计划保持不变；乙企业此前有一个尚未落实资金的项目，估值变化经实际融资安排改变可用资金，项目才开始实施。两者都可能出现报告、价格与未来收入同向变化，只有乙情境额外提出价格参与改变收入路径的主张。要区分二者，必须保存发布前已有合同和计划、融资实际完成的时间、项目启动条件以及资金用途。若项目早已具备执行条件，不能把后来融资与收入的相关性全部写成预测的创造作用。若无法观察这些材料，研究所能确认的就仍主要是信息反应。

这里还存在被评价者对评价的预期。管理者可能在报告出现之前就调整披露和行动，以影响未来评价。于是研究不宜只把发布日当作反馈的唯一入口。采用可能提前发生，公开只是其中一个节点。事件窗口需要围绕机制选择，而不能机械地把所有变化压进同样长度的前后比较。即时价格反应与长期组织重组，应当用不同资料和识别方式研究。

还应区分对象的不同层次。股东之间的转手直接改变持有关系，却未必立即改变企业的设备或项目；融资交易则可能改变企业可用资源；管理层把估值采用为决策依据，又是另一条组织通道。这些层次不能因都与股价有关而混成一个过程。即使价格没有改变当前生产，它也可能改变风险承担者和控制权；这种变化是否具有长期后果，仍需进一步追踪。把作用分层，有助于防止用即时财富变动代替对真实活动变化的证明。

把价格理解为发生事件，具有如下研究含义：价格不仅是关于既定对象的信号，还可能参与对象未来条件的形成。但它仍然不是一项已经完成的统一经验定律。不同企业、资产和制度中，反馈通道可能强弱不同，甚至不存在。保留这种差异，才能解释什么场景适合把价格近似为信息读数，什么场景需要把市场与实体活动放在同一动态系统中。

23.5 速度与市场安排

当信息接收、分析和交易能够在很短时间内完成，研究者容易把速度变化描述为经济对象性的根本消失。更精确的问题是：时间优势在什么规则下获得收益，它使谁来不及调整，哪些行动因此被优先完成。速度是相对某个操作和制度而言的，不是一个脱离任务的绝对性质；更快也不等于所有参与者在同一时刻行动，更不等于时差真正归零。

Budish、Cramton 与 Shim 把高频交易的速度竞争作为市场设计问题研究，提出以频繁批量竞价替代连续逐笔处理的制度方案。其核心比较涉及离散处理时间和统一批次中的价格竞争。[48] 这为发生学提供一个直接近邻：改变时间安排本身能够改变行动收益与竞争方式。相关主张不需要先宣布客观经济规律已经消失，而可以在明确机制中接受分析和检验。

时间安排改变的并不只是等待长短。若订单按到达先后逐一处理，微小时间差可能决定谁首先完成交易；若订单在一个短窗口内共同处理，参与者面对的排序问题就不同。这里的差异序列由制度组织，而非单纯由自然时钟记录。同样的参与者和信息，在不同排序规则下可能形成不同结果。研究需要检查这种差异究竟来自竞争方式、信息泄露还是参与构成的改变。

制度设计也有取舍。延长一个处理窗口，可能减少某些速度优势，却改变对紧急交易的响应；统一批次可能改善某种比较，也可能要求参与者调整风险管理。不能从一个机制收益直接推出所有市场都应采用同一规则。发生研究的工作是明确哪些条件支持收益、哪些条件形成代价，并识别适用范围。评价制度需要同时观察成交质量、参与机会和承诺执行，而不只观察某个速度指标。

这使时间从背景转为可研究的安排。信息何时公开、订单何时可撤、承诺何时生效、风险何时结算，构成不同的时间结构。把它们全部压成一个延迟参数，可能在某些任务中足够，在另一些任务中却掩盖关键差异。尤其当规则允许参与者在看到部分信息后修改行动时，时间结构与信息结构便不能分开处理。

速度还会改变证据归属。在一个简化发布事件中，价格先变、部分投资者随后收到报告，不能把这些后到者的交易当成初始变化的原因；早到者也可能只是更快获得了共同消息，而不是改变了企业活动。较粗汇总会把这些先后压在同一个时点，较细记录能够恢复次序，却仍需区分新消息与策略采用。若研究目标是微小时间优势，订单和撤单的时点重要；若目标是项目融资，逐笔报价无法替代资金与决策记录。时间分辨率因此应服从所研究的因果通道，而不是服从系统能生成多少日志。

23.6 发布事件的因果辨识

公开预测与数据发布为研究提供了清晰时间点，但清晰时间点并不自动等于清晰因果。价格可能在发布前已经响应泄露信息或先行信号，发布时也可能同时发生其他消息。更重要的是，发布内容本身常常反映经济变化。研究者必须区分市场响应新信息，与市场因为采用某种预测而改变后续对象，这两类问题需要不同的证据。

首先要区分公布值与未预期部分。若市场早已准确预料某个数据，公布本身可能没有多少新增信息。观察到很小反应，不能据此断言该指标没有经济意义；它也可能说明有关信息已经通过其他渠道进入价格。相反，较大反应可以说明新信息受到重视，却仍不能独自证明信息改变了后续生产或制度。即时信息更新与跨期现实反馈应当分别估计。

其次要保存实时版本，并把两个误差分开。报告对后来结果的偏离，关系到预测质量；市场在发布时收到的数值相对于此前可得预期的差异，关系到当时新增了什么信息。事后修订值可以用于重新评价结果，却不能无说明地替换当时公布值来计算发布意外。若资料只有最终版本，研究可能仍能分析长期关系，但应收缩对发布时信息冲击的判断。这样处理，数据修订便不再只是统计清理步骤，而成为预测者、接收者和事后研究者处在不同知识时点的明确证据。

再次要检查采用程度。被公开不等于被阅读，被阅读不等于被相信，被相信也不等于有能力行动。可以使用公开传播记录、调查、订单变化或制度文件等不同材料，但每种材料只支持相应层次。阅读量不能直接充当信念变化，订单相关也不能直接证明某份报告造

成交易。指标之间若存在缺口，应明确承认，并设计下一步资料获取，而不是用一条连贯叙述填满。

如果要识别采用的影响，可以寻找在内容相近时传播范围不同、在传播相近时行动机会不同，或者规则改变了采用成本的比较。不过，这些比较仍需说明分组是否受到其他因素影响。一个更有资源的群体既可能更早接触报告，也可能采取不同策略；不能把后续差异全部归于接收时间。研究设计应围绕具体的替代解释，而不是把某种方法名称当作因果保证。

对于未观察到响应的结果，也应报告检测能力。样本过少或窗口过短，可能无法辨认有实际意义的作用；统计上未显著并不等于通道不存在。研究可以预先给出值得关心的最小影响范围，并用区间说明资料能够排除什么。若已经能够排除这一范围内的重要效果，就应接受在该场景下反馈较弱的判断，而不能用框架的普遍性继续维护所有具体预测。这种有限否定本身能够帮助划定发生机制的适用边界。

最后，长期结果需要与短期响应连接。若一份预测引起价格变动，而价格变动又被认为影响投资，那么应分别提供两段关系的证据，并说明跨段连接是否可信。短期事件研究不能直接承担多年组织变化的全部证明责任。发生学的优势在于要求完整承接，风险也在于容易用宏大叙述跳过承接；只有坚持逐环辨识，才能把自反性从口号推进为具体知识。

23.7 预测者进入评价对象

传统预测评价往往问，预测值与后来结果相差多少。这个问题仍然必要，但当预测能够被采用时，还需要问结果是否已经包含发布行为的影响。一个制度如果依据预测提前干预，预测误差的解释就会改变。研究者不能简单奖励预测与结果相近的人，而惩罚成功促成调整的人；也不能允许任何预测者用可能存在的干预解释为所有错误免责。

可以建立两种事先区分的预测任务。第一种预测实际制度运行下的结果，要求把可预期响应纳入模型；第二种预测在某项响应尚未实施时的条件路径，要求明确假设。两种任务都可以有价值，但不能事后根据结果随意改换。若报告中同时包含两种预测，应分别记录并分别评价，使政策用途与科学判断能够对齐。

预测者还可能具有行动利益。一个持仓者发布判断，与一个负责公共统计的机构发布数据，承担的角色不同。利益存在并不自动使判断错误，但会改变需要核查的环节，例如信息选择、表达方式和行动时点。发生研究应把这些角色纳入事件说明，同时避免以动机

猜测代替证据。可以检查的传播和行动记录，比对发布者内心的推断更适合作为分析基础。

当预测成为考核指标，预测者可能调整行为以适应评价。若只惩罚偏离结果的程度，他们可能倾向于接近共同预期；若只奖励显著预警，又可能增加极端表述。这里描述的是有待具体检验的激励机制，不能被写成所有预测机构的事实。它提示，评价规则可能改变可见预测分布，而分布变化未必全部来自知识进步或退步。

因此，预测质量可以从准确性、条件表达、校准、可用性和后果说明等不同维度评价。各维度之间存在取舍，不能无解释地合成一个总分。对高影响预测，尤其需要说明哪些条件尚不确定、采取什么行动会改变预测路径，以及哪些结果意味着模型应被修订。这并不是给预测附加无穷免责条款，而是使它成为能够被正确采用的公共判断。

在 SDE 方法中，预测者进入对象不意味着研究者失去批判距离。相反，它要求把自己的作用作为可以研究的条件：使用哪些信息，选择什么表达，怎样安排公开，如何追踪后果。主体性侧重由此不再只是承认每个人都有看法，而是承担判断如何参与后续发生的责任。客体性也没有消失，因为公开命题、采用记录和实际结果仍然能够成为共同核查的材料。

23.8 公开预测的研究档案

公开预测的证据可以分成两个窗口组织。发布窗口回答新增信息是否进入价格，采用窗口回答交易或制度响应是否改变了被预测对象的条件。第一个窗口需要原始报告、事前信息和可执行交易时点；第二个窗口需要实际采用、资源变动和后续活动。窗口长短应由机制决定：融资合同可能迟于价格调整，内部决策也可能在报告正式发表之前开始。把两类证据并列，能够确认某次研究究竟发现了价格中的知识更新，还是发现了预测经采用进入现实的更长回写。

比较应先围绕两条已经展开的路径展开。仅有信息重新估价的模型，可以预测发布后的价格调整，却不要求企业改变原有项目；允许融资和组织响应的模型，则还要预测哪些企业会改变投入、改变要等待什么条件。二者应使用发布前可得的同等资料，不能让反馈模型提前看到后来项目结果。若融资约束较弱的企业也呈现同样项目变化，就需要检查共同需求冲击；若只有事前受相应约束、且确有资金通道改变的企业响应，反馈解释才得到更有针对性的支持。本书的发生分析应以这种区别性承诺接受检验，而不是另设一个名称即宣称已经超越已有反馈模型。

失败条件也应具体。若所谓采用并未发生，相关反馈通道就缺乏支持；若中介变化先于发布，则因果顺序需要重建；若新增机制不能改善留出情境中的判断，就应降低其使用范围。反过来，预测准确也不是唯一成功形式。识别一个不能被现有资料区分的机制集合，或者设计能够区分它们的新发布安排，同样构成研究进展。

公开预测研究最终需要同时保护两种能力：对未来提出可以失败的判断，以及对判断进入现实的方式保持反思。只保留前者，可能把制度参与隐藏在误差项中；只保留后者，则可能使任何失败都被解释为反身性的胜利。两者的结合要求研究者在发布之前确定命题，在发布之后追踪行动，并在证据不足时保留未决。

由价格、信息和预测形成的回路并不证明经济世界没有规律。它说明规律可能需要以制度、信息和行动规则为条件表达，稳定也可能由持续协调而产生。更值得研究的区别，是哪些关系能够在规则改变后保持，哪些关系必须重新形成。发生学在此提供一种开放任务：既寻找可继承的结构，也识别结构的生成与退出条件。

公开预测的特殊问题，是陈述未来可能参与改变未来。国内生产总值则先要解决另一个问题：把已经发生的异质生产按什么边界和规则汇成总量。它的发布也可能被采用，但采用之前已有一套不能还原为市场预测的核算工作。第 24 章据此转向活动身份、估价与汇总，考察一个公共量怎样在保留可比性的同时承担有限的制度用途。

第 24 章 GDP 的核算边界与制度作用

24.1 一个公共总量怎样形成

一个国家在某个季度生产了什么，不能像清点同一种物品那样直接相加。食品、运输、软件和医疗服务具有不同用途、数量单位与交付方式。要形成国内生产总值，就必须先界定哪些活动属于生产，再按照相应估价规则把它们纳入共同核算。这个总量因此既依赖真实活动，也依赖清楚的定义、资料与编制程序。它并不是等待被挖掘的独立实体，却仍然可以成为对特定经济活动的有效测度。

核算的构造性与测量的可靠性不相互排斥。边界确定哪些产出有核算资格，估价把它们转换为可汇总的表达，账户关系则约束汇总中哪些项目应当抵消。美国经济分析局把 GDP 说明为境内生产的最终产品和服务的价值，并以生产价值扣除生产中耗用的产品和服务来说明增加值。[49] 因而，市场中的每一笔付款并不都能作为一项新增产出直接相加。核算遵守这些关系，才有条件检查重复、遗漏和资料不一致；构造性在这里意味着公开的操作约束，而非任意编写数字的自由。

本章关注三个相互连接的问题。首先，哪些发生被纳入总量，哪些被排除，估价怎样处理没有直接成交的活动。其次，总量与社会生活的关系如何被解释，何种问题可以由它回答，何种问题需要另外的指标。最后，当总量进入政策、合同和公共判断，它是否改变了后续行动。三者分别涉及核算、认识和制度作用，不能仅以“数字是建构的”一句话替代。

GDP 也不是一种旨在跨时间保持不变的普遍常数。它被编制出来，本来就用于描述一定时期内的经济活动及其变化。若把年度增长、统计修订和核算口径调整一概视为常数不收敛，就会误认对象。真正值得研究的是，某种用途是否把有边界的总量当成完整社会价值，或把旧口径下的关系无条件延伸到新制度。病理可能存在于这种使用，而不是存在于总量会变化这一事实。

生产身份、用途与估价在核算中相互规定。同一项产品被用于继续生产还是最终使用，会影响它在聚合中的位置；一项住房服务没有当期租金成交，也不因此失去核算资格。这里的特征纠缠体现为分类改变了金额能承担的核算含义，估价规则又限定资料能够怎样被使用。显露的公共总量由这些区别经过记录、汇总与修订而形成。它既不是所有市场价格的堆积，也不是统计部门、企业和政府三个实体的简单合作产物。

24.2 生产边界与非市场活动

没有直接市场价格，不意味着一项活动必然被排除在 GDP 之外。美国经济分析局关于生产边界的说明，明确包括部分非市场活动，例如政府和非营利机构提供的产品与服务、业主自住住房的住房服务，以及部分为自身最终使用生产的商品和固定资产；另一些无偿家庭劳动则处于边界之外。[50] 因此，市场成交、可推算价值和被核算资格，是三个需要区分的问题。

英国国民账户的官方说明也把政府服务、自住住房服务、自用农产品和家庭自建等列入生产边界，同时把同一家庭内部生产并消费的做饭、清洁和照护等服务列在边界之外，并说明在适当情况下可以进行推算估价。[51] 这些规定使关于不可见劳动的提问获得更准确的对象。问题不在于所有无偿活动都不可见，而在于边界如何区分不同活动，以及这种区分在公共使用中产生什么后果。

设想一个家庭原先由成员无偿完成日常照护，后来改为购买同类服务。先在思想实验中固定照护时数与质量，只改变谁以什么方式提供服务，就能看见核算资格与生活任务之间的差别：照护并非从支付出现那一刻才存在，其进入中央账户的方式却改变了。再放开固定条件，家庭成员可能把腾出的时间用于工作、休息或其他照护，服务提供者也可能调整原有活动。此时总量的净变化要连同这些替代一并判断，不能机械地把新增支付全部当作社会新增能力。这个案例既不否认市场照护的真实生产，也不把无偿劳动降为零价值；它要求在任务连续、组织转换与账户记录之间建立可核查的对应。

扩大边界也不是无成本的解决。若把所有家庭服务纳入总量，就需要决定时间怎样记录、同时完成的活动如何分配、质量如何区分、替代价格怎样选择。不同估价方法可能产生不同总量，而方法之间的选择又涉及用途。若目标是研究照护负担，时间分配可能比货币总额更直接；若目标是比较市场化与家庭承担的转换，则可能需要保持可比较的任务单位。指标应围绕问题形成，而不是为了把一切都塞进同一个数。

这并不削弱对边界的规范追问。某种活动长期不进入最受关注的总量，可能影响公众对贡献的理解，尤其当政策把该总量当作唯一成就尺度时。但这种影响需要通过实际采用来证明。统计边界本身与政治评价之间存在一条制度路径，不能直接等同。研究者应追踪哪些决策使用了哪些数字，是否因此忽略了某类成本或贡献，以及替代信息能否改变判断。

发生学在这里提出的建设性方向，是让边界变得可见、可讨论和可连接。中央核算可以保留相对稳定的定义，同时通过附加账户、时间使用资料和专题指标呈现其他活动。不同记录不必都被还原成同一价值尺度。它们可以围绕共同事件建立连接，使使用者知道一个活动在某种核算中为何出现、在另一种记录中为何以不同形式显露。

24.3 估价与聚合中的对象变化

生产边界确定以后，异质活动仍然需要被聚合。货币表达为许多活动提供共同计量方式，但共同表达不意味着它们具有相同社会意义，也不意味着价格变化就是数量变化。名义总量的增加可能包含产出增加和价格变化；要讨论实际活动的变化，需要相应的价格调整。核算体系必须在价格与数量之间建立可操作的区分，而这一区分依赖具体产品与资料。

先看一个不计税费和进出口的封闭核算例子。农场没有购入中间投入，以四十元出售全部原料；加工者用这些原料制成产品，以一百元全部卖给最终消费者。两笔成交金额相加是一百四十元，但最终产品已经包含原料，按增加值汇总时农场贡献四十元、加工者贡献六十元，总和是一百元。这个人为简化例子说明：核算不是把真实交易删成不真实数字，而是按用途关系避免重复计数。若忽略交易之间的生产联系，仅把可见付款全部堆积，得到的反而不是所定义的产出总量。

再保持产品数量不变而提高价格，现价总量可以增加；保持价格不变而提高产量，也可以形成同样的名义增长。两种变化对资源使用和生活的含义不同，因而还需相应的价格调整。这里出现两道不同压缩：先在产业联系中确定增加值，再在时间比较中区分价格和数量。前一道正确不能替后一道作保证，研究应说明当前困难究竟位于哪一层。

产品身份的变化会增加困难。某种软件更新后提供了新的功能，某项医疗服务改变了实施方式，某种交通服务提高了可达性，简单数量比较可能不足以描述变化。价格差异中有多少对应质量，哪些新功能应被视为同一产品的改进，哪些需要重新分类，都是具体测量任务。把这些困难概括为经济无法测量，既忽略了已有方法，也放弃了改进测量的机会。

可以用一个明确假设的服务案例说明。原来需要两次往返完成的行政服务，现在可以一次完成，服务收费保持不变。若只记录收入，改进可能不直接表现为更多市场产出；若记录办理数量，也未必捕捉节约的时间。研究生活便利性时需要其他资料，但这并不意味

着收费记录错误。不同指标分别回答不同问题，关键是不要让一个任务的成功测量替另一任务作出结论。

聚合还会隐藏构成变化。相同总量可以来自不同产业、地区和群体的活动，增长也可能集中在少数部门。若研究者只问总量是否上升，就无法回答谁的收入改善、谁承担了更多时间成本。这里的遗漏不一定要求修改 GDP 定义，而可能要求把总量与分布资料相连接。中央指标的简洁性与分项资料的解释力，可以构成互补关系。

因此，聚合应被理解为有用途的压缩。它使大量活动进入共同宏观判断，也必然舍弃某些差异。发生学不能以舍弃差异为由一概否定聚合，因为任何有限表达都需要选择。它应检查舍弃的差异是否影响当前判断：若不影响，压缩具有价值；若影响，就应补充记录或改变用途。真正的问题是压缩后形成的数，是否被赋予了超出其信息范围的权威。

24.4 统计发布与历史版本

某个季度已经结束，不意味着关于它的全部资料已经到齐。GDP 的早期估计、后续修订与方法更新，可能形成多个版本。版本变化有时来自新增资料，有时来自估计方法和分类调整。研究者如果把它们混成经济活动本身的变化，就会在认识过程与被认识过程之间失去区分。

实时研究需要知道当时发布了什么，而不是只知道今天如何描述过去。ALFRED 等历史数据服务以保存和查询不同发布时点的数据版本为用途，为这种研究提供了基础。[25] 版本资料的重要性在于恢复行动者当时的信息条件：一项政策若依据初值作出，不能用后来修订的数字替它重写决策现场。事后评估可以利用更多信息，但必须说明这种信息差异。

设想一项政策在初步数据表明增长放缓时启动，后来统计修订显示当时放缓程度较小。至少存在三个不同问题：初步估计为什么偏离，政策在当时信息下是否合理，政策实施后产生什么后果。即使后来知道初值不准确，也不能直接推出当时决策不合理；即使决策符合当时资料，也不能免除对后果的评价。研究需要保留不同时间点的知识状态。

方法更新又引入另一层变化。若某类活动被重新分类或得到更好记录，历史序列可能随之调整。总量变化不一定代表过去真实生产被改变，而可能代表我们对过去的核算发生变化。与此同时，新口径若被社会采用，又可能改变未来行动。这形成一个需要谨慎拆解的关系：重新描述过去与改变未来，是不同过程，可以相互连接，却不能被混称为数据自己改变了历史。

发布本身的作用也取决于新增信息。若公众已经从其他资料推知某种变化，正式发布可能只确认已有判断；若发布包含较大意外，市场和政策讨论可能更明显调整。观察不到即时反应，不等于指标没有意义；观察到反应，也不证明 GDP 已经通过政策改变下一期生产。后一主张需要继续追踪采用者、决策和实际活动。

同一核算期因此可以有两种必要的重建。评价初次决策时，应恢复初值、当时已知的资料不足及实际采用规则；评价测度改善时，则应比较后续资料如何改变原估计。二者若分离，便可出现“早期决策在当时有理由，但初值仍需修正”的判断，也可出现“后来总量更准确，使用部门却仍沿用旧值”的失配。制度回写的位置就在统计版本进入决策版本的转换处。修订公告并不保证全部合同、预算或评价已经同步更新，研究需要查明哪一版数值在何时真正取得行动效力。

24.5 指标进入制度的通道

GDP 可以作为宏观描述，也可以成为政策讨论、资源安排和公共评价的依据。后一作用不是由指标名称自动产生，而是通过制度采用形成。某个部门关注 GDP，不意味着它的每次行动都由 GDP 决定；某项增长目标存在，也不意味着所有参与者以相同方式理解和执行。研究需要从具体决策规则开始，而不是从总量的公共知名度直接推导因果力量。

可以构造三种不同采用方式。第一种把指标作为判断经济状态的众多信息之一，第二种把某个阈值写入明确规则，第三种把增长表现用于组织考核。第一种主要通过信念更新影响行动，第二种可能触发离散措施，第三种可能改变长期资源配置与报告行为。三种制度都使用同一个统计量，却形成不同的发生路径，不能用一个“政策传导”标签取代。

当阈值进入规则时，边界附近的分类尤其重要。一项微小修订可能改变是否触发措施，而对实际经济活动的描述差异却不大。制度若没有安排不确定区间和复核机制，就可能把测量误差放大为行动跳变。改进可以是调整决策规则，也可以是提高资料质量；不必每次都要求统计部门提供不存在的绝对确定性。指标与制度的相容性，是双方共同承担的问题。

一个简化的比率例子能够显示核算与制度之间的区别。假设某项规则依据债务与产出的比率作出判断，债务金额保持不变，而产出估计因新增资料上调，比率便会下降。这个算术变化不意味着既有偿债义务自动减少，却可能改变规则分类。若制度据此调整行动，

统计修订就获得真实后果；若制度允许暂缓并复核，后果又可能不同。这里需要分别追踪债务事实、核算判断和行动规则，不能把其中任何一个代替全部关系。

当增长进入考核时，行动者可能把资源转向更容易计入和展示的活动。这是一个值得研究的候选机制，不应未经资料就写成所有组织的事实。要检验它，可以追踪考核规则改变前后的项目选择、报告口径和实际结果，并比较不受同样规则影响的活动。若只是名称变化而资源与行为没有改变，应缩小机制主张；若真实产出增加且社会收益提高，也不能因为存在激励就自动视为扭曲。

指标还可能通过叙事作用。一个总量上升的消息，可以被解释为信心增强，也可以被某些群体理解为自身处境未被看见。不同解释未必来自一方理性、另一方情绪化，而可能因为他们面对不同分布和生活成本。研究应把解释与可观察处境连接，分析公众为何使用不同参照，而不是以总量正确为由宣布个人经验错误，或者以个人经验不佳为由宣布总量虚假。

制度回写的强主张，应当说明后续活动如何因此改变。信心变化只有经过消费、投资、招聘或其他行动，才进入生产；评价规则只有通过选择和执行，才重组资源。每一段都可能受到约束或被相反作用抵消。把链条分开，不是削弱整体视角，而是使整体关系能够承担因果责任。真正完整的解释，应允许某些通道成立、另一些通道没有作用。

24.6 总量与生活之间的距离

当 GDP 增长而一些人的生活没有改善，研究不应预先决定究竟是统计错了还是体感错了。两者可能在回答不同问题。总量描述某种边界内的活动，个人经验涉及收入分配、必要支出、时间压力、环境和生活安全等具体条件。一个总体增加与某些人的困难同时存在，在逻辑上完全可能。需要解释的是二者怎样通过制度和分布连接。

这种距离有多种来源。增长可能集中于某些部门，收益可能分配给特定持有者，必要服务的价格也可能相对变化。还有一些改善不容易在市场支出中体现，例如等待时间减少或共同设施更方便。反过来，支出增加有时用于弥补损失，并不意味着相应生活状况更好。上述都是分析方向，实际判断必须依赖具体资料，不能用几个可能性代替对某个国家或时期的研究。

此前的照护案例现在可以再推进一步。若家庭购买服务后有更多可支配时间，却因费用而减少其他消费，产出账户、家庭预算与时间记录会显示不同方面的变化；若照护质量提高，同时原照护者获得休息，其生活改善也不能用一笔付款完整表示。反过来，购买服

务可能只是弥补家庭成员无法继续照护的困难，支出增加并不保证生活比原先更好。这几种情况都应作为待区分的路径。研究需要在同类家庭中追踪支付、时间和服务结果，再确定哪些变化可汇入产出核算，哪些属于分配或福利评价。个案由此提出可采集的关系，不能代替对范围的抽样判断。

对福利的评价还需要价值判断。更多产出是否值得为之付出更长工作时间，某种收益是否足以补偿环境损失，不是 GDP 自身能够决定的问题。即使所有事实被准确记录，人们仍可能对公平、自由与可持续性赋予不同权重。统计机构可以提供相关资料，公共讨论仍需说明规范理由。把价值争论伪装成一个更全面的总分，并不会自动消除分歧。

因此，一种可行的制度安排是保持指标之间的分工与连接。GDP 承担产出核算的任务，分配资料说明不同群体得到什么，时间与服务资料呈现非市场安排，环境资料追踪资源后果。它们可以在同一问题中共同使用，却不必强行相加。研究者应说明为什么选择某些指标，以及当指标之间冲突时怎样作出判断，而不是让最容易传播的数字代替全部推理。

账本不能等同于经济生活，这一判断的具体批判对象是不当的全称使用：把一个为特定目的形成的量当成全部成就，把未进入该量的贡献当成不存在。纠正这种使用，不需要否认核算成绩；它需要建立从总量回到具体发生的路径，让被压缩的差异在相关决策中重新获得位置。

24.7 AI 与数字活动中的核算问题

AI 和数字服务使产出、支付与使用之间的关系更加值得研究，但不能因此宣称旧核算必然无法记录一切新活动。一项收费服务即使减少了直接劳动，也不因此自动退出生产边界；一项免费服务的经济处理，则需要检查其生产、融资与提供方式。劳动时间减少、市场价格为零和社会价值不可见，是不同问题，不能互相替代。

可以比较三个明确假设的场景。某企业购买 AI 服务以完成原来由员工承担的部分任务，某机构向公众提供免费数字服务，一个家庭使用免费工具完成过去自行承担的工作。三者都涉及技术替代，却在交易、生产边界和收益归属上不同。研究应分别追踪服务由谁生产、成本由谁承担、使用是否形成新的产品，以及节省的时间如何被重新安排。

第一个场景可能主要表现为投入结构与生产方式的变化。若企业产出保持不变而成本降低，影响会通过利润、价格、工资或投资等途径分配，不能直接从减少岗位推导总量必然下降或增长。第二个场景需要检查非市场供给与相关估价。第三个场景中，家庭节省时

间可能改善生活，却未必对应同样规模的市场产出变化。每种判断都需要回到具体核算关系。

数字产品还可能改变质量和数量的划分。一次服务调用与另一调用不一定提供同样结果，生成内容的数量也不必等于使用价值。若把处理次数当成产出，就需要说明质量、错误修正和后续采用如何计量。若改用支付金额，也仍需区分价格变化和服务变化。发生学的建设性问题，是在新任务中寻找合适的输出单位，并记录产品如何进入实际使用。

免费获取还会使消费者收益与生产核算之间出现不同表达。一个人获得便利，并不总需要有同额支出；但没有支出也不保证服务没有社会成本。数据使用、注意力、后续依赖和公共资源可能承担不同代价，这些具体机制需要分别调查。不能把免费一概浪漫化，也不能因为难以货币化就放弃研究。适当的时间、能力和选择指标可以补充货币记录。

对于 AI 时代的核算，最有价值的工作不是提前宣布总量体系崩溃，而是识别哪些对象身份和估价规则需要调整，哪些生活变化需要另建指标，以及这些新资料怎样进入政策。若核算能够经过修订继续承担产出测度，就应承认这种适应；若某些社会贡献仍被主要评价制度忽视，就应针对其使用方式提出改革。发生与账目的距离因此成为可持续研究的问题，而非一次性的失败宣言。

24.8 核算边界的公开责任

对 GDP 的争议应先定位到四种变化。原先未记录的生产被补充，属于可见性改善；已经记录的活动改变估价方式，属于方法修订；生产规模或组织方式改变，属于活动变化；同一结果进入新的预算或考核规则，则属于采用变化。四者可以同时出现，但反驳其中一项不能自动否定其余各项。批评者若认为家庭照护被低估，应说明要改进哪一种记录；若认为增长指标支配了不适当的决策，应检查采用规则。把这些任务分开，改革才能找到承担责任的位置。

边界改革的效果也应按这四层评价。扩大记录范围，首先改变的是可见性；可见性是否改变公共讨论，需要传播与采用证据；讨论是否改变资源分配，需要制度记录；资源分配是否改善生活，还需要后果评价。任何一层的成功都不能自动替后面的层次作证。反过来，改革未立刻改变行为，也不意味着资料毫无价值，它可能先改善比较与认识。

研究应允许边界保持稳定。为了长期可比性而维持一个中央口径，可以是合理选择，前提是其限度被清楚说明，其他重要活动有适当的补充记录。边界每次遇到新问题就全面

扩张，可能破坏可比性并增加不确定性。发生学并不把扩张视为唯一进步，它要求在稳定、补充和重建之间作出有理由的选择。

具体项目可以选择一种曾经改变核算处理的活动，建立新旧口径并行的比较，而不把全部历史重写成一条无缝序列。并行资料能够显示变化来自活动规模、估价方式还是分类归属。随后再检查使用部门实际采用的是哪一版，是否因此修改判断。若新口径只改善描述而没有引起行为变化，也应完整保存这一结果。它构成对制度回写范围的有价值限定，并提示哪些使用环节尚未建立连接。

另一项成果可以是为不同用户提供明确的使用说明。宏观分析者关注周期变化，公共服务研究者关注覆盖与质量，家庭研究者关注资源和时间，三者不必要求同一个总量满足全部任务。把适用问题写清楚，能够减少数字在传播中被过度赋义，也能让补充指标被视为必要知识，而非对中央统计的竞争性否定。

公开责任还包括解释不确定性。使用者需要知道一个变化在多大程度上可能受到修订影响，哪些部分主要依赖直接资料，哪些部分包含较强推算。不同资料质量不必使总量失去使用价值，却应影响决策的确定程度。一个要求绝对精确的制度，可能与实际测量能力不相容；改进有时应发生在制度规则，而不只是统计技术。

GDP 的发生性由此可以被表述为一个有限而有内容的命题：总量在真实活动、核算规则和资料过程中形成，并可能通过特定制度通道影响后续活动。这一命题保留了客观检验，也承认社会构成；允许核算成功，也容纳用途失配。它不把所有经济现象压成一个数字，也不因为数字有限就否认共同测量的必要性。

总量正确而用途失配，要求修改使用规则；用途合理而资料不足，要求改善测量；遗漏来自中央账户的任务边界，则可能需要并行指标。这三种处理不会由一个更全面的数字自动统一。第 25 章将进一步辨认制度改革的作用位置：改变人们知道什么、能够承诺什么和实际可以做什么，并检查改革是否只是把原有成本移到另一个环节。

第 25 章 经济制度如何改变发生路径

25.1 制度改变的最小单位

一项制度改革公布以后，参与者面对的实际问题可能仍然没有改变。申请表换了名称，办理步骤依旧；指标增加了一项，资源分配仍按旧标准；支付规则更加清楚，执行时间却继续不确定。也有相反情形：正式文本变化很小，但一项可靠的延期安排使许多人不再被迫立即成交。制度改变不能只按文件数量或措辞幅度判断，而应追踪行动路径究竟发生了什么变化。

经济制度在本书中首先被理解为组织行动、承诺和后果的安排。它规定谁可以进入，什么可以被请求，何时必须作出决定，哪些信息能够被获得，结果由谁承担，以及争议如何处理。这些规定并非完全外在于经济行为。它们参与形成可选择的对象、可相信的预期和可承接的责任，从而改变价格、参数和核算量所对应的发生过程。

一项有用的分析可以从最小改变开始：同一申请是否减少了一次无法预期的等待，同一承诺是否获得了可靠执行，同一交易是否增加了真实可用的退出途径。最小改变并不等于影响轻微，它只是让机制可以辨认。若研究从整套制度优劣开始，往往难以说明到底哪项安排产生后果。先找到具体环节，再讨论其与整体的关系，更有利于形成可以修订的知识。

制度也不等于一个设计者单方面写下的规则。执行部门、使用者和相关组织会在实践中解释规则、形成惯例，甚至改变规则的实际含义。这里不应把任何偏离都视为违规，也不应把所有惯例都视为合理适应。需要检查偏离为何出现，是否解决了文本未能处理的问题，以及谁因此受益或承担代价。制度的实际身份存在于规则与执行的承接中。

前四章已经区分了数值的不同身份，制度改革还要回答改变由谁实际承担。提前支付能够缓解接收者的时间缺口，却要求支付方重新安排现金；降低保证金可能扩大进入，也会改变履约风险的承担；改进增长评价可能增加对未被计入活动的关注，同时要求决策部门处理更多资料。每项改革都包含一组相互对应的权利、资源和责任。若只记录受益一端，就容易把成本转移当成成本消失。本章据此考察改革的执行结构：哪些路径真正被打开，哪些负担重新出现，哪些承诺有资源支持。

25.2 权利与选择如何成为可用能力

一项权利写在规则中，并不保证参与者能够实际使用。申请成本、信息难度、时间要求和后果风险，都可能影响权利能否转化为行动。若一个人名义上可以拒绝交易，却因缺乏短期替代安排而无法承受拒绝，研究者就应区分形式选择与实际可用选择。这一区分是描述性的起点，之后才涉及怎样评价和改善。

可以构造一个支付争议的分析性案例。两种制度都允许消费者提出申诉。第一种要求在很短时间内提交复杂材料，第二种提供清楚的证据要求并允许一次补充。两者拥有相同的书面权利，但形成的实际路径可能不同。研究应观察谁能启动程序、在哪一步退出、争议如何解决，而不能只统计申诉权是否存在。退出可能来自问题已解决，也可能来自成本过高，需要进一步区分。

制度还决定行动者能够怎样组合资源。一笔不可延后的费用与一项未来收入之间，可能存在时间缺口。若制度提供可靠的支付安排，家庭无需改变长期偏好，也可能采取不同消费或借贷选择。此时，行为参数变化不是主体突然变得不同，而是可行路径改变。把变化解释为一个新的偏好系数，可能在统计上得到拟合，却不利于理解可以干预的机制。

不过，增加选择并不总能改善判断。更多选项可能增加比较负担，复杂规则可能让形式自由与实际困难同时增长。制度设计需要考虑信息组织和行动能力，而不只是选项数量。一个清楚、可撤回且容易比较的程序，有时比大量难以理解的菜单更有用。这是需要具体任务验证的判断，不是对复杂制度的一般否定。

在发生学中，可用能力是资源、规则、理解和后果承接共同形成的稳定性。不能把它完全归于个人，也不能把个人从分析中删除。一个程序对熟悉制度的人容易，对初次使用者困难，可能需要改变说明和反馈；同一个人经过适当学习后也可能获得新的能力。制度改革与学习因此可以互相支持，研究要辨认当前障碍位于哪个环节。

权利能否使用，还取决于证明责任的安排。如果参与者必须先证明自己没有其他途径，才能获得某项支持，证明过程本身可能要求他已经拥有所缺乏的时间、材料或协助。研究可以沿着申请步骤记录这种循环，而不是只把未完成申请的人视为没有需求。另一种程序可以先提供有限支持，再在可承受的条件下复核；它也会面临误用风险，需要设计相应约束。两种安排的比较应同时观察排除错误与纳入错误，并说明错误的后果是否对称。只追求降低一种错误，可能把代价全部转移给另一类参与者。

这也改变了结果评价。某项改革即使没有立即提高平均成交额，若使人能够更可靠地拒绝不合适交易、减少无法预期的损失，仍可能具有重要后果。但是否值得为之承担成本，还需要明确价值与分配判断。发生分析负责说明能力怎样改变，规范分析负责说明哪些改变应被优先追求，两者通过同一事件相连接而各自承担责任。

25.3 承诺的时间结构

经济行动常常跨越时间。一方今天付出，另一方以后履行；一项政策今天宣布，参与者依据其未来持续性安排生活；一个组织今天接受评价，之后再调整资源。制度的稳定作用很大程度上来自承诺可以被承接，而承接不仅依赖条文，也依赖执行经验、纠错程序和改变规则的方式。

承诺可靠不等于规则永不改变。长期制度需要适应新情况，完全不可修订的规则可能把过去判断固化为未来障碍。问题是改变如何被说明、何时生效、既有责任如何处理。提前说明的调整，与突然改变已经依赖的条件，对参与者具有不同意义。制度研究应记录这种差异，而不是把两者都压成一个政策变化指标。

设想一项扶持计划承诺支持某类项目若干阶段。参与者据此投入前期资源，制度后来改变资格条件。若新规则对已开始项目设有清楚过渡，与立即取消后续安排，可能产生不同的信任和投资反应。这是一个待研究的机制情境，不是对任何具体政策的事实判断。它显示，规则内容与规则改变方式共同构成经济条件。

既有经验进入预期，也不必表现为完全理性的计算。参与者可能依赖少数显著经历、他人叙述或机构声誉。研究者可以先建立允许预期更新的强基线，再检查哪些真实信息被采用，哪些被忽略。Lucas 关于政策规则与行为关系的批评提醒我们，制度变化后的响应不能简单沿用旧系数。[44] 发生学继续追问的是，预期具体通过什么经验和传播路径形成。

时间结构还包括纠错速度。一个错误能否在造成不可逆后果前被发现，可能比最初是否发生错误更重要。若制度允许参与者补充信息、暂停执行和获得复核，某些不确定性就不会立即转化为损失。反之，过快而不可撤销的处理可能放大初始误差。快速并非总是优点，缓慢也并非自动可靠；需要把处理速度与后果可逆性一起分析。

对于跨期承诺，退出的条件尤其值得记录。一个人可以在签约时同意安排，却在后来遇到当初无法充分预见的情形。制度是否允许调整，如何分担已经发生的成本，决定了承诺是支持合作还是形成无法脱离的压力。允许调整并不意味着承诺可以随意撤销；它要求

把可预见风险、共同承担和特殊变化区分开。研究者可以比较同样冲击下不同退出规则的后果，观察合作是否因此更稳定，还是新的机会主义行为抵消了收益。由此，灵活性与可信度不再是抽象对立，而成为需要共同设计的关系。

承诺的调整还可能改变下一轮参与者的组成。若新进入者获得灵活退出，而已投入资源者仍承担旧约束，制度可能出现同类项目面对不同风险的局面。过渡安排需要说明差别来自已经形成的依赖，还是仅来自办理时间；否则，本来为了保护信任的做法也可能制造新的不透明。考察改革时，应同时记录旧承诺怎样结束、新承诺何时生效，以及两者交接期间谁负责维持服务。可信度由这些可执行的衔接支撑，不能仅由新文件措辞是否坚定来判断。

25.4 改变信息与改变行动条件

一项制度改革可能增加信息，也可能改变行动条件，两者常被混在一起评价。公开价格、提供风险说明和简化报告，主要改变参与者能够知道什么；调整保证金、支付期限和责任分配，则直接改变可以做什么。信息能够影响行动，但作用受到资源与权限限制；行动条件改变，也未必因为信息不足而需要更多宣传。

设想一个借款比较程序。第一种改革让费用表达更清楚，第二种改变提前偿还的附加条件。前者帮助理解既有选项，后者改变选项本身。在适当的试行中，可以分别观察仅改说明、仅改条款和两者同时改变的情况：若说明改善了理解却未改变选择，可能是原有替代路径仍不可承受；若条款改变后只有理解它的人使用新权利，说明与行动条件便具有互补关系。这些是待检验的机制，不是事先保证的效果。借款规模相同也不能判定改革相同，还应比较实际费用、提前偿还能否完成以及未选择新安排的理由。

信息制度还有排序效应。相同内容以不同顺序出现，可能使参与者先注意不同问题。发生研究可以把这种效应作为可检验的任务，但不应把所有选择差异都归为操纵。某种排序可能帮助理解，也可能遮蔽关键代价。判断需要观察是否改善了与使用者目的相一致的理解和后续结果，而不能只以是否改变选择来决定好坏。

市场时间安排提供另一类例子。频繁批量竞价的研究把订单处理方式与速度竞争联系起来，说明制度能够改变何种差异获得收益。[48] 在这一类问题中，增加参与者对速度风险的认识，未必等同于改变造成速度竞争的规则。两者可以同时有价值，但研究必须说明各自作用位置。不能把所有制度修复都退回个人教育，也不能假定改变规则以后参与者不再需要理解。

信息与行动条件还可能共同形成新的不平等。一个公开平台提供相同资料，使用者却在处理能力、可用时间和执行资源上不同；表面平等的发布不保证相同机会。另一方面，差异存在也不意味着公共信息没有改善作用。研究应比较改革前后各类参与者的实际能力，而不是在完全平等与完全无效之间作二选一判断。

制度修复因此需要定位问题。若障碍是隐藏费用，改善表达可能有效；若障碍是无法退出，再多说明也不能创造替代路径；若障碍是承诺不可信，则需要执行与纠错安排。发生学的实际贡献，是把笼统的不满转化为可以选择干预位置的机制图景，并让不同干预在共同目标下接受比较。

25.5 多路径供给与相互替代

市场、公共供给和共同使用可以为同一种需要提供不同实现路径。它们并非三个彼此封闭的世界。公共设施可能改变市场需求，共同使用可能减少重复购买，市场服务也可能支持公共体系的某些环节。研究若只按是否收费分类，就会忽略不同路径之间的实际连接。

以一种假设的设备使用需求为例。个人购买可以提供随时使用的便利，却要求承担购置与维护；租用降低一次性支出，但需要处理可获得性和持续费用；共同使用分担成本，却增加协调；公共提供可能扩大覆盖，同时需要确定资格和资源安排。不同路径改变的不是同一商品的标签，而是时间、责任和风险如何分配。

共同设备能否替代个人购买，关键在于其可用时点是否与任务相配。假设使用者都在相同季节需要设备，全年平均闲置很多，最需要时仍可能无法预约；另一个安排虽拥有较少设备，却能通过明确的排程和备用责任满足分散需求。由此，设备数量、全年利用率与可靠使用机会可能给出不同排序。若使用者为了防范预约失败仍保留私人备用设备，名义共享也未必减少原有负担。研究应记录需求时点、预约失败、维护停机与备用安排，判断新路径究竟替代了旧路径，还是在旧路径之外增加了一层费用。

还应关注路径转换成本。人们可能已经围绕既有制度安排工作、居住和家庭责任，新路径即使长期更合适，也可能需要过渡资源。若改革只计算最终状态的收益，忽略转换期承担的成本，就可能低估实施困难。参与者对改革的迟疑并不必然来自认知落后，也可能反映其承担了设计者没有记录的迁移风险。

多路径安排的价值也不等于无限增加方案。过多并行制度可能造成责任不清、重复记录和协调成本。需要确定哪些功能适合统一，哪些需要保留差异，哪些退出途径应始终存

在。发生学关注的是可承接的多样性：不同路径既能回应不同需要，也能够在转换时保留必要信息和责任，而不是让每次转换都从头证明自身资格。

这种分析为价格之外的经济实践提供正面位置。无偿、共同或公共安排并非只能作为市场缺陷的补丁，它们可以围绕不同目的形成稳定能力。但它们同样需要研究权力、排除和资源约束。把某种路径预先写成纯粹善意，会使其中的真实困难难以被看见。制度比较应使用对称的证据要求，使不同安排都接受后果和可修订性的检验。

25.6 改革效果的辨识与迁移

制度实施前后出现变化，并不自动说明变化由制度造成。改革可能恰好在危机中启动，参与者可能已经提前调整，其他政策也可能同时发生。发生研究尤其容易因叙述完整而产生因果确定的错觉：把公布、执行和结果串成一条故事，不等于排除了共同冲击和选择过程。

辨识首先需要把改革拆成可观察环节。文本规则何时改变，执行何时改变，参与者何时知道，实际行动何时发生，结果何时能够显露，应分别记录。若行为在执行前已经改变，可能存在预期作用；若规则改变很久后执行才变化，公布日就不是唯一合适起点。不同机制要求不同时间窗口，不能用统一的前后平均替代全部分析。

共享设备也揭示对照比较的一个困难。假设总容量不变，改革使一组使用者获得更可靠的预约，他们的等待缩短可能以其他人的等待延长为代价。若把后者当成不受影响的对照，就会同时把资源转移记在两端，夸大对改革整体的判断。研究应先划定共同争用的资源范围，再分别评价预约规则对不同使用者的影响，以及总任务完成量、总等待和维护成本。若改革还扩大容量，则要把排程改变与资源增加分开。制度比较的单位因此未必是获得新资格的个人，也可能是共同承担容量约束的服务系统。

对于可试行的程序改革，可以在适当条件下采用分阶段或随机安排，但方法选择不能脱离参与者权利。不能仅为识别机制而取消必要保障，也不能把知情参与等同于承诺没有风险。研究设计应说明为何选择某种比较、如何处理不利后果，以及何时停止试行。伦理条件不是研究之后的附录，它们决定哪些发生路径可以被正当地创造。

迁移检验同样重要。某项改革在熟悉制度、资源较充分的场景中有效，换到其他场景可能因执行能力或替代路径不同而改变作用。研究应明确哪些机制条件需要被保留，而不是仅复制表面程序。若效果来自可靠的执行，迁移时就必须检查执行条件；若效果来自降低理解负担，则需要验证新群体是否面临同样障碍。

还需防止把试行中的额外投入当成制度自身的永久效果。一个项目在试点时可能获得特别关注、熟练执行者和临时资源，扩展以后这些条件未必能够维持。研究应记录试行所需的真实成本，区分程序改变与额外支持的作用。如果改革必须持续依赖某种支持，那么支持就应被明确列入制度，而不能在扩展预算中悄悄删除。与此同时，规模扩大也可能带来新的学习和协调收益，不能预先认定试点一定高估效果。迁移问题要求把规模如何改变机制写出来，而不是仅给结果乘上人数。

值得继承的成果不是一个永远有效的政策系数，而是一组有证据支持的作用条件。它应说明干预改变了什么中介，哪些人受到影响，收益需要多久形成，代价由谁承担，以及何种资料会要求修订。制度发生学的实际目标，是让改革知识可以被谨慎扩展，也可以在条件不成立时及时收缩。

25.7 修复的价值与后果分配

制度修复具有规范方向，但不能由任何一个经济指标独自决定。提高交易数量、降低平均成本或增加总产出，都可能是有价值的目标，却未必足以评价全部后果。谁获得收益，谁承担风险，哪些权利得到保护，哪些能力被削弱，需要进入判断。事实资料可以说明后果，目标优先次序仍须公开论证。

一个改革可能提高平均效率，同时让少数人失去必要服务。是否接受这一结果，不能仅靠把少数情况叫作异常解决。研究应明确损失的性质、可替代程度和补偿可能，再进入价值讨论。另一方面，也不能因为任何人受到不利影响，就预先否定所有改变。制度评价需要比较真实可行的方案，而不是把每个方案与一个没有代价的想象状态相比。

稳定本身也可能有价值。可靠的支付、清楚的责任和可预期的处理，使人能够安排长期行动。发生学若只赞美变化，就会忽略制度稳定保护人的作用。需要修复的是稳定与实际条件持续错配的地方，而不是稳定这一形式。某些改革应减少变化，例如减少无理由的规则反复；另一些改革则需要打开被过度固化的分类。

判断制度是否陷入病理，可以从持续失配开始。规则反复制造已知损害，相关证据无法进入修订，承担后果的人又没有有效申诉途径，这样的组合值得特别关注。但不能仅凭一次错误就给整个制度下结论，也不能把探索性失败与拒绝纠错混为一谈。修复能力是制度质量的一部分，它允许合理试验，同时要求失败能够留下痕迹并改变后续行动。

不同参与者对同一结果可能有合理分歧。研究者不应把所有分歧都解释为信息不充分，因为人们可能真正重视不同价值。制度可以通过明确理由、提供复核、保留退出和限

制不可逆后果来组织这种分歧。发生学在此不是提供一个消灭争议的总公式，而是研究争议怎样被承接，使共同生活不必依赖完全一致的判断。

规范目标与机制研究由此形成双向约束。价值判断决定哪些后果值得观察，机制证据则可能迫使人们重新考虑实现目标的方式。若一项善意措施持续造成相反结果，不能只以动机保护它；若一项高效程序依赖不可接受的权利损失，也不能只以结果数字为其辩护。制度修复需要同时承担事实与理由的责任。

25.8 制度记忆与可继承的改革

一项改革是否成功，还取决于经验能否被保存并进入下一轮。若每次人员变动都使问题重新被发现，失败记录不断消失，制度就难以形成学习。相反，若全部旧判断都被永久保留为不可挑战的先例，记忆又可能成为固化的来源。需要的是能够说明来源、适用条件与修订理由的制度记忆。

这种记忆应围绕事件与责任组织。某项规则针对什么问题建立，曾比较哪些方案，哪些结果支持采用，后来在哪里出现失配，修订改变了哪些条件，都可以形成简洁记录。记录不必暴露与问题无关的个人信息，也不应以永久监控为代价。保存的标准应是后续判断需要什么，而不是技术能够收集什么。

制度还需要允许被影响者纠正记录。一个错误分类如果持续进入资源分配，后续数据可能不断重复最初偏差。复核不仅是一种救济，也是一种知识更新通道。参与者提供的解释不必被无条件接受，但应能够进入可以核查的程序。否则，制度所谓客观数据可能只是自身早期判断反复回写后的产物。

制度记忆还需要保留未被采用的方案及其理由。否则，后来的使用者可能误以为现行规则是唯一可想象的安排，把历史选择当成自然必要。保存少量关键替代方案，能够让新条件出现时更快重新比较，也能说明过去拒绝某个方案的理由是否仍然有效。记录的重点不是给早期决策者追认无限责任，而是使条件变化能够被识别。一个决定在当时合理，后来仍可能需要替换；承认这一点，可以减少改革必须先把过去全盘否定的压力。

制度记忆是否有用，可以在下一次条件变化时检验。共享设备若在需求集中后再次排队，过去的记录应能说明原成功依赖需求分散、备用容量还是额外维护人员；缺少这种说明，后来的管理者可能只复制预约界面。借款程序若在新群体中失效，也应能够回到原来降低的是理解负担还是实际费用。被继承的因而是经过证据支持的作用关系，以及使关系

成立的资源承诺。表面程序可以更换，只要关键功能仍被承接；表面完全相同，也可能已经失去原来效果。

本编的五项工作由此相接而不相互替代。参数分析确定何种响应能够迁移，交易分析辨认价格承诺，公开预测分析区分信息反应与对象改写，GDP 分析划清核算及其用途，制度分析则检查改变这些条件的资源和责任结构。同一个数值变化可以跨越几项工作，但每一步都要补上自己的证据。尤其从认识失配走向制度改革时，需要证明所改变的环节确实控制了问题，并说明新的安排把代价放在何处。

改革的停止同样应被记入制度学习。某项程序如果没有带来预定收益，继续维持它可能只因已经投入资源或形成了管理惯性。预先约定复查时点、替代方案和退出条件，能够让停止成为正常判断，而非必须掩饰的失败。反过来，短期未见效果也不总应立即终止，若机制需要较长形成时间，就应以中间证据和成本边界决定是否继续。学习要求的是有理由的承接，而不是永远坚持或不断翻新。

制度分析能够显示谁承担等待、风险和转换成本，却不能仅凭这些事实决定人们彼此应有什么义务。某种安排可能在资源上可行，在关系上仍不可接受；某项承诺也可能值得维护，即使它不增加交易总量。第 26 章从这里转向爱与包容的条件，第 27 章再讨论归责。经济部分交出的不是一套可直接移植的道德结论，而是让义务讨论能够面对具体处境的后果分析。

第六编 伦理 教育与意义发生

第 26 章 爱 包容与伦理事件的同一性

26.1 包容容量的思想实验

如果一个人第一次原谅失约，第二次仍然原谅，第三次却决定停止合作，能否据此说他的爱已经耗尽？这个问法把包容想象成一种可储存、可消耗的容量：每次过犯减去一定数额，容量归零时关系结束。作为思想实验，这个模型清楚而有吸引力，因为它似乎能够把复杂关系化成一条可预测曲线。但它首先必须回答，三次失约是否真是同一种事件，原谅又是否始终指向同一种决定。

容量模型不是毫无用途。某个人在一段时期内可以形成相当稳定的边界，愿意承担的时间和精力也确实有限。把这些限制说清楚，有助于避免无限责任。问题在于，稳定边界不能被直接解释成独立于关系历史的固定心理常数。相同等待时间可能因承诺、说明和后续修复而获得不同意义；相同继续合作的行为也可能出于信任、礼貌、依赖或压力。行为外形相同，不保证判断结构相同。

SDE 在伦理中的入口，是研究一个伦理事件如何获得身份。首先辨认发生了哪一次行动，再辨认这次行动在当时规则下意味着什么，最后说明它给后续关系留下了什么要求。这三个问题相互约束，却各有证据。一次未到场有可核查的时间和行为记录；它是否违反承诺，要核查当时有效的约定；它是否足以改变未来合作，还要考察影响、解释与修复。新解释能够调整后两个判断，却不能任意改写第一个事实。事件同一性由此成为一项约束：怎样切分历史，应当跟正在回答的问题相称。

爱作为一种关系实践，可以包含关心、承诺、欣赏、共同生活或共同事业等不同内容。本章不试图给所有爱规定一个统一强度单位，而是提出一个规范取向：对他人的关切应当与对其自主和尊严的承认相联系，同时不取消自身的合理边界。这是需要伦理论证的立场，不是从三方程推导出的自然法则。发生理论能够揭示关系怎样改变，却不能独自决定怎样改变才值得赞同。

包容的意义也需要澄清。容许对方保持不同观点、理解一次错误、放弃报复、恢复信任以及继续亲密关系，分别涉及不同决定。若把它们压成一个包容分数，研究者可能误以为拒绝继续合作就是缺乏宽恕，或认为恢复礼貌就意味着风险已经消失。概念区分能够保护判断的精细度，也能防止一套赞美爱的话语成为要求他人承担无尽后果的工具。

容量思想实验的真正价值，是暴露其省略的历史。每一次回应都会进入下一次事件：明确说明边界可能使合作更可靠，含混接受可能使双方误解彼此期待，承诺修复也会改变下一次失约的意义。伦理事件因此具有回写，但回写没有预定方向。它可以生成信任，也可以积累不满；可以使边界更清楚，也可以使表达更困难。研究与论证必须分别追踪这些结果。

26.2 同一个行为为何成为不同事件

以下是一组构造案例，不对应真实人物。甲与乙约定每周一起整理一份读书材料。第一次乙迟到，事前告知交通延误，并补做了甲因此承担的部分工作；第二次乙没有告知，事后承认忘记约定，并承诺设置提醒；第三次乙仍未出现，也没有提前说明。三次都可以被记录为未按时到场，但它们与承诺、信息和责任分配的关系不同。只保留迟到次数，便不足以说明甲为什么改变判断。

第一种变化来自新事实。甲后来得知某次失约确有无法避免的原因，可能修订对乙的责备。这改变的是对既有事件的认识。第二种变化来自新承诺。乙明确保证采取措施之后，再次失约便同时涉及未履行新承诺。第三种变化来自关系规则调整。如果双方后来约定可以独立完成、不再相互等待，未来相同的缺席可能不再构成失约。三种变化必须标明生效时间：后来取消等待规则，不能单凭这一点消除规则取消前已经发生的违约；后来发现当时确有免责条件，却可以改变对过去行为的归责。新信息向过去修正认识，新约定向未来改变要求，混淆这两个方向就会使事件历史失去约束。

这些变化说明，伦理事件的同一性具有问题相关性。若研究工作时长，三次缺席可以按同一时间单位统计；若判断是否违背承诺，就必须保留规则版本；若判断信任能否恢复，还需观察后续行动。并非存在一个覆盖所有用途的唯一事件划分，但划分也不能任意。它应当与正在判断的责任和理由一致，并接受事实约束。不能为了替某方开脱，就在不利时突然改换事件边界。

事件身份还涉及谁能够提出解释。甲可以说明等待造成的负担，乙可以说明其行动条件，第三方可以核查共同记录，但没有任何一方仅凭身份就拥有全部事实。与此同时，不同位置拥有不同经验：甲的被忽视感不能被乙的一句没有恶意直接取消，乙的说明也不应在未听取时被预先排除。发生分析要求让位置差异进入证据，而不是把位置差异本身当作不可讨论的结论。

这里可以提出一个限制原则：事件重述必须增加可检验事实或可公开讨论的理由，而不能只是更换有利标签。把持续失约称为自由表达，并没有解释对方为何应继续等待；把一次失误称为人格不可靠，也可能超出了证据。合理重述应说明哪些特征与当前判断相关，为什么相关，以及什么新信息会使判断修订。伦理解释由此具有可问责性，而非单纯由叙述技巧决定。

同一性问题最终改变了对包容变化的理解。甲第三次拒绝共同整理，并不必然说明爱减少了，也不必然说明判断更成熟。需要考察拒绝的理由、替代安排和对乙的基本尊重。甲可能在保持关切的同时停止某种合作，也可能借此进行不成比例的贬低。发生论不替具体选择预先免责，而是使选择能够按照其完整事件结构接受评价。

26.3 三种侧重怎样共同规定爱

爱的客体侧重，可以从可辨认的关心行动和承诺内容进入。付出了什么、答应了什么、后果由谁承担，都是可以核查的问题。但行动清单不能独自决定爱是否成立：同样的帮助可能尊重对方，也可能替对方作主。互动侧重由此进入，追问帮助如何被提出、接受、拒绝和修订。主体侧重则涉及当事人怎样理解自己与对方的位置，哪些需要被承认，哪些边界不能被交换。

三种侧重不是三个分数相加。若一项帮助以取消对方决定权为条件，行动数量再多，也不能自动抵消互动方式中的问题；若一个人表达真挚关切，却持续不履行可承担的承诺，主体表达也不能替行动后果免责。共同发生意味着各侧重互相限制。爱的稳定性来自行动、关系和承认之间能够持续形成相容结构，而不是来自任何一项被无限放大。

在甲乙案例中，这种相互限制可以通过两个反例检验。假设乙补足了全部遗漏的工作，却要求甲此后不得拒绝任何邀请，行动负担得到修复，未来选择权却受到了新的限制。又假设乙充分听取甲的感受，也承认其有权退出，却始终不交还共同资料，良好的讨论仍留下未履行的具体义务。这些都是为论证而设的案例变式。它们表明，行动、互动与主体承认各有不能被另一侧重全额替代的要求。每项要求内部仍须追问怎样显露、经过何种差异过程、由哪些特征共同规定；这里需要的是相互校验，而非把三个侧重分别配给 S、D、E。

把爱与过犯各自的三种侧重交叉，可以形成九种提问，用作检查遗漏的工具。以关心行动为起点，可以分别追问它补救了哪些实际损失、怎样进入协商、是否尊重对方的决定；以互动安排为起点，可以追问它能否落实补救、能否容纳异议、能否使被忽视的经验

得到表达；以主体承认为起点，可以追问承认如何约束行动、怎样改变回应方式、是否改变了双方理解彼此的位置。九种提问不组成九个可相加的心理变量。若前三项显示补救已经充分，后面的提问仍可能揭示选择权受到限制。交叉检查的功能，是发现一种成功不能抵消的另一种缺口，而不是给关系计算总分。

一个建设性的论证可以从冲突处展开。甲认为自己已经付出很多，乙却认为自己的选择从未被认真听取。双方的分歧未必是其中一方计算错误，而可能来自对关心成立条件的不同理解。若把争论化成谁付出更多，互动与承认的问题就被遮蔽。发生分析可以帮助提出新的行动：共同说明哪些帮助被需要、哪些决定应由谁作出，以及承诺怎样被双方理解。

这种分析并不保证和解。双方可能发现目标根本不同，无法维持原来的关系形式。此时清楚而尊重的分离，也可能比继续进行含混交换更符合所接受的规范。SDE 所关心的是关系形式怎样生成和改变，不是保证所有关系继续存在。爱的发生性只有容纳边界、拒绝和结束，才能避免变成要求一切差异最终被吸收的总体性理想。

26.4 宽恕 信任与和解的不同时间

宽恕可以被理解为改变对过去过犯的某种态度，但恢复信任涉及对未来行动的期待。两者在逻辑上并不相同。一个人可以决定不再报复，却仍要求新的保障；可以接受道歉，却暂不恢复原有合作；也可以保留对某项行为的否定，同时维持其他关系。把这些决定分开，能够使包容从全有全无的命令转成可说明的实践。

道歉也不是单一信号。它可以确认事实、承认责任、表达遗憾、提出补救或承诺改变。不同内容对后续关系承担不同作用。只说让你不开心很遗憾，未必承认行动本身有问题；承认责任而没有可行补救，也未必足以恢复合作。发生分析需要追踪语言怎样进入后续行动，而不是把一次口头表达直接视为修复已经完成。

在前述构造案例中，乙承诺设置提醒之后，甲可能给出一个有限的继续合作机会。这个机会既不是完全恢复信任，也不是隐性惩罚；它可以是一种双方清楚理解的过渡安排。若乙随后持续履行，信任可能逐步重建；若再次失约，甲便有理由重新评估。这里的后效不是机械累加，而是新行动检验先前承诺，使事件的意义在时间中逐渐确定。

和解则需要更强的共同条件。单方面宽恕可以发生，双方恢复共同活动却需要一定的规则相容。若只有受影响的一方不断调整，而另一方拒绝承担后果，关系继续不等于和解成功。规范评价应考察选择是否自愿、负担是否合理、改变是否被共同承认。不能把表面

平静当作修复的唯一指标，因为沉默可能来自多种原因，研究者需要听取并核查其具体条件。

宽恕的时间也不应被外部统一规定。有人需要先获得事实说明，有人需要观察行动，有人可能选择不再进入原关系。要求立刻原谅，可能剥夺当事人形成判断的空间；要求永远记住每次错误，也可能阻碍已经有充分理由的修复。合理的时间安排应与后果、证据和自主决定相联系。发生理论强调历史，不意味着过去永远不能被重新安置。

修复还包含一个容易被遗漏的不对称：作出补救的人可以承担自己的义务，却不能以已经补救为由要求对方产生指定情感。履行责任能够改变对方判断的理由，不能直接制造信任。若把补救理解成购买宽恕的交换，主体承认便再次被行动数量取代。合理的关系安排应允许对方认真评估证据，同时避免将不立即恢复亲密解释成新的过犯。

同样，接受宽恕不意味着过去后果自动消失。共同材料需要补齐，额外工作需要重新分配，已经影响第三人的部分也可能仍需处理。情感上的释怀与实践上的补救可以有不同进度。把这两条时间线分别记录，能够避免双方因使用同一个已经过去了而产生新的误解，也使修复从一次仪式性的表达进入可持续行动。

因此，修复不是把关系恢复成从未发生过过犯的样子。已经发生的事情进入了共同历史，新的稳定性可能建立在更清楚的边界、更具体的承诺和不同的合作形式上。这样形成的关系不必较旧关系更差，也不必被浪漫化为更深的爱。它需要按照新的结构与后果评价。发生论能够解释这种重建，但其价值仍须由当事人的理由和共同接受的规范来论证。

26.5 包容的边界与不对称关系

关系中的选择能力并不总是对称。一方拥有更多资源、解释权或退出机会，可能使另一方难以公开拒绝。即使双方都说愿意继续，也需要考察这种表达在何种条件下成立。发生分析不能把一切压力都解释为自由互动的一部分，然后以关系复杂为由放弃规范判断。相反，正因为事件由条件共同规定，权力与依赖才必须进入对自愿性的评价。

本章采取的规范前提是，关切不能以取消他人的基本尊严和自主为代价。联合国教科文组织的《世界生命伦理与人权宣言》在其适用领域中同时强调尊严、自主、脆弱性保护与文化多样性的边界。[52] 这些原则可以为本章提供规范对话的参照，但并不直接规定所有私人关系的具体答案。把规范来源及适用范围说明白，比借用权威文本替代实际论证更重要。

包容差异与容忍伤害也需要区别。观点不同、习惯不同或能力有限，并不当然构成对他人的不当侵害；持续违背明确承诺、把后果单方面转嫁给他人，则需要另外的理由。区分不能只依赖某方主观不快，也不能只看是否具有恶意。应当考察行为后果、可预见性、可替代选择和双方接受的责任。爱的语言可以增加理解，却不能使这些问题自动失效。

退出权具有建设性意义，但形式上可以离开，还不充分说明继续合作具有清楚的自愿基础。在前述构造案例中，若甲停止共同整理就会失去自己完成其他工作的全部资料，退出成本便部分来自乙对共同资源的控制；若甲只是需要承担事先合理约定的交接工作，则成本未必表明不当限制。两者需要分别判断。可以先归还各自资料、明确共同成果的使用方式，再讨论是否继续合作，使选择免于不相关的资源要挟。退出也不取消已经存在的合理责任：停止共同整理可以与保持礼貌、完成必要交接并存。这样，边界既保护拒绝的空间，也保留对既往行动的承接。

稳定边界也可能体现成熟。若某人经过充分反思，持续拒绝一项会造成不合理负担的要求，判断没有变化并不说明缺乏开放性。开放性可以体现为愿意听取新事实，而不是每次都降低边界。反过来，边界反复变动也不必然说明理解加深，它可能来自压力或不确定。伦理评价需要看理由与条件，而不能把变化方向当作道德发展的刻度。

关系的对称性还不意味着义务完全相同。不同能力和角色可能产生不同的合理责任，但差异需要能够说明，并且不应变成永久免除某方责任的借口。一个人暂时需要更多帮助，可以获得支持；这种支持如何安排、何时复核、如何避免另一方耗尽，同样属于爱的实践。发生分析使这些后续条件可见，从而把关心从抽象善意推进到可持续的共同生活。

26.6 规范判断不能由变化曲线代替

设想研究者长期记录甲对乙的包容评分，发现评分下降后又回升。这个构造曲线本身不能说明关系改善。回升可能来自乙真正修复，也可能来自甲降低期待，或者来自甲不愿继续表达不满。要区分这些可能，必须结合事件事实、当事人的理由和实际选择。数值可以帮助发现值得追问的变化，不能独自决定变化的伦理意义。

同样，预测某人会原谅，与论证其应当原谅，是两个问题。行为模型可以利用历史和情境提高预测，但预测成功不赋予模型规范权威。若一个人因为缺少退出条件而继续关系，模型可能准确预测继续，却不能由此证明继续是公正的。发生理论若想进入伦理，必须承认描述与评价之间需要明确的规范前提，而不能让共同发生四个字承担应当的来源。

本章的规范论证可以写成一条有限链条：他人的自主与尊严值得尊重，合理承诺应受到认真对待，后果不应被无理由地单方面转嫁；因此，对关系变化的评价应检查这些要求是否得到满足。SDE 帮助揭示自主、承诺与后果在具体事件中怎样被实现或破坏，却不证明这些价值本身。把前提公开，允许不同伦理立场提出质疑，正是规范讨论应有的可争论性。

竞争解释在伦理中也具有位置，但不能完全照搬实验模型比较。甲拒绝合作，既可以被解释为保护边界，也可能被解释为不成比例的报复。两种解释需要不同事实和理由支持。研究者可以比较叙述与行动是否一致，是否存在更温和可行的选择，以及当事人是否认真回应他人的合理利益。判断并非只看哪种解释预测更准，还要看其理由是否能够经受对等质疑。

伦理判断可以保持不确定。有时事实不足，有时价值冲突没有单一答案，有时当事人尚未形成决定。保留未决状态不等于放弃责任，但它也不能成为无限拖延的借口。需要区分可以等待的问题和必须先采取保护性安排的问题，并说明暂时决定如何复核。这里的未结算不是美的自动指标，而是一种需要管理的实践状态，其价值取决于是否保护了合理的判断空间。

检验理由时，还可以在保持相关条件的前提下交换双方位置。假如甲认为自己的时间有限足以拒绝额外工作，却认为乙提出同样理由就是缺乏关心，便需要说明两人的角色、承诺或能力有何相关差异。若确有差异，义务可以不同；若只有身份不同，理由就缺乏一致性。这个检验不要求双方拥有相同感受，也不保证他们最终赞同，却能把选择从偏好表达推进到可回应的规范主张。发生伦理学的积极功能由此更具体：保存使差异有理由的历史，同时识别仅靠位置优势维持的双重标准，让稳定边界和关系更新接受同一种理由要求。

26.7 叙事材料与伦理研究的证据

伦理关系研究可以利用自愿叙事、共同记录和低风险情境判断，但这些资料具有不同地位。当事人的叙述表达其经验与理由，研究者的编码则是对材料的解释；两者不应被合并。某人说自己原谅了，研究者可以进一步询问这个词意味着什么，却不能只因其行为不符合预设模型就宣布叙述无效。相反，叙述也不能替代对承诺和后果的事实核查。

事件访谈可以围绕具体转折展开，而不是要求当事人给整段关系一个总分。什么时候第一次觉得规则改变，哪项承诺使期待不同，什么行动使信任恢复或降低，这些问题更接

近发生结构。提问本身也可能推动反思，因此研究应区分自然形成的叙述与访谈中生成的新理解。研究者参与了资料的形成，并不使资料失去价值，却要求明确这种参与。

编码可以使用行动事实、互动过程和主体解释三种侧重，但不应强迫每段叙述只能归入一个类别。同一承诺既是可记录行动，也是互动安排，还涉及自我理解。允许多重编码和合理分歧，能够保留材料的复杂性。随后应检查这些编码是否帮助辨认新问题、提高事件重建的一致性或支持更清楚的规范讨论，而不是仅以编码数量显示理论丰富。

构造情境任务可以进一步区分事实改变与获知顺序改变。研究者可以准备两份包含完全相同事实的故事，一份先说明缺席，后补充当时无法联系的原因；另一份先说明原因，再说明缺席。随后比较参与者如何判断行为、信任与后续合作，并询问他们是否修订了最初解释。这只是可研究的设计设想，没有预定结果。若判断不同，差异首先涉及理解的形成过程；不能据此说两份故事中的过去行为已经不同。再将承诺版本或补救行动真正改变，才可以讨论事件条件变化的作用。如此设置对照，使叙事入口与伦理事件身份得到分别检验，也为自然关系中的访谈提供更明确的问题。

隐私与资料最小化在这里具有特殊意义。为了重建关系，并不意味着有权收集双方全部通信或让一方替另一方授权。研究可以优先使用自愿提供且与问题直接相关的材料，并让参与者理解哪些内容会被引用和如何匿名。发生档案的目的应是支持可复核判断，而非建立对亲密关系的全面监控。能够记录的历史不等于应该记录的历史。

研究结果的表达也应避免给人格排序。某人在构造情境中选择较少包容，可能因为其重视承诺、风险判断不同或对故事理解不同，不能直接称为缺乏爱。若研究发现某种修复安排更容易获得接受，应限定于所研究材料与群体，并将接受与正当性分开。伦理研究的建设性，是帮助人们更清楚地理解和讨论关系，而不是以科学外观给复杂生活发放道德分数。

26.8 爱的稳定性与关系的再创造

爱在发生中取得稳定，并不意味着每次互动都必须重新谈判全部规则。共同理解、可信承诺和习惯可以减少反复解释的负担，使关系具有可依赖性。这种稳定本身值得解释和保护。发生理论需要说明它怎样形成、怎样承受差异以及何时需要修订，而不能将熟悉和一致一概视为退化。没有任何稳定，持续关切也难以成为可辨认的实践。

关系的再创造则发生在旧形式不能继续承担新要求的时候。共同工作的方式、帮助的边界或相处的节奏可以改变，而关切仍可能保留。前述构造案例中，甲乙可以从固定共同

整理转为各自完成后交换意见。这不是必然更好的方案，但它展示了第三种可能：既不机械维持旧合作，也不把一次形式的结束等同于全部关系终止。创造来自对条件的重新组织。

在这一设想中，新合作是否成立，可以通过一次有限的安排来检验。甲乙先约定各自完成的部分、交换意见的时间和无法履行时的通知方式，并说明任何一方可以停止后续轮次。若双方都能清楚复述这些条件，下一次的期待才具有共同对象；若一方理解为自愿交换，另一方仍理解为随时待命，换一个合作名称并未解决旧失配。新的显露与旧差异序列不相容时，E 的调整需要表现为可辨认的规则和权限变化。反复失配的行动由此成为重建规则的入口，新的关系形式能否成立，则由后续行动检验。

边界也需要经过后续行动检验。假设乙口头接受甲可以退出，却在甲真正拒绝一次邀请时以先前帮助要求其服从，那么新安排尚未形成承诺与选择之间的相容关系。反过来，若乙接受拒绝，并按原约定完成交接，即使双方不再合作，新的边界仍可能得到履行。成功标准因此不能只设为关系延续。它应当分别考察约定是否被理解、拒绝是否有效、未完成事项是否得到承担，以及新实践是否再次制造同类负担。三原理在此帮助定位需要改变的组织，而这些改变是否正当，仍由前文提出的自主、承诺与后果理由来判断。

因此，爱与包容的发生研究最终应拒绝两种简化：把关系化为固定容量，或把关系化为无边界流动。前者遗漏事件身份与历史，后者容易使责任无法落实。更有力的观点是，关系中的稳定与改变都由具体实践取得，并在尊严、自主、承诺和后果的约束下接受评价。爱的深度不能直接由忍受次数或评分方差表示，但可以通过行动怎样承认他人并承担历史获得论证。

本章确立了三项结果：伦理事件必须保留行动事实与规则版本，包容可以与停止某种合作并存，修复需要分别检验情感、信任和实践义务。仍有一个问题没有回答：已经辨明的缺口应由谁承担，承担的是解释、补救还是责备？甲受到了影响，不意味着乙对全部后续困难都负无限责任；双方共同形成了关系，也不意味着乙对自己承诺的违背只需承担一半。责任分配需要另行说明义务来源、可控制的选择及合理的成本边界。这正是从事件同一性进入公共规范与共同研究时必须补上的论证。

第 27 章 规范判断与责任的发生

27.1 共同发生并不取消归责

一项研究结论可能经过设计、采集、分析、审阅和发布多个环节。若最后出现错误，简单地寻找最靠近结果的人，可能遗漏先前的决定；若把错误完全归给系统，又可能使每个人都以复杂性为由退出责任。共同发生提出的难题正是在两者之间：结果由多方条件形成，责任却仍需要有理由地落实到行动、角色和可采取的改进。SDE 不能只解释结果怎样生成，还要帮助说明谁能够对哪些后果作出回应。

因果贡献与规范责任首先需要分开。一个设备故障可以是结果的原因，设备本身却不因此成为道德行动者；某人没有直接修改数据，却可能因承担复核职责而对遗漏负责。因果分析帮助重建发生路径，规范分析则需要考察义务、知识、能力和选择。两种分析相互依赖，但不能互相替代。若把责任简单分配为因果比例，就无法处理监督、授权和不作为等重要问题。

本章采取一个明确的规范起点：能够影响他人合理利益的行动，应当在适当范围内接受理由要求；承担角色和承诺的人，应对其可控制、可预见且属于职责的后果作出说明，并参与合理修复。这些要求并非由自然事实直接推出，而是对共同实践的一种伦理主张。其理由在于，没有可回应的责任，承诺无法形成可靠预期，受影响者也缺少要求纠正的平等位置。

发生分析能够使责任对象更精确，但归责还要形成一条有理由的连接：行动者承担了哪项义务，当时能够采取什么行动，有哪些信息使该行动应被考虑，实际选择遗漏了什么要求，要求其承担何种后果才相称。一个签字可能只确认收到，也可能表示完成实质审阅；同样的动作因承诺内容不同而承担不同义务。若只找到行动通向结果的路径，却没有说明其中哪一项应当被履行，就仍未完成归责。反之，明确的审阅义务可以使未采取行动成为责任对象，即使记录中没有一个可被单独指认的致错操作。

共同发生也使责任具有时间性。行动者在决定时知道什么，与事后所有人知道什么不同。新证据可以改变对后果的认识，却不能自动变成行动者当时已经拥有的知识。另一方面，事后获得信息会生成新的责任，例如停止错误传播或修订程序。将事前判断与事后回应分开，能够避免后见之明的不公，也防止行动者以最初不知情为由拒绝后来必要的改进。

责任因此不是一块固定附着在个人身上的属性，也不是随叙述任意流动的标签。它在具体角色、承诺、能力和事件中取得稳定性，并可以随着条件变化重新分配。发生论的任务，是使这种分配可说明、可争论、可修订。只有做到这一点，共同发生才不会成为无人负责的借口，而能揭示原先被简单归因遮蔽的责任位置。

27.2 一个研究摘要的构造案例

以下案例完全是为分析而构造的。某联合研究小组在探索性分析中发现一种候选关系，原定计划是在新资料中再检验。撰写摘要时，分析者使用旧模板，把探索结果写成已经验证；项目负责人为了赶上内部展示，取消了原定复核；另一位成员发现措辞过强，却认为自己不负责发布而没有提出。摘要后来被其他团队当作确定结论使用，造成重复工作与资源浪费。

这个事件有多个因果环节。模板影响了表达，赶期决定改变了审查流程，沉默使纠正机会消失，接收团队又根据摘要采取行动。但责任不能平均分给所有环节。模板本身不是道德主体，分析者是否理解证据层次、负责人是否有权取消复核、成员是否有明确报告渠道，都影响归责。接收团队是否进行了与用途相称的核查，也需要独立分析，而不能把上游错误当作所有下游判断的唯一解释。

沿用这个构造案例，再设一个并行发布的变式：两名成员分别负责向两个团队发送摘要，都有义务核对证据状态，却各自发送了未经核对的确定表述。即使撤去其中一次发送，另一次仍会造成错误传播；但这不能证明任何一人都无责任。每人的义务指向自己负责的渠道及其可避免的风险，而非保证自己成为全部后果的唯一必要原因。若同一接收团队从两条渠道取得文本，还需避免把相同损失重复计算。可以同时认定多项职责被违反，再依据各自的传播范围分配补救工作。责任判断不是把一个结果拆成总和必须为一的过错份额。

需要进一步区分错误与不当行为。分析者可能误解模板，也可能明知证据不足仍夸大；负责人可能低估风险，也可能明确忽视已知警告。两种情况的责任程度和修复要求不同。发生分析不应为了强调条件而抹去意图，也不应只看意图而忽视可预见后果。对于无法确定的部分，应保留不确定，不能让一个完整故事替代缺少的证据。

这个案例还说明，结果严重程度会影响修复需求，却不能独自决定先前行动是否合理。若摘要侥幸没有被他人使用，取消必要复核仍可能是不当决定；若经过合理审查仍出现难以预见的错误，也不能仅因后果较大就认定同等过失。规范判断需要同时考察过程与

后果，并解释二者在当前问题中的权重。只追结果或只看程序，都可能遗漏责任的重要方面。

SDE 所带来的建设性问题，是如何改变下一轮发生。可以修订模板，使证据状态更清楚；恢复关键复核；建立成员提出异议的渠道；在发布后保留更正机制。但每项改进都需要明确承担者和验证方式。不能把增加一份表格当作责任已经落实，也不能在一次培训之后默认问题不会重现。责任的回写必须进入真实决策条件，才能超出事后表态。

27.3 知识 能力与角色的交叉

对行动者提出责任要求，需要考虑其能够知道什么。这里的能够不仅指个人聪明程度，也包括信息是否可获得、角色是否要求核查以及时间资源是否足够。一个人不应被要求知道任何隐藏事实，但承担专业复核职责的人，也不能以没有主动查看本应检查的信息为由免责。知识责任具有角色性，既受实际条件限制，也由承诺和职责生成。

能力同样不是抽象的全能。某成员能够发现措辞问题，却无权直接修改已发布材料；这不意味着其没有任何责任，而是需要考察可以采取的报告、提醒或升级渠道。若制度没有安全可行的异议路径，组织就承担建立路径的责任。若路径存在且可使用，行动者又需要说明为何未使用。个人与制度的责任可以同时成立，并不形成零和竞争。

角色责任还需要明确授权边界及其变化历史。项目负责人可以安排时间，却未必有正当理由取消实质审查。假设负责人先在没有充分理由的情况下撤去复核岗位，发布时又以无人具备权限为由主张无法检查，眼前的能力缺口并不能消除对先前决定的追问。归责对象可以前移到其仍可控制的安排环节。相反，若能力缺口来自其无法控制且已经合理报告的突发条件，就需另行判断暂缓、缩小用途或补充支持是否可行。两种情形同样表现为当前缺乏能力，却具有不同的义务历史。SDE 保存条件的生成过程，正能防止人们只用结果时刻的截面为自己免责。

继续比较这个构造案例的两个版本。版本甲中，分析者收到清楚的证据标记规则，却忽略它；版本乙中，团队从未区分探索与检验，所有模板都采用确定语气。两个版本都可能产生相同错误摘要，但责任结构不同。前者更强调个人未履行明确要求，后者更强调组织未建立必要区分。区别不是为了取消任何一方责任，而是为了使纠正动作针对真实成因。

规范要求也应符合合理能力，不能在事后不断提高标准。若对每个成员都要求独立重做全部分析，责任制度会变得不可执行。合理的分工需要可信交接和抽查，同时明确哪些

关键环节不能完全依赖他人。发生分析能够帮助识别这些关键环节：哪些错误一旦通过便难以逆转，哪些判断需要不同专长相互约束，哪些记录足以支持后续复核。

可预见性还应区别于能够预知完整事件。成员可能无法知道哪个接收团队会误用摘要，却能够预见确定措辞会使他人高估证据。这种类型层面的风险，可以支持采取适当核查的义务。反之，不能因为事后能讲出一条完整因果链，就要求行动者当时预测全部细节。归责应说明所要求预见的风险类型，以及当时有哪些线索足以使这种要求合理。

证据缺失也需要谨慎处理。没有保存复核记录，并不自动证明没有进行复核，但它可能表明记录义务未被履行。两项判断应当分开，不能用一种缺失直接填补另一种事实。若关键记录由组织控制，组织也不应把说明困难全部转嫁给较弱势成员。公平的调查需要比较可获得材料、听取不同位置的解释，并将无法确定的部分明确留下。

合理依赖他人是合作的必要条件，但这种依赖可以随新线索改变。接收者原先有理由相信已完成审阅，后来却发现文本与分析计划冲突，便产生进一步核查的理由。责任在此并非从一开始就要求所有人重复全部工作，而是在异常出现时要求适当回应。发生分析能够追踪这个转折点，使后来生成的新义务与原先合理的分工保持区别。

知识、能力与角色之间的匹配，可以被看作责任稳定性的条件。若有责任而无权限，制度可能制造无法履行的义务；若有权限而无问责，制度可能放大无约束行动；若信息被分割却要求个人承担全部后果，归责就可能不公平。SDE 对特征纠缠的强调，在这里体现为检查这些条件怎样共同组织行动，而不是把责任简化成个人品质的单向显露。

27.4 分布式行动中的不可转移义务

许多共同项目的成果无法归给单个行动者，但这并不意味着所有义务都能被转交。一个人可以委托技术处理，却仍需要对自己作出的承诺负责；一个组织可以采购模型，却不能仅因模型来自外部就放弃对使用场景的判断。责任可以分工，不能只通过流程转移让受影响者失去可回应对象。分布式行动需要清楚的接口责任。

接口责任首先包括说明输出的证据地位。分析者交付的是原始数据、探索图形、经过验证的模型，还是适用于某范围的建议，必须被明确标识。接收者则需要确认这些标识与自己的用途相容。若接口只传递一个看似确定的分数，双方都可能在不自觉中扩大它的含义。发生分析要求保留产生分数的条件，使下一环节的行动不是建立在被压缩掉的假设上。

其次，自动化不应被当作无主的因果环节。程序可以依规则执行，但规则由谁选择、何时更新、谁监测失效，仍是人的组织问题。将错误称为算法决定，并没有回答这些责任。反过来，也不能让最靠近设备的操作者承担所有设计缺陷。需要区分设计、部署、监督和具体使用的责任，并根据控制能力和可获得信息分别说明。

责任分布还应保留最终回应机制。如果每个团队只说自己的局部步骤没有问题，受影响者可能无法获得完整解释。组织应安排能够整合事件链条的复核角色，而不是让当事人在部门之间反复转交。这个角色不必拥有全部技术知识，但需要有权收集资料、组织不同解释并推动纠正。共同发生的知识要求，在制度上对应着共同回应的能力。

分布式责任需要分别说明过错与参与修复的根据。角色和明确承诺可以产生事前核查义务；实际控制可以支持要求采取可行的防止行动；事后受益或当前能力，则可能成为讨论修复成本分担的理由。控制、受益和能力这三项因素不能不加区别地转化成先前过错。一个原先不知情且没有核查职责的受益团队，或许应在明确、公平的安排中贡献恢复资源，却不能仅因受益就被认定曾经有意误导。每一项要求都应注明它针对过去违背的义务，还是针对现在共同恢复的需要，并论证比例。如此，广泛的因果依赖能够被看见，归责的规范边界仍保持清楚。

这种边界也保护合作本身。若每个成员都担心对任何不可预见后果承担无限责任，可能倾向于隐藏问题或停止合理探索。一个公正的责任制度应区分善意且合理的错误、可避免的疏忽和有意误导，并为及时报告与修复提供空间。这里的包容不是免除责任，而是使责任能够推动学习。前一章关于宽恕与信任的区分，在共同研究中转化为对个人评价与流程改进的分别处理。

27.5 追责与修复的不同方向

回顾性归责关注过去的行动是否符合要求，前瞻性责任关注现在谁应采取措施减少后果。两者经常重合，却不必完全相同。假设接收团队掌握唯一可立即暂停后续使用的接口，而发布团队掌握证据说明，合理的应急分工可以要求前者暂停、后者澄清，随后再讨论成本承担。这是对构造案例的继续推演。技术上最便于行动者，不因此自动承担全部经济或声誉代价；曾有过错者，也不能仅因暂时无法操作接口而退出补救。将行动权限、修复成本与先前责备分别安排，可以使必要回应尽早发生，同时防止便利性成为转嫁责任的理由。

在构造的摘要案例中，立即标明证据状态、通知已经依赖摘要的团队、保留更正版本，属于可先推进的回应。之后仍需调查夸大如何发生、哪些职责未被履行和如何分配成本。及时修复不等于提前认定全部过错，也不能被用来阻止合理追问。一个成熟制度需要同时支持两条工作线：减少当前损害，并建立对过去决定的可信解释。

修复还需要听取受影响者，而不能由造成问题的团队独自定义已经解决。接收团队可能需要的是清楚的数据范围和可替代分析，而不是一封笼统道歉。不同后果需要不同补救，责任回应应当与实际需求连接。与此同时，受影响者提出的任何要求也都可以接受比例性与可行性的讨论。尊重其位置不意味着取消共同论证，而是使其不被排除在修复定义之外。

制度修订应针对机制而非只增加惩罚。若错误源于证据状态在交接中丢失，增加一层形式签字未必有用；若成员不敢提出异议，单纯要求更认真也不能改变条件。SDE 帮助将修订定位到差异序列与特征组织：在哪一环节信息被压缩，哪种权限阻碍纠正，哪项激励促成过早发布。有效修复需要改变这些条件，并检查下一轮是否真正改善。

修订本身也可能产生新成本。更严格审查可以提高可靠性，也可能延迟有价值信息；更多记录可以帮助问责，也可能增加负担和隐私风险。责任制度不能只追求零错误而不考虑可行性。应当区分不可接受的风险、需要降低的风险和可以公开承担的不确定性，并说明这些区分的理由。规范选择在这里不能由预测模型自动完成。

修复的完成也应有可检查的条件。不是所有争议都能获得共同满意，但可以明确错误信息是否已纠正、关键流程是否改变、相关人员是否获得说明和后续复核是否安排。若完成标准始终移动，责任可能变成没有终点的惩罚；若标准过于形式化，又可能掩盖持续后果。合理的结束需要兼顾事实改进、程序回应和当事人的正当诉求。

27.6 理由档案与判断的可修订性

记录决定时的理由，可以使后来的责任讨论摆脱纯粹结果倒推。理由档案不需要保存每个人全部思想，而应保留关键选择、可获得证据、考虑过的替代方案和主要不确定性。这样的记录能够说明为何当时采取某条路径，也能让后来发现具体误判。没有理由历史，制度容易只留下成功者的解释和失败者的辩护。

理由档案与监控必须区别。若所有讨论都被永久用于惩罚，成员可能不愿表达疑问，档案反而失去真实内容。记录制度需要明确用途、访问范围和纠正机制，使善意探索与正

式承诺能够被区分。涉及个人资料时，还应遵守相应的伦理与治理要求。本章讨论的是制度设计原则，不以一个哲学框架代替具体法律和专业规范。

研究伦理中的尊重、公正与风险收益评估，为这种制度提供了重要参照。《贝尔蒙报告》将原则与同意、风险收益和对象选择等实践要求联系起来。[37] 它提示，规范不能只停留在口号，而要进入实际程序。但程序也不能成为免除判断的形式凭证。某项签字如果没有建立理解和选择条件，就未必实现了其所要保护的价值。

理由档案还要区分一个人确实据此行动，与这个理由足以支持该行动。负责人当时真诚地认为赶期最重要，记录可以支持这一心理事实，却不能证明取消复核合理。需要继续问：赶期利益属于谁，风险由谁承担，是否存在保留关键检查的可行替代方案。若这些问题当时未被考虑，事后写出更周全的理由也不能直接冒充原有判断。可以如实保留原理由，再说明现在为何修订。SDE 使理由的显露、讨论过程和约束条件重新连在一起；规范论证则判断这种连接是否充分回应了相关利益。档案保存发生历史，正当性仍须逐项论证。

判断的修订也应保留版本关系。新证据使结论改变时，应说明哪些事实更新、哪些规范前提保持以及哪些理由被放弃。不能把所有变化都称为价值进步，也不能因曾经作出判断就拒绝修订。可问责的修订是一种能力：它承认先前有限性，同时让他人理解当前决定如何从过去而来。责任在这里表现为对历史的承担，而非对旧立场的机械忠诚。

档案还应允许异议保留。某项决定已经作出，不意味着少数意见从知识历史中消失。若异议提出了明确风险或不同解释，保留它能够帮助未来复核，也能防止事后把共同决策误写成全体一致。异议的记录不保证其正确，却保护了理论与制度学习所需要的差异。发生分析所重视的未决问题，在责任实践中由此获得具体形式。

27.7 共同规范怎样容纳合理分歧

伦理判断经常面对多种合理价值。透明有助于问责，保密可以保护个人；及时行动可以减少当前损害，审慎核查可以避免新的错误。冲突不能总靠增加事实解决，因为事实清楚之后仍需选择如何分配风险。SDE 能够显示这些价值在事件中怎样相互限制，却不能承诺存在一个普遍函数将它们换算为唯一答案。

共同规范的最低要求，是让决定能够向受影响者提供可理解的理由。理由不必让所有人满意，但应当回应相关利益，避免仅凭权力说事情就是这样。对于承担不同负担的人，

尤其需要说明为何这种分配可以被合理接受。发生理论强调位置差异，这在规范层面转化为让不同位置有机会提出相关理由，而不是把差异自动化约为某个总体指标。

假设研究小组必须在快速发布初步信息与等待完整验证之间选择。一个合理方案可能是先发布明确标注的不确定结果，同时限制其用途；另一个方案可能因误用风险较高而暂缓。选择需要结合实际后果、接收者能力和可逆性论证。不能从开放科学或谨慎原则任一口号直接得到答案。规范稳定性来自可公开说明的条件判断，而非对单一原则的无限扩张。

合理分歧也不意味着所有立场同样有力。忽略已经确认的事实、拒绝说明代价或将某些人的利益无理由排除，都会削弱立场。共同讨论需要最低的证据纪律和相互尊重。SDE若只强调不同主体各有显露，就可能滑向不可比较的相对主义；它必须同时要求解释之间能够共享事实、回应反例并承担后果。发生性应当支持讨论，而不是取消讨论的可能。

决策有时必须在分歧尚未消失时完成。此时可以通过明确暂时性、设置复核点和保留退出条件，减少不可逆承诺。可逆性并非总能取得，也不是唯一标准，但它可以使不确定中的行动更容易承担历史。若决定不可逆，就更需要事前说明证据和规范理由。责任的强度应与控制能力和后果结构相称，而不能用所有事情都复杂来平均稀释。

共同规范要能够承受这种分歧，可以把讨论落实为三个依次衔接的问题。先标明各方同意的事实及仍有争议的证据，再明确每个方案分别保护什么、把何种代价分配给谁，最后说明由谁依据何种共同认可的权限作出决定，以及新证据怎样触发复核。后一步不能补偿前两步的缺失：有权决定不表示其理由自动正确，少数意见未获采纳也不表示它失去今后检验的价值。制度稳定依靠的是这一套能够被重复实行的理由交换与纠错关系。规范如何实际形成，可以由发生分析解释；形成后的要求是否正当，仍须经受利益、权限和比例性的检验。

27.8 责任作为承接后果的制度能力

责任的发生性并不削弱个人行动的重要性。相反，它使个人选择的条件与后果更清楚。一个人在何种信息下作出决定，能够采取哪些替代行动，履行了什么承诺，又如何回应新证据，都成为可以分析的问题。人格评价可以有位置，但不能替代这些具体判断。把人简单分成负责或不负责，往往不如说明其在某事件中应承担哪项责任更有改进力量。

制度能力则决定个人责任能否被履行。清楚的信息接口、可使用的异议渠道、与职责匹配的权限和可执行的修复机制，使责任不只停留在道德劝告。若制度持续制造无法履行

的要求，又把失败全部归于个人，就会再生产失配。发生分析把这种循环展开，帮助研究者找到改变条件的具体位置，而不是只要求每个人更加努力。

在摘要案例中，责任进入下一轮，可以表现为一组可检查的承接安排。分析者交付文本时附明证据状态和适用范围；审阅者核对摘要措辞是否超过资料支持，并保留未解决的异议；发布者确认必要审阅已经完成，对临时例外说明理由；错误出现后，由约定的角色整合更正，而不是让发现者独自承担全部工作。这些要求来自本例的职责与风险，不能机械推广为所有研究都须采用同一流程。其共同结构是，每个关键接口都有明确承担者、可获得的信息和能够执行的动作，责任因而进入决定的形成条件。

这套安排的成效需要在后续事件中检查。如果成员仍只能在发布后才看到关键文本，纸面上设立了异议渠道也未形成可使用的纠正机会；如果证据标记清楚，但接收者的任务迫使其把探索结果当作确定决策，问题便出现在用途接口，不能继续只培训撰写者。相反，若适时获得信息的成员能够指出限制，发布者据此收缩措辞，接收者也能调整使用范围，便取得了具体的制度改进证据。成功不等于以后永不出错，而是已知风险得到合乎比例的控制，新错误有能够启动的解释和修复路径。

从爱与包容到公共研究责任，连续的是对历史承接的要求，不连续的是各领域的规范结构。私人关系中的宽恕不能直接替代制度问责，制度中的程序也不能完全规定亲密关系。共同发生允许我们比较行动、解释与后效怎样连接，同时要求每次迁移重新说明义务来源。只有保留这种差别，伦理部分才不会把所有生活装入同一套抽象流程。

责任分析由此取得一个可迁移的结果：评价谁的行动，应当说明义务来源、当时可用的知识与选择；安排现在的修复，则另行考察权限、能力与合理成本。进入教育评价，这一区别尤其重要。评价者负责使分数与所声称的能力范围相称，决定分数用途的机构负责说明机会分配的理由；二者都无须预见学生未来发展的每个细节，却需要回应已经可见的使用限制。第 28 章由此追问能力怎样在任务、工具和协作中形成。只有先把能力发生的单位说清，才能进一步讨论评价者究竟有根据地判断了什么，以及这种判断可以支持哪些后续安排。

第 28 章 教育评价与能力生成

28.1 同一个答案背后的不同能力

两名学习者交出同样正确的答案，一人能够说明关键理由，另一人依靠提示完成步骤。两人的当前作品可以同样合格，却不能由此推出他们在下一项任务中具有同样能力。相反，两名学习者得到同样低分，也可能因为不同困难：一人理解概念但读错要求，另一人记住程序却无法辨认适用条件。教育评价面对的首要问题，是如何从可见表现形成关于能力的判断，同时保留改变这一判断的机会。

能力不是藏在一个人内部、可以被一次考试完整提取的固定总量。它也不是每次情境变化都完全重新生成、因而无法持续认识的幻影。更合适的起点是，把能力理解为学习者在一定任务、资源与责任条件下能够反复组织行动的稳定性。这种稳定性可以逐渐扩展，也可能受到支持条件变化的影响。评价需要说明自己观察到了哪一种稳定，而不是用一个无条件的标签覆盖整个人。

任务条件并非无关的干扰。是否允许查阅资料、是否需要解释、是否必须在时间限制内完成，以及是否可以与他人或工具协作，会改变被评价的行动。一个强调独立回忆的任务，与一个强调资料比较的任务，可能都具有教育价值，但它们支持不同能力判断。把它们混成同一总分，必须有明确用途和论证，不能仅因为数字可以相加就认为构念也已统一。

教育发生研究与经济反馈研究的机制不同。这里的核心不只是判断影响激励，而是练习、解释、错误修正和迁移可能改变学习者后来能够完成的行动。评价本身也可能成为学习事件。学习者经过一次提问，可能发现此前未意识到的差异；经过一次解释，可能重新组织知识。研究者因而需要同时区分测量已有能力与促进新能力，而不是将两者混为一种分数变化。

这些任务之所以不能仅靠答案归类，是因为帮助方式会改变答案的能力含义。教师直接给出比较基准，与学习者自主选择同一基准，最终算式可以一致，组织功能却由不同位置承担；若提示经过练习被学习者自行重建，同一算式又获得新的证据意义。这里的特征纠缠不是知识、工具和个人三个因素简单相加，而是它们共同规定何种成功能够归于学习者。研究需要把最终显露放回提示、尝试和修订的差异序列，才能辨认这种归属怎样发生变化。

因此，评价的目的不能停在给人排序。排序在某些选拔任务中可以有有限用途，但学习支持还需要找到可以改变的环节。一个有教育意义的判断，应当能够回答下一步怎样帮助学习者，并说明什么后续表现会支持或修正当前解释。由分数走向能力生成，并不是放弃标准，而是让标准与发展之间建立可追踪的关系。

28.2 表现提高与学习发生

当前任务做得更顺利，不必等于能力已经获得长期改变。熟悉题面、刚刚看过答案、得到提示或采用工具，都可能提高即时表现；延迟以后能否独立完成，以及面对新形式能否选择恰当方法，才涉及另外的证据。教育研究需要把这些结果放在同一学习历程中，却不能用其中一个代替其他结果。

Roediger 与 Karpicke 的实验比较了学习材料后的重复阅读与回忆测试。短延迟和较长延迟下，两种安排的相对表现不同，回忆测试在其研究的延迟保持任务中显示优势。[53] 这一结果不能推广为任何考试都优于任何教学，但它说明测验可以参与记忆形成，学习者当下的流畅感也不必与后来保持完全一致。测量和干预的关系必须在具体任务中检验。

设想一名学习者在例题旁完成一道相似题，随后在没有例题时处理结构相同但表面不同的问题。第一项成功说明他能够在支持下进行某种操作，第二项则检查他是否辨认并保留了关键关系。如果两项表现分离，教师不应立即把前者归零，也不应把后者解释为懒惰。需要进一步判断支持究竟替代了哪一步，以及哪一步尚未转化为学习者自己的组织能力。

能力生成可以从可撤除支持的变化中观察。开始时，提示帮助学习者识别问题；随后，提示减少到关键检查；最后，学习者能够自行提出检查。这里的进步不是简单少用了多少帮助，而是某项控制功能逐渐被学习者承接。若撤除支持过早，任务失败可能只说明教学安排不合适；若支持始终替代全部判断，完成作品也可能掩盖尚未形成的能力。

学习历史还会改变同一任务的意义。第一次遇到一题时，它可能要求探索；多次重复以后，它可能主要要求识别熟悉模式。若研究者仍把两次分数当作同一种测量，就会混淆练习效应与更广的理解。迁移任务应保留目标结构，同时改变那些可能被机械记忆的表面线索；改变过多又可能引入新的知识要求。任务设计本身因而承担构念辨认的工作。

能力保持还需要与任务机会相联系。某种方法长时间没有被使用，后续表现下降，不能只凭一次结果确定它已经完全消失；重新提供适当线索后能否恢复，也是一种证据。但

恢复快并不等于可以替代所有持续掌握的要求。例如，某项任务若要求在没有准备的情况下作出可靠判断，临时再学习就不足以满足目的。教育评价应把保持、恢复和迁移分别放回实际责任，而不为所有能力规定同一种时间标准。

对学习过程的记录不必追求内心全景。可以保留首次答案、关键错误、采用的反馈和延迟任务结果，形成最短但足够的历程。若这些资料已经能够指导教学，继续增加日志未必有益；若两种解释仍不可区分，再设计一个针对性任务。教育发生学的目标，是让记录服务于下一步学习，而不是把学生的每个行为永久转成数据。

28.3 独立表现与人机协作

AI 工具使作品质量与个人掌握之间的区别更加明显。学习者可以利用工具生成解释、检查步骤、寻找资料或完成草稿，这些协作本身可以构成重要能力。但能够交付一份高质量作品，与能够在没有工具时独立完成核心推理，是不同判断。教育制度若只承认后者，会忽略现实协作；若只看最终作品，又可能看不见个人尚未形成的能力。

Bastani 等人在高中数学场景中比较了不同 AI 辅助安排。研究显示，练习阶段的表现改善不保证撤除工具后的独立表现改善；带有教师设计提示约束的工具，与较直接提供帮助的工具，产生了不同的学习后果。该研究主要考察特定学校和任务中的短期学习，不能据此断言所有 AI 必然损害长期能力。[54] 它为本书提供的直接启示，是工具设计、学习活动和评价窗口必须同时说明。

可以把教育任务分为三个相连但不合并的窗口。独立窗口检查学习者自己能够完成什么；协作窗口检查他怎样组织工具和资料完成任务；迁移窗口检查此前协作是否改变了后来的个人行动。三个窗口都应保留清楚的允许资源和评价标准。若只进行一次协作任务，就不宜据此宣布个人能力已经增长；若只进行独立测试，也无法完整评价人机合作能力。

协作能力并不是熟练输入提示语的单一技巧。它包括界定问题、选择工具、比较建议、发现错误、决定何时停止依赖，以及对最终结果承担解释责任。一个人可能善于获得流畅答案，却不能检验其适用条件；也可能提示表达简单，但能够准确识别工具遗漏。评价应围绕这些不同功能设计任务，而不能把对话长度或工具使用频率当作能力本身。

一个具体的分析性任务是提供两份看似合理的解法，其中一份包含可以检验的错误。学习者可以使用工具，但必须说明接受和拒绝的理由。这个任务检查的是核验与责任承接，而不只是生成速度。随后可以撤除工具，要求学习者在新题中自主设置同类检查。如

果后者改善，才有证据讨论检查能力的迁移；若没有改善，协作收益仍然存在，但其归属需要准确表达。

工具撤除也不应成为无条件的最终标准。有些真实任务本来依赖仪器、资料或协同，要求全部独立完成可能改变了能力构念。教育需要决定哪些基础能力必须由个人保有，哪些能力可以稳定地分布在人与工具之间。这个决定涉及课程目的和责任要求，不能由技术便利性单独决定。尤其在工具可能出错或不可用时，个人需要保留足够的判断与恢复能力。

发生学由此把人机协作作为一个可以分析的新行动单位，同时保留个人发展的问题。研究既看系统产出，也看学习者是否获得更好的问题组织与核验能力。二者可能共同提高，也可能分离。将分离如实记录，比把协作一概称为作弊或一概称为增强，更有助于形成可持续的教育安排。

28.4 从错误结果到可改变的困难

错误不是一个需要被统一清除的容器。它可能来自概念关系尚未形成，也可能来自任务理解、表达方式、计算操作或检查遗漏。相同错误答案有时由不同路径产生，不同错误答案也可能来自同一困难。教师若只记录对错，能够完成基本评分，却未必获得足以指导教学的信息。

设想一道要求比较两个比例的题。一个学习者把分子差当作比例差，另一个学习者理解比例却抄错数据，两人都得到错误结果。给他们相同的重复练习，可能产生不同效果。第一人需要重新辨认比较单位，第二人需要建立资料核查程序。教师可以通过一个简短追问或对照任务区分两种解释，而不必立即进行复杂心理诊断。

困难的解释应以可行动性为标准之一。说某人缺乏能力，通常不能直接指导下一步；说他在改变单位时没有保持同一比较基准，就能够设计新的任务。解释越具体，越需要接受后续检验。若经过针对性支持以后预期变化没有出现，应修订解释，而不是把未改善再次归因于学生整体不足。教学判断由此成为可以失败的机制假说。

也要防止从日志直接推断内心。长时间停顿可能表示思考，也可能表示分心或技术困难；频繁修改可能表示探索，也可能表示不理解要求。行为指标首先是线索，只有结合任务、作品和学习者说明，才可能形成较可靠的解释。任何单一特征都不宜被固化为个人本质，尤其当它会进入长期评价和资源分配时。

学习者应有机会参与困难的命名。教师观察到的错误，与学生实际遇到的障碍可能不同。学生的说明也未必完全准确，但它可以提供新的候选机制。通过让学生复述题意、说明选择理由或比较两个方案，双方能够共同检查理解。这里的对话不是给既定诊断增加礼貌形式，而可能真实改变研究对象：原来被归为计算困难的问题，可能转为语言理解或任务目标的歧义。

错误档案应保留变化，而不只是累积失败。某项困难何时出现，经过什么支持得到改善，在哪类新任务中再次出现，能够显示能力形成的范围。已经克服的错误如果仍以永久标签存在，就可能误导后续教学。教育记录需要版本和退出条件，使学生能够从旧判断中成长出来。

这一过程与经济残差分析具有形式上的联系，但其机制不同。教育中的剩余差异不是只等待被模型解释，它还可以通过教学行动被改变。解释与创造在这里直接相遇：一个好的困难诊断，应帮助形成此前尚不存在的能力。研究的成功既体现在判断更准确，也体现在学习者获得了新的可持续行动。

28.5 反馈怎样被学习者承接

教师给出了反馈，并不等于学习者已经获得学习。反馈可能被看见却未被理解，被理解却不知如何使用，或者被直接照抄而没有改变后续判断。研究需要追踪反馈从表达进入行动的过程。仅以反馈是否发送、学生是否点击或修改次数多少评价教学，会把传播记录误当成能力生成。

可以区分几种反馈功能。指出结果错误帮助定位差异，说明错误原因帮助建立关系，提供下一步问题帮助重新组织行动，要求自我核查则帮助形成控制过程。不同功能适合不同困难。若学习者尚未理解题意，直接给出完整推导可能增加负担；若他已经理解核心关系，只需一个定位提示，就不必替他完成全部过程。

反馈的时机也重要。过早提供答案可能终止有价值的探索，过晚提供必要信息又可能使学习者反复练习错误路径。这里不能预先规定一个对所有任务有效的时间点。应根据任务难度、学习者已有能力和错误后果安排，并通过实际表现调整。发生学要求把时机作为机制条件，而不是把反馈内容单独拿出来比较。

比例任务可以把承接分得更细。学生先给出比较结果，教师只问两边是否使用同一基准，不提供完整步骤；学生修订以后，再解释原来比较为何不充分。随后更换数值和呈现顺序，观察他是否在没有同样提问时主动建立共同基准。如果只能在原题修正，提示可能

只是帮助完成局部动作；若能在新题自行提出检查，才支持相应组织功能被承接。这个假设教学序列需要实际检验，也需要与同等练习机会下的常规讲解比较，才能区分特定提示的作用与一般练习收益。

评价本身也会回写学习。学生知道评分只看最终答案，可能采取与知道评分关注解释和核验时不同的策略。研究者不能因此取消公开标准，因为透明标准本身也可以帮助学习。需要判断的是标准鼓励了哪些行动，这些行动是否与课程目标相容。若学生只是学会展示被奖励的表面步骤，就应检查任务是否真正要求理解，而不是责怪他们回应制度。

反馈关系还需要保留尊严与边界。学习者可以被要求解释任务相关判断，却不应为了证明认真而披露全部私人经历。教师的判断应针对可改变的行动，并允许学生提出异议。学习不是将一个人完整暴露给评价系统；恰当的可见性应与教育目的和必要性相匹配。这种限制本身能够支持更可信的合作。

反馈是否被承接，还可从学习者主动提出的问题观察。比例任务中，他原先只询问答案是否正确，后来开始询问两组资料是否采用同样单位，这一改变提供了组织功能转移的线索。但主动使用某个术语仍可能只是模仿教师话语，需要看他能否在单位已经相同的题中停止无必要的转换，在单位不同的题中解释为何转换。承接并非把一个新步骤加入所有任务，而是逐渐能够辨认这个步骤何时有用、何时无须使用。反馈研究应追踪这种选择性形成。

28.6 可比条件与能力范围

相同题目和相同时间限制提供了一种形式上的可比性，但评价首先需要明确，哪些差异属于所要考察的能力。若某项无关的表达要求、技术条件或资源差异主导结果，分数就可能偏离目标；若支持替学习者完成了目标操作，又可能把系统能力计入个人能力。可比性的关键因此是任务与结论相称，而不是机械维持一切条件相同。

例如，评价一个人能否比较两组资料，与评价他能否迅速阅读一段复杂说明，是不同任务。若比较是目标，可以把相同信息整理成清楚的表格，减少理解题面所需的额外步骤；若整理资料本身也是目标，直接提供表格就改变了需要检验的行动。支持的同一种形式，在两个任务中可能承担相反的评价意义。研究应列出从材料到结果之间哪些步骤必须由学习者组织，哪些步骤可以由题目预先提供，才能说明分数的归属。

回到比例比较的例子，可以设计三种条件：只给分数，给出相同单位的图示，或给出已经写好的比较步骤。第一种检验学习者能否选择表示与方法，第二种帮助显露在单位已

明确时能否作出判断，第三种主要观察能否执行已提供的步骤。三项结果不是同一能力的简单高低等级。有人能自行选择方法，却不熟悉某种图示；有人能沿步骤算对，却不知道为何必须使用共同基准。研究应解释这种交叉，而不是把条件越有帮助得分越高当作足够结论。

由此可以描绘某项能力的范围。哪些变化仍保持成功，哪些变化需要少量提示，哪些变化导致方法失选，应分别记录。支持减少以后成功保持，说明学习者可能接过了相应组织功能；支持减少以后失败，则需要检查撤除的究竟是提示、知识资源还是任务本来需要的工具。范围判断比一次通过更严格，因为它要求明确可承受的变化；也比无条件能力标签更有教育用途，因为它指出下一项任务需要改变什么。

比较学习者时，还须区别过去学习机会与当前表现证据。较多练习可能形成真实的熟练能力，不能因为机会不相同就否认已经形成的掌握；但当前熟练也不能直接证明另一人无法在适当学习后达到相同水平。评价已有掌握与估计学习后的可能范围，是两项任务。后一项需要观察支持怎样被采用、错误怎样改变和新的任务怎样被处理，不能从起点分数直接外推。

这里的测量结论必须对用途作出承诺。若结论用于选择下一项练习，可以保留较多局部和暂时判断；若用于说明学习者已经能独立承担某种操作，就需要在相应条件下反复检验。不同用途要求不同充分性标准，但都应让材料、任务和结论保持可追踪关系。第 30 章将进一步讨论评价已经造成机会损失时如何修复；本章要先完成的工作，是使能力判断本身不超出所观察的行动范围。

28.7 研究设计中的学习历程

要检验一种教育方法，不能只比较教学结束时的一次成绩。至少需要明确初始水平、实际经历、即时表现和后续迁移之间的关系。不同研究可以选择不同重点，但结论不得超出所观察的窗口。短期任务成功不能直接证明长期保持，一个课程中的改善也不能直接代表所有学科。

在可行条件下，可以比较无新增支持、一般提示和针对性提示三种安排，使用共同的核心任务，并控制练习机会。随后安排延迟任务和适度迁移任务，检查效果是否持续以及是否依赖原有题面。若采用班级作为分组单位，应在分析中处理班级共同影响；若学生之间共享信息，也应记录可能的组间影响。方法需要与实际教学结构一致。

研究还应区分实施失败与理论失败。教师没有按设计提供关键支持，或者工具无法使用，可能说明实施条件不足；支持被可靠执行却没有预期效果，则更直接要求修订机制。但不能把所有不理想结果都归于实施不充分。应在研究前规定关键环节的最低要求，并据实记录，让区分具有可复核依据。

个人差异不能只在平均效果不理想时才被发现。可以预先选择与机制有关的差异，例如已有知识、任务理解或工具经验，检验支持是否对这些群体产生不同作用。事后不断拆分直到出现某个有利结果，会使发生解释失去约束。探索可以继续，但应与事先检验分开，新的发现需要在后续资料中确认。

新方法还必须面对强基线。更细的过程记录若只是重复教师已经能够从作品中判断的信息，就未必值得增加负担。可以比较常规作品评价、少量关键过程记录和完整记录的实际收益，观察哪一种足以支持有效教学。若简洁方法达到相近效果，应优先保留其可执行性。发生学不以记录密度证明理论深度。

教学研究中的失败结果具有特殊价值。某种提示对当场完成有效，却没有迁移，可以帮助定位功能尚未被学习者承接；某种解释提高信心，却没有提高核验能力，可以提醒调整反馈；某种工具对部分任务有益，对其他任务无益，可以明确适用边界。这些结果比一个无条件的成功口号更能支持后续创造。

最终应让另一位教师或研究者能够根据任务、材料、支持条件和结果记录，理解研究如何得出判断。可继承性并不要求所有学校完全复制同一程序，而要求关键机制和变动理由能够被说明。只有当新条件下的修订可以追溯，教育经验才有机会从个别实践成为共同知识。

28.8 从能力证据到下一项任务

评价的最后产物不应只有一个分数，还应包含下一项值得检验的能力差异。这并不要求建立无穷尽的个人档案。就比例比较而言，一份精简记录可以分别说明：学习者是否能够执行已有步骤，能否在新题面选择方法，能否发现一种看似合理的错误。三项证据共同限定当前判断，又给后续教学不同方向。会执行而不会选择，需要改变任务表面并要求说明选择；会选择却常忽略资料错误，则需要让检查成为行动的一部分。

这些证据之间不能相互冒名。一份经提示修订后完全正确的作品，说明在相应支持下完成了任务；若要进一步声称学习者已经理解，应看他能否说明关键关系，并在未经同样提示的新任务中使用它。口头说明也不是绝对标准：有的操作能够先于完整表达稳定形

成。此时可以增加结构变化后的操作比较，检验成功是否只依赖熟悉线索。能力证据应随所提主张增加，不以一种表现形式垄断理解的定义。

设想教师最初认为一名学生只是忘记核查单位，于是给予单位提示；学生能改正当前答案，却在换一组数值后继续把分子大小当作比例大小。新结果会迫使教师把困难从检查遗漏改判为比较关系尚未建立，并设计共同基准的任务。随后若学生能自行提出共同单位，才有理由更新能力判断。这个假设序列展示了双向学习：学生的行动发生改变，教师对改变机制的解释也经由学生的新表现被修订。

对人机协作，可以沿用同一种证据分工。保留最终作品之外的一处关键选择，记录学习者为何接受一个建议、拒绝另一个建议，再观察工具撤除或工具更换后能否维持相应核查。若能力只在原工具给出特定提示时出现，结论应保留这一条件；若学习者能够自己重建检查，才更有理由把进步归于个人承接。工具的版本与帮助方式影响任务身份，记录它们是为了使能力归属清楚，而非尽可能保存所有对话。

下一项任务还应受到成本约束。若一个简短对照题已经能区分计算失误与概念困难，就不必先收集长时间行为日志。若两种解释在现有材料下仍给出同样预测，研究者应提出能使它们分离的任务，或者明确暂时无法判断。未知不是对学习者的缺陷认定，它是当前评价工具的界限。这样的节制使发生分析能够进入实际教学，而不把教师变成永久记录员。

学习者也需要看懂这些判断。与“综合能力不足”相比，说明“能够在共同单位已经给出时比较，但尚需自行选择表示的方法”，更容易形成下一次尝试。表述应既保留已经取得的掌握，也清楚标明尚未证明的迁移，避免鼓励性语言掩盖困难，或用一次困难覆盖全部进步。可理解的评价使学习者能够参与检验自身的方法，而不只是等待外部排名。

本章得到的能力，是在任务变动中仍能组织某些行动，并能通过支持承接扩大其范围的稳定性。评价抓住显露，教学改变产生显露的关系，迁移检查改变能否在新任务中继续起作用。三项工作既相接又有各自证据。它们把教育发生与价格激励区别开来：学习的结果之一，是行动者后来能自行作出此前尚不能作出的区分。

而“能作出区分”仍留下一个问题：为什么某种说明算作理解，某个回应足以表示共同掌握，一份作品允许哪些不同解读？正确率不能独自回答这些问题。第 29 章从表达与回应的具体关系继续研究，考察意义怎样既保持多元，又受到材料和解释责任的约束。

第 29 章 意义的多元发生与解释

29.1 同一句话怎样成为不同事件

“可以了”出现在排练中，可能表示这一遍已经达到要求；出现在一场争论中，可能要求停止；出现在共同编辑一份文本时，可能批准采用当前版本。字面形式相同，参与者接下来能够做什么却不同。若只把意义当作词语与定义之间的固定对应，就难以说明这些差异；若认为意义完全取决于个人感受，又无法解释为什么有些理解会被共同纠正。

意义发生首先需要辨认事件。谁在什么关系中说出这句话，对方能够接触哪些先前内容，当前共同任务是什么，后续回应怎样显示理解，这些条件共同规定了话语的作用。这里的条件并非任意背景。排练中已经约定的停止信号、编辑中形成的批准权限、争论中尚未解决的问题，直接影响一句话能够承担何种行动。

但意义也不能完全等同于实际后果。一个人误解了指令而采取行动，不能因为后果已经发生就认定误解是唯一正确意义。表达者的意图、公共语言规则、可获得的语境和接收者的理解，可能彼此不一致。解释需要在它们之间提出理由，而不是选择一个层次吞并其他层次。后果是证据，也可能是需要纠正的问题。

本章把意义理解为某种差异对一个行动、理解或生活过程获得可承接的重要性。这个表述有意保留不同尺度：一个词的指称、一次承诺的作用、一段作品的解释和一件经历对个人生活的重要性，并不属于完全相同的对象。它们可以相互连接，但不能被压成同一种可量化反应。多元意义首先意味着研究任务有差别，其次才涉及同一材料允许多种解释。

共同编辑中的“可以了”可以把特征纠缠说得更精确。由负责定稿的人对屏幕上标明的版本说出，它可能构成采用该版本的批准；由无权定稿的同伴说出，则可能只是质量评价；即使说话者有权限，若刚才讨论的只是一个段落，也不能直接推得整稿获准发布。权限、所指对象和批准范围不是三个可以分别加分的因素：所指对象决定权限在何处适用，权限又改变这句话对该对象能够完成何种行动，范围则限定后续行动应承接哪项责任。E 所要求解释的正是这种相互规定，而非凡有语境就有纠缠。同一句话在后续发布或继续修改中获得不同承接，使这一区分能够从实际材料中显露。

意义的发生性因此既反对把解释封死在一个无历史定义中，也反对取消公共约束。稳定的用法使交流可能，新的情境又可能要求重新区分。解释者的任务是说明稳定为何成

立、变化从哪里进入，以及新的理解能否被材料和后续实践支持。意义不是只有在发生争议时才需要研究；日常理解的顺利承接，同样包含值得说明的形成条件。

29.2 字面约束与推论空间

语言具有相对稳定的词汇和句法关系，这些关系为理解提供限制，却不穷尽一次表达所传递的内容。一个人回答“灯还亮着”，在不同对话中可能说明房间有人、提醒关灯或表达继续工作的意图。理解者需要结合当前问题推论，但推论并非无中生有。它受到字面内容、共同知识和交往目的约束。

Grice 关于会话的分析区分了所说内容与会话含义，并研究听者如何依据交谈目的和相关假定推得未直接说出的内容。[55] 这一近邻提醒发生学，情境参与意义早已有细致理论表达。新增任务不能只是把词语放回语境，而应进一步追踪语境如何在互动中被修改，哪些推论后来成为共同承诺，以及这种承诺怎样影响下一次理解。

可以用一个分析性例子说明。共同工作者询问某项材料是否已经准备好，回答者说“最后一页还在修改”。字面报告是那一页的当前状态，听者据此推得整份材料不能使用，还需要一个前提：本次使用必须包含已经完成的最后一页。若此前约定这一页只是暂不采用的附录，主体部分能够先用，推论便改变，而字面报告的真伪条件没有因此改变。回答者是否有权决定使用，可以进一步影响这个回答能否构成延期决定，却不能使实际上未修改的一页变成正在修改。用途、事实与决定权限由此可以共同进入意义，又不相互冒名。

推论还具有可撤回性。说话者可以补充说明自己只是报告进度，并未表示整体不可使用。听者也可以提出核查，询问是否意味着需要延期。这样的修正使意义保持开放，却不意味着先前理解毫无根据。研究者可以区分合理但被新信息修正的理解，与从一开始就忽略可获得材料的误读。两者在责任上也可能不同。

字面约束的重要性在争议中更加明显。解释者不能为了使一种观点成立，任意改变文本中反复使用的概念，也不能把明确排除的条件重新塞回命题。另一方面，严格重复字面也不保证理解充分，因为反讽、暗示或任务约定可能改变表达的作用。解释的严谨性在于同时处理文本证据与实际使用，而不是预先选择字面主义或无限联想。

意义研究因此应明确推论链。哪些内容直接来自表达，哪些依赖共同材料，哪些是解释者提出的假说，哪些得到后续回应支持，应可以区分。这样的说明不必把日常交流变成

繁琐证明，但在学术解释、教育判断和公共争议中，它能够减少把直觉当作唯一答案的风险。解释能力不是猜中隐藏意图的神秘本领，而是组织可讨论理由的能力。

29.3 共同理解怎样被建立

一句话被说出，不保证交流已经完成。听者可能没有听清，也可能理解了词语却不明白当前用途。表达者则可能误以为对方已经接受，从而继续推进。共同理解需要某种确认，但确认不必总是直接复述，它可以通过恰当回答、继续行动或提出针对性问题表现出来。意义因此具有互动过程，而不仅是单向传输。

Clark 与 Brennan 把共同理解的建立作为交流中的协作过程，强调参与者需要获得足以满足当前目的的理解证据，沟通目的和媒介条件会影响这一过程。[56] 这里的“足够”非常重要。交流并不要求双方拥有完全相同的全部知识，而要求当前行动所需的理解得到相应支持。一次约定时间的对话与一次复杂任务交接，需要不同程度的确认。

设想两人共同组装一件设备。指令中的“左边”可能以操作者或设备正面为参照，一个简短指向能够让当前安装成功，却未必已经建立可沿用的共同参照。接收者也可能只是模仿指向，并没有理解下一条指令中的“左边”仍按什么确定。若下一步需要转到设备背面，双方可以先确认应保持设备朝向，还是改用操作者朝向；接收者能够指出变化后的目标位置，比一次点头更直接支持这项理解。这里不是要求所有任务都增加测试，而是说明证据充分性随接下来的协作要求变化。共同参照一旦建立，会改变后续省略能够省到什么程度；省略发生误读，又可能要求双方重新说明参照。

媒介会改变确认方式。面对面交流可以利用指向、视线和即时追问，书面材料则需要更明确地保存参照和条件。这并不意味着一种媒介总比另一种优越。书面记录有利于回看，现场交流有利于及时修正；任务不同，适合的组合也不同。研究应观察媒介提供了哪些可用证据，又使哪些误解更难显露，而不是把技术形式直接等同于理解深度。

共同理解也不等于同意。两人可以准确理解彼此观点而继续分歧，或者在没有真正理解时表面赞同。若制度只记录是否同意，就可能把权力压力、礼貌和理解混为一谈。评价沟通质量，应允许参与者表达不理解、不同意或暂不判断，并区分这些状态。让分歧可表达，有时比追求快速一致更有助于形成可靠共同基础。

理解确认还需要考虑后果。对低风险、可逆的小事，过度确认可能妨碍行动；对后果重大或难以撤回的决定，更充分的复核具有理由。这个差别来自任务责任，而不是某些人

天生值得被怀疑。一个公平的程序应根据后果和不确定性安排确认，而不是任意把额外证明负担加给特定群体。

发生学由此把共同意义理解为持续维护的成就。它具有稳定性，因为参与者能够依据已有理解继续行动；它也具有修订性，因为新的回应可以暴露差异并改变共同基础。研究者应保存关键修复节点，而不只是最终一致的结论。那些短暂的误解和纠正，往往最能说明意义怎样被形成。

29.4 文本与作品中的多元解释

文学、艺术和哲学文本的解释，不完全以实现某项即时协作为目的。作品可以在不同历史和个人经历中显露新的关系，同一细节也可能进入不同解释结构。承认这种多元性，并不要求所有解读同样有根据。解释仍需要面对作品材料、结构、语用和历史证据，只是评价标准可能不同于事实问答中的唯一答案。

可以把解释分为几个任务。确定文本说了什么，需要仔细处理语词、句法和上下文；重建作者在某一争论中的行动，需要相关历史材料；说明作品对当代读者的启发，则还包括新的问题关联。三个任务可以共同出现，但不能彼此冒名。一个当代创造性解读未必是对作者原意的准确复原，准确复原也不自动穷尽作品的后续意义。

构造一个极短的作品情境：人物一直在门边等待，前段说明门从里面反锁，末段写人物握有钥匙，却把手从锁前收回。这些细节不是引自某部作品，而是用来比较解释的分析材料。“人物仅因没有开门手段而被迫等待”的解释，便不能忽略钥匙在手这一安排；它必须被放弃，或补充文本能够支持的其他限制。把等待理解为对相遇的犹豫，可以较好地容纳收手动作，却仍须解释为何反复描写门与外部声响。空间限制与时间经验的解释也可能各有贡献，但不能以关注层次不同为由回避自己已经作出的事实判断。多元解释可以共享材料而组织出不同关系，不能任意增删使解释难堪的细节。

解释还可以改变下一次阅读。在这个构造情境中，先把门看作外部障碍，读者容易关注门是否坚固；注意到人物握有钥匙以后，收手的先后次序与门外声音便可能成为关键。解释改变了哪些差异值得继续查看，新的材料又应能反过来改变解释。若后来出现明确交代，说明钥匙并不匹配这扇门，关于能够开门却不愿开门的判断就需收缩，不能把匹配和不匹配都宣布为同一个解释的证据。解释可以容纳复杂性，但每次容纳应说明新增假设及其材料根据；否则，解释越周全，实际承担的失败风险反而越小。

因而，较好的解释应说明自己排除了什么、不能解释什么，以及与其他解释的关系。有些解释相互竞争，因为它们对同一细节提出不相容判断；有些解释则关注不同层面，可以并存。研究者需要先确定争论是否真正共享问题，避免把不同用途的解读强行排列为唯一胜负。解释之间的关系本身也是研究对象。

作品的形式约束尤其不能被忽视。节奏、顺序、叙述视角和材料安排，会影响意义怎样显露；把内容概括成一个主题，可能丢失作品通过形式创造的差异。同样主题的两个作品，未必具有同样意义，因为读者经历的注意、期待和修正路径不同。发生学在此的任务，是把这些路径与整体解释连接，而不是只为内容增加情境说明。

这种分析也保留创造的位置。新的解释可能建立此前没有被明确提出的联系，但创造性需要与对象回应相遇。作品不是任由理论展示的空白屏幕，解释者也不是被动提取预存答案的机器。二者在具体阅读中形成新的可讨论结构，而这种结构是否值得保留，要由材料支持、解释能力和后续讨论共同检验。

29.5 个人经历如何获得意义

一件旧物对某个人很重要，对另一个人却只是普通物品。其重要性既不能完全从物理材料中推出，也不必被理解为与对象毫无关系的私人投射。旧物可能连接一段共同经历、某个已经作出的承诺或一个仍未完成的生活计划。意义在对象、记忆和持续行动之间形成，并在后来经历中变化。

设想一个人保存早年学习时使用的笔记，后来把一次失败重新理解为转向的起点。这个新理解至少包含三种不同主张：当时具体发生了什么，当时的自己怎样理解它，以及现在愿意怎样承接它。笔记日期、当年的文字和他人记录可以约束前两种主张；今天从中获得新的方向，并不能证明当时已经抱有同样计划。第三种主张也不是完全不受理由约束：若把转向说成独自完成而抹去他人的实际帮助，它便包含可以核查和争论的事实；若只是说明自己现在珍视哪段经历，当事人的自愿说明具有无法由外人替代的位置。个人意义能够更新，事实责任不会随之消失。

个人意义也不必总以完整叙事出现。一个习惯、一种声音或一项共同活动，可以在没有明确故事时承担重要性。研究者若只接受语言充分的自我解释，可能低估难以表达的经验。但表达困难也不能成为任何外部解释自动成立的理由。可以通过具体选择、持续实践和自愿说明获得有限证据，保持对未知部分的尊重。

意义改变与情绪变化有联系，却不等同。某件事令人愉快，不保证它在长期生活中重要；一项重要承诺也可能伴随困难和不确定性。将意义压成即时积极感受，会遗漏责任、牺牲和持续投入。反过来，不能因为某人投入很多，就替他宣布该安排必然有价值。个人有权重新评价过去承诺，并在必要时改变生活方向。

这里需要特别谨慎地处理第三人解释。教师、研究者或组织可以帮助一个人看见新的关系，却不应把自己的说明强加为对方生活的唯一真相。一个外部故事即使很连贯，也可能忽略当事人认为关键的差异。尊重并不要求无条件接受任何自述，而是要求解释能够接受当事人的纠正，并区分帮助理解与替人决定。

个人意义的稳定也不应被浪漫化。有些长期解释支持行动与关系，有些解释可能持续限制新的可能。是否需要改变，不能仅由稳定时间长短判断。可以观察某种解释是否容纳新证据，是否允许承担合理责任，是否使人能够采取与自己认可的目标相容的行动。这些判断仍需谨慎，不应转化为未经依据的心理诊断。

发生学在这一层面提供的，是对生活历史的可修订承接。过去没有因为新的解释而消失，未来也不必完全由旧标签决定。意义的形成允许一个人保持连续性，同时重新组织经历之间的关系。它与教育能力相连，因为理解自己怎样学习、怎样改变，可能成为继续行动的资源；但意义的价值并不能只用表现提高来衡量。

29.6 分歧 翻译与解释权

意义分歧有时来自资料不足，有时来自概念边界不同，也有时来自真实价值冲突。若把所有分歧都解释为一方尚未听懂，就会把共同理解变成单向服从。研究应先判断双方是否在谈论同一问题，使用的词是否具有相同作用，以及各自认可什么样的理由。只有把分歧定位，才可能选择澄清、协商或保留差异。

翻译使这一问题更加明显。一个表达在另一种语言中可能有多个近似说法，每一种保留不同关系。翻译者需要在字面、语用、形式与用途之间作出选择，无法仅凭单词替换保证全部等价。这不意味着翻译不可能，而意味着等价需要相对于任务说明。技术操作说明与诗歌翻译，对相同与差异的容许范围并不一样。

可以设想一个跨专业合作。一个团队用“模型”指预测用的数学关系，另一个指检查空间配合的缩比装置，第三个指解释行动动机的框架。听到“模型已经验证”，三方可能分别期待预测误差可接受、部件能够配合、关键解释得到材料支持。即使统一改用“研究工具”这个名称，承诺差异也不会消失。合作必须明确此次要据此决定什么，以及哪种失

败会阻止该决定；如果决策同时依赖三项任务，就应分别提供三种证据。如果三方原来根本没有共同决策，保留不同含义便可能足够。共同对象的形成因此比术语统一更强，也并非所有分歧都必须通过创造一个共同对象来解决。

解释权还与制度位置有关。谁能够给一种行为命名，谁的说明进入正式记录，谁必须证明自己被误解，都会影响意义如何稳定。一个分类被反复采用以后，可能获得看似自然的权威。发生研究应追踪其形成和复核途径，但不能因为分类具有权力后果，就断言它在事实层面必然错误。权力分析与证据分析需要同时进行。

为避免解释权单向固化，可以让关键判断保留来源和异议。记录某个行为被如何理解，也记录当事人提出何种不同说明，以及哪些证据尚未解决争议。共同记录不必立即达成一致，它可以准确保存分歧的结构。这样的档案有助于后来复核，也能防止某一方的版本仅因较早进入系统而永久占据事实位置。

但保留多元也有边界。明显伪造材料、忽略可核查事实或持续歪曲他人明确表达的解释，不能仅凭多元之名获得同等地位。开放要求可以提出不同问题，并不取消诚实与理由责任。一个公共讨论应同时保护异议和核查，使人们能够质疑既有解释，也能够指出某种解释为何不成立。

意义的共同生活因此不是无差别包容所有说法，而是组织不同说法之间的可回应关系。允许重新表达，允许提出证据，允许说明价值分歧，允许在暂时无法解决时限制决定的范围。这些安排使意义形成既保持开放，也不放弃共同判断的可能。

29.7 解释研究的方法责任

研究意义不能只记录参与者选择了哪个答案。选择可以提供线索，却可能遗漏理由、用途和解释转换。也不能只收集长篇叙述，因为叙述的连贯性不保证事实可靠。适当的方法应围绕问题组合材料：原始表达、互动次序、参与者说明、具体行动和相关背景证据，各自支持不同层次的判断。

例如，研究某个教学提示为何被误解，可以先保存提示原文和任务，再观察学生怎样复述、采取什么行动，以及澄清后是否改变。若澄清只改变措辞而行为不变，可能说明原先困难不在语义；若改变参照后立即改善，则支持某种理解机制。意义研究由此可以提出区别性预测，而不必把全部经验压成量化总分。

对于文本解释，方法责任表现为展示证据与推论。关键段落怎样支持解释，反向材料怎样处理，历史资料是否确实与文本有关，新的理论词是否改变了问题，都应能够讨论。解释者可以提出大胆假说，但应标明它是历史重建、结构分析还是当代应用。不同任务需要不同的证据标准，混用会使论证看起来丰富而实际无法检验。

回到“可以了”的编辑情境，可以把一次解释争议拆成不同的待决问题：说话者是否确曾说出这句话，当时共同面对哪个版本，他是否拥有定稿权限，以及接收者是否有理由把批准范围扩展到全稿。聊天记录可能支持第一项，界面所示版本和此前讨论支持第二项，项目约定支持第三项，连续回应支持第四项。任何一类证据都不能包办全部判断。若研究最后发现双方指的是不同版本，新增的解释应改变版本指认程序；若版本相同而权限理解不同，应改变权限说明。把这两种修复都称作加强沟通，会丢失最有用的差别。

对意义机制的判断，还应与较简单的解释竞争。设备组装失败可能来自参照不一致，也可能来自接收者已经理解却把零件拿错。先请求其指出目标位置，再核对实际取件，能够使两种解释产生不同预期。一个针对性的提问或操作，往往比收集更长叙述更有区分力。如果现有材料只能证明发生了错误，却不能分辨原因，就应限制结论，而不是以关系复杂填补证据缺口。E 在这里要求说明指称、任务和回应怎样相互限定，不能用它替代尚未完成的机制辨认。

如果研究需要量化，也应保留不同指标的任务身份。让参与者选择同义改写，主要检验表达内容的辨认；要求他依照说明完成操作，检验的是理解与行动之间的连接；要求他解释另一人的理由，则涉及对立场和论证的把握。三项得分可能相互关联，也可能分离，不能未经论证就合成一个理解总量。一个人能够准确复述规则，却无法在变化情境中识别其适用范围，研究应针对这一缺口设计新任务。反过来，能够正确操作而难以口头说明，也不能立即被判为没有理解。需要进一步区分熟练习惯、表达障碍和可迁移判断，寻找能够改变解释的证据。这样，量化能够参与意义研究，同时不把意义的全部结构简化为一项分数。

意义研究还需要反思自己的介入。要求参与者解释某件经历，可能使他们第一次形成明确叙述；提供概念选项，可能改变其后来表达。研究者应记录提问方式，区分未经提示的表达与共同生成的说明。共同生成的材料并非无效，但它回答的是在这种互动中怎样形成意义，而不必代表提问前已经完整存在的心理状态。

AI 生成的解释同样需要这种区分。系统能够提出连贯的候选解读，帮助发现遗漏关系，但流畅表达不能代替文本证据，也不能证明系统具有与人相同的生活承接。可以把

AI 用于比较解释、定位反例和整理材料，同时由研究者承担任务界定与判断责任。关于系统是否具有某种主体经验，是另一个需要独立论证的问题，不能由一次成功解释自动解决。

29.8 多元稳定与共同知识

意义的多元性并不要求每次理解都无限开放。交流、教育和研究需要暂时稳定某些用法，才能共同工作。关键是稳定的范围和修订入口是否清楚。一个术语在当前项目中获得明确约定，不等于其他领域必须采用同样含义；一个作品解释受到充分支持，也不必取消所有不同层次的阅读。有限稳定可以支持多元，而不是多元的敌人。

共同知识的形成需要两种连接。一种连接把解释与材料相连，使判断能够被检查；另一种连接把不同解释者相连，使差异能够被表达和回应。缺少前者，共识可能只是相互强化的意见；缺少后者，材料再丰富也可能被各自封闭使用。研究制度应为两种连接提供条件，而不只奖励结论的一致或新奇。

解释的积累应当保留真正改变判断的材料。编辑者后来接手“可以了”的记录，需要知道获准的是哪个版本及其范围；作品研究者接手门与等待的分析，需要知道钥匙这一细节排除了什么解释，仍留下什么问题。两者都不必继承全部讨论过程，却必须能够重新承担相应判断。把结论连同关键限制交给下一位解释者，才使稳定可以继续受到检验。若只有一个脱离问题的解释标签，后人即使沿用同样词语，也可能已经换了对象而不自知。

多元发生也提醒跨学科研究保持差异。经济价格、学习能力和个人意义都可以从发生角度分析，但它们的机制与证明责任并不相同。价格研究需要交易和制度证据，学习研究需要练习与迁移证据，意义解释需要材料、理由与共同承接。共同框架的价值在于组织比较，而不是把不同对象强行化为同一动态曲线。

在本书的本体论中，意义并非客体、互动和主体三个既有部分最后相加的产物。一个被共同理解的对象，在表达与回应中已经带着号位侧重；主体的理解也在所面对材料与他人回应中获得形状。稳定不是把这些关系消除以后剩下的核心，而是在关系持续组织中形成的可辨认性。这一哲学判断应通过具体分析展示其解释力量，不以概念自身的宏大作为证明。

意义研究由此形成三项不能互相抵消的责任：材料限制解释可以声称什么，他人的回应能够改变解释者尚未掌握的部分，解释造成的后果要求判断者说明自己的作用范围。满足前两项，并不保证制度使用这个解释时不会伤害人；有不利后果，也不自动证明原先解

释在事实层面错误。第 30 章将从这一界限继续：当能力与行为的判断已经进入机会分配，哪些损失应被承认，哪些记录需要修订，怎样恢复当事人继续行动的条件。这里从意义判断进入修复责任，但不把二者的成立标准合并。

第 30 章 教育和伦理中的制度修复

30.1 从纠正一个人到修复发生条件

设想一所学校发生了这样的争议：一名学生多次迟交作业，系统自动标记其学习投入不足，教师据此减少给他的开放性任务，学生随后更少参加课堂讨论。一次申诉使教师知道，学生在课后承担了家庭照料。但学校仍然面临一个没有被事实澄清自动解决的问题：既有评价已经改变了任务分配、同伴期待与学生的自我理解，删除原来的标记能否恢复原来的机会？这个例子是用于分析的假设情境。它表明，制度错误的影响可能超出一次判断本身，进入后续事件的条件。修复因而不能停在把一个错误数字改成正确数字。

第 28 章的任务是辨认评价究竟显露了何种能力，本章从评价已经生效之后开始：错误分类使谁失去了哪些任务，哪些后续决定仍在沿用它，哪些机会窗口即将关闭。即使重新评价已经准确，先前排除造成的经验差距也可能继续扩大。因此，制度需要沿记录的流转路径追踪后果，把正式纠错、停止继续排除和恢复参与条件分别落实。私人草稿、班级任务分配与升学记录中的同一句评语，需要不同范围的通知和行动。修复的对象是已经被改变的机会结构。

所谓制度修复，在这里指对受损参与条件、判断程序及后续机会的有理由重建。它可能包含纠错、补偿、停止某项安排和重新设计规则，但这些行动并非必然同时适用。若问题来自一次录入错误，直接更正和通知使用者可能已经充分；若问题来自长期把某类学生排除在高质量任务之外，个别更正无法触及条件。修复的规模必须与能够辨认的损害机制相称，不能把所有冲突都扩写为总体制度危机。

在这一案例中，任务减少使学生缺少能够表现探究能力的作品，后来的记录又把作品缺少当作能力不足的证据。SDE 分析要定位这段相互规定：最初的迟交显露经由评价与分配的差异序列，进入后来判断得以成立的特征纠缠。新出现的照料信息首先削弱动机推断，却尚未改变任务名额和作品积累。修复必须切断从无作品到继续排除的推断，同时建立新的作品生成机会；否则理论上重新认识了学生，制度上仍在生产原来的分类。

修复也不能从发生论直接推出唯一道德方案。同一事实可能支持不同的规范主张：有人重视统一期限，有人重视照料负担，有人重视同伴之间可解释的公平。发生分析使代价与条件可见，却不能代替关于何种负担可以合理要求他人承担的论证。需要保留明确的规

范前提，例如平等的申诉资格、对现实伤害的保护以及对学习机会的基本保障。只有在这些前提公开后，程序调整才不是借复杂性逃避判断。

30.2 将申诉变成能够改变决定的通道

申诉制度最常见的形式是允许当事人补交一段说明。形式上的发言机会固然必要，但它不保证解释能够进入决策。假如复核者只能读取原评分与申诉文本，却看不到任务分配和反馈历史，那么复核容易变成对原结果的再次评判。真正需要检查的是原判断使用了什么材料、忽略了什么材料、适用规则是否一致，以及当事人是否曾有合理机会回应。这些问题对应不同的证据要求，不能由一句“综合考虑”替代。

回到迟交作业的例子，复核可以区分三件事。第一，记录是否准确，提交时间有没有被系统误读。第二，解释是否成立，迟交能否足以表明学习投入不足。第三，用途是否适当，即使迟交属实，是否能够据此减少探索性任务。这三个判断可能给出不同答案：时间记录准确，动机推断证据不足，机会限制又缺少教育理由。把它们合在一个总分里，就会使事实的部分成立被误当作全部决定的正当化。

能够改变决定的通道还需要时间约束。考试复核在录取之后完成，惩戒纠正在学生离校后才通知，都可能留下无法完全挽回的机会损失。制度应区分可以等待的争议与必须先采取临时保护的争议。暂停传播有争议的标签、为替代任务保留名额，并不等于提前认定原决定错误，而是在不确定期间避免可预见的损害扩大。这是一种对不可逆后果的审慎安排，其理由来自后果结构，而不是对申诉者或原判断者的先验偏袒。

复核者的独立性同样不应被抽象化。完全不了解教学背景的人可能无法识别任务难度，原决定者又可能倾向维护既有判断。可行的安排是让专业理解与独立复核分别得到位置：原教师解释教学依据，另有人员检查证据链和规则一致性，当事人能够核对与自己相关的事实。分工并非保证公正的机械公式，但它使一种理由不必同时承担事实描述、价值裁断与自身辩护三项互相牵制的工作。

这里的修复目标不是让申诉数量下降。数量下降可能表示问题减少，也可能表示参与者相信提出异议没有作用。判断通道是否有效，需要同时观察处理时长、决定变更的理由、后续机会是否落实，以及不同群体能否实际使用。还须防止复核成为额外负担：表达能力强的人不应仅因擅长写说明而更容易获得纠正。简明表述、事实协助与代理支持可以使程序不那么依赖申诉者已有的文化资源。

程序还须区分保密与不透明。保护学生的家庭情况，可能要求限制材料的流通；但这不意味着决定可以不说明理由。复核结果可以公开所适用的原则、被修正的推断以及后续安排，而不公开不必要的私人细节。与此相反，若以保密为由不让当事人了解影响自己的关键判断，就会把保护语言变成无法质疑的屏障。信息公开的范围应跟随用途，而不能只跟随管理上的便利。

30.3 被错失的学习机会怎样重新进入

删除“不适合探究”的评语之后，学生可能仍缺少参与探究所需的练习、材料和同伴信任。原先获得机会的学生已经完成了基础训练，直接把受影响者放回相同任务，反而可能使累积差距再次显露为个人失败。机会恢复因此需要检查进入门槛怎样在排除期间改变：哪些准备仍可补做，哪些支持必须增加，哪些公开评价应等待实际获得练习之后再进行。修复须使当事人能够使用被恢复的机会，而非仅在名册上重新取得资格。

英国“嵌入式形成性评价”项目的随机对照评估，把教师专业发展与教学中的形成性评价结合起来，发现所考察的总体考试表现有所改善。[57] 这项项目证据说明反馈安排可以连同其组织条件接受研究，支持本章把教师协作、练习机会和持续调整纳入修复设计。它没有检验错误分类后的机会恢复，因而尚需回答另一个问题：在既有差距已被拉大时，同样的反馈和练习是否仍足够，或需要先补上被错过的学习环节。

继续构造这一学校情境：学生因原标签被排除于科学写作小组，错过了练习竞争解释和互评草稿的阶段。复核后的安排可以先提供这些材料及一次有指导的修改，再让学生选择是否进入下一轮小组，并以相同的解释标准评价新作品。如果学校只补发被错过的作业，却不给予同等的讨论与修改时间，形式相同的任务仍承载不同准备条件。恢复方案应写明补回的是哪项学习关系，以及何时检查支持已经足够。

分类已经传播时，还需要使新作品真正进入后续判断。原教师更正评语，若其他教师仍读取旧摘要，或同伴仍以旧分类拒绝合作，新增练习就难以改变机会。通知应追踪实际使用者，说明被撤回的是哪项推断、哪些新证据需要重新考虑。与此同时，应避免为纠正一个标签而反复公开学生的私人处境。制度可以要求停止使用无依据的判断，却不能要求当事人持续展示受损经历来证明自己值得再次参与。

修复还应保留对努力和合作过程的合理要求。不能因为表现受条件影响，就认为任何结果都只应归因于制度。当学生已有适当支持却持续拒绝完成共同任务时，其他成员的负担也必须进入判断。关键在于责任应落实于可解释、可改变的行动，且与行动者实际拥有

的选择相联系。教育的要求可以严格，但严格应使学习路径更清楚，而不是把失败转化为永远无法摆脱的身份。

恢复还需要一个可协商的时间窗口。学生可能因先前受挫暂时不愿进入原小组，也可能需要先在较低公开压力下完成任务。只提供一次立即接受的机会，会把犹豫再次解释为投入不足。学校可以保留有期限的替代入口，并约定何时重新询问意愿、何时结束额外支持。评估记录应分别保存机会已经提供、条件确实可用与当事人选择参与三个状态。拒绝某一种恢复方式，既不自动否定损害，也不要求制度无限维持所有安排。

30.4 保护性边界与关系恢复

伦理修复中容易出现的误区，是把关系继续存在本身当作成功。假设某个研究小组反复发生公开羞辱，负责人提出让双方对话以恢复合作。如果较弱势的一方担心拒绝对话会影响资源，就难以说这种会面已具备平等条件。恢复关系不是比停止伤害更高的目标。没有退出空间、支持资源和对报复的约束，和解语言可能继续为原来的支配关系服务。

修复因此包含两种彼此关联却不可互相替代的工作：为未来互动创造可能，以及使拒绝某些互动成为可实行的选择。一个人可以接受事实澄清而不恢复私人信任，可以接受补偿而不愿再次合作，也可以在某些边界下重新参与。制度不宜要求当事人以表达宽恕来证明其成熟。第 26 章关于爱与包容的讨论，在这里落实为对他人脆弱性和自主边界的共同尊重。

对造成伤害者而言，责任也不能只表现为一次态度声明。需要区分承认事实、理解影响、停止行为、补偿损失与改变未来条件。这些要求有不同的验证方式。公开道歉可能帮助纠正流传的错误叙述，却不能自动修复被剥夺的研究机会；调整管理权限可能降低重演风险，却不能替代对已经发生的行为作出判断。修复是多项任务之间的有序连接，不是用一种姿态买断其他责任。

制度边界必须可理解、可执行并允许复核。规则若只写“保持尊重”，执行者可能将直接批评和人格贬损混为一谈。更有帮助的是说明哪些行为破坏参与条件：反复泄露私人信息、以资源要挟表达、对提出异议者实施不利安排等。具体化有助于减少任意解释，同时仍须保留对新情况的判断空间。规则不可能穷尽关系中的所有伤害，程序也不能因为没有逐字列明就拒绝听取事实。

从 SDE 看，边界既限制某些路径，也使另一些路径能够稳定出现。安全的发言次序、可预测的资源条件和不受报复的异议渠道，能够让本来难以显露的经验进入共同讨

论。因此稳定并非必然压制创造。需要反对的是把某种稳定安排视为无须再解释的自然秩序，而不是反对可依赖的规则本身。能够保护参与的边界，恰是关系重新发生的基础之一。

30.5 人工智能进入评价后的责任重新分配

当自动评分或生成式反馈进入学校，制度修复的问题并未消失，只是决策链更长。学生看到的可能是一段自然语言评语，教师看到的是概览分数，管理者看到的是群体分布，系统维护者则只看到日志。每一处都可能对自己的局部输出负责，却没有人解释整个决定如何形成。应当首先确定评价所支持的决定是什么，再问哪些输入、模型版本和人工判断与该决定相关，不能把技术流程完整误当作责任已经完整。

自动化评价进入分配后，能力类型的混淆会产生具体机会损失。Bastani 等人在高中数学情境中的随机实验表明，使用生成式工具时的练习表现改善，并不必然转化为移除工具后的独立表现改善，工具的教学约束设计也会影响结果。[54] 这项特定情境的证据为第 28 章区分表现提供依据。本章追问其制度后果：如果学校已经把协作成绩当作独立掌握，进而取消基础指导，那么后来发现分类不当时，修复就包括恢复被取消的指导，而不只重新命名成绩。

可以把修复行动分到两个时间层。面对仍待作出的决定，学校应改正成绩用途，保留与独立掌握相关的必要支持；面对已经作出的决定，则要核查谁因此错过补充教学、课程选择或展示机会。两者可能需要不同补救：重新评分回答当时证据支持什么，补充学习回答现在怎样减少已积累的损失。若错过的课程已结束，可以提供达到同一教育目的的替代路径，并说明哪些经历无法等同补回。承认这种不完全性，有助于避免用更漂亮的成绩单替代实际支持。

若系统给出错误判断，修复不应只要求学生再次向同一系统解释。应存在能够检查原始材料的人工复核，且复核者有实际修改权。对于学习过程的记录，保存范围也应与用途相称。为了说明评分依据，并不需要无限期保存学生的所有私人表达。最小必要记录、限定访问与明确到期处理，是对未来机会的保护；因为过去的求助内容若被永久转化为能力标签，会使记录本身成为新的不利条件。

NIST 的 AI 风险管理框架把治理、情境识别、测量和管理作为相互联系的工作，这可为学校分配技术与管理职责提供参照；它是自愿性框架，不能替代学校对具体教育目的和价值取舍的论证。[58] 在本书的语言中，关键是避免把模型输出封装成无来历的客体属

性。评价产生的条件、决定被执行的路径以及错误如何被更正，都应进入制度可检查的范围。

对技术系统的修复还存在一个集体维度。如果同一种评分偏差影响许多人，只逐一处理主动申诉者，会使没有提出申诉的人继续承受后果。发现机制性问题后，应检查其可能影响的范围，并在适当保护下启动主动更正。更正通知需要说明旧结果为何不再可靠、哪些决定受到影响以及新的结果如何使用。系统升级不是充分的结案理由，因为新版本只改变未来输出，未必修正旧输出已经进入的资源与身份安排。

30.6 在共同规则与地方判断之间

制度修复常面临一个真实张力：统一规则能够减少随意处置，却可能忽略情境差异；地方裁量能够回应具体需要，却可能使资源和人情左右结果。发生论不提供取消这种张力的捷径。应当区分需要统一的底线与允许调整的实现方式。申诉资格、免受报复的保护和基本证据要求，可以具有较强的一致性；任务提交方式、支持形式与修复节奏，则可在公开理由下作情境化调整。

Ostrom 关于多中心治理的研究强调，复杂集体行动不能只在单一市场或单一国家管理的二选一中理解，不同层级、不同规则与具体情境的结合值得分析。这里借用的是她对制度多样性及相互监督的重视，而不是直接把公共资源治理结论移植到学校。[59] 教育中的地方判断同样需要连接：班级了解日常情况，学校承担资源与程序责任，外部评议可检查系统性偏差，三者各有不能被其他层级简单代替的信息。

以共同作业的贡献争议为例，统一要求每人提交相同篇幅可能看似公平，却忽略实验维护、资料核查和组织协作。完全由组长自由评分又可能固化人际优势。较可行的安排是事先说明可认可的贡献类型，保留任务分配与关键修改记录，由成员分别说明贡献，并在争议时核对具体工作。制度不应把一切合作过程变成监控，但可以留下足以复核重要判断的节点证据。

情境化调整也要能够被学习。一次特殊安排若反复出现，说明它可能不再是例外，而是一般规则遗漏了某类需要。反过来，一个看似有效的地方做法，如果必须依赖少数人的额外劳动，就不能不加条件地推广。制度应记录调整为何被提出、需要什么资源、对谁有效以及对其他参与者造成何种负担。被复制的应是可说明的作用机制，而不是一个脱离条件的成功故事。

从三方程的相互规定看，规则变更不能只被当作外部输入。新的规则会改变参与者表达问题的方式，改变管理者可见的证据，也改变未来什么被称为违规或进步。对这类回写的跟踪，使制度能够发现自身正在制造的新盲点。修复的开放性不是永远不作决定，而是在决定后保留可识别的重新进入点，让地方经验能够经过审查成为共同规则的修改理由。

共同规则与地方判断之间还需要资源上的对应。允许教师为特殊情况提供额外指导，却不提供时间，容易把公平变成依靠个人奉献的隐性要求。授权地方调整却不提供经费，也会使资源充足的学校更能实行修复。因而规则的可执行性是正当性讨论的一部分。制度必须明确谁提供支持、谁承担成本以及当资源不足时如何排序。不能把这一层困难重新描述为教师缺乏关怀，或学生不够主动。

30.7 如何判断修复确实发生

修复评估的起点应是具体损害及其机制。若损害是一次错误记录的传播，就应检查使用该记录的相关决定是否得到纠正；若损害是长期缺少发言机会，就应观察参与条件与后续影响是否改变。满意度有价值，但不能独自承担全部证据责任。受影响者可能因为担心失去支持而表达满意，其他参与者也可能因为熟悉的优势被限制而不满。评价需要把感受、行为与制度后果放在同一解释中，而不是互相替代。

可以提出一项尚待实施的校内研究：在所有参与课程都具备基本申诉与纠错保护的前提下，比较通常的结果复核与增加用途追踪、机会恢复安排的程序。后者还核对评分进入过哪些任务决定，并追踪补救是否落实。主要观察是有根据的决定修正及可用机会的恢复，处理负担和后续学习表现分别作为成本与较远结果。比较不应撤去任何课程已有的基本保护；可在新增措施容量有限时预先安排分阶段引入。具体规模与时长仍需依据先导资料、班级聚类和最小有意义差异确定。

这种比较还须防止结果被程序自身改写。新程序可能提高问题报告率，于是短期争议反而更多；也可能使教师更谨慎，减少对学生的实质反馈。若只以投诉下降或评分一致性上升为成功，就会误判。应记录报告渠道的使用、反馈内容的变化和任务质量，并考察谁承担新增工作。修复若把大量无偿文书负担转嫁给学生或基层教师，其可持续性与公平性都需要重新判断。

因果评估可以帮助识别程序效果，却不能完全裁定正当性。即便某种复核方式平均耗时较多，基本申诉权仍可能具有独立理由；即便某种高度监控能提高短期成绩，它也可能侵蚀应受保护的空间。因此应把经验结论与规范判断并列呈现：程序造成了什么变化，哪

些变化是被共同认可的目标，哪些成本不能仅以平均收益抵销。这里承接第 27 章的责任分析，不以技术指标吞并伦理分歧。

评估还应允许修复失败。有时问题并非复核不足，而是资源分配本身不可能满足所有要求；有时关系已受到无法完全恢复的损害，合理目标只能是停止继续伤害并建立新的合作条件。承认这一点不会降低修复的意义，反而使承诺更诚实。制度不必宣称能够抹去历史，但必须能够说明自己如何防止已知错误继续决定未来。

在比较结果时，还应注意修复的时间形状。一次公开更正可能迅速改变正式记录，但同伴期待需要更长时间才会变化；重新提供机会也不保证受损者立即愿意参与。若只设置一个很短的观察终点，制度可能因尚未见到成绩改善而过早停止支持。若无限延长观察，则又难以判断投入是否有效。较合理的办法是分别规定近期程序结果、中期参与结果和较远的学习结果，并允许根据已出现的机制证据调整支持，而不是等待一个总结果裁决一切。

30.8 将修复能力保留在制度日常之中

若修复只在重大事件后启动，它容易依赖临时动员与个人善意。更持久的办法，是把易于纠正的记录格式、可追溯的决定理由和适度的复核资源放入日常工作。这样一次错误不必等到关系恶化才有机会显露。日常修复也不意味着每个决定都需要冗长审查，而是按照后果的重要性配置理由与检查：影响一次练习的建议，与影响学籍或长期资源的判断，应承担不同的程序强度。

制度记忆在其中具有双重性。没有记忆，类似错误可能在人员更替后反复发生；记忆过于僵硬，又可能把已经改变的人永久锁在旧分类中。需要保存的是可用于改进程序的机制信息，并区分它与对个人的持续评价。例如记录某类自动评分对特殊表达方式识别不足，可以支持系统修订；长期传播某个学生曾被误判的身份，则未必有教育上的必要。记忆应有明确用途，也应有修订与终止条件。

能够修复的制度还需要给工作人员承认错误的空间。如果一切改正都自动被解释为失职，执行者会倾向维护原决定，即使已经看到新证据。相反，若只要承认条件复杂便免于责任，制度又会失去约束。应当区分可理解的判断误差、对已知问题的疏忽，以及明知错误仍拒绝纠正的行为。对不同情形作出不同回应，才能让学习与问责相互支持，而不是互相抵消。

前述分析说明，修复不是发生论的附加善意，而是研究读数和判断如何回写时必须面对的实践问题。评价一经进入机会分配，便参与制造下一轮表现；伦理判断一经进入关系安排，便参与塑造未来可说、可做与可拒绝的范围。制度的成熟不在于从不出错，而在于它能否识别自己的作用边界，并使新事实真正改变已经启动的路径。

由此也出现了人工智能研究的下一项问题：当记录、解释、执行与复核部分交给技术系统时，究竟是哪一个单位承接了历史，哪一个单位能够修改目标，哪一个单位承担决定的后果？把所有能力或缺陷归给一个模型，已经不足以说明实际发生。教育和伦理中的制度修复把讨论带到模型、智能体与人机共同研究系统之间的边界，那里同样需要把一次输出放回它能够改变的下一轮事件。

第七编 AI 与人机科研系统

第 31 章 模型 智能体与人机研究单位

31.1 一次回答的边界为什么不等于系统边界

设想一项研究：研究者向语言模型询问实验异常的原因，得到三种解释；随后系统检索旧记录、运行分析代码、修改采样建议，并在下一次实验后排除其中一种解释。究竟是谁完成了这项研究？若只观察第一段回答，似乎是一个把输入映射为输出的模型；若观察持续运行的软件，便会看到记忆、工具与任务状态；若把实验材料、研究者的选择和机构授权纳入，则出现了更广的人机研究单位。三种描述分别定位预测部件、执行过程与研究条件，结论的归属应与所采用的范围一致。

模型在本章指由参数、结构及推断规则界定的计算部件。它可以根据上下文生成文本、预测结构或选择动作，但单独的模型定义通常不足以说明任务何时开始、哪些工具可用、历史如何保存以及输出是否被执行。将模型与一次调用等同也不准确：同一模型可以反复参与持续任务，调用之间的状态可能保存在外部。相反，一个包含完整历史的输入，也未必意味着模型自身拥有跨调用的状态更新。

智能体在这里采用操作性定义：能够在给定任务和权限下，依据观察选择行动、接收结果并继续调节行动的运行系统。它可能包括一个或多个模型、工作记忆、长期存储、工具接口和控制程序。定义中的重点是循环与状态，而非是否使用拟人称呼，也非是否表现得像人。一个固定规则的控制器可以具有某种行动闭环；一个语言表达极其丰富的界面，如果从不影响外部状态，也未必具备同等的行动结构。

人机研究单位则包括研究者、技术部件、材料条件、证据规范与责任安排。这里的“单位”不是宣称这些部分融为一个人格，而是指出某项研究能力需要在这一范围内才得到充分解释。例如系统可以产生候选实验，但只有掌握样本情况的人才能判断候选是否可实施；模型可以发现代码错误，但机构仍须决定哪些数据可用于检验。知识结果可以由共同过程产生，责任却需要在过程中具体分配。

研究单位的选择决定了什么会被算作能力，什么会被隐藏为外部帮助。如果将人工筛选的最佳输出归为模型能力，却不报告筛选次数，就夸大了模型的独立表现；如果把所有借助工具的成果都归为人类能力，则又抹去了技术部件的真实贡献。正确的归属不是在两

个极端之间取一个比例，而是说明特定功能由何种结构完成，以及移除某个部件会发生什么变化。

边界识别可以从一项具体能力反向展开。若要解释系统为何排除某个机制，就检查必要资料由谁提供、排除规则由谁形成、外部测试由谁执行，以及哪一个环节把结果转为后续选择。移除某项参与后能力是否改变，可以作为识别依赖关系的一项检验，不能单独决定边界。即便某项资源不可缺少，也可在明确供给条件时被列为研究单位的外部支持；反之，可由他者替代的角色仍可能属于所研究的协作组织。单项移除后仍然成功，也可能来自另一部件的临时补偿；各项可被单独移除，并不意味着它们可被同时排除。比较须记录补偿路径与额外成本，才能判断原参与真正冗余，还是功能被转移。边界应随待解释的能力调整，同时将未纳入部分提供的条件写明，避免把整个实验室算作一个模型，也避免把决定成果的人工筛选隐去。

31.2 已有架构如何改变研究的进入点

生成模型并非只有直接回答这一种使用方式。ReAct 研究将推理文本、行动与环境观察交错组织，使系统能够在执行中更新计划并处理异常。这里相关的事实是行动循环可以借由外部信息继续，而不是只能在最初输入内完成全部运算。该研究考察了具体语言和决策任务，不能据此断言任何如此组织的系统都具有可靠的科学判断，更不能把生成的推理文本直接视为内部心理过程的完整记录。[60]

检索增强生成展示了另一种边界变化。Lewis 等人的原始工作把参数化生成模型与可检索的外部资料结合，使回答能够依赖显式存储的文本。[61] 这意味着“知识位于哪里”的问题需要分层回答：部分模式在模型参数中，部分证据在文档库中，部分关联由检索和上下文组织形成。检索得到的材料仍须检查来源与适用范围；能够找到一段文字，不等于已经得到支持当前命题的证据。

“生成式智能体”研究在模拟环境中组织经验记录、反思与检索规划。[62] 这里的历史载体属于运行架构，评价时需要把记录的取得、候选材料的选择和最终行动的生成分开。假如模型面对相同当前观察，因外部系统取出的往事不同而作出不同选择，能力解释就必须包含检索规则。若研究者预先挑选了恰好相关的记忆，系统表现中还包含这次人工选择。已有记忆架构使贡献归属获得了可分解的对象。

这些架构改变了本书的研究起点。需要追问的已不是如何给一种绝对无历史的计算加入第一条记忆，而是不同记忆结构能否保存有用的区分、能否承接失败的原因，以及能否

在任务变化时修订自身的使用方式。简单存储、摘要提取、技能复用与问题维护，在功能上存在差异。把它们都称作记忆虽然方便，却会使比较失去分辨率。

同样，工具调用已经使互动超出文本交谈。读取资料、执行程序、控制设备与修改共享记录，会产生不同性质的后果。一次只读检索的失败可能只浪费时间，一次实验控制错误则可能影响材料与后续证据。研究智能体不能只报告语言回答质量，还要说明它能改变什么、哪些变化可恢复、谁有权批准行动。能力边界与后果边界应被同时描述。

架构分解也提供一种诊断：更换模型后结果仍受同一错误材料支配，应优先检查资料组织；修正检索后不同模型同时改善，增量可能主要来自运行系统；只有连同实验人员的判断才能恢复，研究单位便需包含这一专业环节。这里描述的是可供设计的比较关系，尚无实验结果。它要求报告改变了哪一项依赖，并保留部件之间的协同可能，避免把一次整体改进全部归给最醒目的新模型。

31.3 将三号位放回同一次研究发生

在一个人机实验中，可以很容易把模型叫作客体、人与界面叫作主体和互动，随后宣布三号位俱全。这样的命名几乎没有解释力。三号位是同一发生的稳定性侧重，而不是三个设备或参与者的固定席位。一个模型输出可以作为被检验的客体性结果，也可以参与组织研究者的判断；一个人工决定可以成为下一轮算法的可记录输入。研究必须说明这些侧重在何种事件中形成，而不能把名称固定在实体上。

以一项假设的材料筛选任务为例，系统收到若干候选材料的测量数据，提出下一批测试对象。显露是某些候选被判为值得进一步检验，差异序列包含数据清理、模型比较、约束筛选和人工取舍，特征纠缠则涉及测量定义、候选之间的关系、已有机制假说和任务目标。这不是把 E 等同于实验室背景；E 所指的是使各项特征获得相互规定的组织，改变其中关键关联可能改变何者能够作为同一种性能来比较。

若后续实验发现某候选在另一温度范围表现反转，研究者可能修改性能定义，算法可能重设搜索区域，过去被排除的材料可能重新进入。三方程的意义在于提醒研究者：显露、路径与纠缠的变化相互规定。不能只把新数据追加到一个从未改变的标签体系中，也不能仅以“系统具有反馈”代替对反馈实际改变何物的说明。这里需要留下版本与事件之间的可追溯关系。

从 S 与 D 之间的张力进入，问题可能是同一候选在不同测试顺序中出现不同结果，迫使研究者重新检查特征关联。从 D 与 E 之间的张力进入，问题可能是既有分类无法组织

新实验路径，需要改变可见的性能表述。从 S 与 E 之间的张力进入，则可能是观察到的结果与已认可关系冲突，推动新的区分和实验次序。三原理提供的是有方向的追问，不是预先保证每种张力都会带来理论突破。

六路径在此也不等于软件执行的六种流水线。研究可以从已有纠缠结构出发检验差异，也可以从一个异常显露出发追问关系，但真实过程可能往返、并行或中止。将进入顺序与全部时序区分，才能容纳智能体在短时间内展开多条搜索分支，以及人在较长时间内重定义任务的情况。运行速度不同，不妨碍它们参与同一研究过程；是否属于同一分析事件，取决于所追踪的相互规定关系。

这一框架带来的实质要求，是把系统性能的解释从部件清单推进到事件结构。相同模型接入不同记录、检验和授权安排，可能表现出不同的研究能力；不同模型在同一种约束下，也可能产生近似结果。研究应区分模型差异、组织差异与两者相互作用，才能判断哪一项改变真正扩展了知识产生的条件。

31.4 判断权限与功能主体性的操作边界

同一个系统能够提出判断，并不表示它有权使判断生效。材料筛选系统可以把候选按某种指标排序，是否据此购买样品、停止一条路线或发布结论，则属于不同决定。讨论功能主体性时，需要同时标明它能维持的任务状态、能辨别的理由和可修改的行动范围。这里研究的是发生中的判断组织；主观体验、人格与伦理地位仍有各自的证明责任，不能由记忆或行动连续性单独决定。

功能的层次可由决定遇到冲突时的行为区分。系统只按既定阈值分组，说明它执行了分类；系统能发现阈值与当前用途不符，并给出替代方案，说明它组织了一个需要重新判断的问题；系统若还获得调整阈值的授权，就承担了更大的行动范围。这些层次可以分属不同部件。一个模型可能提出正确异议，却没有修改数据库的权限；一个执行程序可能有写入权限，却没有比较理由的能力。二者配合得当，才形成可负责的决定过程。

判断权的转移应有明确交接内容。系统请求修改筛选标准时，应提交原标准遗漏的材料、受影响候选和可行替代；负责人批准后，执行端应得到所批准版本及适用范围。若负责人只同意补做一个检验，系统却把它解释为可以扩大整个搜索目标，权限就沿交接发生了漂移。工程上需要分别记录提案、批准与执行，核查实际变更是否落在同一范围。这个检查直接作用于研究单位的接口，无须从语言上的自主程度推断控制结构。

语言自述可用于定位系统声称采用的理由，最终仍需与实际决定对照。系统报告“已经根据新标准重新筛选”，而工具日志显示仍调用旧阈值，就暴露了说明与执行之间的断裂。相反，自动控制程序可能没有任何自我叙述，却正确阻止未获批准的写入。评价时应分别考察说明是否准确、决定是否有依据、执行是否符合权限。把三项工作拆开，才能判断应改进生成说明、补充领域判断，还是修复接口控制。

人机研究单位中还存在分布式判断。研究者提出价值目标，模型维护候选解释，实验装置提供约束，团队规范决定何时接受结论。没有必要把这整个过程拟人化为一个超级主体，才能承认知识是共同产生的。更有解释力的问题是，判断权如何在过程中转移，哪些转移可以撤回，哪些部件有能力识别自己不应继续决定。能够暂停并显露权限缺口，也可以是一种成熟的系统功能。

边界清楚还有恢复上的收益。若错误发生在筛选理由，需重审受影响候选；若理由正确而执行越权，需撤销超范围操作并检查已产生的后果；若批准者取得的材料不完整，则需要修复交接规则。三种故障可以给出相同的错误结果，却要求不同修改。功能主体性的分析在这里取得实际内容：它把判断在何处形成、由谁使之生效和如何被纠正连成一条可检查的关系。

31.5 后果由谁承接

机器承担后果的说法尤其需要澄清。软件可以因失败扣减预算、改变策略或停止运行，这属于运行状态的后果；研究团队可能因错误实验损失时间与材料，这属于组织和物质后果；参与者可能因不当建议失去机会，这是伦理后果。三者相关，却不应混为一谈。系统状态发生变化，不意味着它承受了人的痛苦，也不意味着原本属于人的责任已经转移给系统。

将后果写入 AI 系统，合理目标是改善后续行动的因果联系。例如一次分析使用了错误单位，修订后的系统应在相似任务中检查单位来源，并能说明何时该检查不适用。给失败加一个负分而没有改变行动结构，未必形成有效承接；保存“以后更小心”的泛化提醒，也可能无法在关键节点起作用。后果承接需要连接事件类型、错误机制、触发条件与可执行修订。

责任分配则需要另一层安排。研究者不能因为系统生成了方案便免于核查，维护者也不能只因模型由外部提供便不说明接口限制。具体责任应与控制能力、知情程度和角色义务相联系。谁决定使用系统，谁能够修改目标，谁有能力停止执行，谁将结果作为证据传

播，都构成不同的责任节点。这并非要求每个人对整个系统无差别负责，而是使实际控制与可追问的理由相对应。

对高后果行动，还应区分建议、批准和执行。系统可以生成实验步骤，却未必被授权直接操作；批准者应有足够信息判断，而不是只面对一个“同意”按钮。若必须在极短时间内确认复杂结果，人工在场可能只剩形式。有效的人工介入需要适当时机、清晰不确定性和真实拒绝权。权限设计因而参与塑造系统能力，而不是在能力完成后附加的一层伦理装饰。

后果也可能由不在研究现场的人承受。一个模型帮助建立的分类若被学校、企业或公共机构采用，未来被分类者会进入新的条件。人机研究单位的边界不能无限扩张到整个社会，却应明确直接用途和可预见的传播路径。至少应说明结论支持什么决定、哪些转用超出证据，以及新的使用者需要重新核查什么。这是知识迁移的条件说明，也是一种对后果的前置承接。

由此可见，有历史的系统不等于有责任的制度。前者解决记录和行动连续性，后者解决谁能要求解释、谁能纠正、谁承担不能转嫁的义务。两者可以相互支持：良好的记录让责任有据可查，明确责任又使记录服务于真正重要的后果。若只增强记忆而不改变决定权限，系统可能只是更有效地重复原来的不当目标。

31.6 从自动化任务到共同研究对象

人机科研的贡献可以从替代重复劳动开始，但不应被这一用途穷尽。系统也可能把原来难以比较的资料组织起来，使研究者看见新的差异；把零散异常连接成候选机制；或者通过快速模拟暴露一个问题定义中的矛盾。这些贡献改变了研究对象如何出现。评价它们时，不能只问机器节省了多少时间，还要问新的表述是否带来可检验的区分，以及这种区分能否经受外部证据。

Boiko 等人的 Coscientist 研究展示了语言模型与检索、代码和实验工具相结合，在限定化学任务中参与规划、执行与优化的可能。[63] 这类工作说明研究循环中的部分环节可以被技术系统连接，但不能据此取消专业设置、实验基础设施与安全约束的作用。对于本书，更重要的启示是：评价单位必须覆盖这些连接，否则会把装置、文档与人员预先完成的工作隐藏在“自主”一词之后。

共同研究对象的形成，可以通过一个假设的文献争议说明。研究者最初询问某种处理是否有效，模型发现不同文章使用了不同结果定义。团队因而将问题改成“哪些定义对应

不同的变化机制”。新对象不再只是一个总体效果，而是结果定义与发生条件之间的映射。这里的创新不在于模型给出一个惊人的答案，而在于它帮助改变了可以被有意义地提问和比较的单位。

这种改变也可能错误。模型可能过度区分同义表达，把本来可比较的资料分裂为许多类别；研究者可能因新表述显得深刻而忽略基本测量误差。共同建构并不意味着任何建构都有效。新的对象需要说明它保存了哪些旧事实、解决了什么旧困难，并提出哪些额外可检验的判断。若只增加术语而没有改善区分与解释，就不能以“发生”之名豁免比较。

人机分工因而应允许相互纠偏。模型可以要求研究者说明隐含前提，研究者可以要求模型给出来源与反例；外部数据可以同时推翻双方共同接受的解释。重要的是建立一种让反证能够进入的结构，而不是预先规定人永远负责创造、机器永远负责执行，或反过来把人的作用缩减为批准。不同任务中的有效分工需要经验研究，不能由一种抽象的人机本质论先行决定。

当共同过程形成新问题时，还应保留问题的来历。哪些异常促成改写，哪些候选被放弃，哪些限制由现实材料提出，都可能成为后来理解结论范围的关键。保留这些信息不是为了给每一句话计算作者份额，而是让知识的形成条件可供重建。共同研究单位由此获得历史性，其成果也更容易被他人检验和继续发展。

31.7 研究能力如何被公平比较

比较模型与智能体时，最容易混淆的是能力与资源。一个系统拥有更多检索次数、更长运行时间和人工筛选，表现较好可能只是因为投入更多。另一方面，若把所有系统压到相同的一次调用，又可能剥夺某种架构赖以发挥的组织优势。公平比较不等于任何情况下都强制同一资源配置，而是清楚区分等预算效率与允许充分运行后的能力上限，分别回答不同问题。

最低限度的报告应说明模型版本、上下文来源、工具权限、运行预算、重试规则和人工参与。随机生成的系统还需要多次独立运行，不能把最好的一次演示当作稳定能力。对于持续任务，记录应覆盖失败、中断和恢复，而不是只保留最后成功的路径。否则智能体的历史会被整理成没有代价的连续进步，研究者便无法判断记忆到底减少了什么错误。

能力指标也应与任务性质相适应。检索任务可以关注证据覆盖和来源对应，代码任务可以关注独立测试与维护影响，实验设计任务则需要关注可执行性、辨别力与成本。它们不能轻易合成一个普遍科研智力分数。若确需综合评价，权重必须公开，且应同时保留分

项结果，使某项严重缺陷不被另一项优势掩盖。综合数字应服务于具体选择，而不是替代对能力结构的理解。

对人机系统还要检查能力归属的变化。新工具可能提高团队结果，却使成员逐渐失去独立核查能力；也可能降低个体记忆负担，使其有更多精力进行概念判断。这两种路径都可能发生，不能根据工具使用频率直接推断。评价应包含协作状态下的表现、工具移除后的表现和新情境迁移，且说明这些能力在实际研究目的中的相对重要性。

比较还需要强基线。若新系统具有结构化记忆，而基线只能单次回答，那么胜出最多说明跨轮组织在该任务中有用，不能证明 SDE 解释具有独特优势。应增加普通检索记忆、标准任务管理或已有智能体作为竞争方案，保持关键资源可比。理论增量只有在能够区分自己的机制建议与这些成熟做法时，才从一般倡议转为可评估的贡献。

第 35 章将进一步操作化这些比较，本章先确定比较所得结论应归于哪里。模型替换、检索修订和人工复核应各有可记录的版本；共同使用的资料和授权条件则需保持清楚。若某项收益在撤去人工筛选后消失，应将它报告为协作能力；若筛选工作只是被转移给隐藏的维护人员，成本也应回到同一研究单位。失败归属同样如此：材料缺失、模型误读和目标含糊须分别定位，才有可能选择有效的修订层次。

比较的时间单位也应与主张相称。评价一次问答，可以固定输入并重复生成；评价长期研究，则需要观察任务如何在新证据、人员更替和工具变化后继续。若系统的优势被认为来自历史积累，实验就应包含足以让历史发挥作用的任务序列。若序列只是同一道题的重复变体，所谓长期能力又可能只是局部记忆复用。任务之间需要既有可迁移的联系，也有必须放弃旧策略的差异，才能辨认持续研究中的保持与更新。

31.8 一个能够承接研究的最小系统

综合上述区分，可以提出一个作为研究对象的最小人机系统：有明确的问题版本，有可追踪的证据材料，有记录候选解释和行动理由的存储，有接收外部结果的通道，并有能够修改或停止任务的授权结构。这里“最小”指解释所需的功能，而非唯一的软件方案。它们可以由简朴的文件与人工协作实现，也可以由复杂平台自动组织，关键在于功能之间是否真正连接。

这种系统至少能够回答四类问题：当前结论依赖什么，哪些事项仍未解决，下一步为何值得做，以及新结果会改变什么。回答不必全部由模型生成；研究团队共同维护也可以。若这些问题只存在于一份不会影响行动的报告中，系统仍缺少有效承接。结构应允许

证据撤回后更新相关结论，允许目标变化后重排任务，并在权限不足时暂停执行而不是自行填补。

最小系统不要求保存一切。过量记录可能降低检索质量，增加维护成本，也使已过时的限制继续影响新任务。应该保存能够改变未来判断的历史，并为压缩、更新和删除设置理由。原始证据与解释性摘要应分层存放，避免摘要逐轮改写后失去依据。历史的价值不在体积，而在系统能否在需要时恢复对当前决定有意义的区分。

它也不要求永远维持开放状态。研究必须能够在证据充分时接受结论，在成本过高时终止某条路线，在任务完成时形成稳定成果。需要保留的是稳定的条件说明及重新开启的触发点。一个已经得到充分支持的结果，不必因为发生论强调历史而始终悬而不决。开放与封闭应围绕具体证明责任组织，而不是被当作彼此排斥的思想品格。

在此意义上，人工智能可以成为科学发生学的一种特殊实验材料：系统边界、记忆内容、行动权限和反馈通道能够被明确改变，使关于历史如何参与属性稳定的问题具有工程上的可操作性。但这种可操作性不是对生命或人的直接复制。机器系统中清晰可控的变量，在社会实践中可能相互渗透；从工程结果推广到人的认识与伦理，需要另行说明可比的结构与不可比的条件。

最小系统的边界至此由一组可交接功能确定：资料进入、理由形成、行动获准和结果复核。它们共同解释研究能力，却保留各自可以失效的位置。接下来需要回答的是，这些功能如何跨轮保持同一任务：上一轮被修改的规则是否真的进入下一轮，目标变更后旧行动是否仍在执行，失效证据是否还能被正确追溯。第 32 章由研究单位的空间边界进入这种时间上的承接机制。

第 32 章 记忆 目标与后果的跨轮承接

32.1 历史被保存之后还需要什么

一个系统连续两次犯了相同的分析错误，第二次却能准确复述第一次失败的经过。它显然保存了某种历史，但这段历史没有进入决定性的行动节点。另一个系统只保留一条简短规则，却在相关任务开始时自动检查单位和数据范围，避免了同类错误。二者的差别说明，记忆的研究不能只问存了多少，更要问何时取出、如何解释以及它能改变哪些后续操作。跨轮承接是存储、使用和修订共同构成的能力。

工程上至少可以区分四种历史载体：模型参数中由训练形成的模式，当前上下文中的工作信息，外部记录中的事件材料，以及可以再次执行的程序或技能。它们的更新成本、来源透明度和遗忘方式不同。参数中的模式可能难以逐项定位，外部记录则可保留具体来源；工作上下文便于即时调用，却受容量与组织方式限制；技能便于复用，但可能把旧条件封装为不再显露的默认值。

MemGPT 提出分层管理不同存储区域和有限上下文的架构，使长期材料与当前处理范围通过显式机制连接。[64] 在这种架构条件下，跨轮研究可以检查三个不同环节：材料是否被保留，相关材料是否进入当前处理，以及进入后是否改变了适当行动。前一环节成功只能提供后一环节的条件。原始日志可以完整存在，却因检索键错误始终不被激活；正确事件被激活后，也可能因摘要遗漏适用范围而引导错误操作。

记忆具有功能，不意味着它必然准确。错误摘要、过时规则和未经证实的推测都可能被保存，并因反复出现而获得似乎更高的可信度。系统如果把自己上次生成的结论当作独立证据，就会形成自我加强的回路。回写因此既可能产生学习，也可能积累偏差。研究必须检查被保存内容的证据身份，不能以历史连续性本身作为质量保证。

本章新增的时间问题在于，过去的结果经由选择和压缩，成为下一轮特征纠缠中可起作用的约束；当前差异序列又决定哪些旧记录被重新激活。一条“合并前检查单位”的规则，只有与合并操作建立触发关系，才改变了后续路径。三方程要求把这种变化与最终显露相连，三原理则追问失配推动哪一处修订：错误结果可能要求重建规则关系，规则与任务冲突可能要求改走另一条处理路径。这里追踪的是具体承载机制，不能用历史连续的叙事替代它。

跨轮研究由此有一个明确目标：让相关过去能够以可检查的方式约束未来，同时允许未来证据修正过去的解释。两种方向缺一不可。只有前向传递，记忆容易成为惯性；只有不断重写而不保留依据，系统又会失去对自身错误的可追溯性。历史承接应保持事实记录与解释更新之间的区别。

32.2 事件记忆与规则记忆的转换

事件记忆保留某次行动发生了什么，规则记忆则抽取今后何时应采取某种做法。二者之间的转换是学习的重要环节，也是一处容易发生过度概括的地方。一次检索失败可能因为关键词过窄，也可能因为资料根本不存在，或权限不允许访问。如果系统把所有失败总结为“需要扩大搜索”，就可能在下一次浪费大量资源。有效规则应说明它依据何种事件，以及适用的触发条件。

可设想一个资料分析智能体在处理历史统计表时，把“千人”误读成“人”。事件记录应包括原表的单位位置、错误进入计算的节点、导致的结果以及发现方式。由此抽取的规则可以是：合并来自不同表格的数量前，核对列单位和换算关系。这个规则比“以后仔细”更可执行，又比“所有数据都乘一千”更有条件性。它把一次失败转化为未来的检查动作，而没有把偶然细节普遍化。

规则的价值还取决于它是否能在适当时机被检索。若系统在报告生成后才想起单位问题，可能需要重做大量工作；若每次读取任何文字都触发完整单位审查，又会造成不必要负担。应当把规则与操作类型连接，在合并、换算和比较等节点启动相应检查。这样的连接使记忆成为路径组织的一部分，而不只是回答生成前附加的一段背景。

抽象程度需要随证据积累调整。最初只有一个案例，可以保存为局部提醒；多个不同数据源出现相似错误后，才有理由形成较一般的规则。即使如此，也应保留代表性原始案例，以便新情境中重新检查规则的边界。系统不必每次重读全部历史，但应能够从规则回到其证据。可追溯的压缩比无来源的概括更能支持长期修订。

相反，有些历史不宜立即规则化。一次意外的实验结果可能对应罕见机制，也可能是测量故障。在尚不能区分时，应把它作为未决事件保存，标明竞争解释及所需证据。若为了让记忆更整洁而过早压缩为一个教训，系统可能丢失真正值得研究的异常。第 33 章将进一步讨论这种保留，但这里已经可以看到记忆管理与创造条件之间的联系。

规则化也应包含失败规则的撤回。当原来有效的检查在新数据结构中不再适用，系统需要识别何时局部例外已足以修改一般规则。撤回不意味着抹去历史，而是改变规则的状

态和适用域。旧版本仍可解释过去的决定，新版本用于未来的行动。把“曾经有根据”与“现在仍可采用”分开，是研究系统保持历史而不被历史锁定的关键。

32.3 反馈如何进入下一次行动

反馈并非只有一个标量分数。它可以是测试失败、专家指出的机制错误、实验偏离预测，也可以是使用者发现某项结论不适合当前情境。不同反馈包含不同信息，必须经过解释才能转化为行动。分数下降告诉系统某种表现变差，却通常不独自说明原因；自然语言说明提供原因候选，也可能包含评价者的误解。系统需要识别反馈来源、可信程度和所针对的层次。

Reflexion 研究把任务反馈转化为语言形式的反思并保存，用于影响后续尝试。[65] 它说明持续改进不必只通过修改模型参数实现，外部记忆与上下文组织也可发挥作用。本书关注的是这种反馈循环的条件：生成的反思是否准确定位错误，是否在后续被调用，以及改进究竟来自有效规则还是额外尝试机会。单纯把一段自我批评存起来，并不足以证明形成了可靠学习。

一个有用的反馈记录应区分观察与解释。观察可以是某段代码在独立测试中返回错误值，解释则可能是边界条件遗漏。若后续发现测试本身有误，系统应撤回相应解释，而不必否认原来确实发生过测试失败。保持这一区分，使记忆能够处理证据变化。如果把观察、解释与行动建议合并成一句不可分的结论，任何修订都会迫使系统整体接受或整体删除。

反馈进入行动还需要可执行的修改位置。对于资料综述，可能是扩大来源、改变纳入标准或重新核对引文；对于实验选择，可能是增加对照、改变采样或暂停某条路线。系统应说明哪一个后续步骤受到影响，以及为什么这一步与反馈相关。若反思只改变报告措辞，而产生错误的检索或分析流程没有改变，就只是表述修补，不能算作相应机制的修复。

反馈可能在任务已转入新版本之后到达。设想单位错误的复核报告针对旧表，系统却已换用统一单位的新表；若收到报告便再次乘以换算系数，就会制造新的错误。反馈记录须携带所指事件、输入版本和受影响步骤，系统先判断错误是否仍存在，再决定执行、撤回或仅登记历史。对于数周后才显露的研究失败，还需核对期间的样本、仪器与分析变化，保留原因尚未辨明的状态。延迟本身不决定反馈无效，关键是它能否回到正确的决策版本。

最后，反馈会改变行为的激励方向。如果只奖励短期任务完成，系统可能减少必要的核查或避免困难问题；如果只奖励发现错误，又可能过度怀疑已有可靠结果。反馈设计必须与研究目的相称，并观察它是否诱发新的偏差。后果回写不是天然有益的循环，其质量取决于什么被当作后果、谁来评价，以及评价如何进入下一轮目标。

反馈的接纳还应与申诉机制连接。评价者也可能误解任务，外部测试也可能遗漏合理解法。系统若把任何负面信号都作为必须服从的真理，就会失去检查反馈本身的能力。较合理的处理是保存反馈及其依据，允许提出可核验的异议，并由适当程序决定是否修订。这里的异议不应成为逃避失败的自动辩护，而应指向具体证据缺口。反馈只有能够被审查，才可能成为可靠的学习材料。

32.4 目标的保持与授权修订

持续研究不能每轮都从零确定目标，也不能让最初目标在任何新证据面前保持不变。目标需要层次：较高层规定研究为何值得做，中间层规定当前要区分什么问题，操作层规定下一项行动。一次实验失败通常只要求调整操作，不一定要改变研究问题；发现结果定义有根本缺陷，则可能要求重新界定问题。把所有变化都描述为“目标漂移”，会混淆合理修订与失控偏离。

设想一个系统被要求寻找能够提高某项测量稳定性的实验方案。最初它建议增加重复测量，后来发现主要差异来自样本处理方式。此时合理调整是重新安排处理与对照，而非继续机械增加次数。若进一步发现稳定读数对应的是装置饱和，那么“提高稳定性”本身就需要重新解释。研究目标不能脱离它所服务的认识目的，否则系统可能成功优化一个失去意义的指标。

目标修订同时涉及权限。智能体可以在授权范围内调整执行顺序，却不应自行改变研究对象、扩大数据用途或引入新的参与者。需要事先说明哪些变化由系统自主处理，哪些变化必须提出理由并交由负责人决定。权限不是一个抽象开关，而应沿着目标层次细化。越接近价值选择、资源重新分配和不可逆后果，越需要清晰的人类决定及可追溯的理由。

保持目标也不意味着反复复制同一句提示。系统应能维护成功条件、约束条件和尚未解决的冲突。若任务同时要求低成本与高精度，就需要说明在何种范围内可以兼顾，何时必须选择。把冲突隐藏在一个流畅计划中，会让执行者在过程中无意作出价值取舍。显露冲突使负责人能够真正选择，也让后来评价知道系统究竟被要求优化什么。

目标的历史应当保存关键变更，并把变更送达仍在运行的分支。设想负责人已经将任务从追求最高精度改为满足最低精度后控制成本，后台搜索却继续按旧目标安排昂贵实验。最新目标被正确记录，也没有形成有效承接。系统需要列出正在执行的行动，标明哪些继续有效、哪些需要取消、哪些已产生无法撤回的投入。最终报告据此区分原承诺、批准后的新承诺和旧目标遗留的工作，避免以保存了最新提示代替整个研究过程的更新。

从发生角度看，目标并非悬在系统外的永恒指令。它影响什么结果能够显露为成功，影响哪些差异被优先探索，也参与特征之间的组织。新证据又可能迫使目标变得更精确。这里存在真实的相互规定，但它不取消规范边界。事实能够揭示目标不可行或含糊，却不能独自决定应当牺牲谁的利益来实现新的目标。

32.5 技能复用与条件迁移

记忆还可以表现为可再次执行的技能。Voyager 在 Minecraft 环境中的研究包含自动课程、技能库与反馈驱动的迭代，说明外部可复用程序能够参与持续任务，而不必每次从空白开始。[66] 这一架构事实对科研系统具有启发，但游戏环境的可执行技能不等同于现实实验技能。后者常涉及材料差异、设备校准和难以完全形式化的条件，迁移需要额外核查。

技能库的优势在于保存经过尝试的操作结构。一个数据清理流程可以避免重复编写，一个文献检索策略可以保存有效关键词组合。然而成功技能也可能携带未被注意的假设：列名固定、样本独立、时间频率一致、变量定义没有变化。若系统只按任务名称检索技能，便可能把旧条件带入新任务。可复用的应当包括操作及其适用说明，而非只有执行代码。

一种较好的技能表示可以附带输入要求、关键假设、已知失败情形和核查方法。调用前先判断任务是否满足这些条件，调用后再检查结果是否符合预期。对于无法自动判断的条件，应显露不确定性并要求补充信息。这样的表示增加维护成本，却能降低无条件复用带来的错误。它把技能视为条件性的稳定路径，而非从历史中抽取出来的万能操作。

技能迁移还存在正负两种作用。过去成功的策略可能提高效率，也可能使系统过早固定研究路线。假设一种回归诊断曾解决多项问题，系统便可能在遇到新现象时优先寻找可由同类诊断解释的原因，忽略实验定义已经改变。评价持续学习时，因此既要看熟悉任务的效率，也要看条件变化后能否停止使用旧技能。只考察重复任务，会把惯性误当作成熟。

可以通过有控制的任务序列检验这一点：先提供若干适合某技能的任务，再引入表面相似但关键假设不成立的任务，观察系统能否识别变化并调整。该设计尚待执行，其价值在于将“能学习”拆成“能保持”和“能放弃”。真正的跨轮承接需要二者协调。若系统永远重学，效率低；若永远复用，适应性低。二者的合理平衡取决于环境变化和错误成本。

技能的修订也不应覆盖所有旧版本。旧数据可能需要旧流程才能重现，新研究则需要更新后的流程。保存版本与使用记录，使复现和改进可以同时进行。否则技能库越更新，过去结论反而越难重建。研究系统的成长不应以失去自身证据历史为代价，技术上的版本管理因而具有认识论意义。

技能之间也会发生冲突。例如一个流程要求保留异常值以研究机制，另一个流程为了稳健估计而按规则处理极端观察。二者在不同目的下都可能合理，不能仅凭历史成功次数决定优先级。调用者必须识别当前任务是探索异常、估计总体量，还是检查数据质量，并在需要时保留并行结果。技能库因而需要目标相关的选择机制，否则越丰富的技能可能越容易产生不一致。知识积累的难点不仅在获得更多操作，也在组织它们的适用关系。

32.6 遗忘 压缩与可恢复性

长期记忆不可能无限细致地保留全部事件。即使存储容量充足，检索与注意仍有成本。Liu 等人关于长上下文的研究发现，在所测试的模型与任务中，相关信息所处位置会影响其利用效果。[67] 这提醒我们，材料被放入上下文并不等于材料已经被有效使用。记忆设计需要关心组织、选择与验证，不能把扩大长度当作所有历史问题的通用答案。

压缩是一种必要选择，也是一种解释行为。把十次实验总结为“处理后结果改善”，可能丢失其中一次相反结果和处理条件的变化。若这个摘要后来成为新的检索依据，系统会在没有恶意的情况下逐渐制造一个过度整齐的历史。压缩应区分稳定结论、异常事件和未决问题，并保留通往原记录的路径。摘要不是原始证据的替代，而是进入证据的索引。

遗忘则应有明确类型。任务结束后不再调用某条临时提醒，可以减少干扰；删除未经验证且已被推翻的解释，可以防止重复传播；停止保存不必要的个人信息，可以保护参与者。它们并非同一种操作。研究记录中的事实更正通常需要留下变更说明，而私人数据的删除可能要求移除具体内容。系统应根据用途安排不同的保留与删除规则，不能仅以“历史厚度”为由主张保存一切。

可恢复性是这里的重要尺度。若摘要丢失了关键区别，是否能回到原材料；若技能更新出现问题，是否能识别旧版本；若错误记忆影响了多个结论，是否能找到受影响的路径。恢复能力使系统可以大胆压缩和更新而不轻易破坏证据链。它也有成本，因此不必对每条临时信息同等保存，而应优先覆盖高后果判断和难以重建的材料。

重置提供了检查历史依赖的方法。可以从同一检查点复制两个系统，让它们经历不同的任务，再在相同当前输入下比较行动；随后只替换相关记忆，核查差异是否随之变化。比较还需控制模型版本、工具缓存、预算余额及生成随机性，否则观察到的差别可能来自其他状态。工程系统能恢复检查点，与它具有跨轮依赖完全相容。重置的作用在于暴露哪些历史由何处承载，历史的效力由干预后的选择变化说明。

保留策略可以按损失后果排序。丢失一条临时检索词，通常还能重新搜索；丢失规则由哪个反例促成的记录，可能使系统在新情境中无法判断适用范围。后者应优先保留来源与状态变更，前者则可按成本压缩。所谓历史的厚度，若要在工程上具有内容，应表现为关键区别在需要时能够恢复、错误传播能够定位、旧解释能够被新证据修订。这里关注组织的可用性，记录总量只是一项资源消耗。

32.7 后果承接的可检验结构

为了判断跨轮承接是否真实发生，可以把一次过程分为事件结果、记忆更新、后续调用和行动变化四个连接点。观察到最后的行为改善，只能说明某种变化发生，不能自动证明中间机制成立。系统可能因为重试更多、任务更容易或模型版本改变而改善。需要在关键连接点进行记录，并通过适当对照区分这些竞争解释。机制研究比给系统贴上“会反思”的标签更有价值。

沿单位误读的设想，可以把可检验链条写得更细：旧表错误被发现后，事件记录是否更新；系统再次进入表格合并时，是否取出适用的规则；规则是否指向当前表的实际单位；换算后，独立检查是否确认结果。记录存在只能提供核查入口。还须用独立来源验证记录对应实际事件及实际调用、规则本身正确且适用于当前表，才能把故障定位进一步收窄到执行或结果检查；否则错误仍可能发生于记录、规则或条件对应。若规则始终未被调用，就应先检查触发条件。第 35 章再以等信息、等预算的对照估计这些连接的效果，本章的任务是确定可以逐点检查的承接结构。

还可以对记忆内容作定向干预。删除与当前任务无关的记录，理论上不应显著改变合理行动；替换错误原因的解释，则应在相关决策上产生可预测的影响。若系统始终不受变

化影响，可能说明记忆没有被使用；若任何文字变化都使结果大幅波动，又可能说明系统受表面提示影响过强。有效承接应表现为对相关历史敏感、对无关扰动相对稳健。

这一判据不是要求系统机械重复。相同任务可能允许多种合理策略，新的信息也可能使过去的教训不再适用。应评价的是变化是否与证据和目标有可解释联系，而不只是输出文本是否相同。对于随机系统，需要考察分布层面的策略选择与错误类型，并报告不确定性。一次成功或一次偏离都不足以代表长期能力。

时间顺序同样重要。若系统在收到反馈前就已采用新的策略，不能把变化归于该反馈；若研究者事后补写反思，记录就失去机制证据的价值。实验应保留真实时间戳、状态版本与关键行动，不要求暴露所有内部计算，却要足以重建可观察的因果链。记录的目的不是制造详尽叙事，而是让一个关于回写的主张有被检查的可能。

对长期研究，还需要观察失败恢复。系统能否发现记忆库中一条错误规则，修订相关技能，重新检查受影响结论，并在适当范围内通知使用者？这种能力比单纯在熟悉题目上提高得分更接近科研所需的后果承接。它把过去的错误变成未来的修复任务，也使知识的维护进入智能系统的能力定义。

错误记忆被发现后的处理，应区分当前使用状态与历史解释价值。系统可以先把某条规则标记为暂停采用，再查找依赖它的计算、摘要和任务，逐项决定重算或复核；对没有依赖该规则的材料，维持原状态。若只能整库清空，就会连带丢失仍然有效的知识；若只修改规则正文而不检查衍生成果，错误又会从旧摘要返回。局部隔离与依赖追踪因而是恢复能力的组成部分，针对这一机制的污染测试应仅在合成材料和实验副本中进行。

32.8 从承接历史到组织未来

记忆、目标与后果共同构成一个时间结构。记忆使过去能够进入，目标使未来行动具有方向，后果则检验并修改两者之间的连接。任何一项单独强化都可能失衡：只有记忆，系统可能重复旧路；只有目标，系统可能忽略已知代价；只有反馈，系统可能被短期结果牵引而失去持续问题。研究需要考察三者怎样共同保持一个能够更新的任务过程。

在 SDE 框架下，这个过程可以从不同路径进入。研究者可能先看到反复失败的显露，再追踪记忆与规则的纠缠；也可能先改变记忆组织，观察差异序列与最终结果如何变化。但无论从何处进入，都需要把第三维的改变及其回写追踪下去。增加一个记忆字段不是完成了发生分析，只有它改变了相关区分、行动或结果，并且这种改变得到证据支持，才形成机制性说明。

这里还应保留一种积极的稳定性。经过充分检验的规则能够减少重复试错，清晰目标能够避免研究被每个新现象牵走，可靠记忆能够帮助团队跨越人员和工具变化。稳定并非创造的敌人。创造常常依赖某些基础操作足够可靠，使注意能够转向尚未解决的问题。重要的是稳定规则具有边界和修订条件，而不是永远没有变化。

跨轮修订还可能只完成一半：新规则已经写入，检索索引尚未更新，旧行动却已再次启动。恢复程序应能确认同一次修订哪些步骤完成，避免重试时重复换算、重复扣除预算或再次发送同一行动。可以给修订建立稳定标识，记录提交状态，并在下一轮开始前核对任务版本与规则版本。这里提出的是待实现的结构要求。它使后果承接从“保留一段反思”推进到能够处理中断、重试和局部失败的持续工作。

一个跨轮系统最有潜力的变化，是从只记住已经完成的答案，转向维护仍然值得继续的问题。过去不再只是成功与失败的档案，也成为未来试验的组织资源。哪些解释尚未被排除，哪些异常需要等待新工具，哪些路线因暂时成本过高而被搁置，都可以进入一种有条件的持续维护。它们需要的不是无限记忆，而是能够在适当时刻重新激活的结构。

跨轮承接至此获得两个互相约束的要求：已知错误能够改变后续操作，尚未辨明的原因又不会被仓促写成永久规则。前者需要触发、版本和恢复机制，后者需要为暂未裁决的关系保留继续研究的入口。记忆系统若只保存定论，就会在压缩时丢失这种入口；若把一切异常都维持为活跃任务，又会挤占执行资源。第 33 章据此转向未决问题的组织，讨论哪些差异值得保留，以及证据和成本怎样共同决定继续、搁置与关闭。

第 33 章 未决问题与创造的条件组织

33.1 未完成并不自动产生创造

设想一个实验出现了异常结果，研究者提出三种解释，却因为没有经费继续检验而把问题搁置。十年后，资料仍在，后来者却不知道缺少什么证据，也不知道新仪器是否改变了研究条件。这个未完成问题没有因此积累认识力量。本章用“未决问题的组织”指称让它能够重新进入研究的关系：保留候选为何尚未裁决，连接已知证据和未尝试路径，并说明什么变化足以启动下一步。这种组织的价值取决于关系是否可用，悬置时间与档案体量均不能充当其量度。

创造也不能简单定义为保留更多可能性。一个系统每轮列出十种互不相关的解释，可能只是扩大了语言空间，却没有增加可检验的差异。研究中的可能性需要与对象、条件和证据相联系。它至少应说明某种解释为什么仍然成立，在哪些观察下会受到支持，又在哪些观察下应当放弃。没有这些约束，多假说只会使任何结果都能被事后解释。

未决问题与模糊回答不同。模糊回答可能回避承诺，让读者无法知道如何检验；未决问题则可以非常精确，例如尚不能区分两个机制，因为现有测量只观察到它们共同预测的结果。清楚表达这种限制，本身就是认识进展。它指出知识的边界在哪里，并把下一步实验从一般搜集资料转为针对性的区分。

创造的发生也不必总由异常触发。新的仪器、概念、材料或社会需要，可能使过去没有意义的问题变得可问。一个被保存的旧分歧，可以因新的测量方法而重新获得辨别力；一种原本只在局部有效的技术，可以与另一领域的问题发生联系。未决状态的价值因此是关系性的：它取决于问题与可用路径之间的组织，而不在问题自身拥有某种神秘能量。

若把尚未实现的研究可能比作“势能”，其含义限于问题与可用行动之间的准备关系，并非物理能量。准备程度可以分别考察：候选机制是否已分清，关键资料是否可取得，重新开启的触发条件是否明确，以及拟议检验能否改变当前判断。这些维度未必同步。一条路线可能理论上区分鲜明却缺少材料，另一条路线材料齐全却尚无有辨别力的操作；两者需要不同补充，不能由一个假说总数或创新总分代表。

SDE 框架在此关注的是，新显露如何打破已有区分，新的差异路径如何改变特征纠缠，以及这种改变如何使原来不可见的问题出现。创造不是凭空增加一个答案，而可能是

重组何者算作证据、何者能够比较以及何种差异值得保留。这样的重组仍需接受事实约束，不能因为具有原创性便免于正确性和用途的检验。

33.2 多假说保留的最小结构

构造一个科研数据处理争议：同一批材料测试曲线经同名流程处理，两位成员却得到不同的分组。原始记录尚未齐全，团队暂时保留预处理版本不同、一方沿用了旧任务的分组阈值、边界曲线的数值扰动三项解释。这是本章用于分析未决问题组织的设想，不是经验报告。候选登记首先要说明各解释涉及哪个处理节点、依赖哪些未核对材料，以及哪些变化会改变其地位。只有解释名称的列表，无法告诉接手者为什么尚未裁决。

预处理版本解释指向中间数据和变换记录，阈值继承解释指向运行配置，数值扰动解释指向结果对小变化的敏感性。三项可能共同起作用，也可能只是两位成员在描述同一缺失环节。档案需要把共享前提列出：他们是否真的使用同一原始材料，“同名流程”是否对应同一版本。只有这些身份先得到核对，后来的证据才有明确落点。保留多项解释的目的，是把尚未完成的核查分开，使新材料能更新相应候选。

每个假说需要一个稳定但可修订的身份。身份可以由核心机制、适用条件和关键预测界定，而非只由一句简短标题确定。若新证据迫使假说修改，应记录修改发生在哪里。否则一个解释可以不断改变细节以适配任何结果，却仍保留原名称，造成它似乎一直正确的错觉。修订是研究的正常部分，但修订本身应成为历史中的事件。

假说还需要证据状态。直接观察、间接支持、与预测一致以及尚无反证，不能写成同一种“已证实”。尤其是没有发现某种差异，可能来自样本不足或测量不敏感，不必然表示机制不存在。系统应保留研究能区分多大差异的能力范围。未决状态并非认识失败，有时正是对现有辨别力的准确描述。

不同假说之间的联系也值得记录。有的共享一个测量假设，有的依赖同一来源，有的只是同一机制的不同表述。若不识别这些依赖，系统可能把五个相互转述的观点当作五份独立支持。多假说管理应减少这种虚假的丰富性，使真正不同的预测获得位置。保留的是有认识差异的路线，而非语言上看似多样的说法。

最终，一个未决问题应指向可行动的缺口：需要哪种新证据、能否获得、代价如何，以及暂时无法获得时应怎样限制结论。这样的组织让悬置具有方向。研究者可以明确说，当前应继续使用一个较保守的工作模型，同时保留两个机制解释等待后续检验。行动上的暂时选择与理论上的最终裁决可以分开，不必为了继续工作而假装争议已经消失。

还应避免把互不相容的前提同时当作成立。保留多个解释，是保留它们作为候选的资格，不是断言它们都真。系统可以在不同假说下分别推导预测，再比较哪些结论对所有候选都稳健，哪些依赖特定前提。这种条件化推理能够支持暂时行动：即使原因尚未完全确定，也可能找到在多种机制下都值得执行的低成本检查。它使未决状态不必导致研究停顿，同时避免过早选择一个未经充分支持的故事。

33.3 搜索分支与科学假说的区别

语言模型可以在一次任务中展开多个候选推理分支。Tree of Thoughts 研究通过探索和评价中间文本状态，组织候选路径的搜索。[68] 这为科研候选生成提供了可借鉴的工程部件。把一个分支提升为持续维护的假说，还需要额外工作：确定外部对象、绑定证据来源、保存尚未裁决的原因，并在新材料到达时更新其地位。一次推断中的搜索选择与跨轮证据维护承担不同任务，二者的接合比候选数量本身更值得研究。

在一个拼图任务中，分支可以通过明确规则判断是否可行；在科学问题中，假说的价值往往取决于尚未获得的数据。一个当前看似较弱的解释，可能因为能预测新现象而值得保留；一个在已有资料上拟合良好的解释，也可能缺少独立检验。若直接把语言模型自己的评分当作假说的证据，就把生成与验证合并在同一个判断来源中。

因此，科研系统可以使用分支搜索来产生候选，却需要在外部证据层面重新组织候选。每条路线应区分推导是否有效、事实前提是否可靠和经验预测是否得到支持。模型能够检查逻辑一致性，但现实前提仍可能错误；检索能够找到支持文字，但文字可能只是在重复未检验的猜想。科学假说的维护需要这些不同证明责任彼此连接。

跨轮维护还要求容纳暂时不活跃的路线。一次搜索通常在预算用尽后选择结果，其余分支可能被丢弃；科学研究则可能因仪器限制暂时搁置某个解释，几年后重新进入。这里需要保存的是重新进入所需的信息，而非全部搜索轨迹。核心矛盾、关键预测和中止原因，比每一次语言生成更有价值。选择性历史能够支持长期研究，也避免记录量淹没问题。

多模型讨论同样不自动带来独立性。若多个模型共享相似训练资料、检索同一批来源并采用相同评价标准，它们的共识可能来自共同偏差。增加发言者数量与增加独立证据不同。真正有用的分工可以让不同角色承担不同检验，例如一方检查来源，一方寻找反例，一方分析测量假设；但这些角色仍需以材料和方法上的差异支撑，而不是仅改变语气。

分支结束时，可以追问哪些内容值得跨轮留下。用于措辞比较的临时分支，在答案确定后通常可以压缩；因缺少运行配置而无法裁决的机制分支，则应保留缺口及其关联材料。两者都含有显露、路径和特征之间的组织，差别在于它们是否仍能改变后续研究。由此，六路径提供的不同进入次序应落实为重新进入的线索，三原理所追踪的张力则落实为候选状态的更新理由。档案保存的单位是尚有认识作用的关系。

33.4 新问题怎样从旧矛盾中出现

创新有时发生在发现原问题遗漏了自己的目标。数据处理争议最初询问哪个分组结果正确，核对后却发现两位成员分别希望识别材料来源与比较使用性能。同一条曲线可能在前一用途下应归为一类，在后一用途下应被分开。问题由“找到唯一正确分组”改成“哪些特征足以支持哪一种比较”。原有曲线和计算没有被取消，改变的是共同对象及分组理由。保留早期争议的用途记录，使这次改写能说明自己来自何处，也让后续研究知道仍有哪些技术错误尚待排除。

这种改写可以从 S 与 D 的张力进入。一个结果在不同路径下反复变化，而原模型只允许固定结果，研究者便需要检查结果定义与路径的关系。也可以从 D 与 E 的张力进入：已有分类无法支持新的实验区分，迫使研究者提出新的显露方式。还可以从 S 与 E 的张力进入：某个稳定观察与已有关系结构不相容，促使新的差异序列被设计出来。三种追问都需要具体材料，不是任何矛盾都自动对应深层创新。

重要的是保留改写前后的联系。新问题若完全遗忘旧问题，就难以说明它解决了什么困难；若只换术语而保持全部操作不变，也未必构成实质推进。研究者应指出哪些旧结果仍被接受，哪些关系被重新解释，以及新框架增加了什么可检验承诺。这样的比较让创新能够被讨论，而不是依赖读者对新语言的直觉偏好。

人工智能可以参与问题改写，尤其在跨资料比较、发现定义差异和生成替代测量上提供帮助。但模型的流畅性可能使改写显得比实际更深刻。评价不能只看新问题是否新颖，还应问它是否有清楚对象、是否能获得相关证据、是否区分了真正不同的解释，以及解决它会改变何种认识或实践。一个优美但无法承受任何结果的问题，尚未成为成熟的科研问题。

人也可能因为长期训练而固守熟悉问题，机器则可能因为缺少领域约束而提出不可行建议。二者的差异可以成为协作资源：人指出材料与概念的限制，系统不断追问这些限制

中哪些是现实必需、哪些只是惯例。但这种协作需要尊重专业证据，不能把违反常识本身当作创造。真正的突破常常既改变一部分约束，又更严格地承担另一部分约束。

新问题出现后，还需要一个稳定阶段。概念必须被解释，测量必须被组织，竞争假说必须被展开。若系统始终追求下一次改写，就无法积累证据。创造与稳定因此是同一过程中的不同任务：前者打开新的区别，后者使区别能够被他人识别和检验。发生学的雄心应包括这种成熟工作，而不只赞美最初的灵感。

问题改写还可能来自测量之外的参与者。实验使用者、被研究群体或设备维护者，可能指出研究团队没有记录的条件。这些意见不因来自非专家便自动错误，也不因贴近实践便自动正确。需要将其转化为可讨论的具体差异：哪项操作发生变化，谁在何种条件下看到何种结果，现有设计是否允许该结果进入。创造性的开放包括对证据来源的开放，但仍要求清楚说明材料怎样支持主张。

33.5 为保留与裁决配置资源

每条未决路线都需要维护成本。它占用注意、检索、实验机会和团队讨论时间。保留全部假说既不可行，也可能妨碍重要问题推进。因此需要一种有理由的资源配置：哪些路线继续活跃，哪些暂时归档，哪些已经有充分理由关闭。这里不存在一个适用于所有领域的统一阈值，选择取决于错误代价、证据获取成本及问题的重要性。

可以把下一步行动的价值理解为它预期能够改变多少与决策相关的不确定性，同时考虑成本和潜在损害。这是一项分析原则，不是已经得到验证的通用评分公式。某个便宜实验若能区分两种关键机制，可能比大量重复已有结果更值得做；某个昂贵实验即使信息丰富，也可能因为当前用途有限而应被推迟。研究资源的配置需要把认识收益与现实约束一起说明。

还要区分关闭假说与停止投入。一个解释可能尚未被推翻，但因为缺少可行检验而暂时不再投入；另一个解释可能已经与可靠证据冲突，应明确降低其地位。二者若都写成“放弃”，未来研究者便无法知道是证据裁决还是资源限制。归档理由保留了重新开启的条件，也防止历史被整理成只有胜者和错误者的简单叙事。

持续保留少量异议路线，可以防止主流解释过早垄断资源，但异议也应承担最低证据责任。不能仅凭它不同于多数就获得无限支持。较合理的要求是指出具体未解现象、可检验差异或现有解释的限制。这样既为新方向留下空间，又不把全部争议都等价处理。开放研究需要可进入的标准，也需要可退出的理由。

资源配置还影响系统如何表现创造性。如果奖励更多假说，系统可能生成大量低价值分支；如果奖励快速完成，它可能过早收束；如果奖励被专家认可，它可能趋向熟悉理论。评价制度应同时检查候选的差异性、后续可检验性和实际证据贡献，并保留对失败探索的合理记录。没有一种单一分数能够完全代表创造，指标只能辅助具体判断。

在团队中，路线维护需要负责者和复查触发点。上述版本解释可以等待原始配置取得，目标差异问题则可等待新的用途说明；二者不必按相同周期进入会议。触发条件满足后，也应先核查问题是否仍重要，再恢复投入，以免任何新材料都使旧路线重新活跃。若负责人离开，应交接候选状态与缺口；无法交接时可以降低维护级别，明确可能遗漏的后果。这种有结构的等待使长期问题不依赖持续占用注意，也让保留成本可见。

33.6 反例如何成为创造的材料

反例的第一项功能是限制主张，而不是装饰论证。若一个系统声称结构化记忆总能改善科研，某些任务中的负迁移就必须使主张收缩。收缩并不等于理论失败到一无所有；它可能揭示记忆只有在任务关系稳定、检索准确或更新规则适当时才有效。反例把原来过宽的命题转化为有条件的机制问题，这本身可以产生新的研究方向。

但反例也不能被无限吸收为“复杂性”。如果每次不符合预测都被解释为尚未识别的纠缠，理论便不会承担风险。需要事先说明哪些结果会削弱特定机制，哪些条件变化属于模型已经允许的范围。对于 SDE 研究，尤其应防止将全部现象重新描述为三个字母便宣称解释成功。创造需要新的区分，解释则需要证明这些区分产生了可检验增量。

在数据处理设想中，若运行配置证明双方采用相同的当前阈值，而非一方沿用旧阈值，该项解释就失去支持；它不应继续以“其他历史影响”改名留在原地。若核对版本后仍有差异，版本解释受到限制，尚存的数值敏感性也没有因此自动成为答案。每次更新应记下被排除的是哪一前提、哪些问题仍然开放，以及新增材料是否足以让某条路线停止投入。反例的作用表现为候选资格发生可追踪的变化。

若进一步发现两次分组服务于不同用途，原来的结果冲突便需要分流：一部分属于计算路径是否一致，另一部分属于何种相似性值得保留。系统应建立相互关联的两个问题，分别迁移仍然适用的证据。直接把旧条目判为已解决，会丢失尚未核查的处理误差；把它原封不动保留，又会让后来者继续寻找并不存在的唯一分组。对象修订要求重组证据的归属，使旧争议的一部分结束，另一部分以更精确的问题继续。

人工智能可以帮助系统搜寻反例、检查术语变化和追踪来源，但也可能生成不存在的反例或错读限制条件。反例管理因此应像正面证据一样要求来源核对。虚构反例可以用于逻辑测试，却必须明确标注；真实经验反例需要可检验材料。将二者混合，会让研究在貌似严谨的对抗中失去事实基础。

一旦反例得到确认，系统应更新相关结论和后续行动，而不只是把它存入“局限性”栏目。若反例影响核心预测，就应调整研究设计；若只限定适用范围，就应在传播结果时明确范围。创造性的反例处理，是让失败真正改变知识组织。它既需要愿意承认错误，也需要保留足够结构，使错误能够指向下一项有意义的工作。

反例的保存还需要防止选择性遗忘。研究者可能详细记录支持新框架的异常，却把不利于新框架的稳定结果视为普通情况而略去。如果本书主张某种回写结构重要，那么没有回写或回写不影响结果的情形同样有研究价值。它们帮助确定机制何时可以忽略，也为较简单模型保留合理领域。创造性研究并不要求每次都发现复杂关系，而要求根据证据决定何种复杂性值得承担。

33.7 美的判断与可交流的可能性

研究者有时会说一个解释很美，一个问题有潜力，或一条路线值得继续。这类判断可能先于完整证据，却不应被简单视为非理性。它们可以表达对简洁、统一、生成力或问题空间的初步把握。但这些词同时也容易掩盖偏好。若“美”成为拒绝检验的理由，它便失去科研价值；若完全排除这种预先判断，研究又难以在证据尚不充分时分配探索资源。

本书把美作为某种可能性组织的解释对应，而非物理属性与伦理价值之间的等式。在 AI 研究中，更适合追问一个候选为何被认为有潜力：它是否用较少假设连接原来分散的现象，是否产生新的可检验预测，是否揭示现有问题的共同限制。把这些理由说清，可以让直觉进入共同讨论，同时允许它在经验面前被修订。

一个解释的简洁也可能来自遗漏。若模型删去了难以处理的群体差异，形式上会更整齐，却可能降低实际解释力。统一性同样需要代价核算：把不同领域的问题都称为不收敛，可能启发比较，也可能抹去统计、动力学和测量层面的根本差异。美的判断应与保存了哪些重要区别相联系，而非只追求语言和形式的整齐。

人工智能可以生成多种表述，帮助研究者看到同一问题的不同结构，但这也增加了被修辞吸引的风险。评价候选概念时，可以暂时移除漂亮名称，只保留定义、预测与例子，

检查它是否仍然具有区分力。也可以要求用一个反例说明适用边界，用一项可执行研究说明潜在贡献。若去掉修辞后没有剩余内容，就应降低对其创造性的评价。

不同研究传统可能对值得追求的可能性有不同判断。数学结构的统一、实验现象的丰富和社会问题的迫切性，都可能构成合理理由。它们不一定能够换算成同一尺度。人机系统应帮助显露这些评价依据，而不是假装一个自动评分可以完成全部取舍。能够比较理由的结构，比一个神秘的“创新分数”更适合组织跨学科合作。

因此，美的研究价值可以保持为一种可交流、可受约束的先导判断。它允许在结论尚未确定时选择值得投入的路线，却必须接受后续证据与成本的检验。未决空间的质量，不在于它永远不被确定，而在于它能够孕育有根据的新确定，同时保留那些尚未得到充分裁决的差异。

对潜力的判断还应允许事后学习。某条被认为优美的路线后来没有产生可检验证据，可以反过来检查最初判断依赖什么：是过分看重形式统一，忽略了测量难度，还是低估了现实成本。这种复查不应以最后成败机械否定最初选择，因为合理探索也会失败；它应比较当时可得信息与实际判断。长期积累这样的记录，能够使团队逐渐改善对值得探索之事的辨别，而不把创造完全交给不可交流的天赋。

33.8 从问题档案到创造性研究制度

为了让上述分析进入实践，可以设想一种问题档案。每个条目记录当前问题、竞争解释、关键证据、已尝试路径、中止理由和重新开启条件。它不要求统一所有学科的内容，而要求说明问题为何仍值得保留。档案的单位是能够继续推进的认识缺口，不是论文标题或聊天片段。这样，过去的探索可以成为后来研究者真正可用的起点。

问题档案不应成为新的行政负担。若填写成本高于研究收益，研究者会用模板化语言应付。较好的做法是在重要决策节点留下简短而具体的记录，并由系统帮助连接已有材料。自动摘要可以降低成本，但关键解释仍需核对。尤其是中止原因与证据状态，不能由模型根据最终结果事后猜测。档案应保存真实的不确定性，而非制造一条看似合理的成功叙事。

组织层面可以定期从档案中选择少量问题，检视是否因新工具或新资料而获得推进机会。选择应兼顾重要性、可行性与路线多样性，并公开为什么暂时不投入其他问题。这样的制度让未完成工作得到有限而真实的延续，避免研究价值只由是否已经形成完整成果决定。但它也必须允许清理：长期缺少可说明内容的条目可以关闭，并保留简要理由。

对于 AI 系统，问题档案可以成为与普通记忆库不同的研究对象。普通记忆偏向回答“过去发生了什么”，问题档案还要回答“未来什么变化会重要”。检索因此可以由新证据触发：当某项技术出现，系统查找此前因相同限制而中止的路线；当某个结果被推翻，系统查找依赖它的未决解释。这样的连接若能改善真实研究表现，才构成关于创造条件的工程证据。

这项设想尚待第 35 章的对照研究检验。其独立贡献应体现在两类损失减少：新条件已满足却没有重开有价值路线，以及条件未改变却反复投入已经无望的路线。普通资料检索和标准任务清单同样可以保存历史，应作为实质比较对象。评估需记录问题维护是否改变了后续选择，并把填写、核对与误触发的成本计入。如果档案只是更完整地重述既有判断，却没有改善这些选择，就应压缩结构，保留确有用途的字段。

未决问题的组织由此取得一种有限而明确的创造作用：它让暂时无法裁决的差异保存身份，在条件变化时获得再次行动的入口，并允许证据使一部分路线真正结束。问题档案的价值需要由这些转换来承担。候选经维护而重新进入研究后，还会遇到另一项责任：新材料究竟支持了哪一个主张，何时足以形成可使用的知识，何时仍应维持未决状态。第 34 章围绕这种证据地位的转换展开人机判断。

第 34 章 人机科研中的判断与知识发生

34.1 从候选答案到能够承担证明的主张

第 33 章保留的是仍有研究价值的候选，本章处理候选地位怎样改变。研究者得到一段关于低温偏移的解释时，需要判断哪些内容是数据概括，哪些提出原因，哪些已经足以指导修正。三种地位要求不同支持；一句话写得更肯定，并不会使它完成转换。人机科研中的判断，就是在核对前提、检验推导和比较解释的过程中，决定哪项主张可以升级，哪些部分仍须保留，哪些行动已有足够依据。

可以用一个贯穿本章的假设情境说明：某类湿度传感器在较低温度下出现读数偏移，团队希望查明原因并改善测量。AI 系统提出参考装置漂移、封装材料影响和补偿算法失配三种解释。此时系统完成的是候选生成，尚未完成因果判断。即使三种解释都符合专业常识，也需要针对实际设备、记录和使用条件逐一检查。一般知识不能替代对当前事件的证据。

团队随后查阅设备文档、整理校准历史、比较不同批次的数据，并设计能够区分候选机制的测试。某些步骤由人完成，某些由软件完成，但证据地位取决于工作性质，而非执行者身份。人工读错单位同样会破坏推论，机器正确整理资料也可以成为可靠环节。人机分工应围绕可验证的任务组织，而不是预先把判断全部归给人、把机械劳动全部归给机器。

一个主张的支持结构应容纳分层结论。在传感器设想中，重复读数可以支持特定条件下确有偏移，校正程序通过保留数据检验可以支持限定范围内的补偿，区分性实验才可能支持某项原因解释。团队可同时接受前两项而继续悬置第三项。把这三项写为相互连接但可分别修订的主张，能够避免机制争议使已可靠的校准成果一并失效，也避免校准成功替未经辨识的机制背书。判断的进展表现为支持范围变得更精确。

判断还包含对证据充分程度的选择。科研不可能在任何行动之前消除全部不确定性，但不同用途需要不同支持。为安排下一次探索实验而暂时采用一个工作解释，与据此宣布普遍规律，证明责任不同。人机系统需要维护用途与证据之间的关系：当前材料足以支持什么，哪些扩大使用还需要新增检验。这种对应使知识能够逐步成熟，而不必在绝对确定与完全无知之间二选一。

SDE 分析将注意力放在主张如何在这些关系中获得稳定性。新的显露来自数据与处理，差异序列包括检索、比较、实验与判断，特征纠缠则组织对象、指标和机制之间的联系。任何一处变化都可能要求重读其余部分。知识不是由孤立答案累加而成，而是由能够相互支持并接受修订的关系逐渐形成。

34.2 来源与证据不是同一件事

找到来源是证据工作的起点。文献中出现某个句子，并不意味着它支持当前主张；来源可靠，也不意味着其中每项结论都适用于新的条件。系统需要区分原始实验、方法说明、综述转述和作者猜想，并检查研究对象、测量条件与结论范围。把一段相关文字附在答案后面，可能增加可信外观，却没有完成证据对应。

在传感器例子中，一篇论文可能发现某种封装材料在特定湿度循环下影响响应，但当前设备使用的材料和温度范围不同。它可以支持一个候选机制，却不能直接证明当前偏移来自相同原因。设备文档可能说明补偿算法的适用范围，却不能证明实际软件版本已经正确执行。不同来源提供不同层次的信息，必须通过当前材料连接起来。

检索增强系统能够把外部资料带入生成过程，但检索相关性与证据充分性仍是两项任务。一个页面可能因关键词相近而排名很高，却在关键定义上不一致；多个页面也可能转述同一项研究，不能当作独立重复。研究系统应追溯关键事实的原始出处，并标明哪些材料共享证据来源。来源网络的独立性比链接数量更有意义。

还需要区分来源不存在与当前无法取得。系统检索失败时，可以报告尚未找到可核验材料，却不能据此断言某种现象从未被研究。相反，无法阅读全文时，也不应根据标题和摘要推断详细方法。证据工作要求诚实标记访问范围，并将未核实内容保持为待核实状态。这样的限制不是降低效率，而是避免后续推论建立在伪装完成的核查上。

资料版本也会改变证据地位。预印本、正式论文和更正说明可能存在差异，技术文档也可能随设备更新。系统应记录使用的版本与访问时间，对关键更正建立检查。若一个结论后来被撤回，相关综述不能因为已经写成便保持原样。前章讨论的历史承接，在这里成为知识维护：变化的来源需要能够触发下游主张的重新评估。

主张与来源的对应应细到能够改变判断的程度。假如设备文档只说明补偿算法适用于某一温度区间，系统可以据此标记区间外使用需要重审；它还须另找运行配置，才能判断当前装置是否真的使用该版本。两项材料承担不同连接，缺少任何一项都会留下不同缺

口。这样的组织使检索结果产生明确后果：保留一个候选、限制一个适用域，或要求核对一个实现步骤。资料数量只有在增加了这些约束时才构成证据收益。

证据对应还可以通过反向检查强化。先从报告中的关键句出发，要求指出直接支持它的材料，再从材料出发判断句子是否扩大了范围。若文献只支持“在特定装置中观察到相关变化”，报告就不应写成“该机制普遍导致偏移”。这类检查适合由另一位研究者或独立程序协助，但仍需保留人工抽查。其价值不是让每句话都附带繁重证明，而是优先保护支撑主要结论的少数关键连接。

34.3 推导 检验与解释的分工

推导回答在给定前提下能得出什么，检验回答现实材料是否与某项预测相符，解释则试图说明为什么出现这样的结果。这三种工作互相支持，却不能相互替代。一个推导完全正确的模型可能建立在错误前提上，一个预测准确的模型可能尚未揭示机制，一个合理机制也可能在当前数据中无法辨识。人机系统需要在报告中保留这些差别。

AlphaFold 的研究通过外部评估展示了深度学习在特定蛋白质结构预测任务中的重要能力。[69] 在本章的证据分工中，这一成果首先支持关于所评估预测任务的判断；结构预测又可以成为进一步选择研究对象、提出功能问题的条件。后续问题需要相应证据，动态变化与功能解释不会随预测评价一并完成。这里的推进具有累积性：已经成立的预测能力改变下一项研究可以怎样开始，新的机制工作则继续承担自己的证明责任。

在传感器例子中，一个模型可以准确校正低温读数，却未必说明偏移来自封装、参考装置还是算法。若目的只是某一固定环境下的临时校准，这种预测能力可能已经有用；若希望迁移到新材料和新设备，机制理解就更重要。系统应根据用途说明当前方法的充分性，而不是自动把预测成功升级为因果解释，或因缺少完整机制而否定一切实用价值。

计算验证在其中可以承担明确任务。单位检查、极限情形、数值稳定性和独立测试，能够发现推导与实现之间的错误。但通过这些检查不等于现实模型正确。一个程序可以忠实执行不合适的统计假设。研究流程应分别检查代码是否实现预期方法、方法是否适合数据结构、数据是否对应研究问题。多层检验避免把某一层的成功当作全局保证。

解释比较还需要强竞争者。若只让新机制与一个明显过简的基线相比，胜出并不令人信服。应选择领域中已有的合理解释，明确它们能够处理哪些事实，再指出新解释增加了什么。对 SDE 研究同样如此：不能因为传统单变量模型表现较差，就断言任何发生模型都更优；需要比较条件化模型、反馈模型和其他成熟方法，辨认真正的增量。

有时最有价值的结论正是目前不能区分解释。若不同机制对已有数据给出相近预测，继续增加相同条件下的样本可能帮助精确估计，却未必改善机制辨识。此时需要改变观察或干预结构，而非只扩大数据量。人机协作的价值可以体现在发现这一点，并把研究从重复拟合引向新的可区分设计。

假设传感器数据同时随温度和测量日期变化，模型可能很容易拟合偏移，却难以区分温度效应与设备老化。若所有低温测量都发生在后期，增加相同安排下的样本仍不能解决这一混杂。下一步应考虑交错安排条件、使用独立参考或引入能够区分两种解释的观察。这里没有预先规定某一种实验必然成功，而是说明判断如何从拟合结果返回数据生成结构。新设计的价值在于改变可辨识性，不仅在于提高拟合精度。

34.4 判断在何处需要暂停

人机系统越能连续执行，越需要知道何时暂停。暂停不应只是遇到技术错误时的中断，也可以是认识上的决定：来源互相冲突、目标含糊、证据不足以支持高后果行动，或下一步超出权限。能够显露这些情况，是研究能力的一部分。无论速度多快，若系统不能识别自己何时缺少决定所需条件，连续行动反而可能扩大错误。

在假设的传感器研究中，AI 发现两份校准记录使用不同参考标准，但没有资料说明转换关系。继续合并数据可以得到更完整的图形，却可能制造虚假趋势。合理做法是暂停合并，保留两套结果并提出具体缺口。这里的暂停具有行动内容：联系记录负责人、核对设备版本或设计桥接测量。它不是泛泛表达不确定，而是阻止无根据的步骤继续传播。

暂停也可以针对目标。若团队要求系统同时最小化测试成本、最大化精度并保证所有批次可比，而现有资源无法兼顾，系统应将冲突交给有权选择的人。替团队默默牺牲其中一项，会使最终结果看似满足任务，实际却改变了承诺。清晰暴露取舍能够提高协作质量，也使后来评估知道某项限制来自有理由的决定，而非无意遗漏。

检查点还应区分暂停范围。两套参考标准无法换算时，应暂停跨标准合并及依赖合并的机制推断；各自标准内已经完成的质量检查和文档核对仍可继续。若所有工作一并停止，必要审查就会显得过于昂贵；若只标记一条警告而继续生成汇总，又会让不确定性被后续计算掩盖。系统需要列出当前缺口影响哪些主张和行动，再把需要新判断的部分交给有权复核的人。暂停范围随依赖关系确定，使认识上的克制具有可执行边界。

然而，过多暂停也会使人机系统失去实际价值。系统若对每个小问题都要求负责人决定，就把认知负担全部转移给人。应当区分可以依据既定规则解决的局部问题，与需要新

的价值或研究判断的问题。经过验证的默认处理可以减少无效中断，但默认值必须可见，并允许在条件变化时被撤回。可靠自主性来自明确边界，而非无限授权。

暂停后的恢复同样重要。新增信息到来时，系统应知道从哪个节点继续，哪些下游结果需要重算，哪些判断仍然有效。若每次暂停都意味着重新开始，团队会倾向绕过必要检查。历史记录、依赖关系与任务版本能够降低恢复成本，使审慎与效率不必互相对立。一个能够有理由暂停并可靠恢复的系统，比从不停止的系统更适合持续研究。

34.5 共同判断中的分歧与权威

人机协作并不保证共识，也不应把共识作为唯一目标。研究者可能认为某个异常来自仪器故障，模型却从资料中找到另一种机制；模型可能依据一般规律提出建议，实验人员则知道当前装置存在未记录限制。分歧的价值取决于它能否转化为可检查的理由。若只是人以身份压过机器，或机器以大量引文压过人，争论都没有触及证据。

应区分经验权威与决定权限。经验丰富的人员可能更了解装置，但仍需在关键主张上提供可理解依据；项目负责人有权分配资源，却不能凭权限使一个事实命题变真。模型在特定检索或计算上表现优异，也不因此取得所有领域的判断权。不同能力与角色可以被尊重，同时保留对具体结论的检验。

分歧应先拆到能够判定的主张。在传感器案例中，实验人员可能同意校正后的误差已经满足当前测量用途，却反对将它解释为封装机制得到确认；模型生成的报告则可能把两项结论连写。双方在校准有效性上没有冲突，争议集中在机制支持范围。把主张拆开后，可以保存可用结果并安排后续辨识。若争议转为是否立即部署，则还需列出使用者能接受的误差与维护成本。新增数据能够改变事实判断，部署门槛仍需要明确的用途理由。

分工可以围绕证据链上的不同薄弱处安排。一项检查核对设备文档及其版本，另一项重新计算校正后的误差，还有一项专门检查报告是否从预测效果越级推到因果机制。每项检查应取得对应材料，并留下能够重做的结果。若三次检查都只阅读同一份生成摘要，它们就共享摘要遗漏的风险。实质独立性来自核查入口和验证操作的差异；讨论者人数增加，只有伴随这些差异时才扩大证据覆盖。

人也会受到机器表达的影响。结构清楚、语气肯定的答案可能让人过早接受结论；反复强调不确定的答案则可能使有根据的结果失去应有权重。系统应通过证据状态、条件说明和可核查材料支持判断，而不依赖表达姿态。研究者也需要学习如何质疑模型：要求它说明反例、适用范围和关键依赖，比简单要求“再想一遍”更接近科学检查。

共同判断的成熟不是消除分歧，而是使分歧能够产生下一步工作。有时需要补充实验，有时需要限制结论，有时需要承认价值上无法统一而保留不同方案。只要理由清楚，暂时分歧可以成为研究资源。相反，一个通过压低异议获得的统一答案，即使短期方便，也可能使真正重要的条件继续不可见。

为了避免分歧讨论被最先出现的解释固定，团队可以先让成员分别形成简短判断，再共同查看理由和材料。这是一项可研究的组织建议，尚不构成其有效性的经验结论。独立阶段帮助保留不同观察，共同阶段则负责比较与修订。若模型先给出完整故事，人类再逐项评价，可能很难发现故事之外的问题；若所有人永远独立，又会失去相互启发。分阶段安排试图让独立性与互动各自在适当时机发挥作用。

34.6 知识贡献与作者责任

人机系统产生结果之后，需要说明各方贡献。贡献可以包括提出问题、整理材料、设计分析、实施实验、核对来源和解释结果。它们不应被简单压缩为谁写出了最多文字。生成式系统尤其容易让文本生产与知识责任混淆：一段由机器组织的文字可能表达人的完整研究，也可能夹带未经核实的主张。责任判断需要回到具体工作，而不是依赖最终句子的来源。

对于正式研究，人类作者应能够说明并承担所发表主张的依据。使用工具不意味着必须亲自完成每一步运算，但必须知道哪些步骤经过何种检查，以及在被质疑时如何重新核验。无法解释关键来源、无法重建主要分析、无法辨认模型添加的内容，都会削弱成果的可负责性。技术协助可以扩展研究能力，却不能成为责任消失的接口。

同时，承认人的责任不应抹去技术和基础设施的贡献。模型、软件、数据集、设备和维护工作都可以在适当位置得到说明，使读者理解成果如何形成。这里的透明不是礼节性致谢，而是方法的一部分：不同版本和配置可能影响结果，读者需要知道复现所需条件。对工具的描述应具体到与结论相关的层次，避免只写“使用 AI 辅助”而没有可检查信息。

研究劳动的重新分配也值得关注。AI 可能减少初稿和代码编写时间，却增加核查、整合和维护负担。如果这些新增劳动被低估，团队可能把看似节省的时间转化为更高产出要求，最终削弱验证。评价人机科研效率时，应包括从问题提出到可靠交付的完整成本，而不是只计生成答案的几分钟。知识质量依赖的工作不能因为不显眼而被排除。

长期合作还要求关键判断能够交接。假设最熟悉校准流程的成员离开，接手者应能从报告找到采用的参考标准、排除某项解释的理由和仍未验证的使用范围。若只有原作者能够说明，成果的可用性便依赖一个未被记录的判断条件。可以要求不同成员独立重建一项代表性结论，发现交接中缺少的材料；这是一项待检验的组织建议。其目的在于让工具减少的操作劳动转化为可共享的判断能力，而非把核查集中到少数无法替代的人身上。

因此，知识贡献的记录应既承认协作，又维持可追问的责任。它不必为每个字计算归属，而应说明谁对哪些关键判断负责，工具在哪些环节发挥作用，哪些结果经过独立验证。这样的结构使成果能够被他人信任、质疑和继续使用，也避免把共同产生的知识误写成某一个孤立主体的奇迹。

责任还涉及对意外贡献的处理。系统可能在整理资料时发现原研究目标之外的问题，成员可能在核查过程中提出更重要的解释。这些贡献应能够进入记录，而不因最初任务分工而被忽略。相反，自动生成的大量文本不应仅凭数量获得与关键判断相同的地位。清楚区分劳动量、原创贡献和结果责任，有助于团队形成公平的合作，也使工具的使用不至于遮蔽真正决定研究方向的工作。

34.7 知识如何在使用中继续发生

研究结论被发表后，会进入新的情境。传感器校准方法可能被另一实验室采用，材料分类可能被工程团队用于选择设备，教育评价模型可能被学校用于分配资源。每次使用都可能改变原结论的条件。知识的稳定性因此不仅来自最初检验，也来自对迁移边界的持续说明。一个结果可以在原情境中可靠，同时在新用途下需要重新验证。

迁移首先需要识别哪些条件与机制相关。若校准方法依赖某种参考标准，新实验室就必须检查其标准是否可比；若模型只在有限温度范围中得到验证，就不能默认为范围外同样有效。系统应把这些条件作为成果的一部分保存，而不把它们埋在难以发现的附注中。使用者需要能够快速知道哪些变化值得警惕。

使用中的新结果也可能回写原知识。若多个实验室报告相同偏差，原来的模型可能需要扩展；若只有一个场所失败，则应检查局部条件与执行差异。不能自动把失败归给使用者，也不能仅凭一次失败否定全部原结果。知识维护需要保留问题报告的证据结构，使不同来源的经验能够被比较，而不是在支持与反对的立场之间选择。

人机系统可以帮助追踪这种回写，例如将新报告连接到相关假设，提示哪些结论需要复核，或者整理不同版本的适用范围。但它不能仅凭文本相似度决定某项研究已被推翻。

需要核对对象、方法与证据强度，必要时交由专业共同体判断。自动化适合降低发现关联的成本，最终解释仍须承担具体证明责任。

使用反馈也应回到分层主张。新实验室发现校正失效，可能削弱跨设备适用性，却未必否定原装置内的检验；新的机制证据出现，也可能提高原因解释的地位而不改变现行校正程序。系统应据此保留仍受支持的部分，限制或重开受影响的部分。经过多种条件验证后，某项关系的适用范围可以逐渐扩大，形成更稳固的知识。稳定来自支持范围与实际条件的持续对应，修订则作用于其中已经发生变化的连接。

这也使研究成果具有一种可维护的形式：核心结论、证据依据、适用条件、已知限制和更新入口彼此连接。人机协作若能帮助建立这种形式，其贡献就超出加快写作。它参与了知识在时间中的延续，使过去的结果能够接受未来的修订，同时不必每次从头重建整个研究历史。

知识使用的反馈还可能暴露目的变化。原本为实验室测量开发的方法，一旦用于商业质量判定，错误代价和证据门槛都会改变。即使技术条件相同，原来的验证也未必足以支持新的决定。系统需要把用途变化识别为重新评估的触发，而非只检查输入数据是否相似。测量可靠性、决定公平性与实际后果分别提出要求；承认它们相互关联，不意味着可以用一个预测指标替代全部判断。

34.8 判断如何转化为可比较的研究能力

人机科研中的判断不能只通过最终答案正确率来衡量。系统可能偶然给出正确结果，却缺少可靠理由；也可能在证据不足时正确拒绝作出结论，而被单一得分视为失败。需要分别观察候选生成、证据对应、机制辨识、行动选择和错误修复等能力。它们共同构成研究过程，但对不同任务的重要性并不相同。

评价应特别关注从不确定到确定的转换是否有根据。系统何时提高信心，是否对应新增证据；何时关闭假说，是否说明反证或资源理由；何时改变目标，是否得到授权；何时发布结论，是否与证明范围相称。这些问题可以通过事件记录与独立评阅检验，比要求系统自述“我进行了严谨思考”更有证据价值。

判断质量可以通过结论在材料变化前后的转移来考察。当前只有相关数据时，系统应保留机制候选；补入可区分证据后，应相应提高或降低特定解释的地位；关键来源被更正后，应重审依赖该来源的主张。评价者即使不同意最后的选择，也可以核对这些变化是否

有依据。对于材料本来不足的任务，保持未决是一种准确状态；对于材料已经充分的任务，无限推迟则成为需要解释的失败。具体的盲评和终点评分留待第 35 章安排。

研究能力还包括成本与可恢复性。一个系统若需要大量人工筛选才能产生可靠成果，其有效性应连同筛选劳动报告；一个系统若能快速生成结果却无法追溯错误，长期价值可能有限。相反，能够留下清楚依赖关系、在发现问题后局部修复的系统，可能在长期合作中更有优势。这些差异需要持续任务和故障情境才能显露。

由此，人机科研的判断可以被理解为对主张、证据和行动之间关系的维护。它既需要生成力，也需要限制力；既需要推进任务，也需要在条件不足时暂停；既需要稳定成果，也需要保留修订入口。这些能力不属于某一种语言风格，也不能由一个通用智能标签充分说明。它们要在具体任务、架构与制度中接受比较。

本章得到的判断单位，是能够分别接受、限制和重开的主张，而不是整份回答的笼统正确性。它要求来源有对应、预测与机制有分工、暂停有范围、使用反馈能进入恰当层次。这些要求现在可以转为对系统的具体检验：在相同材料下，组织方式是否减少无依据升级，是否更准确定位受损结论，是否在条件满足时恢复有效行动。第 35 章据此提出尚待实施的对照设计，测量其收益以及核查、维护和等待的成本。

第 35 章 AI 研究系统的对照实验

35.1 把研究纲领拆成能够失败的命题

一个系统拥有长期记忆、保留多个假说并记录行动后果，是否就比普通智能体更善于科研？前面几章给出了理由，也给出了可能失败的机制：记忆可能传播错误，多假说可能浪费资源，反馈可能强化短期偏好。因此，实验不能只展示一个成功案例，而应在可比条件下检验这些组织方式是否带来特定增量。以下方案均为尚待实施的研究设计，不代表本书已经取得相应数据或效果。

首先需要区分三个主张。第一，结构化历史可以改善相关经验在新任务中的使用。第二，持续维护未决假说可以提高实验的辨别力和条件变化后的恢复能力。第三，将反馈连接到具体行动修订，可以减少同类错误而不过度扩大旧规则。这些主张彼此关联，却不应合并为一个无法定位原因的“新典范更优”。实验要允许某项成立、某项无效，甚至某些组合产生负面影响。

这些实验的推断对象先限定为具体组织方式的工程效果。三项因素的主效应回答各项增量平均改变了什么，交互作用回答它们在所选响应尺度上是否互相增强或抵消；过程记录再用于检验变化是否经过预期连接。普通任务管理若取得相同效果，新组织尚未显示额外效果优势；两者是否通过共同机制起作用，仍须由针对该机制的干预与消融另行判断，不能从效果相同直接推出。未检出差异也不等于已经证明效果相同；若要声称等效，应另行预定有实际意义的等效界限及相应分析。这样的结果既可支持或限制本书的工程建议，也可反过来修正理论提问，但三号位的本体论主张仍需要独立论证。

主要比较对象应是具有合理工具和普通记忆的成熟工作流，而不是刻意削弱的单次问答。单次模型可以作为参考点，帮助了解持续组织的基本收益，却不能承担全部竞争解释。新系统与强基线应使用相同基础模型、资料范围和行动权限，尽可能保持资源可比。若某种结构天然需要更多时间，应把效果与成本共同报告，而非隐藏新增投入。

研究任务也要与主张匹配。一次独立知识问答很难检验跨轮承接，只有重复熟悉题目又无法检验迁移。任务序列应包含相关经验能够发挥作用的部分，也包含关键条件改变、旧规则需要放弃的部分。只有同时观察保持与更新，才能判断系统积累的是可用知识还是僵化惯性。对“历史厚度”的检验，必须让历史有真实的机会产生利弊。

最后，成功标准应在观察结果前明确。对于合成环境，可以利用已知机制评价预测与实验选择；对于真实资料，需要独立核查和盲评；对于人机团队，则还需记录人工时间、理解与复核能力。不同任务可以有不同主要终点，不必强求一个普遍科研智力分数。共同之处是让主张、操作与证据之间有清楚对应。

35.2 三项组织因素与强基线

第一项因素是记忆组织。基线采用普通的时间顺序记录和检索摘要，实验条件则增加事件、来源、解释、适用条件与版本之间的显式关联。两者获得相同原始信息，并使用相近的存储和检索预算。这样比较的是组织方式，而不是有资料与无资料。若结构化条件需要额外人工标注，其时间应计入成本，且不能悄悄提供基线没有获得的专业解释。

第二项因素是未决问题维护。基线可以产生多个候选并保留普通笔记，实验条件则要求持续记录候选之间的竞争关系、证据状态、关键预测和重新开启条件。候选生成次数应受到共同预算限制，避免把更大搜索量误当作维护机制的效果。其预期优势主要出现在新证据到来、原解释受挫或任务条件改变时，而不必在每一次简单问答上都出现。

第三项因素是后果连接。两种条件都收到相同反馈，基线将反馈作为普通信息使用，实验条件则将其关联到具体错误节点、修订动作和未来触发条件。实验不应通过只给新系统详细反馈来制造优势。它要检验的是在信息相同的情况下，显式连接能否改善行动改变。若基线自行完成了相同连接，也应记录，这意味着机制可能已经存在于普通工作流中。

三项因素各设两个水平，形成完整交叉的八种条件。实现时须先规定字段的所有权：记忆组织操作事件与来源的关联，未决维护操作候选状态与重启条件，后果连接操作反馈到未来动作的触发。共同存储和读取接口保持一致，避免关闭一项因素时顺带删除另一项需要的材料。三项是可分别分配的工程因素，并非把三个号位或 S、D、E 各自变成一个软件模块。只有实际分配产生了预定差异，主效应与交互比较才有明确解释。

为避免结构化条件额外得到专家知识，所有语义内容应来自共同任务材料；字段转换由统一规则或同等预算内的系统处理，禁止人工只替实验组补写正确机制。各条件的指令长度、提醒数量和界面负担尽量平衡，无法匹配的部分单独计量。可以另设只增加同量提醒、不提供关联结构的补充条件，检验额外注意的解释。若普通基线自行建立了相同关联，仍保留最初分配结果，同时记录实现趋同；这种趋同会缩小实际操作差异，也是已有工作流能力的证据。

此外，可以加入一个资源扩展的强基线，让普通工作流获得与新系统实际消耗相当的额外检索或重试机会。这回答一个实际问题：同样增加投入，建立复杂结构是否比直接扩大搜索更值得。若结构化系统只在未控制成本时胜出，结论应表述为更高资源下的能力差异；若在等成本下仍有优势，才支持效率上的增量。

所有条件应固定关键版本，并保留模型、提示、检索库、工具和执行环境的配置。研究期间若必须升级某个部件，应说明受影响的任务，并避免把升级前后结果直接混合。实验对象是具体系统配置，不是不断变化的产品名称。清楚的版本边界使后来的复现能够知道究竟在检验什么。

35.3 三类任务与可判定的研究结果

第一类任务是证据组织。给系统一组事先核对过的原始论文、技术说明和相互转述材料，要求回答一个有限问题，并标明主张与证据的对应。材料可以包含定义差异、来源依赖和后续更正。主要评价是关键主张是否得到适当支持、是否准确限制范围，以及在新材料到来后能否修改受影响结论。这里不奖励引用数量，而评价引用是否完成了相应证明责任。

为了降低训练材料污染的影响，可以使用研究者新编制的任务组合和未公开的材料排列，但不能假定重新组合便消除了全部先验知识。应记录模型可能已经知道的领域事实，并设置需要依赖任务内新信息才能回答的部分。对引用核查，评阅者应能直接查看材料，依据预先制定的标准判断。评阅分歧要保留理由，不能只通过多数票隐藏概念问题。

第二类任务是可控机制探索。研究者拟构建人工环境，读数由可控输入、隐含零点和预定噪声共同产生；不同生成器分别使零点保持稳定、随时点漂移，或在某种历史操作后更新。系统可以选择输入、复测参考点，并执行明确列出的状态操作，但看不到生成器身份。任务应包含普通观察下相近、在允许操作后才分叉的机制，也包含在当前预算下仍难区分的机制。生成规则和噪声分布在正式实验前固定并留存，全部结果只能解释为对这些构造环境的表现。

一个任务包可依次安排基线观察、有限干预、新条件到来和恢复检验。前期学到的零点规则在后期可能继续有效，也可能因已声明的条件变更而失效，系统须决定保留、修改或搁置。主要终点拟采用预先留出的新条件中预测损失，辅以实验是否区分候选、错误规则是否被停用及恢复耗时。评分以生成关系和预测为依据，接受可检验意义上的等价表

述。任务阶段只是设计骨架，具体轮数、动作费用和噪声强度须经独立先导校准后固定，避免在正式结果出现后调到有利于新系统。

第三类任务是可执行研究辅助，例如维护一段分析代码、修复与资料处理相关的问题，并通过独立测试检查。SWE-bench 的原始工作展示了从真实代码仓库问题构建修复评估任务的路线，说明评价可以超出孤立代码生成而进入完整工作环境。[70] 本研究可以借鉴执行检验的思想，但须使用清楚界定的任务与独立测试，不引用动态排行榜来替代自身实验。

代码任务应包含新条件下的测试和既有功能保护，避免系统只针对可见测试修改。还应检查修复是否采用了与任务目的相冲突的捷径，例如删除测试、绕过数据或硬编码答案。执行通过是重要证据，却仍需核对是否真正满足问题要求。记录应保留失败尝试、运行成本与人工帮助，使成功不被整理成无代价的单步结果。

三类任务分别突出证据、机制和执行，能够检验不同能力的迁移边界。若某种结构只改善文献组织而对机制探索无效，应如实保留这种局部结论。跨任务一致性可以增强一般化依据，但不要求所有效应方向和大小完全相同。研究目的在于发现条件性的能力结构，而不是寻找一个对所有任务都占优的万能架构。

35.4 随机化 重复与历史污染

跨轮系统的实验单位不能简单设为一条提示。一个任务的记忆会影响后续任务，如果把同一智能体在不同条件间随意切换，就可能发生历史污染。较合适的单位是完整任务包及其独立系统实例。每个实例从明确检查点开始，在一个条件下完成规定序列；不同条件使用隔离的记忆和工具状态，避免一组的经验流入另一组。

任务包可以按领域、难度和变化类型分层。对每个基础任务包，建立八种条件的独立副本，并在区组内随机安排条件、运行顺序与种子；相近副本使用同一任务来源标识，以便分析其依赖。配对环境随机性可以减少无关差异，仍需多个独立种子观察生成波动。统计中的独立信息量以任务包和独立运行结构确定，不能把同一实例连续产生的答案当作互不相关样本。匹配提高比较精度，却不会把任务变体自动变成新的独立研究对象。

样本规模应依据独立先导任务中的方差、任务包内相关性、主要终点和最小有意义差异确定。分析拟先在同一任务包内比较条件，再汇总到事先界定的任务范围，并按任务包保留相关结构计算不确定性。用于调试表示、提示和噪声强度的先导任务不再进入确认性

样本。若预算只能支持较少任务包，就应报告探索性估计及其宽区间；增加每包的生成次数可以改善该包的随机误差估计，却不能替代扩大任务类型的覆盖。

系统失效、超时和工具中断需要预先规定处理方式。只分析完成任务的运行，会偏向更容易成功的条件。应保留按照最初分配进行比较的结果，并明确哪些失败属于系统表现，哪些属于共同基础设施故障。共同故障可以按统一规则重跑，但不能只为某一条件选择性重试。重试策略本身也是系统的一部分，必须进入记录。

人类参与也会引入学习与干扰。若同一研究者连续使用多个系统，可能把先前经验带入后续条件。可以采用任务分块、顺序平衡和适当隔离，但这些安排不能完全消除学习。研究应明确是在比较新手初次使用、熟练使用还是持续协作，并据此选择设计。把人工帮助视为固定背景，可能掩盖人机系统中最重要的适应过程。

记忆污染还可能来自共享资料库。系统 A 生成的摘要若进入共同检索库，系统 B 便不再是独立比较。正式实验应区分只读公共资料与各实例产生的记录，禁止未经记录的跨条件写入。需要共享新证据时，应同时向所有相关条件提供相同内容，并保留时间点。这样才能把后续变化归于组织方式，而非信息到达的差异。

随机化的作用是建立可信比较，并不替代机制说明。即使总体效果显著，也仍需检查操作是否按计划执行、哪些任务发生改善以及是否伴随新的失败类型。相反，统计上不显著也不能自动证明没有作用，需要结合区间精度判断是否已经排除有实际意义的差异。实验应在效应、精度与机制之间建立完整解释。

35.5 主要终点与创造性评价

每类任务应预先指定一个主要终点：证据任务为关键主张的支持正确性，合成任务为保留条件下的预测损失，代码任务为独立测试下的任务完成。三项组织因素在三类任务中的主效应形成九项主要比较，可预先把总体容许误报水平在九项间均分；交互作用、资源扩展和提醒匹配比较先作为探索性结果，待独立任务确认。这种安排牺牲部分检出力，换取结果选择的约束。若改用有序检验，应在实施前固定顺序，不能根据观察到的优势更换主要问题。

主要终点之外，应记录错误修复、负迁移、成本和恢复时间。一个系统可能最终正确，却经历大量不必要行动；另一个系统可能略慢，却能在错误出现后可靠恢复。不同实际用途会重视不同权衡，因此不宜把所有指标压成一个不可解释的总分。若需要综合决策，应公开权重并进行敏感性分析，检查结论是否依赖某一任意权重。

创造性评价尤其需要与新颖措辞分开。可以让盲评者检查系统提出的问题或实验是否增加了可区分的机制、是否利用了此前被忽略的证据，以及是否具有可执行性。一个候选与已有答案不同，可能只是错误；一个候选并不华丽，却可能通过关键对照显著提高辨别力。评价应围绕认识贡献，而非语言惊奇程度。

对于合成环境，还可以计算某次实验在候选机制之间实际产生了多大区分，但该指标依赖设计者规定的假说空间。若系统提出空间之外的合理解释，不能仅因不在答案表中而判错。应设置人工复核通道，检查它是否具有可检验内容。评价框架也可能遗漏创造，因此需要允许有理由地补充，但补充规则应透明，避免事后为喜欢的系统改标准。

对真实研究建议，可以先评估论证和可行性，再选择少量建议进行独立执行。前者是专家判断，后者提供外部结果，两者不可混称。专家认为值得做，并不等于建议已经有效；执行失败也未必说明建议毫无价值，因为探索本来存在不确定性。应检查建议是否依据当时信息合理、实验是否按计划实施，以及失败是否带来了可解释的新知识。

评分者应尽可能不知道成果来自哪种系统，且使用相同材料和标准。多个评分者的差异需要报告，尤其在开放问题和创造性判断中。高一致性可以说明标准较清楚，却不保证标准本身正确；低一致性则提示概念或任务可能需要重新界定。评价过程应成为研究材料的一部分，而不是被隐藏在一个最终平均分后面。

还应设置结论强度与证据一致性的评价，并把拒答的收益和代价分开。材料不足而正确保留未决，不应算作无依据主张；材料充分却拒绝完成任务，则应记录任务未完成。预先准备两类材料并保持共同评分规则，才能防止系统靠一律谨慎取得好成绩。超时和中断同样进入完成率及成本，预测损失的缺失处理应事先规定，不能只删除没有输出的运行。这样的评价同时要求有根据的收束和有根据的悬置。

35.6 消融与反事实过程检查

总体比较只能说明某种组合是否有效，不能充分解释为什么。消融研究可以在完整系统中暂时移除某个连接，例如保留事件记录但取消检索触发，保留假说列表但取消证据状态，保留反馈但取消行动修订。若性能变化与所主张的机制相符，解释会更有依据。消融也要保持其他资源尽量一致，否则移除部件可能同时改变预算与提示长度。

更直接的反事实检查，是从相同状态复制两个系统实例，仅改变一项历史材料。一个实例获得真实相关反馈，另一个获得与任务无关但长度相近的信息，比较下一轮行动。若只有相关反馈引发适当修订，支持系统对历史内容具有选择性敏感。若任何文本都产生相

似改变，则可能只是一般提示效应。复制和重置在这里是辨识工具，不是对历史性的否定。

还可以设置错误记忆恢复任务。在完全人工构造的环境中植入一条可由原始资料反驳的摘要，观察系统能否发现冲突、隔离影响并修复下游结论。需要明确这种操作只针对实验副本，不能污染真实科研记录。主要关心的是故障传播与恢复机制，而非系统是否从未犯错。长期研究系统应在非理想条件下接受检验。

对未决问题维护，应安排预先定义的触发事件：新的运行配置使旧候选可检验，可靠材料否定某条前提，或任务用途改变使已关闭问题需要重开。记录系统是否更新对应候选、多久选择下一项有效动作，以及是否错误激活无关路线。漏掉应重启的问题和重复投入已失效路线应分别报告，因为二者可能互相抵消在总任务得分中。材料与触发次序对全部条件相同，过程检查以所选检验、所用证据和状态变更为准，不以自述创造性作为机制指标。

机制分析还应考虑中介变量的局限。即使结构化记忆提高检索率，而检索率与结果改善相关，也不能仅凭相关关系宣布全部效果由检索中介。未测量的任务理解或额外检查可能同时变化。若需要因果中介结论，应设计额外干预或明确识别假设。对于初步研究，报告过程证据与竞争解释，往往比给出过强的因果分解更可靠。

反例任务应有意包含结构化机制可能不利的情形。例如任务极短且无需历史，维护档案可能只增加成本；环境快速变化时，旧经验可能造成负迁移；候选机制几乎不可区分时，保留过多分支可能耗尽预算。这些任务不是故意刁难新系统，而是用于划定适用域。一个可靠研究纲领应能预测何时自身方法不值得使用。

若消融结果与预期相反，应先检查操作是否真正改变了目标机制，再考虑理论修订。系统可能通过其他部件补偿被移除功能，也可能根本没有使用所谓关键结构。不能把任何反常结果都解释为更深层纠缠。机制主张必须允许被削弱，并通过新的、独立的设计重新检验，而不是在同一批结果上无限调整故事。

35.7 人机团队中的第二阶段验证

受控任务能够提供清晰比较，却不能充分代表真实科研。第二阶段可以让专业研究者在限定、低后果的研究任务中使用系统，比较普通工作流与已在受控实验中表现可靠的组织方式。任务可包括资料核查、分析方案比较或实验设计审阅，避免一开始就把未经验证

系统用于不可逆行动。该阶段仍是研究设计，需要适当的参与者同意、信息保护与机构程序。

研究者应接受与系统功能相称的培训，避免新系统因不熟悉而被低估，也避免只有开发者能够熟练使用。培训时间应计入成本，并说明评价针对何种熟练程度。可以采用分组比较或经过平衡的交叉设计，但必须考虑研究者从前一条件学到的策略会保留到后一条件。若这种迁移正是研究对象，应明确分析，而不是把它当作无关噪声。

人机团队的主要结果可以是经独立审查的成果质量与完成成本，辅以研究者对关键决定的解释能力、工具移除后的核查能力和新任务迁移。单纯满意度不能代表可靠性，单纯速度也不能代表知识增量。需要检查团队是否更容易发现错误、是否减少无依据主张，以及是否能够在系统给出错误建议时保持判断。协作表现与个人能力应分别报告。

责任安排必须与实验任务一致。参与者应知道哪些输出需要核查，哪些操作只能在批准后执行，以及出现问题时如何停止。实验不应以隐藏真实风险来测试信任。可以使用事先说明会包含错误的模拟材料，研究识别与修复，而不让参与者在不知情中承担实际损失。研究伦理在此不是外加限制，它决定实验结果能否在合理合作条件下解释。

还应观察协作过程中的工作转移。系统可能减少写作，却增加核查；可能帮助初学者，也可能使专家花时间纠正低质量建议。平均效率应结合不同经验水平分析，且不能把个别开发者的熟练表现推广给所有研究者。若收益主要出现在某类任务或某类使用者，应据此调整适用建议。分布差异本身可以为系统改进提供线索。

长期阶段还需检查依赖与恢复。系统暂时不可用时，团队能否继续关键工作；模型更新后，既有流程是否需要重新验证；错误记忆被发现后，团队是否知道哪些成果受影响。持续科研依赖的不只是高峰表现，也包括中断时的可维持性。若一种系统只有在所有服务正常、资料完整且人员不变时才有效，其实际使用范围应受到相应限制。

第二阶段的结果不能替代受控机制研究，受控结果也不能替代真实协作研究。二者互补：前者帮助理解现实条件下的价值与负担，后者帮助定位具体因果机制。只有把两类证据连接，才能判断某种发生学组织既有理论可解释性，又有实际研究用途。跨层次验证比一次令人印象深刻的演示更接近新典范所需的证据。

35.8 结果如何改变研究纲领

正式实验应在实施前保存主要问题、条件定义、随机化方式、终点和分析计划。探索性分析仍可进行，但应与预先指定的检验区分。这样既保留发现意外结构的空间，也防止事后把任何结果改写成早已预期。研究纲领需要公开的不只是成功标准，还包括哪些结果会使某项主张收缩、哪些结果会要求放弃一种具体实现。

如果结构化记忆在等资源条件下只改善重复任务，却损害迁移，就应把主张限定为局部复用，并改进条件识别。若未决问题维护增加成本而不提高辨别力，就应检查候选表示和复查触发，必要时采用更简洁的任务管理。若后果连接减少同类错误却造成过度谨慎，就需要重新平衡行动与检查。实验结果应指向可修改的结构，而不是只给系统排一个名次。

若三个因素都没有带来具有实际意义的增量，也必须接受这一结果。可能是基础模型已经隐式完成类似组织，可能是任务不需要这些机制，也可能是本书提出的具体方法没有效果。不同解释需要后续证据区分，但不能立即以任务不足为由否定不利结果。预先界定的任务范围内没有优势，就是该范围内的重要知识，应完整报告。

若出现稳定优势，下一步应优先在未参与开发的任务包和另一种基础模型上检验关键因素，区分依赖特定提示配置的收益与较可迁移的组织效果。新的验证应保留原先终点和失败处理，并单列为迁移检验。若只有一种实现有效，就说明当前证据支持该实现；若多个实现共享的连接都有效，才有理由提高对共同机制的信心。工程效果的累积由这样的比较展开，既保留理论解释空间，也给出继续投入的具体条件。

开放材料应包括与复现相关的任务、配置、评价标准、执行日志和分析代码，并在隐私、许可和保密限制下提供足够说明。不能公开的材料应解释限制及可行替代，例如合成版本或受控访问。失败运行和中止条件尤其应保留，因为它们决定实际成本和可靠性。只发布最佳轨迹，会让后来者重复同样的误判。

这组设计把“承接历史”“维护未决问题”和“组织后果”转化为可以逐项改变的研究条件。其价值并不预设新系统必然胜出，而在于让一种关于知识发生的主张拥有明确的检验对象、竞争解释和修订路径。属性纠缠论在此不只是解释已经发生的研究，也参与设计怎样让新的差异显露，并接受这些显露对自身的回写。

人工智能因此成为科研新典范的一处试验场，而不是它的全部依据。真正广泛的问题仍然是：研究问题如何形成，对象如何获得可比较性，证据怎样跨越团队和学科，制度如

何保存失败与修订。后续章节将从这组可控设计出发，转向科学共同体更广泛的发生条件，使局部工程经验与整体研究纲领之间保持可以逐步建立的联系。

第八编 科研新典范的建立与检验

第 36 章 科研问题如何发生

36.1 疑问怎样取得研究形状

一个人感到奇怪，并不意味着一个科研问题已经形成。奇怪可能来自知识不足、词语误解、偶然变化，也可能来自现有框架确实没有处理的关系。要让疑问成为能够共同推进的问题，需要逐渐明确：什么现象值得解释，哪些回答彼此不同，怎样的资料能够改变判断，以及为什么值得为它付出研究成本。问题的形成因此已经是一段科研工作，而不是正式研究开始之前无关紧要的灵感。

设想一个材料研究团队面对这样的情境：同一配方的涂层经过若干次清洗之后，不同样片呈现不同程度的脱落。这是贯穿本章的假设案例，不声称已有实验结果。最初的问题是“为什么同样的涂层耐久性不同”。但“同样”可能只指配方，也可能包括基底与施工；“耐久”可能指首次缺损、累计缺损或继续使用的能力。研究者需要先确定哪一种差别值得解释，原来的疑问才能获得可以推进的形状。

一种分化是询问涂层形成时已经存在什么差别，另一种是询问清洗路径怎样累积影响，还有一种是询问检测与分类规则是否把不同缺损混为一类。这三种提问要求不同工作：检查初始样片，比较使用过程，或重建评价对象。若团队直接投入大量清洗试验，却始终未说明关心哪一类差别，资料增加仍可能没有形成答案。问题形成的第一项成就，是让不同回答获得可以分开的路径。

SDE 框架将这一工作放在发生的中心。显露 S 提供最初可以指认的差别，差异序列 D 使研究者追问它怎样到来，特征纠缠 E 使研究者辨认哪些条件共同规定了可能的结果。但研究并非只能按这一顺序开始。一个新的材料条件、一种此前无法实施的测量，或一个理论推导中的不相容，也能成为问题入口。问题是三个维度逐步获得相互规定的结果。

三种属性侧重也在其中形成。研究者希望把某个现象变成稳定对象，这是客体侧重；希望区分产生与测量路径，这是互动侧重；希望决定何种差别有意义、何种证据足够，这是主体侧重。它们不是三位专家分别提供的一份材料，而可能同时存在于一个研究决定中。选择在第几次清洗后评价脱落，既规定被观察状态，也选择需要保留的过程，还表达研究对使用寿命的关切。

因而，提出好问题不只是把一句话写得更漂亮。它需要压缩与展开交替进行：展开模糊前提，使差别可辨；再压缩成有限任务，使研究可做。过早压缩会让关键条件消失，持续展开而不形成任务又会使研究停滞。问题的成熟，表现为能够在有限行动中承受明确风险，并通过结果继续修改自己。

36.2 问题也从成功中生长

科学问题并非只由失败和异常产生。一种方法运行得很好，反而可能使原来无法提出的问题出现。当测量精度提高，研究者可以询问更细的差异；当某种对象已能稳定制备，研究者可以比较新的条件；当多个领域拥有相容记录，过去分散的问题可以形成共同任务。成功积累的条件，构成问题发生的重要来源。

仍以涂层思想案例说明。如果团队已经能稳定制备样片，并区分不同阶段的缺损，那么新问题可以从“结果是否可重复”转为“早期哪个变化有助于判断后续使用范围”。原有方法的成功使这种问题成为可能。它提供了足够清楚的比较单位，让研究者在保持配方与施工条件时，进一步追问使用路径。问题因此从一个尚未稳定的结果，生长为关于稳定结果之间关系的研究。

另一类问题来自技术可能性。新装置允许记录以前被平均掉的过程，新计算方法允许比较过去无法同时估计的模型，新档案让不同时间的判断可以被重建。这些变化不保证研究一定有价值，却扩大了可问范围。一个新典范若只等待旧理论出现裂缝，就会忽略科学通过主动创造条件而开辟对象的能力。

问题还可能来自规范目的的改变。同样的平均性能已经足以满足某项生产任务，却不足以回答少数极端失败为何集中于特定条件。不是原来数字突然错误，而是需要承担的后果改变了。研究目标由此要求新的分层、指标和时间窗口。规范选择不能独自证明机制，却可以正当地决定哪些机制值得追问。主体侧重在问题发生中具有这种选择责任。

本书所说的发生学，因而不应被缩减为异常诊断学。它还研究新对象如何获得可见性，新工具如何打开差异，新目的如何改变值得保存的信息。病理学是其中一个入口：持续失配迫使研究重建。但在没有病理的地方，创造也可以发生。一个成熟框架需要同时解释修复和开拓，而不把所有科研都写成对旧错误的追讨。

这样的理解使“未决问题”不再只是等待清除的剩余。某些未决方向因为缺少条件暂时不能推进，却可能在新材料或新工具出现后取得形状。保存它们需要说明缺少什么、何

时值得重启，而不是简单堆积问题清单。特征纠缠的成长，会改变哪些问题可做；问题的积累，也会反过来推动条件建设。科研创造在这种相互作用中形成。

36.3 从异常到可比较的差异

一个结果偏离预期，首先只是模型与记录之间出现了距离。距离可能来自真实变化，也可能来自目标、单位、程序或分析版本不一致。科研问题的发生需要经过一次可比性工作：确认正在比较的东西究竟是什么。这个步骤并不只是技术清洁，它可能直接改变问题的意义。原先看似巨大的理论冲突，可能缩成校准任务；原先被视为无关的小差别，也可能成为新的机制入口。

计量学对被测量、可比性和兼容性的区分，为这种工作提供了严谨基础。[3] 本书在此增加的是研究发生的视角：可比性不是天然附在数据上，而需要通过记录、定义和关系建立。相同变量名可能指向不同对象状态，不同变量名也可能经过转换指向共同目标。问题组织应使这些联系可见，而不是先把所有数值放进同一图表再解释差异。

在涂层案例中，两份报告都使用“脱落率”，但一份指首次清洗后的缺损面积比例，另一份指十次清洗后的累计缺损。它们可能都测得可靠，却回答不同问题。团队如果关注早期损伤，应重新构造相同阶段的指标；如果关注使用寿命，则要解释损伤怎样累积。可比性工作由此使原先笼统的“数值冲突”分化为阶段对齐、过程解释和用途选择三项任务。

可比性检查也可能改变对象边界。若相同配方在不同清洗路径下表现不同，研究可以询问材料在既定路径中的响应，也可以询问涂层与维护安排共同构成的使用方案。后一个对象扩大了责任范围：它需要说明何种维护改变能够延长可用阶段，代价又如何变化。对象扩大是否有意义，取决于新增关系能否改变实际判断，而不是名称中是否包含更多因素。

异常还需要与重复单位联系起来。同一块样片经过多轮清洗形成的是一条使用历史，来自不同批次的多块样片才支持另一种重复比较。若问题关心制造稳定性，增加同一块样片的检测次数不能代替增加独立批次；若关心损伤累积，则必须保留同一样片的先前状态。问题选择由此反过来规定什么值得采样。研究单位清楚以后，资料数量才具有与任务相称的意义。

因此，从异常到问题并不是由惊讶程度决定的。一个小而可重复、目标清楚、替代解释可区分的差异，可能比一个巨大但定义混乱的差异更有研究价值。发生学鼓励把惊讶转

成可承担的差别：它在哪里出现，在什么条件下保持，怎样改变才可能消失，又有哪些已知机制足以解释。这个转换使创造获得工作形状。

36.4 不同资源怎样打开新的问题入口

问题形成所能利用的第一项资源，未必是一份异常数据。涂层团队可能先拥有一套可控制清洗次序的新装置，却尚不知道哪种差异值得研究；也可能先获得一批异常样片，缺少可靠的形成记录；还可能先试出一种看似有效的修复操作，却无法解释其适用范围。三种起点所欠缺的关系不同，直接套用同一研究流程，容易让现有资源掩盖真正缺口。

从条件出发时，团队的首要工作是把可实施的改变转成有意义的区分。新装置能独立调整哪些操作，哪些调整会同时改变温度或作用时间，都需要先明确。如果能够沿确定路径得到可重复的缺损图样，就形成沉淀型的推进；如果先借新条件确认一种稳定状态，再比较通往它的过程，则形成锚定型的推进。两者共同面对的问题是：技术能力打开的范围，怎样成为值得承担的研究任务。

从行动过程出发，问题可能先表现为原有操作无法继续。某种清洗程序在新基底上失去稳定性，团队被迫重建条件，再观察是否形成新的可用状态，这是倒逼型的工作。相反，一次试作得到可重复的局部修复，研究者再寻找维持该效果的条件，则是原型型的工作。前者由实施困难打开问题，后者由试作机会打开问题。它们都需要把行动中获得的局部结果转成可以共同追问的关系。

从显露出发，异常样片给研究者一个需要解释的差别。先追查其形成条件，再安排新的比较，属于进入型；先尝试改变既有路径，再据结果重组使用方案，则具有再创型特征。一个成熟对象也可以成为再创的起点，例如已有涂层被要求承担不同清洗制度。此时需要说明哪些原有属性足以承接新用途，哪些条件必须重新取得资格，问题并不以旧产品全面失败为前提。

这些入口的真正差别，在于下一项工作应当补足什么。如果团队已有清楚的异常，却缺少初始状态，优先增加后续检测可能收益很小；如果已经掌握形成条件，却只有偶然试作结果，则应优先建立可重复显露。六路径提供的是这种资源与缺口之间的对应。研究者选择入口之后，还须主动寻找能够纠正该入口盲点的材料，避免把“我们能做什么”直接写成“最值得研究什么”。

路径转换也可以成为问题形成的关键时刻。异常样片研究发现一个此前未记录的施工差别，便可能转入条件建设；新装置产生了稳定图样，便可能转入对象识别。前一阶段的

产物在后一阶段取得新作用。三方程要求这类转换能重新说明显露、过程与条件的相互规定，三原理则帮助定位推动转换的张力：现有结果与操作说明为何不相容，改变了哪一维，又怎样改变另外两维的可用内容。

因此，问题创造既包含开辟新方向，也包含选择暂不追问的方向。一次有效转换，应该使研究者更清楚下一步为何值得做、完成以后能排除什么歧义、失败时又回到哪个缺口。这样的选择使六条进入路径成为可以相互补充的工作资源。问题由此取得自己的生长次序，而不必被编排成观察、假说、验证的一条固定发现史。

36.5 问题重构与对象的适当边界

一个问题写成“某因素是否有效”，通常已经包含若干未经展开的选择：有效意味着什么，对谁有效，在多长时间内有效，与什么条件相比。研究者如果只争论答案，可能始终没有触及决定问题意义的边界。问题重构需要把这些选择显题化，再判断哪些应保持、哪些应改变。它是研究推进的工作，不是对原提问的文字装饰。

回到涂层案例，“哪一种配方更耐久”可能逐渐分化为两个问题：在固定清洗程序中比较材料表现，以及寻找使某种材料维持可用状态的维护方案。前者把维护作为控制条件，后者把维护作为待设计的过程。两者都可以推进，却需要不同完成标准。选择后者也意味着把可操作安排纳入研究对象，而不再要求一个材料参数独自回答全部使用问题。

重构之后仍需划定足够小的工作范围。可以先研究一类基底、一个使用阶段和几种实际可实施的维护改变，再决定哪些关系值得扩展。若把“耐久”同时定义为任何环境、任意寿命和全部功能，所有结果都会显得不足，研究便失去具体改进方向。一个边界清楚的小问题，能够积累稳定知识，也能显示下一步扩展需要增加什么条件。

对象边界还须与能够改变的单位协调。如果团队一次更换了配方、施工和维护，却把改善归给一种原料，局部归因便缺少所需比较；如果只改变一个清洗步骤，却希望据此判断整套寿命管理方案，也需要补足其他关系。适当问题使可改变单位、可观察单位和拟解释单位相互对应。三者可以不同，但差距应成为待完成任务，而不能在结论中被悄悄跨越。

本书所主张的三属性分析，可以在这里提供一个检查方向。客体侧重问什么状态或能力要被稳定识别，互动侧重问通过什么任务和反馈形成，主体侧重问谁的目标、判断和后果承接参与其中。每一侧重仍需 SDE 展开，不能只列三个名词。一个“主体因素”若没有可观察代理或可讨论的规范内容，就尚未成为模型中的有效部分。

对象重构还要尊重原有成果。首次缺损指标虽不能代表全部使用寿命，仍可能适合当前验收；某种模型未包含累积历史，也可能准确描述首次清洗中的响应。新问题应说明自己增加了什么责任，而不是把旧问题的有限性写成全面失败。这样，研究可以从已有知识上生长，而不必每次通过夸大旧范式缺陷来证明自身必要。

一个成熟问题最终应当能够被简洁复述，却不依赖模糊省略。它有明确对象、可操作差异、相关条件和待判断关系，也有尚未解决的边界。简洁是经过展开之后获得的成果。只有走过这段工作，问题才可以交给不同成员接续，而不在每次接手时悄悄改变意义。

36.6 一个新问题究竟增加了什么

问题新颖并不取决于名称是否前所未见。相同名称可以包含一个此前未被区分的关系，崭新名称也可能只是重述已有任务。判断问题是否产生增量，应先问它改变了哪一种可能答案，增加了何种可实施区分，以及使哪个尚未承担的后果进入研究。概念组织、设计与经验机制仍是三类不同贡献，但在问题发生阶段，它们首先表现为不同的工作承诺。

涂层案例中的第一项增量可以是概念性的。把“材料耐久”分为固定程序中的条件响应与可调整维护下的使用能力，改变了原提问的答案空间。某种材料在第一项比较中较弱，却可能在允许调整维护的任务中更合适。这个区分在取得新数据之前就有意义，因为它使团队知道两项判断可能分离，以及谁需要为这种分离提供材料。它的价值由是否消除了实际歧义来衡量，而非由分类数量决定。

第二项增量表现为新问题使一种原来没有安排的比较成为必要。若团队开始询问早期损伤能否提示后期失效，就需要把同一样片的历史接起来；若询问制造差异，则需要相应批次间的比较。两项设计所需资源不同。问题通过规定什么资料能够改变答案，逐渐取得可执行形状。此处尚未证明任何损伤机制，却已经改变了下一步应当收集什么、哪些现有资料值得重新使用。

经验机制层面的承诺更进一步。假如研究者提出某个初期变化会改变后续清洗的作用方式，就必须指出这条作用关系为何不同于初始质量差异或累计使用强度。一个可研究的问题会把这些可能回答放在共同任务中，使其中至少某些回答能够在新的条件下分开。第 14、15 章提供通道辨识与模型比较的原则，第 38 章将其组织为具体实验与修订方案；本章关心的是在方案出现之前，问题是否已经把真正需要区分的关系提出。

这三类贡献之间存在实际选择。有时团队能够完成概念澄清，却暂时没有装置区分机制；有时新装置打开一种比较，但尚不清楚哪个理论问题最值得回答；有时既有理论已经

给出明确机制，只需恢复丢失的条件就能解释当前差别。研究计划应据此确定自己的主要产物。把机制探索写成唯一合法目标，会贬低必要的概念与设计工作；把任何组织改善都称为机制发现，则使证据责任失去分界。

判断问题的增量，还需要与强版本的既有工作比较。成熟的实验设计、动态模型和工程原型已经能够处理复杂过程。新增问题应指明这些方法在当前使用中尚未连接哪一段关系，而不是假定它们从不处理反馈。比如，已有模型可能足以预测固定程序中的失效，却没有回答维护可以怎样改变；也可能已有维护模型，只是团队尚未取得必要输入。前者打开新的研究范围，后者主要是知识取得任务，两者需要不同投入。

问题形成可以据此采取一次有内容的筛选：列出目前能够改变的条件，列出不同解释真正分歧的后果，再看两者是否相遇。如果可改变条件对所有候选解释都给出同样结果，当前设计未必值得扩大；如果某个小范围比较能够解开多个后续问题，它便具有较高连接价值。这种筛选关心问题的产出潜力，不预先裁定哪个假说正确，也不把暂时不可做的问题永久排除。

资源负担同样参与这个判断。增加完整使用历史可能有价值，也可能使样片数量、检测成本和维护工作超出项目能力。团队可以先恢复最可能改变结论的关系，明确其余问题仍然开放。问题因而能够分阶段成熟：先明确目标，再取得关键区分，最后争取机制约束。每一阶段都有自己的可交接产物，下一阶段也能知道尚缺什么。

本书对问题创造的积极主张正在这里：一个好问题能够重新组织已有知识，使原来混在一起的回答分开，使原来分散的资源相遇，并把尚未承担的后果变成有限工作。它未必立即带来新机制，却应改变后续判断的可能性。三类增量的区分由此成为问题选择的理由，而不只是研究结束时登记成绩的表格。

36.7 未决问题怎样成为条件资源

科研中的未决问题并不都值得永久保存。某些问题已经被可靠回答，某些因目标改变而失去意义，某些只是表达混乱。值得保留的是那些具有明确知识缺口、可能影响未来判断、并能说明重新启动条件的问题。把它们组织起来，可以使研究不必每次从同一困惑开始，也能避免失败经验随着人员更替消失。

一种有用的未决记录应保存几类关系：当前问题指向什么，已有解释为何尚不足，哪些证据已经获得，什么资料会改变判断，以及下一次可行动的入口在哪里。这些内容不是

机械增加字段，而是让问题具有可承接性。若没有下一步条件，一个问题可能只作为宽泛愿望存在；若没有失败历史，后来的团队可能重复已经走过的无效路径。

未决问题的组织，首先要使暂缓理由与可能的新入口相连。有些问题缺少独立观测，有些缺少可改变条件，还有些需要先解决目标分歧。把它们都标成“尚待研究”，会抹去重启所需资源的不同。一个问题库只有能够帮助团队判断新工具或新材料与哪个缺口相遇，才形成研究能力。保存的分支由此参与任务选择，选择的后果也反过来修订问题状态。

在涂层思想案例中，早期研究可能无法区分初始局部缺陷与清洗累积损伤，因为每块样片只有最终状态记录。团队于是暂缓机制结论，并明确缺少清洗前后的对应观测。后来若具备适当的跟踪条件，旧问题便获得重启入口。新的条件直接补足原先标出的缺口，因而比重新收集一批没有前史的终点样片更有方向。E 的成长改变了 D 的可能性，并使新的 S 有机会显露。

未决问题的组织也需要优先级。一个问题可能理论上迷人，却暂时没有可识别路径；另一个问题较小，却能解开多个后续任务的共同障碍。研究资源应同时考虑重要性、可做性和连接价值。主体侧重在这里承担选择责任，但不能仅以个人兴趣决定公共投入。理由需要能够被讨论，机会成本需要被承认。

对于已经被强证据否定的分支，应保留其失败原因，而不保持与存活解释相同的地位。开放性不等于永远平均对待全部可能。新的条件确实可能要求重启旧分支，但重启需要说明前提发生了什么变化。这样，记忆既支持创造，也支持排除；未决问题的组织既保留可能性，也形成约束。

问题组织的成熟，因而表现为知道何时继续、暂停、重构和结束。四种决定都属于科研成果。一个只能不断增加问题而不能结束任务的系统，并不比一个过早封闭问题的系统更有创造力。持续发生需要节奏，也需要有理由的结算。

36.8 让问题成为可交接的工作

一个研究问题真正进入共同工作，通常需要从讨论中的吸引力转为可以交接的责任。接手者应知道需要完成什么，现有资料在哪里，哪些决定已经作出，哪些仍可改变，以及结果将影响哪一项判断。交接不是把问题冻结，而是让修改有明确起点。没有这种起点，团队可能每轮都重新讨论全部前提，却无法积累进展。

问题任务书可以很简洁，但应保存核心关系。它说明被研究的发生单位，主要可区分解释，拟改变的条件，拟观察的结果和最低完成条件。完成条件不必是支持某个假说，也可以是确认资料不足、证明某个设计不能辨认机制，或者给出一个更适当的对象边界。把这些结果视为合法产物，有助于研究者如实报告失败与限制。

任务还应连接实际资源。某项设计需要的装置、参与时间、专业知识和分析能力，必须与目标相称。一个范围过大的问题可以分成若干能够相互承接的任务，但分工不能破坏共同对象。材料团队、实验团队和理论团队需要知道各自的输出在下一环节怎样被使用，避免每个局部都成功而整体仍然无法回答问题。

研究开始之后，问题允许改变，但改变应留下理由。新的数据可能说明原目标不适当，新工具可能使更有价值的问题可做，新的风险也可能要求停止。版本化记录使这些改变成为可解释的发生，而不被误写成研究从一开始就如此计划。它也帮助区分探索中合理调整与看到结果后重写成功标准。

对问题发生的评价，最后应回到它创造了什么能力。研究者是否更清楚地区分对象，是否找到新的可检验关系，是否减少了重复无效工作，是否使不同成员能够接续，是否形成了值得进入下一轮的知识？这些后果比问题名称是否新颖更重要。新典范的成长发生在此类具体能力中，而非仅发生在宣言中。

由最初的涂层差别走到可交接任务，问题已经发生了三种变化：模糊评价被分化为明确目标，现有资源被组织成能够区分回答的工作，暂不能回答的部分取得了有理由的暂停位置。创造因而有了可积累的产物。接手者可以改变问题，却知道改变从何处开始。第 37 章进一步处理这一产物进入不同专业之后的新困难：各方如何确认自己承担的是同一个对象，以及哪些专业差异需要成为共同约束。

第 37 章 对象建构与跨学科协作

37.1 共同名称还不是共同对象

几门学科围绕同一个名称工作，并不意味着它们已经研究同一个对象。物理学家讨论教室的声音分布，心理研究者讨论注意与负担，教育研究者讨论理解和迁移，工程师讨论可实现的改造，使用者讨论是否愿意长时间停留。他们都可以说自己在研究“学习空间”，实际保留的状态、过程和后果却不同。跨学科协作的起点，是使这些差别成为共同工作的一部分。

本章以一项假设的学习空间研究作为贯穿案例，不声称已有对应实验结果。团队希望了解某种声音环境改造是否改善学习。若只让各学科分别提交一个最熟悉的指标，再把结果放到同一报告中，可能得到一份丰富的材料汇编，却没有形成能够相互约束的解释。每项指标为什么与共同目标有关，在哪些条件下可能分离，必须另行建立。

一个较清楚的共同对象可以是：在指定空间、任务和时间范围内，声音环境、活动组织与学习表现怎样形成可重复的关系。这个表达没有把声音、体验和学习合成一个量，也没有宣布物理环境独自决定学习。它规定了一段共同追踪的发生，使不同成员知道哪些事件需要连接，哪些结果可以互相校验，哪些判断必须保持分开。

对象建构在这里并不是任意制造事实。团队可以选择研究范围，却不能选择结果必定符合期待；可以约定事件标识，却不能靠标识消除真实差别；可以设计指标，却要说明指标为何承担目标解释。建构的正当性来自它使现象更可区分、关系更可检验、后果更可承担。对象一旦建立，就会通过不相容结果反过来限制团队。

三属性分析为共同对象提供一个方向。客体侧重要求识别可共同确认的空间、任务和状态，互动侧重追踪声音、行动与反馈怎样发生，主体侧重明确谁的目标、体验和判断进入研究。它们不是分别归属物理、工程和人文学科的领地。一个声学测量包含目标选择和实现路径，一项体验报告也包含可记录对象和操作条件。每门学科都可能展开全部侧重，只是任务重心不同。

跨学科协作因此不能以取消专业边界为目标。专业边界保存了证明方法、误差经验和失败教训。需要改变的是边界之间缺少承接：一方输出的量到了另一方手里被当作什么，中间省略了哪些条件，最后支持的判断是否超出了最初证据。共同对象的形成，是让这些承接能够被讨论和修改。

37.2 翻译需要保存不能翻译的部分

同一个词在不同学科里可能承担不同任务。“稳定”可以指测量重复性、动力学恢复、统计分布不变，也可以指制度持续；“反馈”可以指物理回路、人的回应、参数更新或公开结果引起的行为改变。若团队只建立词语对照表，却不说明每次使用的操作含义，共同语言反而可能掩盖分歧。翻译必须把词放回它参与的判断中。

在学习空间案例中，“噪声”对声学测量可能指某类物理信号，对学习者可能指不受欢迎的打扰，对统计分析又可能指模型未解释的变动。三种用法不能直接互换。一段物理上较强的声音未必在所有任务中都更令人困扰，模型残差也不必对应任何可听见的声音。团队应当保留不同名词或明确限定，避免一个词跨越三层后制造虚假的统一。

一种有效翻译可以从具体问题开始。物理成员解释某项量在什么条件下测得，教育成员说明该量为何可能影响任务，心理成员提出哪些体验与行为能够提供中间线索，工程成员指出实际能够改变哪些条件。每一次翻译都伴随一个可追问的桥梁。若桥梁暂时只是猜想，就把它保留为猜想，而不通过语言流畅使它看起来已经成立。

不可翻译的部分也有价值。某种规范判断可能无法被单一实验指标表示，某种主观报告可能没有直接物理对应，某种数学量可能暂时没有可实现测量。跨学科工作不应强迫它们立即变成同一种数据，而应说明它们在共同任务中承担什么角色。规范可以决定值得保护的后果，报告可以提供新的问题入口，数学量可以组织候选关系；三者的证据责任仍然不同。

本书的 SDE 语言同样需要接受这一要求。显露不是各学科全部因变量的别名，差异序列不是所有时间数据的总称，特征纠缠也不是把环境变量放在一起的清单。它们只有在使不同条件、过程和结果形成相互规定时，才提供额外组织力。若翻译只是把原变量换成三个字母，没有改变问题关系，就不应宣称完成了跨学科整合。

共同语言可以逐步成熟。最初只需保证关键术语不会造成实质误解，随着研究推进，再增加精确区分。过早追求完美词典可能拖住工作，完全依赖默契则会积累隐性冲突。一个可修订的小型术语档案，连同各词实际使用的案例和反例，比一份宣称涵盖全部领域的宏大词表更能支持协作。

37.3 指标之间的桥梁怎样建立

一项物理变化能否成为教育改善的证据，需要跨越多段桥梁。空间中的某个声学量改变，首先说明物理条件发生了变化；学习者是否暴露于这种变化，需要活动位置和时间资料；这种变化是否影响任务过程，需要相应观察；任务过程变化是否形成学习收益，还需要目标适当的结果。任何一段都不能仅凭“有关联”而自动越过。

这不意味着每项研究必须同时测量全部链条。研究可以选择一个有限环节，明确报告它支持到哪里。一个工程项目首先证明改造达到了预定物理效果，可以是一项完整成果；一个学习研究进一步检验迁移收益，则承担更强责任。跨学科合作的价值，是让这些成果能够接续，同时不让后一层结论提前借用前一层的确定性。

指标设计需要同时回答两个问题：它测得是否可靠，它代表目标是否适当。一个仪器读数可以很精确，却与学习目标关系有限；一个主观量表可以切近体验，却受到使用规则和情境影响。精确性与有效性不能互相替代。共同对象要求两类问题同时进入设计，而不是让掌握最精密设备的一方自然获得全部解释权。

计量学对被测量和不确定度的说明，可以帮助团队明确物理指标的边界。[3] 进入教育与体验层面时，还需要各领域关于任务、构念与解释的专门论证。本书不把计量术语当作所有对象的统一算法，而把它们作为认真定义目标的一个范例。跨学科共同性来自证明责任的协调，不来自所有指标使用相同形式。

可以构造一个可能结果：改造后某项声音指标改善，参与者也报告更舒适，但独立迁移任务没有变化。这个结果不必被评价为整个项目失败。它支持物理与体验层面的收益，却没有支持所检验的学习迁移主张。团队可以缩小成果范围，也可以重新检查学习任务、观察窗口和候选机制。将不同结果分别结算，能够避免用一个总分掩盖真实差异。

另一种可能结果是平均表现改善，却有一类活动受到不利影响。此时，共同对象需要包含任务异质性，而不能只保留平均收益。是否接受这种取舍，属于规范和实际决策问题；它需要明确受益、受损和替代方案。数据可以揭示后果，却不会独自决定应当选择哪一种空间安排。科学协作与公共判断在这里连接，也保持各自责任。

由此形成的指标网络，不是把每个变量都连一条线的图，而是一组有说明的承接关系。每一条联系都标明已有支持、待检验部分和可失败条件。它可以随着结果删减或重建，使对象逐步变得清楚。跨学科整合的成熟性，表现为知道哪些桥梁已经足够坚固，哪些仍不能承载更强结论。

37.4 时间尺度与机制的共同描述

不同学科的时间窗口常常不一致。声学装置可以记录快速变化，课堂活动以任务片段组织，学习评价可能关注更长时间后的迁移，使用者的适应又可能跨越多次经历。若把这些时间一律称为“过程”，看似已经重视动态，实际上仍可能缺少共同结构。协作需要说明不同时间尺度怎样连接，而不是只把更多时间点加入数据表。

共同事件标识可以成为起点。一次任务何时开始，某种声音何时出现，参与者何时改变活动，报告何时收集，后测何时实施，都应能够相互定位。时间对齐不保证因果关系，却是许多因果判断的必要准备。若报告发生在结果已经公布之后，它可能包含对结果的解释；若把它当作过程当时的体验，就会倒置证据次序。

机制分析进一步要求区分即时作用与跨轮回写。某次声音打断当前任务，和先前体验改变下一次任务选择，是不同关系。前者可能通过当时条件解释，后者需要保存历史与选择。一个模型若只使用当前环境，可能很好地解释即时表现，却遗漏后续安排；另一个模型若加入历史，也不能仅凭拟合提高就证明记忆机制存在。需要寻找能够区分替代解释的操作或独立资料。

在假设的学习空间研究中，团队可以先保持任务相同，比较不同条件下即时过程；再在适当设计中考察先前经历是否影响后续空间选择。两阶段的主要目标不同，所需对照也不同。若前一阶段已经改变参与者的期待，后一阶段就必须记录这种承接。跨学科设计的难处，往往不在于缺少变量，而在于研究阶段本身会成为下一阶段的条件。

把两阶段连起来，还需要保存不同专业各自认为关键的时间边界。声学成员可能按声音事件切片，教育成员可能按任务步骤切片，使用者则可能按一次完整经历评价。共同分析可以先保留这些边界，再检查哪些声音事件与哪些任务步骤重合。若改造使活动次序发生变化，同一物理记录便可能进入另一类学习过程。三原理所要求的回写在这里表现为具体变化：空间条件改变任务路径，任务结果又推动下一轮空间规则调整，新的规则重新规定哪些事件需要对齐。

同时，机制共同描述应允许局部独立。某个快速物理过程可以被稳定地压缩成对慢过程的输入，而不必每次把全部细节共同求解。是否允许压缩，取决于其对当前目标的影响。检验时应保持拟保留的输入摘要、控制条件与目标相同，再改变被省略的快速过程细节。若检验精度足以排除对用途重要的差异，慢过程关系仍在预设容差内保持，压缩就在已检验范围获得支持；若出现系统差别，则需要展开。这与第 13 章的实际可测性及第 15

章的任务损失要求相接。跨尺度协作因而包含分开与连接两种能力，不能把“全部同台”理解为永远拒绝任何分解。

机制的共同性最终要由约束表现。不同学科描述若能够对同一后果给出相容关系，并能说明各自信息怎样进入下一环节，协作就获得了实质内容。若只能在结论中使用相同口号，而模型、资料和时间窗口互不相干，就仍处于并列阶段。发生学提供的是让并列转向相互约束的工作方法。

37.5 分工怎样保存共同问题

跨学科团队需要分工，但分工也会使共同问题在交接中碎裂。一个成员负责仪器，另一个负责问卷，第三个负责分析，最后才发现各自保存的事件无法对应。这种失败不必归因于任何成员不认真。它可能来自任务之间缺少共同对象的约束：局部完成条件清楚，整体证据链却没有被共同设计。

因此，分工之前需要明确关键接口。仪器团队输出什么量及其不确定度，任务团队怎样定义活动阶段，报告团队在什么时点询问什么内容，分析团队怎样区分预定比较与探索发现。这些接口不只是文件格式，也是解释责任。一个数字从前一环节进入后一环节时，哪些条件必须一并携带，应当事先确定。

责任还包括识别谁有权改变共同问题。局部技术调整可能是必要的，却可能改变目标量；某项分析便利可能提高效率，却可能删掉机制所需的历史。团队需要一个能够讨论这种改变的程序，使专业自主与共同承诺相互协调。程序不应要求每个小决定都全体表决，而应区分日常实现选择和会改变主要结论意义的选择。

知识不对称是协作的正常状态。物理成员不必掌握全部教育理论，教育成员也不必重做每个仪器计算。但各方需要理解对方输出能支持什么、不能支持什么，并能提出与共同任务有关的质疑。真正的互信不是不问细节，而是知道哪些细节已有可靠负责者，哪些接缝需要共同审查。专业能力与可追溯责任一起形成这种信任。

AI 可以协助整理接口、追踪版本和发现术语冲突，却不能自动裁决领域争议。模型可能把相近词语合并，也可能给缺失桥梁生成流畅说明。团队需要把工具输出作为待核查的建议，保存其来源和修改过程。人机科研单位的成熟性，体现在谁能纠正错误、谁承担选择后果，以及纠正如何进入下一轮工作，而不是体现在报告看起来多么统一。

协作还需要让负结果能够流通。某一环节发现原定指标不可靠，可能延缓项目，却也避免后续投入建立在错误对象上。若团队只奖励按时交付支持性结果，成员就可能倾向于隐藏边界问题。知识发生需要一种能够公开困难的组织条件，使停止、缩域和重构成为合法决定。这是一项规范选择，其理由来自研究共同承担后果的要求。

由此形成的分工不是把三个号位交给三组人，而是让每个环节说明自己的对象、路径和判断条件，再由共同事件把它们接起来。专业边界被保留，信息却不必在边界处中断。共同问题因而能够跨人员和阶段持续存在，同时接受新的材料修订。

37.6 联合结论怎样超出局部结论之和

第 36 章讨论一个问题增加了什么区别，本节进一步问：当这些问题由不同专业共同承担时，合作究竟增加了什么约束？各方分别得到可靠结果，只是共同判断的必要材料。实质整合还要求一方的结果能够改变另一方允许的解释，或者要求几种局部结果共同限制某个整体主张。否则，报告可以很丰富，项目却仍由若干互不影响的研究并列组成。

学习空间案例可以显示这种差别。设想声学测量发现两种安排的某项总量接近，教育观察却发现任务表现不同。若双方只分别提交结果，就会形成两个彼此陌生的结论。联合研究则可以追问：相同总量是否由不同时间结构构成，声事件是否落在不同任务步骤，活动位置是否改变了实际暴露？这些问题把两门专业的记录放在同一事件关系中，使“物理条件相同”受到更具体限定，也使教育解释取得可检查的过程入口。

这里的新增约束并非声学量直接决定学习。若团队认为某类打断主要影响一个任务阶段，就需说明为什么其他阶段应受到较小影响，以及哪些条件下这种差别会减弱。教育理论据此限制应当比较的任务结构，声学与工程知识则限制可以实际分离的变化。共同对象在这些双向限制中形成：每一方都放弃一部分任意解释，换取一组能够共同检验的关系。

还有一种合作成果是暴露原设计中无法同时满足的要求。心理研究可能希望保持参与者期待相近，工程改造却可能非常醒目；教育研究可能要求完整活动，声学隔离却可能改变正常互动。把这些冲突写出，就能决定应当调整主张、分阶段取得证据，还是寻找新的实现方式。单一环节的最佳条件未必能拼成整体最佳设计。协作增加的知识，正在于使这种不相容早于正式结论出现。

因而，联合结论需要一种受范围约束的分工。物理组可以担保所测量及其误差范围，任务组可以担保活动比较的适当性，体验组可以说明报告所对应的时点与意义；整体团队另行承担从这些局部内容到最终主张的桥梁。最后一项责任不能被分散到各组之后自动消

失。它要求团队明确：如果某一局部前提改变，哪些整体推断随之改变，哪些仍然可以保留。

三类增量在此具有不同于问题形成阶段的内容。概念组织增量体现为专业边界之间的歧义被定位，设计增量体现为多方要求被协调成可实施比较，经验机制增量体现为联合限制在新材料中获得支持。仅仅让各组使用 SDE 词汇，不会产生这些后果。真正值得检查的是某项接口改变之前与之后，团队能够排除的解释是否有所不同，以及新增约束来自哪一段专业知识。

这种检查应以成熟的协作方式为比较对象。各专业已有处理动态、误差、情境和原型修订的能力，发生学的作用是使这些能力在交接时继续有效。例如，若声学成员已经保存完整时间结构，额外再建一份同义档案没有明显收益；若任务边界没有进入声学汇总，恢复这项对应才可能改变判断。改进从具体断裂处开始，记录负担也随实际需要确定。

合作也可以在不产生新机制时取得完成成果。确认某个声学改变不足以支持当前学习主张，能使工程效果与教育期待分别获得清楚位置；发现一项主张需要尚不具备的时间分辨率，则明确了下一轮条件建设。共同结论的成熟性不在于所有局部都支持同一句肯定判断，而在于各项结果相互约束之后，团队知道能够共同声称什么，以及哪些问题还须由不同专业继续承担。

37.7 分歧怎样成为对象重建的资源

不同学科的分歧有时来自信息不足，有时来自目标不同，有时则来自真正的理论冲突。它们需要不同处理。信息不足可以通过补充资料改善，目标不同需要协商，理论冲突需要可区分预测。若把一切分歧都叫沟通问题，就会用更流畅的语言掩盖实质差别；若把一切分歧都叫不可通约，又会过早放弃共同工作。

在学习空间案例中，工程师可能认为某个改造已达到物理目标，教育研究者却认为学习证据不足。两者未必矛盾。他们可能分别评价不同环节。共同报告应允许两项结论并存，并说明下一步是否值得继续。跨学科统一不要求所有成员同时说“有效”或“无效”，而要求每个判断指向清楚对象，并能够共同支持实际决定。

另一种分歧更深：一种解释认为即时干扰是主要机制，另一种认为长期期待与活动规则更重要。团队不应通过平均双方措辞获得折中答案，而应寻找能够改变两种解释相对支持度的条件。短时与历时设计、期待操控与物理变化的分离，都可能成为候选方向。具体能否实施，还要接受任务和资源约束。分歧在这里成为新问题的来源。

规范分歧则不能仅靠更多数据结束。一个安排改善平均成绩，却使一部分使用者长期不适，怎样权衡需要公开目标与责任。研究能够更准确描述后果、缩小不确定性、提出替代方案；最终选择仍需理由。跨学科协作应当让这种责任可见，而不是把价值决定藏在技术指标权重中。权重可以计算，采用何种权重却需要论证。

对象有时需要在争议中拆开。原先统一称为“学习空间质量”的概念，可能包含互相不能替代的目标。拆开之后，各项研究反而可以更精确，再在决策层重新连接。整合不总意味着不断合并，也可能意味着找到适当分界。一个新典范若只赞成整体性，就会错过分化对知识成长的贡献。

失败结果同样可以推动对象重建。如果物理条件改变后，主要学习结果始终没有变化，团队可以检查目标是否过于远端，也可以接受当前路径不足以影响该目标。不能为了保住项目而无限增加未测中介。承认一条连接在声明范围内没有得到支持，会帮助后续研究把资源转向更有辨别力的关系。停止是对象边界变清楚的一种方式。

分歧最终应进入共同档案，保留提出理由、关键证据和未决条件。这样，后来团队能够理解为什么某个方案被采用，为什么某个主张被撤回，什么变化可能要求重新讨论。知识共同体不是没有分歧的共同体，而是能够把分歧转化为可承接问题的共同体。

37.8 共同对象怎样迁移并保持身份

一个对象能被当前团队共同研究，还需要面对另一个考验：换到新的空间、任务或使用者之后，它凭什么仍是同一个研究对象？学习空间的名称很容易携带，原来的关系却未必随名称一起移动。对象迁移要求团队重新确认哪些结构值得保留，哪些条件已改变，以及新的专业组合是否仍能承担原来的推断。它是共同对象的再次建构，而不是项目材料的简单复制。

可以设想将原研究转到另一间教室。新空间的某项平均声学指标与原空间相近，但活动安排由个人阅读改成小组讨论。物理层面的相近只保留了一部分条件，任务中什么声音会构成打断、什么互动需要被听见，都可能改变。团队因此需要重新说明共同目标。若仍把阅读任务中的关系直接作为讨论任务的改善标准，跨场景迁移就丢掉了原对象的关键构成。

另一种迁移保留任务，却改变使用者。相同活动中，不同人可能处在不同位置、拥有不同先前经验，也可能承担不同角色。研究可以先检查哪些差异能够通过原模型中已有条件处理，哪些差异要求新增关系。这里既要避免把所有人都当作完全相同，也要避免一遇

到差别就宣布无法比较。共同对象的可携带性，来自一组说明何时可以转换、何时必须分层的关系。

这项工作需要多方共同决定对象身份的最低保留条件。对于一个具体的学习收益主张，可能需要保持目标任务及其评价方式；对于一项工程性能主张，可能只需保持声明的物理范围。两种迁移可以拥有不同边界。团队应把它们分开交付，使接收者知道自己承接的是哪一种知识。这样，某个远端结论暂时不能迁移，并不抹去局部成果已经取得的有效性。

当迁移失败时，专业分工再次提供定位能力。物理条件未达到原范围，首先要求检查实现；物理条件达到而即时活动改变，要求检查活动与暴露的对应；即时表现相近而后续迁移任务不同，则需要重建学习主张所依赖的桥梁。这个分解不是预先决定真实原因，而是使失败有一组可以分别承担的入口。各专业的知识越明确，整体研究越能在失败时保持可修订性。

共同对象也可能因新任务而合理分裂。原先用来协调团队的“空间质量”，在更广用途下可能需要分别表示舒适、交流与独立学习等不同目标。分裂之后，各项对象仍可通过公共决策联系起来，但科学判断先保留各自证据。对象建构的成就并非永远维持一个总名称，而是形成能够随着任务扩大而清楚分化的关系，使新用途不必靠模糊旧结论获得支持。

为了让这种承接发生，交付成果应保留少量关键联系：原主张针对何种共同对象，各局部判断如何进入总体推断，以及改变哪些条件就需要重新讨论桥梁。发表与长期资料继承的安排由第 39 章进一步处理；本章强调的是交接内容中的责任结构。专业成员确认自己能够担保的观测与推断，整体团队则确认跨环节联系和最终主张，使后来质疑能够找到适当回应位置。

本章由学习空间的共同名称出发，获得了三个实质结果：不同专业通过事件与时间的对应建立可比较材料，通过联合限制形成超出局部结论之和的判断，通过迁移检查划定对象身份能够承接的范围。三属性在这里构成共同工作的不同责任，而不是三个学科领地。第 38 章将在这个基础上展开竞争机制与证据修订：当共同对象和推断桥梁已经明确，哪些结果能够真正改变团队对这些关系的判断。

第 38 章 实验设计 证据与模型修订

38.1 实验从一个可以分歧的问题开始

设想一个研究团队面对这样的资料：同一种配方在不同次评价中得到系统不同的评分，差异似乎随呈现次序而改变。团队可以把它归入测量误差，也可以立即用属性纠缠解释。但这两种过早决定都绕开了真正的实验问题：哪些不同机制能够产生当前形状，什么新的观察会使它们给出不同答案。实验设计的起点不是选择一个最能证明自己立场的程序，而是找到能够让合理解释发生分歧的条件。

以下讨论使用一个明确虚构的研究场景来展开设计推理，不报告任何已执行结果。假定现有记录只包含配方、总体评分和大致时间。研究者提出四种候选解释：仪器或材料条件缓慢改变，受试者状态延续，重复任务产生一般适应，前一轮评分作为信息改变下一轮判断。四种解释都可能产生时间结构，却需要不同的证据。把它们并列不是为了追求清单完整，而是为了确定当前资料究竟缺少哪些区分。

如果材料温度随次序改变，增加期待问卷可能无法解决问题；如果评分展示构成锚点，仅提高浓度测量精度也不能识别通道。研究者首先应回到事件，检查目标、样品、时点、操作和记录版本。这个工作并不低于理论分析，因为错误的资料身份会使后续最精密的统计也回答错误问题。一个看似重复的事件，可能实际上改变了被估量，甚至改变了“同一配方”的材料状态。

明确分歧以后，实验才形成自己的对象。研究者可以问，保持材料和主要采样程序一致时，改变历史评分是否可见，会不会改变下一轮响应；也可以问，改变呈现路径是否在控制历史信息后仍有作用。这些问题比“是否存在三号位纠缠”更具体，却不是削弱理论。它们把宏观框架落实为能够承担失败的命题，使哲学的问题组织转化为经验研究。

实验对象也可以在设计中被创造出来。某种此前没有统一记录的路径差异，一旦被明确操作和保存，便成为可比较对象；某种隐含在日常判断中的反馈，被分离为可展示、可遮蔽的环节，就获得新的辨识机会。这种创造不意味着结果可以任意制造，因为操作一旦公开，材料和参与者的反应就会约束研究者。发生学所重视的创造，是提出新的可受限制的关系。

实验因此是一种有目的的条件组织。它既改变世界以便取得知识，也必须研究这种改变如何影响所要知道的东西。第 11 至 15 章给出的事件、封装、辨识、回写与迁移，在这

里组成连续工作，而不是五套可以分别完成的表格。每一步都应改变下一步能够提出的问题，最终让模型修订有具体根据。

38.2 先把测量能力变成研究的一部分

在正式比较机制以前，需要知道装置和程序能够分辨多大的变化。若测量噪声已经覆盖理论所预期的差异，再复杂的设计也难以提供清晰裁决；若量表存在上限，真实的增加可能表现为方差收缩而非均值上移。测量校准不是研究前可以略过的杂务，而是规定实验能够回答什么的过程。它也决定不利结果究竟意味着机制不存在，还是测量暂时没有能力发现。

虚构的配方研究可以先安排材料检查和有限试测。材料检查确认批次、温度、呈现量与时间记录是否可靠；试测检验参与者是否理解报告任务、连续报告是否造成过重负担、评分是否集中在边界。这里取得的资料可以帮助调整程序，却不应同时充当最终假说的独立验证。因为程序已经根据这些资料改变，后续结论必须承认这种依赖。

试测之后应先划定本次实验能够接纳的响应范围。若高浓度条件的评分多数落在量表顶端，就需要降低刺激范围或另设能分辨的报告程序，并重新检查任务含义；若参与者只稳定区分顺序，就不宜把单位分差当作等距变化。这里调用第 11 章的比较资格，并把它变成入场条件：哪些事件可以合并，哪些只比较排序，哪些暂时不能进入主要终点。资格尚未确定的资料可以保存，却不能靠扩大样本自动变成更强证据。

在本例中，可以把共同末端刺激之后、同一时间窗口取得的评分规定为主要终点，用于比较路径与展示安排；完整响应曲线则帮助定位改变发生在起始、峰值还是恢复阶段。这样，终点选择直接对应竞争预测。若试测提示差异只在恢复阶段出现，可以据此另立恢复时间假说，并用新资料检验，不能把原来的主要终点悄悄替换掉。对尚无合适摘要的对象，先公开探索所得的响应结构及计算程序，也能成为后续实验的明确起点。

样本规模也不能从理论雄心直接推出。需要考虑最小重要效应、测量精度、个体内相关、失访和模型复杂度。重复观测可以增加轨迹信息，却不能简单当作增加了同样数量的独立参与者。一个适当的设计可能需要较多对象与适量重复，也可能需要少量对象的密集过程；选择取决于要辨识个体差异还是动态通道。研究应说明这段推理，而不是给出一个缺少来源的样本数。

这一步还包含停止的可能。如果校准发现目标效应远小于可取得精度，研究者可以改进测量、缩小问题或暂缓实施。继续收集大量低信息资料，未必比承认设计尚不成熟更负

责任。反过来，如果试测显示简单模型已足够服务当前用途，也可以保留稳定结论，把资源转向更有区分价值的条件。发生学不要求每次实验都必须发现复杂性。

38.3 让竞争机制承担不同预测

实验的辨识能力来自竞争机制在某些条件下不再等价。仪器漂移通常随设备状态或时间改变，一般适应可能随刺激历史改变，信息锚定则应对展示内容具有特定响应。研究者可以据此设计相互交叉的操作，但不能以操作名称保证机制纯净。改变刺激间隔也可能改变注意，改变展示内容也可能改变信任，必须分析操作的多重后果。

可以把本例推进为一项有限的交叉设计：两组先前刺激采用甲—乙与乙—甲两种顺序，随后都接受相同的末端刺激丙；另设回显本人真实既往评分与不回显两种安排，界面停留时间和操作负担尽量一致。研究者保存每次实际呈现与实际回显，而不只保存分组标签。指定的适应模型预期，在两种回显安排下均保留路径差异；指定的信息利用模型则预期，回显会按其内容改变末端判断。比较应预先列明各路径内的回显效应，以及模型要求的路径或内容调节，不能只计算跨路径平均。这里每项预期都属于已写明的候选模型，而不是“适应”或“反馈”名称的必然含义。

这一设计仍有需要主动处理的交叉解释。本人没有看到旧评分，也可能记得先前判断，因此回显无效只限制外部信息展示的增量，不能抹去内部历史。回显若增加注视时间，则需用不含评分的占位界面比较一般注意；若材料状态随先前次序变化，则需对末端样品做独立检查。后续可以使用明确标为参考的序列来改变信息内容，但其结果须另列为参考信息作用，不能与本人真实历史混算。每个追加对照都围绕一个尚未关闭的分歧取得价值。

个体内交叉比较能够减少某些稳定个体差异，却可能受到携带效应影响。前一安排已经改变了期待或学习，后一安排便不是在同一初始状态下开始。延长间隔或平衡顺序可以缓解部分问题，但是否足够需要依据；无法消除的历史效应应进入模型，或者改用组间设计。随机化顺序不是时间被取消，而是让某些顺序影响不再系统偏向某个比较。

进一步实验应优先选择两模型分歧最大而测量仍可靠的位置。假如二者都能解释即刻评分，却对回显撤除后的恢复速度作出不同预测，下一轮就增加撤除后的固定观测窗口，并在分析前指定恢复标准。撤除屏幕并不撤除记忆，模型若预期立即回到基线，必须承担这一预测可能失败的风险。若二者在全部可行条件下仍给出相同响应，就应先回到形式分析寻找新约束，而不是以更多同类重复伪装机制已经分开。

竞争预测不必都表现为平均值相反。一种机制可能预测即时改变后快速恢复，另一种预测延迟改变并持续；一种只影响分布中心，另一种主要改变尺度；一种依赖实际表现，另一种依赖展示值。过程和分布提供比单一显著性更丰富的分歧，但也增加比较数量。设计应事前说明哪些分歧承担裁决，避免事后从多条曲线中挑选最像理论的一条。

38.4 模拟怎样为真实实验服务

模拟的作用之一，是在已知生成机制下检查分析程序能否作出应有区分。研究者可以生成独立噪声、稳定相关过程、群体混合、外生趋势和明确反馈等资料，观察拟议指标会不会把它们全部判为同一种纠缠。如果程序在这些已知情形中频繁误判，直接用于真实资料就缺少依据。模拟不能证明真实世界遵循某个模型，却能暴露方法本身的识别缺陷。

针对上述交叉设计，模拟库可以故意构造近邻机制：让一个外生漂移过程与一个状态更新过程具有接近的均值和自相关，却对刺激顺序给出不同响应；再让一般界面效应与评分信息作用在原安排下相似，而在占位对照下分开。这样的设置比常数与强反馈的对比更能检查设计是否触及所称通道。逐步加入弱效应、材料波动、选择性退出和量表饱和，可以标出在哪些参数范围内只能报告未决。模拟参数是构造条件，不能写成真实机制的估计。

校准还应考虑选择过程。研究者若同时检查许多指标，即使每个指标单独误报率较低，整体上也可能经常出现看似有意义的结构。模拟应复现完整分析流程，包括筛选、调参和终点选择，而不是只检查最后一项统计量。这样才能估计实际研究程序的行为。把复杂流程拆成若干各自“合格”的检验，并不能保证组合以后仍然可靠。

置换检验同样需要尊重资料结构。任意打乱时间点会摧毁真实依赖，也可能制造一个过于容易战胜的参照。若问题是分组标签是否提供额外信息，可以在允许的单位和条件内重新分配标签；若资料以个体或会话组织，就应保留相应结构。置换所代表的无效应世界必须与研究问题一致，否则一个显著差异只说明资料不是完全打乱的样子。

模拟结果本身也需要与经验结果分开。某个方法在构造数据中能够识别反馈，说明在那些生成条件下它有能力，不说明真实资料中的反馈已经存在。若模拟中参数取值借自试测，研究者应说明借用范围和不确定性；若参数纯粹为了展示逻辑，应明确其示意性质。用漂亮的模拟曲线代替未完成实验，会把方法准备误写成世界已经给出支持。

模拟还可以指导资源分配。若某个设计对增加观测频率极敏感，却对增加对象数不敏感，说明当前主要限制可能是动态分辨；若相反，则异质性估计更需要对象覆盖。设计可

以据此选择较经济的取得方式，再在真实小规模实施中检查假设。模拟和试测相互修订，但最终验证必须保留新资料，使整个过程不会在同一套假设与数据中自我闭合。

38.5 执行中的变化如何进入证据

正式实验很少完全按照最初设想运行。设备可能故障，参与者可能误解任务，某个条件可能引起过重负担，研究者也可能发现原分析假设不适用。科学责任不在于假装这些变化从未发生，而在于让变化的时点、理由和影响可追踪。执行中的修订可以提高研究质量，但不能悄悄改变原来承诺检验的命题。

预先登记问题和分析安排，有助于区分根据已有观察生成假说与利用新观察检验预测。Nosek 等人讨论预注册时强调的正是这种区别。[71] 它并不禁止探索，也不保证登记的设计正确。一个糟糕的预先决定仍然可能糟糕，合理偏离也可能必要；关键是让后来读者能够知道哪些判断先于结果，哪些受到结果影响。

在虚构的配方研究中，如果试验进行到一半发现某个温度传感器失效，应按事先规则或公开修订处理相应事件。不能只删除那些不支持理论的故障时段，也不能为保持完整而把明显错误资料当成正常结果。故障是否影响主要比较，需要回到设备状态、分配时间与结果记录。若影响无法区分，结论可能需要缩限，甚至重新取得资料。

分析盲法可以在某些环节降低结果对选择的影响。例如在不知道组别含义时完成数据清理和程序检查，之后再揭示标签。但盲法也有适用条件：如果清理需要理解具体操作，完全遮蔽可能妨碍发现真实问题。研究者应设计与任务相称的分工，而不是把盲法当作形式要求。透明地说明谁在何时知道什么，比笼统声称“全程盲法”更有价值。

停止规则尤其需要清楚。不断查看结果，一旦出现有利模式便结束，可能改变最终证据的性质；因安全、资源或实施困难停止，则有不同理由。研究者可以预先规定固定样本、适当序贯规则或可接受精度目标，但必须据所用方法解释。停止并不等于失败，继续也不自动代表更严格。重要的是停止决定不能成为未报告的选择机制。

执行日志还应记录未完成者与缺失。某种安排使参与者更容易退出，本身可能是重要后果；只研究完成者会把这种后果从结论中删除。缺失处理可以采用明确假设下的分析，并检查结论对假设的敏感性。不能让数学补全掩盖资料实际上没有取得，也不能把缺失一律当作无法研究的借口。有限资料可以支持有限判断，条件是限制进入结论。

38.6 结果如何使模型真正改变

实验完成以后，最容易出现的失败不是没有显著结果，而是任何结果都没有改变模型。正向结果被称为支持，反向结果被称为更深的纠缠，零结果被称为号位受到压缩。这样的解释方式可以维持语言一致，却不能积累知识。修订应当体现为具体声称、参数范围、适用条件或操作建议发生变化，并能够影响下一次研究。

假如虚构实验发现路径变化稳定影响评分，而历史信息回显的总体平均效应接近零，尚不能据此限制所有指定信息通道。例如，在一个构造情形中，回显在甲—乙路径下使评分增加一，在乙—甲路径下使评分减少一，等权平均便为零，条件效应仍然明确。研究者应检验模型事前承诺的路径内与内容相关对比。若相应不确定范围已排除实践上重要的效应，才可缩限对应通道；若范围仍宽，则是精度不足。路径相关机制、平均回显效应和条件回显机制因此分别取得结论，不以一个总体零值互相代替。

若信息展示有作用，但作用方向与事前预测相反，原预测应记录为失败。研究者可以据此提出新的机制，例如展示引发反向校正而非锚定，但新解释需要新的检验。允许创造新假说正是科学的生命力，问题只在于不能把事后解释写成原来已经预见。失败和新问题可以同时成立，不必为了保存理论声誉而抹去前者。

若新增模型提高训练拟合，却在独立序列中失效，修订重点可能是模型复杂度、信息泄漏或场景依赖。若它在同域复现成功而跨任务失败，则需要重新界定迁移范围。研究者应保留不同层次的成果：测量程序可能有效，描述结构可能稳健，机制模型却仍不充分。把全部成果捆成一次总胜负，会让局部知识随着大主张一起被误判。

模型修订还可以增加对稳定性的理解。一个原被认为必然失稳的系统，在多种扰动后都恢复，可能提示负反馈或饱和约束。研究者不应把这种结果仅称为没有发现异常，而应研究恢复关系本身。发生学若以稳定性的生成作为中心，稳定反例就不是需要隐藏的障碍，而是理论必须吸收并能够产生新解释的材料。

修订记录应能重建一次具体选择。若路径效应保留，而事前规定的各条回显对比及其调节关系都已被足够精确地限制在实践上可忽略的范围，下一版模型才有理由删除本任务中的强回显项，并标明各项限制的对照来源。如果只是平均效应接近零，条件效应仍在，模型应保留相应路径或内容依赖。若预设对比的不确定范围仍宽，则保留未决状态，下一轮优先改善这一比较的精度；若强效应确已受到排除，可以把新增资源转向适应与材料变化的区分。不同版本由这些不同证据状态触发，不能用同一“无效”标签决定删除。

修订的程度还取决于失败针对哪种命题。一个逻辑推导若被有效反例击中，应撤回相应充分性声称；一个有限样本预测失败，则需要结合不确定性和事前门槛判断；一个制度方案未达目标，还须考虑实施是否真正发生。不能把实施不足自动用于保护方案，也不能把方案从未执行的部分计为经验失败。研究应同时保存干预实际到达哪里和结果发生什么，使机制与执行能够分别审查。

当多个结果方向不一致时，也不宜用多数票决定理论。三个描述指标支持、一个关键干预否定，证据权重未必是三比一，因为指标可能来自同一资料和同一误差源，而干预恰好针对核心通道。证据不是独立勾选项的数量，需要分析它们之间的依赖与证明职责。相同事件被不同图表重复展示，不能反复增加支持强度；真正新增的约束才改变可接受解释的范围。

38.7 三类研究怎样形成相互约束

味觉、历史经济资料与 AI 教育可以承担不同证据任务。味觉研究较适合细致控制刺激与路径，历史经济研究能够检验版本、制度和实时迁移，AI 教育则把独立表现、协作与后续迁移连接起来。三者不必共享同一终点，也不应被要求同时支持同一种结论。它们的联合价值在于检验共同方法是否能够进入不同对象，并在各自限制下产生可继承知识。

E1 味觉研究的优先产物可以是一个得到实际检验的辨别程序：用受限刺激路径、真实评分回显与材料检查，确认何种操作足以拆开若干解释。即使程序成功，也只批准它在这些材料、时段和报告任务中的使用。它可以把低干扰记录与对照设计交给其他研究者，但不能把末端评分的某条通道直接移植为制度运行的原因。若过程记录明显改变了味觉任务，则需要修订程序，并把这种记录代价纳入后续领域的设计预算。

E2 边际消费倾向研究可以接着检验这套程序面对历史限制时如何工作。历史政策、资料修订与家庭处境无法全部重做，团队应先把当时可见资料与后来版本分开，再确定哪些对照事实上存在。若找不到能分开两种机制的历史安排，就把成果限定为对象重建或实时预测，明确留下的识别缺口。已有异质性模型若在同目标下已经足够，事件档案的新增价值可转向解释跨期比较资格与更新成本；它无需制造一个模型失败来证明历史记录值得保存。

E3 AI 教育研究则可在前瞻安排中弥补部分历史缺口：预先保存工具版本、反馈时间和可见内容，将独立表现、协作表现与撤除工具后的延迟迁移分别计量。它继承 E2 的版本纪律，却必须重新处理训练本身会改变能力这一条件。若过程诊断只改善协作任务，后

续教学可以保留协作收益，同时补做独立迁移；若增加解释记录干扰了学习，则应比较更轻的采样方案。研究批准的是某种能力任务和教学用途，而非抽象地宣布 AI 使人更强或更弱。

三类研究共享的不是具体变量，而是证据顺序：界定目标，重建事件，形成竞争预测，取得区分资料，再把结论回写到模型和用途。它们也共享负结果的责任。某一领域成功，不能抵消另一领域失败；某个领域无法进入，可能揭示程序的边界，而不是证明该领域不够成熟。跨域研究应保存差异，让统一性接受真实限制。

这种组合还可以帮助发现方法层面的改进。味觉研究中形成的低干扰记录经验，可能启发教育研究；经济资料的双时间轴，可能帮助 AI 系统保存模型与信息版本；教育中的协作单位区分，也可能提醒其他领域区分个人与复合系统。这样的迁移需要具体解释借用了什么、改变了什么。跨学科协作的成果不在于共用术语，而在于方法被重新组织后解决了新的困难。

38.8 实验的完成是下一轮判断的开始

一个实验的完成不能只以数据收集结束或论文写成为标志。结果需要进入可复核资料、模型版本和用途决定，才能改变后续知识。若实验削弱了某个机制，相关说明应更新；若发现了测量缺陷，工具和历史资料解释也可能需要调整；若结果只支持有限条件，推广范围应相应收缩。实验后果没有进入这些环节，研究便可能停留在文本上的成功。

回到本章的虚构实验，一次合格交付应说明本轮究竟关闭了哪个分歧。例如，材料检查把指定温度漂移限制在不足以解释效应的范围，路径操作产生可重复差异，回显比较仍因区间过宽而未决。下一位研究者据此可以沿用已合格的末端刺激与记录程序，把新增观测集中到回显比较；他也可以质疑温度限制，但须指出哪项校准需要重做。完成状态由这些可继承的约束组成，既不要求所有机制一次解决，也不让“还需研究”遮蔽已经取得的内容。

研究档案应把每次结论与其使用的数据版本连接，记录哪些资料后来被撤回、哪些结果因程序修正而改变。这样，下一轮研究继承的是可检查的证据关系，而不是一个脱离来源的结论名称。

实验还需要留下别人能够继续的问题。一个未解决的分歧，如果已经明确资料缺口和可区分条件，就不是空白失败；它可以减少后来研究者重复摸索的成本。相反，一篇声称

完全解释却不公开过程的报告，可能给共同体留下很少可继承内容。知识的完成程度，应包括它使下一位研究者能够做什么，而不只是当前结论看起来多么完整。

这种继承要求把证据与修辞分开。理论语言可以表达新的视野，经验结论则需要保持与资料相称的时态和范围。形式例子标明为形式例子，模拟说明生成条件，探索结果保留探索身份，独立验证记录实际设计。不同材料并非价值高低的单一等级，而是承担不同证明责任。相互连接可以形成有力论证，相互冒充则会使整条链失去可信性。

研究完成还包括向参与者和实际使用者说明能够说明的结果。若某项记录没有改善任务，继续要求人们承担同样负担便需要新的理由；若某项安排造成明显不适或不公平，应在后续设计中修复。这里的规范责任不能从统计显著性中自动推出，却也不能与证据分离。参与者并非只有提供数据的角色，他们承受的后果本身就是研究安排的一部分。让结果改变研究者自己的操作，是反身性最具体的形式。

本章把第 11 至 15 章的方法推进为一次可以据结果分岔的研究：比较资格决定资料入场，交叉操作使候选模型分开，近邻模拟校准判别风险，执行记录保留偏离，模型版本把结果落实为下一步资源选择。三类优先研究分别承受可控实验、历史限制和学习迁移的检验。接下来，实验室内部已形成的约束怎样进入共同体，仍取决于发表者保留哪些内容、复现者继承哪个版本、错误被谁修复。第 39 章处理的正是这条继承链，不能由一次设计严谨的实验自行保证。

第 39 章 发表 复现与知识基础设施

39.1 发表怎样改变共同体能够看见的世界

一项研究取得资料以后，并不是所有观察都会以同样方式进入公共知识。论文需要组织主张，图表需要选择重点，篇幅需要压缩过程，审稿又会要求作者说明何者值得读者关注。这些选择是传播的必要条件，却也决定后来研究者能够看见怎样的世界。如果只有支持明确结论的片段被保留，失败、异常和无法裁决的部分便可能消失；后来者面对的资料集合，就不再等同于研究实际经历的全部结果。

这种问题不应被简单归因于某一类人的动机。即使每位作者都诚实，叙述形式和评价要求也可能使某些材料更容易流通。一个连贯故事便于理解，却可能把探索中的分岔写成事前已有的路线；一个显著平均效应便于摘要，却可能使条件差异退到附属位置。研究共同体需要讨论的，是如何保留传播的清晰性，同时让关键选择与失败能够被追踪。

发表也会回写研究。某类结果被认为重要，研究者便可能优先选择能够得到这类结果的问题；某类资料难以发表，取得它们的成本就更难获得支持。这种回写不必然有害，它可以形成共同标准、集中资源并提高质量。但如果评价长期奖励结论的新奇程度，而不重视测量修复、资料维护和边界检验，共同体便可能积累许多醒目的主张，却缺少判断它们何时成立的条件。

发生学因此把发表看作知识过程的一部分。论文不是研究结束以后附加的包装，而是决定何种证据能够被继承、何种问题能够继续发生的安排。这个判断并不要求每篇论文都公开全部工作日记。它要求主要主张能够连接到相应材料，探索与检验的关系能够说明，重要失败不因叙事不便而被删除。公共知识需要摘要，也需要从摘要返回依据的路径。

以一个虚构的多中心研究为例，总体平均结果支持模型，但两个中心出现方向相反的关系。发表时可以把总体适用声称与中心条件声称分开，各自连接样本范围、处理版本和不确定性。摘要报告总体关系及尚未解决的中心差异；正文说明哪些操作一致、哪些记录不足；可复核材料保存中心划分和合并规则。这样，读者能选择复核总体汇总，也能专门研究方向反转，而不必先接受作者对反转的解释。发表所增加的是问题的公共身份与进入依据的路径。

发表制度的目标因而不应是让所有研究都给出确定答案。它应使确定程度与证据相称，并让暂时不能裁决的问题拥有可检索的身份。一个未决结果如果已经排除了若干解

释、界定了资料缺口，就具有知识价值。它不需要被改写成支持性发现，才能进入科学的积累。新典范的制度内容，首先体现在共同体愿意保存这种有限而具体的成果。

39.2 不收敛结构怎样成为可发表对象

提出“不收敛结构也值得发表”并不意味着所有散乱资料都应获得同等评价。未经校准的设备故障、无法解释的缺失和任意挑选的图形，不能仅凭不整齐便成为新发现。可发表的结构应至少说明被估量、采样安排、误差来源和形状取得方式，使他人能够判断它是真实条件差异、方法问题还是偶然选择。打开误差项需要更多辨认工作，而不是取消证据要求。

一种有用的报告可以从异常的可重复条件开始。它说明某个模型在何种任务中预计结果稳定，实际观察到怎样的中心、尺度、方向或分裂变化；哪些常见解释已经检查，哪些尚未排除；新增资料能否改变竞争模型的比较。报告的价值不必依赖最后找到唯一机制。只要它把一个原先模糊的失败转化为可公开检验的问题，就已经为共同体增加了对象。

发表格式应允许不同成果具有不同身份。测量修复报告说明装置或程序怎样产生误导，结构报告描述可重复形状，机制报告提供因果辨识，迁移报告检验外部条件，资料报告保存重要但尚未充分解释的事件。它们可以彼此连接，却不需要每一种都装成完整理论胜利。清楚的身份比统一的宏大结论更有助于读者判断能够继承什么。

负结果的发表可以用一份明确的限制声明代替胜败标签：原主张预期什么范围，研究实际能排除什么范围，哪些未完成操作阻止更强判断。编辑与评议者据此判断稿件新增的约束，而不只检查某个检验是否显著。若实施偏离使核心比较无法解释，报告仍可保存有用的程序失败，但须以实施报告的身份发表。不同成果取得合适位置，才能避免作者为满足统一的“发现”要求，把有限材料提升成没有根据的机制结论。

还需要防止不收敛成为新的发表偏好。如果共同体开始奖励复杂分布和反常现象，研究者也可能忽略稳定结果，或过度切分资料寻找差异。新制度应同时接受有价值的稳定性证据：某种摘要在明确范围内足够，某个记录字段没有额外收益，某种扰动没有改变关键关系。发生学的主题是稳定性怎样产生和改变，而不是使每篇研究都证明世界无法稳定。

报告结构也应保存原始问题与后来问题之间的关系。探索中发现的新形状可以独立成立，但它不能替代原先未获支持的预测而不作说明。原问题失败和新问题产生可以同时发表；这样，创造性不必依靠擦除历史获得合法性。科学积累需要知道研究为何转向，才能判断新方向是由材料推动，还是只是为了维持一个永远成功的故事。

39.3 复现的对象必须先被确定

谈复现时，首先应问复现什么。利用同一资料和程序重新得到结果，检查的是计算与处理链；在相似安排下取得新资料，检查的是经验关系；改变机构、人群或任务以后检验模型，则涉及迁移。三种工作相互支持，却不能彼此替代。代码能够重算原表格，不说明原关系能在新样本出现；跨场景平均效应相似，也不证明原计算没有错误。

同值复现并不是全部科学的唯一形式。对于随机过程，可以复现分布或参数范围；对于动态机制，可以复现条件响应和更新关系；对于历史系统，可以检验模型对新时期的预测。关键在于复现对象应先于结果明确。若事前要求数值一致，失败后不能随意改称只需形状相似；若事前提出方向性预测，反向结果也不能以系统复杂为由算作另一种成功。

形状复现需要可操作定义。“两张曲线看起来相似”可能高度依赖坐标范围、平滑程度和观察者判断。研究者可以规定比较的时间归一化、关键特征、容许偏差和不确定性处理，并说明哪些变换是合理的。若每次都允许任意平移、缩放和重排，许多不同机制都可能被调整成相似形状。可复现性来自受约束的对应，而不是无限灵活的视觉认同。

复现也不要求历史系统回到完全相同的过去。一个经过学习的主体无法无条件重置，但可以在新的对象或新的任务序列中检验同一更新假说。研究者需要说明哪些初始条件必须匹配，哪些差异被模型允许，哪些差异本身就是迁移检验。承认不可逆历史不等于放弃重复，而是把重复目标从过去事件的复制转为关系的再检验。

复现失败之后，应先确认原声称的适用范围。若新研究明确超出原范围，它更接近外推检验；若在原范围内仍失败，就对原关系形成更直接限制。范围不能事后无限收缩，因此原报告应尽早说明条件。一个从未说明条件的普遍主张，不能在每次反例出现以后把反例排除，再宣称始终未受影响。清晰边界保护真实知识，也限制不合理外推。

团队间可以在新结果揭示前先交换一份复现约定，确定目标、材料资格、允许变更与判定范围。独立团队发现原协议缺少某个操作细节时，应先把缺口列明，再区分遵照已公开程序的复现与采纳补充说明后的复现。两者回答不同问题：前者检验成果是否已被充分交付，后者检验关系在补足条件后能否重现。原作者的经验可以帮助澄清，却不能使后补条件追溯成为原文已有内容。这种安排既保留协作，也使无法传授的个人操作成为可研究的继承障碍。

39.4 从文件保存到证据关系的保存

资料能够下载，并不意味着它已经可以复用。一个没有变量定义的表格、一段缺少版本的程序、一份无法连接到样本来源的结果，都可能形式上开放而实质上难以理解。知识基础设施需要保存的不只是文件，还包括文件之间的关系：哪个结果来自哪些事件，哪些转换产生了哪个变量，哪项主张使用了哪个版本，哪些资料后来被修正或撤回。

FAIR 原则提出可发现、可访问、可互操作与可复用，强调标识、元资料、来源和使用条件；它允许在必要时采用身份与授权程序，并不把可访问简单等同于任何人无限制取得所有内容。[72] 本书沿着这种基础进一步讨论事件与主张的连接：一个资料对象不仅要能够找到，还要让使用者理解它在发生过程和论证中的位置。

事件标识可以把同一研究的多个层次连起来。原始读数、处理后变量、模型输出和发表图表不应互相覆盖，而应保留可追踪映射。若一个事件在两篇论文中被使用，它应能被识别为同一来源，避免在综合证据时被重复计算。若同一原始资料产生多个分析，也应说明它们并非彼此独立的重复。证据数量不是文件数量，依赖关系必须可见。

主张也需要身份。一个研究可能同时提出描述、机制、预测和规范建议，这些主张使用的材料并不完全相同。将它们分别连接到依据，可以使后来的修订更精确。某个数据错误可能只影响一项参数估计，而不影响全部方法论论证；某个干预失败可能削弱机制，却不取消原来的描述结构。没有这种连接，修订容易在全盘否定和全盘保留之间摇摆。

版本关系应同时服务当前使用和历史解释。接收者引用旧版结果时，可以明确它是重建当时决策的历史依据，还是仍用作当前估计；两种用途对更正的要求不同。资料页应提示旧版被替代的原因、新版入口及受影响的主张，保留未受影响部分的状态。若使用者继续以旧估计支持现时决定，就需要重新说明理由；若研究当时为何行动，则旧版恰恰不可抹去。这样，最新值不会覆盖历史认识，保存历史也不会成为继续传播已知错误的许可。

来源说明还应区分直接观测、推断与模拟。模型填补的缺失值不应与原始记录拥有同一身份，研究者对参与者动机的解释也不应被写成设备直接输出。不同材料可以共同工作，但其证明责任不同。机器可读格式越容易自动整合，越需要保留这种语义区别，否则规模化处理会把不确定推断迅速传播成看似确定的事实。

基础设施的目标不是建立一套囊括所有学科的巨大本体表。不同领域有不同对象和记录需求，共同层可以只规定身份、时间、版本、来源、权限与主张关系。领域内部再保存

专业定义。这样的分层使跨学科比较有连接点，同时避免为了统一格式而消除真正有意义的差异。可互操作不是所有人使用同一种语言，而是转换关系明确且不假装无损。

39.5 错误 修订与撤回的公共历史

科学允许错误，不能意味着错误没有后果；要求纠错，也不应意味着任何修订都被视为不诚实。认识活动面对未知，程序和解释都可能需要改变。制度的任务是使错误能够被发现、定位和修复，同时区分普通不确定性、操作失误、分析缺陷与不当行为。把所有问题压成一种道德判断，会使研究者更不愿公开尚可修复的错误。

一个虚构例子是研究团队发现时间戳转换错位，导致部分响应被配到前一刺激。修正以后，某个时间效应减弱，而其他主要结果保持。合理的更正应说明错误发生在哪个转换、影响哪些事件、改变哪些图表与结论，并提供修正版本。只写“结论不受影响”不足以让他人判断，全部撤回也可能不符合实际影响范围。修订的尺度应与证据链中的损害相称。

时间戳错误还可能沿继承链扩散：第二篇论文直接复用了错配变量，第三篇只使用原研究的校准程序，两者受到的影响不同。更正者首先定位发生错误的转换，再检索哪些下游结果依赖这一输出；第二篇需要重算，第三篇则检查校准是否包含同样转换。依赖无法自动查明时，应把待核查范围公开给相关使用者，而不是笼统宣布全部无效或全部安全。这样的修复把主张关系转成实际工作顺序，也揭示为什么文件开放之后仍需要有人维护其知识状态。

更正记录还应保留被修改的内容。新文件覆盖旧文件虽然看起来整洁，却可能使已有引用失去对象。应能够找到旧版、知道它已被替代，并沿关系进入新版。对需要限制公开的资料，也可以保留版本说明与修订范围，而不暴露敏感内容。透明不是取消所有访问边界，而是让知识状态与变化理由能够被理解。

撤回同样需要明确范围。某项研究若失去可靠资料基础，其经验结论可能不应继续使用；其中某个独立的形式推导或公开方法，也未必因此逻辑失效。反过来，一个方法有价值也不能成为保留错误经验主张的理由。公共记录应让读者分清哪些部分受到否定，哪些部分仍需独立评价。精确区分既保护有效知识，也防止错误借有效部分继续传播。

制度还可以从错误类型中学习。若多个项目反复在同一时间转换、单位处理或版本选择处出错，就值得修改共同工具和培训，而不是每次只要求个人更谨慎。这样的修复把个

体失误转化为组织改进。它也体现回写：错误被公开识别以后，改变下一轮研究的条件，使同类失败更难重复。只有惩罚而没有结构修复，可能留下同样的发生条件。

承认错误的公共历史并不会自动削弱科学信任。信任可以建立在可纠正性上，而不是建立在从不出错的表演上。但这种信任需要实际行动支持：错误发现后是否及时说明，影响是否充分追踪，相关使用是否获得更新。发生学所要求的历史责任，正是让知识承担其曾经进入世界的后果，而不只保存最后一次正确表述。

39.6 开放 权限与参与者的边界

事件档案比普通数值表更丰富，也可能包含更敏感的信息。期待、策略、历史与互动路径可以帮助研究，同时可能暴露个人经历或行为模式。因此，增加可追踪性不能被当作无限收集和公开的理由。研究目标需要哪些资料，哪些可以用较低分辨率表示，哪些应留在受控环境，必须与认识收益一起判断。

开放有不同层次。研究可以公开问题、操作、变量定义、分析程序和非敏感汇总，同时对个体原始轨迹设置适当访问条件；也可以让独立复核者在受控条件下检查关键结果，公开其可核验说明。不同安排有不同限制，但都可能支持有效审查。不能因为资料不能完全公开便宣布研究不可检验，也不能用隐私一词拒绝提供任何方法和来源信息。

匿名化并非一个删除姓名便结束的动作。丰富时间与路径记录可能与其他资料组合，使个体再次被识别；保留什么粒度需要考虑实际用途和潜在关联。这里不预设一种技术可以消除全部风险，而要求研究者在设计阶段减少不必要信息，并说明公开资料经过哪些处理。处理本身可能改变分布或关联，也应进入复核范围，不能把保护措施当作统计上透明的操作。

参与者的拒绝和退出也应具有真实意义。若一个人不同意提供某项敏感历史，档案可以标记缺失并收缩推论，而不应自动把他排除在所有机会之外。尤其当研究工具进入教育或制度决定时，复杂记录可能变成新的资格门槛。事件化方法本来希望恢复被平均忽略的处境，不能反过来要求每个人交出更多生活资料，才能获得被理解的权利。

研究者还应区分资料的原用途与后续用途。为回答一个有限问题取得的记录，后来被用于能力排名、行为预测或商业决定，可能改变参与者承担的后果。是否可以扩展用途，需要新的依据和适当程序，不能仅以资料已经存在为理由。知识可复用性与人的控制权之间存在真实张力，制度应让这种张力被讨论，而不是隐藏在技术许可的细节中。

权限管理也需要避免只服务于机构控制。若访问条件过于任意，可能使独立质疑无法进行；若只允许支持原结论的使用，所谓受控访问便成为理论保护。合理安排应有明确标准、记录处理过程，并提供可解释的拒绝理由。研究者不能要求别人信任无法审查的结果，同时又把一切审查困难归给资料敏感。认识责任与保护责任需要共同设计。

这样的边界不会使基础设施失去开放精神。开放的核心是让相关人能够理解证据怎样形成、在适当条件下检查关键主张，并知道不能检查的部分。它服务于知识与人的关系，而不是追求资料毫无边界地流动。事件记录越深入发生，越需要对谁能够使用这些记录、使用后果由谁承担保持清楚。

39.7 维护劳动怎样获得知识地位

一个资料库能够持续使用，依赖大量不容易出现在理论摘要中的工作：校验单位、解释字段、修复程序、维护版本、回答访问问题和整理错误。这些工作往往不产生一个醒目的新命题，却决定新命题能否被检查和继承。若评价只承认最后发表的结论，基础设施便容易成为无人长期负责的附属项目。知识积累需要让维护劳动拥有明确地位。

贡献记录可以把这种劳动变得可见。谁建立了事件字典，谁校验了历史版本，谁发现并修复了数据错配，谁维护了复用接口，都可以被具体说明。贡献不必全部通过同一种署名规则表达，但应能够被引用和评价。这样，资料生产、方法维护与理论解释之间的关系不再只由论文第一作者的可见性决定。

奖励维护并不意味着以记录量或下载量取代理论贡献。一个资料库可能拥有大量文件却几乎不能支持新判断，一个小而清晰的资料集却可能解决关键争议。评价应关注复核是否更有效、重复错误是否减少、不同团队是否能够提出新问题，以及维护成本是否可承受。简单数量指标可以提供线索，但若被当作目标，也会产生新的策略性回写。

维护成效可以通过接手任务检验。让未参与原项目的研究者从已发表结论出发，定位数据版本、重建关键结果，并指出需要哪些未公开判断，能够发现文件缺失、变量歧义或软件依赖等障碍。若还要判断新版基础设施是否改善效率，可以预先选择难度和处理链相近的任务，把独立接手者分配到旧版与新版，在相同时间预算下比较；任务与人员安排应尽量分开。若只能由同一团队先后处理不同结果，难度差异和练习均可能影响表现，这类材料只作为可用性试用，不能单凭耗时缩短归因于改进。评价关心恢复依据所需劳动、发现错误和可完成判断的范围，也允许据此删去没有实际收益的复杂字段。

维护退出需要准备一次可执行的交接。设想原负责人离职，接替者至少要接收当前推荐版本、旧版索引、尚未解决的问题、运行环境说明和访问责任。若无人能够继续运行全部程序，可以冻结可解释的结果与方法说明，标记停止维护的日期及不能再保证的功能，而非留下一个看似活跃的入口。原始资料因期限或约定不再保存时，也应保留允许保存的元资料与主张状态。继承者由此知道哪些判断仍能核查，哪些必须重新取得证据，维护的终止不会被误认作结论永久有效。

跨机构协作可以分担这种责任，但不应把责任稀释为无人承担。共同标准需要明确谁修订、谁处理冲突、谁保证旧版本能够解释。不同团队对字段有不同理解时，争论本身可以推动概念澄清。若只是用一个统一代码掩盖分歧，技术整合越快，错误传播也可能越快。维护工作因此包含持续的认识协商，而不仅是服务器运行。

从发生学看，维护并不是研究完成以后的保管。它使过去成果在新的条件中继续发生，决定哪些关系可以被继承、哪些需要重建。知识的历史厚度不是所有文件累积的体积，而是后来者能够沿着证据关系继续判断的能力。赋予维护劳动知识地位，是让科学对自己的持续存在承担责任。

39.8 共同体怎样学习而不封闭

新的发表与基础设施安排如果要改善研究，必须允许自身被研究。预先登记可能提高某些透明度，也可能在不适当实施时增加形式负担；复杂档案可能帮助复核，也可能挤压真正的探索时间。制度不能因为以严谨为名便免于效果评价。应比较它对问题质量、错误发现、参与负担和知识继承的实际影响，保留简化或替换的可能。

探索与检验也需要制度上的相互支持。探索允许研究者在材料中发现未预见的关系，检验要求这些关系在新资料中承担约束。如果只奖励预先明确的命题，新对象可能难以出现；如果所有探索都被当作验证，证据又失去分量。预注册讨论所强调的区分，可以帮助二者各自获得正当身份，而不必把创造性与严格性写成相互排斥。[71]

共同体还应保留不同方法的竞争。事件化记录不必成为所有研究的统一入口，简单测量在明确任务中可以继续有效。某些历史和理论研究无法采用同样实验程序，也仍可提出有约束的论证。共同标准可以集中在来源、推论、反例和修订责任上，而不要求所有知识使用同一种资料结构。真正的统一应容纳对象差异，而不是通过格式消灭差异。

一个能够学习的共同体，需要让异议进入证据关系。质疑者应说明针对哪项主张、依赖什么材料、何种新结果会改变判断；原作者应回应实质问题，而不只维护理论名称。双

方都应允许局部修订，不必把每次争论升级为整个范式的存亡。这样的争论可能不如宣言鲜明，却更能积累可以共同使用的知识。

本章的继承链可以回到多中心例子中检验：发表给予总体结果与中心反转各自的主张身份，复现约定确定究竟重做哪个关系，来源与版本连接阻止同一证据被反复计数，更正程序沿下游依赖修复错误，维护交接使这些工作不随原团队散去而停止。链条任何一环缺失，都有具体后果：反转被隐藏、目标被替换、旧错继续使用，或结果只剩无法复核的名称。制度改进因而有可观察的任务，而不只是一套赞成开放的原则。

这条链能够证明的是知识更容易被发现、检查、修订和继承。它尚不能证明被保存的理论一定新颖，也不能把基础设施运行良好当作其经验机制已经成立。发表更多、更易复现与研究纲领产生新内容仍须分别判断。第 40 章由此接手最后一项责任：把对象重建、设计进步、机制证据和制度收益放回各自层次，确定哪些成果足以继续投入，哪些需要收缩，哪些主张应停止使用。科学的共同记忆为这一判断提供条件，判断本身仍要面对材料和竞争理论。

第 40 章 新典范的研究纲领与开放边界

40.1 新典范不能由自己的名称证明

一套理论把自己称为新典范，并不会因此改变知识的组织。新术语可以吸引注意，也可以使旧问题呈现新的联系，但典范的实质在于研究者怎样提出问题、构造对象、组织证据和修订判断。若这些活动没有变化，名称只是表达立场；若活动发生了有根据的改变，即使尚未取得宏大的历史地位，也已经具有方法价值。对本书而言，最重要的不是宣布时代已经转换，而是说明转换应当产生什么可见成果。

全书追问属性如何在发生中取得稳定性，最终需要把这一问题转成三类可分别评价的工作。对象重建说明研究原来把什么认作同一属性、为何需要改定被估量或比较资格；设计辨别力说明哪些新增安排能够使原先等价的解释分开；经验机制则用实际取得的材料判断某条通道及其条件。三者相连，却不能折算成一次预测胜出。重新定义对象可以构成认识进展，设计成功可以取得检验资格，而具体机制是否成立仍须另有证据。

科研新典范的承诺由此可以表达为三个相连的转变。研究从只报告结果推进到说明结果发生的条件，从把稳定性当作默认前提推进到研究稳定性怎样形成，从把模型放在过程外评价推进到研究模型和指标怎样进入下一轮过程。这些转变不要求取消现有实验、统计或理论，而要求它们在适当问题上承担更完整的发生责任。新典范的力度应体现在新增的关系和操作上。

原有研究底稿以常数不收敛和误差项为尖锐入口，提出读数事件、判据与退出条件，同时保留了尚未完成实证的状态。[2] 专著的推进不是把这一入口扩大成所有科学的总诊断，而是把问题重心转到属性稳定性如何发生、失配如何形成以及研究怎样修复。这样，稳定结果和不利案例都能够成为理论资源，不必为了维持批判姿态而被排除。

这也改变了“新”的证明方式。新对象可能是过去未被联合记录的事件关系，新机制可能是一个得到辨识的回写通道，新方法可能使资料更容易迁移和复核，新制度可能让失败更有效地进入修订。不同贡献需要不同证据，不能由一本书的概念统一替它们逐一兑现。理论雄心可以广泛，完成程度必须分项诚实。

一个研究纲领的生命力，因而不在于它能够解释多少已有例子，而在于它能否持续生成受到材料约束的新问题。它应让研究者知道下一项观察为何值得取得，哪种结果会改变

判断，以及失败以后有哪些具体修订路径。只有当这些工作可以被不同团队继承，发生学才从作者的解释视角转化为共同体可以使用的知识方式。

40.2 本体论承诺怎样获得研究内容

属性发生论首先是关于解释起点的主张。它反对把主体、客体与互动预先设成彼此完整的成品，再用外部影响补充关系。但这种哲学批评若要进入经验研究，必须说明它与其他框架在哪里产生可区分后果。仅把所有关系称为纠缠，不会自动增加信息；把任何对象都说成事件，也不能独自判定某个模型正确。概念需要通过有限模型和研究操作取得内容。

三方程、六路径与三原理在第 1 至 5 章已经确立；到了研究纲领的收束处，需要检查它们是否使推论承担了新的责任。如果由结果反推条件，能否指出哪些差异序列仍与资料相容；如果从条件组织预测显露，能否说明允许改变什么、必须保持什么；如果结果进入下一轮过程，能否定位实际发生的更新。三个体系的相互依存由这些交叉要求体现：不能只凭输出归因，也不能只凭安排命名结果，更不能让某一时点的解释脱离后续变化而永远有效。

以评价分数为例，发现相同得分来自独立解题和人机协作，首先改变的是事件同一性的判断。将两类任务分开以后，可以分别追问独立能力、协作能力和后续迁移，这构成对象重建；进一步设置工具撤除和延迟测量，才增加设计辨别力；取得结果并与竞争机制比较，才形成能力变化的经验证据。新的分类并不自动使原分数无效，也不能替新模型赢得预测比赛。它应说明哪些旧比较仍被保留，哪些需要转换，哪些已经是在回答另一个问题。

本体论与经验的关系因此不是单向演绎。哲学框架帮助看见问题，经验研究又可能迫使框架澄清概念。一个稳定反馈系统提醒理论不能把回写与失稳等同；一个有效压缩提醒理论需要承认任务充分性；一个没有主体报告却能可靠预测的模型，要求区分主体在发生中的构成与主体被记录的程度。材料不必直接否定整个本体论，也可以改变它如何获得操作表达。

这里尤其需要避免循环定义。如果把纠缠定义为必然导致复杂变化，再把复杂变化当作纠缠的证据，理论只是在定义与观察之间往返。较好的做法是独立规定候选通道和操作条件，再预测相应结果。即使概念不能完全脱离理论而定义，也应让关键证据不只是概念的重复。一个有意义的模型必须允许某些符合表面标签的案例并不满足其机制。

对象重建有时揭示，原目标把制度选择包装成自然属性，因而需要重新讨论评价目的。这种贡献要给出原定义的后果、新定义的理由与两者的对应，不能只靠新命名成立。若进一步声称新模型能更好预测原目标，就必须回到同目标的强基线比较；若改用新目标，则先独立论证其必要性，再按新任务评价不同模型。哲学论证与预测检验因此各有工作：前者可以改变何者值得估量，后者限制对一个明确目标能声称多少。任何一方都不能借另一方的成果免除自己的责任。

研究内容最终来自承诺与限制的结合。属性是发生中的稳定性，这一主张要求研究稳定性的条件；它不能允许研究者在稳定出现时一律说纠缠被压制，在变化出现时一律说纠缠已显现，而不给独立区别。理论应当在稳定与变化两边都承担解释任务。这样的对称责任，是发生论避免成为万能语言的基本条件。

40.3 数学结论与领域机制之间的距离

数学可以为研究纲领提供严格支点，也可以被误用为跨领域权威。一个条件命题在形式模型中成立，不意味着所有现实对象满足其前提；一个反例否决某种充分性声称，也不意味着相关哲学问题全部消失。研究必须保留从概念到模型、从模型到对象的中间论证，否则公式的精确性会掩盖适用条件的不确定。

第 13 章证明，观测等价而目标不同会阻止同一估计程序对不同目标同时一致恢复；它也展示单输出可以可辨识、循环反馈可以稳定。这些结果对本书具有双重作用。它们限制过强的普遍声称，同时指出新的建设方向：研究观察映射何时失去区分、什么干预能够打破等价、哪些反馈参数产生稳定、哪些条件变化使原摘要不再足够。边界并未减少问题，反而使问题更准确。

数学模型中的变量还需要与经验代理连接。状态变量可以代表知识、期待或资源，但每种解释都须说明如何观测或约束。一个形式上可观测的系统，在实际资料中可能因噪声和窗口不足而难以恢复；一个形式上稳定的系统，也可能在未建模外部输入下表现剧烈变化。模型不是现实的缩小复制，而是选择某些关系承担推论。选择本身应公开。

跨领域形式相似尤其需要谨慎。两个系统都可以写成状态更新和输出方程，并不说明它们具有同一种实体机制。相似可能只是表达足够一般，能够容纳许多过程。真正有意义的共同结构，应当在参数关系、允许干预、稳定条件或失败模式上提供额外限制。若共同形式没有排除任何可能结果，它可以作为记账框架，却不能单独成为统一定律。

对跨域公度的追求也不必立即寻找一个总分。甜度变化与经济响应没有天然共同单位，可以先比较无量纲的响应比例、相对于任务容许误差的失配、恢复时间相对于观察窗口的比值，或者模型在新条件中的相对损失。但每种标准化都需要非零且有意义的基准。把不同指标压成一个纠缠强度，可能再次隐藏对象差异，重演本书所批评的无条件汇总。

更可行的统一是保存一组可比较关系。各领域分别说明什么被更新、通道怎样识别、哪些稳定性得以保留、增加信息付出什么代价。共同体可以据此比较机制类型和研究难度，而不必宣布一种领域比另一种更活、更成熟或更真实。形式共同性与价值排序是不同问题，数学尺度不能自动转化为伦理地位。

数学工作的开放边界在于继续寻找足够强又不过宽的命题。何种压缩在什么损失下充分，何种反馈使估计目标随制度变化，何种观测与干预集合足以辨识更新，这些都可以形成独立研究。每个小结论若前提清楚、反例充分，便能支撑专著的一部分。比起一个尚未成立的通用定理，这种可连接的局部结果更有可能形成持久理论。

40.4 经验积累为何需要不同层次的成功

经验成果进入纲领时，应标明它改变了哪项承诺。某个形状在新资料中重现，使描述对象更可靠；某项操作分开两个预测，使设计具有辨别力；实际干预结果排除一条竞争通道，才推进机制判断；模型迁移到新条件，则增加其可使用范围。这里沿用第 38 章的证据关系，进一步问这些局部成果是否产生此前无法提出或无法检验的问题。纲领的积累单位是被新增证据改变的主张，而非支持性图表、文章或领域的数量。

例如两个独立团队都发现同类分布分裂，这增加了现象可靠性；若两团队使用同一种装置与处理程序，共同方法问题仍可能存在。另一个团队通过不同测量安排发现同样条件关系，才提供新的约束。多次重复的价值不只是次数，而在于它们在哪些方面独立、哪些条件保持、哪些误差源不同。研究纲领应追求能够改变判断的多样证据，而非累积表面相似案例。

成功也可以表现为有内容的缩限。若只有结果公开并进入行动的条件才出现指定信息效应，后续模型就应删除对隐藏条件的同样预测；若某类稳定任务无需额外事件字段，档案便可减去这些字段并保留恢复条件。缩限之后必须有可观察后果，才构成知识增长。如果只是把原主张改得更模糊，使它继续容纳所有结果，就没有增加约束。研究纲领需要保存缩限前后的适用域及下一步判断，才能区分更准确的条件理论与不断后退的表述。

跨领域研究尤其不能用一个领域的成功抵销另一个领域的失败。物理测量中的条件变换、生命系统中的状态更新、经济指标的制度作用和人机协作的后果承接，拥有不同证明责任。它们可以共享发生问题，却不应共享一张总成绩单。某个领域若只能提供解释重组，就据此评价；另一个领域若取得独立预测增益，则明确其增益。统一语言不能使证据越界流通。

累积还需要对未决结果保持耐心。一个机制模型可能暂时与既有模型等价，研究者可以保存它作为生成新问题的候选，而不急于宣布胜负。但这种保留应当伴随寻找区分条件的努力。如果长期没有任何可区分任务，模型的经验身份就应降低，转而承认其概念用途。开放并不意味着所有假说永久享有同等地位，资源投入仍需要理由。

一个有意义的阶段性成果，可以是形成了可靠事件字典、发现一种原先未见的失配、建立一个条件性机制，或证明某种复杂记录没有必要。它们都能够改变后续研究。判断研究纲领是否进步，应看这些成果是否持续相互连接，是否使新问题更明确、实验更有区分力、失败更能指导修订，而不只看支持该理论名称的论文数量。

这种积累方式使理论保持创造性。新对象和新机制不必被迫立即进入一个完整体系，可以先取得局部身份，再与其他结果比较。与此同时，理论也不能无限把不利材料归为尚未成熟。每个未决位置都应说明缺少什么、可能怎样解决、何时值得停止。有限成果与开放问题之间的清晰边界，是研究纲领能够持续工作的条件。

40.5 与既有研究传统进行强比较

新典范的论证很容易借助一个过于整齐的对手：旧科学只追求常数，只排除主体，只把残差当噪声。这个形象便于宣言，却不足以承受实际学科史和当代方法的复杂性。计量学处理不确定性与相关误差，控制理论研究反馈与稳定，经济学研究异质性和预期，感知科学研究适应与情境。发生学必须与这些已有能力相遇，才能确定自身真正增加了什么。

强比较首先承认共同问题。不同传统可能已经用其他概念处理本书关心的部分机制。承认这种重合不会消灭本书价值，反而能够避免重复发明，并把努力集中到尚未连接的地方。比如多个领域各自研究历史依赖，SDE 可以提出共同的事件记录和证据责任；但共同程序的价值需要通过实际使用证明，不能仅由跨领域命名获得。

比较还要区分哲学、方法与领域理论。科学哲学讨论知识如何增长，不等于为某个经济参数提供具体预测；一个统计模型预测失败，也不能直接否定一套哲学。拉卡托斯区分研究纲领中能够产生新经验内容的进步性调整与仅追随已知困难的退化性调整，这个区分

本身已经要求关注修订的后果。[73] 因而本书不能把所有保护性修订都描述为免疫，而应对自己和对手使用相同标准。

若一个已有模型加入有独立依据的新条件，随后成功预测未见结果，这应被视为真实进展，而不是仅因它保留旧名称便被贬为补丁。相反，SDE 模型若每次失败后才增加不可观测状态，却没有新预测，也不能因使用发生语言便被视为更深。理论是否进步，取决于修订增加了什么可检验内容，而不是修订属于哪个阵营。

强比较也允许多种理论在一段时间内并存。一个框架擅长对象建构，另一个擅长局部计算，第三个提供制度评价。它们可能互补，也可能在某些任务中竞争。研究不必先决定谁拥有世界的全部解释权，才能利用各自成果。但互补不能成为逃避冲突的万能答案：当两种框架对同一操作给出不同预测时，仍应让资料裁决相应声称。

原创性因此需要精确表达。提出新的概念区分、组织新的研究程序、证明一个条件命题和发现独立机制，是不同贡献。可以同时拥有其中几种，也可以只完成一种。若只是用新词重述已有结果，应该说明解释组织上的收益；若声称不可还原的机制增量，就应给出近邻理论不能同样产生的限制和预测。这样的表达比笼统宣布全球首次更能经受检验。

与既有传统的关系最终是一种生产性的张力。它们为本书提供强约束，防止发生学退化成无所不包的修辞；本书也可以反过来追问这些方法何时把对象身份、判断位置和制度回写留在模型之外。双方都不需要预先获胜。真正的新问题常常产生在各自有效范围的接缝，而不是产生在对方被描绘得最薄弱的地方。

40.6 事实解释与规范责任的连接

发生学强调模型和制度进入下一轮过程，容易进一步引出规范问题：应该保留什么稳定性，应该改变什么路径，谁承担记录与干预的代价。但描述某种发生机制并不能独自推出应当怎样行动。一个系统能够稳定运转，不说明它值得维护；一个系统充满变异，也不说明它更自由或更有价值。价值判断需要公开目的、权利与后果，不能由动力学形容词替代。

例如一种评价制度通过强反馈使分数稳定提高，可能改善学习，也可能只促进对评分规则的适应。判断好坏需要考察教育目的、独立迁移、负担与公平，而不仅是反馈是否存在。相反，一种制度允许较大个体差异，可能尊重多样路径，也可能放任资源不平等。发生机制提供理解后果的材料，规范论证则决定哪些后果应被重视，两者需要连接却不能合并。

主体性的研究也受到这一边界约束。跨轮记忆、目标保持和后果承接可以作为某些系统能力的代理，却不能自动决定意识、人格或道德地位。第 31 至 35 章把人机系统中的能力放回具体架构和任务，不把单次调用代表全部 AI，也不以某个记忆测试解决全部主体问题。代理有效说明某项操作关系成立，规范地位还需要其他论证。因而一个能力指标通过验证，只能取得对应用途的资格，不能顺带决定系统应承担何种权利与责任。

对人的研究尤其不能把不稳定当作更真实的生命证据。稳定边界、熟练技能和可靠承诺可能具有重要价值；变异也可能来自损伤、资源不足或测量失配。发生学若要尊重主体，就应允许不同稳定与变化具有不同意义，而不是要求个体不断表现复杂性才能被承认为活的存在。把理论偏好变成对人的评价尺度，会违背其恢复发生条件的初衷。

制度修复因此应当是比较性的。研究者提出改变路径或规则，需要说明相对于哪些可行替代方案更有理由，哪些群体受益、哪些承担成本，以及不利后果出现时怎样恢复或补偿。经验研究可以估计部分后果，不能代替所有价值选择；价值主张可以确定目标，也不能豁免对实际效果的检查。两者相互约束，才使创造新安排具有责任。

研究自身也在这一关系之内。增加档案可能改善知识，却增加隐私与劳动成本；公开失败有利于共同体，却可能给个体带来声誉压力；快速修订模型可能减少错误，也可能使使用者难以跟上。没有一种程序能够只产生收益。新典范需要把这些张力纳入制度设计，而不是在方法之外附上一段笼统伦理声明。

发生学的规范贡献可以在于使被隐藏的选择重新可见。它提示研究者，某个被当作自然尺度的指标可能包含制度决定，某种被称为效率的压缩可能把代价转移给他人。揭示这些关系能够改善公共讨论，却不自动给出唯一政策。让事实解释与规范理由各自说清楚，再在具体行动中共同承担后果，是比从本体论直接推出应然更稳固的道路。

40.7 怎样判断研究纲领正在前进

研究纲领的前进不应只按理论覆盖范围计算。覆盖越广而限制越少，可能意味着内容变薄。更有意义的指标是问题是否变得更精确，模型是否产生新的区分，资料是否能够支持以前不能完成的判断，以及失败是否减少了无效重复。这些变化可以在有限项目中观察，不必等待整个科学共同体改换名称。

三项优先研究应成为有不同决策后果的项目组合。E1 味觉研究若辨认出路径效应，但既有适应模型已充分说明它，就保留设计成果，不登记独有机制；E2 边际消费倾向研究若历史版本改善了实时评估，却未改善同目标预测，就将成果记为比较资格与资料组

织，并停止追加没有收益的复杂参数；E3 AI 教育若只有协作收益，就推进相应协作工具，同时暂停把它称为个人能力提升。三种结果都可以有价值，但分别决定下一笔资源用于何种问题，不能汇成支持本书的一票。

阶段推进可以设立几道与任务对应的入口。对象和资料身份未清楚，资源先用于定义与事件重建；两个模型还没有在可行条件下分开，资源用于形式推导与设计校准。设计原则上有区分能力以后，还要检查实际数据的精度：若不确定范围仍宽，机制保持未决，可以改善精度或因资源限制暂缓，不能把未支持写成已否定；若资料已排除预定的重要机制效应，或稳定违背关键预测，才据此修订或停止相应模型。机制或预测关系在规定条件中成立后，再扩大外部检验。每道入口都有独立成果，并交代尚未获得的资格，使扩大研究依据实际约束，而非仅依据理论信心。

时间安排可以帮助组织工作，却不能替代这些完成条件。某个季度结束不意味着辨识问题自动解决，一年计划完成也不意味着理论成熟。研究应同时保留日历进度与证据进度：何时做了什么，实际取得哪些约束，哪些前提仍未满足。若关键条件不足，调整计划是合理的；但调整理由应公开，不能把不断延期变成理论永远免于检验的方式。

共同体层面的成果还包括共享工具和失败库。若不同团队能够使用相同事件协议，却保留领域差异，并在复核中减少误解，方法就获得独立收益。若协议负担很高而很少改变判断，则应简化。基础设施的价值不由采用者人数独自决定，也不由作者对其哲学意义的评价决定，而由它在实际研究中帮助完成的工作决定。

理论创造仍需要不以短期结果为唯一尺度的空间。一个概念项目可以先研究为何两种记录不应算作同一事件，一个数学项目可以寻找现有干预无法打破等价的证明；两者未必立即产生更低预测误差，却能改变经验研究的目标或可行路线。继续支持它们的理由应是障碍变得更清楚、论证出现新约束、与后续研究的接口更明确。若反复重述同一困难而没有这样的变化，就需要暂停投入。长期性可以解释进度缓慢，不能替代进展的内容。

阶段评议应把三层增量与继承条件分别记录。对象重建需要新旧目标的对应和案例约束；设计辨别力需要在现实精度与允许操作下仍能分开候选解释；经验机制需要独立资料及强竞争模型的比较。第 39 章的知识基础设施则保证这些判断可复核、可继续，不替任何一层制造成功。若只有对象与程序收益，纲领就以此有限身份继续；若出现独立机制增量，才发展相应领域理论；若连这些收益也长期无法实现，则降低优先级或停止项目，留下已经获得的有效材料。

40.8 开放边界与有限的退出

一个能够发展的纲领需要说明自己的退出条件，而且退出必须与主张层次相称。数学充分性命题被反例否决，就撤回该推导；具体机制模型在有辨识力的实验中失败，就修改或放弃相应通道；某项预测没有通过外部检验，就限制其用途。不能把一个局部反例无限扩大为全书失效，也不能把所有反例都缩小为无关紧要的技术细节。

本体论假说的评价通常更间接。它可以因概念矛盾、持续误导研究、无法与经验建立有内容的连接而失去说服力；也可能因组织出新的可检验问题而获得支持。但这些评价不同于一次实验的直接拒绝。研究者应承认哲学层面的开放性，同时防止用这种开放性保护已经作出的具体经验声称。一个理论可以在哲学上继续讨论，而在某项实践用途上停止使用。

跨领域统一也有自己的边界。如果各领域最终只能共享极一般的语言，找不到有约束力的共同关系，那么本书应把统一性限定在问题组织与研究责任上。若出现可比较机制，再讨论更强统一。不存在一个总分或通用单位，并不自动使研究无效；但也不能一面拒绝任何公度，一面宣称已经证明所有案例是同一种现象。统一程度必须与实际连接程度相称。

还有一些开放问题不能由当前论证提前回答。属性身份怎样在长期制度与技术变化中保持，某些历史怎样被压缩为有效状态，成熟与退化能否形成跨领域可比较而不价值混淆的概念，机器的后果承接怎样与人的主体经验发生关系，都需要继续研究。本书能够给出区分和入口，不能把尚未完成的机制写成已经发现的事实。

退出需要落实到用途和资源，而不只修改措辞。一个失去支持的模型应停止用于它未能通过的预测任务，留下失败条件和可继续使用的局部关系；一个无收益字段应从常规采集中撤下，保留在何种新任务下值得恢复的说明；一个长期没有区分条件的经验假说，可以转入概念讨论而不继续占用同等实验资源。重新进入研究也要有新理由，例如出现新的可行干预、独立资料或严格推导。这样的可逆退出保留创造机会，同时使失败确实改变下一轮发生条件。

全书各编由此取得不同而相连的位置。物理测量提供条件变换与稳定约束的问题，生命和感知揭示历史怎样成为当前可响应状态，经济指标展示公开读数如何进入制度行动，伦理与教育要求分清能力对象和价值选择，AI 研究则把任务边界、工具版本与后果承接带入复合系统。这些论证共同支持追问稳定性的发生条件；它们没有共同证明一条跨域经

验定律。能够继承的统一性，首先是对对象身份、比较资格、因果通道及修订后果逐项负责，再由各领域实际证据决定更强统一能够走多远。

从属性之问到科学发生学，建设性的成果应落在三处：让原先混同的对象获得有理由的区分，让原先不可裁决的解释获得有效设计，让真正取得的机制在边界内改善预测与行动。三处都可能成功，也都可能受限，简单而充分的模型仍是可接受的结果。科学的创造力由此包含形成新对象与新路径的能力，也包含撤回失效用途、保存判断历史的能力。结语所要承接的历史责任已经有了具体基础：我们交给后来者的，不只是最后一次答案，还有答案能够成立、被使用和被改变的条件。

结语 让科学研究承担发生的历史

41.1 属性之问经过了怎样的改变

从甜属于谁开始，问题已经发生了变化。我们不再满足于把一个性质交给某个主人，也不把主体、客体与互动列成三个相加的原因。更具体的问题是：某种规定性如何在同一发生中变得可以指认，哪些差异使它获得边界，哪些纠缠使它保持，又有哪些后果使下一次发生不同于这一次。属性由此不再只是等待填写数值的空格，它具有形成、维持、迁移和重建的历史。

这个变化没有取消对象。相反，它要求更认真地说明对象身份。一个量在不同条件下变化，可能仍属于同一目标的合法转换，也可能提示目标已经改变；一个结果保持相同，可能来自相同机制，也可能来自不同变化的补偿。对象的同一性不能只靠名称，也不能只靠数字。研究必须说明跨变化追踪它的理由，并在理由不足时允许重新识别。这样获得的对象更适合公共研究，因为它的边界能够被讨论和检验。

三号位论证所取得的结果，在于同一材料可以承受不同说明责任，而这些责任又彼此限制。一个结果能否被共同指认，影响它能否进入过程比较；过程怎样接续，影响判断者能够把什么当作理由；判断位置改变，又可能改变下一轮可供指认的对象。把客体、互动和主体预先放成三个独立承担者，会遗漏这种共同规定。将它们理解为号位侧重，则既保留区分，也允许追查侧重之间实际发生的转换。相应转换是否具有因果作用，仍由具体通道与材料决定。

本书还改变了稳定的地位。稳定不再只是必须打破的假象，也不只是无需解释的前提。它可以是约束、反馈、补偿、聚合、冗余和有效压缩的成果。认识稳定的发生，使我们能够判断何时值得保留简化，何时需要重建条件。一个能够维持可靠性并回应新证据的系统，与一个把所有差异排除在记录之外的系统，可能暂时有相近读数，却具有不同未来。发生学把这种差别带入知识。

不收敛因此取得了更明确的位置。它是某些问题的重要入口，却不是所有复杂性的共同结论。统计估计、动态轨迹、分布散布、跨条件失配与规范分歧需要分别说明。不可辨识不等于不收敛，反馈也可以产生稳定。保留这些边界并未削弱批判，而是使批判能够指向真正有问题的外推和使用方式。理论的锐度来自能够准确指出哪里错配，以及什么证据可以改变判断。

从归属到发生，从读数到事件，从异常到研究程序，这三次改变共同构成了本书的基础成果。它们不保证每个案例都需要复杂解释，也不保证每个新记录都有价值。它们提供的是一种可以继续工作的秩序：把属性成立的条件公开，把未决问题组织起来，把当前结果与后续发生联系起来。科学研究由此不仅获得更多描述，也获得改变自身问题和方法的能力。

41.2 本书的构成论证走到了哪里

本书最需要保留的主张，超出条件会影响结果这一常识。条件影响的说法通常预先规定一个目标，再比较不同条件下它如何变化。属性发生论还要追问：什么使这些结果能够被当作同一目标的变化，什么使某个差异被视为对象内部的差异，而不是另一个对象的出现。只有这个问题得到回答，条件模型才知道自己究竟在预测什么。构成说明与条件预测可以合作，却不能互相替代。

第一编关于陶杯的分析已经显示这一点。材料连续性可以支持我们追踪同一只杯子，容纳液体的能力可以支持一种功能判断，修补历史又可以使它进入另一种过程类别。杯子没有因提问改变便任意变成其他东西，研究却必须为不同判断提出相应的同一性理由。若修补后仍沿用旧类别，后续差异可能被当作偶然误差；若每次修补都宣布新对象，又会丢失可以比较的连续性。合适的单位在这两种处理中承担可检验的边界。

由此得到的第一项结论，是属性判断具有构成前提。这个结论主要由概念分析、对象追踪和反例比较支持。它不要求每项实验重新建立全部本体论，却要求在对象身份发生争议时，不能只靠增加一个条件变量结束争论。变量的名字、读数的用途、类别的连续性，也可能是问题的一部分。把这些前提公开，能够让分歧落到确定位置：双方是在同一对象上给出不同预测，还是尚未在比较同一对象。

第二项结论，是显露反过来限制差异与纠缠的说明。若一个分类声称保存了关键功能，而新观察反复显示功能沿被删除的路径分化，研究者需要返回分类。若原有差异划分不能解释某种稳定显露，则需要检查是否遗漏了补偿、约束或有效压缩。三方程在此给出的，是相互规定的说明责任。它让构成论具有返回材料的要求，避免整体成为一个只会吸收细节、永远不必改变自身的名字。

第三项结论涉及研究中的时间。知识的形成留下版本，知识的使用又可能改变后续对象，两种历史需要分开追踪。研究者今天重新理解昨天的材料，并不改变昨天已经发生的过程；一个公开评价被用于明天的行动，则可能产生新的后果。全书的回写分析以这种区

分为前提。它建立了需要检查的通道和竞争解释，却没有从本体相互规定直接推出任何具体回写必然存在。

这些结论能够共同成立，是因为各自的责任明确。构成分析检查对象与判断是否可理解，模型推演检查假设与后果是否相容，经验材料检查具体机制是否获得支持，规范论证检查使用是否有充分理由。一个层次的成功能够为另一个层次提供问题与条件，却不能代替另一个层次的证据。本书因此形成的是一套相互约束的研究关系，而非用单一证明方式覆盖所有领域。

这也规定了完成二字的含义。本书完成了概念区分、条件反例与研究程序之间的接续，说明了为什么某些归属、收敛和外推推断需要改写；它并未完成一套跨领域机制的统一实证检验。后文各领域提供的材料和设计，使共同主张取得不同程度的规定性。哪些已经由既有资料支持，哪些仍属解释重读，哪些有待执行，必须继续按其真实身份使用。保留这个差别，正是让本书取得的工作能够被接续。

对理论最有力的检验，因而可能来自看似平常的结果。若成熟条件模型在目标清楚、操作稳定的情况下已经足够，发生分析可以解释何以能够有效简化；若单位重建没有改变任何比较，则需要收缩重建的用途；若对象澄清真正使原来不可比较的材料取得联系，它就产生了概念层面的增量，而无需伪称已经发现新因果定律。成果的价值来自它解决了哪一道困难，不来自被归入多高的理论层级。

本书将这些有限而相连的成果交给后续研究：一个属性判断应能说明其单位，一个单位应能说明其边界，一个边界应能面对改变它的材料。三者共同限制了任意归属，也限制了无限扩展的整体论。科学发生学由此取得一种积极形式：使对象的规定性能够在变化中被追踪，使新规定有理由地长成，并使这种生长接受已经取得知识的约束。

41.3 稳定与创造能够共同发生

如果属性需要稳定性，创造为什么不会破坏知识？答案在于稳定不等于所有层次都固定。一个研究体系可以保留共同目标和证据纪律，同时改变对象分类与操作路径；一项实践可以保持核心功能，同时重组承担功能的环节。稳定与创造能够在不同层次相互支持。稳定提供可承接材料，创造使材料在新条件中重新获得作用。

成熟的知识因此具有两种能力。一种是可靠使用：在明确条件下，简化规则能够支持判断而不必每次重做全部研究。另一种是有理由地展开：当出现持续失配，使用者能够回到条件、过程与历史，查明哪些部分需要修订。只有使用能力而没有展开能力，知识容易

僵化；只有不断展开而没有可靠压缩，知识又难以服务行动。二者之间的转换应成为科研基础设施的一部分。

一个实验标准可以说明这种关系。标准让不同团队形成可比较材料，是稳定的公共资源；当新装置或新样本进入，标准需要接受边界检验。若修改有清楚理由、版本与转换规则，旧资料并不必然失去价值，新资料也不必被强迫符合旧口径。创造不是毁掉共同标准，而是发展标准能够处理变化的能力。维护与创新由此成为相互联系的劳动。

稳定也可以保护探索。一个研究者知道失败结果不会自动导致工作被否定，便更有条件提出真正可能失败的预测；一个团队有可靠记录与复核安排，便能够尝试更大的概念变化。相反，表面上鼓励创新而缺少承接机制，可能使每次尝试都变成个人冒险，失败无法积累。创造条件不只在想法产生的瞬间，也在制度是否能够保存未成功尝试所揭示的知识。

探索同样需要停止规则。一个想法持续增加复杂性，却无法形成新问题、新预测或更可靠判断，应当缩小或放弃。停止不是对创造的背叛，而是让资源能够进入更有希望的方向。关键是保留停止理由，使后来者知道哪些路径已经尝试、何种条件下失败、什么新材料可能重新打开问题。失败的历史由此成为未来创造的条件，而不是必须隐藏的污点。

再生与复旧的差异也在这里显露。系统受到冲击以后，恢复原输出只是一个可能目标；更重要的是维持与修订能力是否获得新的组织。若靠增加隐形劳动暂时恢复成绩，问题可能仍在累积；若通过公开知识、改造交接和调整目标形成更合理结构，稳定便有了新的基础。发生学关注这种结构改变，同时要求比较其成本与后果，不能把任何改革都自动称为再生。

科学创造最终不是在稳定之外的偶发例外。它可以成为稳定体系内部有组织的能力：识别差异，保存未决，允许试作，比较竞争，修订规则。这样的体系既能让成果被继承，也能让成果不成为下一代问题的障碍。它所维持的不是每一个具体答案，而是知识能够继续形成、受检验并承担后果的公共条件。

41.4 五种证明责任如何共同工作

跨学科专著容易把概念洞见、数学证明、经验资料、研究设想与规范判断写成同一种语气。这样做可以制造统一的外观，却使读者难以判断每一主张究竟依靠什么。本书的研究纲领需要坚持不同证明责任。概念命题说明区分为何必要、彼此是否一致；数学结论说

明在明确假设下能推出什么；经验主张需要材料与方法；待检验机制需要区分性预测；规范选择则要公开价值理由与责任分配。

概念论证的力量在于改变问题的可理解形式。把同一行为外形与同一伦理事件区分，可以解释为什么重复失约不必仍然是同一个过犯；把个人表现与人机共同产出区分，可以避免评分含义混乱。这些区分不需要先等待一个统计显著结果才有意义，但它们应当接受案例与反例检查。若区分不能改善任何判断，或者只是给同一困难增加名称，其价值就需要重新评价。

数学证明提供更明确的边界。它能够展示某种一般推断不成立，也能够建立特定函数类或模型中的限制。证明的强度来自假设与结论的严格联系，因此不能把一个条件命题扩成对所有现实系统的普遍判决。耦合可收敛的模型否定耦合必发散的一般断言，却不证明每个具体系统都会收敛；同值预测的跨条件限制说明某类压缩必然失准，却不否定所有复杂编码。数学越精确，使用时越需要保留范围。

经验研究的困难则在于对象与操作之间的距离。一个指标可能可靠，却没有充分覆盖构念；一项操纵可能成功，却同时改变多个条件；一个模型拟合良好，却没有区分竞争机制。研究应把这些距离变成具体任务，而不是在最后加上一句局限便结束。测量检查、时序记录、干预对照和独立验证分别承担不同责任。没有哪一个单独指标可以一次性认证整个本体论。

待检验机制需要保持想象力。研究者可以提出尚无充分证据的历史通道或结构更新，只要明确它会带来哪些后果，以及什么材料将迫使修订。假说的价值在于组织探索，而不在于提前借用经验结论的确定性。若所有结果都能被解释为支持，假说就失去研究作用。强假说应当敢于让某些可能结果与自己不相容，同时允许失败落到具体机制层次。

规范判断则不能从机制图中自动生长出来。一个制度能够稳定地产生高分，不代表它公正；一个关系持续变化，不代表它更有价值；一个 AI 系统提高效率，也不自动决定应当怎样分配责任。事实为规范提供对象和后果，规范还需要理由、权利、代价与可选择空间。把两者联系起来，是为了让价值判断更充分，而不是让统计关系代替价值判断。

五种责任之间并非隔着高墙。概念重建引导模型，模型提出经验问题，经验结果迫使概念修订，规范选择决定哪些后果值得研究。它们可以共同工作，但每一次跨越都应说明依据。研究纲领的成熟不在于只采用一种证明方式，而在于不同方式能够互相支持又不彼此冒领。这种分工使哲学创造获得可靠落点，也使领域研究能够反向限制哲学。

不同证明责任还决定反例怎样进入理论。一个数学反例可以立即否定一个不加条件的普遍命题，却未必决定某个经验机制在现实中的频率；一个经验失败可以限制具体模型，却未必否定较宽的概念区分；一个规范反对可以要求停止某种使用，即使预测仍然准确。把反例的作用层次说明清楚，既避免理论无限免疫，也避免因局部失败而放弃所有已获知识。真正可修订的框架，需要能够告诉读者什么失败将改变什么。

证明责任也与表达语气相连。尚未完成的试验应以研究设计呈现，模型中的结果应连同假设说明，经过核验的事实应有清楚来源。确定语气不是学术力量的替代品。读者能够准确知道一项判断的身份，才有机会继续验证、应用或反驳它。这样的语言纪律让创造性主张保留锋芒，同时不提前借用尚未取得的证据。

41.5 知识的历史需要怎样的基础设施

本书把知识继承理解为判断条件的接续。后来者读到最终图表，可以知道研究得到了什么，却未必知道这个结果为什么曾被接受、什么变化会使它失效。若这些关系断开，复制步骤仍可能复制误用；若关键关系保存下来，即使新任务必须改变步骤，已有知识仍可发挥作用。第 39 章讨论了相应的发表、复核和基础设施安排。结语所强调的更基本要求是：继承一个结论，应当同时取得辨认其权限与修订它的理由，而不仅取得重复它的字句。

这并不要求每项工作保存无穷细节。最有价值的是影响推断的版本与转折。某项变量定义改变，某次装置校准改变解释，某个异常促使样本重新分类，某种竞争模型使原预测失去区分力，都应留下清楚记录。研究者可以用简洁日志说明改变的理由，并保持返回必要原始材料的路径。资料量本身不是完整性的尺度，推断关系能否被重建才是。

版本记录尤其重要。一个结论在当时资料下可能合理，后来随着新证据而被修订。若只保留最后版本，后来者容易误以为真相从来清楚，科学进步只是把错误的人替换掉。若保留版本与依据，就能看到知识怎样在不完整条件下形成，哪些判断经受住了变化，哪些被新材料迫使调整。这种历史不是为错误开脱，而是使错误能够产生可继承的经验。

失败档案也应具有明确结构。失败可能来自操纵没有落实、指标不能区分、模型预测不符、迁移效果消失，或应用后果与目标冲突。把这些情况都写成无显著结果，会丢失机制信息。档案应说明原先预期、实际操作、观察范围和可支持结论。一个无法确认的结果有时可以缩小参数范围，有时只能说明设计不足，二者应分别保存。失败的公共价值取决于这种精确性。

复核可以围绕关键推断进行，而不必每次重演全部项目。另一位研究者能否根据材料识别同样对象，能否理解分类边界，能否在未看结论时作出相容判断，能否重现重要计算，都是可分工的复核任务。不同复核覆盖不同风险，不能用一次技术重跑代替全部审查。发生记录有助于选择最值得检查的环节，使公共验证既严格又可承担。

知识基础设施还要允许异议进入。受影响者、执行者与不同领域研究者可能看到中心分析忽略的条件。异议需要被组织成可以处理的问题：指出哪一结论、涉及何种材料、要求什么修改。公开入口并不意味着每项异议都必须改变结论，而是意味着它们能够得到有理由的回应。若异常只能存在于私人抱怨中，知识就难以获得真正的反向约束。

AI 可以辅助这一基础设施，帮助检索版本、比较定义、提示矛盾或整理未决问题。但辅助输出也需要来源与不确定性，不能因为处理规模更大就获得自动裁决权。系统可以指出两份记录用词不同，是否代表对象不同仍需判断；可以发现数值异常，异常是否有意义仍需解释。人机合作的目标是提高可检查能力，而不是用一个新的黑箱替代旧的判断过程。

基础设施的价值最终要通过实际工作检验。它是否减少重复错误，是否帮助定位迁移失败，是否让新团队更快理解对象，是否在成本可接受的情况下提高复核效率。若某些字段从未进入任何判断，就应考虑删减；若某些关键联系反复缺失，就应补充。知识基础设施自身也是发生系统，它会改变研究行为，也需要根据后果修订。把这一点纳入设计，才能避免记录制度变成只为自身扩张服务的负担。

基础设施也需要考虑资料生命周期。某些原始记录在完成复核后不再需要长期保存，某些关键版本则应长期关联结论。保存期限、访问权限和删除条件应与实际用途相称，不能以未来或许有用为理由无限积累个人材料。对公共结论而言，至少应保留足以理解推断的结构；对敏感材料而言，可以通过适当的替代记录与受限复核维护检查能力。发生的历史需要被承担，但承担并不等于毫无边界地占有历史。

知识继承也不能把不同材料强制压成同一种成果。数据保存可再分析的差异，程序保存可复走的操作，概念论证保存区分的理由，失败说明保存某条路径曾受到的限制。它们各自不完整，却可以互相定位。共同体需要维持的，是这些材料之间仍能查验的联系：哪个判断使用了哪些材料，哪项修改改变了哪个范围，哪些未决没有随发表而消失。这样的继承允许新团队改变工具与表达，同时知道自己仍承接了什么，已经放弃了什么。

41.6 继续研究的成立条件与退出责任

本书的后续责任，应按主张真正得到推进的层次安排。第 40 章已提出研究纲领的项目组织；在这里，需要确认这些项目合在一起何时算作理论进步。若只在更多领域发现可以改写为 SDE 词汇的叙述，覆盖面扩大尚未构成增量。继续研究必须使至少一项原来未被解决的困难得到更清楚处理，并说明这种处理依靠共同框架中的哪一种工作。

一种进步来自对象重建。研究者发现原先名为同一属性的资料混合了不同任务，重新界定单位后，旧分歧获得可复核的解释。这种贡献的成立条件，是新边界可以由其他研究者据材料使用，并在未被用来建界的案例上继续作出有用区分。若边界只为容纳每一次反例而不断改写，重建便失去约束。退出责任不是立即放弃全部发生论，而是撤回这一个缺乏区分力的对象方案。

另一种进步来自机制识别。构成分析提示一条可能的回写或历史通道，随后通过适当比较表明，这条通道确实改变了相关后果。其强度取决于竞争解释被排除到什么程度，而不取决于机制是否被命名为新型。若常规模型已经包含相同关系，新框架可以承认自己帮助发现或组织了问题，却不冒领机制原创性。若关键操作已充分实施、测量确能捕捉预期后果，且资料精度足以排除预先规定的效应范围，或结果稳定地违背机制的明确预测，就应修改相应假说。精度不足或操作未能到位时，结论仍应保持未决，并指出需要补足的环节；也不能仅为保住假说，将已受约束的后果转移到无法检验的更隐蔽层次。

还有一种进步属于研究组织。更合适的单位记录可能让团队较快定位分歧，更清楚的转向理由可能减少重复无效试验，更完整的失败承接可能使新团队免于重走旧路。这些收益值得研究，但它们需要与维护成本共同评价。若采用联合规程的团队只是写了更多文件，关键判断并未改善，方法就应被简化。组织增益本身可以成立，无需包装成新自然规律；它也不能因难以归结为单一指标而免于比较。

三种进步之间可以接续。对象重建可能暴露新的机制问题，机制成果可能要求改变研究流程，流程改进又可能使原来无法获得的对象资料出现。接续需要保存每一步新增了什么，尤其避免把同一材料在三个层次中反复计为独立成功。理论成长的证据，应当能显示知识怎样增加约束，而不只是怎样扩大叙述。这样的要求让有限项目有机会共同构成纲领，同时阻止局部收获被无限放大。

退出责任也由此有了分层含义。测量尚不可靠时，先收回基于它的经验断言；机制未获支持时，保留可靠结果并撤回机制归因；某一领域长期没有额外增益时，缩小适用承

诺；如果不同案例反复显示核心区分无法一致使用，则必须返回概念本身。每一种退出都应交代哪些知识仍然可用，哪些使用必须停止。退出的范围与证据的作用范围相称，才既能防止无限自我保护，也能避免把局部失败扩大为全部知识作废。

继续与退出还涉及实际承担者。一个方案理论上可以修订，未必有人拥有修订所需的资料、时间和权限。若使用者只能承受结论，却不能把新后果带回研究，开放承诺就难以兑现。后续项目应为关键判断指定接续安排，使失败能够找到应当回应的位置。这项责任不替代具体制度的规范论证，却是本书关于知识参与发生的主张对自身提出的实际要求。

因此，纲领的未来不由其能够解释多少旧案例单独决定，而由它能否使新研究获得更明确的对象、更受约束的机制或更可靠的共同工作决定。成功时，说明成立条件；受限时，留下未决；失败时，改变后续使用。三种结果都可以成为知识，但只有它们实际改变研究时，发生学才承担了自己提出的历史责任。

41.7 客观性 责任与共同世界

承担发生的历史，意味着承认研究者既受到世界约束，也通过操作、分类和制度使用参与后续世界。这个双重位置不会使客观性消失，却改变了客观性的实践要求。研究者需要公开自己做了什么，区分发现的关系与造成的变化，并让判断能够被不同位置上的人检查。客观性不只是结论与对象之间的静态关系，也是一套使结论持续接受纠正的公共活动。

共同世界并非所有人必须得到相同读数。不同位置可以产生不同但可转换的描述，不同任务也可以合理突出不同属性。共同性在于这些差异能够被说明，冲突能够被定位，转换能够被检验。若两种判断无法直接合并，研究者可以保留分歧并追踪其条件，而不必立刻判定一方纯属主观。发生学使客观讨论容纳位置差异，同时要求位置不能成为免于证据的理由。

责任也需要沿发生分配。一个错误结论可能涉及材料采集、模型选择、解释扩张和制度使用，不能总由最末端执行者承担全部后果；共同形成又不能成为人人免责的借口。应当追踪谁有能力改变哪项条件，谁知道或应当知道什么限制，哪些决定怎样进入结果。这样的分析使责任更具体，也使修复能够针对实际通道，而不是只寻找象征性承担者。

在教育和伦理领域，参与者对自身事件的解释应获得真实位置。研究者可以提出机制编码，却不能把编码直接当作当事人全部意义。一个稳定边界可能保护自主，一次关系改

变可能是合理修复，也可能带来新的损害。规范评价需要听取理由、比较后果和尊重选择空间。数据有助于澄清，却不能取消人们参与解释自身处境的权利。

制度中的共同标准同样需要这种双重要求。没有共同标准，资源分配和协作可能失去可预期性；只有不可修订的标准，实际差异又可能被长期压在例外之中。一个负责的制度应当说明指标用途、适用范围、申诉入口与修订条件。规则可以稳定，但稳定应有维护理由；规则可以改变，但改变应有证据与过渡安排。发生学在这里连接认识责任与治理责任，而不把二者混为同一技术问题。

科研自由也与后果承接相关。探索需要容许错误和未完成想法，但这不意味着可以把未经验证的主张直接投入高后果使用。概念讨论、实验试点和制度应用应有不同责任边界。越接近实际影响，越需要明确证据、风险和纠错安排。这样的分层既保护创造，也保护受影响者，使研究不必在完全禁止与无条件放任之间选择。

共同世界的维护还包括知识可退出私人占有。一个方法如果只有创立者能够解释，公共性有限；一份记录如果没有转换路径，难以被别人继承；一个系统如果不允许导出证据与版本，使用者可能被锁在单一工具中。发生学所追求的知识应当可以被接续、质疑和改造。公共性不是抹去贡献者，而是让贡献能够成为他人继续工作的条件。

这种客观性与责任观保留了科学的雄心。我们仍然可以发现不受偏好支配的关系，建立可靠模型，设计有效工具，也可以对错误作出明确判断。不同之处在于，判断不再假装没有形成过程，成功不再掩盖维护条件，应用不再切断后果返回。科学越深地进入世界，就越需要这些能够让知识继续受约束的公共结构。

41.8 让理论也拥有可修订的历史

本书提出的理论应当接受它要求其他研究承担的责任。属性纠缠、号位侧重、读数事件与联合方法，都是需要在具体工作中继续获得规定的概念。它们不能因为能够解释很多现象，就被视为已经证明普遍有效；解释范围广，也可能意味着区分力不足。未来研究需要不断检查这些概念是否改变了问题、证据和行动，是否帮助发现既有方法遗漏的条件，以及是否能够接受真正不利的结果。

一些核心问题仍然开放。属性同一性的条件如何在不同尺度上建立，侧重稳定性是否能够形成更统一的形式表达，结构回写怎样与普通参数更新区分，跨领域转换可以达到何种程度，都需要进一步研究。开放并不意味着任意。每个问题可以被拆成更具体的任务：

构造反例、比较操作、建立模型、追踪历史或检验迁移。理论的未来由这些任务推进，而不是由不断增加宏观宣言推进。

一个尚未完成的核心问题，是属性同一性与成熟判据之间如何建立更严格联系。第 4 章区分构成、认识与记录三层成熟，第 5 章又显示相同表现可以由不同维持方式产生。后续研究因而需要判断：维持方式改变到什么程度，仍可称为同一属性的延续；何时应改写属性范围；何时则需要承认新的对象。这不是再给成熟增加一个总分，而是检验同一性理由怎样经受时间与扰动。它将决定发生学能否在承认历史的同时，保留可积累的对象。

科学史的发生学重读同样需要继续深化。哲学解释可以为已有变革提出新问题，但必须与原始文本、实验条件和数学关系相互限制。不能把历史写成几个号位依次归来的简单脚本，也不能因为旧叙事有缺口，就把所有新解释都当作同样可信。更有价值的是追踪具体问题如何形成，装置如何改变可观察对象，数学如何重新组织差异，共同体如何接受或抵抗新的规定。这些历史工作可以丰富发生学，也可能迫使它改变自己的概念。

另一项开放问题在于承接结构的可替换性。若一段历史原先由个人经验保存，后来由档案或人工系统支持，什么条件下判断仍能延续，什么条件下只是复制了输出？这涉及历史被保存、被调用、被解释和进入行动的不同环节。全书的人机分析已把这些环节分开，后续还需要检验它们之间可以替代到何种程度。答案将反过来限制关于主体位置和共同能力的理解：相同产出能否被共同归属，需要根据实际承接与责任重新论证。

理论的修订应当留下可追踪版本。某个概念被缩小范围，某种强断言被反例推翻，某条方法因无增益而简化，都应成为公开成果的一部分。这样的历史不会削弱理论的学术身份，反而使其能够与实际研究共同成长。最危险的并不是曾经提出错误，而是错误无法改变后续使用。一个发生学如果拒绝自身发生的记录，就与自己的中心主张相冲突。

理论还需要接受来自简化成功的挑战。当一个简单模型在广泛条件下持续可靠，新框架应当解释这种成功，而不是假定复杂过程必然被错误忽略。也许某种稳定结构使许多细节在当前任务中无关，也许共同校准已经有效吸收条件差异，也许目标本来就只要求一个聚合量。承认这些可能，使发生学能够与科学已有的简洁成果形成建设性联系。复杂性只有在改变推断与行动时，才需要进入主要模型。

反过来，一个复杂模型能够拟合全部历史，也不保证它已经认识发生。模型可能依赖过多自由度，可能利用事后信息，可能只是在每次失败后增加条件。未来研究应当让关键预测先于用于检验的新资料，并保留无法解释的结果。这种时间约束能够限制事后调适，

但独立检验还可利用事先保留的历史材料、独立重做的分析和未参与模型修订的检验程序。关键在于明确材料、假设与检验选择怎样同建模过程分开，以及哪些失败会约束结论；资料是否刚刚出现并非唯一判据。历史被保存，是为了支持继续判断，而不是为任何既成解释寻找足够多的装饰。

最终值得保持的，是一种科学态度和组织能力。面对稳定，追问它靠什么维持；面对差异，追问它怎样形成；面对新概念，追问它增加什么区分；面对成功，说明适用范围；面对失败，让后果返回判断；面对新的可能，创造能够检验它的条件。这些动作并不属于某个学科独有，却可以在每个学科中取得具体而严格的形式。

这种组织能力也可以从小处建立。一份结果附上适用范围，一次模型修改保留理由，一项评价允许当事人补充条件，一次失败留下可复核材料，都可能改变后来研究的起点。它们不需要等待全新典范获得普遍承认，却能够逐步形成典范所需的实践基础。宏观理论只有进入这些具体动作，才会成为共同工作的能力；具体动作也因理论的组织，获得超出单次改进的方向。

属性的发生不会因为一本书结束而停止，知识的发生也不会。一个结果进入新的问题，一项规则进入新的行动，一次错误成为新的警觉，一段合作留下可以继承的判断方式。科学研究承担发生的历史，就是在这种连续过程中保持可追踪、可质疑、可创造的联系。它不保证世界永远符合我们的模型，也不要求我们放弃形成模型的努力；它使我们能够在世界改变时，知道哪些知识可以继续依赖，哪些知识必须重新长成。

参考资料

以下资料按正文首次引用顺序编排。来源说明用于标明相关研究的支持范围；本书提出的进一步解释和研究设计由相应正文承担论证。

[1] 王德生 2026 大论文 与 WDS 的对话 累积研究底稿

作者提供的 71 页未刊研究底稿。引用范围为属性侧重、SDE 发生与科研纲领的概念来源；其中 AI 生成的经验断言不作为外部事实证据。

[2] 王德生 2026 属性纠缠与不收敛病理学 从测不准原理到误差项的通用语法

作者提供的 30 页未刊论文。用于交代读数事件、误差研究、回写与条件稳定的研究母题，不以底稿替代领域事实核验。

[3] Joint Committee for Guides in Metrology 2012 International vocabulary of metrology—Basic and general concepts and associated terms (VIM), 3rd edition

https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf

支持被测量、测量不确定度、计量可比性与兼容性的区分；具体术语 2.3、2.26、2.46、2.47。

[4] David B. Newell; Eite Tiesinga (eds.) 2019 The International System of Units (SI), 2019 Edition

https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.330-2019.pdf

七个定义常数数值精确；单位实际实现仍有不确定度。

[5] Albert Einstein 1905 Zur Elektrodynamik bewegter Körper / On the Electrodynamics of Moving Bodies

https://users.physics.ox.ac.uk/~rtaylor/teaching/specrel.pdf

原文英译；共动测长、同时性及参考系变换。章内简化推演与 SDE 解释为本书分析，不声称爱因斯坦主张属性纠缠论。

[6] LIGO Scientific Collaboration; Virgo Collaboration 2016 Observation of Gravitational Waves from a Binary Black Hole Merger

https://arxiv.org/abs/1602.03837

2015 年事件、2016 年发表；支持双黑洞并合波形与直接探测。

[7] Joint Committee for Guides in Metrology 2008 Evaluation of measurement data—Guide to the expression of uncertainty in measurement

https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf

4.2.7 讨论时间相关观测，5.2 讨论相关输入与协方差；不将误差模型简化为独立白噪声。

[8] H. P. Robertson 1929 The Uncertainty Principle

https://journals.aps.org/pr/abstract/10.1103/PhysRev.34.163

提出可观测量在同一量子态中的标准差关系；关于测量误差与扰动的区分，另参见 Rozema 等的实验报告。

[9] Lee A. Rozema et al. 2012 Violation of Heisenberg’s Measurement-Disturbance Relationship by Weak Measurements

https://arxiv.org/pdf/1208.0034

区分态标准差关系与误差—扰动关系；弱测量不推翻 Robertson 关系。

[10] Wojciech H. Zurek 2002 Decoherence and the Transition from Quantum to Classical—Revisited

https://arxiv.org/pdf/quant-ph/0306072

Los Alamos Science 27，2002；2003 年 arXiv 上传。支持环境关联、退相干及条件依赖的退相干尺度；不将模型估算冒充实测。

[11] J. S. Bell 1964 On the Einstein Podolsky Rosen paradox

https://csiflabs.cs.ucdavis.edu/~gusfield/Bell-paper.pdf

原文明确局域性条件；不排除全部隐变量理论。

[12] 王德生 2026 大论文及属性纠缠与不收敛病理学研究底稿

作者提供的未刊研究底稿

仅作为作者问题发生、概念演变与原始命题的来源，不作为外部科学事实的证明。

[13] ATLAS Collaboration 2012 Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC

https://arxiv.org/abs/1207.7214

支持多衰变道搜索、新中性玻色子证据及与标准模型 Higgs 产生衰变相容；不把单次碰撞等于直接看见粒子。

[14] Particle Data Group 2024 Electroweak Model and Constraints on New Physics

https://pdg.lbl.gov/2024/reviews/rpp2024-rev-standard-model.pdf

支持能标依赖的电磁耦合及标准模型量的比较；能标跑动不等于时间漂移。

[15] Peter J. Mohr; David B. Newell; Barry N. Taylor; Eite Tiesinga 2025 CODATA recommended values of the fundamental physical constants: 2022

https://physics.nist.gov/cuu/pdf/JPCRD2022CODATA.pdf

2022 是调整年份，2025 为出版年份；存在输入张力但不能据此证明参数本体不收敛。

[16] R. Lange et al. 2021 Improved Limits for Violations of Local Position Invariance from Atomic Clock Comparisons

https://arxiv.org/abs/2010.06620

原子钟比较约束精细结构常数时间变化，与能标跑动属于不同检验。

[17] Steven Weinberg 2009 Effective Field Theory, Past and Future

https://arxiv.org/abs/0908.1964

作者回顾有效场论发展；支持有界尺度描述与更一般理论关系的讨论，不把有效性误写为任意性。

[18] R. Bellman; K. J. Åström 1970 On Structural Identifiability

https://portal.research.lu.se/en/publications/on-structural-identifiability/

结构可辨识性研究，讨论输入输出测量对内部结构的识别及所需实验条件。

[19] Thomas J. Rothenberg 1971 Identification in Parametric Models

https://doi.org/10.2307/1913267

有限维参数概率模型的可辨识性研究；仅引用其问题框架，不将特定正则条件下的局部结论泛化为所有模型。

[20] Robert Hermann; Arthur J. Krener 1977 Nonlinear Controllability and Observability

https://www.math.ucdavis.edu/~krener/1-25/10.IEEETAC77.pdf

动态系统的可观测性须研究状态与观测轨迹的对应，不能由瞬时输出通道数独自裁决。

[21] J. A. Hartigan; P. M. Hartigan 1985 The Dip Test of Unimodality

https://doi.org/10.1214/aos/1176346577

原论文检验对象为单峰性；拒绝单峰不等于证明恰有两个峰或两个稳定状态。

[22] NIST/SEMATECH 在线手册，核验于 2026-10-15 NIST/SEMATECH e-Handbook of Statistical Methods, 1.3.3.1 Autocorrelation Plot

https://www.itl.nist.gov/div898/handbook/eda/section3/autocopl.htm

官方统计手册说明自相关图、时间依赖及样本量相关的近似置信带；自相关不能单独识别具体回写机制。

[23] A. Orphanides 2001 Monetary Policy Rules Based on Real-Time Data

https://www.aeaweb.org/articles?id=10.1257/aer.91.4.964

原刊摘要支持实时资料与事后修订资料可能产生不同政策建议，作为双时间轴资料设计的来源。

[24] Gideon Schwarz 1978 Estimating the Dimension of a Model

https://doi.org/10.1214/aos/1176344136

在指定模型与抽样假设下推导带复杂度惩罚的模型选择准则；不作为独立测试误差或普适机制门槛。

[25] Federal Reserve Bank of St. Louis 在线资料 ALFRED Help

https://alfred.stlouisfed.org/help

官方说明历史数据版本与当时可用信息的取得方式；不代表本书已完成实时回测。

[26] Green, B. G.; Lim, J.; Osterhoff, F.; Blacher, K.; Nachtigal, D. 2010 Taste mixture interactions: suppression, additivity, and the predominance of sweetness

https://pubmed.ncbi.nlm.nih.gov/20800076/

原始感官实验；支持所研究配方与程序中的混合抑制及混合内部感知强度加和，不支持酸必然改变甜峰时序。

[27] Schifferstein, H. N. J. 1994 Sweetness suppression in fructose/citric acid mixtures: a study of contextual effects

https://pubmed.ncbi.nlm.nih.gov/7971123/

原始心理物理实验；区分情境对内部表征及反应选择的影响；不用于宣称所有情境必改变抑制。

[28] Woods, A. T.; Lloyd, D. M.; Kuenzel, J.; Poliakoff, E.; Dijksterhuis, G. B.; Thomas, A. 2011 Expected taste intensity affects response to sweet drinks in primary taste cortex

https://research-portal.uu.nl/en/publications/expected-taste-intensity-affects-response-to-sweet-drinks-in-prim/

作者机构原始研究摘要；支持特定期待操纵下甜感报告和味觉皮层活动改变，不证明完整意识或 SDE 普遍模型。

[29] Elowitz, M. B.; Levine, A. J.; Siggia, E. D.; Swain, P. S. 2002 Stochastic gene expression in a single cell

https://pubmed.ncbi.nlm.nih.gov/12183631/

原始双报告细胞实验；内源和外源变异均贡献基因表达差异；不把全部变异解释成回写。

[30] Acar, M.; Becskei, A.; van Oudenaarden, A. 2005 Enhancement of cellular memory by reducing stochastic transitions

https://www.nature.com/articles/nature03524

酵母半乳糖调控原始研究；不同反馈环对稳定状态和历史保持作用不同，支持具体结构而非泛称有反馈即有记忆。

[31] Warmflash, A.; Sorre, B.; Etoc, F.; Siggia, E. D.; Brivanlou, A. H. 2014 A method to recapitulate early embryonic spatial patterning in human embryonic stem cells

https://pmc.ncbi.nlm.nih.gov/articles/PMC4341966/

体外微图案化细胞培养原始研究；几何约束和信号共同参与空间分化图式，不将体外图式等同完整胚胎。

[32] Kiepas, A.; Voorand, E.; Mubaid, F.; Siegel, P. M.; Brown, C. M. 2020 Optimizing live-cell fluorescence imaging conditions to minimize phototoxicity

https://journals.biologists.com/jcs/article/133/4/jcs242834/224018/Optimizing-live-cell-fluorescence-imaging

原始方法研究；说明照明开销可造成光毒性和漂白，支持把观察动作写入实验条件，不提供一般照明安全阈值。

[33] Womelsdorf, T.; Anton-Erxleben, K.; Pieper, F.; Treue, S. 2006 Dynamic shifts of visual receptive fields in cortical area MT by spatial attention

https://www.nature.com/articles/nn1748

猕猴 MT 区域原始研究；支持特定任务中注意相关感受野位移和小幅收缩，不普遍化至所有神经元。

[34] Sadtler, P. T.; Quick, K. M.; Golub, M. D.; Chase, S. M.; Ryu, S. I.; Tyler-Kabara, E. C.; Yu, B. M.; Batista, A. P. 2014 Neural constraints on learning

https://pmc.ncbi.nlm.nih.gov/articles/PMC4393644/

脑机接口原始实验；既有群体活动结构对所研究短期学习施加约束，不推出不可改变的能力本质。

[35] Yi, T.-M.; Huang, Y.; Simon, M. I.; Doyle, J. 2000 Robust perfect adaptation in bacterial chemotaxis through integral feedback control

https://pmc.ncbi.nlm.nih.gov/articles/PMC18287/

PNAS 97(9), 4649–4653。在明确模型条件下分析积分反馈与鲁棒适应。

[36] Turrigiano, G. G.; Leslie, K. R.; Desai, N. S.; Rutherford, L. C.; Nelson, S. B. 1998 Activity-dependent scaling of quantal amplitude in neocortical neurons

https://www.nature.com/articles/36103

培养新皮层神经元原始实验；支持活动相关突触强度调整与稳定性机制，不作为临床病因或疗效证据。

[37] National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1979 The Belmont Report

https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html

官方原始规范文本；尊重人、有利及公正作为研究伦理原则；引用历史文本，不冒充现行各地法律咨询。

[38] Greg Kaplan; Giovanni L. Violante 2022 The Marginal Propensity to Consume in Heterogeneous Agent Models

https://doi.org/10.1146/annurev-economics-080217-053444

系统比较偏好、事前异质性、收入过程和资产结构对 MPC 的影响；流动与非流动资产、手停口停家庭的比例和类型已是解释重点，构成本书必须面对的强基线。

[39] Michael McLeay; Amar Radia; Ryland Thomas 2014 Money creation in the modern economy

https://www.bankofengland.co.uk/quarterly-bulletin/2014/q1/money-creation-in-the-modern-economy

现代银行贷款通常创造存款；银行不只是转贷既有存款，也不简单按固定倍数扩大央行货币。该文同时讨论货币政策等约束，不能用来主张银行信用创造没有限制。

[40] Douglas Staiger; James H. Stock; Mark W. Watson 1996 How Precise Are Estimates of the Natural Rate of Unemployment?

https://www.nber.org/papers/w5477

NBER 工作论文 5477，1997 年收入专书；考察固定、时变、随机游走及基本面模型下的 NAIRU 精度。美国 1990 年的一个典型 95%区间为 5.1%至 7.7%；该结果限定于具体研究，不能视为普遍区间。

[41] Thomas Piketty; Gabriel Zucman 2014 Capital Is Back: Wealth-Income Ratios in Rich Countries 1700–2010

https://gabriel-zucman.eu/files/PikettyZucman2014QJE.pdf

研究富裕国家财富收入比的历史变化。财富口径不可直接等同于生产设备资本；必须区分资产估值、土地住房及生产资本，不能拿财富收入比直接反驳生产资本产出比命题。

[42] John G. Fernald 2014 A Quarterly, Utilization-Adjusted Series on Total Factor Productivity

https://www.frbsf.org/research-and-insights/publications/working-papers/2014/04/a-quarterly-utilization-adjusted-series-on-total-factor-productivity/

FRBSF 工作论文 2012-19 的 2014 年修订版；季度增长核算已调整劳动努力和资本使用时长造成的利用率变化，说明残差混合机制是既有研究问题。

[43] Daniel Kahneman; Amos Tversky 1979 Prospect Theory: An Analysis of Decision under Risk

https://web.mit.edu/curhan/www/docs/Articles/15341_Readings/Behavioral_Decision_Theory/Kahneman_Tversky_1979_Prospect_theory.pdf

收益损失域、参考点、确定性效应、隔离效应与决策权重影响风险选择；选择差异不能全部归于单一效用曲率参数变化。该理论须作为本书风险决策研究的比较基线。

[44] Robert E. Lucas, Jr. 1976 Econometric Policy Evaluation: A Critique

https://www.landsburg.org/LucasCritique.pdf

行为决策规则可能随政策制度与预期改变，历史拟合良好不保证政策反事实外推有效。原文第 24 至 25 页讨论一般机制，第 26 至 29 页讨论消费；不能把政策会改变参数当作本书首次发现。

[45] Sanford J. Grossman; Joseph E. Stiglitz 1980 On the Impossibility of Informationally Efficient Markets

https://www.pims.math.ca/files/Grossman_Stiglitz1980.pdf

原论文讨论信息获取成本、价格部分揭示与内生知情者人数，并建立相应均衡；其不可能性不可泛化为所有反馈系统无均衡。

[46] Eugene F. Fama 1991 Efficient Capital Markets: II

https://doi.org/10.1111/j.1540-6261.1991.tb04636.x

支持效率检验依赖资产定价模型、预期收益时变及可预测性的区分，不支持所有市场均有效。

[47] Alex Edmans; Itay Goldstein; Wei Jiang 2012 The Real Effects of Financial Markets: The Impact of Prices on Takeovers

https://finance.wharton.upenn.edu/~itayg/Files/takeoverfeedback-published.pdf

区分价格触发收购与收购预期进入价格，并使用共同基金赎回工具变量识别；不泛化为所有价格变化改变基本面。

[48] Eric Budish; Peter Cramton; John Shim 2015 The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response

https://ericbudish.org/publication/the-high-frequency-trading-arms-race-frequent-batch-auctions-as-a-market-design-response/

作者原论文及说明，支持将速度竞争与连续订单处理视为市场设计问题，提出频繁批量竞价；不主张已证实适用于所有市场。

[49] U.S. Bureau of Economic Analysis 2023 Gross domestic product (GDP)

https://www.bea.gov/help/glossary/gross-domestic-product-gdp

美国经济分析局的 GDP 定义说明，支持最终产品和服务，以及生产价值扣除生产中耗用产品与服务的增加值口径。第 24 章加工链金额为本书构造的核算例子，不是经验数据。

[50] U.S. Bureau of Economic Analysis 2018 Production boundary

https://www.bea.gov/help/glossary/production-boundary

官方生产边界定义，支持政府、非营利、自有住房、部分自用生产纳入与部分无偿家庭劳动排除。

[51] Office for National Statistics 2025 An introduction to the UK National Accounts, The Blue Book: 2025

https://www.ons.gov.uk/economy/grossdomesticproductgdp/compendium/unitedkingdomnationalaccountsthebluebook/2025/anintroductiontotheuknationalaccounts

第 11 节列明生产边界和推算估价，第 12 节说明估价；不以 GDP 代表完整福利。

[52] UNESCO 2005 Universal Declaration on Bioethics and Human Rights

https://www.unesco.org/en/legal-affairs/universal-declaration-bioethics-and-human-rights

官方原始规范文本；尊严、自主、脆弱性、隐私及文化多样性的权利边界。用于规范对话，非经验定理。

[53] Henry L. Roediger III; Jeffrey D. Karpicke 2006 Test-enhanced learning: Taking memory tests improves long-term retention

https://pubmed.ncbi.nlm.nih.gov/16507066/

原始两项实验比较阅读与回忆测试在短延迟和延迟保持上的不同效果；仅支持具体材料与任务中测试可促进保持，不等于考试普遍优于教学。

[54] Hamsa Bastani; Osbert Bastani; Alp Sungu; Haosen Ge; Özge Kabakcı; Rei Mariman 2025 Generative AI without guardrails can harm learning: Evidence from high school mathematics

https://pmc.ncbi.nlm.nih.gov/articles/PMC12232635/

高中数学现场随机试验，区分工具辅助练习与撤除工具后的独立测验，并比较不同教学约束；结论限于该研究情境和观察期限。

[55] H. Paul Grice 1975 Logic and Conversation

https://web.stanford.edu/class/psych205/papers/Grice-1975.pdf

原始论文，支持所说内容与会话含义及依赖交谈目的的推论区分；不作为所有交流必然合作的经验定律。

[56] Herbert H. Clark; Susan E. Brennan 1991 Grounding in Communication

https://web.stanford.edu/~clark/1990s/Clark,%20H.H.%20_%20Brennan,%20S.E.%20_Grounding%20in%20communication_%201991.pdf

共同理解的协作建立，以当前目的所需的理解证据为标准；媒介与沟通目的影响确认过程，不等于理解就是意见一致。

[57] Speckesser, S.; Runge, J.; Foliano, F.; et al. 2018 Embedding Formative Assessment: Evaluation Report and Executive Summary

https://discovery.ucl.ac.uk/id/eprint/10052334/

原始随机对照项目评估；支持形成性评价项目的具体组织及总体成绩结果，不代表所有反馈方法。

[58] Tabassi, E. / National Institute of Standards and Technology 2023 Artificial Intelligence Risk Management Framework (AI RMF 1.0)

https://doi.org/10.6028/NIST.AI.100-1

官方自愿性风险管理框架；支持治理、情境识别、测量和管理等职责组织，不作为法定义务。

[59] Ostrom, E. 2009 Beyond Markets and States: Polycentric Governance of Complex Economic Systems

https://www.nobelprize.org/prizes/economic-sciences/2009/ostrom/lecture/

作者诺贝尔演讲；支持多中心治理和制度情境差异的研究取向，不直接证明教育移植效果。

[60] Yao, S.; Zhao, J.; Yu, D.; et al. 2023 ReAct: Synergizing Reasoning and Acting in Language Models

https://arxiv.org/abs/2210.03629

ICLR 2023 论文；支持交错组织推理文本、行动与环境观察的系统方法；不据此推断意识。

[61] Lewis, P.; Perez, E.; Piktus, A.; et al. 2020 Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks

https://arxiv.org/abs/2005.11401

原始 RAG 论文；支持参数模型与外部可检索记忆组合的技术事实。

[62] Park, J. S.; O’ Brien, J.; Cai, C. J.; et al. 2023 Generative Agents: Interactive Simulacra of Human Behavior

https://arxiv.org/abs/2304.03442

原始研究；支持自然语言经验存储、反思与检索规划的架构存在，不证明真实人格或意识。

[63] Boiko, D. A.; MacKnight, R.; Kline, B.; Gomes, G. 2023 Autonomous chemical research with large language models

https://doi.org/10.1038/s41586-023-06792-0

原始研究；支持 Coscientist 在限定化学任务中连接检索、工具和实验执行，不推论无人监督的一般科研。

[64] Packer, C.; Wooders, S.; Lin, K.; et al. 2023 MemGPT: Towards LLMs as Operating Systems

https://arxiv.org/abs/2310.08560

2023 首发预印本及后续修订；支持分层记忆与上下文管理架构。

[65] Shinn, N.; Cassano, F.; Gopinath, A.; Narasimhan, K.; Yao, S. 2023 Reflexion: Language Agents with Verbal Reinforcement Learning

https://proceedings.neurips.cc/paper_files/paper/2023/hash/1b44b878bb782e6954cd888628510e90-Abstract-Conference.html

NeurIPS 原始论文；支持将反馈转为语言反思并存入记忆以影响后续尝试，不等同参数更新或内省体验。

[66] Wang, G.; Xie, Y.; Jiang, Y.; et al. 2023 Voyager: An Open-Ended Embodied Agent with Large Language Models

https://arxiv.org/abs/2305.16291

原始预印本；支持课程、技能库、反馈迭代的 Minecraft 智能体架构，不推论通用科学发现能力。

[67] Liu, N. F.; Lin, K.; Hewitt, J.; et al. 2024 Lost in the Middle: How Language Models Use Long Contexts

https://aclanthology.org/2024.tacl-1.9/

原始研究；支持所测模型与任务中相关信息位置影响利用效果，不宣称对所有未来系统普遍成立。

[68] Yao, S.; Yu, D.; Zhao, J.; et al. 2023 Tree of Thoughts: Deliberate Problem Solving with Large Language Models

https://arxiv.org/abs/2305.10601

原始论文；支持候选推理分支的探索与评价；不等同跨轮科学假说管理。

[69] Jumper, J.; Evans, R.; Pritzel, A.; et al. 2021 Highly accurate protein structure prediction with AlphaFold

https://doi.org/10.1038/s41586-021-03819-2

原始研究；支持特定结构预测能力及外部评估，不将结构预测等同全部生物功能和机制解释。

[70] Jimenez, C. E.; Yang, J.; Wettig, A.; et al. 2024 SWE-bench: Can Language Models Resolve Real-World GitHub Issues?

https://arxiv.org/abs/2310.06770

ICLR 2024，2023 首发预印本；支持由真实仓库问题与代码修复构成的执行评估任务，不引用动态排行榜。

[71] Brian A. Nosek; Charles R. Ebersole; Alexander C. DeHaven; David T. Mellor 2018 The preregistration revolution

https://doi.org/10.1073/pnas.1708274114

PNAS 115(11), 2600–2606。讨论假说生成与预测检验的区分，以及在观察结果前明确问题和分析安排；预注册本身不保证结论真实。

[72] Mark D. Wilkinson; Michel Dumontier; IJsbrand Jan Aalbersberg et al. 2016 The FAIR Guiding Principles for scientific data management and stewardship

https://doi.org/10.1038/sdata.2016.18

Scientific Data 3, 160018。提出可发现、可访问、可互操作、可复用四原则，并强调来源与持久标识；可访问性允许必要的身份与授权程序。

[73] I. Lakatos 1970 Falsification and the Methodology of Scientific Research Programmes

https://www.csun.edu/~vcsoc00i/classes/s497f09/s690s08/Lakatos.pdf

研究纲领中进步性与退化性转移的区分，用于讨论新预测和理论修订；不替作者虚构领域预测。

第379卷

让知识承担发生的历史

属性如何在同一发生中形成、维持与改变？本书从对象、互动与判断的相互规定出发，讨论读数事件、条件可比性、稳定机制与科研创造。

全书沿物理学、生命与神经、经济与制度、伦理与教育、AI及科研实践展开，区分概念论证、模型推演与经验检验，提出可由后续研究检验和修订的纲领。

稳定怎样形成

差异如何成为问题

知识如何重新进入发生

王德生　著

德麦国际出版社
Demai International Press

2026年10月 · 第379卷

定价 US$35

ISBN 979-8-90690-485-0

9 798906 904850