天文学思想史

天文学世界如何发生

阳涌 著

指导老师 王德生



德麦国际出版社

DEMAI INTERNATIONAL PRESS

天文学思想史

天文学世界如何发生

阳涌 著

指导老师 王德生

德麦国际出版社
Demai International Press

版本与阅读说明

书名：天文学思想史——天文学世界如何发生

作者：阳涌　　指导老师：王德生

版本：2026 年 10 月 · 数字阅读版 v1.1

德麦国际专著第 377 卷 · 定价：US$20
ISBN：979-8-90690-953-4
出版与平台：德麦国际出版社 · SDE Universes

开本：190 mm × 250 mm

规模：导论、十二编六十章及结语；正文 297,328 汉字。

本版提供完整正文阅读，并配套逐章学习任务及智能体“问穹”。正文汉字统计不含标题、目录、参考文献与学习材料。数字阅读版保留研究初稿的论证边界；原典核读范围和待精校事项见配套研究记录。

本书由 AI 协作起草与统稿。历史事实、现代重构与 SDE 解释在阅读中应分别核查；概念的可应用性不替代观测证据。封面为 AI 生成的主题艺术图，不作为历史仪器复原或真实天文观测图。

作者与指导老师

作者 阳涌

阳涌，SDE 学员，英语语言文学硕士。长期从事一线教学与课程改革，关注语言、数学、思维与人工智能教育的融合。自 2019 年起跟随王德生博士研修三视角、321 智慧、SIO 及 SDE 本体论，研究生成式学习、数学底层结构、创新阅读与教育智能体，探索知识在学科、生活与真实问题之间的生成和迁移。

指导老师 王德生

王德生，数学家、SIO 三态本体论及 SDE 本体论的创立者，获中国科学院计算数学博士学位，曾在英国斯旺西大学与新加坡南洋理工大学工作。持续研究知识、能力与主体性的发生机制，推动 SDE 方法论与人工智能研究和教育实践结合。

作者背景据 SDE 学员专栏公开简介整理。

怎样读这本书

先追问一个对象如何成立

按历史顺序读导论与第一至十编，观察“同一颗星”“轨道”“恒星成分”“距离”“黑洞图像”等对象怎样获得可重复辨认的条件。不要把当代分类直接投射到古代实践。

再比较三条贯穿全书的主线

天体身份、宇宙尺度与证据制度相互牵连。第十一编讨论巡天、统计模型与 AI，第十二编比较不同历史转折，检验 SDE 分析能增加什么解释，也明确它应当收缩的地方。

用学习包保留自己的变化

每章先写初答，再读原文，留下复答、异议、迁移计划和观察记录。学习任务是讨论的起点，不是历史事实测验的标准答案，也不以赞同本书作为学习成功。

与问穹对话时回到出处

智能体以选入章节和知识骨架为依据。要求它标明章与节，区分原文、推论和猜测；对任何新数值、引文或当代研究进展，继续回到原始资料核查。

目录

导论　天文学世界如何发生 10

第 1 编　天空经验与可记录天象

第 1 章 夜空怎样成为公共的研究对象 21

第 2 章 从月相到历法：时间单位的发生 25

第 3 章 食、彗与客星：异常何时能够被保存 30

第 4 章 行星为何成为“游走者” 35

第 5 章 如何比较不能同时看见的天空 40

第 2 编　古代几何宇宙的建构

第 6 章 从天球到数学坐标 45

第 7 章 希腊宇宙论中的几何与因果 51

第 8 章 托勒密体系：预测成功与解释代价 57

第 9 章 测量地球如何改变天文尺度 64

第 10 章 模型怎样跨越语言与时代 70

第 3 编　跨文明星表与观测改革

第 11 章 印度天文学中的计算体系 76

第 12 章 中国历法与星官体系的长期积累 82

第 13 章 伊斯兰世界的天文表与观测台 89

第 14 章 图西偶与行星运动的几何改造 95

第 15 章 乌鲁伯格星表与精度共同体 101

第 4 编　日心体系与望远镜新证据

第 16 章 哥白尼：行星秩序与解释联系的重构 107

第 17 章 第谷：精度为何改变理论竞争 113

第 18 章 开普勒：一个残差怎样迫使轨道重写 120

第 19 章 伽利略：望远镜图像如何获得可信地位 127

第 20 章 为什么革命不是一次性胜利 133

第 5 编　轨道动力学与精密位置

第 21 章 牛顿：从天上轨道到普遍动力学 138

第 22 章 预报回归的天空：哈雷与星历实践 145

第 23 章 天王星残差与海王星：预测怎样生成新对象 151

第 24 章 恒星光行差、视差与“地球会动”的证据 158

第 25 章 标准天球、时间与测量误差的制度化 164

第 6 编　光谱学与恒星物理的诞生

第 26 章 太阳光谱暗线：星光怎样变成可测特征 172

第 27 章 实验室与恒星之间：光谱化学的归属方法 178

第 28 章 哈佛摄影与恒星光谱分类 185

第 29 章 赫罗图：两种测量怎样生成恒星演化问题 192

第 30 章 恒星为什么发光：从辐射到核反应 199

第 7 编　宇宙距离阶梯与星系身份

第 31 章 距离尺度：校准链、独立锚点与可迁移性 206

第 32 章 勒维特：周期—光度联系的建立 211

第 33 章 “岛宇宙”争论：星云究竟是什么 217

第 34 章 从速度与距离到宇宙扩张关系 222

第 35 章 距离再标定：为什么一个旧数值的修正改变整个宇宙 228

第 8 编　膨胀宇宙与精密宇宙学

第 36 章 动态宇宙：场方程、假设与观测约束 234

第 37 章 宇宙微波背景：噪声怎样成为宇宙证据 241

第 38 章 轻元素丰度与早期宇宙：两条证据链怎样交叉 247

第 39 章 不可见质量：引力证据怎样改变物质分布图景 253

第 40 章 宇宙加速与参数拟合：理论何时需要再打开 259

第 9 编　多波段、多信使与不可见宇宙

第 41 章 无线电天空：新波段如何生成新的天体目录 266

第 42 章 X 射线与伽马射线：突破大气遮挡的观测 272

第 43 章 脉冲星与中子星：时间序列如何确证对象 278

第 44 章 引力波：非电磁信号怎样成为一类公共事件 284

第 45 章 EHT 黑洞图像：从干涉数据到稳健特征 292

第 10 编　行星系统、系外行星与生命问题

第 46 章 太阳系探测：从远观到到场测量 300

第 47 章 行星、矮行星与类别边界：定义会怎样回写探索 306

第 48 章 第一批系外行星：径向速度怎样支持看不见的伴星 312

第 49 章 凌日、成像与系外行星人口统计 320

第 50 章 生命迹象与天文可居住性：证据距离有多远 327

第 11 编　大规模巡天、AI 与证据责任

第 51 章　巡天不只是更多数据：选择函数重写对象集合 335

第 52 章　Gaia：百亿观测怎样成为视差、星表与银河结构 341

第 53 章　模拟宇宙与统计模型：模型结果能成为何种证据 348

第 54 章　AI 巡天候选：发现、分类还是制造标签 353

第 55 章　宇宙学张力：共享误差与理论重开的判据 359

第 12 编　天文学思想发生的跨时代检验

第 56 章　天体身份的四种发生：观测、推断、确认、再分类 366

第 57 章　宇宙距离与时间：两种不可直接取得的尺度 371

第 58 章　用三方程比较五次天文转折 376

第 59 章　六路径与三原理：创新为何有时发生、有时受阻 382

第 60 章　新的天文学世界：已成能力、盲区与开放条件 389

结语　已支持、应收缩与开放问题 394

导论　天文学世界如何发生

0.1　抬头之后，世界才开始成为问题

夜晚抬头，人可以看见亮点、明暗、疏密以及彼此的位置。可是，“看见了天空”与“已经拥有一个天文学世界”之间，隔着漫长而复杂的工作。一个亮点怎样与昨夜的亮点认作同一个？两次位置不同，是它在运动，还是观察者换了地方？一片微弱的光是一个对象，还是许多对象未被分开的显露？这些问题并非在对象已经完整摆好之后才附加上去。对象能够以何种身份进入研究，已经受到这些问题及其处理方式的规定。

设想两个人分别在不同地点记录同一颗亮星。这不是一个历史事件的复原，而是帮助我们拆开问题的简单思想实验。甲用山峰作参照，乙用屋顶作参照。甲说星在左边，乙说星在右边。他们的说法可以同时诚实，却不能立即比较。把左右译成共同方向仍然不够，因为日期、时刻和所在地都影响可见位置。只有在这些条件之间建立转换，两个地方的天空才获得一种可共享的对应。共享并非观察者的消失，而是观察差异经过组织之后产生的新能力。

这一步很容易被成熟知识遮蔽。今天我们用一个天体编号或一组坐标，便能把不同人的讨论连接起来。编号与坐标仿佛天然附在天空上，读者也往往将星表当作已经存在的对象清单。然而每一个稳定条目，都以某种方式解决了复认、区分、定位和记录问题。表格压缩了它的形成史。越是方便复用的成果，越容易让人忘记支撑这种方便的劳动，以及它在哪些条件下不再可靠。

天文学思想史由此具有双重任务。它需要讲述历史人物提出什么、观察什么、计算什么；也需要解释这些活动如何构成了可以继续操作的世界。仅仅列举不断增加的发现，无法说明为什么某些时代可以询问恒星的成分，另一些时代却连这个问题的证据入口都尚未建立。反过来，仅仅讨论观念如何变化，也可能忽视仪器、计算和实践如何改变了观念所能涉及的对象。

本书以 SDE 发生学尝试重新组织这个问题。S 指显露，D 指差异序列，E 指特征纠缠。我们关心的不是给每段历史贴上三个字母，而是追问：一种显露通过哪些不断承接的差异得以稳定，又与哪些不能独立解释的特征相互构成？当新的显露成立以后，它如何改变后续可走的路径，以及对象能够在其中获得意义和作用的条件？这组问题需要用具体历史回答，而不能由抽象形式替历史回答。

因此，本书所说天文学家创造了新的天文学世界，具有明确的解释内容：他们参与形成新的显露单位、差异操作与特征纠缠聚合体系统。肉眼星区、轨道体系、光谱恒星、河外尺

度、干涉成像和多信使事件，不仅是对同一张完成的目录不断添加注释。它们使不同的属性、对应关系、可计算量与研究行动得以成立。哪些变化达到这种构成上的强度，哪些只是精度提升或名称更改，必须分别判断。

这也意味着，创造世界不能靠修辞升级完成。换一个名字，未必改变研究对象；发表一个新模型，未必使它取得可靠身份；增加一台仪器，也未必产生新的属性系统。需要说明新增的构成关系在哪里，什么操作因此成为可能，什么判断因此必须修正。我们既不因术语新颖就宣布新世界已经诞生，也不因为旧编号被保留便否认对象的研究构成发生了深刻改变。

在这一意义上，历史连续与世界新生可以同时成立。新的研究系统会继承旧数据、旧坐标和一部分算法，也会拆开旧系统曾经混为一体的东西。过去称为同类的对象可能分化，过去被看作彼此无关的信号可能取得共同事件身份。连续性提供比较的基础，差异性使比较成为必要。思想史的责任正是说明两者怎样纠缠，而不是预先将历史写成彻底断裂或平滑累积。

读者不必在翻开本书时接受全部 SDE 主张。只需允许这些问题进入历史，并检查它们是否让某些通常被压缩的环节重新显露。如果它们能够指出一个未被说明的转换、一种被忽略的依赖，或者一次误把结果当起点的推理，便取得了具体价值。如果它们只把已有解释换了说法，本书就应当承认这一限度。理论进入历史，首先要承担被历史修改的责任。

0.2 遥远不是单纯的距离

天文学面对的遥远，首先是一种操作处境。对于绝大多数研究对象，人们无法像移动桌上的器物那样改变它的位置，无法让一颗恒星为了复核而重新演出过去的事件，也不能随意更换一个星系的组成。能够调整的，常常是探测器、观测时段、波段、计算方法和比较对象。研究者对对象的作用方式受到限制，但这并不意味着研究只能消极等待。它迫使科学把检验安排在不同的联接之中。

例如，一种解释若认为信号来自某个方向，就可能对别的仪器、别的时间或别的波段提出可以检查的要求。某些成分判断需要与实验室的物理过程比较；某些距离推断需要另一种尺度关系提供锚点；某些图像解释必须同时符合实际采样到的信号。可以改变的条件越有限，就越需要明确哪些约束彼此独立，哪些只是同一个假设的重复出现。

这里的有限不能只理解为信息太少。信息很多，仍然可能只覆盖对象的一部分；测量很精细，仍然可能共享同一种系统误差。把同一份资料经过不同程序处理，可能有助于检查算法依赖，却不能因此被当作多次完全独立的观测。证据力量来自差异的组织方式，而不只是条目数量。增加十个相同来源的支持，有时不如获得一个真正不同来源的限制。

由此可以理解天文学为什么不断依赖转换。角度转成位置，光强转成光度判断，光谱差异转成运动或成分约束，时间序列转成周期，若干观测转成可共享的源身份。每一种转换都有适用条件。条件不是结论之外的附加说明，而是结论能够以当前意义成立的一部分。脱离条件保留一个数值，很可能同时丢掉这个数值究竟在说什么。

特别需要警惕的是“间接”的贬义。间接推断不天然弱于直接观看。视觉也有光学条件，照片也有响应范围，现场读数也有仪器误差。关键不是中介有多少层，而是各层如何受约束，错误怎样被识别，结论在条件变化时是否保持稳定。一个推断链可以很长却可复核，也可以很短却无法排除混淆。思想史不应以“终于亲眼看见”作为所有证明的共同终点。

另一方面，承认中介的构成作用，不等于说任何结果都可任意制造。仪器有可以校验的响应，模型需要承担不符合资料的后果，新的分类必须面对无法容纳的个体。科学世界的生成包含阻力。正是这些并不听从研究者愿望的差异，使世界的形成有别于任意编写的故事。本书使用“发生”一词时，始终保留这种能够使判断失败的约束。

遥远还带来时间的结构。一次短暂事件结束以后，能够留下的可能只是记录和信号。后来的研究者无法回到原初观测处境，却可以重读记录、重新处理数据、换用模型或发现其他关联材料。历史因此不仅存在于研究对象的过去，也进入科学证据的现在。一张旧底片的作用，可能在新的分类和计算条件出现以后才扩展；其记录时刻与其知识作用的形成时刻并不相同。

同样，一种被保留下来的古代记录，可以进入远晚于记录者时代的问题。利用这种材料，需要尊重原有语言、纪年和记录目的。我们不能因为某条记录后来有助于研究天体事件，就假定记录者已经在从事同样的研究。记录的第一次使用和它后来的重新使用，属于不同的差异序列。它们能够联接，但联接需要被解释，不能靠后见之明直接抹平。

这使天文学思想史具有一种特殊的长时段结构。研究成果可以跨越很久的时间成为后来工作的条件，但这种潜能并不保证实现。资料可能散失，单位可能不明，工具可能失传，意义也可能无法翻译。把保存当作已经完成的传播，会忽视这些中断。真正进入后续研究的，是得到重新定位、校验和组织的材料，而不只是物理上继续存在的纸张或刻痕。

天文学的远距处境于是将一个普遍问题放大：人怎样在无法穷尽对象、无法消除所有中介的条件下，形成仍然值得信任的世界？本书的回答不从一项终极保证出发，而从具体成功和具体失败出发。我们将看见不同形式的可信性怎样形成，怎样扩大，又怎样暴露边界。这些边界不是知识之外的黑暗背景，而是知识继续发生的切口。

0.3 三条主线与竞争解释

全书第一条主线是天体身份。“这是什么”看似是最简单的问题，却可能同时涉及位置复认、运动规律、成分、结构、来源和分类目的。两个信号被判为同一个对象，需要条件；一个旧类别被拆成多个类别，同样需要条件。对于某些研究，稳定编号已经足够；对于另一些研究，编号稳定而物理解释仍可大幅改变。对象同一性不是一个在所有问题中保持相同内容的印章。

第二条主线是宇宙尺度。距离使许多属性获得不同意义：表观明亮与本征发光能力需要区分，角大小与物理大小需要联接，远近判断也可能影响对象类别。尺度不是只在研究开始时测一次的数字。校准变化会向下游传播，下游出现的不一致又可能促使研究者返回尺度来源。沿这条线，可以看到一个结果怎样变成条件，一个条件怎样再次成为问题。

第三条主线是证据制度。这里的制度不仅指学会或行政机构，也包括记表方式、比较规则、公开程序、错误报告和资料共享。一个结果能够被另一人检查，需要某种稳定传递形式。这种形式可能帮助差异积累，也可能让不符合既有栏目和分类的现象难以进入。因而制度既能增强研究能力，也能塑造研究盲区。必须具体检查，不能预先把制度只写成促进或压制科学的单向力量。

三条线并非各写二十章之后在结尾会合。身份可以决定哪些对象被当作尺度工具，尺度可以改变身份判断，证据制度又影响哪些身份与尺度得到复用。以变星研究为例，周期和亮度之间的关系只有在具体样本与比较条件中取得意义；这种关系后来能否成为测距资源，还需要额外校准和适用性判断。1912 年勒维特相关原始论文呈现的表格和图形，为区分关系形成与后续工具化提供了明确入口。[导-1]

但这种研究方向并非无人踏足。科学史早已超越单纯的英雄名录；专业天文学史讨论观测、计算和模型竞争；科学哲学讨论观察的理论条件与对象的可靠性；统计学研究校准、误差与可辨识性。若只把这些成熟问题重新排成 S、D、E 三栏，就不能声称完成了理论创新。本书必须承认对手已经解释的部分，才能判断自己是否提出了更多需要回答的问题。

第一类竞争解释是以科学革命、研究传统和具体模型史为中心的解释。这类研究提醒我们，理论转换涉及评价标准、共同体和旧问题结构。本书与它的分歧不能被简化为“他们只看观念，我们看实践”，因为成熟史学本身就研究实践。真正应当比较的是：在某个具体案例中，解释是否明确到对象属性如何成立、不同操作如何联接，以及新成果怎样改变后来研究的可行条件。[导-2]

第二类竞争解释是表征、仪器和干预的哲学研究。它能够说明科学对象的可靠性如何与实际活动相连。天文学尤其适合检验这类讨论的适用范围，因为对遥远对象的干预常常受到强限制，而对信号取得与处理的干预却十分丰富。我们不能把实验室中有效的判断标准原样

搬到所有天体上，也不能因为不能直接操纵恒星，就否认关于恒星的强证据能够成立。[导-3]

第三类参照是现代测量与统计方法。它们不是等待哲学替它们发现关联的简单工具。联合估计、模型比较和误差传播已经揭示许多深刻依赖。例如 Gaia 天体测量的核心求解，将源参数与姿态、校准等参数置于共同估计问题中，原始技术论文明确介绍了相应的分块处理。[导-4]本书不能把这一事实说成 SDE 新发现的算法。可以继续追问的是，这种联合成立方式如何进入对象属性的解释，以及它对历史叙述提出什么额外要求。

与这些解释和方法比较，意味着本书可能得到程度不同的结果。某些章节只能提供较清楚的整理；某些章节可能纠正把先后次序当成因果的叙述；某些章节可能辨认出以前未充分展开的跨系统构成关系。它们不必被强行放在同一高度。把一项细小但可核查的修正写清楚，比把每一次说明都命名为本体论革命更有价值。

同样，SDE 理论文本是解释 SDE 概念的重要原始材料，却不单独证明天文学事实。科学家的论文、观测记录和仪器资料承担另一种责任。前者不能替代后者，后者也不能被任意改写成理论作者没有提出过的哲学主张。本书保持这两种来源之间的区别，正是为了让思想重构与史料研究能够真正相遇，而不是相互替代。

0.4 何种变化值得称为世界的发生

为了判断不同转折，本书首先区分精化、确立、重构与联接。精化是在既有研究关系中提高分辨力或修正参数；确立是让一种身份或联系获得较可靠的支持；重构是改变对象单位或属性系统；联接是让原来分立的系统形成新的复合研究能力。这不是按高低排列的四级荣誉，而是帮助我们避免把所有变化都称作同一种进步。

一次参数修正也可能非常重要。若某个尺度被广泛复用，它的调整就会影响大量后续推断。然而影响范围大，不自动等于对象构成方式全变。相反，一个只涉及较小对象集合的新分类，也可能深刻改变这些对象如何进入研究。变化强度不能只按影响人数、媒体关注或数值改动大小衡量，而应追踪哪些关系被保持、取消、增添或重新定义。

一个有用的判断办法，是问旧系统中的操作能否不加修改地继续使用。如果原有分类、校准和比较仍然成立，只需更新某个数值，主要变化可能属于精化。如果一项测量开始回答以前无法提出或无法检验的问题，就需要进一步检查新的属性是否成立。如果不同通道之间出现可操作的对应，还要问这种对应有没有增加单一通道所没有的限制。判断依靠这些差异，不依靠章节标题的宏大程度。

这里必须分清天体的物理形成与科学对象的构成。恒星的形成属于具体天体物理问题，需要关于物质、动力学与演化的证据。恒星作为具有某种谱型、距离、成分和演化地位的科

学对象，又涉及这些属性如何在研究系统中成立。后者并不只是给前者贴名字，因为属性的定义和联接可以改变我们能够进行的解释与操作；但它也不能被偷换成“研究者的命名引发了恒星物理诞生”。

为了分析两者的联接，本书区分自然过程、观测操作、知识构成和共同体活动四个层次。它们不是彼此隔绝的四个世界。自然过程产生的信号进入仪器，仪器输出经过处理形成研究资料，资料进入判断，判断又影响后续观测安排。不同环节具有不同的证据责任。当我们说一种模型改变了世界，必须说明它改变了哪一层，以及通过什么中介影响其他层。

如果不做这种区分，发生学语言可能轻易制造解释幻觉。把星光当 S、把论文争论当 D、把整个社会当 E，形式上似乎什么都有，实际却未必说明这些层怎样连接。一个范围无限扩大的 E，几乎可以吸收任何缺失原因；一个随结果变化而重新指定的 S，也几乎不可能遭遇反例。因此，每章分析都需要在一个主要问题上保持研究单位稳定。

研究单位稳定并不意味着永远不能改。恰恰相反，身份转换有时会迫使我们承认原来的单位不合适。但这种改变要作为结果被报告，不能悄悄发生在论证中间。一个信号原先被视为单一来源，后来被分解为多个来源，便应解释分解证据如何出现、原单位为何失效，而不是在叙述前后分别使用不同意义的“这个对象”，假装从未发生转变。

本书还将区分构成关系、推断关系和历史因果。构成关系关心一个属性离开某些条件后是否仍有当前意义；推断关系关心证据怎样支持某个结论；历史因果关心某项实际行动如何改变后续研究。三者可以联接，却不能互相代替。一个现代分析发现两个参数需要联合估计，并不能由此证明历史人物已经因此修改研究制度；制度的改变需要相应历史材料。

例如，一种新尺度在逻辑上能够支持更远对象的研究，只能说明新的可能性。要说它实际上推动了某次观测，就应当找到研究方案、论文使用或其他可追溯联接。反过来，某项观测发生在新理论发表之后，也不自动表明由该理论引起。它可能早已计划，或者出于其他任务。把“能够”写成“已经”，把“随后”写成“因此”，是思想史特别常见而又容易被顺滑叙述遮住的跳跃。

这些区分服务于更强的正面论证。只有知道一种变化确切发生在哪里，我们才有能力解释新世界的构成，而不是始终退回“认识更深了”或“工具更好了”的泛泛评价。本书所期待的世界发生，是有明确单位、有操作内容、有属性关联，并能在继承和差异中展示出来的发生。它的哲学分量，应由这些具体工作承担。

0.5 从关系到路径，再到改变的动力

三方程提供本书分析的共同骨架：S＝F(D,E)，D＝G(S,E)，E＝H(S,D)。它们要求同时考虑显露如何成立、差异序列如何受已成结果和条件规定、特征纠缠如何被结果与行动重

构。SDE 原著明确把它们放在元方程层次，具体算子需要由研究领域给出。[导-5]因此，本书不会把这三行形式当作已经求解的宇宙方程，也不会假定将历史材料放进三栏就自动得到解释。

第一式使我们返回结果的构成。拿到一个已经被称为天体的来源，要问它通过什么比较和限制取得身份。这里的 D 并非任意差别的集合，而是差别如何被承接、排除、重试与稳定的序列。第一次异常与第二次复核之间，究竟保留了什么、改变了什么？如果只有散乱资料而无可说明的承接，便不能因为它们时间相邻就称为一条有效的发生路径。

第二式使我们看到成果具有方向作用。一个已成立的类别可以规定下一轮应当寻找的特征；一个候选目标可以限制观测安排；一个尚未证实的模型也会组织需要检查的差异。这里不存在未来实体向过去施力的问题：起作用的是当时已经存在的模型、目标表征和判断标准。它们可能指导研究，也可能使某些差异被忽略。结果对路径的规定并不保证方向正确。

第三式则使条件不再永远停在背景位置。一次成功的测量可能留下基准，一次失败可能改变校准，一次分类争议可能重写目录规则。研究通过行动改变了后来的研究条件。但是条件的改变并非全能：资金增加不能自动解决不可辨识问题，算法更快也不能自动补回未被记录的数据。回写需要具体说明改变了什么，哪些限制仍然留下。

三式的联合尤其要求警惕 E 被当作清单。说研究依赖仪器、数学、制度与人才，只说明因素很多。要表现特征纠缠，需指出因素的特征如何相互规定。一个谱线位置必须相对于标尺解释，一个视差估计可能与姿态校准共同受限，一种分类边界会随可比较属性改变。改变其中一项，会使另一项的意义、估计或稳定性受到影响，这才进入本书要分析的关系深度。[导-6]

在这个关系骨架上，六路径使研究入口可以改变。由 S 出发，我们可能从现成结论追查构成；由 D 出发，可能从异常与操作序列研究新的结果怎样形成；由 E 出发，则可能先分析哪些约束使某些结果可行、另一些不可行。每一个入口还可选择先进入另外哪一维，由此形成六种基本顺序。实际研究可以往返、并行和中断，六路径没有把它压成六种固定人格或六个历史阶段。

路径的价值要以判断是否改变来衡量。若从成图出发，人容易接受图像显得清楚的部分；从采样条件出发，则可能重新检查哪些细节未受数据充分约束。若从成熟距离工具出发，人容易忽略原始关系与绝对标定之间的距离；从最初表格出发，则能看到后来补入的条件。路径转换不是为了把同一段话再讲一次，而是为了暴露一个入口看不见的限制。

三原理把研究推进到改变与受阻。按照本书使用的定义，D 与 E 的失配可能引出 S 改变，S 与 E 的失配可能引出 D 改变，S 与 D 的失配可能要求 E 重构；新的改变还会回到其他

维度形成后续作用。[导-7]这些是动力分析的候选相位，而不是只要出现冲突就必然成功的保证。错误模型可以长期稳定，异常也可以被搁置，甚至形成更多错误。

因此，动力分析必须保留分支。一项残差可以促使修改仪器、调整参数、增加对象，也可以促使改变原有理论。哪种响应在当时成为可能，受到材料、工具、问题结构和判断标准限制。只有在结果出现以后才说“这正是第三维发生”，无法区分成功机制与事后命名。我们要比较实际走过的路径及当时有理由考虑却未实现的路径，并说明哪些证据使某个分支终止。

这三种方法的结合，不是每章先列公式、再画箭头、最后补一句“回写”。三方程帮助明确关系，六路径检查入口偏差，三原理解释改变和停滞，再由新的关系检验此前判断。它们在实际论证中可以交错出现。有的章节以仪器排错为中心，有的以模型竞争为中心，有的以文本传承为中心。统一的方法责任可以承载不同的叙述形态，这也是本书组织长篇历史的基本选择。

0.6 怎样阅读一部仍然允许自己被改写的思想史

本书按十二编推进，但不把人类天文学写成一条从某个文明出发、经过单向传递而抵达现代的道路。不同地区的历法、星表和计算传统有各自任务，也可能发生交流。能够证明的传播，需要文本、算法、器物或其他可追溯中介；无法证明的相似性，则应保留为比较问题。让多个传统进入同一部书，不意味着提前将它们压成一个必然汇合的故事。

前几编从天空经验、记录与几何体系展开，研究重复显露怎样取得公共身份。随后的章节进入跨文明修订、日心体系和望远镜证据，再讨论动力学、精密位置与恒星光谱。中间几编围绕距离和宇宙结构，将局部观测与大尺度模型连接起来。后面转入多波段、多信使、行星探测和数据化研究，最后回到整个解释框架，检查哪些结论得到支持，哪些必须收缩。

这种安排有历史次序，也有主题回返。恒星视差与后来校准的讨论相隔若干章，并不因为中间缺少连续性，而是因为它们承担不同责任。前者研究一种操作如何成立，后者研究其成果怎样与其他方法连接。读者可以沿天体身份、尺度或证据制度中的任意一条线阅读，但每次返回旧问题，都应看到新的层次，而不是重新经历同一段人物故事。

史料在书中承担的责任也有所不同。原始论文可以告诉我们作者提出了什么、使用了什么数据，却未必完整呈现研究过程；通信和笔记可以补充过程，却也不是毫无选择的内心记录；后来的专业研究能够汇集材料，却需要注意解释立场。本书尽量让结论与其支持材料相称。没有足够材料的细节，不以“可能就是这样”的故事感填补。

对公式也采用同样尺度。现代读者熟悉的表达可以帮助说明历史问题，但必须知道哪些是当时使用的形式，哪些是为解释而作的现代重写。数学推导正确，不代表其前提在历史情

境中已经取得支持；一个模型拟合某组资料，也不代表所有参数都有唯一物理解释。形式化越清楚，越应帮助我们看见假设，而不是把假设藏在符号背后。

本书会保留失败和未决问题，因为成功结果本身不足以说明发生机制。如果同一类异常曾经导致可靠新对象，也曾经引出未能确认的对象，就需要比较两种路径，而不能只选择胜利的一支。如果一个框架在多数方面有效，却在某个问题上存在困难，也不能为了叙述整齐把局部困难清除。真正的统一，应当能够容纳不同程度的成功。

对于 SDE 本身，同样必须执行这一要求。理论可以增加研究问题，也可能使研究者偏好某类材料。特别是借助 AI 进行长篇综合时，语言的一致与论证的流畅容易被误认作证据的充分。SDE 原著关于自我修订的讨论也指出，文本规模不能代替检验，概念来源的权威与经验支持需要区分。[导-8]本书因此把竞争解释、反例与不确定性放入正文，而不是只在结尾作一般性声明。

统一性、多样性与和谐性在这里具有具体含义。统一性是概念不随章节任意改变、同类主张采用相近证据标准；多样性是不同历史对象保留自己的问题与操作，不被迫重复一种公式化故事；和谐性是各章互相提供条件，又明确自身边界，不通过抹平分歧取得表面一致。它们之间存在张力，需要在写作与阅读中反复调整。

读者可以用一个简单问题持续检验本书：读过这一章以后，我是否更清楚某种天文学世界怎样成立，而不只是多记住一个名字或年代？若答案是否定的，那么无论概念多么整齐、例子多么壮观，这一章仍未完成它的责任。若答案是肯定的，还应继续追问：这种解释留下了什么尚不能说明的部分，哪些材料出现时应当改写它？

天文学思想史所展开的，最终是一种面对有限处境的创造能力。人并不能随意支配宇宙，却能够改变信号如何进入比较，改变问题怎样形成，改变结果怎样成为下一次研究的条件。一个科学世界的形成，不以所有困难已经结束为标志，而以新的、受约束且可继续检验的行动开始成为可能为标志。我们将从最初的复认与记录开始，追踪这种可能性怎样一次次获得具体形态。

本章参考文献

[导-1] Leavitt, H. S.; Pickering, E. C. Periods of 25 Variable Stars in the Small Magellanic Cloud. Harvard College Observatory Circular 173, 1912, pp.1–3. 原始扫描：https://articles.adsabs.harvard.edu/pdf/1912HarCi.173....1L 。本导论使用其研究对象和材料形式。

[导-2] Kuhn, T. S. The Copernican Revolution: Planetary Astronomy in the Development of Western Thought. Harvard University Press, 1957. 书目核对：https://search.worldcat.org/title/1022740307 。本导论将其作为比较对象，未据此给出未经核定的页码或引语。

[导-3] Hacking, I. Representing and Intervening: Introductory Topics in the Philosophy of Natural Science. Cambridge University Press, 1983. 出版社书目与内容说明：https://www.cambridge.org/core/books/representing-and-intervening/F6506B708BB5A8B6A5D884BDCF28E7B7 。本导论以其问题领域作为哲学对照，不宣称完成逐章文本评述。

[导-4] Lindegren, L. et al. The astrometric core solution for the Gaia mission. Astronomy & Astrophysics 538, A78, 2012. DOI:10.1051/0004-6361/201117905。https://arxiv.org/abs/1112.4139 。联合估计与分块处理参见论文摘要及正文方法说明。

[导-5] 王德生：《SDE 三方程导论》第二章“三方程的标准形式与理论地位”。https://sdeuniverses.com/books/m/51/chapter-02/ 。

[导-6] 王德生：《SDE 思想体系入门》第九章“E：特征纠缠维度”。https://sdeuniverses.com/books/m/395/chapters/ch09 。理论概念来源；本导论中的天文学分析由本书提出。

[导-7] 王德生：《从三视角的 123 原理到 SDE 的三原理》，第四、八篇。https://sdeuniverses.com/column/ontology-grid/123-to-three-principles/ 。三原理概念来源；其天文学解释须另行检验。

[导-8] 王德生：《SDE 三方程导论》第三十九章“反例、批评与理论修订宪法”。https://sdeuniverses.com/books/m/51/chapter-39/ 。参见来源分层、AI 协作和修订责任的讨论。

第 01 编

天空经验与可记录天象

第 1 章 夜空怎样成为公共的研究对象

第 2 章 从月相到历法：时间单位的发生

第 3 章 食、彗与客星：异常何时能够被保存

第 4 章 行星为何成为“游走者”

第 5 章 如何比较不能同时看见的天空

第 1 章 夜空怎样成为公共的研究对象

1.1 地平线不是天然坐标

一个人抬头看见亮星，另一个人也抬头看见亮星，这还没有形成两人共同研究的天空。他们可能站在不同的山坡，选择不同的时刻，把不同的光点叫作同一个名字。公共研究对象的第一个困难，并不在于遥远恒星的内部结构，而在于怎样使“我刚才看见的那个”成为“你能够再次找到的这个”。天文学思想史必须从这种复认的成立讲起。

地平线看似是最现成的参照。山脊、树木、建筑物，把开阔天空分出了边界；一个亮点从某个缺口升起，便有了容易记忆的位置。然而，缺口属于特定观察地点。向旁边移动，近处山脊的轮廓便会相对于远处星空改变。若没有同时保存观察点，“在山峰右侧”只能是一条局部说明，不能直接成为跨地点的天体坐标。

这里已经包含一种值得重视的相互规定。一个方向标记之所以能帮助认星，是因为观察者在适当时刻从适当地点使用它；而这个时刻是否适当，又要根据目标在天空中的位置判定。位置、时刻和地点不是各自完备之后偶然相遇的三个信息。它们共同规定了这一次显露能否被下一次观察接续。移走其中一项，原来的名称未必还能指向原来的可见配置。

现代位置天文学依旧保留这种责任。美国海军天文台的天体导航服务要求输入地点和时刻，并区分计算高度、方位以及折射等改正项。[1-1] 这里不能反过来假定古人已经拥有现代修正理论；它只让一个基本界限变得清楚：从地上看到的天空位置，并非脱离观察条件而独立存在的一串数值。早期观察虽然没有这种数值工具，却同样必须在实际活动中处理位置依赖。

因而，固定观察点本身就是一项知识操作。它减少了不必要的差异，使某些差异值得记录，某些差异暂时可以忽略。当观察者今天和昨天站在相近位置，选择相近天光条件，比较同一星群附近的亮点，天空开始呈现有结构的重复。“稳定”不是把变化完全取消，而是在能够控制或辨明的变化中维持复认。

用本书的语言说，这里的 S 是可复认星区的显露，D 是跨时刻、跨夜晚的比较序列，E 则是位置、出现时序与地平参照的相互规定。说 S 由 D 与 E 成立，不是说天空中的天体由观察者制造，而是指出：作为可以接续研究的星区，它的身份并不等于单次视觉中的一个亮点。形成这种身份之后，观察者才有理由再次回到特定地点寻找它。

1.2 口述与图像记录

复认并不以文字为唯一前提。一个人可以通过指示、动作和反复练习，把星群的辨认方法教给另一个人。名称与故事能够组织记忆，使散乱亮点形成容易传承的形状。若把一切非文字传统都视为模糊印象，就会把记录媒介的差别误当成知识有无的差别。

但是，口述的有效性要落实到具体能力：听者能否在合适时刻找回目标，能否指出可能混淆的邻近光点，能否在外观改变时仍认出相应星区。故事保存了称谓，并不当然保存了这些能力。同样，写在器物上的星形符号也不自动是一幅精确星图。符号排列可能表达秩序、礼仪或身份，也可能兼有观测用途；决定判断的是材料之间能够建立怎样的对应。

这使我们面对史前材料时必须采取不对称的谨慎。认定一件器物含有天空意象，通常只要求辨明图像语汇；认定它用于具体天象计算，则需要更多信息，如稳定的数值对应、方向关系、使用痕迹或同类材料。如果把一种可能的解释提升为确定用途，后来的论证便会围绕未经证明的起点不断加固，最终使解释显得比证据更完整。

与含义争议较大的图像相比，有文字结构的星表提供了另一种入口。大英博物馆对 MUL.APIN 一块泥板的馆藏说明，列出了天空区分、主要星体升起日期、同时升落的星体以及月亮路径中的星群。[1-2] 本章使用的是馆藏释读信息，并非自行校读楔形文字。即使只在这个范围内，材料也足以表明：保存下来的不只是孤立星名，还有名称之间的时序和位置联系。

这种联系具有操作意义。只记住一个名字，观察者仍可能不知道什么时候寻找；若同时知道它与另一星群的升落关系，就可以把一个已认出的现象作为寻找另一个现象的线索。知识在这里不是名词数量的增加，而是辨认活动之间出现了接续能力。星名从记忆单位进入操作网络，才能构成可重复使用的天空知识。

现代国际天文学联合会将星座组织为具有明确边界的天空区域，与古代借亮星构成的形象并不相同。[1-3] 这个差别提示我们，不应把今天的星座地图投射回全部过去。古人可能认识若干同样的亮星，却并未把整片天空划成今天这样的区域。名称沿用可以掩盖对象单位的改变；反过来，名字不同也未必意味着辨认的是不同星群。

因此，“大家看见同一个天空”必须拆成几个不同问题。是否接收到了来自相同方向的光？是否组织成相同图形？是否用相同边界划分对象？是否能够相互核对？第一项成立，并不使其余三项自动成立。文化差异不是后来附着于纯粹天空的装饰，它参与了星区如何成为研究单位；但这种参与也不能取消方位和时序对解释的限制。

1.3 重复出现的力量

天空经验最重要的资源之一，是重复中仍有差异。月亮会重新呈现相似外观，某些亮星会在相近季节重新出现，但每一次观察都受到天气、地形、视力与选择时刻的影响。重复的价值不在于保证每次所见完全相同，而在于允许观察者把多次经验放进同一问题之中。

假设连续数夜都在天色接近时观察某个星群，记录其中一个亮点相对邻星的位置。这是一个解释性例子，不是对某位古人行为的复原。第一次记录只能描述位置；第二次开始提供差异；更多次记录才可能分辨持续变化和偶然误认。如果某夜没有看见目标，缺失本身也要解释：它是被云遮住，还是接近太阳方向而难见，抑或先前辨认有误？

由此可见，观察序列不是把一串现成事实按日期排好。序列的形成需要持续维护同一性。要说“它移走了”，先要说明这个“它”跨越几次目击仍然成立；要说“它没有移走”，也要说明比较的精度允许这种判断。差异与同一不是彼此排斥的终点，而是在一项追踪活动中互相规定。

农业活动确实能够提出季节性需要，但仅仅指出农业用途，还不能解释观测为何变得可靠。种植和收获提供了研究动机；固定参照、传承规则和复查机会提供了实现这种动机的条件。同样的实用需要可以容许不同程度的精度，也可以由天气、植物或动物征候部分满足。因此，不能从“需要掌握季节”直接推导“必然形成精密星表”。

这也是 SDE 分析在这里试图增加的一层区分。若只从社会功能进入，会容易把记录的出现归因于需求；若从已经形成的星名进入，又容易假定稳定对象始终存在。把入口改为实际比较序列，就可以追问：需求究竟通过什么操作转化为可以复认的星区？对象的稳定在哪里获得支持，在哪里仍依赖熟练者的经验？

重复还会反过来改变观察的组织方式。一旦某星群被认作有用的季节标志，观察者便不再平均地关注所有亮点，而更可能在特定时段寻找它。已经形成的 S 规定新的 D，新的比较又要求维护相应参照 E。这个相互规定可以形成知识积累，也可能形成注意的收窄：未被纳入已有命名系统的微弱或短暂现象，更容易被忽略。

因此，公共天空的形成从来不是把一切差异完整收集起来。它建立了一种取舍：哪些差异值得持续保存，哪些可以在当时任务中放下。这种取舍使知识成为可能，也为后来被称作异常的现象准备了背景。没有稳定预期，便很难说某次所见偏离了什么；没有保留意外的能力，稳定预期又可能把新现象消解为误认。

在这种重复中，记错也可以成为有意义的经验。只要其他观察者能够指出寻找步骤中的偏差，错误便不再只是某个人的失败，而成为完善公共指认的方法材料。可纠正性比从不出错更适合衡量一个早期研究传统是否已经形成。

1.4 第一种公共天空

公共性不要求所有人具有同样技能，也不要求他们同时站在同一地点。它首先要求，一项指称能够通过某种可传授的操作被另一个人接续。不同观察者可以有不同判断，但他们必须能够说明差异发生在哪里。只要差异能够定位，就有修正的机会；若名字、时间、地点全部混在一起，争论便难以转化为共同研究。

这种公共性具有层次。村落内部熟悉山形的人可以共享一套指认方法，远方来访者却未必能使用它。把地方参照转换成更一般的方向说明，会扩大可交流的范围；把观测时刻和星群关系一起保存，又会扩大跨代复认的可能。这里没有理由假定所有传统都沿着相同顺序前进，也没有理由把后来形成的统一坐标当作早期知识的唯一价值。

本章所说的第一种公共天空，因而不是一个可指定唯一诞生日的全球事件。它指一类构成：眼前显露能够进入重复差异，重复差异又由共同参照保持。这样的构成可以在不同地方形成，可以中断，也可以只限于少数承担特定任务的人。其存在需要材料支持，而不由现代研究者对“文明起源”的期待决定。

它的创造性在于形成了一种以前未必存在的可持续关系。人不再只是在天光变化中即时反应，而能够等待、寻找、比较、纠正，把一次目击接到另一人的未来行动中。作为这个关系聚合体的天空，已经具有研究世界的初步形态。星体的物理存在与公共对象的构成，在这里必须分别说明，又不能完全割裂。

然而，持续研究还需要回答一个更难的问题：什么时候算同一时刻，多少次重复算一个周期，怎样把观察到的周期变成大家可以执行的日期？星区复认提供了对象的接续，历法则将这种接续扩展为共同时间。两者互相支持，却各有自己的误差和决定。

本章参考文献

[1-1] US Naval Observatory. Celestial Navigation Data for Assumed Position and Time. https://aa.usno.navy.mil/data/celnav

[1-2] British Museum. Tablet 1899,0610.108，MUL.APIN 馆藏条目。https://www.britishmuseum.org/collection/object/W_1899-0610-108

[1-3] International Astronomical Union. The Constellations. https://iauarchive.eso.org/public/themes/constellations/

第 2 章 从月相到历法：时间单位的发生

2.1 自然周期何以不齐

月份最容易造成的一种错觉，是它好像一开始便是一格等待填写的时间。人们打开历书，便能看到整齐的行列，误以为天空也按同样的格子运行。实际上，观察月相、划分日期和安排年度事务，是三个需要协调的活动。历法的困难产生于它们不能自动重合。

太阳带来的昼夜交替、月亮相位的重复和季节的往复，分别提供了不同尺度的经验。美国海军天文台的历法说明将年不能整除为日、也不能整除为朔望月，列为历法复杂性的基本来源。[2-1] 这不是某一文明技术不足造成的偶然麻烦。即使观察越来越准确，也不会把本来不成整数比的周期变成整数比；准确观察反而使不整齐显得更清楚。

以现代平均值作一个说明，朔望月约为二十九天半，十二个朔望月约三百五十四天，明显短于一个季节年。这里的数字只用于揭示问题结构，不代表某种古代历法采用的参数。若坚持每年十二个月而不作协调，月序就会相对于季节移动。若坚持月份始终与季节相近，则必须改变某些年份所含的月份数量，或者放弃月份对月相的紧密追随。

由此产生不同历制，并不意味着其中只有一种真正理解天象。太阳历重视年与季节的对应，阴历重视月相周期，阴阳合历尝试使二者共同得到照顾。它们选择不同的保持条件，因而拥有不同的误差意义。阴历月份沿季节移动，并不当然是算法错误；如果历制没有承诺保持固定季节位置，便不能用这种外部要求判它失败。

这个区分直接触及 E 的含义。某个日期“偏早”或“偏晚”，必须相对于具体关系才能成立：是偏离实际月相，偏离某个季节位置，还是偏离已公布的社会安排？同一观测差异，在不同历制中可能被认作需要修正的误差，也可能是规则允许的变化。历差不是脱离历制而漂浮的一项裸事实。

但这并不意味着历差任由制度宣布。若一种历制承诺某月对应某季节，又在长期运行中持续偏离，便产生了可以观察和比较的张力。制度规定了保持什么，天象变化检验这种规定是否实现。规则与经验既不能互相替代，也不是各自独立的两半；历法的有效性就在二者的持续对应中成立。

2.2 从观测到规则

看见月亮，并不等于知道一个月何时开始。月相连续改变，历日却必须在某处切分。以天文合朔所在日作为月首，和以初见新月作为月首，是两种不同操作。前者依赖计算和时区或地方时间规定，后者还取决于能见条件。它们可以接近，却不能保证每次同日。

初见新月也不是一个只有天体几何的判断。暮光、地平遮挡和大气状态会影响实际可见性。若两个地方报告不同，不能立即推定某方观察失误。报告可能真实描述了各自条件。要建立公共日期，就必须决定如何处理这种合法差异：允许地方月首，接受指定地点的报告，或者采用统一算法。每项选择都改变了何种观察具有决定权。

规则把多次比较压缩为可以重复执行的操作。观察者不必每次重新解决整套天体运动问题，便能够按规则确定下一月的安排。这是知识封装的一种早期形式，但封装不是使天象停止变化，而是规定哪些变化已经被规则吸收，哪些变化需要重新检查。没有这一区分，历法只能在不断临时决定中运行。

置闰尤其显示这种结构。它把累积的周期不协调转化为有规则的调整，而不是简单增加一个无关月份。在一个阴阳合历中，月序、季节位置与置闰相互规定：置闰改变以后月份的编号和年度安排，季节位置又决定何时需要置闰。单独问“闰月是否正确”没有完整意义，必须同时明确历元、月首和季节标志的算法。

现代历书对中国历法的技术说明，便把月首、冬至、中气和闰月分配放在相互联系的规则中。[2-2] 本书使用这一现代说明来展示规则联立的结构，不把其具体标准时间和算法投射到所有古代朝代。中国历法经历过长期变化；某一时期实际执行什么，仍需查阅该时期材料。结构上的连续性不能代替制度史上的逐项核对。

规则还需要处理“哪一种年”。从春分再到春分，从太阳相对恒星回到相近方向，所强调的重复并不完全相同。早期天文学未必用现代术语划清这些年长，但思想史不能把不同任务所用的周期一概混称为一个没有条件的“年”。一项参数为什么被采用，必须放回它希望保持的现象联系之中。

对规则的评价因此至少有两层。第一层是内部一致：给定相同输入，算法能否产生相同日期。第二层是外部对应：这些日期与承诺保持的天象关系是否相符。一个计算毫无歧义的历法可能逐渐偏离季节；一个紧贴地方观察的历法也可能难以提前远距离发布。两种困难不能相互抵销，必须分别解决。

还可以用一个纯粹算术模型看清置闰为什么既必要又不能随意。设月相平均周期为 M，季节年为 Y，连续 n 年的季节长度为 nY。若安排 m 个月，差值便是 mM−nY。寻找小整数

m 和 n，使差值不至于过大，是一类有明确限制的任务。这里没有要求天体周期必须服从整数；整数来自月份必须能够逐个编排的制度。

若周期参数估计有微小偏差，其影响会随使用年数累积。一次月份开始差一小段时间，也许不足以使使用者察觉；经过许多次重复，同一偏差便可能表现为明显错位。因此，长时段记录的价值不同于多次相近时刻的观察。前者扩大累积差异，后者主要帮助辨认短期波动，二者不能互相完全替代。

这种算术模型还解释为什么规则有时采取周期性修正。一个简单周期便于记忆和公布，可让使用者预先知道何时增加月份；更依赖实际天象的方案则能追随不均匀运动，却需要更复杂计算。历史选择涉及可预报性、长期对应与计算负担之间的组合。不能仅凭规则简单，就判断制历者不知道天象复杂。

再考虑日的离散化。一个天文事件发生在连续时刻，而历书必须把它放入某一天。若事件接近日界，时刻估计的微小变化可能造成日期相差整整一天。这是一种由分界规则放大的差异，并非天体运动突然跳跃。比较两部历书时，先检查是否位于这种边界，比立即比较整日误差更有解释力。

对月首而言，地方时间选择也会产生类似效应。相同合朔时刻在不同经度或不同标准时间下，可能被归入不同历日。这说明统一日期需要对地方差异作制度处理，但没有消除地方差异本身。历法从地方实践走向较广区域时，必须维护天文时刻与公共日期之间的转换。

平均周期与实际周期也应分开。平均值适合长期安排，不保证每一次间隔都相等。若用平均数预测单次事件，其剩余差异可能有规则性，并非纯粹噪声。研究者可以据此增加修正，也可以在任务容许范围内继续采用平均表。问题在于误差是否超出所承诺的用途，而不是表中是否出现任何偏差。

如此看来，历法精度不是单一指标。月首准确、季节位置稳定、提前公布容易、跨地域执行一致，各自需要评价。有时提高一项，会增加另一项成本。公共时间的形成，是把这些要求组织成可以持续使用的关系，而不是寻找一个在所有方面都自动最优的数字。

这种结构也为辨认改历提供了更具体的问题：改的是周期参数、日期分界、置闰规则还是公布程序？如果只调整某个平均值，便不应描述为全部宇宙观改变；如果日界或权威观测地点改变，即使天体理论保持，公共日期的构成也可能发生重要转变。

2.3 制度如何介入

一张历书要成为公共时间，不仅要算得出来，还要让足够多的人知道并采用。公布日期、解释闰月、安排祭祀或行政活动，都会使天文计算进入共同生活。制度在这里不是给现

成知识盖章的最后一步，而参与了哪些问题必须被提前回答、允许多大不确定性、如何处理地方差异。

不过，“历法与权力有关”不能成为不再研究算法的理由。权力可以规定一个月叫什么，却不能任意改变太阳和月亮的相对运动。历法的合法性有时来自权威，实用性却仍会受到长期偏差的限制。若把所有改历都解释为政治象征，就难以区分真正改变周期参数的改革与只改变纪年的行政措施。

反过来，把改历纯粹写成精度竞赛，也会遗漏实际实施的困难。改变规则可能打断已有契约、习惯和节日安排；算法更贴近某项天象，并不代表所有使用者立即接受。研究者必须分别问：新法是否提出，是否颁行，哪些地区实际采用，旧法又在哪些活动中延续。这些问题不能由一部新历书的存在一并回答。

现代复活节日期的计算提供了清楚的结构性例子。美国海军天文台区分用于教会计算的表定月相和实际天文满月，并说明儒略历、格里高利历相关规则的差异。[2-3] 这里不需要展开完整宗教历法史，就能看到“满月”在一个制度中可能是一种可计算的规定单位。它与实际天空有来源关系，却不是每年逐次观测的简单抄录。

这类例子让“时间是自然的还是社会的”显得过于粗略。自然周期提供限制，算法提供可执行转换，制度提供共同采用的范围。离开周期，规则可能失去其天文目的；离开规则，连续变化难以化为共同日期；离开采用共同规则的人，历书不能成为公共行动的时间。三者的联系必须由实际运作维持。

在三方程的审查中，S 是可执行的历日系统，D 是朔望、节令与长期偏差的比较，E 是各周期如何与日期规则联立。已有 S 会规定观察者在哪些日子检查月相，怎样记录误差；观察结果又可能要求调整规则与参数。若只得到计算建议而没有实施证据，便只能说改法具备可能性，不能说社会时间已经改变。

历法也可能出现一种误稳定。某项规则长期被沿用，使使用者把与规则不符的报告当作个别观察错误；这些报告因此无法积累成修订证据。这不是天文学从来服从权威的证明，而是一个可检验机制：现成历表可能规定了哪些差异能进入档案，进而影响反证形成的机会。判断它是否在具体历史中发生，仍要比较原始记录与修订过程。

2.4 历法的反作用

历法建立之后，观察的时间组织发生改变。人们能够提前约定某日观察、比较不同年份相同月日的天象，或者检查预期事件是否出现。历书因此不是过去知识的静态容器，也是一种未来行为的安排。它使分散观察获得同一时间入口，扩大了可以相互核对的范围。

但这种安排也带来一个循环：历法用观测形成，新的观测又按历法记日。若后来发现历法偏差，研究者必须把记录日期重新转换，才知道原来究竟观察了什么。不能把历书中的日期直接当作现代绝对时间。尤其在闰月、年首或日界规定不同的材料之间，一项看似简单的对照可能需要整套历制的知识。

这不意味着记录失去价值。恰恰相反，保存了历制信息的记录，可以在转换之后恢复其观察约束。问题在于转换责任必须保留。若现代数据库只留下一个公历日期而删除原纪年、算法和不确定范围，方便检索的同时也可能切断复核路径。越是精确的现代显示，越容易使读者忘记背后存在历史换算。

从成品历书进入研究，容易看到整齐规则；从逐次月相和季节对照进入，则更容易看到为什么必须修改。两种入口应当互相纠正。前者让我们理解公共时间怎样执行，后者让我们辨认执行中哪些差异真正累积为问题。六路径在这里的价值，不在于给古人指定一个心理顺序，而在于避免从一种入口把整个过程看成理所当然。

P2 所关注的张力在此具有明确含义：已有历日显露与所要维持的季节、月相关系不协调，于是比较和计算方法可能改变。但“可能”不能省略。使用者也可能忍受偏差，局部修补，或更换所重视的目标。并非每次历差都产生新宇宙论；相当多的修正可以在既有天体模型中完成。

历法思想最深的贡献，是把不同节律变成可以共同操作的时间关系。它没有消除自然周期之间的差异，而是使差异能够被计算、分配和讨论。由此成立的公共时间既有稳定性，也有修订条件。当天空突然出现预期之外的阴影、尾迹或亮点时，历法又提供了保存异常所必需的时间骨架：这件事发生在何年何月，持续多久，与哪些重复现象相邻。

历法因此也是一种约定未来的技术。别人尚未观察到的月相，已经可以进入共同安排；但安排的可靠性仍须在未来接受检查。它把尚未发生的经验纳入现在的行动，使预测既有实用价值，也承担可以失败的责任。

本章参考文献

[2-1] US Naval Observatory. Introduction to Calendars. https://aa.usno.navy.mil/faq/calendars

[2-2] Explanatory Supplement to the Astronomical Almanac, 3rd ed., Chapter 15, Calendars；美国海军天文台公开章节。https://aa.usno.navy.mil/downloads/c15_usb_online.pdf

[2-3] US Naval Observatory. The Date of Easter. https://aa.usno.navy.mil/faq/easter

第 3 章 食、彗与客星：异常何时能够被保存

3.1 异常为何留下痕迹

天空中的异常有两种不同的生命。一次遮蔽、一道尾迹、一颗此前未见的亮星，可以短暂改变观看者的经验；而它能否成为后世继续研究的事件，还取决于另一组过程：有没有人记录，记录采用什么语言，文本怎样保存，以及后来的人能否把文字重新接回天空。物理事件的发生和档案事件的成立，不能混为一谈。

日食特别容易提醒人们注意这种差别。它是一种具有地点依赖的天象。同一次日月相对运动，可以使某地看到全食，另一地看到偏食，还有一些地方没有可见遮蔽。[3-1] 若后世在两份文献中找到不同叙述，不应先把它们处理为相互冲突。需要先检查地点与时刻，判断差异是否恰恰来自同一事件的不同观测条件。

彗星与客星则提出另一种困难。异常光点是否移动，是否有尾迹，持续多少天，出现在哪个星区，这些特征比一个总名称更能约束后来解释。然而，古代分类不是现代天体物理分类的译名。“客星”表达突然出现的性质，不自动等于超新星；某些含混记录可能容纳不止一种现代成因。

为何这样的记录会被留下？政治、礼仪与占候需要可以使异常获得持续关注。但保存动机与记录效力不同。一个人可能出于今天不接受的解释框架记录了准确日期，也可能出于求知目的写下过于模糊的位置。我们不能依据动机正确与否，直接决定其中每项观察是否可靠。

历史超新星研究正建立在这种分层处理之上。Green 与 Stephenson 对相关记录的研究同时讨论较可靠的历史事件和虚假候选，并未把所有异常星光汇成一份没有等级的超新星名单。[3-2] 这种做法的重要性在于，历史材料的利用并不是“发现一条古文便确认一个天体”，而是逐步收缩可能解释的过程。

从 SDE 看，异常首先冲击的是已有差异序列。平常的星区复认或历日预期，使某次目击显得不寻常。但要形成可研究的异常 S，必须把它放入位置、日期、持续时间等 E 中。一个没有位置的惊异叙述，和一个能够在连续日期中追踪的光点，具有不同的后续研究能力。差别不在于哪一位观看者更真诚，而在于留下了怎样的可复查约束。

3.2 记录的坐标困境

古代异常记录最容易在现代加工中获得过度精确的外观。一条文字本来只说在某星官附近，数据库却需要填入一个位置；原纪年可能存在传抄差别，数据库却要求一个日期。转换如果没有保留不确定范围，原来的含混就会被隐藏在整齐字段中。

位置换算首先需要知道星官指什么。某个名称可能代表一颗主星，也可能代表一组星或一个较宽区域。用现代星图找到名称对应后，仍须解释“旁”“守”“犯”等关系词在该文本中怎样使用。方向和距离语词未必等同于现代望远镜测量的精确角量。文本学上的一字之差，可能改变天文学上的候选范围。

1054 年客星记录提供了清楚的提醒。Pankenier 重新检视相关东亚记录时，讨论了译文、纪日与所报位置的困难，并对若干旧解释提出修正。[3-3] 本章不据此宣布所有争议已经解决，而强调一件事：即使一项历史天象与现代遗迹的联系总体很强，组成证据链的每个细节仍然可能需要重读。

这里必须区分两种置信。第一种是某历史事件与某现代对象相联系的整体可靠性，第二种是某一抄本中特定方向或日期的准确性。整体联系强，不使每个字都无误；个别字存在问题，也不必推翻全部联系。若把二者捆在一起，研究者就会陷入非全真即全假的错误选择。

纪年问题同样如此。原记录的年号、干支、月序和闰月，需要相互校验。如果某个日期在所述月份不可能出现，可以考虑抄误、改写或历制误解，但每一种修正都需要依据。不能只因改一个字就能与现代候选相合，便宣布这必定是原文。为匹配而修正，与由独立文本证据支持的修正，具有不同的证明力。

多个文本也不自动等于多个独立目击。一部后出的汇编可能转录早期材料，几部史书又可能共同依赖已佚来源。若把所有传本作为独立观测累计，会制造一种虚假的多数支持。考察文本源流，因此不是天文学之外的辅助装饰，而直接决定证据数量能够怎样使用。

E 的关键就在这些相互规定处。日期范围决定哪些天体候选能够进入比较，位置范围又决定日期解释是否有实际意义，持续时间则排除某些短暂现象。三者联合收缩候选集合，但每一项都可能带有相关误差。若来源相同，一项抄误也许同时污染多个字段，不能把它们当成独立证据相乘。

同一条记录还可能混合不同时间。初见日期、正式上报日期与后来的总结日期，不应合并成一个事件时刻。Pankenier 的文本讨论提示，1054 年资料中既有天象描述，也有后来提交的解释报告；其中纪日传抄与方向词的问题，需要分别处理。[3-3] 若把报告日期当作爆发现象首次发生日期，模型会在错误时间上寻找物理对应。

从观察逻辑说，初见通常只给出一个界限。若前一夜观察可靠而未见，后一夜见到，才可以把出现限定在较窄时间内；若此前连续阴天，“初见”便不能单独确定出现时刻。同样，最后一次看见不等于物体随即消失，只表明后来缺少满足可见条件的目击。持续时长因此也有不同定义。

可以设计一个不指向具体古人的例子。某记录说亮星持续数十日，另一记录只列出三次目击。前者可能是概括，后者可能是抽样；若直接把三次目击间的空白当作不可见期，就会制造虚假的亮度起伏。恢复时间序列必须先问记录频率，而不是把文字空白填成数值零。

亮度比较同样如此。“如某星”可能表达大致等级，未必是严格等亮；白昼可见则还涉及太阳角距、透明度与观察条件。现代研究可以利用这些描述建立宽约束，但若把它们转换成没有误差条的精确星等，便将语言近似误装成仪器读数。资料的丰富性来自它能限制某些解释，不来自每个词都能变成高精度数字。

原记录中的方向也需要观察姿态。晨昏目标可能很低，附近星区难辨；后来的编纂者又可能将较宽位置表达压缩成一个方向词。说明这种可能性，并不许可任意改文，而是要求把定位作为需要多条支持的推断。史料解释越依赖单个方向字，越应检查其版本和语境。

当现代研究给出候选关联时，最好把证据链拆成几条可以独立阅读的线：日期范围、天区范围、持续性描述和现代对象的物理性质。若某一条很弱，其他线能否仍支持关联？若某一条与候选明显相冲突，应否调整解释，还是降低整体把握？这样的分解使不一致能够被定位，而不必将整段古史当作一团可以任意修饰的文字。

它也有助于保存真正的负例。某条记录即使很生动，若没有可转译位置，就不能在身份确认上发挥与精确星区同样的作用；某条简短记载若日期和地点可靠，反而可能更有约束力。资料的文学感染力与科学判别力，可以很不一样。

3.3 旧解释如何保存有效数据

把古代占候文本拆成“科学事实”和“迷信解释”两袋，虽然方便，却可能损失理解记录如何形成的机会。某种解释制度之所以重要，恰在于它规定了观察者应该关注什么、报告什么、保存多久。今天不接受其因果主张，不代表它在资料生产中没有实际作用。

例如，对异常星是否移动的持续关注，既可能服务于天象分类，也可能服务于占候解释。其动机不改变位置比较本身能够保留的信息。但同一关注方向也可能造成选择：某种现象被反复报告，另一些不符合关切的变化则很少进入档案。留下来的记录不是天空所有活动的随机样本。

这使“没有记载”尤其难以解释。没有文字可能表示没有发生，也可能表示没有看到、没有写下、没有进入官方汇编，或保存下来的文本恰好缺失。只有在了解观测覆盖、报告制度和传存状况之后，沉默才可能成为有分量的反证。把今天未找到的记录一概当作古人未曾看见，会把档案边界误当作天空边界。

同样，解释制度也不应被写成一个无所不能的操控者。即使记录服务于特定目的，持续的天象差异仍可能迫使报告者区分不同情况。一个光点几夜之间明显移动，与长期停在相近星区的亮点，不能无限地用同一观察描述替代。制度选择与现象约束共同塑造资料，任何单向说明都不充分。

本书因此采用两道审查。先问记录在其当时制度中完成什么任务，再问它留下的哪些字段能够被今天的问题重新使用。第二道审查不能取消第一道，因为不知道记录目的就难以判断缺失与选择；第一道也不能垄断第二道，因为原作者的解释并不穷尽文字所保留的信息。

这种再利用是一种新的 SDE 聚合。原始目击、文本传承、现代日期换算、星区重建和物理候选共同进入新的研究活动。新产生的是一种可以支持当代推断的历史事件身份，并不是现代研究者让过去的天体重新发生。区分这两者，才能既承认资料的创造性转化，又不混淆物理历史与知识历史。

3.4 史料的第二次生命

日食记录还能够成为研究地球自转历史的资料。NASA 有关历史日食与自转的说明指出，古代记录在约束长期时间差问题中具有价值，同时其时间和地点必须得到足够确定。[3-4] 这是一种特别鲜明的用途转换：原先可能关注征兆的记载，后来被用于研究一个完全不同的问题。

转换的核心不是把古文翻译成现代术语，而是建立新的约束关系。计算给出在一定自转假设下阴影如何经过地面，史料提供某地何时出现某种遮蔽；二者相互检验，限制可能参数。如果史料只说某年天下有食，而没有可用地点与时刻，这种限制就弱得多。材料价值取决于新问题所需的结构，而不仅是其年代古老。

现代模型在此也有双重作用。它帮助识别候选事件，但可能诱导研究者偏爱能与既有预测相符的读法。因此，可靠研究需要保留被排除候选，说明排除依据，并将文本校勘与天文拟合尽量分开。若先认定目标，再反复改动日期和位置以求相合，拟合结果就会掩盖解释自由度。

从现代遗迹反查古记录，是一种有效入口，却有确认偏向的风险；从全部同期异常记录进入，能够检验哪些材料被留下、哪些没有进入候选名单。这两种路径并用，才可能辨认一

个联系是否过于依赖选择。六路径的作用在这里不是机械轮换六遍，而是使关键判断接受另一种入口的挑战。

一种负结果也可能具有知识价值。若某条客星记载无法可靠定位，结论应是目前不能确定其现代身份，而不是为保留精彩故事勉强指定一个对象。这种克制保存了未来重新解释的空间。档案中的空白并非研究失败的羞耻，它标示现有差异序列和纠缠条件尚不能支持进一步显露。

食、彗与客星因此显示了天文学世界的一个重要特点：可持续知识不只来自重复，也来自对不可重复事件的可复查保存。重复提供常态，异常暴露常态的边界，档案则使短暂偏离能够穿越时间进入新的比较。后来天文学家追踪那些并不突然消失、却不断改变位置的亮点时，也要解决同样的身份问题：变化中的这个对象，怎样仍然是同一个？

最后还须区别记录日期与现存文本日期。一部较晚汇编可以忠实保存早期观察，也可以在转录中改变语境；一份较早文本也不保证作者亲见天象。材料年代提供重要限制，却不独自决定目击资格。需要把成书、抄写、所述事件和作者可接触来源分别排列。

例如，后世编者可能把一次事件的初见、解释和消失分别收在不同栏目。现代整理者若按栏目逐条计数，就可能误认为多次事件；若一律合并，又可能混入其他候选。只有把日期和内容关系一起检查，才能决定哪些条目属于同一档案单元。事件数量本身，也是一项需要论证的结果。

这样的区分使档案中的每个事件都有可检查的形成过程。记录被合并还是保留为多个候选，应当能由读者复查，而不能只依赖整理者给出的编号。

本章参考文献

[3-1] NASA. Why Do Eclipses Happen? https://science.nasa.gov/eclipses/geometry/

[3-2] Green, D. A.; Stephenson, F. R. The Historical Supernovae, 2003.

https://arxiv.org/abs/astro-ph/0301603

[3-3] Pankenier, D. W. Notes on Translations of the East Asian Records Relating to the Supernova of AD 1054, 2006；作者公开

稿。https://www.lehigh.edu/~dwp0/Assets/images/Pankenier%20JAH2%20Final.pdf

[3-4] NASA. Eclipse Predictions and Earth's Rotation.

https://eclipse.gsfc.nasa.gov/SEhelp/rotation.html

第 4 章 行星为何成为“游走者”

4.1 游走者的辨认

今天的人很容易凭一张太阳系图理解行星：它们是围绕太阳运行的一组天体。但古代观察者面对的不是这张图，而是星空中一些明亮光点。要在这些光点中辨认出一个独特类别，需要先找到某种可以反复成立的差异。行星思想的起点因此不在名单，而在追踪。

亮度并不足以提供稳定边界。恒星之间本有明暗差别，大气也会改变所见亮度；一个目标很亮，并不因此成为游走者。关键在于目标相对于其他星的缓慢位移。天空整体在一夜中转动，而某些光点在多夜比较中改变了与周围星群的关系。两种变化的时间尺度不同，必须在观察中被拆开。

这种拆分要求背景具有相对稳定性。“恒星不动”在这里是一种有限时段与有限精度下的工作近似，不是现代意义上否认恒星空间运动。对肉眼连续追踪而言，星群相对形状能够作为参照；相对于该参照持续偏移的目标，才成为不同性质的对象。这一身份是差异稳定形成的结果，而非单夜视觉可直接读出的标签。

大英博物馆保存的金星泥板条目，将相关器物明确列为后世抄本，并说明它包含占候预言。[4-1] 这个事实提醒我们，天象追踪与解释文本可以共存，观察内容与现存抄本年代也须区别。不能把今天拿到的一块泥板，直接视为最初观察者当夜留下的原始日志。

但这类资料仍表明，行星具有能够被长期辨认和保存的出现方式。早晨与傍晚的可见性、消失与再现，可以组成一个有联系的序列。要把分隔开的目击归于同一个目标，需要已有类别，也需要新观察不断检验类别。身份既支持追踪，又由追踪维护，这种循环不是逻辑错误，而是经验知识得以稳定的实际条件。

古代“游走者”的范围与今天太阳系行星分类并不相同。《至大论》传统把日月与五颗肉眼可见行星置于可计算运动的系统中。[4-2] 因此，不能拿今天以太阳为中心的对象名单，去判断古人为什么“误把”日月纳入同类。我们首先应还原当时的分类标准：哪些对象相对星空背景表现出需要单独追踪的运动。

4.2 逆行作为难题

如果一个光点只是匀速朝同一方向穿过星空，建立它的模型相对容易。更棘手的情况是，它在一段时间内改变运动方向，经过近乎停留的阶段，随后再恢复原方向。逆行不是天空中一条肉眼同时可见的轨迹，而是由许多离散位置拼接出来的变化关系。

观察间隔在这里具有决定作用。若间隔过长，短暂转向可能被漏掉；若只观察一两夜，位置误差又可能被误认成停留。要说发生了逆行，需要足够连续的参照，使长期运动方向与局部反向都能够辨认。记录频率不只是资料数量问题，它决定什么类型的变化能够成为对象。

用现代说明式可以表达这个区别：设某目标相对于共同参照的角位置为 λ(t)，逆行对应一段时间里角位置变化的符号与通常方向不同。这个写法不是古代原式，也没有给出造成逆行的物理原因。它只是揭示，逆行首先是一个序列性质，而非一个孤立位置的性质。单个 λ 值无法证明逆行，至少需要比较不同 t。

对于地面观察者，逆行显然可以被描述，却不会自行告诉他地球在运动。一个地心几何模型可以组织这类角位置，后来的日心模型也可以解释它。相同的视运动现象容许不同构造，模型竞争因此必须超过“能画出回环”这一最低要求。还要比较时刻、持续时间、亮度联系及多个天体之间的秩序。

《至大论》将行星逆行作为专门问题处理，说明它已经不是偶然惊异，而成为需要系统运算的现象。[4-2] 这里的思想转折在于，异常运动被纳入规律目标：所谓规则并不一定意味着轨迹表面简单，也可以意味着复杂变化能够按同一方法被反复推算。

如果从 S 开始，我们看到一个已经命名的行星，随后追踪它；如果从 D 开始，我们先审查位置序列能否支持“同一目标在反向运动”。第二种入口会暴露第一种入口可能隐藏的假设：观测间是否误认了目标，背景参照是否改变，缺测如何处理。两种入口的区别，不是改变天空，而是改变什么需要被证明。

逆行序列还要求研究者选择怎样连接离散位置。假设某目标在三个观察日分别位于参照点东侧、较远东侧、稍近东侧，至少显示一次方向变化；但若中间隔了很长时间，就难以确定何时转向。补充观测可以收缩转向区间，却不应该把平滑绘出的曲线当作原来已经看见的轨迹。轨迹包含资料，也包含插接资料的假设。

“留”因此不是绝对停止。它表示在所选参照和可分辨精度下，短时间角位置变化很小。观察间隔更短或精度更高，仍可能测出变化。若把留理解为天体在空间中完全停住，就把视运动性质移成了物理速度判断。古代问题的数学重要性，恰在于可以准确描述前一种性质而暂不决定后一种原因。

亮度变化可以作为辅助，却不能取代位置身份。一个目标变亮，可能帮助下一夜重新找到它；但同样的亮度也可能属于邻近恒星。可靠追踪应使位置与亮度互相核对，同时承认两者精度不同。用其中一个指标独自承包身份，会在云层、暮光或目标接近其他亮点时变得脆弱。

金星的早晚可见性更显出时间空缺的困难。晨空目标与晚空目标不会总在同一夜同时供比较；把它们置于同一循环，需要追踪消失、再现及相对于太阳的关系。这里所强调的是身份确认的逻辑要求，不是为某一文明编造第一次认同两者的日期。相关泥板的存在支持长期关注，却不能替我们回答全部起源问题。

行星表的形成也使缺测拥有新意义。过去没有看见，可能只是无人关注；有了预期位置与持续任务，未见才可能被登记为需要解释的结果。这种转变提高了资料价值，因为观察者能够区分“没有实施观察”和“实施后没有发现”。两种空白在事后看似相同，实际支持的结论不同。

从一个目标扩展到多个目标，比较能力又会发生变化。不同游走者表现出不同的快慢、可见时段和回转范围。类别不是消灭这些差异，而是让它们成为同类内部可系统研究的差异。若所有游走者都被要求具有完全相同的周期，类别反而失去描述能力；共同身份应允许规律上的多样性。

这为后来的模型设计提出了具体任务：某些参数可以各体分别给定，某些关系也许能够共同组织。独立处理越多，模型越容易局部灵活；共同约束越强，跨目标检验越有力。古代数学天文学发展的一个重要方向，就是在这两种要求之间寻找能够工作的平衡。

4.3 共同背景的选择

行星的运动之所以能够比较，需要建立共同背景。星座名称提供一种局部指认，沿黄道组织位置则提供更系统的方式。黄道不是天空中发光的轨道线，而是从太阳周年视运动组织起来的参照。把它用于行星比较，意味着不同对象的路径可以进入相关的坐标问题。

现代术语把黄经、黄纬与相应参考平面和零点联系起来。[4-3] 本章借此说明坐标意义，不能据此认定所有古代记录使用同一套成熟定义。早期材料可能以邻星、星区或出现方位表达位置，坐标化则逐渐使这些局部联系进入更广的计算体系。转换过程中，原来可用的经验并未全部消失，但它的组织方式发生了变化。

背景也不是完全被动的尺子。为了追踪游走者，研究者需要更可靠的星群位置；更可靠的星群位置又使更细小的行星变化能够被辨认。已成目标反过来提出参照精化的要求。这种相互促进，在 SDE 语言中可由 S、D、E 相互规定来审查，但真正的历史改进仍需由星表、观测或仪器材料证明。

例如，一份行星位置表可以显示计算任务已经形成。大英博物馆另有馆藏条目将一块泥板辨认为木星天文表。[4-4] 它支持“相关运算资料曾被制作和保存”这一有限判断，却不足以独自回答制作过程、实际精度和使用范围。材料是什么，与它能支持多少历史推论，应始终分开。

当目标位置与参照星场联合定义视运动时，改变参照会改变运动的表达。一个在地平坐标中迅速向西移动的光点，可以同时在恒星背景中缓慢向东移动。这不是两种互相排斥的事实，而是两个不同差异序列。若忽略参照，就会把同一个“运动”词用于不相同的问题，制造表面矛盾。

这一点也使本书对“新的天文学世界”的解释更加具体。游走者世界不是把几个原有光点换上新名字，而是形成了可以与背景星群分离追踪、具有会合与逆行等序列属性的对象系统。新属性属于这种相互规定的研究构成。天体在空间中的运动并不因此才开始，但“逆行周期”作为能够稳定计算和比较的属性，需要这种构成才能成立。

4.4 从对象到预测

一旦游走者成为可追踪类别，观察任务就发生重组。观察者可以等待它在某时出现，记录它与相邻星体的关系，比较是否符合已有预期。对象身份使观察不再只是对当前天空的浏览，而成为有目标的返回。预测由此将未来目击纳入已经建立的问题。

预测成功的含义仍需仔细规定。能够说出大致出现季节，与能够给出某日的角位置，不是同一种能力；能够预报一次特殊会合，也不代表整条轨迹已经解释。评价古代方法必须使用其实际任务和可用精度，不能因未达到现代标准便否定所有贡献，也不能因一个成功案例便宣布体系完备。

失败则提供了不同层次的诊断。可能是目标误认，也可能是日期换算错误，或参数已经不适用。只有把这些可能分开，偏差才会成为模型问题。若每次不符都被归给天气或记录者，模型便不会承受压力；若每次不符都宣布宇宙论失败，又会忽略操作层面的普通误差。

对行星的持续追踪因此创造了一种双重能力：稳定地保持对象，又允许其位置不断变化。它打破了“只有不变的才是知识对象”的狭窄想法。天文学可以研究的，不仅是固定形状，还包括能够通过差异序列保持的身份。逆行、留、会合这些属性都在这一能力中获得意义。

然而，这样的对象首先是在地方观察中成立。一个地方记录的出现时刻和方向，能否与另一个地方的记录相接，仍须解决纬度、时制与地平参照的问题。游走者已经可以跨夜辨认，却未必已经能够跨地域无损转换。由地方稳定走向更广公共天空，需要新的数学和记录条件。

会合提供了另一种可见联系：两个目标在天空中接近，可以让其相对位置更容易比较。但角度接近并不等于空间接近，视觉相遇也不意味着发生物理接触。古代位置天文学能够研究会合，而无需先知道两个目标的真实距离。

由此可见，一种研究能力可以在另一种知识尚未形成时成立。行星的稳定追踪并不以完整太阳系尺度为前提，却为后来讨论尺度和结构提供了资料。历史上的先后依赖，应按具体能力辨认，不能用今天知识体系的章节顺序倒推。

本章参考文献

[4-1] British Museum. Tablet K.160，Venus Tablet of Ammisaduqa 馆藏条目。https://www.britishmuseum.org/collection/object/W_K-160

[4-2] Ptolemaeus Arabus et Latinus. Ptolemy, Almagest (Greek)，作品与卷次目录。https://ptolemaeus.badw.de/work/19

[4-3] US Naval Observatory. Astronomical Almanac Glossary，黄道及位置术语。https://aa.usno.navy.mil/faq/asa_glossary

[4-4] British Museum. Tablet 1876,1117.1945，木星天文表馆藏条目。https://www.britishmuseum.org/collection/object/W_1876-1117-1945

第 5 章 如何比较不能同时看见的天空

5.1 同一星的不同方位

两位观察者给出不同的天空描述，并不意味着至少有一人看错。北方的人说某星很低，南方的人说它更高；一地记录太阳已经落下，另一地仍处于白昼。公共天文学的任务不是抹掉这些差异，而是说明在什么转换下，它们能够属于同一个研究问题。

地方天空有自己的几何结构。观察者的地平面、天顶和可见星区与其位置相关。亚里士多德在《论天》第二卷讨论地球形状时，已经把南北旅行中可见星体的变化作为论据之一。[5-1] 这里重要的不只是他得出的形状判断，也在于地理位移被转化为天文比较：地方天空的不同，可以共同约束地球的性质。

这提供了一个与简单“统一”不同的路径。若要求每个地方都看见完全相同的高度和升落，就会把真实差异当作观察失败；如果承认地点进入观察关系，差异反而能够成为共享知识的来源。公共性不是一致读数，而是能够解释读数为什么有系统差异。

用现代球面近似说明，在忽略折射等因素时，北天极的地平高度与观察地点的纬度相联系。这个关系不是本章用来替代古代原文的引文，而是帮助读者看清：同一方向结构可以在不同地平框架中具有不同坐标。若星体从某地始终不可见，不能凭另一地的可靠目击要求它也应在当地出现。

山岭遮挡又增加局部条件。几何地平线以上的目标，可能仍被近处地形遮住。天气和天光条件进一步决定实际能见。于是“出现”至少可能指几何上升起、越过地形边缘、或亮度达到可辨认程度。若不同记录使用不同意义，即使日期完全一致，直接比较也会产生假矛盾。

现代导航计算仍需区分几何高度与折射等改正，说明这些条件不是早期知识独有的缺陷。[5-2] 现代方法的优势在于可以明确计算和登记更多转换，而不是彻底脱离地方观察。跨地域天空的建立，始终需要把观察者所在的位置纳入共同解释。

5.2 天象记录的可转译条件

一份记录若只说某星在南方，并不能充分回答另一个地方的研究者。南方可以是一片宽阔区域，描述可能对应不同夜间时刻，星名还可能在当地指向另一组亮点。要进行可靠转译，至少要知道说的是哪个对象、什么时候、从何处、相对于什么参照。

这些条件之间不能完全独立补齐。一个含混星名可以由其出现季节缩小候选；一个模糊时刻可以由与太阳的相对方位得到约束；某个位置解释若使目标在当地不可见，就必须重新检查名称或日期。真正的转换通常是往返校验，而不是按四个字段依次机械填表。

数学坐标提供了一种强有力的转换方式，但坐标本身也需要说明。现代天文术语明确区分不同参考中心、参考平面和位置定义。[5-3] 同一个数值，若定义不一样，并不是同一个位置；不同数值，经过合适变换，也可能指向同一方向。读者应把这种区别视为古今比较的基本纪律，而不是只属于精密天文学的技术细枝。

时间尤其容易被默认。两个地方都说“某日夜间”，未必意味着同一个绝对时间段；历制的日界、地方时刻及换算规则都可能不同。对变化缓慢的星区，时刻误差也许不妨碍粗略复认；对持续时间短的遮蔽或接近，则可能决定是否为同一事件。可接受转换精度随研究问题改变。

因此，资料能够用于一类比较，不保证能够用于所有比较。一份记录可以支持某季节某星可见，却不能支持分钟级事件同步。把它用于前一个问题是合理利用，用于后一个问题则超出证据能力。不是史料等级一经确定便永不改变，而是资料字段与问题需求之间存在具体适配。

本书把这种适配看作 E 的一个实例：记录的意义由对象名称、时间制度、方向参照和用途共同规定。移去其中一项，剩余信息仍可能有价值，但可支持的结论范围会改变。SDE 分析的收益在于要求我们指出这种改变，而不只在材料清单中加上“文化背景”四个字。

5.3 传播不等于复制

一张星图从一个地区传到另一个地区，纸上的线条可以保持，实际使用却可能改变。接收者的地平、季节任务和名称传统不同，原图中的某些区域难以观察，某些称谓需要重新解释。传播并不以完整复制原来活动为常态，它往往涉及筛选、组合与重新校准。

同名也不能单独证明传播。同样显眼的星群可能在不同地方得到相近的形象描述；相似的计算需求也可能产生相近操作。反过来，明确的传播可以伴随术语变化，甚至只留下算法结构而改变解释语言。因此，思想史需要手稿年代、译者记录、术语痕迹与具体运算的共同支持。

《至大论》的后世传播展示了这种复杂性。有关阿拉伯语与拉丁语译本的专业研究指出，文本传统存在混合与校补，某些手稿把另一译本的内容加入页边。[5-4] 这些材料使我们看到，传承者并不只是搬运一件封闭成品。他们也在比较、修正和重新组织可用知识。

然而，一份译本存在，仍不能证明当地观察者已经能够按它计算。理解定义、掌握数值步骤、获得合适仪器、确认地方参数，都有各自门槛。文本接触、运算能力和观测实践应分别立证。如果只用译本年代叙述知识何时“到达”，就会遮蔽到达之后怎样才能真正使用。

传播还可能扩大错误。某个原本只适用于特定地点的表值，被脱离说明抄录后，可能在其他地方继续沿用。材料保存得越整齐，错误有时越不容易显露，因为后来的使用者把其统一格式误当作普遍适用性的保证。地方校验因此不是对经典的不敬，而是使其进入新环境所必需的工作。

从这个角度看，知识的统一性与多样性并不矛盾。统一性是相互可查的定义与转换，多样性是不同地点、任务和传统所保留的差异。和谐并不意味着所有地方采用同一表达，而是各自的结果能够在明确条件下发生联系，并在不能转换时保留边界。

跨地点比较还有一种特别有价值的用途：让彼此不同的记录互为检查。若某星在较南地区能够升得更高，而在较北地区更低，这种有方向的差异比两条完全相同的报告更能支持一个几何关系。差异不再只是需要消除的障碍，而成为推断所需的信号。

但要实现这种用途，需要避免偷偷改变比较单位。甲地记录山脊上方高度，乙地使用理想地平，二者之差就混入地形；甲地以日落为起点，乙地以固定夜间时刻记述，差值又混入时间。转换首先要分离这些操作，否则所谓纬度效应可能只是记录方式不同。

记录中保留原始表达具有长期价值。现代换算结果方便使用，原名称、原纪年和原方向词则允许后来研究者重新解释。两者应并存。只保留旧文字，可能限制跨资料计算；只保留现代字段，则可能让换算假设不可见。一个良好档案应让读者在结果与来路之间往返。

这种往返也改变了“标准”的意义。标准不只是指定一种唯一说法，还应提供从其他说法进入的办法。它可以统一输出格式，同时保存输入的差异。当某条资料不能转换时，标准应允许标注不确定，而不是强迫给出一个伪精确位置。能够容纳边界的标准，才更适合长期知识积累。

因此，全球公共天空不能被倒推为古代已经存在的整体。早期地方传统可以拥有真实而精细的知识，同时尚未与远方传统建立充分转换。思想史要研究的，正是这些联系何时、以何种范围成立，怎样在保留地方差异的同时扩大共同研究能力。

5.4 公共性及其限度

我们可以用一个思想实验辨明公共天空的最低条件。甲地报告某星在日落后不久出现，乙地的档案说同一日期没有见到。若只比较“有”与“无”，两份资料似乎互相排斥；一旦

加入当地天气、地平高度与天光，二者可能同时成立。共同解释保留差异，而不是用权威决定删除哪一份。

另一个结果也完全可能：名称和日期信息太少，无法确定是否同一目标。此时，应保留两条地方记录，暂不合并。承认不能统一，是可靠公共性的组成部分。强行制造全球一致会减少文本表面的复杂，却增加知识结构中的错误。

当已有结果 S 与原有比较方式 D 发生冲突，新的转换关系 E 可能成为解决途径。但并非每次冲突都推动转换。有时没有足够记录，有时测量工具不足，有时研究者并没有把差异视为需要处理的问题。动力解释必须说明这些具体条件，不能把任何矛盾都写成历史自动前进的发动机。

跨地方比较真正增加的能力，是使一个人不必亲自重走所有观察地点，也能检验某些共同关系。记录、规则和转换承担了部分经验的接续。但这种能力没有取消证据责任；每次跨地点推断，都需要知道哪些条件已经被转换，哪些依然留在原处。

天文学由此逐渐形成一种能够容纳差异的公共世界。地方视野不是最终边界，也不是必须抛弃的障碍，而是更广知识的组成部分。把天空方向组织到球面上，将使这种转换获得更强的数学表达；与此同时，球面的可计算性又会带来新的问题：数学结构是否意味着天空真的由物理球壳构成？

如果共同标准只接受某地已有格式，它可能把其他地方有用但尚未转译的信息排除。因此，扩大公共性还包括发展新的转换办法。资料不符合旧表格，有时是资料不足，有时却是旧表格没有为真实差异留下位置；两者应通过实际试译和核对来区分。

本章参考文献

[5-1] Aristotle. On the Heavens, Book II，J. L. Stocks 英译，第二卷第十三、十四节。https://classics.mit.edu/Aristotle/heavens.2.ii.html

[5-2] US Naval Observatory. Celestial Navigation Data for Assumed Position and Time. https://aa.usno.navy.mil/data/celnav

[5-3] US Naval Observatory. Astronomical Almanac Glossary. https://aa.usno.navy.mil/faq/asa_glossary

[5-4] Grupe, D. The Latin Reception of the Arabic Tradition of the Almagest. Qatar Digital Library. https://www.qdl.qa/en/latin-reception-arabic-tradition-almagest

第 02 编

古代几何宇宙的建构

第 6 章 从天球到数学坐标

第 7 章 希腊宇宙论中的几何与因果

第 8 章 托勒密体系：预测成功与解释代价

第 9 章 测量地球如何改变天文尺度

第 10 章 模型怎样跨越语言与时代

第 6 章 从天球到数学坐标

6.1 把天空映到球面

不知道星星有多远，并不妨碍比较它们朝向哪里。这种不完全知识能够成为研究优势：先把距离问题暂时悬置，集中处理方向，天空便可以在一个球面上获得共同表达。天球的数学力量，来自这种选择性的抽象。

以观察者为中心，把朝向各目标的视线延伸到同一个想象球面上，远近不同的天体就对应不同方向的点。球面半径可以任意选取，因为这里保持的是夹角，而不是实际长度。两个目标在天球上相近，只说明方向接近，不足以证明它们在空间中彼此接近；同一方向也可以对应完全不同距离的对象。

这一现代说明使我们能够区分两个命题：天空方向可以用球面数学表达；天体必须镶嵌在一个真实球壳上。前一个命题并不逻辑地推出后一个。古代宇宙论可能把二者联系得很紧，但思想史若不加区分，就会误以为后来的宇宙结构改变必须连带废弃全部天球计算。

《至大论》的结构提供了重要线索：它先处理宇宙论、三角关系和球面天文学，再展开日月、恒星与行星模型。[6-1] 这种安排体现了共同计算工具的基础地位。不同目标的运动需要各自解释，但其方向必须先进入能够共同计算的结构。

天球并不是对视觉的被动描摹。肉眼所见常因地平、天光与心理距离而呈现不均匀的空间感觉；数学球面则要求统一角度、确定极点和参考圆。它在经验中选取可重复关系，并建立新的运算可能。作为科学世界的构成，这里增加的不只是一幅图，而是一组允许比较不可同时看见位置的操作。

与此同时，抽象必须保持边界。若用天球上的角距直接代替天体之间的真实距离，就把为特定任务建立的表征扩张成未经证明的物理结构。天文学的成熟并不只是获得更强模型，也包括知道模型保留了什么、舍弃了什么。方向世界的形成具有独立价值，不需要假装已经解决宇宙深度。

6.2 角距与经纬

要给球面上的一个点赋予数值，必须先有参照。选哪个平面作为零纬度，沿哪个方向计量，经度从哪里起算，都会决定数值的意义。坐标不是点本身穿在身上的标签，而是点与一套定义之间的关系。

黄道坐标把位置与太阳周年视运动相关的参考平面联系起来；赤道坐标则与地球自转轴所规定的赤道方向联系。现代术语对黄纬、方位、天极和不同位置类型的区分，说明这些数字只有在定义完整时才可交换。[6-2] 本章并不以现代规范替代古代材料，而用它揭示坐标思想所需承担的最基本责任。

一个例子足以说明这项责任。把同一目标的经度零点转过一段角度，记录中的经度值会整体改变，但目标方向没有因此移动。若两份星表采用不同零点而未说明，直接相减就会制造系统偏差。反之，数值相同也未必代表相同方向，因为参考圆和历元可能不同。

球面坐标还使两个地方的地平位置能够进入共同关系。用现代符号，目标高度 h、地点纬度 φ、目标赤纬 δ 与时角 H 在理想球面条件下满足 sin h＝sin φ sin δ＋cos φ cos δ cos H。这个式子是现代说明式，并非声称古代作者使用了相同符号；它不包含折射、地形等修正。其意义在于：高度不能脱离地点与时刻被理解。

同一个赤纬目标，在不同纬度或时角会有不同高度。坐标转换因此并非把一个真数换成另一个不真数，而是在不同参照中保存同一方向关系。现代天体导航服务将地点、时刻与高度计算共同处理，正体现了这种联合责任。[6-3] 古代球面方法的历史价值，也应从它实现了多少这类转换来衡量。

在 SDE 分析中，位置显露 S 不是孤立的经纬度对。D 包含测角、比较和转换，E 则是参考平面、零点和目标方向的联立规定。改变平面会改变经纬数值的意义，改变历元又可能使跨表比较需要附加处理。没有这些关系，数字虽然可以抄录，却无法可靠使用。

必须强调，这不使坐标任意化。人们可以选择不同定义，但选定之后，几何关系与观察结果会限制运算。约定性存在于参照选择，非任意性存在于转换一致性和实际检验。把两者混为一谈，既可能把坐标看作天然绝对标签，也可能误以为任何数字都同样有效。

角距本身也值得细看。两个方向之间的夹角不因天球半径的选择而改变；把想象球画大，球面弧长会增加，但中心角保持。这正是它适合比较未知距离目标的原因。若用图上的厘米直接代替角度，而不同图比例不一，就会丢失这种不变量。

球面上的经度差也不总等于两点实际角距。靠近极点时，经线聚拢，两点即使经度相差很大，也可能彼此接近。只有明确纬度与球面关系，才能把两组坐标转成角距。这个细节说明，建立坐标并不意味着所有比较都可以直接相减；坐标是通向几何的工具，不是几何的替代。

用现代单位向量表达，两个方向 u 与 v 的夹角 ψ 满足 cos ψ＝u·v。它把方向比较与实际距离分离，也说明坐标转换应保持相应夹角。若一种所谓转换改变了应保持的夹角，便需要检查它是否用了不同观察中心、不同历元，或根本出现计算错误。这样的关系为资料统一提供了可检验条件。

一个数字例子可以说明方向与尺度为何不能混用。假设两颗目标的视线相差十度。若它们离观察者的距离相同，空间间隔可以由该距离与十度角推算；若一个近、一个远，真实间隔可以大得多。相同天球图容许多种三维分布。因此，天空图上的视觉邻近不能单独支撑物理关联。

赤道与黄道参考也不是二选一的真伪争论。面向每日升落问题，赤道关系具有便利；面向太阳周年位置及行星运动，黄道关系具有便利。不同任务选择不同入口，可以减少计算复杂性。只要转换明确，两套系统能够互补，不必把某一套提升为唯一真实天空。

然而，便利也会引导注意。以某参考平面组织资料后，偏离该平面的变化更容易作为一类问题集中显露，而别的结构可能不那么直观。坐标选择不是决定事实的任意行为，却参与了哪些规律容易被发现。研究者需要在计算效率与探索盲区之间保持自觉。

时间变化给坐标增加另一项责任：目录中的数值对应哪个时段。即便在短时间内将星场当作稳定背景，跨越很长时间仍可能遇到参照方向变化。若不注明历元，后来的比较者无法区分目标变化与参照变化。在实际史料中，这种区分达到什么程度，需要检查星表说明，而不能从现代习惯推定。

从记录格式看，历元的出现把坐标从永久名称变成带时间条件的数值。一个名字可以沿用，而位置表必须在时间关系中解释。对象同一性与数值同一性于是进一步分离：同一个对象可以拥有不同坐标值，不同对象也可能在某时接近同一方向。这种区分为后来的精密位置天文学准备了思想条件。

坐标误差还可以具有共同结构。如果整个星表的零点偏了一小段，许多目标会朝相同方向偏移；如果某一目标误认，则可能表现为个别离群。比较多颗星而不只看一颗星，便有助于区分共同参照问题和局部记录问题。这是一个现代诊断说明，不是对古代统计术语的追认，但其几何基础已经包含在共同坐标之中。

因此，坐标化的主要收益不只是把名称换成数字。它允许不同差异具有可比较的形态：整体平移、旋转、个别异常或随时间变化。问题从“这张图像不像天空”转为“哪些关系系统地不符”。这种转变扩大了纠错能力，也使新的仪器与数学要求能够被清楚提出。

6.3 模型之外的可用工具

一种宇宙论被修正，为什么它的许多数学方法仍可继续使用？答案在于，不同知识部分承担不同解释责任。球面角度负责组织方向，天体轨道模型负责说明位置怎样随时间改变，物理宇宙论还要说明为什么如此运动。它们互相联系，却不具有完全相同的成败条件。

地心与日心构造可以在某些任务中使用相同角量和星区参照。当地面观察者描述太阳的视位置时，并不需要每一句都重述地球围绕太阳运动。参照的便利性与宇宙中心的物理主张可以分离。若把使用地心坐标直接当作相信地球静止，就会混淆表述工具与因果承诺。

这种分离也使历史评价更公平。古代球面方法即使服务于后来被替代的宇宙结构，仍可能有效组织观测。它的价值不能只从最终模型的真假追溯，也不能因仍有应用便宣称原宇宙论全部正确。需要逐项检验：哪些关系在转折后保留，哪些参数要重估，哪些解释必须放弃。

数学表征的持续性还依赖教育。一个几何命题可以被证明，却不保证使用者会把实际观察转成正确输入。仪器读数、单位换算和定义理解，都需要训练。数学工具的传承因此既是命题的传承，也是操作能力的传承。只有前者，没有后者，文本可能被保存而方法难以运行。

本书所谓“世界构成”，在这里并不要求每次都产生全新宇宙。球面坐标可以生成一种可计算方向世界，它既能嵌入不同宇宙模型，也能在模型更替中保持部分稳定。发生具有层次：有时是旧关系精化，有时是对象单位改变，有时是不同系统新建立联系。不能把所有变化都写成同样强度的革命。

竞争解释可能把这一切概括为抽象表征的成功。这种说明非常有力。SDE 若要增加价值，必须进一步指明：表征怎样通过反复测量成为稳定操作，已成坐标又怎样规定新的差异采集，以及参考关系为何需要维护。若只把“表征”换成“S”，而不展示这些相互规定，就没有额外解释收益。

因此，本章不把三方程当作另一套球面公式。三方程审查的是科学活动的结构，球面公式执行的是特定数学计算。前者不能替代后者，后者也不自动解释其历史形成。把两个层次分清，才能让方法论帮助理解数学天文学，而不抢占它已经能够完成的工作。

6.4 坐标的历史回写

一旦观测习惯于记录角量，许多以前难以提出的问题便变得清楚。不同年份的星位可以比较，不同星表的系统差可以检查，某些缓慢变化可能从名称的稳定外壳中显露出来。坐标化使差异进入可累计形式，这是一种实质能力增长。

但它也改变了记录中的可见与不可见。只留下位置数字，可能丢失天光、地平遮挡或观察者对目标身份的犹疑。一个十分规范的星表，可能比叙述记录更适合计算，却不一定保存更多关于观测过程的信息。知识形式的精炼总有取舍，不能把数字化本身等同于信息完整。

坐标的反馈还体现在仪器要求上。既然目标被定义为某个角位置，测角装置的刻度、对准与零点就成为研究能力的一部分。新位置 S 要求更细的比较 D，比较又要求调整参照和仪器 E。但这些要求不必立即实现：材料工艺、观测时间和误差意识都可能限制改进。

我们也应注意一种反方向情况。整齐坐标可能赋予旧数据超出实际精度的权威。若数表显示了多位数字，读者容易误以为原测量同样精确。历史上某个参数究竟来自观察、计算还是转录，需要单独查明。数值的书写精度与观测的可信精度不是一回事。

从天球图像进入研究，容易追问古人想象了怎样的宇宙壳层；从坐标运算进入，则能发现不同宇宙观共享的方法资产。两条入口共同使用，可以避免把数学工具完全物理化，也避免把古代宇宙观完全消解为方便计算的假设。具体作者究竟如何理解模型，仍需阅读其自然哲学和数学论证。

球面坐标最终留下一个富有生产性的张力：可计算的方向结构已经强大，而造成这种结构的宇宙原因仍可争论。天文学不必等到所有原因清楚，才开展有效计算；但计算成功又会不断诱发关于原因的要求。古希腊宇宙论中的几何与因果，正是在这两个要求之间形成多种路线。

这种转变还可以用地图与目录的差别说明。地图让整体形状较易看见，目录让单项位置较易比较；一幅图可以隐藏小偏差，一张目录则可能使整体结构不直观。二者结合，才更容易同时发现局部误记与系统变化。显示形式参与研究，却不能代替被显示关系的检验。

当星位从口述进入目录，读者不必掌握原来的完整故事，也可以进行某些计算；但若目录没有保存星名对应与辨认说明，数字也可能失去目标。抽象提高可迁移性，同时要求维持与实际显露的联系。数学坐标的成功，正来自这种联系被可靠保留，而不是因为数学能够完全离开观察。

因此，天球模型的稳定并非静止不变的封装。它可以不断更换参考对象、精化定义和增加改正，同时保留方向比较这项核心能力。旧工具在新体系中继续运行，是知识发生的一种重要形式，不应仅以是否出现全新名词判断创新。

本章参考文献

[6-1] Ptolemaeus Arabus et Latinus. Ptolemy, Almagest (Greek)，卷次及模型范围目录。https://ptolemaeus.badw.de/work/19

[6-2] US Naval Observatory. Astronomical Almanac Glossary. https://aa.usno.navy.mil/faq/asa_glossary

[6-3] US Naval Observatory. Celestial Navigation Data for Assumed Position and Time.
https://aa.usno.navy.mil/data/celnav

第 7 章 希腊宇宙论中的几何与因果

7.1 同心圆的吸引力

为什么圆周曾经如此有吸引力？如果只用“古人迷信完美”回答，就无法理解它为何能同时支持数学构造、自然哲学与实际计算。圆周无须终点便能继续，旋转能够重复而不离开整体位置，球形又将各方向纳入一个统一结构。这些性质，使圆与球不仅是美感对象，也是组织天空规律的资源。

在亚里士多德《论天》中，运动形式与物体性质紧密相连。直线运动有趋向中心或远离中心的区别，圆周运动则具有另一种地位；第一卷由此展开对天体性质的论证。[7-1] 我们不必接受这套自然学，仍应注意其推理责任：他并不满足于用一条曲线追踪位置，而要说明何种物体为什么适合某种运动。

这种责任决定了什么算作满意答案。一个算法若准确给出位置，却要求物体进行自然学上难以接受的运动，可能仍然不够；一个宇宙图景若在原因上统一，却无法解释可重复的角位置差异，也会承受压力。几何与因果的成功指标可以重合，也可以分离。

同心构造的吸引力正在于，它尝试用围绕共同中心的规则旋转解释复杂天象。把多个运动叠合，可以使简单成分产生复杂结果。这里的“简单”指成分满足统一规范，并不意味着最终轨迹一定简单。若只数模型里有多少圆，而不问圆承担什么关系，就无法公平评价其解释经济性。

然而，几何上的构造可行不等于物理上的实现清楚。若多个球面或圆周共同决定目标运动，它们的联结怎样理解？为什么采用这些方向和周期？怎样与不同天体的明暗和远近相容？每一种成功都留下新的责任，而非自动解决全部问题。

在 SDE 框架下，圆周规范可被看作 E 的一部分，但不能只把它列为背景信念。它实际限制哪些 D 能够被采用：观察差异需要用什么运动组合表达，哪些构造被接受，哪些即使方便也受怀疑。同时，某种组合的预测结果 S 又可能促使研究者重新理解规范。规范与操作相互规定，才构成这里的纠缠。

这一分析也提醒我们，不能把现代偏好的最小参数数目直接当作古代唯一评价标准。古代作者可能更重视运动形式一致、物体性质相容或宇宙整体秩序。现代读者可以批判这些标准，却应先明确它们如何起作用，否则所谓“解释失败”就只是把另一套任务强加给历史对象。

7.2 物理原因与位置算法

一张位置表回答“某时朝哪里看”，自然哲学则可能回答“为何这种物体如此运动”。两者并非毫无关联，却不能相互替代。前者要求可计算和可核对，后者要求原因与其他自然现象协调。天文学长期处在这种双重要求之中。

亚里士多德《物理学》第二卷对自然物的讨论，将运动与静止的内在原则作为重要区分，并区分多种原因。[7-2] 把这种原因观简单等同于现代力学的作用力，会造成误读。他关心的解释范围包含材料、形式、变化来源与目的等问题；现代人只问受什么力，不能穷尽其“为什么”的含义。

这也说明，古代几何天文学与自然哲学之间的张力，不只是“科学对抗哲学”的冲突。数学研究本身需要判断何种结构可取，哲学论证也会调用观察。真正的区别在于它们怎样分配证明责任，而不是一方纯有数据、另一方纯无经验。

例如，球形地球的论证可以使用月食阴影和纬度变化等观察，同时嵌入自然运动的解释。[7-3] 多种论证在同一个结论上汇合，不代表它们具有相同强度。如果其中某个自然学论据后来失效，其他几何或观察支持仍可能保留。分析一项历史结论，应拆开它的支持结构，而不是将整套哲学一起判真或判假。

同样，一种位置算法可以在有限范围内有效，却不唯一确定宇宙原因。有限角位置资料往往允许不同参数或构造给出相近结果。于是，研究者需要额外要求来选择模型：简洁性、运动规范、与其他现象的联系。模型竞争不仅比较已经算出的数值，也比较愿意承担怎样的解释代价。

一个现代思想实验可以说明这种差别。设两个几何模型在现有观测精度内产生相同位置。若模型甲把各行星彼此独立处理，模型乙建立跨行星的共同关系，二者在拟合上可能相同，在解释组织上却不同。但若共同关系尚无独立检验，乙的统一也不能仅因形式漂亮便获得决定性优势。

这意味着，“预测成功”与“原因正确”之间既不能画等号，也不能断绝关系。预测成功是原因论证的重要支持，却需要排除替代解释；原因框架则引导新的预测，但必须接受实际检验。只有持续往返，二者才可能互相校正。

本书的三方程在这里承担的不是裁判所有古代思想，而是要求说明这种往返。新位置显露如何由观测差异与模型关系成立？已得结果怎样规定下一轮比较？比较又怎样使解释规范需要维持或改变？如果没有发现实际变化证据，就应保留长期并存与分工的可能，不把张力预写为革命。

7.3 竞争宇宙形态

古希腊宇宙论并不是一支意见一致的地心队伍。《论天》第二卷在论述地球位置和运动时，记录并批判了地球不居中心、中央为火等主张。[7-3] 这些争论本身已经足以反驳“古人都自然相信同一宇宙”的概括。共同的观察并未生成唯一的解释形态。

不同候选首先改变了什么被视为需要说明。如果地球静止，天空的日运动需要归给天体或天球；如果允许地球运动，就必须解释人为什么没有直接感觉到相应运动，以及观察现象如何保持。每个方案都重新分配解释负担，而不是只改变图上的一个中心点。

我们今天知道地球在运动，不能因此把所有古代地动方案都当作现代天文学的缩小版。某种方案承认地球移动，仍可能在中心、轨道、物质或证据上与后来理论相差很远。用结论关键词寻找先驱，容易掩盖思想结构的差异，也容易使失败方案仅因“像现代”便得到额外宽容。

反过来，地心方案也不是一个没有内部差别的概念。有的着重自然位置，有的着重运动几何，有的把不同计算工具组合起来。评价其中某一模型，必须明确所评的是宇宙的物理中心、计算的参考中心，还是某条轨道的几何中心。三个“中心”可以在历史论证中相关，却不应被现代叙述含混使用。

在比较模型时，观察证据的强度还取决于可探测条件。若某方案预期一种极小变化，而当时仪器不足以测出，没有看到便不能简单证明变化不存在。另一方面，研究者也不能无限放大未知距离来保护假说，而不承担新的检验责任。模型可以在观测界限内生存，但其自由度需要被说明。

这种情况提示一种重要的历史判断：一个方案未被接受，不必意味着所有人都被权威蒙蔽；它可能缺乏足够的计算细节，也可能尚未给出比竞争者更有力的可检验联系。社会条件当然重要，但需要与模型在当时的可工作性共同考察。

从本书的角度，新的天文学世界不是靠宣布一种宏大图景便完成。图景必须进入可持续差异操作，使对象身份、属性联系和预测条件形成稳定聚合。一个宇宙猜想可以极富启发，却尚未形成可运行研究系统。承认其思想价值，与夸大其实践完成程度，是两回事。

因此，思想史应保留未展开的可能性，也保留未成功的原因。某些方案的存在表明古代问题空间比后来叙事更宽；某些方案未形成长期研究，则说明可能性与能力之间有距离。创造性的发生，需要选择之外的物质、数学和检验条件，而不只是思想家脑中的一瞬灵感。

《论天》对地球静止原因的批评，可以使上述分歧更具体。亚里士多德讨论过把地球想作浮在水上的解释，并追问：如果水能够承托地球，水本身为什么保持在那里？[7-3] 这项批

评并不是新的位置观测，而是检查原因是否只是把原问题移到另一物体上。它要求解释停止于真正有说明力的关系，而非无限添加支撑物。

他还讨论以地球平坦和空气支撑解释静止的方案，并区分形状与体量在论证中实际承担的作用。如果一个解释主要依赖大量空气不能逸出，那么仅仅把地球形状改成球形，是否就会破坏机制？这一追问展示了概念分析的力度：作者试图辨认真正起作用的前提，而不是只看结论是否与自己一致。

这些批评不能直接视为现代流体力学，却值得作为解释纪律来阅读。一个模型说原因是“形状”，计算或类比中起作用的却可能是“大小”；若二者没有分开，解释就会在术语之间滑动。天文学思想史中的不少争论，也需要这种前提审查，而不仅是增加一个新观察点。

另一条讨论涉及旋涡。即使旋涡曾经把某些物质带到中心，仍须解释为什么现在普通重物继续朝地面运动。宇宙形成的起因与已经形成之后的持续运动，不是完全相同的问题。亚里士多德在批评中正要求这两种说明相互衔接。[7-3] 这也提醒我们，起源叙事不能自动代替运行解释。

从本书的方法看，这一区别特别重要。描述某个 S 怎样首次成立，与描述它怎样被持续维持，使用的差异序列可能不同。一个初始聚合机制可以结束，而后续稳定依靠另一组关系。若把“曾经发生”当作“所以一直如此”的充分原因，发生学反而会退化为起点故事。古代争论在这里提供了可以对今天方法反向施压的资源。

亚里士多德对中心对称性的批评也有独特意义。他转述某种主张：地球在中心，朝各方向没有优先性，因而保持不动。随后追问，若论证只依赖居中，那么把另一种物体放在中心，是否也应静止？如果不同物体表现不同，就需要加入物体性质。[7-3] 这不是现代对称原理的完整表达，却清楚区分了位置关系与对象性质的解释职责。

今天可以从中提炼一个模型检验问题：某个解释是否过度宽泛，以至于对本应不同的情况给出相同结果？若是如此，它遗漏了有区分力的条件。用 SDE 语言说，仅有某种位置 E 而没有相关差异 D，可能不足以规定所讨论的 S。但这只是本书的分析映射，不等于亚里士多德已经提出三方程。

在地球形状上，他又把几种支持并列：部分物质趋向中心的自然学论证，月食时地球投影边缘的形状，以及南北旅行中的星空变化。[7-3] 三条线并非同一证据的重复。它们分别从物体运动、遮蔽几何和地点变化进入，因此可能在理论修订中具有不同命运。

这组材料可以用来纠正一种常见评判：只要结论正确，所有论证都值得接受；或者只要某些前提错误，整个结论便没有知识价值。真实的支持网络比这更复杂。不同证据可以在同一结论处汇合，又在更细问题上发生冲突。研究者需要保留这种内部层次。

也正因为如此，古希腊宇宙论不能只被写成等待近代替代的错误清单。它形成了对原因、类比、前提和反例的多种要求，其中一些与具体宇宙结构一同被修订，一些则继续构成理论研究的纪律。承认这些方法资产，并不要求降低对其经验局限的审查。

7.4 哲学约束如何回写计算

一种运动规范一旦被接受，会改变研究者的搜索方式。他们更可能尝试哪些组合，更少考虑哪些轨迹，以及如何判断一个近似方案可以继续。这就是哲学约束对计算的具体作用。它不必表现为禁止命令，也可以表现为熟练方法的默认边界。

这样的边界具有双重效应。它减少无穷候选，使研究可集中推进；也可能排除后来证明有用的构造。若没有任何规范，每次拟合都可能任意变形，研究难以积累；若规范永不可改，持续偏差又无法转化为新方法。可靠研究需要一种既能保持又能重新打开的约束。

这里的 P2 张力可以明确表述：位置计算的结果 S 与关于可接受自然运动的 E 不协调，促使构造方法 D 发生改变。但历史结局并不只有一种。研究者可能修改几何，使其更符合自然规范；可能暂时把计算与物理解释分工；也可能调整规范本身。哪条路线发生，要由具体文本和后续工作判断。

若从现代残差大小开始评价，会看见某些模型在数值上的优势；若从古代因果标准进入，则会理解为什么同一模型仍受批评。两条路径的差异不是让评价相对化，而是把不同评价对象明确分开。只有知道争论双方试图解释什么，才能判断分歧是真实困难还是互相错位。

SDE 本体论对此提出的重构是：科学世界由显露、差异运行和特征纠缠共同构成，因而计算结构与解释标准也会参与对象属性如何稳定。这个主张并不证明某一古代宇宙论正确，也不要求历史人物已经使用 SDE。它需要展示的，是哪些属性在某套运行中具有明确意义，换一套运行后哪些关系必须重新组织。

在这一意义上，几何宇宙的创造不是把已知天空再描绘一次。它使圆周、球面、中心、角位置和自然运动进入能够相互要求的体系。其力量来自这种体系的连续工作，其局限也产生于同一体系对候选方式的约束。创新的条件和保守的条件，有时正是同一组已成能力。

托勒密式数学天文学将把位置预测推进到更系统的程度，也把这种张力表现得更加具体。一个模型可以提供强大的计算资产，同时在均匀运动和物理实现上留下争论。理解这一点，便不必在“伟大成就”和“错误宇宙”之间二选一；历史贡献可以在明确边界中成立。

还可以注意《论天》第二卷对天体运动产生和谐声音的批评。亚里士多德没有因为一种解释优美便接受它，而追问其关于巨大声音及人类听觉的后果。[7-3] 无论我们如何评价他的

具体反驳，这都表明古代自然哲学内部已经区分解释的吸引力与其证据责任。将全部古代宇宙论归入未经检验的美学，会遗漏这种内部批评。

不过，批评者也可能把自己的前提视作无需再打开的常识。例如，他从自然运动区分出发反驳某些地动方案，论证力度依赖自然运动理论本身。一个论证可以对对手提出尖锐问题，同时留下自身未充分检验的基础。这种不对称不只属于古人，也是理论研究普遍需要警惕的风险。

公平对读因此应当采用同样尺度：既查候选方案是否解释观察，也查裁判标准依据什么成立。若只审查异己模型而让自己的自然学退出检验，争论就会固化；若所有标准都不允许暂时使用，研究又无从展开。可持续的理论工作需要暂时采用标准，并保留重审其适用范围的机会。

本书对古代论证的评价，也应接受同样要求。SDE 解释必须说明其新增加的区分在哪里，不能因它能够把任何争论写成张力便宣布解释成功。这里真正有内容的区别，是几何位置、自然原因与论证前提各承担什么责任，以及它们如何在具体文本中互相支持或抵触。

标准本身进入检验，才使理论竞争成为开放研究。它并不要求每次从零开始，而要求已成判断仍有能够重新讨论的来路。

本章参考文献

[7-1] Aristotle. On the Heavens, Book I，J. L. Stocks 英译。https://classics.mit.edu/Aristotle/heavens.1.i.html

[7-2] Aristotle. Physics, Book II，R. P. Hardie 与 R. K. Gaye 英译。https://classics.mit.edu/Aristotle/physics.2.ii.html

[7-3] Aristotle. On the Heavens, Book II，尤第十三、十四节。https://classics.mit.edu/Aristotle/heavens.2.ii.html

第 8 章 托勒密体系：预测成功与解释代价

8.1 模型为什么能预测

如果把托勒密体系画成许多小圆套着大圆，很容易得到一种印象：只要资料不合，就随意再添一个圆。这样的图像虽然流行，却没有解释这套数学天文学为什么能够长期工作。一个可以任意迎合过去资料的模型，未必能够在未来时刻提供有用位置；托勒密体系的历史力量，必须从其实际计算结构理解。

《至大论》将球面方法、日月理论、恒星和行星运动组织成一个可运算整体。[8-1] 它不是只有一张宇宙图。观察日期需要转换为相对于历元的时间，平均运动把时间转换成角量，几何修正再把平均量转换为可与观测比较的位置。每一部分都要求明确参数与计算步骤。

本轮与均轮可以先用一个简化的现代几何说明。令一个小圆的中心在较大圆周上运动，目标又围绕小圆中心运动；观察者看到的是两种位移共同形成的方向。在平面直角坐标中，可写成 $x=R\cos\alpha+r\cos\beta$，$y=R\sin\alpha+r\sin\beta$。这里的 R 与 r 是两个半径，α 与 β 是随时间变化的角度。该式是现代说明，不是托勒密原文，也未包括其全部行星修正。

这个简单模型已经能够说明一个关键点：即使两个组成运动各有规则，合成后相对于观察者的方向也可以出现快慢变化乃至局部反向。因此，本轮不是专门为遮掩错误而发明的语言，它是一种将复杂视运动分解为可计算成分的数学方法。它是否有效，要比较它在特定参数下产生的整段位置序列。

偏心构造进一步把观察者与轨道几何中心分开。物体可以相对于圆心以规则方式运动，却从偏离圆心的位置看来时快时慢。这里必须区别实际沿圆周的运动规定与观察方向的角速度。没有这种区别，“匀速”二字会让读者误以为任何中心看到的变化都相同。

均衡点使这种区分更加尖锐。就外行星模型的典型构造而言，本轮中心到均轮中心的距离保持固定，而相对于另一点的角度按均匀方式改变。美国国会图书馆展示的《至大论》拉丁手稿相关页，正呈现了均匀运动中心与等距中心分离的结构。[8-2] 它保留了圆形轨迹，却改变了匀速旋转所依据的中心。

从现代几何角度看，若 Q 为均衡点、C 为圆心、P 在圆上，条件 CP 长度固定与 QP 方向均匀旋转，通常不会使 CP 方向也均匀旋转。三个点不重合，两个角度就不具有相同时间变化。这是明确的几何后果，不需要先诉诸某种审美判断。模型的数值收益与运动规范的困难，都由这组关系产生。

8.2 “拟合”不等于“乱编”

任何位置模型都需要从资料确定某些量。把这种过程叫拟合，本身并不是批评。关键在于参数受到什么限制，能否同时处理多次观测，是否对未用于确定参数的情形提供正确预期。一个模型的解释责任随着自由度增加而增加，但自由度的存在并不等于任意性。

考察一个只有半径比和初始相位的简化本轮模型，便可理解约束。改变半径比会同时影响视运动幅度和某些转向特征；改变相位会移动这些特征出现的时间。参数不是只修正一个点而不碰别处。若用一个参数改善某段轨迹，却使另一段偏差扩大，这种冲突就会限制可接受取值。

这正是 E 需要具体化的地方。半径比、相位、观测时刻与允许误差共同决定模型是否成功。它们不是模型旁边的四项背景，而是在预测过程中互相约束：时间解释改变，所需相位会变；观测精度提高，过去可忽略的差异会变成系统残差；运动规范收紧，某些原来可用的参数化方式会被排除。

公平比较应保持任务相同。不能拿某模型预报黄经的成功去抵销它对距离变化的失败，也不能用后来更精密的观测直接宣判当时方法没有价值。一个体系可能在某种角位置任务上有效，在另一种物理量上不充分。有效域应当细分，而不是压缩为“对”或“错”一个字。

天文表又增加了一个容易忽略的层次。几何模型说明怎样计算，但使用者实际常通过表格获得结果。表格的步长、插值方法和数字精度，会影响最终误差。即使几何关系没有改变，数表抄误或舍入也可能造成不同结果。因此，对模型的检验应尽可能辨明误差来自构造、参数、表格还是使用过程。

以一般线性插值作现代例子：若表格给出两个相邻角度对应的量，使用者可以按中间角度所占比例估计中间值。这样的近似节省计算，却只有在函数变化足够平缓、间隔合适时才可靠。增加表格密度可以减少一类误差，却增加抄写、存储与核查负担。计算传统的可用性，是精度与操作成本共同形成的结果。

托勒密的弦表便是这种计算基础的一部分。现代对其数表的研究指出，几何论证、表中数据与插值项之间存在需要细查的关系。[8-3] 在这里，史学问题不是赞叹数字多么精确，而是查明这些数字怎样产生、怎样在后续运算中使用。数表使理论获得执行能力，也让理论承受来自实际执行的新检验。

这使托勒密式体系超出一套宇宙意见。它是一种能够训练使用者、生成预报、比较位置的研究能力。该能力在地心构造中形成，但并不能被“地心”二字完全概括。若不把计算资产拆出来，就无法理解后来的批评者为什么仍然需要学习并使用其方法。

8.3 被保留的计算资产

一套理论最持久的部分，有时并不是它对宇宙中心的断言，而是它提供的工作方式。明确历元、区分平均量与修正量、把几何关系编成数表、通过已知关系检查抄误，这些方法能在新的解释框架中继续发挥作用。理论变迁因此既有断裂，也有重新组合。

我们可以把这种连续性分为几个层次。符号和名称可能保留而意义改变；数学关系可能保留而物理解释改变；计算步骤可能局部保留而参数全部重估。只有具体比较，才能知道所谓继承究竟发生在哪一层。笼统说后人“沿用托勒密”，既可能低估创新，也可能夸大不变性。

弦与角度之间的关系尤其适合说明这一点。对于半径 R 的圆，现代形式给出弦长 $c=2R\sin(\theta/2)$。这是一条几何关系，与太阳系采用哪个中心无关。使用者可以把弦表转换为另一种三角表达，却仍承接同一计算问题。方法资产的迁移，不要求旧体系所有本体假设一并迁移。

另一方面，具体数表并非只由这条普遍关系决定。它还需要选取半径单位、角度步长、保留位数与插值规则。换一种表达后，如果不处理这些约定，数值就无法直接复用。因此，保留数学结构与复制表格不是同一回事。新的计算传统常常要重新制作旧关系的可执行形式。

后世译本中发生的重算，给出了很好的例证。Zieme 对克雷莫纳的杰拉德拉丁译本所作研究认为，若干基础表经过按文本方法重新计算或调整。[8-3] 本章只取这一有限发现，不把所有译者都描述为同样的改革者。其意义在于：翻译可以包含对计算一致性的检验，而不是词语替换完成后便结束。

托勒密体系还留下了一种研究规模。太阳、月亮、恒星与各行星不是完全孤立的故事，它们共享球面工具、时间框架和观测比较。某些成功可以为其他研究提供基础，某些参数误差也可能跨多个计算环节传播。一个体系之所以强大，正在于它把局部能力组织成可互相调用的整体。

用 SDE 表达，这种整体不是把多个 S 简单相加。某一表值作为另一运算的输入，改变了后者可以处理的差异；共同时间与坐标又使不同结果能够比较。E 因此包含计算之间的依赖结构。调整一个基础参数可能影响多个输出，局部修正的意义需要在系统中检查。

这也带来稳定性的代价。越多任务依赖同一套基础表，更换它的成本越高。研究者不只是比较两张宇宙图，还要考虑已有数据、训练和预报能否转接。模型长期存在可以来自确实有效的工作能力，而不必全部归于惰性。与此同时，这种能力也可能让体系内部的问题被较长时间局部处理，而不促成整体重建。

因此，本章把托勒密传统理解为一种带有特定边界的强研究系统。它能够保存成果、复用运算并扩大比较；这些优点并不消除其不足，反而使不足以更明确的形式出现。当角位置预测与物理尺寸或运动规范发生张力时，批评者面对的是一个可工作的对象，而不是一个可以轻易丢弃的空壳。

为了更明确地理解本轮模型的能力，可以继续分析前述现代简化式。若 R 固定而 r 减小，小圆所造成的方向扰动会减小；若 r 相对于 R 较大，合成方向可能出现更显著的局部转向。这种连续变化说明模型不是只会画一种回环，它通过受约束参数表达一族可能轨迹。观察所需确定的，是这一族中哪些成员与资料相容。

但是，参数相容不必唯一。只观察短短一段轨迹，几组不同参数可能给出几乎相同的位置；观察覆盖更长时段，尤其包括转向与不同相位，才能更有效地区分。模型的可确定性因此取决于观测如何分布，而不只是资料点有多少。许多集中在同一阶段的点，可能不如少量跨关键阶段的点有约束力。

这也能解释为什么特殊位置在几何推算中有用。当目标、太阳方向或某个模型中心满足特定排列时，复杂合成中的某项影响可能被简化。研究者选择这些时刻，不一定是挑选有利数据，也可能是为了分离参数。关键是所用简化条件是否真实成立，以及其他时刻能否验证所得参数。

一旦模型有了固定参数，未来预测便成为不同于解释既有资料的检验。旧资料参与了参数选择，新的独立资料则检查这些参数能否持续工作。如果模型每遇新点都可以自由更改全部设置，预测的约束就会减弱；若参数基本保持而仍能预报一系列位置，才显示了可迁移能力。

还可以把均衡点的数学后果说得更精确。设 P 沿以 C 为圆心的圆运动，从偏置 Q 看过去的角度在相等时间里增加相同数值。由于 QP 长度随位置改变，相等视角未必对应圆周上相等弧长。由此产生的非均匀圆周速度，是构造本身要求的，并非观察误差。若坚持绕 C 匀速，就必须改动其他部分来保持同样位置效果。

这里出现了真正的理论张力：一项提高预测组织力的几何规定，同时使另一项自然运动规范更难满足。解决方式不能靠把“匀速”一词说得含混。必须明确绕哪个中心匀速、保持什么距离、哪一种实际运动需要解释。词语澄清本身就能把模糊不满转化为可计算的问题。

数表的运行还有另一种误差传播。假设一个平均运动参数每单位时间只偏少量，短期预报可能很好，长期历元外推却会累积差异。若某表值只是局部插值误差，则影响可能局限于特定区间。区分随时间增长的偏差和随相位重复的偏差，有助于判断应该重估平均运动还是修改周期修正。

这样的残差形态是现代分析语言，但它指向模型中真实不同的责任。时间参数、几何形态与计算近似各自留下不同类型的后果。历史研究若把所有不符统称为“不准确”，便无法说明后来改进为什么针对某个参数或某个构造。更细的误差分类能够使创新动机具体化。

对于月球视直径问题，也可以做一个不依赖现代观测数值的推导。固定直径 d 的球体，离观察者距离 r 足够大时，角直径近似与 d/r 成比例。若模型让 r 大幅改变，角直径也应明显改变。它对角位置的成功因此会被另一个观测量挑战。这个检验的力量来自同一物理对象应当同时满足多项属性关系。

若辩护者说模型只负责黄经，不负责距离，那么它的有效域就得到收缩；这并非毫无意义的退让，因为一个有限任务模型仍可以有用。但收缩必须明说，不能在需要物理解释时把模型当真实宇宙，在出现困难时又将其说成纯计算工具。模型地位应当在不同论证中保持一致。

这一整组分析使“解释代价”具有可辨认内容：参数可能不唯一，特殊几何可能违背运动规范，基础表误差可能传播，位置与尺寸约束可能不相容。它们分别要求不同修正，不能由一句理论过时一并替代。只有这样，后世对托勒密的改造才会呈现为具体创造，而不是不断添加无意义小圆的漫画。

8.4 未解决的代价

均衡点最直接的代价，是圆形轨道与匀速旋转不再由同一个中心共同保证。如果规范要求每一组成运动都绕其圆心匀速，均衡点就会造成解释困难。但如果主要任务是组织角位置，它又可以具有实际优势。双方争论并不必然来自一方懂数学、另一方不懂，而可能来自不同的优先要求。

月球模型则展示另一种代价：改进一类位置解释，可能给另一类可观察量带来问题。Krisciunas 讨论了托勒密月球模型所隐含的过大距离变化，以及这种变化对视直径的后果。[8-4] 这里关键不是用今天的月球距离数字嘲笑古代，而是指出模型内部不同输出应当接受共同检验。

若假定月球实体尺寸在短时间内近似不变，距离显著改变便应造成相应的角直径变化。这是一条几何联系。一个模型不能把距离当作只为位置计算服务的自由变量，同时拒绝它对可见大小的影响，除非明确限制模型的物理解释。如此一来，模型究竟是运动计算装置，还是完整空间结构，就成为无法回避的问题。

这个案例也说明，发现可检验后果与历史上实际执行检验仍有区别。今天可以设计裸眼或简单仪器的角直径比较，不等于已经证明某位古代作者曾按同样方式系统测量。思想史应

把“当时原则上可检验”与“史料显示已经检验”分开。但即使保持这种区别，跨量检验仍清楚揭示模型的边界。

本轮、偏心和均衡点因此应分别评价，不能捆成一个贬义词。某些构造是合成周期的有效数学工具，某些规定是对不均匀运动的近似表达，某些则引出物理实现问题。将它们一概称为补丁，会失去判断不同改造为何出现、为何被保留的能力。

从残差入口进入，我们问模型在相同数据和精度下做得怎样；从运动规范入口进入，我们问它要求怎样的宇宙构成；从距离与视大小入口进入，又出现新的约束。路径变化使评价对象更加完整。三原理的分析则需要进一步查明，哪些张力实际引发了新构造，哪些长期被容纳于不同任务的分工。

托勒密体系没有因为存在代价而立即失去工作能力。相反，正由于它足够系统，后来的研究者才能明确指出需要修改的是哪种关系。它既为继承提供方法，也为批评提供对象。天文学思想的发展，常不是从全无到全有，而是从一个有强能力也有具体边界的世界，重组出另一种能力结构。

在方向预测之外，地球本身的尺度将使天文学获得另一种检验资源。角度可以组织天空，但把角度转化为长度，需要可靠基线。当地面测量和天空观测建立联系时，宇宙不再只是一张可计算的方向图，而开始获得可以讨论绝对尺度的条件。

由此还可以理解为什么“增加复杂度”不是充分指控。有时多一个几何关系能够同时解释许多独立差异，复杂度增加获得了明确收益；有时增加一个参数只是吸收一个孤立点，几乎没有新的预期。应比较新增自由度带来的约束与预测，而不是只统计圆的数量。

若把两种模型放到同一组资料上，还须使用相同的日期解释和坐标转换。否则，一个模型使用修正后的数据，另一个使用未经修正的旧数值，比较就混入了材料差异。历史评估中最容易被忽略的公平性，正是让候选方案承担相同任务并接触相同证据。

这种比较可以保留多项结果：某方案位置较准，另一方案结构较统一，第三种方案更便于制表。没有单项绝对优势，并不使研究停滞；它明确指出下一步需要什么证据或新构造。托勒密传统的丰富性，正在于它让这些不同尺度的成功可以被分开讨论。

因此，对它最有意义的批评，不是从结局宣布过去不该存在，而是重建当时哪些能力已经成立、哪些连接仍然不足。模型越具体，改造的方向越能落实；笼统贬低旧理论，反而无法解释新理论究竟创造了什么。

这种评价既保留模型的真实贡献，也让它承担具体的失败责任。

本章参考文献

[8-1] Ptolemaeus Arabus et Latinus. Ptolemy, Almagest (Greek). https://ptolemaeus.badw.de/work/19

[8-2] Library of Congress. Mathematics—Rome Reborn: The Vatican Library & Renaissance Culture；《至大论》第十卷相关手稿说明。https://www.loc.gov/exhibits/vatican/math.html

[8-3] Zieme, S. Gerard of Cremona’s Latin Translation of the Almagest and the Revision of Tables, 2023. https://journals.sagepub.com/doi/full/10.1177/00218286221140848

[8-4] Krisciunas, K. A Mixture of Ancient and Modern Understanding Concerning the Distance and Motion of the Moon, 2013. https://arxiv.org/abs/1311.2798

第 9 章 测量地球如何改变天文尺度

9.1 观测角度如何变成长度

角度可以告诉我们两个方向相差多少，却不能独自告诉我们距离有多大。一条小圆弧和一条大圆弧可以张开同样的角度。要从天空进入长度世界，必须找到已知或可测的地面尺度，使角度与长度在同一几何关系中相遇。

埃拉托色尼测量地球周长的故事，最适合揭示这种转变。我们今天熟悉的叙述主要经由后世作者传下，而不是一份完整保留的原始观测报告。克莱奥米德斯所述的构造，把两地太阳方向与地面距离联系起来；美国天文学会对该古代转述的介绍，也明确列出同子午线、基线和近似平行太阳光等条件。[9-1]

方法可以用现代形式简洁表达。若两地点处于同一子午线，相隔地面弧长 s，对应地心角 θ，那么球形地球半径 R 满足 s＝Rθ，其中 θ 用弧度表示。地球周长 C 于是为 2πs/θ。若用度数，则 C＝360s/θ。这个写法只是把几何结构呈现给今天读者，不是声称古代原著采用了现代符号。

为什么太阳影长可以提供地心角？在太阳光线近似平行的前提下，两地竖直方向不同，造成太阳相对天顶的角度不同；如果两地位于适当的同子午线关系，比较当地正午天顶距便能约束纬度差。几何中起决定作用的是地方竖直方向的差异，而不是两根杆子本身长短的差异。

一根杆子的影长必须与杆高一起使用。若杆高 g、平地影长 l，理想情况下天顶距可由 tan z＝l/g 表示。只有影长而没有杆高，便没有完整角度；杆未竖直或地面不水平，又会引入额外偏差。一个被讲述为“看一眼影子”的故事，实际隐藏着明确的测量条件。

两地数据的合并也不能只靠同一个日期。如果太阳赤纬在比较间隔中变化，或观测没有对应当地正午，就需要处理附加差异。对于古代转述，我们不能在材料缺失处自动补出完美同步的实验；应区分几何方案要求什么，与现存文献证明怎样实施。

正是在这组联立中，地球周长成为新的 S。D 不只是在天空测角，也包括地面距离比较；E 是球形近似、光线方向、当地竖直与基线之间的约束。长度与角度缺一不可。这里的世界发生不是地球获得了此前没有的体积，而是地球尺度进入一个能够计算、比较和传承的研究聚合体。

9.2 基准与误差

一个精妙公式不会自动使输入可靠。地面距离究竟指道路行程、河流路线，还是对应所需纬差的弧长？两地是否确实在同一子午线附近？使用哪一种长度单位？这些问题决定了结果能被解释到什么程度。

道路长度通常大于两点间的最短地表路径，沿曲折路线累计的步数也不能直接代替子午线弧长。若把它们混用，计算仍可得到一个整齐周长，但这个数值已经与所宣称的几何对象不同。误差不只来自仪器不够精细，也可能来自测量对象定义错误。

史料中的 stadion 单位尤其需要谨慎。关于埃拉托色尼所用 stadion 究竟对应多少现代米，存在不确定性；美国数学协会的相关讨论将这一点置于问题中心。[9-2] 因而，不能先选择一个最接近现代结果的单位长度，再把所得接近程度当作古代测量精确的证据。那样会让结论参与决定前提。

用比例分析可以把几种误差分开。由 C＝360s/θ 可见，基线估计增加百分之一，在角度不变时会使周长增加百分之一；角度估计增加百分之一，在基线不变时则使周长略减约百分之一。这个现代敏感性说明不需要知道古人具体误差，就能显示两类误差如何进入结果。

如果两项误差方向相反，它们可能部分抵销。最终数值接近现代值，并不保证每一步都准确；最终数值偏离，也不等于几何方法没有价值。思想史若只用一个百分比评判整项成就，就会丢失方法结构与实际测量之间的区别。

更细的审查还会区分随机波动与系统偏差。重复测量有助于减少某些偶然读数误差，却不能自动消除杆始终倾斜、基线定义不当或单位换算有误。若所有测量共享同一偏差，多次一致反而会使人更加确信错误结果。共同条件的可靠性，不能由重复本身完全证明。

太阳并非数学点，阴影边缘也不是绝对锐利。选择太阳中心还是边缘，如何处理半影，都会影响测角。本文不据此断言某个古代实验采用何种具体技术，而指出：当我们声称重现古代精度时，就必须说明这些细节。教室演示能显示几何原理，却不能替代对历史操作的考证。

还有一种容易被忽略的尺度问题。球形近似适合解释古代测地的基本结构，但现代地球形状与局部重力方向更复杂。不能把现代椭球参数和精密测地坐标悄悄代入古代数据，再宣称古人已获得同样意义的结果。不同精度层级的“地球半径”需要说明定义。

9.3 测地与天文的相互支持

地面测量给天文学长度基准，天空观察又为地面方向和纬度提供约束。这种相互支持，比单向“用天文学测地球”更准确。当地理位置可以由天象比较加以组织，地理资料本身也会改变天文学能够使用的基线。

斯特拉波《地理学》保存了关于埃拉托色尼与喜帕恰斯的批评争论，其中涉及行程距离、直线近似、子午线关系以及应当使用怎样的几何严格性。[9-3] 这些讨论说明，古代研究者并非不关心数据质量；他们会争论一种地理材料能够承受多严格的数学检验。

这项争论有一个至今仍有启发的核心：资料的生成方式决定它能支持的推断强度。商旅传闻中的距离可能对区域描述有用，却未必足够精确，不能未经处理就作为严格三角计算的一边。反过来，几何不一致也可以提示资料需要修订，但必须考虑原叙述本来是否只是近似。

斯特拉波对过度严格检验近似地理资料的批评，并不意味着拒绝数学。他同时指出某些错误过大，不能借近似为其辩护。这就形成一个精度层级问题：不是所有差异都重要，也不是任何差异都可忽略。判断必须与任务和材料可靠性相称。

对于本书而言，这是一项具体的 E 分析。地理叙述中的“同一子午线”若只是宽松近似，在天文尺度计算中便不能不加说明地当作严格相等。一个词在不同操作中承担的精度责任不同，转换时必须重新规定。知识跨学科连接，不是把两个词表拼在一起，而是让这些责任取得相容性。

地球尺度一旦被估计，又会成为更远距离研究的单位。即使暂时不知道现代公里值，研究者仍可能用地球半径表达月球距离的倍数。相对尺度与绝对尺度因此可以在不同阶段取得进展。用哪个单位计算，不影响某些比例关系，却影响把结果转成实际长度的可靠性。

这也使“古人不知道距离”显得过于粗糙。我们应区分只知道方向、知道距离比例、知道以地球尺度表示的距离，以及知道可靠现代长度。这些不是同一成就，也不应混为全有或全无。尺度知识可以逐层建立，前一层既支持后一层，也可能把自身误差传进去。

三方程中的回写在这里可以获得明确内容：一个被采用的地球尺度 S 改变了后续距离推断 D，而这些推断对基线、单位和地理位置 E 提出新要求。但某个尺度被写进书中，不足以证明所有后续使用者已经采用；具体传播和运算痕迹仍然重要。测地的思想影响，应通过实际调用来辨认。

斯特拉波的批评材料还允许我们看到，一个几何词怎样在跨领域使用时改变负担。他讨论“直线”“平行”和“同一子午线”时，并不总要求地理叙述达到仪器测定的严格程度。

[9-3] 但天文测地若要由弧长反推周长，又必须知道这种近似可以容忍多大范围。相同词语进入不同任务，需要重新评估精度。

设两地纬差已知，而实际基线偏离南北方向。若直接使用全部路线长度，便可能把东西方向的行程也算入所需弧长。即使每一步长度都测得很好，结果仍会系统偏大。这个现代例子说明，测量精细不能补偿几何对象取错；首先要确认究竟在测哪一段。

同理，纬度差与地表距离具有不同的证据来源。前者可以来自天象角度，后者可能来自测量或既有地理资料。把两个来源连起来时，应该保留各自不确定性。若最后只剩一个周长数字，后来研究者就难以判断修正应该发生在哪个环节。可追溯性因此是尺度可积累的条件。

由周长换成半径，虽然数学上只需除以二π，历史上却仍要说明采用怎样的圆周率近似与长度单位。对于粗尺度判断，这些细节也许不是最大误差源；对于宣称极高精度的比较，它们却不能省略。精度要求越高，越需要完整交代计算链，而不是只选择最有名的两项输入。

还有一个思想层面的变化：地球不再只是观察者所站立的地面，而成为一个可以从局部关系推算整体的对象。人无法沿每一条经线直接测遍全地球，却可以利用模型把有限资料扩展为整体参数。这种能力与任何外推一样，需要条件；但它展示了数学科学怎样超出直接接触的范围。

这种超出并不是凭空跳跃。圆周关系提供形式，天象比较提供角度，地面基线提供尺度。三者缺少一项，就不能得到同样意义的结果。说“数学算出了地球”，若被理解为数学独自制造答案，便遮蔽了测量与单位；说“只是观察出来”，又遮蔽了从局部到整体的推断结构。

这一案例因此比单纯列出一个伟大数值更适合展示主客互动。观察者不是对地球毫无影响的透明窗口，测量活动也不是任意主观意见。特定操作让地球以整体尺度的方式显露，结果又要求他人能够沿相应关系复核。对象的科学属性在这种复合过程中成立，同时受到地面与天象的共同限制。

如果以后取得更好的基线，周长数值可以改变而方法结构保持；如果发现球形近似不足，则需要改变模型层级。两种修正具有不同强度。把它们区分开，就能看见知识发展既可以是已有能力精化，也可以是对象结构重组。古代测地的价值，不只在某次答案接近今天，而在于为这些不同修正建立了入口。

9.4 不确定的古代数据

埃拉托色尼故事的吸引力，常在于一个简单比例就能得到地球周长。这个简洁性值得保留，却不能把历史也简化成同样干净的四个步骤。现存转述、不同数值和单位争议，都是历史材料的一部分，不能为了叙事流畅而删去。

我们可以很有把握地评价几何思想：把地上距离与天空角差联系起来，使地球整体尺度成为可推算对象。我们对某个具体测量数值与现代值究竟差多少，则可能没有同样把握。两种评价可以同时成立。方法成就不必依赖一个被过度精确化的百分比来维持。

从著名结论往回看，容易把后世教科书中顺畅的版本当作原始实验；从基线、角差、单位和文本源流逐项进入，则会发现哪些部分有材料，哪些部分是教学简化。这不是拆毁经典故事，而是让故事的真正力量显露：它提出了一种把局部观察联结到整体尺度的可检验结构。

一种思想实验还能显示其边界。若只在同一地点测得一个影长，且不知道太阳方向与日期，通常无法由此唯一决定地球大小。增加第二地点并不魔法般解决一切；还必须知道两地关系与基线。新信息之所以有效，是因为它被嵌入合适几何，而不是数量增加本身。

同样，若两地影长相同，也不能立即断言地球平坦。可能观察时刻不同、地点关系不适当，或太阳相对位置恰好使两个局部结果相同。一个反例要对准模型的真实预测，而不是对准简化故事中的口号。几何方法越清楚，支持与否定的条件也越清楚。

本章因此把尺度知识理解为一种联合成就。地面行走、长度制度、竖直测量、太阳观察和数学关系共同构成结果；任何一项单独存在，都无法提供完整周长。它们不是任意堆积的背景，而是互相规定的测量系统。

这样的系统可以被传承，但不会靠一个答案数字自动传承。后来的使用者需要知道数字从什么单位、什么模型和什么操作而来，才能重新校验或移用。地球尺度的故事于是把我们带向另一个问题：一套模型怎样穿越语言与时代，使其方法仍可继续运行，而不只留下被反复抄录的结论？

历史数值的保留还影响后来的比较。若一个传统只记住某个周长，忘记它的单位与测量前提，数字可能逐渐成为权威常数。重新询问单位并不是琐碎考据，而是恢复这个常数与实际测量之间的联系，使其再次成为能够讨论的结果。

在这一点上，误差分析与文献分析相互需要。前者说明输入变化会如何影响输出，后者说明哪些输入实际上有依据。只有数学，没有文献，会把假定参数计算得极其精确；只有文献，没有数学，则可能不知道某个看似小的歧义会造成多大后果。两者结合，才构成可靠的历史尺度研究。

尺度的可靠性必须沿整条证据链理解。

本章参考文献

[9-1] American Astronomical Society. This Month in Astronomical History: June, 2019；转述 Cleomedes《天文学讲义》I.7 及其前提。https://aas.org/posts/news/2019/05/month-astronomical-history-june

[9-2] Walkup, N. Eratosthenes and the Mystery of the Stades. Mathematical Association of America. https://old.maa.org/press/periodicals/convergence/eratosthenes-and-the-mystery-of-the-stades

[9-3] Strabo. Geography, Book II, Chapter 1, §§38–41，H. L. Jones 英译，LacusCurtius 公开文本。https://penelope.uchicago.edu/Thayer/E/Roman/Texts/Strabo/2A3*.html

第 10 章 模型怎样跨越语言与时代

10.1 翻译中的变与不变

一部天文学著作被翻译之后，什么东西真正到达了另一种语言？书名和作者可以到达，几何图形可以到达，数表也可以到达；但让这些部分共同运行的能力，需要另一种更细的传递。使用者必须理解定义，辨认图中的点，按顺序计算，并知道所得结果适用于什么情形。

《至大论》的传播不是一条没有分支的直线。关于其阿拉伯语版本的研究区分多个译本和修订传统，其中部分已经失传，只能经书目和引文获知。[10-1] 这样的材料格局意味着，今天所见文本不能无条件代替全部历史文本，也不能因为某个版本未存，就断言它当时没有影响。

书名本身可以保存传播痕迹，但不能承载完整传播史。某个名称经过中间语言进入另一语言，说明接触链的一部分；它不证明全书每一段都走同一路径。标题、术语、星名和数学表可能有不同来源，后来又被编入同一份手稿。作品的统一外观可能覆盖其内部的复合历史。

翻译者首先要解决术语等值。一个词可以在普通语言中有相近含义，却在几何中要求严格区分。例如，圆心、观测点和匀速运动参照点若在转译中混淆，读者仍可能看懂图形的大意，却无法重现模型。定义的稳定因此比字面顺口更重要。

另一方面，过度追求逐字对应也不保证数学清楚。原语言省略的关系，可能在新语言中必须补明；原文依赖的数制习惯，可能需要解释。翻译中的“增加”不一定改变理论，也可能是在新读者条件下重新建立同一操作。判断改变性质，需要比较它影响了哪些推论。

图形也有自己的转译问题。字母换成另一套书写系统，正文指向必须同步；图形旋转或镜像，方向词也需检查；一个点的位置若因抄绘略有偏移，哪些关系仍由文字保证，哪些会被读图者误解？数学图不是完全脱离语言的透明通道，它与定义、标记共同构成可读对象。

因此，模型的连续性不能仅由公式外观判断。相同图形可能配有不同运动规定，相同术语也可能指向不同参照。只有实际执行一次计算或证明，才能看出译文在关键关系上是否保持原操作。思想传承的一部分证据，藏在那些看似枯燥的步骤里。

10.2 教育与仪器的条件

一份手稿的物质形态能够揭示传承并非一次完成。卡塔尔数字图书馆所列英国图书馆 Add MS 7474，是《至大论》前六卷的阿拉伯语抄本。馆藏说明记载其有页边注、行间异文

和改正，且若干表格、图形未完成。[10-2] 这里依据的是馆藏编目，不能冒称已独立校读全部阿拉伯语页。

这些具体特征至少表明，“拥有这部书”不是一个单一状态。一份文本可以在正文上相当完整，而表值尚未填齐；可以有丰富批注，却不是每一部分都同样使用；可以保留多种读法，等待读者选择。手稿是持续工作的载体，不必等同于现代印刷本式的封闭成品。

空白表格尤其值得认真对待。它可能反映抄写流程、材料缺失或未完成工作，不能仅凭空白判定作者不会计算。但空白确实限制了后来的读者可以直接调用什么。如果正文给出方法而表格未成，使用者需要更多计算能力；若表格齐全而方法难懂，则可能形成能查表却难重建的另一种能力。

教育由此进入模型生命。学习者必须掌握数制、角量、插值和几何阅读，才有可能把书中操作变成自己能够执行的程序。一个传统能长期延续，通常不仅要保存权威文本，还要维护解释文本的训练方式。没有这种训练，经典可以被尊重，却难以成为日常研究工具。

仪器同样不是附加附件。文本规定测量某个角度，使用者还须能够对准目标、读取刻度并辨明误差。仪器在不同工艺条件下产生的读数未必具有相同精度；复制外形也不保证复制性能。理论进入当地观测，需要把数学量与实际读数重新接通。

这正是 SDE 所强调的复合能力。S 不是单独存于书中的结论，D 包括读图、计算与观测的差异运行，E 包含术语、图形、数制和仪器尺度的相互规定。若这些联系不能被接续，模型可能以文化权威形式存活，却未充分恢复其研究能力。

但我们也不应预设每位读者都必须掌握全部环节。共同体可以分工：有人译释，有人计算，有人制造仪器，有人实施观测。关键是这些环节能否可靠衔接，而不是每个人是否成为全能天文学家。知识能力可以属于复合协作系统，其成败需要沿接口检查。

10.3 不同地区如何重新校验

一套天文表即使在译文中完全无误，也可能不适合直接用于另一个地点。地方纬度、时间起点和历法单位改变，都会要求调整输入或表格。传播进入实践的标志，因此不只是引用经典，还包括重新计算、观测核对和处理不一致。

可以用一个一般例子说明。某表以特定历元为起点，给出天体平均运动；新使用者采用另一纪年，就必须把日期转换成同一种时间间隔。若只改年号而没有改基准，误差会被误认成模型不准。另一个表若涉及升起时间，纬度改变便可能使旧表失去适用性，即使天体轨道参数没有变化。

这种调整说明模型有不同可迁移部分。某些几何关系在地点改变后保持，某些参数须重新估计，某些输出必须重新制表。把它们拆开，才能知道一次“地方化”究竟只是日期换算，还是涉及观测、算法乃至对象结构的重建。

译本传统确实可以包含主动重算。Zieme 对杰拉德译本的研究发现，基础数表与文本运算的一致性经过调整；其中自含表结构使部分数值可以彼此核查。[10-3] 这一发现不宜泛化为所有中世纪译者的共同做法，却足以说明表格差异不总是抄写失误，也可能是计算性修订。

从操作上看，若一列数值由另一列差分决定，两列便构成相互检查。读者发现一项异常时，可以检查相邻变化是否连贯，而不只相信抄写权威。这里的知识可靠性发生在数值关系内部：一个结果既供使用，也成为另一结果的约束。

这种自检也有范围。两列都按照同一错误参数计算，仍可能彼此一致；内部一致不能完全代替外部观测。相反，外部差异也不总要求改变基本几何，可能首先暴露日期、单位或表格使用的问题。可靠校验应当沿不同层级寻找原因，而非把任何偏差立即归入整个理论。

地方化还有一种不以数值精度为唯一目标的作用。不同地区可能优先关心历日、时刻、方向或占候任务，由此选择不同的教材组织。简写本、注释本和完整论著可以各有用途。不能把非完整形式一概理解为知识衰退；应检查它是否保存了其任务所需的关键关系，以及删除造成什么限制。

一个可运行传统的形成，因此是一种新的聚合。它吸收传入模型，也加入当地时间制度、教学方式和观测条件。新的 S 可以与旧 S 建立可靠对应，却不等于全部条件原样复制。传播既有保持，也有生成；其创造性应通过具体改动判断，而不由文明名称预先分配。

保留原表与重算表的并置，还能形成一种特殊的历史证据。如果若干数值只是无规律错写，可能反映传抄；若差异沿某种运算规则系统出现，则需要考虑主动重算。这里的“系统”不能凭肉眼印象决定，应与具体算法比较。数字的差异因此可以帮助重建知识劳动，而不仅是等待校正的污点。

反过来，计算结果相同也不必意味着使用了同样方法。不同算法可以在给定精度内取得一致输出。若要判断方法传承，还需要观察中间步骤、舍入方式、表格布局或独特参数。只比最后答案，可能把独立工作误认为复制；只比文字，也可能漏掉隐藏在数值中的再创造。

手稿中的批注则可能保留这种中间状态：读者提出疑问、列出另一版本，尚未形成统一答案。它们不是成书正文之外的无关噪声，有时恰是知识正在发生的痕迹。但批注归属与年代仍需独立判断，不能因为它写在古书旁，就自动归给原作者或最早译者。

由此，文本、算法与实践具有不同的传承速度。一部书可以很快被带到另一地区，培养能理解它的人需要更久，建立相应观测条件又可能更慢。模型的历史生命，是这些不同速度如何重新衔接的结果。看见这种时间差，才能解释为什么接触并不立即等于能力形成。

10.4 对单线传播史的修订

关于拉丁语接收的研究指出，科学翻译在不同地点展开，手稿也可能结合不同阿拉伯语版本。[10-4] 这使“希腊传给阿拉伯，再传给欧洲”的简图只能作为极粗略的索引，不能承担完整解释。真实传承包含往返校读、多种中介、部分失传和新的数学工作。

问题不在于一切影响都无法确认，而在于每条影响需要适当证据。相同术语可能支持语言接触，相同特殊错误可能提示文本亲缘，相同算法结构则需要再排除独立形成的可能。不同证据的力度不同，不能用一句“受到影响”将它们全部压平。

尤其不能从后来的知识相似，反向编造没有文献支持的完整路线。有时只能确认某种方法在两处出现，尚不能确定中间接触。保留这段缺口，不会否定两处各自的贡献。相反，虚构顺畅路线容易把接收者的重新计算写成被动复制，也把发送者的多样传统写成单一源头。

还应区分文本传存史与思想作用史。保存下来的抄本数量受收藏、战争、材料和偶然性影响，不能简单作为某种学说受欢迎程度的统计。一个版本存本少，可能曾经重要；一个版本后世抄本多，也可能主要服务特定教育需求。现存数量需要与内容和使用痕迹合读。

本章的路径比较由此具体化。从书名与作者名单开始，容易得到接力式故事；从表值、批注和局部定义进入，则能看见实际工作者怎样维持或改变模型。前者提供传播范围，后者揭示传承机制。两者相接，才有可能把文明交流写成可考察的知识过程。

当传入结果与当地观测差异不合时，新的参照关系可能被建立。这里的 P3 不是抽象地说“文化碰撞带来创新”，而是指明：哪一项日期、纬度或定义使原比较失效，怎样改动后恢复可计算性。若只有译本，没有相关运算痕迹，就只报告文本接触，不越界宣称实践已经改变。

古代几何宇宙因而没有停在它最初成书的时刻。它通过翻译、注释、重算与观测进入多个新的研究系统。统一性存在于可追踪的数学关系，多样性存在于不同任务和重新组织，和谐性则体现为这些差异能够互相校验、接续，并在不相容处提出新的问题。跨文明星表与计算改革，将继续在这样的具体工作中展开。

传承的成果因此不只是同一本书在更多地方出现，而是更多人能够接续其中的研究关系。判断这种能力，应追踪计算、批评、重测和改写的实际痕迹。由文本流动走向能力发生，才足以说明跨文明交流怎样参与天文学世界的持续创造。

本章参考文献

[10-1] Kunitzsch, P. The Arabic Translations of Ptolemy’s Almagest. Qatar Digital Library, 2018. https://www.qdl.qa/en/arabic-translations-ptolemys-almagest

[10-2] British Library / Qatar Digital Library. Kitāb al-Majisṭī, Add MS 7474，馆藏说明。https://qdl.qa/en/archive/81055/vdc_100000004884.0x000001

[10-3] Zieme, S. Gerard of Cremona’s Latin Translation of the Almagest and the Revision of Tables, 2023. https://journals.sagepub.com/doi/full/10.1177/00218286221140848

[10-4] Grupe, D. The Latin Reception of the Arabic Tradition of the Almagest. Qatar Digital Library. https://www.qdl.qa/en/latin-reception-arabic-tradition-almagest

第 03 编

跨文明星表与观测改革

第 11 章 印度天文学中的计算体系

第 12 章 中国历法与星官体系的长期积累

第 13 章 伊斯兰世界的天文表与观测台

第 14 章 图西偶与行星运动的几何改造

第 15 章 乌鲁伯格星表与精度共同体

第 11 章 印度天文学中的计算体系

11.1 计算对象而非人物传

一颗行星的位置怎样成为能够计算的对象？这个问题看似只要求输入日期，查出一个角度，实际却把历元、日的起点、周期、运动修正和观察地点压缩在同一次运算之中。计算者如果不知道哪一天被算作起点，哪一种日被用作单位，即使拥有正确的三角关系，也不能得到可以同天空比较的结果。印度天文学的重要性首先在这里：它发展出将这些条件联结起来的计算传统，使天象进入可传授、可重复、可批评的操作过程。

阿耶波多的《阿耶波多历数书》适合作为入口，但不宜被当作整个印度天文学的缩影。这部简练的韵文作品包括数学、时间计算和天球问题。克拉克的英译本提醒读者，作品并非把全部步骤写出的初学手册，许多常用运算被视作读者已经掌握的知识。[11-1] 因而，现代读者如果从一首短偈找不到完整算法，不能立即认定古代计算者同样缺少这一环节；反过来，也不能任凭今天的数学知识替原文补出任何想要的步骤。

天象的计算化以区分不同周期为前提。月亮相对于恒星转过一周，与月相重新回到相似状态，不是同一个周期。行星返回同一恒星背景附近，与它同太阳重新构成相近位置关系，也不是同一件事。这些差异并非抽象数学强加给天空的繁琐分类，而是长期比较过程中不同重复现象逐步取得不同身份的结果。只有先建立这些身份，周期数才有明确的意义。

以合朔为例，太阳与月亮在相应坐标中的经度相合，是一个可供计算的条件，但合朔不能单独保证发生日食。还要考虑月亮相对黄道的偏离，以及观测地点对应的几何关系。同一个预测程序因此包含两种性质不同的工作：先确定一个事件候选，再判断这个候选是否满足更严格的发生条件。如果把它们合并，便会误以为历算只是寻找一个神秘的周期数字。

行星位置的推算也具有层次。先用周期建立平均运动，再施加修正，得到可与观测方向比较的位置。这里的“平均”不是宣布真实天体始终匀速，而是提供一个可操作的基准。修正不是可以随意增减的装饰，它受到模型中的几何关系和参数约束。原典中的平均位置与真位置之分，提示我们将算法看作一组有组织的转换，而不只看最后出现的角度。[11-1]

用现代语言说明，一个历元位置加上平均角速度与经过时间的乘积，可以构成平均经度。这个表达只为帮助理解，不冒充古代原式。若历元前移，起始位置必须随之改变；若日的定义改变，经过时间的数字也必须相应换算。单独搬动其中一个数，会破坏整个程序的一致性。可计算天象的形成因此不是数表、时间和几何的相加，而是它们相互限定后得到的整体。

这也解释了为什么印度传统中既有不同论著，又有不同计算用书与注释活动。提出一套宏观周期结构，与为具体年代提供便利运算，是相联而不相同的工作。前者追求广泛的统一，后者承担实际计算的负担。如果史家只重视壮阔的宇宙周期，便容易忽略把它转化为可用历算的中间劳动；如果只重视预测成败，又容易丢掉数值系统赖以组织的理论联系。

本书所谓新的天文学世界，在这里指一种新的对象构成：太阳、月亮和行星不再只作为可指认的亮体存在于叙述中，还作为具有不同周期、修正规则和地方可见条件的计算对象相互规定。这并不意味着算法使月亮在物理上诞生，而是说，在这种主客互动复合体中，“同一事件”“同一周期”和“同一位置”获得了新的具体内容。

11.2 三角工具的作用

理解印度天文学，不能绕过与半弦有关的计算。现代人熟悉无量纲的正弦函数，古代计算者所使用的量却往往同所取圆的半径相联系。若用现代记号写作 $R\sin\theta$，R 的选择便是数值系统的一部分。这种写法有助于理解原理，但必须保留它与古代语言和操作形式的距离。把古代数表直接排成现代函数表，容易让真正需要解释的构造过程消失。

阿耶波多相关表式以正弦差的方式组织数值。它的意义不只是节省记录空间，还在于把一系列角度上的量联系成可以逐步恢复的序列。研究者对这些差值的解释与分数部分进行过专门分析，表明原典短句与后世注释之间存在值得认真分辨的运算层次。[11-2] 一列数字的形成，可能依赖几何推导，也可能包含近似、舍入和实际计算的取舍。

考虑一个简单而明确的现代说明。若已知相邻两项的差，逐项累加便能够恢复表值；但早期的舍入误差也可能沿着累加过程传播。这不意味着差分表不可靠，而是说可靠性的判断必须同时看步长、存储精度、重建方式和所需预测精度。只说古人“已经知道正弦”，没有说出计算系统如何运行；只按今天的小数位比较，也没有说明当时的任务究竟需要什么。

正弦关系一旦进入天文运算，就不再只是孤立的数学定理。角度、半径、方位和时间之间需要明确对应。一个角度究竟表示黄道上的经差、天球上的高度，还是某个模型圆上的偏角，会改变运算的解释。数学上相同的乘除步骤，放进不同问题中可能承担不同功能。对象的身份依靠整个运算语境取得，而不是由一个公式独立携带。

由此可以区分三种变化。第一种是改善数表精度，保持基本几何关系与用途不变；第二种是更换算法，让相似输出以更少的劳动获得；第三种是重组问题，使原来不能处理的对象进入计算范围。这三种都可能构成贡献，但其思想史意义不同。若把它们一律写成“计算更准确”，便会抹掉工具、对象与解释之间的区别。

关于阿耶波多正弦构造的现代研究，常用相似三角形或有限差分来帮助读者理解。[11-3] 这种重构具有教学价值，却不能单凭公式相似就断言古人已经拥有现代微分观念。两者可

能在某个关系上等价，却并不具有相同的问题范围、证明语言和运算工具。承认这一点，不是降低古代数学的地位，而是使其原创性摆脱“提前做出今天的东西”这一狭窄标准。

在 SDE 分析中，数表的一个值可以成为显露 S，但这个值的意义依赖差异序列 D 与特征纠缠 E。D 不仅是表中前后两项之差，也包括从角度到几何长度、从平均运动到修正结果的一系列转换。E 则表现为这些量的相互规定：半径选择限定数值尺度，角度划分限定表的采样，目标天象限定容许误差。把这些要素一一列出，还不够；必须指出改变某一项为何要求其他项一起改变。

例如，将半径扩大一倍，不会改变抽象正弦关系，却会改变所存长度数值。若运算中某个分母仍沿用旧半径，最终改正就会失去一致性。这是一个可以检查的构成依赖，不需要用宏大的语言加以包装。它说明数学工具的可迁移性有明确条件：不变量可以保持，表达它的单位与接口却必须重新配合。

11.3 注疏和修订

一部简短原典为什么能够维持长久的研究生命？答案不能只归于作者的天才。作品必须被解释、示范、批评和重新用于计算。注疏把凝缩语句展开成可执行步骤，又把实际运算遇到的问题带回文本。它既可能保护既有体系，也可能使体系内部的歧义暴露出来。因此，注释传统不是原创思想之后的静止尾声，而是知识继续发生的重要场所。

克拉克译本的导论讨论了不同引文、注释与存世文本之间的困难，并明确承认其材料与翻译的局限。[11-1] 对思想史而言，这种自我限制极其重要。我们今天面对的“阿耶波多”并不是未经传抄的声音，而是通过文本、解释和编辑活动显露的历史对象。不同版本的歧义不能用一个顺畅的现代故事抹平。

注疏可以承担至少三类工作。它解释词语，使学习者知道某一技术名词所指；它补出运算，使简写的规则变得可用；它说明理由，使一个规则不只是需要服从的指令。三者可能出现在同一段文字中，却应当分辨。解释者补出的证明，未必就是原作者当年的发现过程；但即便如此，这个证明仍然是后来传统的一项真实成果。

历算修订同样需要区分问题来源。预测与观测不合，可能来自周期参数，也可能来自历元、地方经度、记录时间或运算错误。只有将这些可能性分开，残差才能具有诊断意义。如果每次偏差都被当作整个宇宙模型错误，研究将失去稳定基础；如果每次偏差都被归咎于抄写或操作者，体系又会失去改变机会。成熟的计算传统必须在这两种极端之间建立分层纠错。

设想一个现代说明性的检验：同一算法用于相邻年份时表现尚可，跨越较长年代后偏差持续扩大。这样的模式提示平均周期或长期改正可能需要检查。若误差在所有年代大致相

同，则起点或常量偏置更值得怀疑。若只在某些方位显著，地方观测条件可能不可忽略。这些并非对某位古代作者实际操作的复原，而是展示史家为何不能把所有“历差”压成一个数字。

三方程在这里要求我们追踪两种反向作用。已经形成的预测结果 S，会改变下一轮比较什么、在哪些日期观察，从而影响 D；新的比较与运算经验，又可能改变历元、参数和教学程序之间的配合，使 E 重新组织。可是，只有在文本或使用记录中看到这些变化，才能把它们写成历史回写。逻辑上可能发生，不等于历史上已经发生。

传统的延续也可能保存不一致。某种公式因便于教学而继续使用，即使更复杂的计算已经出现；一种解释因经典地位而被保留，同时实际运算悄悄采用其他规则。这类情况不宜简单判为虚伪或落后。理论论证、计算效率和教育传播的成本不同，其变化并不会自动同步。思想史必须说明这种不同步具体发生在哪里。

阿耶波多地球自转主张及后世对相关文本的解释，是需要特别谨慎的例子。原典与注释的关系表明，即使某种运动解释受到争议，其他计算资源仍可能被吸收。[11-1] 因而，一个传统不是整包接受或整包拒绝。天体运动的物理说明、数表的便利性与数学规则的有效性，可以有不同命运。

这使“同一学派”成为需要研究的对象，而不是现成答案。学派的一致性可能来自共享文本，也可能来自共同算法或共同教学。若这些方面并不完全重合，就不能仅凭名称认定所有追随者接受相同宇宙论。真正有内容的统一，常常是在差异中维持有限的共同操作，而非把全部解释都固定下来。

11.4 跨文明证据纪律

印度天文学与其他传统之间的交流，是不能回避的问题。但交流史有一种常见诱惑：一旦发现两套算法相似，就迫不及待决定谁借自谁。这样的结论跳过了时间、文本、中介语言和使用方式。算法相似可以提出传播假说，却不能独自完成证明。历史联系需要比数学等价更多的证据。

反过来，过度强调独立原创也会使研究失真。一个传统完全可能接收已有材料，然后改变问题、表式和用途。接受与创造不是互斥选项。若一个几何工具进入新的时间制度和历算结构，其抽象关系可以保持，而实际可计算的对象已经发生变化。把这种工作说成“照搬”，同样没有解释知识如何成为当地实践的一部分。

六路径的价值在此可以落到具体判断。若从外部影响的成品 S 出发，研究者容易寻找名称、图形和数值的相似，再将印度材料归入传播线路。若从印度计算中的 D 出发，先问如何

处理特定天象，再检查所用工具与文本 E，则能够看出某种外来资源进入之后究竟承担什么功能。两条路径不必相互排斥，但它们应当互相纠正。

一条可取的论证链，应先确认比较对象：究竟是同一几何关系、同一数值表，还是相似的宇宙安排？随后确认年代与版本，避免把后世修订投回早期。再寻找可达性证据，说明材料如何可能接触。最后比较改造内容，确认接受者改变了哪些单位、参数或解释。没有后一步，传播史即使证明了来源，也仍然没有写出思想发生。

以三角计算为例，半弦与整弦之间存在数学关系，这是可独立验证的事实。它是否通过某一具体文本在某时传入某地，是另一类历史命题。而进入一套不同历算之后是否形成新的计算便利，又是需要操作比较的问题。混淆这三层，容易把数学证明当成传播证明，把传播证明又当成贡献已经解释完毕。

本章采取的判断因此是有限而积极的：印度天文学的思想贡献，必须在其可执行的计算结构中辨认。它的创造性不靠隔绝交流来维护，也不靠宣布每一种现代观念早已具备来抬高。它体现在对象、单位、周期、数表与解释不断重新配合，使新的问题能够被提出，新的差异能够被稳定比较。

这一判断也给 SDE 分析设置了边界。如果某项所谓“纠缠”只是把时代背景列得更长，没有揭示计算要素之间的相互规定，那么它并未比普通技术史增加解释。如果三方程只把输入叫 D、背景叫 E、输出叫 S，也没有说明修订如何反过来改变操作条件，那么它只是换了名字。真正的增量，是辨认一个可计算对象赖以成立的依赖结构，并指出这种结构怎样被使用、修补与迁移。

因此，印度天文学不能被放在希腊与欧洲之间，充当一段等待跨越的走廊。它拥有自己的问题组织和修订历史。进入相邻文明的材料时，应保留这种方法：先看一个传统怎样让天象成为可处理的对象，再考察它如何同其他传统发生联系。由此，比较不再是给文明排座次，而成为追踪天文学世界多种发生方式的工作。

最后还要问，什么证据会迫使我们收回上述解释？如果一个算法在文本中存在，却没有可辨认的使用痕迹，就不能直接把它写成广泛实践；如果后世注释只重复权威话语，而没有补充运算与比较，就不能称之为持续创新；如果一种数值变化完全能够由抄写差错说明，就不应为它构造一场理论革命。发生学解释的强弱，不取决于描述是否宏大，而取决于它是否接受这些更简单的反解释。

这样的限制使比较更有力量。它允许同一传统在某些问题上创新，在另一些问题上停滞；允许接受者改变来源材料，也允许某次传播没有留下可辨认成果。多样性因此不只是列举不同文明，而是承认不同知识环节具有不同变化速度。统一性则体现为共同遵守证据要

求。两者的和谐，不靠先验宣布每个传统都同样重要，而靠准确呈现各自实际解决的问题及仍然保留的困难。

评价这些实践时，还应让计算者自己的问题保持可见：他们需要在特定日期作出判断，而不仅是在后世的知识谱系中占据一个先驱位置。这一差别决定了我们应如何选择材料和安排叙述。

本章参考文献

[11-1] Āryabhaṭa. The Āryabhaṭīya of Āryabhaṭa. Walter Eugene Clark 译注，1930。核读在线译本导论及时间、球面部分；旧译局限按译者自述保留。https://en.wikisource.org/wiki/Āryabhaṭīya_of_Āryabhaṭa

[11-2] R. C. Gupta. Fractional Parts of Aryabhata’s Sines and Certain Rules Found in Govindasvami’s Bhasya on the Mahabhaskariya. Indian Journal of History of Science, 6(1)。https://insa.nic.in/writereaddata/UpLoadedFiles/IJHS/Vol06_1_5_RCGupta.pdf

[11-3] Vijay A. Singh, Aneesh Kumar. Aryabhata and the Construction of the First Trigonometric Table. 2023。本文仅据其公开摘要采用现代教学重构的性质，不据此复原古代发现过程。https://arxiv.org/abs/2309.13577

第 12 章 中国历法与星官体系的长期积累

12.1 星官是什么单位

打开一幅古代中国星图，现代读者很容易先寻找自己熟悉的星座。找到几颗亮星之后，再把古代名称安放到现代星座轮廓上。这种做法便于入门，却可能错过最重要的问题：星官本来就是另一种组织天空的单位，它不必服从现代星座的划界原则。比较的第一步应是辨认差异，而不是立即把差异消除。

《史记·天官书》以星群、方位和社会秩序相互联系的方式展开天空。其开篇围绕天极及相邻星体安排不同身份，可以看出命名不仅用于识别亮点，也用来组织这些亮点之间的关系。[12-1] 这里存在一套能够被记忆、传授和调用的天空结构。政治语汇参与了结构的表达，但不能因此认定其中没有可用的观测知识。

星官至少具有三种相联的作用。它使观察者能够在星空中寻找位置；它为事件记录提供参照，如某天体出现在某星官附近；它让不同记录之间有机会建立同一性。三个作用都需要共同学习，但共同学习并不意味着所有时代的图像完全相同。名称可以稳定，所指恒星的辨认却仍可能需要版本与位置证据。

现代星座是一套包括天空区域边界的制度化划分。古代星官则不能简单理解成同样的区域容器。某条记载说天象在某星旁边，与说它发生在一个具有精确定义边界的区域内，提供的信息并不相同。若把古代语句直接翻译为现代区域归属，就可能制造原记录并未提供的位置精度。

敦煌星图提供了一个具体入口。对该星图的研究将手稿、星点、文本与投影方式结合分析，而不只欣赏它的图像外观。[12-2] 其意义在于，天空秩序已经能够以可携带、可复制的图面显露。图面使观察者不必每次从零开始组织星空，也使后来研究者能够比较不同部分的制图方式和可能误差。

可是，图上一个点并不等于当时一次独立实测。它可能来自既有目录、临摹、改绘或混合材料。颜色、文字和图形的不同，也需要先查明手稿本身的性质，不能随意解释为不同观测队伍。思想史研究若忽略这些材料层次，就会把知识产品直接当成制作过程。

星官的构成可以用 SDE 关系说明，但须避免空泛。某个可辨认星群作为 S，依靠反复指认、比较邻接关系和辨别明暗的 D 而稳定；E 则体现为星名、相对位置、观测方位与使用规则的相互规定。若名称指向改变，旧记录的定位方式也会改变；若参照星被误认，整组事件的位置理解都可能偏移。这里的纠缠有具体后果，不只是多个因素同时存在。

命名尤其不是给一个完全完成的对象贴标签。一个星官在观测、叙述和教习中的可辨认性，部分依靠名称所组织的联系。但名称也不能任意制造天空的几何关系。亮星位置与季节可见性会限制组织方式，而既定组织方式又影响哪些细节受到关注。这个双向限制，使天空既不是纯自然数据的堆积，也不是文化想象的无限投射。

从这里可以看出中国天文学世界的一个特点：天体身份经常嵌在可叙述的空间秩序中。某颗新出现的星之所以能够被记下，不仅因为它明亮，也因为已有一套能说明它“在哪里”的参照系统。稳定背景为异常提供了位置，异常又可能暴露背景描述的不足。对常态的组织与对例外的发现，因而并不相互排斥。

星图还是一种训练视线的工具。没有受过训练的人面对密集星点，未必能够立即看出稳定图形；熟悉某套星官的人，却能够把若干微弱星点同附近亮星联系起来。训练不改变恒星的物理位置，却改变可以稳定完成的辨认任务。这种变化不宜只叫主观想象，因为它必须在多次观察中经受实际位置关系的限制，也必须能被他人学习。

如果两位观察者对同一个星官的边缘成员意见不一，分歧可能来自视力、天空条件，也可能来自所学图本。此时，比较不能只问谁“看得更真”，而应检查各自调用的参照与判据。只有把这些条件展开，分歧才有机会转化为可处理的差异。科学世界的构成，正在这些看似细小的指认操作中发生。

星官与历算也不可直接合并。星官提供空间参照，历算组织时间与运动，二者可以合作，却不等于同一任务。一幅出色星图不能自动给出准确合朔时刻，一套有效日月算法也不能代替对某个客星位置的细致记录。把两类成果分别评价，再研究它们如何连接，比笼统称作“古代天文发达”更能说明实际能力。

12.2 官修资料的连续性

少见天象的研究需要超出个人生命的时间长度。一次显著客星可能只出现一段时间，观察者无法等待同一事件再次发生。记录的保存因此成为天文学能力的一部分。中国长期积累的天象资料，为后世提供了许多现代仪器无法返回当时重新取得的观察线索。

这种价值来自连续劳动，却不能被简化为一份无缺口的长时间序列。观察可能因天气、战乱、组织变化和选择标准而中断。记录写下之后，还要经过汇报、编纂、抄写和保存。今天所读到的历史书，往往不是当夜原始日志，而是经过多重处理的文本。每经过一层，某些信息可能被保留，另一些可能被压缩或改写。

因此，“有制度”与“有连续数据”必须分开。一个机构名义上长期存在，不保证每天采用同样方法；一条书中记载的日期，也不保证就是天象首次出现的物理时刻。首次发现、

首次上报与后来被编入史书的日期，可能相关，却不能未经论证地等同。史料工作所要做的，正是保持这些差别，而不是将其整理得过于平滑。

记录目的也影响内容。若某种天象被视作异常征兆，它可能得到特别详细的记载；若某类变化不符合当时分类，便可能被忽略或写成另一种现象。这不意味着记录失去科学价值，而是说价值必须通过理解选择机制来取得。一个有偏的资料集合仍然可能非常珍贵，前提是不把它假装成没有选择的随机样本。

同时，制度不只施加偏差，也创造可比较性。固定的纪年、岗位分工和报告程序，使分散观察有机会被放进共同时间尺度。它们减少了个人记忆的任意性，形成能够跨代调用的知识资产。若只谈权力如何塑造观察而不谈这种保存能力，会低估制度的认识功能；若只赞美记录丰富而不谈筛选，又会把制度理想化。

三方程在此最值得追踪的是记录如何改变后续观察。一个已被认定的重要异常 S，可以使机构加强某类巡视 D；长期巡视又会改变哪些现象被认为值得报告，从而调整分类与岗位规则 E。但具体历史中也可能出现相反情况：一种固定解释过于强大，使观察者反复追寻同类迹象，忽略其他变化。回写既能打开问题，也能缩窄问题。

这使“积累”不再只是条目越来越多。真正可用的积累，还要求名称、日期和出处能够彼此衔接。大量无法确定来源的记录，未必优于数量较少但层次清楚的记录。一个传统的历史力量，既体现在长期保存，也体现在后人能否重新检查保存下来的内容。

假如两部史书保存了同一天象，不能立即把它们算成两次独立观测。它们可能共同取自一个已经失传的来源。独立性必须通过文本关系考察：用语、日期、错误及叙述结构的相似，可能提示复制；不同观测地点和不同细节，则可能提供另一个方向的线索。现代统计若忽略来源依赖，会把证据数量夸大。

相反，两条描述略有差异，也不必然意味着其中一条错误。它们可能记录不同时间或不同地点的可见状态。史料批判不能预先要求所有文字完全一致，而应寻找能够解释差异的最小条件。若这种条件需要不断增加不可核实假设，则应降低匹配的可信度；若能够由已知时空条件自然说明，则差异反而可能增强材料的具体性。

这种工作同时连接自然科学与人文研究。天体计算帮助判断某种叙述是否可能，文献学帮助判断叙述究竟说了什么。任何一方都不能吞并另一方。精确计算不能修复一个被误读的星名，细致校勘也不能独自决定物理类型。两种能力在共同问题中相互约束，正是一个复合研究主体形成的过程。

12.3 历法改制的证据

历法把天文计算带入共同生活。它既要协调日月周期，又要形成可以发布和使用的时间安排。天文学中的某个参数改变，未必立即转化为一套新历法；反过来，一次政治上的改历，也不能自动证明全部算法发生根本革新。要理解中国历法，必须把测量、计算、制度颁行和实际使用放进同一个问题中，同时保持它们各自的证据责任。

十三世纪的授时历改革适合作为观察窗口。席文关于这次改革的专门研究，将技术文本的翻译同制度、政治与数学分析结合起来，提示我们不能凭一个著名年长度数字评判整个体系。[12-3] 年长度回答的是一个周期问题，历算还必须处理位置转换：同一个天体的运动，怎样在不同的天空参照之间表达？

黄道与赤道之间的换算提供了一个具体案例。Wagner 依据钱宝琮的描述重建授时历的一段计算：从黄道上的已知位置求赤道参照下的相应位置，计算利用一种与沈括弧长近似有关的方法，并通过求弧矢推进转换。[12-5] 这一论证的史料层级须保持清楚：它是现代学者依据专史描述所作的计算重建，作者明确说明自己没有直接核读《明史》的原计算文本。

这里的困难不在于给同一个数换名字。黄道与赤道并不重合，在一个参照上等量推进，在另一个参照上不必等量推进。若只保留平均步长，便可能遗漏随位置改变的几何关系。要把两类位置放在可比较框架内，需要一个随输入位置而变化的转换程序。这样的程序使不同观测和历算表达能够接合。

今天可以用三角函数重述这些步骤，但这种重述不是郭守敬已经使用现代三角函数的证据。现代符号有助于识别计算的功能，历史分析仍须保留原有弧、弦、矢及近似方法的组织方式。两种表达之间有可说明的对应，不等于它们拥有相同的概念语言。比较的价值正在于说明不同数学资源如何承担相近任务。

因此，评价改历中的数学工作，需要追踪输入如何经过一系列近似而成为输出，而不能只摘取某个常数。单独提高一个常数的精度，未必已经改善整段程序；各步骤的误差可能相互加强，也可能部分抵消。这是对组合算法的一般判断，不是对历史作者意图的推定。只有把实际步骤展开并重算，才能确认一种改变对最终位置究竟产生什么影响。

这一案例也限定了我们能够作出的历史结论。它表明授时历的研究不只关乎周期长短，还涉及天空参照之间可操作的数学联系；它不能独自证明所有预测项目同时优于前历。冬至、朔望和食的推算具有不同输入与判断条件，准确性必须附着于具体任务。把一项换算的成果扩大为全体系的绝对精确，同样会掩盖改革的真实结构。

但是，算法上的优越还需要实施。新表要被制作，使用者要学会，日期要被换算，地方机构要获得同一版本。颁行只是其中一环。若有人沿用旧本或混用参数，实际结果便不等于

设计者的结果。因此，一部历法的接受史需要与其数学结构并行研究，不能把中央认可日期当成全国计算实践瞬间改变的日期。

SDE 分析能够把这里的相互规定写得具体。历法输出 S 不是脱离程序的日期标签，它依赖观测与推算的 D，以及历元、单位、改正规则和制度用法之间的 E。新的输出可能使旧程序的不足暴露，也可能迫使教材和表格改变。于是，S 不只是终点，还会改变后续的 D 与 E。发生过程若停在一个著名人物“制定新历”，就会丢掉这一循环。

不过，循环不能被当作自动进步机制。一套新历法可能在初期适合，在长期使用后积累偏差；制度上的权威又可能增加修订成本。原先帮助统一操作的规则，后来可能成为难以改变的负担。稳定与创新的张力不是制度外部偶然添加的障碍，而是一个成功体系持续存在所必然面对的问题。

因此，评价授时历需要两个尺度。一个是它在自己的历史任务中怎样组织观测、算法和发布；另一个是今天依据可核资料重算时，它在哪些条件下表现如何。前者不能被今天的仪器精度抹去，后者也不能被历史赞誉替代。两种尺度结合，才能既理解历史成就，又保留可检验判断。

本章不将改革解释为“政治推动科学”或“科学服务政治”的单向关系。时间制度需要天文结果提供可用安排，天文实践又需要组织支持、共同标准和长期保存。这里的关系不是所有因素都同样重要，而是某些任务缺少其中一个条件就不能以原有方式成立。识别这些具体依赖，才使“纠缠”超过一般背景叙述。

现代重算也应当尊重历史输入。若给旧算法提供当时不可能得到的高精度参数，得到的结果只能说明算法在现代输入下的表现，不能直接代表古代实际能力。若故意采用最差版本的旧表与最优版本的新表比较，同样会夸大进步。公平比较需要同时记录算法、参数、文本版本和计算条件。

统一性、多样性与和谐性在历法改革中不是抽象口号。统一意味着共同日期必须有稳定规则；多样意味着不同地点和任务不能被一个无差别近似吞没；和谐则要求共同规则与地方条件之间存在可执行的转换。若统一取消了必要修正，预测可能失效；若地方修正失去共同接口，公共时间又难以维持。历法的技术价值就在这种张力中具体显露。

12.4 旧数据的新用途

第三章已经讨论古代客星如何取得对象身份，以及现代研究怎样检查历史事件与天体遗迹的匹配。本节不再重复一〇五四年客星的定位细节，而转向这种匹配赖以成立的长期条件。今天能够提出那些问题，首先因为名称、日期和叙述跨越了原观察者的生命，被保留下来。[12-4]

长期保存的认识价值，并不等于每份文献都是同样精确的数据表。一条记录可以保留可靠日期，却只给出宽泛位置；另一条可以描述位置关系，却缺少连续可见时间。将不同记录放在一起，首先要辨明它们各自保存了什么，不能为了统一数据库而补出原材料没有的精度。共同格式应当容纳信息的不齐整。

数字化也延续着这种选择。把名称转换成标准字段，便于检索和比较；保留原文、版本与转换理由，才能让后人检查标准化有没有改变原义。前者使材料能够大规模使用，后者保持材料接受重新解释的可能。倘若只留下换算后的数值，新的整齐可能以丢失证据关系为代价。

旧资料的新用途显示了两类世界之间的接续。古代星官世界中的一个客星，以相对位置、时间与可见状态获得身份；现代天体物理世界中的一个遗迹，以形态、辐射与动力学获得身份。将二者连接起来，需要建立桥接关系。桥接成功并不意味着两套对象体系从来完全相同，而是表明不同构成方式能够通过共同约束形成更广的解释。

这正是本书所说的发生性继承。新的研究没有简单取消旧的观察方式，而是让旧显露进入新的差异序列。S 变成后续 D 的资源，旧记录与现代测量一起参与新的 E。其结果既不是把古代思想现代化，也不是把现代解释退回古代，而是形成一种跨时代的证据联合。

联合也必须允许失败。一个候选遗迹如果在位置、年龄或性质上不合，应当被排除，而不能因某条古籍记载珍贵就勉强对应。某个古代事件也可能暂时找不到现代对应物。没有对应不等于记录虚假，有对应也不等于所有细节都已解释。保留这些中间状态，是历史天文学区别于传奇故事的条件。

中国历法与星官传统的长期意义，由此获得较准确的表达：它不仅保存了关于天空的意见，还保存了经由特定分类、操作和制度形成的比较能力。这种能力有选择、有局限，也有其他材料难以替代的时间深度。今天的任务不是取消其历史性质来取得数据，而是通过理解其历史性质，使数据真正可用。

从现代结论反查古代资料，是一种研究入口；从古代文本的分类与记录用途出发，再寻找现代可用约束，是另一种入口。前者善于提出匹配候选，却容易只选取有利句子；后者能够恢复资料语境，却可能迟迟不进入可检验问题。两条路径交替使用，才能兼顾解释力与材料纪律。六路径在此的作用，不是为研究贴上箭头，而是使单一路径的盲区被另一种入口暴露。

一个具体的失败条件是：如果某种匹配只能在删除相反记载、固定有争议译法并忽略日期误差后成立，它就不应被当作稳固发现。相反，若在合理的版本和误差范围内，主要位置与时间约束仍指向同一候选，结论才获得稳健性。所谓稳健，并非所有细节都被消除，而是关键判断不依赖某个任意选择。

保留这种限度，才能让古代观察继续参与现代研究，而不必被改写成现代观察。历史材料的尊严与科学推断的严格性，可以在同一工作中得到维护；它们相互支持，而不是彼此抵消。

本章参考文献

[12-1] 司马迁：《史记·天官书》。中国哲学书电子化计划，在线文本。https://ctext.org/wiki.pl?chapter=574568&if=gb

[12-2] Jean-Marc Bonnet-Bidaud, Françoise Praderie, Susan Whitfield. The Dunhuang Chinese Sky: A Comprehensive Study of the Oldest Known Star Atlas. 2009。https://arxiv.org/abs/0906.3034

[12-3] Nathan Sivin. Granting the Seasons: The Chinese Astronomical Reform of 1280, With a Study of Its Many Dimensions and a Translation of its Records. Springer, 2009。此处核读出版社研究说明与目录；未据未读细节填入参数或原典引文。https://link.springer.com/book/10.1007/978-0-387-78956-9

[12-4] D. A. Green, F. R. Stephenson. The Historical Supernovae. 2003。https://arxiv.org/abs/astro-ph/0301603

[12-5] Donald B. Wagner. Guo Shoujing’s conversion of ecliptic to equatorial coordinates. 2021。核读网页文字；作者依据钱宝琮描述作现代重建，本文未逐式重算图示公式。https://donwagner.dk/gsj/gsj.html

第 13 章 伊斯兰世界的天文表与观测台

13.1 观测台为何重要

一座观测台能够建成，并不意味着新知识已经完成。建筑提供稳定空间，仪器提供测量条件，资助维持人员工作，但这些条件需要进入具体程序，才能成为可以比较的结果。如果新仪器的读数未经校准，新观察仍被旧参数覆盖，宏大投入也可能没有产生相应的天文学改变。观测台的思想史意义，必须落实在这个转化过程中。

十三世纪的马拉盖是一个重要案例。图西参与建立并主持这一观测机构，研究者、仪器、藏书和教育活动在这里汇集。有关该项目的研究把它放进蒙古统治时期的赞助网络，同时注意到其活动超出了简单编表任务。[13-1] 这里的关键不是把它称作现代研究所的提前诞生，而是辨认它如何使原本分散的工作能够持续关联。

共同使用仪器能够建立相近的读数条件，但也可能把同一种系统误差传给所有人。因此，协作的价值不只是人数增加，还包括比较、复查和发现共同偏差的机会。若所有测量都依赖同一错误零点，多次一致不会自动变成正确。这个简单事实说明，知识共同体必须组织差异，而不只是制造一致。

仪器的几何结构、观察者的身体操作和数据的书写方式，构成实际测量链。瞄准一颗星、读取刻度、记下时刻、进行坐标转换，各步骤的错误性质不同。有些可以通过重复减小，有些需要重新确定基准，有些则来自计算规则。把最终误差归给某个“落后的仪器”，可能掩盖问题真正出现的位置。

观测台也改变研究的时间组织。单次观察可以由个人完成，跨季节、跨年的比较却需要持续安排。有人必须保存旧记录，使后来观察能够与之相接；有人必须懂得相同的符号和算法，使数据不随个人离去而失去可用性。这些工作构成一种跨人的记忆，其可靠性既依赖制度，也依赖技术细节。

然而，稳定资助不等于稳定研究方向。宫廷需要的预测可能推动特定表格编制，却不一定奖励所有理论问题。理论家对古典模型的批评，也可能有自己的数学动机，不能都解释为使用者要求“更准”。马拉盖相关成果中，理论构造与实际历表的创新程度并不必然相同，专业研究对此有明确区分。[13-2]

从 SDE 看，观测台不是位于知识外部的一片背景。共同仪器的零点、表格的历元和观测时刻相互限定，改变其中之一可能要求其他环节一起改算。这种相互规定才是 E。研究人员在不同日期、不同目标和不同模型之间进行比较，形成 D；可以发表、传授和重用的结果构成 S。建筑和资助只有通过这些具体关系进入过程，才具有可解释的知识作用。

因此，衡量一个观测台不应只问它拥有多大的仪器或多少学者，还要问它能否把差异保存下来，使后人知道某个参数为何被改变。一个结论如果只留下权威署名，没有可追踪的操作，传播时容易变成不可挑战的成品。若保留了推算与比较的接口，它便可能成为新研究的起点。

这也改变了对个人贡献的理解。主持者可能提出方向，制作者保证仪器可用，计算者把读数转为位置，抄写者保存结果。不能因为后世书名只保留一个名字，就认定全部能力属于此人；也不能因为需要多人，就抹去某个构造或判断的个人原创。复合主体的分析必须同时保存分工与责任，而不是把一切笼统归于共同体。

一个现实的检查方法，是看成果能否离开原主持者继续使用。如果别人只有在作者在场时才能得到结果，说明许多操作尚未被明确；如果表格与说明足以支持重算，知识便获得更强的可传递性。但是，可传递并不等于永远正确。恰好因为更多人可以使用，新的适用边界才可能被发现。

13.2 zij 的知识形态

通常被称为 zij 的天文表书，并非只有孤立数字。不同作品的内容并不完全相同，但常把时间换算、三角工具、天体运动和使用说明组织在一起。它们像一组相互配合的工作资源，使计算者能够从具体日期进入天象预测。其历史力量在于把复杂知识压缩成可调用形式，同时保留一定的更新空间。

《伊儿汗历表》的存世抄本提醒我们，表书也是具体物件。惠康收藏所登记的相关材料以波斯语保存，作为有特定馆藏编号和叶面范围的手稿存在。[13-3] 这些书目事实并不能代替内容研究，却能防止把一部著作想象成从诞生起便只有一个固定版本。表格在抄写、使用和再编过程中，可能发生局部变化。

读取一张表，首先要知道它回答什么问题。一个角度值可能是平均位置，也可能是对平均值的改正；一个时间差可能指经度差造成的地方时差，也可能属于另一种时间换算。若只抽取数字而不读表头和说明，现代复算很容易把不同性质的量混在一起。表格的可压缩性，恰好以使用者掌握接口为前提。

其次要知道它的起点。某个历元、某个地点和某套单位，经常共同决定表值的解释。把表移到另一地点，并不意味着所有关系都失效，但需要明确哪些部分保持不变，哪些需要校正。正是这种可分解的结构，使一部表书能够旅行：数学关系可能迁移，地方参数则必须重新安排。

可以用一个简化的现代例子理解。若表给出基准地点某时刻的天体位置，另一地点的计算者需要将自己的时间转换到同一基准。若只改地点名称而不改时间接口，得到的结果可能

系统偏移。这个偏移不证明几何理论错误，而说明知识的迁移不是复制页面那么简单。表的适用性由内部一致性与外部接口共同决定。

表书因此具有一种双重性质。对于日常使用者，它封装了大量计算，使结果可以方便获得；对于修订者，它又是一组可以拆开的假设、参数与规则。封装提高效率，却可能隐藏来源。当后人不知道某个数字来自新测量还是旧表沿用，就难以判断新的误差应该怎样处理。

思想史必须区分“新书”与“新数据”。一部新编表书可以在编排和使用便利性上有创新，同时继续使用大量旧参数。这并不使它毫无贡献，只是贡献应被准确命名。相反，旧式表格中一个关键参数的改变，可能对预测产生广泛影响，不能因为外观传统就被忽视。

三方程在表书上具有清楚的操作含义。一个输出 S 若在应用中持续偏离观察，会促使使用者改变比较程序 D；若偏差源于历元或参数联系，E 也需要修订。但如果表书已经成为教学权威，人们可能优先怀疑新观察。回写并不总向创新方向流动：成功形成的可信度，也可能在新情境中延迟纠错。

因此，表书既是知识保存工具，也是研究选择的分配工具。容易从表中获得的量，更容易被拿来比较；未被表格组织的现象，则可能需要额外劳动才能进入议题。一个新的表格结构，有时比一个新的单独数值更能改变研究方向，因为它改变了什么问题可以方便地提出。

表书中的错误也具有不同层次。单个数字误抄，可能通过邻项关系或其他版本发现；整列使用错误单位，则可能保持表面规律，却在应用时持续偏离。更隐蔽的是一项默认前提已经改变，而说明仍沿用旧说法。这些情况要求修订者不仅检查数字，也检查表与任务之间的联系。

因此，对古代计算能力的评价不应只取一条最准确结果，也不应只举一个明显错误。需要考察误差怎样分布，是否集中在特定时间、角度或天体，以及使用者有没有办法识别它。一个体系的成熟，有时不表现为完全没有错误，而表现为错误能够被定位，修正能够被传递。

13.3 模型修订的具体位置

对托勒密传统的继承，不应被理解为长时间的被动保存。伊斯兰世界的天文学家既处理参数与观测问题，也批评某些模型构造同所要求的运动原则之间的不一致。两类工作相关，却具有不同判据。新测量可以要求调整参数；即使数值预测尚可，理论一致性问题也可能推动几何重构。

例如，若一个模型要求天体由均匀圆周运动组成，那么它怎样使用偏心和均衡点就不只是计算技巧问题。研究者会问，模型中的运动是否能够由所认可的球体或圆周安排实现。这

里发生的张力，来自预测效力与物理—几何规范之间，而不是单纯来自“数字不准”。[13-2]

需要分辨参数修订和结构修订。改变某个平均速度或起始角度，仍可在原有构造中进行；改变运动中心、组合方式或容许的运动种类，则改变了模型空间。前者可能明显改善预测，后者也可能主要改善一致性。两者没有固定的价值等级，应按解决的问题判断。

把理论创新全部归因于新观测，会忽略数学问题本身的生成力。反过来，把几何重构全解释为纯粹理性，也会忘记模型仍要处理具体天体。一个构造如果不能接入经度、纬度或距离问题，再漂亮也未必成为天文学工具。数学效力与天文适用性需要分别证明，然后说明它们如何结合。

马拉盖传统的重要性正可以在这里把握。专业研究区分图西在实际表书与理论天文学中的工作，并指出理论构造并非简单由更近代的观测直接产生。[13-2] 这为思想史提供一个反例：机构投入、新资料和理论突破之间不存在机械的一一对应。某些创新可能来自长期继承的问题，在新组织条件下获得继续工作的机会。

SDE 的分析若要有增量，就应识别这种非同步关系。D 可以是观测差异序列，也可以是不同几何候选之间的推演比较；E 可以包含匀速运动规范与可计算性的相互约束；S 则可能是一种新的构造，而不一定是一份更准确的数据。若预先把 D 仅理解为经验误差，便会错失理论内部的发生机制。

但这也要求避免把所有逻辑比较都泛称创新。一个新构造必须说明它解决了什么旧问题，又产生了哪些新负担。若为保持一个原则而增加许多难以解释的结构，研究者需要评估代价。理论一致性可以提高，同时计算便利性下降；某一方向改善，不等于整体无条件更优。

一个有用的反事实问题是：如果没有新增观测，这项几何修改是否仍有提出的理由？若答案是有，便说明它部分由已有模型的内部张力推动。再问：如果放弃匀速圆周要求，这个修改是否仍然必要？答案若改变，就揭示了规范条件如何参与创新。这样的比较比宣布“数学推动天文学”更具体。

另一个问题是，修订是否进入实际使用。一篇理论论著提出新模型，不保证历表立即按它重算。使用者可能继续采用熟悉表格，因为它足以完成任务，或者新构造尚未转成便利算法。理论与计算之间的距离，本身就是知识发生的重要部分，不应以“成果已经发表”将其跳过。

若将模型修订看成研究者在候选空间中的选择，就能进一步问：哪些候选根本没有进入比较？一个共同体若坚持某种运动规范，其搜索可能在规范内非常丰富，在规范外却十分有

限。我们不能因此简单说它缺少创造力，因为创造本来依赖某种可工作的限制；但也不能把内部丰富性当成不存在盲区的证明。

同样，后来的成功理论也不能成为唯一尺度。某项几何修订即使没有通向现代天文学，仍可能解决了当时一个明确问题。它的历史价值可以通过问题与解法的关系来判断，而不必借后世胜利来授权。这样处理，才使思想史真正容纳多样路径，而不是把所有支线写成偏离正道。

13.4 传播的证据链

天文表和模型能够跨越语言与政治边界，但传播不是一条自动向西流动的河。它包含选择、翻译、抄写、讲授和再使用。某些内容被保留，另一些被删去，还有一些在新问题中获得原来没有的功能。因此，真正的交流史应当追踪具体物件与操作，而不只画文明之间的箭头。

最直接的证据包括可确定年代的译本、明确引用以及可以识别的文本对应。图形和算法的相似也很重要，但其强度需要结合复杂程度判断。一种简单构造可能独立产生；一组不常见的结构、符号和排列同时对应，则更值得追查。即使如此，仍需区分一般传播可能与某个人实际接触某个版本的证明。

不能因为一条传播路径尚未完全恢复，就断言没有交流；也不能因为交流在历史上可能，就替缺失环节编出故事。两种极端都把不确定性提前消除了。合理做法是分层陈述：哪些对应已经确认，哪些中介有证据，哪些环节仍只能作为假说。读者由此能够看见判断的承重位置。

从结果 S 向前寻找来源，是有效但危险的路径。它容易把早期人物写成后来成果的准备者。换成从当时 E 和 D 出发，先看研究者自己的问题，再追踪方法后来怎样被吸收，便能恢复早期工作的独立意义。六路径的互查，在这里不是装饰，而是阻止后见之明吞没历史问题。

传播也可能发生在一个整体内部的局部。接受者可以借用三角工具，却不接受相同宇宙论；可以使用星表，却重写时间起点；可以吸收几何装置，再放进另一种行星安排。这类部分继承使知识具有延续性，也使“同一理论”的边界变得复杂。工具的生命往往长于最初赋予它意义的体系。

由此可见，观测台、表书与理论不是互不相干的三项成就。观测台组织持续比较，表书封装可调用关系，理论为修订提供候选结构。它们相互支持，却也可能步调不一。某份表可以长期流传，而原机构消失；某个理论构造可以在新地区重获用途，而最初的使用目标被遗忘。

本章主张的新的天文学世界，正是在这些可分离又可重接的环节中形成。其新，不只是得到更多数字，而是形成新的对象处理能力：可以用共同基准比较，可以将结果带到另一地点，可以让后人指出某个环节何以失效。这样的世界既有稳定结构，又容许内部改造。

评价它不需要把整个时期描写成单一黄金时代，也不需要以欧洲后来采用多少来决定价值。更好的尺度，是它在自己的问题中建立了什么能力，以及这些能力后来能够怎样继续发生。图西偶将提供一个更集中的例子：一个数学构造怎样在保留若干旧原则时，改变天体运动的表达空间。

本章的解释还有一个可否定条件：如果具体历表并未使用某项新理论，就不能仅凭同在一个机构出现而宣布二者已经融合。人物之间见过面、机构之间有联系，都只是进一步研究的入口。知识关联需要在算法、引用或使用上找到痕迹。这个要求使社会史与技术史能够相互制约，而不是用一方的材料替另一方填空。

由机构进入研究，能够看见资源和训练；由表格进入研究，能够发现实际沿用；由理论进入研究，能够揭示构造目标。三种入口所得结论若不一致，正应当成为分析重点。它可能说明新思想尚未进入实践，也可能说明实践变化没有被理论叙述充分反映。差异不必被统稿抹平，它可以成为新的问题。

本章参考文献

[13-1] Hadi Joráti. The Maragha Observatory Complex in Ilkhanid Iran. Max Planck Institute for the History of Science，项目说明。https://www.mpiwg-berlin.mpg.de/research/projects/maragha-observatory-complex-ilkhanid-iran

[13-2] George Saliba. Ṭusi, Naṣir-al-Din ii. As Mathematician and Astronomer. Encyclopaedia Iranica。https://www.iranicaonline.org/articles/tusi-nasir-al-din-mathematician-astronomer/

[13-3] Wellcome Collection. Zīj-i īlkhānī，MS Arabic 867，馆藏记录。只据目录确认材料身份，未声称完成波斯语手稿校勘。https://wellcomecollection.org/works/xnnszz6u

第 14 章 图西偶与行星运动的几何改造

14.1 几何装置的生成

两个圆周运动能否合成沿直线往返的运动？今天可以很快用坐标证明，历史上却是一个具有天文学意义的问题。它关系到研究者能否在保持某些基本运动规范的同时，表达原本难以容纳的变化。图西偶的重要性，正来自这一局部而深刻的转换：不先放弃整个宇宙框架，而先扩大框架内部能够构造的运动种类。

纳西尔丁·图西研究托勒密理论中的几何困难，提出后来被称作图西偶的构造。相关专业研究区分它的不同形式和用途，不能把所有形式都看作同一张简单动画。[14-1] 本节只解释其中最容易明确验证的平面直线形式，再讨论这种数学效力怎样进入天文学。

设大圆半径为 2r，小圆半径为 r，小圆在大圆内部相切。取大圆圆心为原点，小圆圆心以角 θ 沿半径 r 的圆运动。若小圆上选定点相对自身圆心的方向为−θ，则该点的位置是两个长度为 r 的向量之和。这里的角度以固定坐标系为参照，须与相对旋转坐标的角度区分。

用现代坐标表示，小圆圆心为(r cosθ，r sinθ)，选定点相对小圆圆心的向量为(r cosθ，−r sinθ)。两者相加得到(2r cosθ，0)。纵坐标始终为零，因而该点沿大圆的一条直径往返。这个证明是现代说明，不是声称图西使用笛卡尔坐标。它展示了构造为何有效，也使适用条件清楚可查。

若以连接两圆圆心的转动方向作局部参照，小圆上的选定方向相对它转过−2θ；这就是常见描述中“反向两倍”的来源之一。若不说明参照，人们容易把绝对转角与相对转角混在一起，画出不满足直线性质的图。一个简单数学例子已经表明：运动特征不是孤立的速度标签，而由参照关系共同规定。

构造的精确性依赖半径比和运动配合。把小圆半径任意改变，或让两个向量角度不再对称，纵向分量便不一定抵消。直线运动不是两个圆随便叠加后“涌现”的神秘结果，而是特定约束下的严格结果。这一点对发生学尤其重要：新显露需要具体生成条件，不能只用整体性一词代替证明。

同样重要的是，图西偶输出的直线往返并不是匀速直线运动。现代说明中的位置为 2r cosθ，即使 θ 随时间匀速变化，点沿直线的速度仍会改变，在端点处转向。这里改变的是轨迹表达能力，而不是得到任意想要的直线运动。若忽略这一限制，就会夸大构造能够解决的天文问题。

把这个例子写成 S=F(D,E)，S 是沿固定直径的往返轨迹，D 是两种圆周运动的相位变化，E 是半径比、方向与参照的相互规定。E 不是旁边另列的一组背景，因为缺少这些规

定，所谓 D 也不再是同一个差异运行。新的轨迹特征属于复合构造，而不能归给任何一个圆周运动单独拥有。

还可以通过改变条件来检验这个判断。若两个向量具有不同长度，纵向分量不再完全抵消，点的轨迹一般不再是那条直线。若长度相同而相位关系出现固定偏移，直线方向或轨迹性质也需要重新计算。由此可见，“两个圆生成直线”不是一句普遍口号，而是一组参数关系的结果。发生学在此应当承担把条件写全的责任。

这一构造还有教学上的启发。只观看一幅静止图，读者可能只看到两个相切的圆；只有跟踪选定点的连续位置，才能认识新轨迹。因此，新特征不在任何单一瞬间的图形中充分呈现，它依赖差异序列被组织和比较。这里的 D 具有不可省略的作用：没有连续关系，就没有所要证明的往返特征。

14.2 在旧宇宙框架内创新

为什么天文学家要费力把直线往返写成圆周组合？如果只看现代数学，直接写出直线运动似乎更简单。但历史问题不只要求得到一条轨迹，还要求轨迹由可接受的基本运动组成。图西的创新同均匀圆周运动的规范相联系，这个规范同时限制和激发了几何创造。[14-1]

一种规范通常被描述为阻碍创新，图西偶却显示另一种可能：规范也能形成明确难题，使研究者寻找此前没有的构造。若允许任意曲线任意变化，拟合某些位置可能不难，却会失去对模型组成的约束。保留限制，反而迫使数学工作更具体。问题不在于限制必然有益，而在于它是否能够产生可检验的新关系。

这里不能把“局部创新”理解为不重要的小修补。一个新的组合方式，可能扩大整个模型体系的表达范围，让原来不兼容的要求在某些条件下共同满足。可是，它也不自动构成整体宇宙论革命。图西没有因为使用这个构造就成为日心论者；平面几何定理的成立，也不意味着地球运动已经得到论证。

区分这两层，能够避免后见之明。一种写法从哥白尼出发，把图西看成尚未走完道路的先驱；另一种写法从图西自己的问题出发，考察他怎样处理既有理论的困难。后者并不否定后续影响，而是先给予早期创新独立的解释位置。它让研究者的目标不被未来结局重新定义。

图西偶进入月亮或行星模型时，还需要额外结构。哪个点代表哪个几何部件，哪种往返对应所需变化，以及组合后能否保持整体模型的一致性，都不是直线定理本身已经证明的。数学装置可以正确，具体天文使用却仍有问题。就像一块零件尺寸合格，不等于装入任何系统都会改善性能。

因此，需要分别评价三种成功：构造在数学上成立，构造满足模型中的某项要求，整个模型在天文学上可接受。第一种最容易严格证明；第二种需要检查接口；第三种还涉及观测、其他运动和物理解释。把三种成功混为一谈，会把一个真实但有限的创新夸张成包办一切的解决方案。

SDE 方法在这里提出一个具体问题：新显露 S 出现之后，是否改变了研究者允许比较的 D 与能够组合的 E？若后来工作能够在不改变全部前提的情况下使用这种装置，则候选模型空间确实扩大了。这是可追踪的回写。至于这种扩大是否最终使旧框架被超越，则需要另外的历史证据，不能从几何定理直接推出。

这一机制也可称为内部生成：新事物并非总要从彻底否定既有秩序开始。保持部分共同约束，改变其他关系，可以生成具有新特征的整体。其局限同样清楚：若某个根本困难来自被保留的前提，局部构造可能只能延缓问题，而不能解决它。承认这种两面性，才能公平评价旧框架中的创新。

这也说明审美判断不能简单等同于圆的偏好。保持均匀运动可以被看作一致性要求，增加构造则可能降低简洁性。不同价值在一个模型中发生张力，研究者需要决定何者值得优先满足。我们不能预先把旧模型的复杂都判为丑，把后来模型的复杂都解释成深刻；必须用相同尺度比较。

若一种构造在保留核心规范的同时解决一个明确矛盾，它可能提高体系的和谐性。但这种和谐是有范围的，不应扩大成宇宙整体已经协调。部分问题仍然存在，恰恰说明创新没有结束。新构造建立的统一，可以成为下一轮差异发生的条件，而非封闭所有提问的终点。

14.3 传播的史料门槛

图西偶在哥白尼工作中的对应，长期引起历史研究。重要问题不只是两个图形像不像，而是相似性应该支持多强的历史结论。专业学者对传播证据和独立发现可能提出过不同判断。[14-2][14-3] 本章保留这种争论，不用一个简单的“直接借用”或“完全独立”替代材料分析。

最容易确认的是数学对应。如果两个构造具有相同半径关系和运动组合，它们可以在数学上等价。接着才能问，是否存在文本、图像或论证次序的进一步对应。再往后，才是某个版本在特定地点与年代是否可达，以及当事人是否有接触证据。每上升一层，所需证据都会增加。

相似性强弱也不能只凭直觉。两个常见圆相切，很容易独立画出；若同样的特殊标记、论证顺序和非必要细节反复对应，就更难用简单巧合解释。可是，强相似仍不能自动补出全

部传播线路。它可以提高传播假说的可信度，而具体经由谁、哪种语言和何时传递，仍可能未明。

萨利巴有关知识传播的讨论指出希腊语手稿等材料的重要性，同时也明确涉及哥白尼能否直接阅读阿拉伯语的问题。[14-2] 这提醒我们，传播不必由一部完整拉丁译本完成。图形、口头讲授、节录和中介语言都可能起作用。承认多种渠道，使研究更开放；但渠道可能性仍须同实际证据分开。

反对传播推断的研究则要求检验相似性是否足以排除独立构造，以及论证是否选择性使用材料。[14-3] 这样的质疑值得认真对待，因为历史解释不能仅以符合宏观交流图景为理由成立。但质疑也不能把“没有找到完整路径”当成“没有发生传播”的证明。材料残缺是历史研究的常态，结论应当与证据强度相称。

可以将争论拆成几个不同命题：某种方法已在伊斯兰世界出现；其相关内容后来进入其他语言；哥白尼工作中存在相应结构；哥白尼实际通过某条可指认路线获得它。前三者各自获得支持，并不使第四者自动完整。反过来，第四者未完全解决，也不抹去前三者。这种分层能够减少争论中不必要的二选一。

对 SDE 而言，传播也不是成品 S 沿管道移动。接受者必须把构造接到自己的 D 与 E 中，才能成为有效工具。即使来源已经证实，仍要解释用途怎样改变；即使独立发现成立，也要比较两个发生过程为何在某些约束下得到相似结构。两种历史可能都可以研究，不能让理论框架预先决定结论。

因此，本章采取克制的判断：图西偶的数学性质可严格说明，其在不同传统中的出现和对应值得实质比较，具体传播程度须依文献证据分别陈述。这样的表述并非回避问题，而是把可以证明的内容与仍需研究的内容放在各自位置。只有这样，跨文明研究才能避免变成贡献归属的竞赛。

关于地球运动的讨论还构成一个相邻但不同的证据问题。拉杰普比较图西与哥白尼相关论证，关注的不只是几何装置，也包括如何评价地球静止的证明。[14-4] 这提示我们，思想传播可以涉及论证方式，而非仅仅一幅图。即便如此，也要分别核验每一种对应，不能用一个领域中的联系替所有领域作担保。

同理，不能把某一争论尚未解决解释成跨文明研究没有成果。研究能够缩小可能范围、排除某些过强说法、找到中介材料，本身就是进展。只有把成功定义为得到唯一且完整的故事，才会迫使史家在材料不足时过早定案。保留分层的不确定性，反而更有利于后续发现进入并修正判断。

14.4 工具如何跨理论生存

宇宙框架变化时，某些数学工具为何继续有效？原因之一是工具具有相对可分离的接口。只要新的模型仍需要某类轨迹关系，就可以重新使用相同构造，而不接受它最初服务的全部宇宙论。数学不变量提供了延续，新的任务则赋予它不同意义。

但可分离不等于完全中性。工具总会让某些问题容易处理，让另一些问题困难。圆周组合擅长表达周期性关系，也会使研究者优先在这种表达空间中搜索。它既是能力，也是搜索范围的约束。只有当目标、资料和其他理论要求发生变化，这种约束才可能显露为不足。

由此可以理解一种看似矛盾的历史现象：同一工具既能帮助维护旧体系，也能参与新体系的形成。这不是工具具有两种神秘本质，而是它进入了不同的特征纠缠。局部数学关系保持不变，整体对象的解释联系却改变了。统一性在局部成立，多样性在整体用途上展开。

图西偶尤其说明，思想创新不能只按“是否改变中心”来评价。一个地心框架中的几何构造，可以具有重要原创性；一个日心安排也可能大量使用继承的数学资源。旧与新并不是整齐分隔的两个包裹，而是不同层次的保留、改造与放弃。把这些层次写清楚，比用革命或保守给人物分类更有解释力。

在三原理的分析语言中，这里存在规范与表达能力之间的张力，构造改变使某些张力暂时得到处理，后续使用又把成果回写为新的研究资源。但不能把这一结构写成历史必经三步。可能有构造被提出却无人采用，也可能被采用但未增加预测力，还可能在不同问题上产生相反效果。动力解释必须容纳这些分支。

如果一个后续模型采用图西偶，只是因为它方便表达，而并不提高观测符合程度，仍可以承认其工具价值，却不应宣称经验上更真。如果它改善一致性却增加运算负担，便需要说明这种交换是否值得。如果新的理论最终不再需要它，也不意味着此前贡献被取消，而是它的适用范围发生了变化。

这个例子还提示我们如何理解知识的寿命。知识不必以原来的完整形式永久保存，才算具有生命力。一个构造可以从原宇宙论中分离，进入新的教科书或模型，继续产生作用；也可以在某些任务中退出。传播、转用和退出共同组成其历史，而不是只有“影响越来越大”才算结局。

因此，图西偶给本书的贡献不是一个漂亮的跨文明符号，而是一项可验证的发生学分析：新轨迹来自特定运动关系，天文意义来自模型接口，历史影响来自可追踪的使用。数学证明、天文学解释和传播证据三者相接，却互不代替。新的天文学世界，正由这种既能拆分又能重组的能力不断形成。

知识工具的这种生命方式，也使统一性与多样性获得具体含义。同一个几何证明提供共同核，多种模型提供不同用途，适用条件则限制这些用途不能任意扩张。和谐性表现为共同核与具体用途之间的接口成立，而不是所有用途都被说成一样。这个关系可以被验证，也可能被推翻。

若后续使用只保留外观相似而破坏关键半径比，就不能称作有效继承；若保持数学条件却改变物理解释，则应称作重新应用。两者不同，说明思想史必须同时检查形式和功能。图西偶之所以值得单列一章，就在于它让这种双重检查变得异常清楚。

因此，形式可以延续，功能可以改变；二者之间的具体联系，正是历史解释需要承担的工作。

本章参考文献

[14-1] F. Jamil Ragep. Ṭūsī. The Biographical Encyclopedia of Astronomers, 2007，1153—1155。麦吉尔大学公开作者文本。https://islamsci.mcgill.ca/RASI/BEA/Tusi_BEA.htm

[14-2] George Saliba. 关于图西偶与哥白尼的传播问题，哥伦比亚大学研究展示。https://www.columbia.edu/~gas1/project/visions/case1/sci.2.html

[14-3] Viktor Blåsjö. A Critique of the Arguments for Maragha Influence on Copernicus. Journal for the History of Astronomy, 2014。本文据公开摘要确认反对论点的存在，不声称解决全部争议。https://journals.sagepub.com/doi/10.1177/002182861404500203

[14-4] F. Jamil Ragep. Ṭūsī and Copernicus: The Earth’s Motion in Context. Science in Context, 14, 2001, 145—163。https://www.cambridge.org/core/journals/science-in-context/article/abs/tdotuusi-and-copernicus-the-earths-motion-in-context/5C1FCB3528509213BD5CB1B32CBD1D72

第 15 章 乌鲁伯格星表与精度共同体

15.1 大型仪器与星表

把古代仪器做大，为什么可能看得更准？一个直接理由是，相同角度在更大的圆弧上对应更长距离，刻度因而更容易分辨。但这个优点附带条件：支架必须稳定，几何形状必须可靠，视准与读数程序必须配合。尺寸放大只改善某些误差来源，不能消除所有误差。大型仪器的历史意义，要在这种有条件的能力提升中理解。

乌鲁伯格及撒马尔罕天文学工作的声誉，经常同宏大观测设施联系。但更能接受检验的成果，是留下来的星表与计算资料。相关天文表书包含时间、三角、天体运动和恒星目录等内容；研究者列出的参与者也表明，这不是一人凭个人视力完成的工程。[15-1] 目录因此是团队操作凝结成的知识产品。

一份星表看上去由许多独立条目组成，实际依赖共同基准。每颗星的坐标都要相对于规定的坐标系统与历元表达。如果基准存在偏差，大量条目可能一起偏移。相同错误在许多星上重复出现，并不意味着每次观察都独立犯错，而可能意味着整个系统共享了一个错误条件。

以下将精度分解为多环节能力，是今人的分析；不能据此推定撒马尔罕研究者已经逐项采用现代误差分类。这使精度成为复合能力。观察者辨认目标的能力、仪器测角的能力、计算者转换坐标的能力，以及记录者保留数值的能力，共同决定最终结果。把全部精度归于某位君王的天才，遮蔽了分工；把全部结果归于仪器，又忽略了仪器必须经人操作和校验才能发挥作用。

在 SDE 语言中，星表位置 S 依赖跨夜测量与比较 D，也依赖基准、历元和对象身份之间的 E。一个位置数值并不是某颗恒星独立拥有的无条件标签，而是特定参照关系中的显露。恒星没有因此由测量者创造，但“可进入共同目录并同其他星比较的坐标对象”，确实在这一复合实践中形成。

因此，精度共同体的关键不只是共同相信一个数字，而是能够共同说明数字怎样取得。保存条件，使异议有入口；保存误差，使后人知道结果的限度。这样的共同体并非没有分歧，而是使分歧能够转成更细的检查。

在仪器层面，重复读数与校准也不可混同。重复能够显示波动，却未必暴露共同偏置。若每次都从同一错误起点开始，一百次读数可能非常集中，却仍然整体偏离。因此，一致性是一项证据，但不是充分保证。需要不同角度、不同目标或不同程序的比较，才可能看见同一种方法自己遮蔽的问题。

这为“共同体”一词增加了要求。它不只是共享同一设施的人群，而应能够容纳部分独立的检查。独立并不意味着完全没有共同条件，而是至少让某些关键误差来源不同。如果所有结果经过同一处理而被统一，原有差异可能在发布前消失，后人便无法再判断一致究竟来自天空还是来自处理程序。

15.2 重新测量旧星

重新测量旧星，与把旧星表换算到新年代，是两种不同工作。后者可以在原有数据上施加岁差等修正，前者则重新建立观测联系。两者都可能必要，但其独立证据价值不同。如果把换算误认为新测量，就会高估资料的独立性；如果把重新测量看成仅仅抄录，则会低估实际劳动。

现代研究将托勒密与乌鲁伯格星表整理为机器可读形式，并同现代星表比较。[15-2] 这种工作显示，评价古代目录不能只读序言中的成就宣称，还要进入条目、版本和对应恒星。目录中的不同量也可能具有不同来源，不能因为位于同一行，就假定全部出自同一次观测。

恒星同一性是比较的第一道门槛。古代描述可能以星座图形中的部位指认目标，现代目录则使用编号与坐标。两者的对应需要考虑邻近星、亮度与历史描述。若目标认错，即使坐标换算非常精确，得到的“测量误差”也没有意义。身份判断先于误差统计，但又会受到位置差异的反向检验。

历元则是第二道门槛。天球坐标随参照系统的长期变化而改变，恒星自身也可能有位置变化。不能把几百年前的坐标直接减去今天的数值，再把全部差额归为古代错误。比较必须先把相关条件统一，或在模型中明确处理它们。这个步骤看似技术细节，实际决定了比较对象是否成立。

第三道门槛是文本版本。抄本中的数字可能因传写发生变化，后来的编辑也可能作出校订。现代研究采用哪一版，便应说明其判断适用于哪个版本。若一个极端离群点来自抄写错误，它对原观测能力的含义，与真正测量偏差不同。把二者混合，会歪曲历史评价。

这里可以看出 E 为何不能只是“文化环境”。历元规定坐标的比较方式，身份决定条目配对，版本决定数值是什么。三个环节相互限定，任何一项变化都可能改变误差分布。史料学与数学计算因此不是两种先后分开的装饰性工作，而是在具体比较中互相构成。

重新测量的价值，也不只在于替换旧数字。它建立了一个可能反对传统权威的操作途径：既有星表不再是最终答案，而成为需要与新观察比较的资源。与此同时，新表一旦被广泛接受，又可能成为后来人的权威。知识从突破旧封装到形成新封装，具有连续的双重运动。

比较还有一个容易忽视的方向：旧目录能够帮助认出新观测目标，新观测又能帮助纠正旧目录。两者互相支持，也可能互相牵制。如果观察者预先相信某颗星应位于某处，便可能在附近选中符合预期的目标。没有独立检查的对应，会把旧错误带进新资料，制造“重新测量已经确认”的假象。

因此，现代重建应保留不确定对应，而不是强迫每个古代条目都有唯一现代编号。某些疑难条目暂时无解，比分配一个方便编号更诚实，也更有研究价值。它们指出目录身份系统的边界，提醒我们“同一颗星”本身也是需要证据的历史判断。

15.3 精度评估的历史尺度

说一份星表“更准确”，必须回答准确在哪个意义上。平均偏差较小，随机散布较小，极端错误较少，或某类恒星表现较好，都是不同判断。它们可以同时成立，也可能相互分离。一份目录的总体平均接近零，仍可能包含很大离散；一份有固定偏置的目录，校正之后也可能具有较好内部一致性。

韦尔邦特与范根特的研究正是通过明确版本、现代比较目录及误差统计来评价古代星表。[15-2] 本章不把某个单独统计量扩展为文明排名，而关注它展示的方法：历史赞誉需要落实为可以复核的比较，统计结论也必须携带样本和处理条件。

误差分布尤其不能默认完全理想。少量严重偏差可能来自错认、抄写或特殊观测困难。如果只报告一个总体宽度，容易把这些差别压平；如果为了得到漂亮结果而任意删除离群点，又会高估可靠性。合理分析应说明哪些点被排除、依据是什么，以及保留它们后结论如何改变。

同样，亮星和暗星、不同天空区域的对象，可能具有不同观测条件。一个目录在易测目标上表现优秀，并不保证在所有目标上同样优秀。公平比较应尽量匹配同类对象，而不是让一份目录包含更困难样本，另一份只取最清楚部分。历史精度需要历史条件，也需要一致的现代统计口径。

这种评价并非把古代工作置于不可能达到的现代标准下受审。现代高精度资料提供比较基准，不意味着要求古代仪器达到同样水平。恰恰因为有更可靠基准，才可能辨认古代成果究竟在哪些方面已经非常好，哪些误差来自共同零点，哪些属于个别异常。

从 S 出发看成品星表，我们容易只比较数字；从 E 出发看仪器与组织，又容易由宏大条件推定优秀结果。让两条路径相互检查，可以识别投入与产出的不一致。如果大仪器没有带来预期改善，应追问其他误差是否占主导，而不是怀疑所有测量价值；如果较简单工具取得良好结果，也应寻找操作与校正的贡献。

这是一种有反例能力的发生学解释。它不预设共同体越大结果越好，也不预设技术越新精度越高。它要求找出真正限制结果的环节，再看改变是否触及该环节。只有这样，制度与技术才从叙述背景转成能够支持或否定解释的机制。

若以一个现代说明例子比较两组测量，第一组全部偏高但散布很小，第二组平均无偏而散布很大，那么“谁更好”取决于任务。能够识别并校正共同偏置时，第一组可能更有用；无法确定偏置时，风险又不同。这个例子不代替古代目录的实际统计，却解释了为什么单一排名无法充分呈现精度。

同样，不应由今日重算推定古人已经知道相同的误差结构。现代分析揭示一项系统偏差，属于我们的认识；要说古代共同体已经发现并处理它，还需要当时记录。将两者分别说明，可以同时承认现代研究的贡献与历史主体的实际能力，避免把后人的聪明投射回前人。

15.4 共同体遗产

星表完成之后，其生命并未结束。它可以进入教学、导航相关计算、仪器设置和后来目录比较。不同使用者关心不同部分，表中的一个共同基准可能因此成为许多研究的隐含前提。知识影响并不总以新理论名字出现，也可能通过日常计算持续发挥作用。

乌鲁伯格相关表书的多种语言抄本与后续研究，说明其成果具有长久的传播生命。[15-1][15-3] 但传播数量不能直接证明每次使用都更准确。一份表可能因权威、便利或可获得性而持续流传。接受史需要与性能评价分开，随后再研究二者怎样相互影响。

以下关于版本与检查的要求，是由案例引出的现代评价尺度，不是对当时制度已经全部具备这些程序的断言。目录的统一有巨大收益：不同观察者可以指向相同对象，使用相近坐标，减少每次重新建立基准的成本。统一也带来风险：共同错误可能广泛扩散。真正可持续的和谐，不是禁止改动共同标准，而是建立能够识别错误、发布修订并保存版本关系的方式。

若新观察只在私人笔记中纠正旧值，却没有进入公共表格，局部进步便难以成为共同能力。若新版本删除旧痕迹，后人又难以理解变化原因。因此，知识遗产不仅是保存最终答案，也包括保存修订关系。现代数字目录的版本管理，与古代抄本比较虽有巨大差异，却都面对这一基本问题。

这也重新规定“伟大观测者”的意义。个人的判断力可以启动改进，但只有当改进进入可传递程序，才能跨越个人生命。与此同时，程序并不能替代所有判断：目标识别、异常处理和参数取舍，仍需要具体解释。复合能力不是把人的作用取消，而是让个人判断同可复用结构结合。

本章的中心判断因而是：星表精度属于一个有历史条件的整体能力。恒星、仪器、读数、历元和文本并非随意拼接，它们在共同坐标任务中相互规定。重新测量改变显露，显露又改变后来训练与比较的条件。这种回写可以积累可靠性，也可能封存盲区。

从印度计算到中国记录，再到伊斯兰世界的表书与观测台，我们看到的不是一条单线接力，而是不同问题组织方式。它们分别强调可计算关系、长时段保存、几何一致性与共同测量，并在交流中部分连接。后来的欧洲天文学改变，进入的正是这样一个已经富有技术、文本与争论的世界。

一份目录最深的价值，可能是让后来某个偏差变得无法忽视。没有共同基准时，每次观察都像孤立事件；有了稳定基准，反复出现的差异才可能被识别为问题。于是，精度不是只让答案更细，而是使新的问题获得存在条件。这种从结果到问题的回写，连接了本章与近代天文学的改变。

因此，星表的遗产既包括它已经回答的事，也包括它使后人能够继续追问的事。只有把这两方面一起看，才能理解持续积累如何成为创造的条件，而非创造的对立面。

从这一点出发，精度共同体的成功不是让所有人不再怀疑，而是让怀疑可以指向明确条目、明确条件和明确计算。可靠性与可修正性因而能够共同增长，这才是目录长期生命的基础。

本章参考文献

[15-1] Robert H. van Gent. The Astronomical Tables of Ulugh Beg. Utrecht University。https://webspace.science.uu.nl/~gent0113/ulughbeg/ulughbeg_zij_i_sultani.htm

[15-2] Frank Verbunt, Robert H. van Gent. The Star Catalogues of Ptolemaios and Ulugh Beg: Machine-readable Versions and Comparison with the Modern Hipparcos Catalogue. Astronomy & Astrophysics, 544, A31, 2012。核读全文的资料说明、识别方法与结果部分。https://arxiv.org/html/1206.0628v1

[15-3] CDS. Star Catalogues of Ptolemaios, Ulugh Beg，J/A+A/544/A31，机器可读资料说明。https://cdsarc.cds.unistra.fr/viz-bin/ReadMe/J/A%2BA/544/A31?format=html

第 04 编

日心体系与望远镜新证据

第 16 章 哥白尼：行星秩序与解释联系的重构

第 17 章 第谷：精度为何改变理论竞争

第 18 章 开普勒：一个残差怎样迫使轨道重写

第 19 章 伽利略：望远镜图像如何获得可信地位

第 20 章 为什么革命不是一次性胜利

第 16 章 哥白尼：行星秩序与解释联系的重构

16.1 改变中心不等于抛弃所有旧数学

把太阳放到中间，把地球移到一条轨道上，这幅图容易记忆，却不足以说明哥白尼做了什么。中心位置的变化只有进入周期、距离和观测的组织，才成为一套天文学。若只是把一张图旋转或换一个原点，同一组相对位置可以保持不变；哥白尼希望建立的，却是各部分相互联系的宇宙秩序。

《天体运行论》于一五四三年出版，它既有总体安排的论述，也有大量技术内容。[16-1] 将它概括成一句“地球绕太阳”，会使其数学劳动从历史中消失。另一方面，若只把它看作方便计算的技巧，又会低估作者对世界结构的主张。文本内部关于假说地位的不同声音，也提醒读者不要把书前所有文字都当成作者同一立场。

哥白尼保留了圆周运动的重要地位，并没有一步进入开普勒的椭圆体系。[16-1] 这使他的改变具有选择性：重新安排地球与行星的联系，同时继续使用某些古典数学规范。所谓新体系并非全部零件都新，而是一些继承资源在新的相互规定中发挥作用。

这种选择性是理解创新的关键。如果用“是否彻底摆脱古代”来评判，就会把历史人物放进一个不可能完成的考试。新问题必须用某些已有资源展开，完全无继承的创造并不存在。真正要问的是，哪些关系被保留，哪些关系被取消，哪些对象因此获得新身份。

地球从静止中心成为行星，是一种身份改变。它不再只是观察一切的固定支点，还参与解释其他行星的表观运动。太阳与行星的关系也随之重组：不同轨道不再只是各自拟合天空现象，而能够通过地球共同运动建立联系。这个联接，是比“图画更简洁”更具体的思想成就。

然而，联系更强也意味着牵连更广。改变地球运动，会影响对昼夜、季节、逆行和恒星尺度的解释；它也要求重新回答地面物体为何没有显出预期运动效果。一个总体安排获得解释收益，同时承担新的证明负担。创新不是只拿走旧体系的困难，还可能生产此前不显著的新困难。

在 SDE 框架下，新显露 S 不能只写成“日心说”三个字。它是地球、行星周期、相对距离与视运动之间的新组织。D 中的比较不再把地球当作绝对不动的参照，E 中的轨道关系也因此重新规定彼此。这里创造的是新的可计算天文学世界，而非凭思想使太阳或地球在物理上重新形成。

书前奥西安德的文字与哥白尼自己的序言，应分开阅读。前者强调假说的计算用途，后者追求宇宙各部分的联系。[16-1] 两种立场放在同一出版物中，已经说明“这本书意味着什

么”并非从印刷之日起就只有一种答案。出版把思想送入新的解释互动，也可能改变它的接受方式。

这并不允许我们任意选择喜欢的一段代表全书。更好的阅读是承认层次：技术模型能否算出位置，宇宙安排是否被作者当成真实结构，读者又怎样使用它。三者分别有证据。把技术接受当成完整信仰，把作者目标当成读者实际理解，都会制造虚假的一致。

此外，圆周资源的保留也不宜简单视为妥协。它们是当时成熟的数学语言，拥有可用推算与几何证明。要放弃这种语言，不仅需要不满意，还需要替代办法。后来的椭圆成功，不能证明哥白尼当时只要“勇敢一点”就能完成同样改变。创造需要新的差异与可工作结构，不只是更强决心。

16.2 逆行的重新表征

逆行是行星在恒星背景上暂时改变移动方向的现象。对于地面观察者，它是真实的视位置变化，不是错误观察。需要改变的是这种变化的解释：它是否必须对应行星自身运动方向反转，还是可以由观察者与天体共同运动产生？哥白尼把地球运动纳入之后，使这一问题获得新的组织。

可以用现代简化模型说明。设地球与外行星都沿同一方向运动，地球轨道较小且运行较快。当它追近并超过外行星时，两者连线在遥远恒星背景上的方向可能暂时反向变化。外行星不必在自身轨道上真的掉头，视运动仍然可以逆行。这个说明省略了真实轨道细节，只展示相对运动机制。

关键不只是某一颗行星得到解释，而是多个行星的相似现象同地球同一运动联系起来。原先需要分别组织的某些周期关系，可以由共同参照变化统摄。统一性因此表现为不同现象之间新增约束，而不是单纯减少图中的圆数。一个模型即使仍有复杂计算，也可能在解释联系上更统一。

对于内行星，最大离日角等现象也进入相对轨道关系。它们总在太阳附近出现的限制，可以同轨道位于地球轨道内部的安排联系起来。由此，行星顺序与可见行为相互规定，不再完全依赖另行安排。现代教学模型能够帮助检查这些几何联系，但不能把简化模型的漂亮结果直接当成十六世纪全部数值表现。[16-2]

周期关系也可以用现代说明式表达。在理想化的同向圆周运动下，会合频率等于两轨道频率之差的绝对值。这个关系使观察到的会合周期同轨道周期相联系。它不是一项孤立算术技巧，而是新的参照安排所带来的约束：一旦地球周期参与，多颗行星的表观变化便共享同一个量。

这种共享同时提供检查机会。若对每颗行星都需要随意改变地球相关参数，统一便只是表面。一个有内容的体系要求共同部分保持一致，再解释各行星的差异。统一性与多样性由此不是敌对关系：共同运动提供联系，各自轨道条件保留差别。

然而，几何解释并不自动决定物理解释。若将运动重新表达为另一种地心混合安排，某些相对角位置仍可能相同。观察能够排除哪些模型，取决于它们是否真正给出不同预测。不能因为一个解释自然流畅，就宣布所有替代结构已被观测淘汰。

这里六路径可以改变判断。从日心图像 S 出发，容易把逆行看成新体系已经获胜的证据；从观测差异 D 出发，比较多个模型怎样产生相同角位置，则会发现判别不足。从 E 出发再问哪些共同约束使一个体系更有解释力，才能同时保留其优点与未完成之处。

另一个有区分力的问题，是模型能否把原本独立的自由选择变成共同约束。若每个行星的某些参数可以独立调整，那么拟合能力可能很强，但解释联系较弱；若一个共同运动使多组现象必须一起配合，体系便承担更严格要求。它可能因此更容易失败，也可能在成功时获得更强支持。

这不等于参数越少理论越真。过度简化会丢掉真实差异。关键是哪些自由度有独立理由，哪些只是为保住结论临时增加。哥白尼体系值得研究的，是它重新分配了解释自由度：某些表观变化成为共同运动的结果，另一些差异仍需各自处理。统一并没有消灭多样，而是改变了二者的边界。

16.3 视差难题与恒星尺度

如果地球每年改变位置，为什么恒星没有表现出容易观察到的年度位移？这个问题在当时不是愚昧的反对，而是新安排必须面对的几何后果。观察地点移动而目标有限远，两次视线通常会有差异。没有测到差异，需要解释，而不能因理论吸引人就被忽略。

一种解释是恒星极远，使地球轨道相对而言很小。用现代小角近似，视差约与基线长度除以距离成比例。距离越大，角差越小，便可能落在观测能力之外。这是一个有效的可能性说明，却不是对距离的独立测量。它能够解除直接矛盾，同时把新的负担转移到宇宙尺度上。

哥白尼讨论天空相对于地球的巨大尺度。[16-1] 但对当时读者而言，需要接受的不只是“大”，而是一个难以由已有观察直接约束的巨大间隔。现代人熟悉恒星距离，容易觉得这一推断毫无困难；历史参与者没有这些后来材料，不能被要求预先知道答案。

更复杂的困难来自恒星表观大小。若将肉眼或早期器物所见星点的宽度当成真实角直径，再同极远距离结合，便会推出非常巨大的恒星实体。关于第谷与开普勒讨论恒星大小的

研究说明，这曾经是有内容的争论。[16-3] 后来的光学理解会改变问题，但不能被提前放入十六世纪的证据库。

这里出现一个典型的纠缠结构：视差上限约束距离下限，表观宽度与距离共同决定推定实体大小，而光学解释决定表观宽度能否如此使用。任何一环改变，整体论证都会改变。若只说反对者“不愿接受大宇宙”，便忽略了他们实际上在协调多项当时看来相关的证据。

这也说明未检测结果的逻辑。没有检测到视差，不等于证明地球静止，除非同时拥有足够强的恒星距离约束；它也不等于证明恒星极远，除非接受地球运动。未检测支持的是一组条件关系，而不是一个无条件结论。天文学的许多争论，正发生在不同条件组合之间。

SDE 的发生性分析在此应避免把选择描写成任意意志。研究者受到可用仪器、光学认识和宇宙结构假设的共同约束，能够提出的解释有范围。新的主体—仪器—天体复合体一旦改变，例如光学误差获得新解释，原来支持某个结论的关系也会改变。不是事实任意变化，而是事实怎样承担推断作用需要重新规定。

从这一角度看，恒星尺度问题不是哥白尼体系之外的枝节，而是地球身份改变后必然出现的关联问题。新的统一扩大了问题域，也提高了解释成本。理论的力量与它制造的未决问题一起增长，这正是思想革命常有的形态。

用一个条件论证可以看得更清楚。若地球运动且恒星不太远，应出现可测视差；现在没有可测视差，便至少需要放弃其中一个条件，或重新评估测量能力。不同研究者可能选择不同分支。只有获得额外证据，才能减少分支。把这种分歧写成理性与非理性的冲突，会使真正的推理结构消失。

同样，观测能力不是一条抽象的精度数字。不同星、不同季节和不同大气条件下，可靠程度可能不同。一个未检测结果能排除多大范围，需要说明这些条件。科学史中的“没有看见”并不空白，它是一种由仪器、程序和理论共同确定含义的显露。

16.4 新体系的真实成本

一套新体系能解释更多联系，为何仍不能立即取代旧体系？因为接受它需要的不只是赞同一幅宇宙图，还包括新的计算能力、物理说明和可信证据。旧体系已经拥有教学习惯、表格和问题处理方法。它们构成现实成本，不能仅用“思想顽固”解释。

计算采用与宇宙论接受也可以分离。有人可以使用新参数或新表，却不承认地球真实运动。这并非逻辑上必然矛盾，因为不同计算形式可能产生相近预测。思想史若只按“支持或反对哥白尼”分人，就会丢掉这种部分接受。一个体系的传播往往先发生在某些操作环节，再影响整体理解。

物理成本尤其突出。地球运动要求解释抛落物体、空气和日常经验为何仍呈现熟悉结果。现代惯性概念可以帮助我们理解，但哥白尼不能直接调用后来完整动力学。新天文学世界已经开始构成，相应的物理世界尚需进一步重组。两种变化的不同步，是长期争论的重要原因。

审美与一致性同样不能独自裁决。一个体系可以因整体联系而显得美，却仍需要独立检验。美感能够引导研究者继续投入，也可能使其低估困难。科学史不必把美排除在创造之外，但必须区分它作为动力与作为证据的角色。前者可以启动研究，后者需要更多约束。

三方程揭示这种相互推动：日心安排的 S 改变了应该计算和观察的 D，同时使 E 中的参照、运动规范与尺度关系重新组合；新的 D 与 E 又反过来检验最初 S。但这不是一个保证成功的闭环。若新增假设不断增加而无法获得独立支持，体系可能陷入保护自身的循环。

因此，评价哥白尼应采用双重判断。他确实重构了行星世界的解释联系，创造出使地球与其他行星共同进入运动比较的新体系；他也没有独自完成地球运动的所有证明，更没有取消所有旧数学。前一判断保证我们不把贡献缩成坐标更换，后一判断保证我们不把后来的成果全部提前归给他。

这是一种具有生成力而尚未封闭的统一。它把许多现象接到共同关系上，也让未解决的物理与观测问题变得更集中。第谷、开普勒和伽利略面对的，不是只需宣布胜利的成品，而是一个必须继续构造、检验和修正的新问题世界。

这种成本还体现在训练中。已经熟练使用旧表的人，转向新方法需要时间；教师必须把新安排写成学生能掌握的步骤；印刷与流通决定哪些版本可以获得。它们不能证明某个理论正确，却影响理论能否接受充分检验。一个思想若没有形成可工作的学习路径，便可能长期停留在少数人的讨论中。

不过，不能把任何迟缓接受都解释成传播障碍。若物理难题尚未解决，暂缓接受也可能是合理判断。反过来，某位支持者可能因统一性和未来研究潜力而先行投入，即使证据尚未完备。这两种选择各有风险。思想史应说明他们掌握的材料和承担的代价，而不以最终胜负决定当时是否理性。

本章据此提出一个检验标准：如果日心安排只是在原有观察上换名，却没有新增任何相互约束，那么“重构世界”的说法就是过强的；如果它改变了对象身份、比较程序和问题联系，同时可以指出哪些结果仍未被证明，这一说法才有具体内容。这里的本体论主张，需要由实际机制承担，而不是靠口号维持。

哥白尼的突破正适合这样的审慎肯定。他让观察者所在的地球进入解释结构，使天空的许多变化成为复合运动的显露。新世界从这里发生，但它的可靠性还要由后续观测、数学与物理共同建立。生成与确证相互联系，又有不同时间节奏。

后续章节将沿着这些未决问题展开，不把不同人物写成预先分工的胜利队伍。他们各自维护不同约束，也作出彼此不完全相容的判断。正因为如此，资料、方法和对象的重新组合，才具有真实的历史难度与创造意义。

本章参考文献

[16-1] Nicolaus Copernicus. On the Revolutions，第一卷与序言，Edward Rosen 译，Dartmouth 大学在线节录。https://math.dartmouth.edu/~matc/Readers/renaissance.astro/1.1.Revol.html

[16-2] Todd Timberlake. Modeling the History of Astronomy: Ptolemy, Copernicus and Tycho. 2013。现代模型教学研究，不作为哥白尼原始发现过程的证据。https://arxiv.org/abs/1301.2119

[16-3] Christopher M. Graney. Of Mites and Men: Johannes Kepler on Stars and Size. 2018，含开普勒《新星》相关章节译文。https://arxiv.org/abs/1802.03313

第 17 章 第谷：精度为何改变理论竞争

17.1 大象限仪与准确坐标

两个模型如果都能在粗略观测下得到相近结果，如何判断它们是否真正等效？一种可能是，它们的差别小于现有误差；另一种可能是，它们对所观察的量确实给出同一预测。提高精度能够处理前一种情况，却不能自动处理后一种。第谷的工作使这一区分变得具有历史力量。

第谷建立持续观测，重视仪器与资料组织，取得了后来研究者能够重新分析的位置数据。他不是望远镜天文学的代表，成果主要依赖肉眼瞄准的测角实践。[17-1] 因而，不能把近代观测进步写成望远镜出现之后才突然开始。仪器尺度、刻度、支承、校准和系统记录，已经能够显著改变比较能力。

精确坐标不是直接从天空落到纸上。观察者先瞄准目标，读取相应角度，再结合时间与地点转成可用位置。大气折射、参照星和仪器偏差都会参与。若忽略这些过程，数据就像没有来源的纯事实；若把它们全部归为社会建构，又会丢掉天空实际施加的约束。数据的可靠性恰好来自对这些环节的控制与检验。

对第谷星表不同版本的现代研究，区分手稿、较短印本和后来较完整的编辑本，并分析坐标误差。[17-2] 这一事实说明，“第谷数据”不是一个无差别整体。恒星目录、行星观测和不同编辑版本，不能互相替代；某类资料的总体精度，也不能直接当成每条火星记录的误差界。

从数学上说，若模型差异为一个小角量，而测量不确定性更大，观察暂时无法稳定区分。缩小不确定性后，差异可能显露为重复的系统残差。但是，误差界必须有理由，不能为了宣布模型失败而事后缩小。需要通过重复观察、仪器比较及其他检查建立可信范围，再让模型接受检验。

这改变了误差的地位。粗略条件下可以容忍的小偏差，在新条件下可能成为必须解释的问题。偏差的绝对大小没有改变，证据意义却改变了，因为它进入了不同的测量体系。SDE 所说的显露依赖 D 与 E，在这里有明确内容：同一角差在不同可靠性条件下，承担不同排除力量。

因此，第谷贡献的不只是更多坐标，而是一种更严格的研究门槛。模型不能再以“差不多”轻易过关。与此同时，他的资料也没有让一切问题自动得到答案。精度扩大了可排除范围，却仍需研究者决定比较哪些候选、哪些共同假设可以保留。

还应区分准确与精密。多次读数聚集得很紧，说明随机波动较小，却不保证接近正确基准；整体偏差小，也可能掩盖某些观测条件下的大误差。古代人物未必使用今天同样的术语，但现代史家必须把这些性质分开，否则“高精度”便会成为没有内容的赞词。

对行星研究而言，时间组织尤其重要。同一位置差若对应错误时刻，便会被误解为轨道问题。观测时刻、角位置和太阳参照需要共同配合。钟表、日影或其他计时方式的可靠性，因此进入轨道推断，而不是天文学之外的辅助细节。看似纯粹的空间数据，实际上已经包含时间条件。

这也解释为什么单次漂亮测量不等于一种新能力。只有持续工作让不同季节和不同位置都进入比较，模型的适用范围才能接受检验。研究者可能在某一位置恰好得到好结果，却在其他位置偏离。跨情境的差异序列，才使局部巧合与较稳定关系有机会分开。

17.2 一五七二年新星与一五七七年彗星

一五七二年的新星与一五七七年的彗星，提供了另一种挑战。问题不只是它们出现了，而是它们位于哪里。如果变化现象都被放在月下区域，天界不变的主张仍可维持；若测量将它们置于月球之外，这种划分就受到压力。[17-1] 位置判断因此直接关系到宇宙的性质。

这里的工具是视差。观察点变化时，近处目标通常比远处目标表现出更大的方向变化。若一个对象处于月球以内，原则上应当有相应量级的视差。第谷相关观察和比较，使把这些对象简单归为近地现象越来越困难。推断不是“看起来很远”，而是通过未出现预期位移来约束距离。

未检测视差依然是条件判断。它需要确认测量足以发现所预期的位移，并考虑观测时刻、参照与误差。一个含糊的“没看见”不能承担相同力量。将未检测变成有效证据，是仪器、程序和几何共同完成的工作，不是观察者个人确信就够了。

新星对不变天界的压力，与彗星对某些天球安排的压力，也不是完全同一件事。新亮体的出现与消退，使变化进入原被视作不变的区域；彗星的路径则使某些固体天球结构难以维持。不能把这两种挑战混成“一次观察推翻整个古代宇宙”。不同假设有不同的失败理由。

关于彗星的历史研究还提醒我们，当时并非所有人立即同意。测量结果、现象分类和自然哲学的权限关系，都影响结论怎样被接受。[17-3] 这些因素不能使任意解释同样成立，却说明证据从形成到获得公共地位，需要经过实际争论。

如果数学家能够由测量指出一个变化对象在月球之外，位置知识就开始对有关天体性质的传统划分施加约束。这是证据制度的变化：某类过去可能被视为只处理数量的工作，获得了影响宇宙论的资格。它并不取消哲学，而是改变哪些判断必须回应测量。

从 SDE 看，天体身份在这里发生重组。彗星不再只是归入近地变化的一种显象，而成为必须在更广天空结构中定位的对象。对象的物理形成当然没有由第谷决定，但它在科学世界中的属性联系改变了：距离、运动和变化性不再按照旧划分固定配对。

这类变化还会反过来改变观察。若彗星可能属于远方天空，就值得持续记录其路径，比较不同日期和不同观察者的结果。新的身份 S 改变 D 的安排，新的 D 又要求更合适的仪器与共同标准 E。由一次异常到一种研究对象的转变，正是在这种回写中完成。

应当特别避免把现代“超新星”类别直接放进第谷最初观察。他看到的是新出现的亮体，并通过位置与变化建立判断；现代天体物理解释则包含后来形成的物理知识。二者可以连接，却不属于同一历史时刻的认识。保留这种区别，才能看见对象身份如何逐步增厚。

彗星也是如此。确定它不在近地范围，并没有立即确定其轨道形状、物质构成或运动原因。一个问题的解决打开新的问题，而非一次性完成所有解释。若史家只写“证明彗星在天上”，读者容易误以为现代彗星观念已经完整出现，实际上仅有某些属性联系被重建。

这种有限突破具有特殊力量：它不必先给出完整替代理论，就能使旧框架中的一个普遍判断失效。可变天象被放到月球之外之后，原有分类至少需要重新说明。反例的价值在于迫使问题改变，而不在于它自己已经包含全部新答案。

17.3 观测与理论并不一一配对

第谷自己没有因精密观测而接受地球绕太阳运动。他提出的地日混合体系，使太阳绕地球，其他行星绕太阳。[17-1] 将这一体系称作“失败的日心说”，既不准确，也不利于理解它当时的竞争资格。它试图保留地球静止，同时吸收行星围绕太阳安排所带来的若干联系。

在一定几何条件下，日心体系与相应地日混合体系可以给出相同的地面行星视方向。用现代向量语言说明，观察方向取决于行星位置减去地球位置。若重新安排整体坐标，使地球成为固定原点，同时保持相对向量，角位置就不变。这个现代说明并不等于两套物理理论完全相同，而是揭示某类观测的判别限制。

因此，即使把行星角位置测得无限精确，也未必能单靠同一类量排除所有经过适当转换的运动学安排。需要加入其他约束，例如物理机制、恒星视差或能够区分参照安排的现象。精度与证据种类是两个维度，不能用一个代替另一个。

这也解释了为什么金星相位后来能够排除某些传统安排，却仍能同地日混合体系相容。一个新观察的力量，取决于它同每个具体模型的关系，而不是取决于发现者站在哪一边。科学争论中，支持某个模型与排除全部对手之间，存在必须填补的距离。

第谷对地球运动的保留意见，也涉及恒星距离与表观大小等问题。相关原文研究显示，这些问题并非单纯来自习惯偏好。[17-4] 当时的光学与测量理解，使某些今天看来容易消除的困难具有实际分量。公平历史判断应恢复这种分量，而不是要求参与者使用后来才形成的知识。

但恢复理由也不意味着所有选择等价。地日混合体系在保留地球静止时，需要承担自己的物理和结构问题。它能适应某些观测，不代表全部代价都已解决。比较应当对称：不能只列日心体系的困难而不问混合体系，也不能只赞美新体系的统一而忽略其未证前提。

第谷案例由此揭示一条重要规律：更强证据有时不会立即产生唯一理论，而会重排竞争格局。它可能让某些候选退出，使剩余候选更接近，也可能迫使争论从旧问题转向新问题。科学发展不是每增加一条数据就减少一个理论的简单计数。

在三方程语言中，精密资料 S 改变后续 D 的比较范围，却不一定使 E 中的所有世界安排唯一化。新的张力可能出现在尚未被测量直接触及的地方。发生学若预先假设数据必然导向正确理论，便会失去解释第谷的能力；只有允许多种候选暂时共存，才符合这段历史的结构。

模型比较还应区分经验等效与解释等效。两套模型对一类角位置相同，不意味着它们对所有现象相同；对所有当时可测量现象相近，也不意味着它们组织未来研究的方式一致。一个体系可能更自然地提出新的测量任务，另一个可能更容易同既有物理接合。这些差别可以影响研究选择，但需要明确其性质。

不能用未来研究潜力逃避当前失败。若模型已经同可靠观察矛盾，就必须修正；也不能因尚无决定性检验，就宣布所有理论工作无意义。科学发展常处于两者之间：部分约束已经成立，部分判断仍需探索。第谷体系的存在，迫使我们承认这种中间状态具有真实认识价值。

从三原理的角度，新的精度释放了一些旧张力，却同时形成更细的张力。原先被容忍的差异开始要求解释，原先被忽略的候选可能因此受到重视。动力不来自数据数量自动增加，而来自差异取得了能够改变研究选择的地位。没有可信误差界，再小的残差也可能只是噪声；有了这种界，微小差异便可能承担重大作用。

17.4 数据的跨代继承

第谷资料后来成为开普勒工作的重要条件，这个联系广为人知，却不应被写成一场无摩擦的接力。资料拥有者、使用者、编辑者和资助者的角色不同，数据进入新理论也不是简单交接一叠纸。开普勒需要理解记录条件，把观察转成适合轨道问题的约束，并对其中关系重新组织。[17-5]

一份观测首先可能只给出从地球所见方向，而研究者希望知道行星绕太阳的路径。两者之间需要几何反演。地球位置若不够清楚，行星轨道的推定也会受到影响。观察结果并不是轨道的直接照片，而是一个需要与其他关系共同求解的问题。

因此，数据遗产包含显性与隐性部分。显性的是日期、数值和目标；隐性的是仪器习惯、改正规则和对可靠性的判断。若后人不了解后者，就可能对某些数字过度信任或不当舍弃。可持续积累需要尽量把这些条件保存下来，使资料能够脱离原操作者继续被正确使用。

现代对第谷目录版本的研究，还表明编辑本身会改变知识产品。[17-2] 补充、排序和改写可以提高可用性，也可能引入新错误。故不能将后来的印本无条件当作当年观测的透明窗口。数据从产生到出版再到再利用，有自己的历史链条。

开普勒使用这些资料时提出的问题，也与第谷不完全相同。同一数据可以在不同理论项目中获得不同角色：对于一人，它帮助建立和维护某种宇宙安排；对于另一人，它成为迫使轨道构造改变的约束。资料的连续不等于解释的连续，但新的解释仍必须尊重资料实际允许的范围。

这说明科学世界的创造不只是增加新数据，也包括改变旧数据进入问题的方式。S 在新的 D 中被重新比较，原先隐藏的冲突得以显露；E 中的误差判断、几何与物理要求重新结合，产生新的候选对象。所谓继承，不是原封不动保存成品，而是让可核查的成果继续参与发生。

第谷留给后人的，因而既是位置记录，也是严格对待小差异的条件。他并未亲自走向后来全部结论，但他的工作使某些后来结论成为可能。历史贡献不必以作者本人接受最终理论为前提，也不该被未来成果吞没。

从这一点看，第谷是全书三条主线的交汇处：异常现象改变天体身份，视差判断涉及距离尺度，精密与持续观测改变证据制度。三者没有在同一时刻整齐完成，却在同一个研究实践中相互牵动。这种不完全同步的结合，正是近代天文学发生的重要形态。

数据继承还涉及责任。后人可以重新解释原观察，但不能任意删除不合新理论的条目；若认为某条记录错误，应提出可检查理由。否则，新理论看似获得全部资料支持，实际只是选择了有利部分。反过来，也不必把每一条旧记录神圣化：明显的抄写或计算错误可以修正，只是修正过程需要留下痕迹。

这种责任使复合研究跨越时间。第谷的观察者团队与开普勒的推算并非同时进行的同一个行动，却通过保留的资料建立了合作关系。后人既依赖前人，也能够指出前人局限。知识因此不是沿着人物名单传递的宝物，而是一个可以被继续操作的条件系统。

若没有这些条件，开普勒即使具有强烈的物理兴趣，也难以从粗略资料中迫使轨道改变。若只有资料而没有新的问题组织，许多差异又可能长期停留在表格中。两者相互需要，

但不能因此抹去各自独立的工作。统一性来自共同约束，多样性来自不同研究目标，和谐性则表现为资料能支持彼此纠正。

本章因此不以“第谷最终错了”作结。他的某些宇宙论判断后来被放弃，他建立的测量与资料能力却成为新理论的重要条件。这种分层的命运，比胜者与败者的二分更能解释科学如何发生。

还可以提出一个反事实检验：若保持同样模型，却把测量不确定性扩大，原本显著的差异便可能重新落入容忍范围。这样，数据是否迫使改变，不取决于研究者有没有看见某个偏差，而取决于偏差是否具有足够判别力。这个关系解释了为什么相近天象可以在不同历史条件下引起不同理论后果。

相反，若两模型在所测量上严格等效，单纯缩小误差不会打破等效。必须改变问题，寻找新的量或新的条件。这一点提醒现代研究者，更多同类数据不一定解决所有理论分歧。测量提升与问题重构需要协同，任何一方都不能被无限替代。

在第谷的实践中，异常天象与精密位置分别推动了这种协同。前者挑战对象分类，后者提高模型要求。它们可以互相支持，却没有把宇宙安排唯一固定。正是这种既强有力又有限的证据状态，使后续创造仍然必要。

如果我们从最终日心胜利出发，第谷容易显得只是一个提供原料的人；若从当时未决问题出发，他便是主动重组证据与候选体系的研究者。两种叙述的差别，不在于是否尊敬人物，而在于是否保留当时真实的选择空间。

本书选择后一种展开，同时用后来的知识检查其局限。历史理解不等于放弃判断，现代判断也不必取消历史理解。二者在共同材料上互相限制，能够产生比赞颂或贬低都更有内容的科学史。

一个共同标准的形成，也意味着判断责任重新分配。使用者可以暂时依赖经过反复检查的数据，而不必亲自重建每一次观察；但当新问题超出原来的可靠范围，封装就需要重新打开。第谷资料之所以能够继续产生力量，正在于它们既足够稳定以支持推算，又能够被新问题重新检验。稳定与开放并非互相排斥，而是高质量积累的两个条件。

这是一种可以跨代延续的研究能力，其价值不被任何一项个人判断的成败完全决定。

本章参考文献

[17-1] Albert Van Helden. Tycho Brahe. The Galileo Project, Rice University。https://galileo.library.rice.edu/sci/brahe.html

[17-2] Frank Verbunt, Robert H. van Gent. Three Editions of the Star Catalogue of Tycho Brahe. 2010。https://arxiv.org/abs/1003.3836

[17-3] Albert Van Helden. Comets. The Galileo Project, Rice University。https://galileo.library.rice.edu/sci/observations/comets.html

[17-4] Christopher M. Graney. Of Mites and Men: Johannes Kepler on Stars and Size. 2018。https://arxiv.org/abs/1802.03313

[17-5] George E. Smith. Kepler’s Astronomia Nova and the Orbit of Mars. Tufts University 课程研究资料，2014。https://dl.tufts.edu/downloads/s4655t006

第 18 章 开普勒：一个残差怎样迫使轨道重写

18.1 火星数据的约束力

八角分为什么能够改变天文学？若回答“开普勒特别认真”，便把一个涉及仪器、几何与物理的复杂问题缩成个人品德。认真当然重要，但只有当偏差超过可信误差，并且不能由其他可接受调整消除时，它才具有改变模型的力量。残差不是独自行动的原因，它必须进入一套判断关系。

《新天文学》以火星为主要对象，展开大量模型比较和推算。它不是先宣布椭圆，再附几个支持数字，而是把困难与尝试纳入论述。[18-1] 今天的教科书可以直接列出定律，思想史却必须解释这些定律为什么在当时成为有理由接受的关系。

地面观测首先给出方向，轨道研究却要求知道行星在太阳周围怎样分布。方向与距离不是同一个量。一个模型可能很好地拟合经度，却在推定距离方面出现问题。开普勒研究中的关键复杂性，正是不同量互相检查，而不是拿一串点同圆和椭圆简单比赛谁贴得更近。

被称作替代假说的构造，在某些经度拟合上表现很好，但将其与其他几何约束结合后，问题显露出来。现代细致重构强调，经度成功不能无条件保证轨道结构正确；当结构调整以满足距离关系时，又可能产生不能忽略的经度差异。[18-2] 因而，“一个圆差八分，于是换成椭圆”的故事遗漏了最重要的相互检验。

这个例子可以用一般逆问题说明。观测者获得某个方向，许多不同距离的点都位于同一视线上。若只拟合方向，模型可能保留较大自由度；加入另一时间的方向或其他几何关系，候选范围才缩小。不同资料不是重复投票，而可能分别约束不同自由度。

火星的困难还同地球有关。观察者所在位置若被错误描述，推定的火星轨道也会偏移。因此，研究火星不能把地球轨道当成永远无误的外部条件。观察者的位置进入对象解释，正是主客互动复合体的一个具体机制。它不是抽象宣称主体影响客体，而是一个可写出几何关系、可检查误差传递的问题。

第谷资料提供较严格的条件，但开普勒仍须判断哪些结果可信，哪些偏差可能来自运算或其他环节。误差界不是一个神圣数字，而是由既有测量能力和相互检查支持的工作判断。正因为它有根据，不能轻易把令人不舒服的残差当作无关小事。

SDE 分析在此得到一个明确结构：诊断阶段的中间显露 S 是模型预测与观察之间的残差，差异运行 D 的含义依赖误差界、参照选择和轨道假设之间的 E。这里的研究单位是一次模型诊断；后文讨论新轨道结构 S 时，研究单位转为轨道体系的构成，不能将两个层次的 S

直接等同。改变任何一项，残差的解释可能改变。因此，残差的存在与残差迫使哪个假设改变，是两个不同命题。

“八角分”还需要放回论证的位置。它并非每一轮初始拟合都同样出现的标签，而与不同约束之间的不相容相关。开普勒在《新天文学》第十九章强调不能忽略这一差异。[18-4]若将它改写成某个圆对所有观测都统一差八分，就会使原来复杂的诊断变成一个错误的简图。

更一般地说，残差具有形状。它可能在某些位置偏正，在另一些位置偏负；可能随时间积累，也可能随几何配置重复。研究者若只保留最大值，就可能丢掉能够辨认原因的信息。一个误差序列比单个误差数字更有诊断力，因为它显示错误与条件之间的关联。

这里也能理解为何同样数据未必产生同样发现。研究者可能只检查通常用于拟合的观测，因而看到模型成功；也可能寻找不同几何条件，让原先未约束的自由度显露。创造不只来自更多资料，也来自改变资料之间的比较方式。这种变化需要明确问题，而不是任意组合。

当然，寻找有区分力的资料不能变成挑选有利资料。如果研究者只保留使圆失败而使椭圆成功的记录，就无法建立可靠结论。必须说明筛选的几何理由，并用其他资料交叉检查。由此，选择的自由同检验的约束相互配合，才可能产生可公共确认的新关系。

还应注意，发现叙事本身经过作者组织。《新天文学》呈现的论证顺序，不必等于每一次实际尝试发生的先后。我们可以依据文本分析它向读者提供了怎样的理由，却不能没有其他材料就把它当作逐日实验日志。这个区分避免把一部精心论证的作品误读成透明的心理记录。

它也使本书的方法保持自我约束。SDE 可以重建条件之间的相互规定，但不能声称因此已经知道作者每一刻如何思考。可验证的是文本、计算与资料所呈现的关系；更细的心理过程只能在有证据时讨论。发生学的严密性，正取决于愿意在材料边界处停下。

18.2 从圆到椭圆的试探

圆并非被一个漂亮椭圆图瞬间击败。轨道形状、运动速度和参照中心共同参与预测，改变其中一项可能改善某些结果，却恶化其他结果。开普勒必须在这些相互牵制中工作。真正困难的不是想象一种非圆曲线，而是使它同连续运动与多类资料同时相容。

首先要区分轨迹和运行规律。轨迹说明天体经过哪些位置，运行规律说明它何时到达这些位置。同一条圆轨道，配不同速度规则，可以产生不同时间序列；同一组少量位置，也可能被不同曲线和速度组合经过。只比较曲线外观，无法判断完整模型。

替代假说的价值正具有双重性。它作为完整结构存在困难，却能在有限任务上提供有用近似。一个失败模型并非只能被丢弃，它的局部成功可以成为后续研究的资源。开普勒的工作体现了这种分解能力：不把所有旧结果一并取消，也不因局部成功而保护全部假设。

第二类尝试涉及非圆的卵形候选。承认圆不能满足要求，并不会唯一指定椭圆。许多形状都能偏离圆，关键在于它们怎样同距离与运动规则配合。现代重构展示了从候选卵形到椭圆之间的进一步比较。[18-2] 这里的失败不是徒劳插曲，而是缩小可能空间的实际工作。

用现代几何说明，椭圆具有中心、长短轴与焦点等不同位置。太阳在一个焦点，与太阳在椭圆中心并不是同一模型。即使画出相同椭圆，如果运行规律或太阳位置不同，预测仍会改变。教科书的一句话“轨道是椭圆”，只有连同这些关系才具有天文学内容。

可以进一步作一个思想比较。若只观察轨道的一小段，圆弧与某些椭圆弧可能非常接近；若只检查几个特殊位置，某种错误速度规则也可能暂时过关。扩大时间范围并更换观察几何，才可能暴露差异。因此，模型检验需要有意识地寻找候选分歧最大的条件，而不是只在最容易成功的地方重复确认。

这正是六路径互查的意义。从椭圆成品 S 回看历史，容易把所有尝试写成接近答案的台阶；从失败差异 D 出发，检查当时可用约束 E，则会看到不止一条可能道路。两种入口结合，既能够解释最后关系的力量，也保留发现过程的开放性。

必须避免把开普勒描述成完全摆脱一切审美与形而上信念的纯数据机器。他对宇宙秩序和物理原因有强烈追求，但这些追求没有使全部候选免于失败。创造力在这里表现为让偏爱的构造接受具体约束，并在不能兼容时改变它，而不是事先没有任何偏好。

S=F(D,E)在本章不意味着把数据机械投入公式便产出椭圆。F 所代表的，是一系列受到约束的构造、比较与判断。若不给出候选、操作和排除理由，这个符号没有解释力。只有说明为什么某些自由度被保留、某些被取消，发生学才触及真实研究。

失败构造还具有一种方法价值：它让研究者知道哪种成功只是表面。某模型可以通过调整两个参数补偿彼此错误，在少量观察上表现优秀；当加入新的条件，补偿关系就可能崩解。这样，失败不是简单宣布“模型不对”，而是暴露不同部分如何错误地相互支持。

在 SDE 语言中，这是一种病态的稳定化：某个 S 在有限 D 中稳定，却依赖一个不能扩展的 E。扩大比较范围后，原稳定性失去支持。这个判断不需要给历史人物贴标签，它只说明为什么局部拟合不能承担总体真实性。可扩展的稳定，必须经受不同任务的检验。

圆到椭圆的改变也不宜被理解成多样性无条件胜过统一性。任意曲线越多，拟合选择越多，解释也可能越空。开普勒需要的是新的受约束统一：既容纳观测差异，又不允许任意修改。创新由此不是不断增加选择，而是在探索之后形成更有内容的限制。

这一点与“自由就是没有约束”的理解不同。研究自由表现为能够改变曾被认为不可动的前提，随后还要承担新的约束。若只放弃圆，却不给出时间关系和太阳位置，就没有完成新天文学。自由律的分析必须落在候选生成、选择理由与可操作结果上，不能停留于赞美突破。

18.3 运动规则与物理原因

椭圆解决轨迹问题，面积规律则把时间重新带入几何。在现代表述中，行星与太阳连线在相等时间扫过相等面积。这意味着行星在靠近太阳时通常运行较快，在远离太阳时较慢，但不能把这种定性描述当成完整定律。面积与时间的比例关系，才提供可以计算的约束。

用现代记号写，面积变化率保持恒定。这个式子是现代说明，不是开普勒原始表达。它把位置、角变化和时间联系在一起，使速度不再只是沿某个圆的固定角速度。由此形成的运动对象，其性质属于轨迹与时间的共同结构，而不是一条静止曲线独立具有。

还要区分描述规律与物理原因。面积规律告诉我们运动怎样组织，并不单凭自身说明为什么如此。开普勒尝试从太阳作用等方向寻求原因，但不能把他的解释直接写成牛顿万有引力。[18-3] 后来的理论会重新理解这些规律，早期探索却有自身语言与局限。

物理兴趣仍然能够产生有效研究选择。把真实太阳而不是某个方便的数学点作为重要参照，会改变模型组织。这个选择最初可以部分由物理理由推动，随后必须接受观测与几何检验。[18-2] 理论直觉与经验约束因此不是只能二选一：前者提出值得尝试的结构，后者决定它能否继续承担解释。

这里存在一种容易被忽视的创造机制。新物理想象未必完全正确，却可能促使研究者提出更有区分力的问题。若它引导的比较取得稳定结果，某些关系能够被保留，原先物理说明则可能后来被替换。知识不是只有整套全对才有价值，但必须明确哪些部分通过了检验。

反过来，经验规律也可能先于完整原因获得可靠地位。不能因为当时原因解释不足，就否认它的描述成绩；也不能因为描述准确，就宣布原因已经确定。同一运动规律可以在一段时间内同不同机制设想相联。思想史需要保留这种暂时开放的状态。

对 SDE 而言，E 在这里表现为轨道几何、时间规则和太阳参照之间的相互规定。若改变太阳位置，面积规则的表达与预测也会改变；若改变时间分配，同一椭圆不再给出相同观测序列。这不是把几何和物理放在同一段文字中的简单并列，而是它们共同构成一个可检验运动系统。

这一系统带来的新世界，不只是人们脑中多了一条正确曲线。新的行星成为具有焦点关系、连续速度变化与可比较周期的对象，研究者可以提出此前不同的问题。对象特征、计算方法和解释目标一起改变，正是科学世界构成的实质内容。

但这个判断仍有边界。开普勒没有因建立新的运动对象就制造行星本身，也没有消除所有扰动和长期变化。新的体系具有适用范围，后来还要进入更广动力学。承认这一范围，能够使本体论解释同科学工作的实际层次保持一致。

轨迹与时间的区分，还能说明为什么不能把定律当作单纯图像。读者看见椭圆，可能以为已经理解运动；实际上，要预测某日位置，还需要从经过时间求出相应位置。这一计算并不等于沿椭圆均匀走过相同弧长。面积关系改变了实际求解任务，也改变了“知道轨道”的含义。

从现代角度，可以把这种任务看成几何与时间参数之间的转换。转换要求明确起始条件、轨道尺度和运动周期。只有这些量共同确定，椭圆才成为一个可运行模型。可见，科学对象的形成不是先有一个完全独立的几何体，再给它加上时间，而是几何与时间在计算中相互规定。

同时，不能把后来普遍采用的教科书顺序当成发现顺序。今天先讲椭圆再讲面积，便于系统教学；历史研究却可能在运动规则、距离与形状之间来回推进。知识最后的组织方式，与它取得的路径不同。承认这种不同，是保护创造过程多样性的必要条件。

现代证明也不能替代历史理由。我们可以由后来的动力学推导面积规律，却不能因此说开普勒当时已经具有那套证明。后来的解释增加了关系的统一性，早期论证则展示它如何首先获得可信地位。两种工作都重要，但它们回答不同问题。

18.4 新轨道的回写

一项新规律只有进入进一步计算与检验，才能从令人信服的论证变成持续研究能力。椭圆和面积规律改变了星历制作，也改变了后来研究者对原因的要求。旧问题是如何组合圆周以表征位置，新问题逐渐包括为什么行星应服从这种轨道与时间关系。

这种回写并非自动发生。新方法可能计算复杂，使用者需要学习，表格需要重新制作。理论上优越的关系若尚未被转成便利程序，旧方法仍可能在实际任务中继续使用。接受节奏因此不应只按出版日期排列，而应考察何种能力何时进入实践。

开普勒后来参与的星历工作，使新的轨道资源得到更广使用。[18-3] 本章不把这种后续成功倒灌为一六〇九年所有论证已经无懈可击，而是强调：生成、检验与封装需要时间。每个阶段既继承前一阶段成果，也可能发现前一阶段没有处理的困难。

新规律还重写了误差解释。此前某些残差可以归于模型基本形状，新的主要关系成立后，剩余偏差可能促使研究者寻找更细原因。一个层次的成功没有结束差异，而是改变差异出现的背景。理论进展因而往往不是使所有误差消失，而是使误差更有结构、更值得追踪。

这里的 D=G(S,E)非常具体：椭圆关系 S 成立后，研究者不再任意比较所有曲线，而是在新的约束 E 中检查参数、时间分配和剩余异常。E=H(S,D)也有具体含义：新的计算成绩与反复应用，改变了什么模型被认为值得信任，什么偏差必须认真对待。三个方向不是重复一句“互相影响”，而是不同任务的重组。

如果只把这一过程写成“事实战胜偏见”，还会遗漏一个重要条件：事实需要被组织成能够区分候选的证据。方向、距离、时间和误差的相互检查，才使八角分承担力量。孤立数字不会告诉研究者究竟该改圆、改速度、改地球轨道还是改测量。

因此，开普勒的创造性既在提出新关系，也在改变证据怎样被联合使用。他没有满足于一种量的良好拟合，而要求不同量之间相容。统一性在这里具有风险：连接越紧，一个地方的失败越可能迫使其他地方调整。也正因如此，体系成功时不再只是多个孤立拟合的集合。

对本章而言，还有三个可能否定过强解释的情形。若某种偏差仍可由可信的测量修正解释，就不能立即归因于轨道形状；若新曲线只在用于构造它的资料上成功，就不能宣称已经通过独立检验；若后来的计算程序才使某项成绩实现，就不能把全部成绩归到最初发表时。这些限制使发生学叙述能够接受检查。

从反面看，真正有力的证据是多重约束在同一结构中相容：方向支持的关系没有破坏距离，时间分配没有破坏位置，误差处理没有依赖任意删改。统一性不是少说几句话，而是让更多判断彼此承担责任。多样性则保留在不同资料、不同候选与不同检验方式中。和谐性表现为这些差异能够共同支持一个有限而稳定的运动结构。

这个结构仍然可以继续改变。后来研究者会询问其他行星、卫星及扰动是否服从相同关系，何种近似可以使用，何时需要更深解释。每个新问题都可能使原规律的范围更清楚，而不必立即将其全部取消。知识的成长因此包括巩固、限制和重构，不只是增加正确句子。

对读者而言，最值得从开普勒工作中学到的，不是迷恋某个微小差异，而是识别差异何时真正有意义。一个数字只有同测量能力、候选结构和独立约束结合，才可能指向需要改变的地方。没有这种联系，执着小数位也可能只是精致的误判。

反之，一项看似不起眼的偏差若反复出现，并且无法由已知误差解释，就值得让它改变研究安排。让失败结果回写问题，而不是只让既有目标筛选结果，是本章所揭示的创造机制。新的轨道世界由此产生，不是因为旧知识被全盘遗忘，而是因为旧知识被迫进入更严格的相互检验。

由此，开普勒案例既支持关系重构的解释，也限制任何万能解释。三方程能够帮助组织问题，六路径能够暴露入口偏差，三原理能够追踪张力与改道；它们是否有效，最终仍要由具体分析是否比原先更清楚、更可检验来判断。

这种判断也适用于后续章节：一个方法的名称不能代替它实际完成的工作。只有当它帮助区分了原先混淆的层次，找到了可以反驳的依赖关系，并使材料获得新的可检查联系，方法才真正进入了思想史。

这一章给出的发生学结论是有限而可检验的：残差只有在可信误差、候选结构与交叉约束中，才取得改道力量；新结构的形成又重新安排后续比较。若删去这些中间环节，只剩“八分带来革命”，便把历史创造变成了传奇。保留它们，才能看清一个新的行星运动世界怎样具体发生。

本章参考文献

[18-1] Johannes Kepler. Astronomia nova，1609。Smithsonian Libraries 原版数字化馆藏。https://library.si.edu/digital-library/book/astronomianovaa00kepl

[18-2] George E. Smith. Kepler’s Astronomia Nova and the Orbit of Mars. Tufts University，2014。核读模型、经度与距离交叉检验相关部分；其重构不视为发现过程逐日复原。https://dl.tufts.edu/downloads/s4655t006

[18-3] Albert Van Helden. Johannes Kepler. The Galileo Project, Rice University。https://galileo.library.rice.edu/sci/kepler.html

[18-4] Johannes Kepler.《新天文学》第 19 章八角分相关段落，MacTutor 所收译文与出处。正文不作长引。https://mathshistory.st-andrews.ac.uk/Biographies/Kepler/quotations/

第 19 章 伽利略：望远镜图像如何获得可信地位

19.1 望远镜是否可信

第一次把望远镜指向天空，看到的不只是放大的旧世界。新的亮点、明暗边界和形状进入视野，但它们首先也是问题：哪些属于天体，哪些来自透镜、视线或操作？如果未经区分便全部相信，仪器可能制造幻象；如果坚持只有肉眼才能作证，新能力又无法进入知识。望远镜天文学在这道门槛上开始。

伽利略不是望远镜的最初发明者。他在《星际信使》中叙述自己听闻相关器物后进行制作与改进，并说明使用条件。[19-1] 这一点很重要：天文学突破不等于凭空创造全部技术，而是把已有器物发展为能够承担新任务的观察系统。改进、校验和问题选择共同决定其研究价值。

地面测试可以帮助确认放大与辨认能力。远处可知物体提供某种对照，让使用者判断仪器是否只是任意扭曲。但地面有效不能自动证明所有天体图像都可靠。亮度、角尺寸、背景和观测条件不同，可能使某些缺陷在天空中更显著。因此，验证需要针对具体任务继续进行。

放大率也不是唯一指标。将模糊影像放得更大，不会自动增加可辨细节；过强光晕可能掩盖近旁暗点；视场狭小会增加寻找和追踪难度。使用者需要在清晰度、亮度、视场和稳定性之间作实际选择。人—器物复合体的能力，不等于某个产品参数独立决定的能力。

月面明暗边界提供了一个重要检验场。伽利略观察到边界并非处处平滑，并将孤立亮点与逐渐连成的明亮区域联系起来，解释为不同高度表面在太阳照明下显露的结果。[19-1] 这里的推断不只来自一张斑驳图像，还来自明暗随时间变化的有规律序列。

若某个亮点只在镜片某处出现，不随月面移动而保持相应关系，就值得怀疑为器物问题；若变化同照明几何一致，并能在其他时候或其他仪器上重现，天体解释就更有力量。不同解释提供不同预期，观察者可以据此安排比较。这比笼统问“相信眼睛还是相信理论”更具体。

月面类比地面，也有边界。阴影和照明的共同规律支持表面起伏解释，却不意味着月球上的每种地貌都与地球相同。早期某些进一步猜测后来没有保留，不能因此取消核心观察；同样，核心观察成立也不能为所有附加猜测担保。仪器时代的知识仍需分层判断。

本章的 SDE 单位，是观察者、望远镜与天体共同构成的观察过程。明暗图像是中间显露 S，连续观察与条件变化构成 D，光学响应、照明关系和识别规则形成 E。后来将某种稳定特

征认定为月面地形，是另一个层次的知识显露。两层相接，但不能把第一眼图像直接当成完成的天体身份。

一个特别有说明力的变化，是月面阴影的伸缩。若暗斑只是固定颜色，随照明变化应有怎样表现；若它是地形阴影，又应有怎样表现？两种解释并非完全没有共同现象，但对连续变化给出的关系不同。把这些预期带入观察，才能让图像超出“看起来像山”的类比。

图示也参与这种比较。它可以保留明暗边界和亮点位置，使不同时间不必只靠记忆连接；但绘图会选择和突出某些特征，也可能省略细节。因此，读者需要把图与文字、日期及观察条件一起读。图像并不因经过人的手就失去证据资格，关键在于其选择是否仍允许别人检查。

19.2 木星卫星的连续记录

木星附近几个小亮点，为什么最终取得卫星身份？在一夜的图像中，它们可能只是背景恒星，也可能受到光学缺陷干扰。决定性工作是连续记录，让不同解释面对时间变化。《星际信使》保留了按日期排列的观察与图示，使读者能够看到身份判断所依赖的序列。[19-1]

伽利略最初并非已经知道自己看到四颗卫星。他持续观察亮点相对于木星的排列，逐步发现它们没有像普通固定背景星那样被木星的移动远远甩开，而是在木星附近改变位置。数量的辨认与运动关系的辨认一起推进。一个完成后的分类，不能被投回首次观察。

把一六一〇年一月最初几次记录连起来，判断的转折便十分具体。一月七日，伽利略记下木星东侧两个亮点、西侧一个亮点，起初把它们当作恒星。八日，三个亮点都在木星西侧。这样的改变首先使他怀疑木星的运动是否与预期不同，而不是立即给亮点安上卫星身份。九日云层阻断观察，序列本身留有空缺。[19-1]

十日，他只看到木星东侧两个亮点，并提出另一个可能被木星遮住。随后记录使他将位置变化归于这些亮点；到十三日，他写下第一次同时看见四个的排列。这里至少有三个不应合并的判断：观察到亮点，判定亮点自身运动，确认可辨成员不止三个。它们相继加强，却不是第一眼同时完成的知识。

《星际信使》将这些观察组织成回顾性叙述。因此，书中的清楚转折不宜直接当作未经整理的内心实录。但日期、所见数量和位置关系仍让读者有机会检查结论如何获得支持。作者还说明同一夜内也可能出现明显位置变化，因而记录时刻有意义；只按日期汇总，会抹去对象运动所提供的部分约束。

这个过程包含两个层次的同一性。首先，需要判断不同夜晚的亮点属于相同一组对象；其次，需要解释它们为什么伴随木星而又相对移动。若只保留“每夜都有几个点”，不足以

建立卫星；若只看其中一个点的位置，也可能因遮掩或重合造成误解。整体序列提供了单个瞬间没有的约束。

视线投影又增加了困难。绕木星运动的天体，在地面看来可能位于东侧或西侧，也可能接近木星而难以分开。没有看见某个点，不必表示它消失；但也不能每次缺失都随意解释成遮掩。一个好的运动解释，需要对何时可见、相对位置怎样变化给出可以继续检查的预期。

这使“卫星”成为有内容的对象身份。它不仅是一个新名字，而是对位置变化、共同移动和相互关系的压缩表达。命名之后，研究者可以提出周期、距离和预测问题。新对象 S 由 D 与 E 形成，又反过来组织新的 D：哪些夜晚值得看，哪些排列值得记录，怎样区分四个成员。

木星卫星也改变了宇宙结构讨论。它表明并非所有可见天体都围绕地球直接运行，且一个运动中心可以伴随自己的绕行天体。[19-2] 这削弱了某些反对地球携带月亮运动的论证，但没有单独证明地球绕太阳。一个局部结构的存在，可以解除某种不可能性主张，却不等于已经证实所有类比。

这里要分清发现与推广。确认木星有伴随天体，建立它们的精确运动模型，以及将其用于时间或位置测定，是不同阶段。新身份打开应用可能，并不保证应用立即可靠。若把后来成熟用途提前写进最初发现，就会抹掉长期计算和校验的劳动。

《星际信使》的出版也改变了证据形式。私人观察转成别人能够阅读、比对和尝试重复的记录。图示不是天体自己印在纸上的照片，而是观察者选择后形成的表达；但只要保留日期、方向与相对关系，它就能够成为公共检查的入口。表达的选择性与证据价值可以同时存在。

从 SDE 看，望远镜没有简单替眼睛增加倍数，而是形成一种新的视觉 SDE 聚合体。透镜、身体操作、图示与时间序列共同构成了原来不能稳定取得的特征关系。这个新世界具有现实约束，不能任意编造；也具有发生条件，不能假装始终以相同方式完整摆在任何观察者面前。

可以比较三种候选解释。若小点是普通背景星，其相对木星的变化应主要跟随木星经过背景的运动；若是器物固定伪影，其行为应与器物操作产生某种稳定关联；若是绕木星的天体，则应在伴随木星整体移动时保持有界的相对运动。连续记录的作用，是逐步使某些候选难以维持。

同一组卫星的个体编号也带来新的任务。近旁亮点可能交错、遮掩或暂时无法分辨，要持续追踪某一成员，就需要运动预期与观察相互校验。这种对象同一性不再只靠固定空间位置维持，而靠变化中的关系维持。它为后来更复杂天体系统的识别提供了一种有力量的操作形式。

19.3 金星相位的解释范围

金星相位是另一类证据。观察者看到的不是亮点数量，而是受光部分形状与表观大小的变化。它涉及太阳照明、金星位置和地面视线之间的几何关系。新图像要成为宇宙模型的检验，必须说明每种模型实际允许哪些相位，而不是仅凭月牙形状宣布某一学说获胜。

在传统的某些托勒密式安排中，金星总被置于地球与太阳之间的相关区域，因而不能产生所观察到的完整相位范围。若金星绕太阳运动，便能够在几何上解释较满与较亏状态的转换。可是，第谷式地日混合体系同样允许金星围绕太阳，因此也能容纳这一证据。模型排除必须逐个进行。

相位与表观大小的联系，比单看一种形状更有约束。一个模型若把较满状态放在较远位置，把较亏状态放在较近位置，就应同时解释大小变化的方向。多个特征相互配合，能够削弱某些任意图像解释。这里的 E 是照明、距离与投影之间的相互规定，而非把太阳、金星与地球列成三个背景要素。

帕尔米耶里的研究重审伽利略发现与解释金星相位的过程，并讨论同时代不同反应。[19-3] 它提醒我们，观察到变化、理解几何含义和公开表达结论并非必然同一时刻完成。通信与发表材料需要细读，不应以一个英雄发现日期替代整个证据形成过程。

从现代角度，几何关系很容易画出，历史困难却还包括仪器是否足以显示相位，图像是否受到缺陷影响，以及观察者对天体受光的理解。一个现代演示可以说明模型后果，却不能据此认定当时所有参与者都立即面对同样清楚的证据。可见性本身有技术与训练条件。

这不意味着证据是任意的。只要在不同仪器和时间条件下反复出现相容变化，某些模型就必须改变。历史语境解释为什么判定需要时间，不应成为任何解释永不失败的借口。事实的公共力量，正来自不同观察能够在可检查条件下形成稳定联系。

六路径在这里能够防止一种常见误判。从日心体系 S 出发，人们倾向于将金星相位看成其直接证明；从观测 D 出发，列出所有相容安排，则能发现混合体系仍在；从仪器与照明 E 出发，又能检查图像本身的可信条件。三种入口互相补充，使支持程度不被夸大。

因此，金星相位的结论既强又有限。它能够迫使某些传统几何安排退出，增强围绕太阳运动的解释，却不能单凭自身唯一决定地球是否运动。承认这一限制不会减小伽利略的贡献，反而使贡献落在真正有力量的地方：新观察改变了候选体系必须满足的条件。

19.4 证据公共性的产生

望远镜图像从私人所见到公共证据，需要可重复操作。别人必须能获得足够合适的器物，知道观察对象和时机，并能用相近的规则辨认结果。单纯邀请人看一眼，不保证他们具备相同观察条件；单纯诉诸发现者声望，也不能替代复查。

早期器物差异使这一过程尤其困难。有人看不到预期细节，可能因为解释错误，也可能因为仪器、天空或操作不同。真正的复查要尽量分离这些原因。若任何失败都被解释为观察者无能，主张就失去可检验性；若任何失败都被当成发现不存在，又会忽略技术条件。

公共性因此不是多数人投票，而是可进入的检验关系。观察者可以提出条件，说明何种器物和程序应当显示什么；反对者可以检查这些条件是否足够、是否过度依赖临时补丁。共同标准在争论中形成，同时需要保留异议指出条件之外问题的能力。

出版、通信与器物制作在这里相互回写。记录提出复查目标，复查暴露器物需要改进之处，改进又使更多记录可能。人的视觉能力不再只封装在个人眼睛中，而延伸为能够传授和复制的操作系统。所谓复合主体，并非比喻，而是这种能力分布的实际结构。

但公共确认也有层次。月面起伏得到较广接受，不意味着所有宇宙论争议结束；木星卫星被确认，不意味着其周期已精确掌握；金星相位稳定显露，不意味着日心运动唯一得到证明。不同成果应分别记录其支持范围，才能避免以一项成功为全部主张作担保。

伽利略的力量在于使新的视觉关系进入公开论证，而不只是让大家“看到更多”。新的图像要求重新组织天体属性：月球可以有起伏，行星可以有自己的伴随系统，金星的相位可以约束轨道安排。这些关系构成新的天文学世界，并进一步改变什么问题值得问、什么证据必须回应。

这个世界的形成仍受实际天体与光学过程限制。它不是语言任意赋义，也不是仪器自动吐出真理。观察者、器物与对象在反复互动中形成稳定特征，理论与公共检验再使这些特征获得明确身份。统一性来自可复查联系，多样性来自不同观察与解释，和谐性则在彼此纠正中逐步形成。

本章参考文献

[19-1] Galileo Galilei. The Sidereal Messenger. Edward Stafford Carlos 译，1880，包含 1610 年《星际信使》及相关材料。核读器物说明、月面与木星记录。https://www.gutenberg.org/files/46036/46036-h/46036-h.htm

[19-2] Albert Van Helden. Satellites of Jupiter. The Galileo Project, Rice University。https://galileo.library.rice.edu/sci/observations/jupiter_satellites.html

[19-3] Paolo Palmieri. Galileo and the Discovery of the Phases of Venus. Journal for the History of Astronomy, 32,
2001。https://sites.pitt.edu/~pap7/galileo%20and%20the%20discovery%20of%20the%20phases%20of%20Venus.pdf

第 20 章 为什么革命不是一次性胜利

20.1 谁与谁在竞争

把哥白尼、第谷、开普勒和伽利略依次排列，很容易形成一条胜利接力线：一人提出，一人提供数据，一人修正数学，一人用望远镜证明。前面几章已经显示，这种排列省去了实质分歧。第谷不接受地球绕太阳，伽利略的望远镜证据并未唯一排除混合体系，开普勒的椭圆也不是哥白尼原有圆周体系的简单完成。革命发生在这些不完全一致的工作之间。

首先要区分宇宙安排、运动学预测与物理原因。宇宙安排讨论哪些天体围绕哪些中心运动；运动学预测讨论某时刻应出现在哪里；物理原因讨论运动为什么如此。一个体系可以在第一方面大胆改变，在第二方面继承许多旧方法，在第三方面仍保留困难。若只用“新”与“旧”分成两组，就会掩盖这种组合。

哥白尼与开普勒都让地球进入绕日行星序列，但对轨道和运行规律的处理并不相同。第谷保持地球静止，却能吸收一些围绕太阳组织行星的关系。托勒密式计算的某些资源可以被不同宇宙安排使用。理论边界因此不能仅凭一张中心位置图决定。[20-1]

其次，参与者可能在不同任务上采用不同资源。计算某个日期的天象、讲授宇宙结构、讨论自然哲学与回应宗教解释，并不是同一个场景。一个人在实用计算中采用新表，不必意味着已经接受其全部宇宙论；在宇宙论上支持新安排，也不保证每次都使用最新算法。接受不是一个只能取零或一的变量。

《天体运行论》卷首并置的两种声音，使这一差别直接进入读者眼前。奥西安德的前言把计算与观察相符视为足够条件，不要求假设已经表达真实原因；哥白尼自己的序言则把行星次序及整体关系当作新安排的重要理由。[20-5] 面对同一部书，读者得到的并非只有一种使用说明。

这种并置不证明每位读者都按前言划出的边界行动，却说明出版物中的技术资源与解释承诺可以发生分离。一本书流通之后，作者难以完全规定它将被怎样阅读。有人能够使用其中的计算关系，同时保留对地球运动的疑问；另一些人则可能先被整体安排吸引，再投入改进计算。思想传播因此不只是一个完整学说获得多少赞成者，也包括学说的不同部分怎样进入不同实践。

这并非宣布研究者可以随意矛盾，而是要求我们检查其选择的理由。某种计算形式如果与多种物理解释相容，采用它并不必然承诺唯一世界观。只有当研究者同时提出彼此不相容的实质主张，才构成需要解释的矛盾。史家不能把自己划定的阵营边界强加给所有实践。

SDE 分析因此需要把单位保持清楚。一个表格的 S、一组观测判断的 S 和一种宇宙结构的 S，属于不同层次。它们能够相互回写，却不必同步改变。所谓革命，可能是多个层次经过长期相互推动后，逐渐形成新的稳定组合，而不是一个命题宣布之后所有环节立即更新。

20.2 证据不足是否允许多模型

多模型共存不必表示科学无力，也不表示任何解释同样好。它可能意味着现有证据已经排除了某些候选，却仍不足以区分剩余候选。第谷式体系与日心体系对某些角位置或金星相位的相容，正是这种情形。争论需要寻找新的判别，而不是反复宣布已有观察只属于自己。

一种常见错误，是把“这个模型能解释现象”推成“只有这个模型能解释现象”。前一命题只要求展示相容关系，后一命题需要检查竞争模型。望远镜的新观察具有巨大力量，但其力量必须按实际排除范围计算。木星卫星削弱只有一个运动中心的要求，金星相位排除某些安排，二者都不单独完成地球运动的唯一证明。[20-2]

另一种错误，是把未检测结果当作绝对否定。没有测到年度视差，可以同地球静止相容，也可以同恒星极远相容。要进一步选择，需要独立距离约束、测量能力或其他物理依据。观测没有给出唯一答案时，合理态度不是随便选一个，而是说明不同选择分别依赖什么条件。

恒星表观大小问题展示了这种条件纠缠。若将所见星点宽度理解为实体角直径，极远距离会引出巨大实体；若光学解释改变，推断链也随之改变。现代人知道某些早期假设不成立，但当时争论需要实际解除这些联系，而不能靠事后答案宣判。[20-3]

比较理论还需要区分已实现成绩与研究潜力。一个体系可能目前预测相近，却具有更统一的结构，值得进一步研究；另一个可能更符合当时物理理解，因而暂时更稳妥。两种判断可以同时有理由。潜力不是正确性的证明，稳妥也不是永久拒绝改变的许可。

因此，多模型阶段具有建设性任务：找出相同预测来自何种共同结构，找出差异出现于什么条件，设计能触及这些条件的观察或论证。它不是等待一个天才突然选对，而是逐步改变证据能够约束的范围。统一性在这时首先表现为共同接受某些约束，多样性则保留在尚未决定的部分。

20.3 共同体怎样改变判据

新观察之所以能够改变宇宙论，还因为共同体逐渐承认新的证据方式。精密测角、望远镜图像和复杂轨道计算，并不是一开始就拥有不需解释的权威。它们必须显示可重复性、适用范围与排错能力，才能进入公共判断。技术创新与证据制度因而相互牵动。

然而，这不应被简化成旧共同体拒绝事实、新共同体拥抱事实。有人可以确认望远镜现象，却不接受发现者的全部宇宙论推断；有人可以支持某种新安排，却对某项新现象解释错误。共同体不是两个完全同质的阵营，分歧经常沿着具体问题重新组合。

宗教与科学的关系同样不能用一条固定对立线概括。具体机构、文本解释、政治环境和个人判断各有作用。承认这一复杂性，不是淡化限制研究的现实事件，而是防止将所有技术争论都归结为宗教，将所有宗教人物都写成同一种立场。具体阻碍需要具体材料，不能靠宏大标签代替。

判据改变还发生在教学中。学生需要学习新坐标、新轨道和新仪器操作，教师需要把复杂成果组织成可理解课程。教科书往往删去发现过程中的失败，提供较稳定版本。这有利于训练，却也可能让后人误以为新理论从诞生起就如此完整。思想史需要重新打开被教学封装的过程。

出版使成果更容易传播，也把比较扩大到作者无法控制的场域。读者可以采用局部方法，指出错误，或将材料用于另一理论。出版因此是知识继续发生的条件，而不只是把已经完成的思想复制出去。与此同时，印刷错误、版本变化和获取不平等，也会影响传播质量。

三方程在此呈现为共同体层次的反馈。新成果 S 改变训练和检验程序 D，新的程序又改变可信证据与合格解释的联系 E。这个过程可能形成更可靠标准，也可能产生新的排除。一个成熟标准应允许有根据的挑战进入，而不是把已经形成的共识当成不再需要说明的终点。

因此，证据制度的改变不只是权威从一个群体转给另一个群体，而应体现为判断可以承担更明确责任。什么观察能够反驳主张，什么误差必须报告，什么推断仍是候选，应越来越清楚。只有在这个意义上，制度改变才与知识可靠性的增长建立实质联系。

20.4 从一次革命到长期过渡

长期过渡意味着不同问题以不同速度解决。行星顺序的重新组织，可以早于地球运动的充分物理说明；某些旧宇宙安排被观察排除，可以早于新动力学建立；新算法进入计算，可以早于广泛教学接受。把这些变化压成一个年份，会失去革命真正的结构。

旧方法的继续使用也不必全部解释为落后。如果它在某个有限任务中足够准确、成本较低，保留它可能合理。关键是使用者是否知道适用范围，是否把方便的近似误当成总体真理。新体系同样需要这样的自我限制，不能因为具有更大解释力，就宣称所有旧成果全部无效。

用本书的方法说，革命不是 S 孤立替换，而是 S、D、E 之间的关系逐步重构。一个新显露只有改变后续比较，并在新条件中稳定下来，才成为可延续能力。它的稳定又为更细差异

提供背景。于是，发生并非一次从混乱到秩序的单向跃迁，而包括形成、检验、修正和再次开放。

统一性、多样性与和谐性在这段历史中各有具体承担。统一性体现在不同天象越来越能进入共同关系；多样性体现在不同仪器、模型和研究入口继续提供检验；和谐性则要求这些差异在共同约束下相互校正。若只追求统一，可能过早封闭竞争；若只赞美多样，可能拒绝已有证据的排除；若把和谐理解为没有冲突，又会失去改变的动力。

从哥白尼到伽利略的变化，因而应被理解为一个新的天文学世界持续构成的阶段。地球身份、行星轨道、月面属性与可见证据相继改变，它们没有自动拼成最终体系。牛顿将面对的，正是如何把已经取得的运动关系放进更广的动力学，同时让新的统一继续接受观察检验。

本章参考文献

[20-1] Todd Timberlake. Modeling the History of Astronomy: Ptolemy, Copernicus and Tycho. 2013。https://arxiv.org/abs/1301.2119

[20-2] Galileo Galilei. The Sidereal Messenger. Edward Stafford Carlos 译，1880。https://www.gutenberg.org/files/46036/46036-h/46036-h.htm

[20-3] Christopher M. Graney. Of Mites and Men: Johannes Kepler on Stars and Size. 2018。https://arxiv.org/abs/1802.03313

[20-4] Albert Van Helden. Copernican System. The Galileo Project, Rice University。作为 16—19 章事实脉络的综合参照，不据此宣称所有地区同步接受。https://galileo.library.rice.edu/sci/theories/copernican_system.html

[20-5] Nicolaus Copernicus. On the Revolutions. Edward Rosen 英译，奥西安德前言与哥白尼序言。https://math.dartmouth.edu/~matc/Readers/renaissance.astro/1.1.Revol.html

第 05 编

轨道动力学与精密位置

第 21 章 牛顿：从天上轨道到普遍动力学

第 22 章 预报回归的天空：哈雷与星历实践

第 23 章 天王星残差与海王星：预测怎样生成新对象

第 24 章 恒星光行差、视差与“地球会动”的证据

第 25 章 标准天球、时间与测量误差的制度化

第 21 章 牛顿：从天上轨道到普遍动力学

21.1 开普勒定律能说明什么

一条轨道可以画得非常准确，却仍然留下一个问题：物体为什么不断沿着它运动？开普勒把行星运动组织成椭圆、面积与周期的数学关系，使天空不再只是需要逐项描述的位置集合。然而，从轨道的几何秩序到能够解释运动改变的动力学，尚有一段不能省略的距离。给出某行星在某时刻的位置，与说明改变它的速度需要何种作用，并不是同一种知识任务。

这一差别不宜写成“开普勒只描述，牛顿才解释”的简单分工。开普勒也追问物理原因，太阳在其研究中并非毫无作用的几何标记。重要的区别在于，某种原因说明是否已经与运动变化建立可计算、可迁移而且可以接受检验的关系。如果一个解释只能说太阳推动行星，却不能约束行星快慢如何随位置改变，它所承担的论证责任仍与成熟的动力学不同。

对于椭圆运动，太阳位于焦点这一关系，把位置序列组织成一个整体；等时扫过等面积，则将图形中的面积变成时间的尺度。周期与轨道尺度之间的关系又把不同的行星联系起来。三种规律的结合具有特殊力量：它们不再允许研究者对每颗行星分别指定任意速度表。然而，它们仍然没有单独决定行星之间是否相互作用，也没有直接给出潮汐、彗星或落体的共同处理方法。

牛顿《自然哲学的数学原理》的序言明确说明了两个方向的工作：由运动现象研究作用力，再由这些作用力推出其他现象。第一、二编给出一般数学研究，第三编将其用于世界体系。[21-1] 这种往返尤其值得注意。如果只保留从定律推出轨道的方向，科学史就会变成一套成熟理论向事实的单向覆盖；如果只保留从轨道概括定律的方向，则会看不见数学构造如何使原本没有明确形式的问题成为可求的问题。

用本书的术语，开普勒规律已经是某种稳定显露。牛顿研究面对的并非未经组织的“原始天空”，而是经过观测、制表、几何重构的天空。新的差异序列将轨道曲率、面积增长、周期和地面运动的差别置于同一研究之中。新的特征纠缠则使质量、运动量、作用力、空间位置和时间间隔相互规定：离开相应测量和推导关系，一个孤立的“力”字还不能生成动力学世界。

这不是把牛顿力学归结为研究者的主观命名。在这里，“新的天文学世界”具有很具体的含义：行星成为可相互扰动的质量体，轨道成为初始条件与作用规律共同规定的运动结果，观测残差成为可能揭示未计入作用的差异。新的对象身份与新的运算能力同时形成。其有效性须在推导、观测与实际预测之间反复经受限制，不能由本体论表述替代。

还应看到，统一并不意味着先把所有差异抹掉。落体有空气阻力，摆有支点摩擦，月球同时受到太阳和地球影响，行星也并不严格沿着孤立二体椭圆运动。动力学要建立共同关系，恰恰必须判断哪些作用在一个问题中可以暂时忽略，哪些差异到了另一个精度就不能再忽略。理想化因此具有双重身份：它使问题可解，也给出了不能无限扩张的条件。

可以用一个现代教学对照理解这件事。若只掌握一辆车每天到站的时刻，我们可能归纳出一个稳定时刻表。若要解释它载重变化、道路坡度或发动机输出改变以后会怎样，仅有时刻表便不足够。这个对照不是把行星想象成有发动机的车，而是说明规律之间的责任差别：描述已经稳定的序列，与研究使序列维持或改变的条件，必须连接却不能混同。

同样，牛顿的普遍性不能从“天上与地面相似”直接得出。它必须说明相似究竟发生在哪一个量的关系上，以及为何不同尺度下仍能采用共同度量。月球与地面重力之间的联系，依赖地球尺度、月球距离、运动时间等条件。测量、几何与动力学任何一个环节改变，都可能改变比较的结果。所谓普遍，首先是一项跨问题的严格约束，而不是一个宏大的形容词。

21.2 轨道与向心力之间的推导

《原理》第一编开头的面积论证，并不是现代课本中对一个现成微分方程的简单积分。牛顿以几何构造处理运动：物体在一段短时间中沿直线前进，在指定时刻受到朝向某一中心的冲量，随后沿另一条直线前进。连接中心与运动位置，可比较连续三角形的面积。在相等时间下，中心指向的冲量不破坏相应面积关系；再通过极限过程将折线运动联系到连续曲线。[21-1]

这个构造的重要性在于，它没有先假定轨道是椭圆。面积性质首先联系于作用方向，而非某一种特定的距离依赖。因而，看到等时扫过等面积，并不足以单独证明逆平方力。它约束的是加速度的中心性；要进一步确定作用强度如何随距离变化，还需要轨道形状等信息。把这两步合成一句“面积定律证明万有引力”，就会丢失推理中最有价值的分层。

第一编第十一命题考察以椭圆焦点为力中心的运动，建立相应的逆平方关系。这里处理的是一个条件问题：给定轨道及中心位置，求使这种运动成立的作用规律。反向问题则是给定作用规律和初始运动，求可能轨道。两者相关，但在历史论证与数学证明中仍有区别。现代研究对牛顿几何证明的细节及其极限论证有专门讨论，不能把今天熟悉的符号体系当成原书逐字采用的形式。[21-2]

为了显示条件关系，可以另写一个现代说明式。在理想的孤立二体系统中，将相对位置记为 r，将两个质量记为 M 与 m，相对加速度满足与 $G(M+m)/r^2$ 相应的中心规律。由此得到轨道周期平方与半长轴立方之比，由系统总质量决定。这种写法是现代重述，不是对牛顿

原始符号的摹写。它让一个常被忽视的限定显现出来：不同系统的比例常数并不可以无条件地当成同一个数。

质量、距离和时间在这里不是三个互不相干的观测标签。假如轨道尺度的换算发生改变，而周期保持不变，从运动反演得到的引力参数便会相应改变。又如，只凭一颗卫星的轨道测量，常常首先约束的是引力常数与质量的乘积，而不是分别独立地知道两者。我们今天使用的具体测量方式已经高度发展，但这种参数之间的约束关系，从根本上说明了动力学量为何要在整个关系中获得意义。

这里的 E 因此不能写成“时代、数学、望远镜、社会”四项背景的集合。真正承担推导的纠缠是：怎样定义运动变化，怎样把几何距离与时间对应，怎样区分引力作用和惯性运动，以及怎样选取可以忽略其他作用的范围。改动其中一项，其他量的可比性和解释身份也会改变。共同体、仪器与出版当然重要，但它们如何进入这一关系，需要分别说明。

牛顿式统一的一个生产性结果，是让偏离本身取得了更精确的身份。如果行星不是严格二体系统中的粒子，那么椭圆就不应是不可违背的绝对模板；它可以成为进一步摄动计算的出发点。这样，理论一方面提供稳定规律，另一方面也使不稳定部分能够被定位。若观测位置偏离计算位置，研究者可以追问初始条件、质量估计、其他天体作用或观测改正，而不必立即从头另造一套宇宙图景。

但这也带来风险。可以调整的参数越多，理论越容易在事后吸收差异。若所有偏差都可以通过未加约束的新参数消除，那么“符合”就会失去检验力量。动力学的可靠性并不等于它可以拟合每一段已知记录，而在于同一组约束能否支持新的位置、新的时段和不同对象。数学统一与经验支持因此始终互相需要，却不能互相冒充。

从六路径看，由轨道差异进入作用关系，和由既有定律回查历史资料，会产生不同警觉。前一种路径看见新关系如何生成，后一种路径容易把原来尚不稳定的环节误认为早已确定。换用路径的实际价值，在于使研究者重新问：这一项结论在当时究竟是由哪些资料支持，哪些步骤仍然依赖理想条件？路径不是给历史套一个箭头，而是改变检查的次序和所暴露的问题。

21.3 彗星与潮汐的新问题

若一个理论只重新解释产生它的行星规律，它的扩张能力尚未充分显示。彗星提供了另一类运动：轨道倾角可能很大，运动方向未必与主要行星相同，显著可见的时间又很短。它们不像整齐地排列在近似同一平面的行星，因此特别适合检验共同动力学是否只能处理熟悉的规则样本。

在牛顿的体系中，圆锥曲线不只是描述行星的椭圆。不同初始条件可以对应不同的轨道类型，彗星的短观测弧于是有机会进入同一个动力学研究。问题随之变得尖锐：在地球也运动的情况下，望远镜取得的方向记录怎样变成彗星相对太阳的轨道？有限时段的数据能否区分很长周期的椭圆与近似抛物线？这些都是新体系创造的研究任务，而不是写下逆平方关系后自动解决的细节。

彗星说明“允许存在”与“已经确认”之间的差别。数学理论可以允许某一类轨道；要证明具体彗星采取哪一条轨道，还需要日期、方向、地球位置及误差估计。即使观测拟合得很好，对远离可见弧段的运动也可能仍有很大不确定性。把短时间的位置记录外推几十年，恰好把这一不确定性放大到公共检验面前。

牛顿在序言中把彗星轨道与月球不均匀运动、多个物体相互作用等列入其研究扩展，同时承认月球问题的处理存在不完善之处。[21-1] 这一自我限定很重要。它显示《原理》的力量并不在于交付了一本已经算完全部天空的答案册，而在于建立了一套能持续产生、约束并修订问题的方式。一个统一理论可以在总体上极其成功，在某些计算上仍然需要几代人的改进。

潮汐则把统一带到了另一种困难之中。海水的涨落与天体位置存在联系，但要从引力写到某港口的实际潮时和潮高，必须跨越海盆形状、水深、海岸、地球自转和流体运动。太阳与月球对地球不同部位作用的差别提供了潮汐解释的关键；港湾中的海水却不是一个完全自由、均匀覆盖地球的理想薄层。

现代说明中，潮汐常以引力的空间变化率来讲解。距离较近一侧与地心受到的作用不同，距离较远一侧亦然。这种差别不能等同于“月球把所有海水同时拉向自己”的简单图像。更重要的是，即使驱动力的基本性质已经知道，局地响应仍受海洋结构约束。建立共同原因，并不取消不同地方的动力响应。

这一点对本书的整体性分析具有直接意义。若把 S 只指定为某港口每天出现的高潮，就必须把相关流体条件纳入构成；不能拿太阳和月球的引力解释直接覆盖一切。若把 S 指定为潮汐驱动力的周期结构，则研究单位不同，所需条件也相应不同。同一个现象名称之下可能包含不同粒度的问题。若粒度不明确，三方程就会变成任何因素都能往里放的容器。

因此，对牛顿统一的评价需要区分三个层次：共同动力学关系的提出，特定自然问题的模型化，以及具有可用精度的数值预测。第一层成功不保证第二层无需调整；第二层成立也不保证第三层在现有观测与计算能力下已经完成。潮汐和月球研究都说明，求解能力的不足可以暂时限制预测，而不立即等同于基本理论错误。但这种区别也不能成为免疫装置，长期、系统、无法消除的偏差仍然可能迫使研究者重新检查理论。

在发展动力上，新体系所释放的能力反过来积累了新的张力。过去只能说“运动不规则”的地方，现在出现了具体残差；过去没有统一计量的对象，现在需要共同质量尺度；过去只被当作偶然天象的彗星，现在要求判定它是否能够回来。这种回写并非抽象的“知识影响知识”，而是研究工具、采样时机和问题优先级发生了可辨认的改变。

若用三方程压缩这一过程，显露的动力学秩序由差异运行与数理条件共同支持；一旦成立，它又重排下一轮差异运行；当新问题不断显现，构成计算与测量的关系也被修订。但正文的价值不在于写出这三句，而在于能够指出每一次变化的具体内容。彗星把等待时间延长到跨世代，潮汐把几何天文学带入流体与地理条件，它们是统一之中真正不同的研究世界。

比较原件与转录还有另一层价值。剑桥保存的《原理》相关材料使版本身份可以被追查。[21-3] 不同版次包含修订，研究者引用“牛顿说”时，应说明自己依据哪一个文本。成熟教科书中的规范表达有助于理解，却不能自动充当一六八七年论证过程的直接证据。

21.4 《原理》如何进入共同体

一部书的印刷完成，与其中的方法成为共同体能力，并不是同一个时刻。《原理》的几何论证要求读者掌握相当高的数学训练；它又跨越运动原理、流体问题和天象应用，使读者不容易一次把握所有环节。即使赞扬它的人很多，能够继续补充计算、检验具体天象的人也仍然需要成长。这正是知识传播不能被当作内容简单搬运的原因。

哈雷的作用在牛顿自己的序言中已有明确记录：他推动研究结果向皇家学会传播，促成出版，并参与校对与图版事务。[21-1] 这种贡献不能被缩减成伟大理论旁边的一段友谊故事。手稿要成为可以共同查阅、引用和批评的书，符号、图形、印刷与发行都必须稳定。一个图形关系排错，可能直接影响读者理解证明；一个命题引用不一致，可能使推导链无法追踪。

不过，传播环节的重要性不意味着证明真伪由出版制度决定。哈雷能帮助定理进入共同体，却不能凭支持把错误定理变成正确；共同体能组织观测，却不能通过投票让彗星按预报回归。这里有两种需要同时保持的约束：知识的公共可用性依赖组织与媒介，知识主张的具体成立还要接受其对象和方法的检验。

后来的研究者通过不同数学语言重写天体力学，使原先不易处理的问题获得新的求解形式。分析方法、摄动展开和数值计算的发展，不只是给相同结论换外衣；它们会改变哪些问题可以被计算、哪种精度可以被达到，以及怎样分辨误差来源。牛顿的世界因而不是一个在一六八七年封闭完成的整体，而是一个不断发展其内部能力的开端。

这也解释了为什么不能以“牛顿发现了一个本来就在那里的公式”概括全部贡献。公式的自然适用性需要独立支持，但公式进入研究后所产生的对象组织和行动结构，是历史上新

形成的。行星质量、摄动、轨道反演和长期预报，逐渐成为能够由不同研究者接续完成的公共任务。这里的新世界不是任意想象的宇宙，而是新显露、新差异操作与新特征纠缠构成的天文学复合体。

也不宜把旧世界说成从此消失。几何星表、实用历算与肉眼观测仍有作用；在某些任务上，不必每次重新计算全部动力学才能获得可靠结果。新体系能够包含旧方法的有效部分，同时改写其解释地位。一个轨道公式可以继续有用，却从终点变成近似；一项经验改正可以继续使用，却需要说明它代表何种未展开的动力作用。

统一性的真正考验由此出现：它是否容许不同问题保留必要差异，又能在共同关系中互相约束？若要求所有现象都呈现相同表面形式，统一便会压制多样性；若每遇到一个复杂问题就另立完全独立的规则，统一又会失去内容。牛顿力学提供了一个重要历史答案：在共同运动关系下，以明确的近似、不同初始条件和逐步增加的作用项组织差异。

本书的 SDE 解释在这里需要保持自己的边界。关于牛顿证明的正确性，应由数学和史料检验；关于引力模型的适用性，应由物理与观测检验。SDE 所追问的是，这些检验怎样与新的对象身份、研究操作和相互规定的特征整体形成联系。它若只是把已有动力学换成三个字母，就没有额外贡献。它能够提供的，是把“定律成立”进一步展开为“何种天文学世界由此获得持续发生能力”。

在这个世界中，知识的完成形式不再只是已经解释过去。它开始要求未来天空承担检验责任。彗星是否回来、回到何处、何时经过近日点，成为任何赞誉都无法代为作答的问题。牛顿动力学最深远的后果之一，是使这样的等待具备了可计算的内容，也使尚未发生的观测成为理论生命的一部分。

还有一个常被统一叙述遮蔽的条件，是研究者选择的坐标和所讨论的运动对象。太阳在近似行星模型中可以作为便利的中心，但这不等于太阳在所有严格计算中绝对静止。当多个质量体相互作用时，共同质心的处理能够使运动关系获得更一致的表达。若把计算上的便利中心当成宇宙中绝对不动的实体，就会让新动力学重新落入旧的中心争论之中。

同样，观察到的天球方向并不是动力学方程中的空间位置本身。观测者需要把地球的运动、光传播以及仪器读数联系起来，才能让理论轨道与可见位置相比较。在不同精度层次上，哪些改正必须计入会发生变化。理论与观测的符合因此总包含中介步骤，而不是两列天然同类的数字直接相遇。这些步骤越明确，未来出现偏差时就越容易判断责任。

由此可见，普遍动力学并不是一条公式征服所有对象的单向过程。它也依赖其他研究不断提供地球尺度、观测时间、位置标准和质量关系。所谓普遍性，是这些不同研究能够在共同运动关系下彼此约束的结果。它的历史力量来自这种持续协调；它的历史限制，也应在同样具体的协调环节中寻找。

本章参考文献

[21-1] Newton, I. Philosophiae Naturalis Principia Mathematica, 1687；公开原文转录。https://www.gutenberg.org/files/28233/28233-h/28233-h.htm

[21-2] Nauenberg, M. Kepler's Area Law in the Principia: Filling in some details in Newton's proof of Prop. 1, 2001. https://arxiv.org/abs/math/0112048

[21-3] Newton, I. Philosophiae naturalis principia mathematica，剑桥大学数字馆藏原件入口，用于版本与原件识别。https://www.repository.cam.ac.uk/items/4e101d5e-b2d2-4e27-a189-3229f4f0938b

第 22 章 预报回归的天空：哈雷与星历实践

22.1 跨时代认同同一彗星

同一颗彗星能够在相隔几十年的记录中保持身份吗？这个问题看似只要知道轨道便可回答，但对早期彗星研究者而言，轨道本身就是需要从分散记录中重建的东西。一次出现时的形状、亮度与尾巴可能不同，观测地点、仪器与使用的星表也会改变。若只凭外观相似认定同一对象，会有太多偶然相似；若要求外观完全相同，又会把一个会改变的天体拆成彼此无关的事件。

哈雷在一七〇五年发表彗星研究概要，将多个时代的彗星运动放入牛顿动力学的共同框架。他特别比较一五三一年、一六〇七年和一六八二年的出现，提出它们属于同一彗星，并预报其再次回来。[22-1] 这里的关键不是把三个年份排出一个近似间隔，而是让各次观测所得的轨道方向和几何关系具有可比较性。相近的时间间隔提供线索，相近的轨道结构才使身份判断获得更强约束。

一颗彗星的历史身份由此具有一种特殊结构。人们没有在几十年的整个旅程中连续看见它，却通过有限可见时段之间的动力学联系，将中间不可见的运动纳入同一对象。对象同一性不再只靠连续感官接触维持，也不完全靠名称维持；它依赖一套能够把分离记录接起来、同时排除任意连接的关系。

这一点与普通档案中的人名对应既相似又不同。不同年代的记录可能指向同一人，但要结合年龄、地点和关系网判断，不能只看姓名。彗星的对应也需要多重特征；其独特性在于，这些特征受到可计算的运动定律约束。轨道平面、近日点方向和出现时机并非可以随意组合的栏目，它们共同规定何种连接在动力学上成立。

SDE 分析因此把“同一彗星”视为一个需要生成的科学身份，而不把它当成观测前已经全部明确的标签。这不否认彗星在未被看见时仍然运动。物理存在的连续性与科学研究中如何确立连续身份，是不同问题。后者需要前者的约束，却不能从一句“它一直存在”自动获得可以计算的证据。

这里的差异运行包括对旧记录的重新读取。原先只是报告某年出现的一次天象，现在可以成为周期检验中的一个数据点。旧记录的用途因新假说而改变，但记录的质量并不会因此自动提高。一项日期含糊的描述、一个没有精确位置的记载，仍然不能与良好测量等量齐观。回写可以改变问题身份，不能追溯性地创造原来没有留下的精度。

由此还可以理解为什么历史记录的缺失很重要。若某一预期回归时期没有记录，可能是彗星位置不利、亮度不足或档案不存，也可能确实影响身份判断。不能把所有缺失都视为反

证，也不能把所有缺失都解释成资料损坏。可靠的身份研究需要把“没有发生”“没有看见”和“没有保存下来”区别开来。

哈雷的预报正是在这一不完整的历史地基上提出。它既不是从零开始的大胆猜测，也不是一项已经由旧记录完全证明的结论。其科学价值来自一种明确的风险：如果同一性假说和动力学解释成立，未来应当出现可以与它们相接的新观测。过去的相似因此转变成了对未来的约束。

22.2 预报为何不是任意命中

预报的成功不能只用“后来果然出现一颗彗星”来定义。天空可能出现不同彗星，如果没有时间和位置的约束，总能从事后找到一个看似相合的事件。真正有内容的预报，需要在观测发生以前说明它期待什么，也需要使可能的失败保持可识别。否则，预言就只是等待一个能够被重新解释为命中的现象。

哈雷预报的回归通常以一七五八年表述；彗星在该年圣诞节重新被观测到，随后于一七五九年通过近日点。[22-2] 这里必须区分首次再次被看见的日期和通过近日点的日期。一个指向回归年份的预期，与后来计算给出的近日点时刻，不是同一种精度的承诺。若混为一谈，既可能不公正地说预报错了一年，也可能夸大成对某一天的精确命中。

在思想史中，这种日期区分不只是校正年表。它关系到如何评价理论的风险。当预报只给出一年左右的范围时，它承受的约束与一个月的窗口不同；当预测位置能够指导定向搜索时，它又增加了新的约束。科学判断应当按预报发表时真正给出的信息评价，而不是把后来获得的精度倒加给最初提出者。

还要区分轨道身份的预测与可见性的预测。彗星走到某一空间位置，并不保证观察者一定能见到它。太阳的方位、天空背景、天气、望远镜能力和彗星自身活动都影响发现。对于位置理论的检验，未被看见只有在具备足够搜索覆盖时才具有相应力度。若根本没有人观测相关天区，“没有发现”并不等于该处没有彗星。

这说明预报需要一个相互规定的整体。轨道模型规定搜索时空，搜索能力规定可检验范围，误差估计规定什么算作偏离，观测报告则使偏离可以被公共讨论。把这些因素分开列出，尚不足以说明它们的关系；必须指出改动一项如何影响成功或失败的判断。例如，把搜索窗口任意扩大十倍，或许提高找到彗星的机会，却降低了预报本身的区分力。

从 SDE 三方程看，周期彗星这一显露并不是研究终点。它立即改变差异序列：观测者在指定时段等待，把注意力放在指定区域，并将新位置与历史轨道比较。与此同时，对预报精度的期待又促使计算者重新审查行星摄动和旧资料。S 由 D 与 E 形成，S 又规定下一轮 D，这一回写在搜索安排中具有可见的形态。

不过，定向搜索也会造成一种认识偏向。研究者更容易注意到预报附近的候选，而忽略远处的异常；更容易把大致相合看成支持，把不相合归因于观测质量。因此，成功的预报制度还需要保留搜索范围和失败记录。检验并非理论与天空的一次抽象碰面，而是对何时看过、看到了什么、哪些区域仍未检查的完整交代。

哈雷彗星的故事之所以重要，正因为它把预报从个人确信变成跨世代任务。提出者不能亲自控制几十年后的观测结果，后来者也不能改写已经发表的基本预期。出版的日期在这里构成了一种时间约束：它把猜想放到尚未发生的观测之前。公共文本因而具有一种特殊作用，使“我早就知道”可以被档案检验。

22.3 预测中断与持续修正

周期并不意味着每次回归间隔严格相等。一颗彗星在太阳系中运动，也受到行星作用；其速度与轨道会因接近不同天体而改变。若把历史间隔直接复制到未来，就会把多体问题误写成理想时钟。哈雷的基本身份假说需要由进一步的摄动计算支撑，才能从回归年份推进到更窄的时刻范围。

克莱罗、拉朗德与妮科尔—雷娜·勒波特共同参与了回归前的大量计算，考虑木星、土星对彗星运动的影响。巴黎天文台的历史资料特别记录了这一合作。[22-3] 它说明成功并非一位理论家在纸上写出一句话，然后天空负责照办。粗略周期、理论修正、人工计算与观测搜索构成了一个分工系统，其中每一项都可能限制最终精度。

多体摄动的困难并不只是算术量大。研究者需要判断哪些作用最重要，怎样近似轨道，以及误差如何沿着长期计算累积。越接近某颗大行星，简单平均近似越可能不足；越长的外推时间，微小速度差越可能变成明显位置差。有限计算能力迫使人们在准确性、工作量和可完成日期之间作出具体选择。

在这一情况下，“不断修正”有两种性质不同的可能。第一种是在新观测发生之前，用明确的动力学理由改进预测，公开修正及其范围。第二种是在结果已经知道之后移动目标，使任何结果都显得符合。两者都叫修正，却有不同证据身份。前者让理论承担更严格约束，后者可能消除约束。思想史应当沿着发表、计算与观测的时间顺序区分它们。

这一差别也是回写理论必须回答的问题。回写并不天然意味着进步。结果可以促使人们改进模型，也可以促使人们保护模型而降低判据。若把“发生变化”直接等同于“正向发生”，就无法识别两者。需要检查改变了什么：是摄动项获得新的独立依据，还是仅仅把允许误差扩大到足以容纳结果？

在彗星问题中，周期不齐并未立即推翻周期身份。恰恰相反，多体动力学允许同一对象在不完全相同的时间间隔后回来。这里的同一性不是所有属性不变，而是变化能够在连续的

约束关系中得到说明。一个稳定对象可以带着不断改变的速度和亮度维持身份；稳定性发生于联系之中，而不是发生于一张永远不变的属性清单。

这种思想对于理解天文学对象尤其重要。我们通常把星名和编号当作稳固标签，但标签能够长期有效，是因为观测与模型提供了足够连续的联接。若彗星出现分裂、轨道受到较大扰动或观测弧不足，身份判断就可能变得复杂。目录中的一个条目并不是自然向人类递交的一张身份证，而是一项能够继续接受检查的科学认定。

哈雷彗星的回归也没有证明牛顿理论的每一个应用已经无误。它支持的是一组具体联系：历史出现可以属于同一对象，太阳引力与行星摄动能够有效组织其运动，相关观测与计算具有足够可靠性。把一次成功扩张成“整个理论从此永远正确”，就超过了证据。真正增强理论的，是它在不同对象、不同精度与不同操作链中的持续表现。

在三原理的分析中，这里最清楚的动力来自稳定身份与非等周期之间的张力。若坚持严格重复，就无法解释差异；若放弃身份，就失去跨时代预报。摄动计算改变了两者之间的关系，使“同一对象的可解释变化”成为新的稳定形式。这种和谐并非没有差异，而是差异获得了可以共同接受的组织方式。

摄动修正还有一个可计算的直觉。假设两个相邻轨道模型在某次经过太阳附近以后只存在很小速度差，远离太阳的漫长运动会让其时间差逐渐积累；再次接近太阳时，空间位置可能重新接近同一轨道区域，而通过该区域的日期却不同。因此，轨道外形看起来相似，并不保证近日点时刻足够准确。时间误差需要独立评估。

对于搜索者，日期误差与方向误差还会互相影响。彗星在天空中的角速度并不处处相同，预报晚了若干天，在某段轨道可能只造成较小方向偏离，在另一段则可能偏离明显。一个合理搜索计划需要根据局部运动速度确定每天覆盖范围，而不能把同样大小的天区窗口机械用于所有时刻。这是将动力学预报转化为观测任务的具体步骤。

旧观测的年代也会改变其约束作用。若两次出现之间间隔很长，周期的微小差别会在下一次回归时累积；但若旧日期本身粗略，便不能假装其周期约束极精确。对日期、位置与轨道的交叉使用，使计算者能够判断哪里值得继续投入工作。预测精度的提高来自找出限制环节，而不只是重复同一计算更多遍。

22.4 星历的集体工程

一次彗星回归具有戏剧性，星历的日常工作却往往没有同样的可见度。星历需要持续给出位置与时间，使观测者、航海者和其他使用者能够依照共同资料行动。它的可靠性不只来自某一条伟大定律，还来自大量重复计算、误差核查、表格编辑与及时发行。若一项正确结果被排进错误日期，它在使用中仍然可能造成错误。

这种工作把理论转变为可调用的公共能力。使用者不必每次重新推导全部轨道，却必须知道表格使用何种时间、坐标和插值方式。封装在这里有明确的内容：复杂计算被压缩为较简洁的查询形式，但适用条件和误差范围必须被保留。若只留下数值而删掉条件，便利就会变成误用的来源。

星历还改变了观测本身。观测者不再只记录天上出现什么，也会比较计算位置与实际位置。这样的残差序列可以揭示观测仪器问题，也可能揭示动力学计算不足。星历既是理论的结果，又是生产新差异的工具。过去零散的偏差被组织成连续的可比较对象，研究的重心因此从单次奇观移向长期一致性。

这里存在一个需要警惕的循环：如果某一批观测已经用于调整轨道，再用同一批观测的良好符合证明预测能力，就没有增加独立支持。可靠评价应当区分拟合资料与后续资料，或至少说明两者并不独立。这种要求并非现代统计才突然发明的额外苛刻，而是预报为何有意义的逻辑条件。

共同体的劳动分工也决定哪些成果容易被看见。理论提出者、计算者、观测者、表格编制者与档案保存者的贡献不同，却彼此依赖。对勒波特等计算参与者的恢复，不应只停留在补充姓名。它要求改变叙述单位：把“一个人的预言”写成具有不同责任的合作过程，同时仍然能够准确区分谁提出身份假说、谁完成计算、谁首先重新观测。

这种多样性并不削弱统一。正因为任务不同，统一需要可以交接的中间结果：统一的日期表达、可检查的轨道参数、说明清楚的观测记录和明确的修正规则。若每个人使用自己的隐含尺度，合作人数越多反而可能增加混乱。共同标准并非消灭个人判断，而是让不同判断能够在同一问题上发生有效碰撞。

哈雷案例因而让天文学中的时间取得双重地位。时间一方面是彗星运动的变量，另一方面也是知识检验的组织形式：旧记录被重新读取，预报被事先发表，计算在回归前完成，观测在未来发生，结果再进入新的星历。自然运动的时间与研究活动的时间相互连接，却不相互替代。

从这一点看，预报能力的发生并不是给已经存在的知识添上一个实用功能。它改变了知识的结构。一个关于过去的解释，被要求承受未来的不同结果；一套关于单次出现的描述，被要求维持跨世代同一性；一种由少数人掌握的计算，被要求成为可以交接和复查的公共工具。这些变化共同构成了“预报回归的天空”。

它仍然是一片开放的天空。成功回归使人们更愿意相信残差能够透露新的天体作用，但这种信任随后也会遇到边界。同样的动力学信心可以帮助发现新行星，也可能支持一个最终没有得到确认的候选。哈雷彗星提供了强有力的成功范例，却没有授予所有未来预测自动成功的资格。

本章参考文献

[22-1] Halley, E. A Synopsis of the Astronomy of Comets, 1705，史密森学会馆藏及数字化书目。https://library.si.edu/digital-library/book/synopsisofastron00hall

[22-2] NASA. 1P/Halley，历史回归与轨道资料。https://science.nasa.gov/solar-system/comets/1p-halley/

[22-3] Observatoire de Paris. Halley's, the first periodic comet，回归计算合作史料说明。https://cometes.obspm.fr/en/comete-halley-premiere-comete-periodique.html

第 23 章 天王星残差与海王星：预测怎样生成新对象

23.1 “剩余差异”有哪些来源

一颗行星没有走到计算要求的位置，并不立即意味着另一颗行星藏在附近。可能是旧观测不够准确，可能是已知天体的质量估计有误，可能是初始轨道没有拟合好，也可能是计算尚未充分计入各种作用。只有这些可能性受到足够约束，剩余差异才可能成为新对象的线索。海王星的发现因此不能从“天王星偏离”直接跳到“必有海王星”。

天王星被确认为行星以后，相关研究还利用了更早记录中未被认作行星的观测。旧位置与新测量结合，延长了可研究的时间跨度，却也把不同年代的误差带进同一计算。早期记录未必使用相同星表，仪器精度不同，记录者也未必特别重视一个普通星点。长基线资料带来更强约束，同时要求更细致的资料辨别。

所谓残差，是观测值与模型计算值之间的差。它不是天空独立送来的一种物体，也不是可以脱离计算程序理解的纯感觉。改变星表、改正时刻、调整已知质量或轨道参数，都可能改变残差。因而，残差的科学身份始终同时依赖两端：观测怎样形成，理论怎样生成可比较数值。

勒维耶在一八四五至一八四六年集中研究天王星运动，逐步提出外部扰动行星的约束；亚当斯也独立开展了相关计算。历史研究对两者的资料、发表与搜索过程有更细致的区分，不能把它们压缩成完全相同、同时完成的发现。[23-1] 对本章最重要的，并非先决定谁应该得到多少荣誉，而是说明预测怎样从一项异常判断推进到能够安排观测的具体对象。

如果观测残差纯属随机误差，它应有与误差模型相称的分散特点；若残差具有随时间变化的结构，则值得检查未计入的系统作用。但有结构也不足以证明新行星，因为仪器系统误差或错误的长期改正同样可以产生结构。规律性是进一步调查的理由，不是实体化的通行证。

由此可以看见特征纠缠的具体内容。天王星的“异常运动”由观测基准、轨道参数和已知摄动共同限定。某一参数的改变可能改善一段资料，却使另一段资料更差；某个新天体候选可能解释一类残差，却要求不合理的质量或位置。不同约束互相牵制，才使问题从任意猜测中收缩。E 在这里不是一个装着历史背景的箱子，而是一组不能各自独立指定的关系。

从 SDE 视角，真正值得追踪的差异序列包括残差如何随计算版本变化。若只呈现最后一张漂亮的残差图，就会掩盖哪些差异曾被视为可靠，哪些资料被降权，以及哪些假设限制了搜索。理论创新常常发生于这些中间判断之中。它不是把一份固定事实清单输入头脑以后自然输出唯一答案。

这也要求公平对待没有选择新行星解释的人。以当时资料的不确定程度，要求先核查观测或已知摄动未必是保守无知。一个能产生发现的方法，也应能说明为什么某些时候暂缓实体判断是合理的。若只有成功以后才能宣布哪种态度科学，方法论就会沦为对胜者的回顾性赞扬。

海王星案例的重要性在于，新增天体假说最终超出了对旧残差的内部解释。它提出可以在天空寻找的方向，并且随后出现了能够独立跟踪的天体。这一步把论证从“这个模型能解释已有误差”推进到“这个模型指引找到一个新观测对象”。两个阶段之间的距离，正是预测发现的核心。

23.2 反演如何约束天空位置

正向动力学可以从已知质量、位置与速度计算未来运动；海王星问题却要求由已知天体的运动差别推断未知天体。这是反演。反演往往比正向计算更难，因为不止一组未知参数可能在有限时间范围内产生相近效果。较大质量与较远距离之间可能存在补偿，轨道偏心率与相位的选择也会改变扰动形式。

因此，勒维耶的工作不是把残差代入一个现成公式，立刻得到唯一的行星位置。研究需要对候选轨道作出限制，再利用多时段资料寻找可接受参数。历史研究特别指出，预报成功不能理解为所有推得的轨道要素都已准确；观测天区的成功约束与完整长期轨道的精确恢复，是不同成就。[23-1][23-5]

这一点可以用现代说明模型解释。设某一组残差由若干参数共同控制，有限观测可能主要约束这些参数的某种组合，而不能把每一个都单独确定。在当前时刻的方向上，不同可接受模型可能靠得很近；向很远的过去或未来延伸，则可能明显分离。因此，一个非常有效的搜索位置不必意味着整个模型各处同样准确。

这种情况不削弱预测的实际价值。把广阔天空缩小为一个可用望远镜检查的区域，已经是一种巨大的能力。它只是要求精确地描述能力所在：理论成功地提供了定位线索，随后还需要观测决定真实轨道。若说成“笔尖已经完整描绘出一颗行星”，就把搜索工具与最终对象参数混在一起。

反演中的先验限制也需要适当评价。研究者必须有一些关于轨道范围与形式的初始判断，否则未知量过多，问题难以处理。但这些限制不是自然保证正确的条件。它们可以帮助找到解，也可能使搜索偏离。合理的做法是保留它们的条件身份，检查结论对它们有多敏感，而不是在成功之后把所有辅助猜测都追认为真。

在这里，六路径的转换有实质意义。从新天体候选出发回看残差，容易关注哪些资料支持它；从观测与误差条件出发重新推算，则会追问有没有其他参数组合也能解释数据。两条

路径的差别不是叙述顺序而已。第二条路径能够限制第一条路径对候选身份的过早确信，要求明确哪一项后续观测可以区分竞争模型。

新的天文学对象在此经历了一个尚未完成的阶段。它已经拥有候选质量、轨道与搜索方向，但没有被直接观测确认为那一颗天体。这个阶段不能说毫无对象，因为研究者已经能够围绕它计算与安排搜索；也不能说对象身份已经完全确立，因为同一组残差尚可在不同假设下解释。候选对象是具有明确检验责任的存在方式。

本书所谓科学世界的发生，在这一阶段尤其可见。未知行星并非由数学运算物理制造出来，但“能够通过其他行星的偏离而被定位的天体”这一对象类型，在相关理论与观测复合体中获得了新的稳定性。天体身份不再只从一个孤立光点开始，也可以从多个既有对象之间的运动关系中发生。

这种新能力还改变了“没有看见”的意义。一个天体没有肉眼可见，并不意味着它不能对已知运动留下可研究痕迹；然而痕迹本身仍需要多重约束。间接证据的力量来自结构，而非来自一种对不可见事物的普遍宽容。只有具体候选能够承担新的预测和排除条件，间接研究才可能继续成长为可靠知识。

若把反演过程写成现代教学问题，可以设想两组候选模型：第一组采用较远、较大的扰动质量，第二组采用较近、较小的扰动质量。它们在短观测段中可能给天王星近似相同的加速度变化，因而都能降低残差。但当观测时间延长，未知行星与天王星的相对方位发生不同变化，两个模型的预期便会分开。延长时间基线的价值，在于获得新的几何构型，而不只是增加重复次数。

不同方向的残差也有不同作用。单纯的整体位置偏移，可能被初始位置调整吸收；随时间持续累积的差别，可能涉及平均运动；更复杂的变化则要求检查周期摄动和几何关系。如果不先辨认残差的形态，就容易把可以由已知参数处理的问题过早归给未知实体。反演不是只看误差大小，更要看误差怎样变化。

还可以把某些旧观测暂时留在拟合之外。先用一组资料得到候选，再检查它能否合理解释未参加拟合的另一组记录，这会增加检验力度。但历史资料质量不一致，留出资料的失败也需要判断是否由误差造成，不能机械采用今天理想数据集的假设。它的原则意义是使候选承担额外责任，而非保证某种分组一定产生最终答案。

搜索范围的宽度体现了理论提供的实际信息。如果几个允许的模型在当前天空方向上分散很大，观察者就需要更多夜晚或更大视场；若它们聚集在较小区域，理论即使尚未确定每个轨道参数，也已经具有较强行动价值。对预测的评价，应同时记录方向约束、时间范围和后续可追踪性，而不只比较一个最漂亮的位置差。

这种分解还能解释为何同一份预测交给不同观测组织，可能得到不同结果。星图是否覆盖相关亮度，望远镜能否区分候选，记录是否及时与第二夜比较，都影响确认。理论贡献不能因此被否定，但理论与观测之间的接口也不能被当作自动完成。预测对象需要经过实际识别，才进入公共天体目录。

在论文和发现报告之间，应当保存候选的版本变化。一个最早估计、一个修正估计和观测后的轨道解，具有不同时间身份。如果只把最接近真实位置的一版保留下来，就可能高估事前约束；如果只盯住最粗略一版，又会忽视公开改进。思想史的严谨并非追求叙述复杂，而是让每项成就和每项限制落在正确时刻。

预报误差也必须用任务尺度判断。一个坐标与后来测得位置相差多少，只有同望远镜视场、搜索区面积和可辨目标数量一起看才有实际意义。若误差足以使目标落在计划搜索范围内，这个预报便可能成功指导发现；它不必同时精确决定未来几十年的轨道。相反，在多年星历中逐渐累积的时间误差，可能要求大幅重算轨道，却不抹去此前搜索行动的有效性。海王星案例中的预报成功与轨道参数偏差因此可以并存。它们分别回答到哪里找和以后怎样运动，判据不能完全相同。将发现后的观测追加进去，便是在保留成功搜索结果的同时，重新约束整个动力学模型。

23.3 从候选到发现的验证

一八四六年九月二十三日，柏林的伽勒与达雷斯特依据勒维耶提供的方向开展观测，找到了接近预报位置的新天体。随后发表的发现报告和位置观测，记录了从计算线索进入观测确认的过程。[23-2][23-3] 这一步通常以一次成功夜观浓缩呈现，但它实际包含多个不同判断。

首先，望远镜视野中出现的一个光点需要与星图比较。星图不是天空的完美复印，却提供了判断“这里是否已有已知恒星”的公共参照。其次，不在星图上并不自动等于新行星，可能是图的遗漏、记录问题或其他天体。进一步的位置变化与连续观测，才使它的行星身份获得更强支持。

理论预测与星图在这里承担互补任务。预测缩小搜索区域，星图帮助识别候选的不同，持续测量则将候选转化为可以计算轨道的对象。任何一环都不能独占“发现”的全部内容。只有预测没有有效搜索，可能停留于纸面；只有望远镜扫到光点而没有辨认，也可能留下未被理解的记录。

由此可见，观察并非一个完全被动的动作。观察者需要知道比较什么，怎样辨别差异，以及下一次测量如何安排。这个主动性不意味着结果可以由期待任意决定。恰恰相反，观察

程序之所以有价值，是因为它可以让期待失败。若候选始终保持恒星位置，或后续轨道明显不符合应有运动，它就不能继续凭预测的声望维持身份。

发现以后，研究还要回答这颗天体是否足以解释原来天王星的异常。找到一个新行星，与证明它在正确程度上承担所需摄动，仍然是两步。观测得到的新轨道和质量估计应回到动力学计算，重新检验旧残差。如果只庆祝找到了天体，却不再检查反演假说，就没有完成论证闭环。

这种闭环也可能修正最初预报。真实轨道要素与初始候选不同，并不一定意味着发现失败；它可能表明有限资料对某些参数约束较弱。重要的是，修正必须提高整体一致性，而且能够支持新的独立观测。预测的局部成功与参数的后续改进可以同时成立，没有必要为了维护英雄叙事而假装最初已经全对。

荣誉争论则使档案问题进入科学史。亚当斯、勒维耶、伽勒、达雷斯特和参与搜索的其他研究者承担不同工作，私人计算、公开发表、有效传递与实际确认不能简单等同。皇家天文学会对这一历史的介绍，也强调档案与贡献认定曾受到后来的叙事影响。[23-1] 公平的写法应说明贡献类型，而不是只排列一个唯一发现者名次。

这里还可以比较“看见”与“认出”。某个天体早已进入观测记录，并不等于其行星身份当时已经发生。照片或笔记中的光点可以在后来获得新的意义，但不能因此把后来的理论理解全部倒加给早期记录者。科学对象的发生包含感知、记录、比较与判定之间的联接，而这些环节不必在同一人、同一天完成。

对 SDE 而言，这个案例比简单的“理论指导实践”更具体。显露不是一个单独的新光点，而是能够在预测位置、星图差异、连续运动和扰动解释之间维持一致的对象身份。差异运行不是泛泛的观察，而是这些操作按照问题要求接续进行。特征纠缠也不是周围条件，而是每项属性获得意义时对其他属性的依赖。

科学发现的和谐性由此具有一种严格含义。计算者与观测者不需要使用相同方法，却需要让中间结果能够互相约束；新行星身份不要求所有预测参数不变，却要求修正后的对象能够与多条证据相容。多样性不是随意性，统一性也不是强制一致。它们在检验过程中相互限定。

23.4 成功怎样创造新难题

海王星的发现增强了天文学家对摄动分析的信心，却也产生了一个危险的诱惑：既然一次残差揭示了一颗新行星，其他残差是否也应如此处理？水星近日点的异常进动后来成为著名对照。寻找水星轨道以内的“火神星”曾被视为可能方案，但没有形成与海王星相同的稳定确认。

一九一五年，爱因斯坦用广义相对论处理水星近日点问题，给出了不同性质的解释。[23-4] 本章不展开相对论推导，只保留比较所需的关键区别：一处异常最终支持了新增天体，另一处异常的成功解释涉及引力理论的改变。同样以残差为起点，研究可以走向不同层次的修正。

从今天回看，火神星常被写成一个愚蠢的错误。然而在牛顿动力学已经多次成功的条件下，先检查未计入的物质分布并非毫无根据。错误不在于提出候选，而在于如果候选长期不能承担独立检验，仍把它当作已经确认的事实。合理性需要按当时可用证据评价，可靠性则需要后续检验不断约束。

这对方法论构成了一个重要反例。不能把“从异常推断隐藏对象”写成普遍发现公式，也不能因为某次失败就否定所有间接推断。真正需要建立的是分支判断：何时优先查仪器，何时调整已知参数，何时增加实体，何时改变基本理论。不同分支之间必须通过新证据竞争，而不是由一个预先决定的偏好永久统治。

海王星本身也继续成为问题。发现一个行星并没有马上给出其精确质量、完整轨道和长期摄动影响。相关参数需要持续观测，也需要与其他天体相互作用结合。新对象进入体系以后，既帮助解释旧差异，也为整个体系增加新的未知量。成功扩大了解释范围，同时扩大了必须管理的复杂性。

因此，新知识的回写不只是“更加相信牛顿”。它改变了具体工作：天王星运动要重新计算，海王星的位置要进入星历，远行星参数要接受长期约束，旧资料需要重新检查。理论声望是一个社会后果，参数、目录与观测计划的改变则是更直接可追踪的研究后果。两者应当分别叙述。

在构成意义上，太阳系也变了。不是说一八四六年以前没有海王星，而是当时的科学太阳系从已知行星及其异常，转变成包含新对象、更多作用关系和新观测任务的整体。这个整体能够做以前做不到的计算，提出以前没有同等明确形式的问题。SDE 所谓新的科学世界，应当落实在这种能力与对象关系的同时改变之中。

一个必要的反向检查是：如果删去 SDE 术语，摄动理论与发现史是否已经解释了一切？对轨道计算的具体正确性，它们确实提供了核心解释，SDE 没有资格抢占这一成就。本书所增加的组织视角，是把残差的构成、候选身份的阶段性、发现确认的多环节以及成功回写的分支放到同一分析中，并持续检查它们能否彼此衔接。

这种视角必须允许零增量。若只是把残差叫 D、候选叫 S、其余所有东西叫 E，而不能指出新的判断差异，那么方法没有获得独立解释力。本章的实质区分在于：预测位置不等于完整轨道，发现天体不等于完全验证原始反演，理论成功不等于新增实体策略永远正确。这些区别决定我们怎样使用历史成功，而非只决定怎样为它命名。

天空中的新对象由此成为一面双向镜子。它显示理论能够越过直接视觉的边界，也显示理论必须重新接受独立观测的约束。海王星的光点之所以成为思想史中的重大事件，不仅因为它增加了太阳系成员，而且因为它把一种研究能力公开地展示出来：从已有运动关系中形成候选，再让天空决定候选能否站住。

这种能力的成熟，要求成功之后仍保留失败的可能。若每一次没有找到都能被解释为“藏得更深”，候选便会失去科学内容；若每一次初始参数修正都被宣布为彻底失败，研究又无法发展。可检验的边界、明确的修正责任和对不同结局的容纳，共同决定了这一方法的生命。

本章参考文献

[23-1] Bowler, S. Finding Neptune；皇家天文学会对 D. Krajnovic海王星发现史研究的介绍与论文入口，2016。https://ras.ac.uk/finding-neptune

[23-2] Account of the Discovery of the Planet of Le Verrier at Berlin. Monthly Notices of the Royal Astronomical Society 7, 153, 1846. https://academic.oup.com/mnras/article/7/9/153/983971

[23-3] Observations of Le Verrier's Planet, In the Meridian. Monthly Notices of the Royal Astronomical Society 7, 154–157, 1846. https://academic.oup.com/mnras/article/7/9/154/983984

[23-4] Einstein, A. Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie, 1915；CERN 原文与译文馆藏。https://cds.cern.ch/record/632319

[23-5] Laskar, J. Des premiers travaux de Le Verrier à la découverte de Neptune, 2017，原始文献索引与数理历史分析。https://comptes-rendus.academie-sciences.fr/physique/articles/10.1016/j.crhy.2017.10.011/

第 24 章 恒星光行差、视差与“地球会动”的证据

24.1 光行差不是视差

同一颗恒星在一年中出现微小方向变化，可以有不同原因。观察者的位置改变，会产生视差；观察者的速度改变，在光速有限的条件下，会产生光行差。两种效应都与地球年度运动有关，也都能形成周期图样，却不是同一件事。若只把它们并列成“地球运动的证明”，便会忽略各自需要怎样测量，以及为什么其中一种先被可靠识别。

布拉德雷在一七二八年向皇家学会报告恒星的新发现运动，解释了年度光行差。[24-1] 研究的重要性不只是取得了一个支持地球公转的结果，还在于预期中的视差研究暴露了不符合原先相位预期的变化。研究者必须比较变化何时达到最大、方向如何随时间移动，才能判断它究竟属于哪一种效应。

现代教学常以雨滴和移动的伞解释光行差：雨相对地面竖直下落，行走者却需要倾斜伞。这个比喻有助于理解观察者运动影响接收方向，却不是光传播理论的完整替代。在现代处理中，光行差需要按照光传播与参考系转换理解；不能把机械雨滴图像的所有细节照搬到光。

视差则来自不同观察位置的几何关系。伸直手臂看手指，分别闭上左右眼，手指相对远处背景似乎移动；背景越远，相对变化通常越小。把这一几何关系移到恒星，地球轨道提供基线，恒星相对背景的年度方向变化提供角度。与光行差不同，视差幅度直接依赖目标距离。

这一区别解释了历史上的一个反直觉结果：支持地球年度运动的明显光行差可以先被识别，而恒星视差仍然难以测出。恒星距离很远，几何角位移极小；光行差主要取决于观察者速度与光速的比例，并不会因为恒星继续远离而按同样方式趋近于零。没有测到某颗恒星视差，不能由此否定已测得的光行差。

证据的力量也不是脱离理论条件的。光行差支持地球运动，需要相应的光传播解释；视差支持几何距离，需要基线、参考方向和误差控制。称它们为证据，不等于说没有任何辅助条件。科学中更有力的支持，常来自不同操作链在各自条件下相互吻合，而不是某一观察完全摆脱解释。

在 SDE 语言中，光行差与视差对应不同的差异序列和特征纠缠。前者关心速度、接收方向与光传播，后者关心位置基线、目标距离与参考背景。若只因结果都指向地球运动，就把它们归为同一操作，方法反而会降低分辨力。统一结论可以由多样机制支持，而这种多样性正是可靠性的来源之一。

24.2 测量微小角的困难

几何道理简单，不代表观测容易。恒星视差要求在不同时节辨认极小的角距离变化，仪器却可能随温度伸缩，机械部件可能存在偏差，光经过大气还会折射。目标的年度信号与许多仪器、环境变化具有相近的时间结构，使“多看几次取平均”无法自动解决全部困难。

随机误差可以在适当条件下通过重复观测减小，系统误差却可能在每次观测中朝相似方向作用。尤其当夏季与冬季温度不同，而视差也需要比较不同季节时，温度影响就可能与目标信号混在一起。若未建立区分方法，重复越多，研究者可能只是越精确地重复同一种偏差。

贝塞尔选择天鹅座六十一星，并使用夫琅禾费制造的量日仪进行相对角距测量。一八三八年发表的结果成为早期可靠恒星视差的重要成果。与此同时，斯特鲁维对织女星、亨德森对半人马座阿尔法星的研究也属于这一时期的关键工作。观测、计算与发表的先后需要分别处理，不能把整个过程缩成某个人在某一天首次想到三角测量。[24-2]

现代对这些早期资料的重新分析表明，仪器温度改正和误差估计确实影响所得视差；研究者可以大体重现早期结果，同时发现一些原来不易识别的系统性。[24-2] 这一事实有双重意义：历史测量不是毫无根据的猜测，但其可靠性也不应靠英雄地位保护。原始数据仍然能够接受新的问题与计算。

量日仪的意义在于，它使相对角距离的精密比较成为可行操作。研究不必直接把恒星在整个天球上的绝对位置全部确定到同等精度，而可以选择附近参照，测量角距离随时间的变化。相对测量降低了某些共同误差，却也引入了对参考对象的依赖。测量优势与新的条件是一同出现的。

仪器不是透明通道。它把目标变化转换为可读取的位移，而转换关系要经过校准。螺旋读数、成像结构和温度响应共同影响最后的角度值。若不同部分不相互协调，所谓“测到一个角”就可能只是读取一个未经充分解释的刻度。精密天文学的重要工作，是把这些转换关系变成可研究的对象。

这里可以看到三方程的具体往返。微小角变化作为候选 S 出现，促使研究者重新检查仪器与季节条件 E；新的改正改变下一轮观测比较 D；重复得到的稳定变化又使恒星距离成为可讨论的属性。过程中的每次回写都可能修改结论，不只是提高小数位数。

对研究动力而言，困难不是外部阻碍那么简单。正是视差目标的微小，迫使天文学把过去可以忽略的误差提升为研究对象。仪器热变形、参照星运动和长期记录的处理因而得到新的重要性。要求更精确的天空，同时生产了更精细的仪器世界与观测者世界。

24.3 选择参照星的代价

在相对测量中，背景星常被近似视为遥远而稳定。但“足够远”与“完全不动”不同。背景星也有视差，也可能具有自行；若目标星和参考星的距离差不够大，测得的相对年度变化就不能直接当成目标的绝对视差。参照不是从自然中随手拿来的一块静止幕布，而是一项需要说明的近似。

选择天鹅座六十一星，与其显著自行有关。较大自行可以提示目标可能较近，因此值得投入视差观测；但自行也取决于恒星横向速度，不能把“大自行”当成“必然近”的定理。选样策略利用的是有根据的概率线索，而不是已经知道答案的确定规则。

这种选样在方法上十分重要。如果从所有恒星中完全随机抽取目标，早期仪器可能长时间测不到足够大的视差。选择更可能靠近的恒星，提高了成功机会。但成功样本因此不是恒星总体的无偏代表。后来研究恒星距离分布时，需要记得这一选择过程，不能把最早被测出的恒星当成普遍样本。

背景星自身的运动也会影响结果。观测需要把随时间近似线性变化的自行，与一年周期的视差区分开来；若观测跨度短、季节覆盖不均，两者就可能互相混淆。更复杂的双星运动还会增加额外变化。因而，测量一个年度效应，往往需要超过一年并具有恰当时相分布的数据，而不只是某两天的精确比较。

可以用一个现代说明式表示这种责任分解：观测方向由基准位置、自行项、年度视差项及其他需要计入的项共同组成，再加上测量误差。这不是贝塞尔论文的原式，而是帮助读者看见各部分承担不同变化形式。模型中的每一项都应有自己的物理或观测理由，不能只因为拟合变好就无限添加。

参考星的选择还涉及一个看似微小却重要的问题：天空上靠近，不代表空间中靠近。两颗星在视野中相邻，可能相隔极远。早期相对测量正是利用这种可能性，在较小视场内比较目标与更远背景，从而减少某些共同影响。但若两颗星实际上属于同一物理系统，预期的背景性质可能不再成立。

这里的 E 具有清楚的相互规定结构。目标与参照的距离关系决定相对视差，视场大小影响共同大气与仪器误差，观测季节决定信号相位，而所采用的运动模型决定怎样从序列中分离各项。没有哪一项可以独立宣称“我已经测得真正距离”。最后的距离属性是在这些限制共同有效时获得的。

从六路径看，若从已经公布的距离数值进入研究，容易把参考系统当作无关的技术附件；若从仪器和参照选择进入，则会看到距离判断的条件。这一换路会改变对误差的评价：

小的随机分散不再足以说明准确，因为所有测量可能共享一个错误的背景假设。路径转换因此使一个原本被信赖的数字重新获得可检查的来历。

参照依赖也不意味着所有距离都是任意的。不同参照、不同仪器与不同处理方法之间的比较，能够逐步限制共同错误。可靠性并不来自完全没有中介，而来自中介足够明确，能够被替换、交叉检验和改进。精密天文学的成熟，很大程度上就是这种可替换性不断提高的过程。

两种年度效应还可以通过相位理解。视差主要随观察者相对基准的位置改变；光行差主要随观察者速度方向改变。在近似圆轨道中，位置方向与速度方向相差四分之一周期。因此，即使两者都表现为一年重复，极值出现的时刻一般不同。年度性本身不能区分它们，时间序列的形状和相位才提供进一步信息。

这是一个重要的观察设计原则：只在同一季节每年测量，可能很适合研究某些长期自行，却不利于完整分离年度信号。若要测视差，应覆盖足以显示相位差的时段；若目标在关键季节太靠近太阳，实际覆盖又会受限制。天文测量不能随意安排所有实验条件，观测几何本身决定可获得的信息。

相对测量的优势可以用误差分解说明。假设目标与附近参照星的读数都带有同一小的仪器零点偏差，两者相减时这项共同偏差可以部分抵消。但若仪器误差随视场位置变化，或两星颜色不同导致传播与成像影响不同，相减就不能全部消除。所谓“差分可以消误差”，只对足够共同的那部分成立。

这使参照选择具有多重要求。角距离过远，共同误差近似较差；过近又可能难以区分或存在物理关联。亮度差过大，成像与读数质量可能不对等；可用参照太少，则难以检查某一颗参照的特殊运动。选择不是寻找完美背景，而是在实际星场中尽量建立可检查的比较网络。

设想一个教学例子：目标的真实视差为某个正值，参照星也有较小正视差，简单差分所得值便小于目标绝对视差。若直接取其倒数，距离会被高估。这个例子不指称贝塞尔某一具体误差，而说明为什么背景星“足够远”的近似需要评估。几何推导完全正确，也可能因为输入量身份误认而产生错误结果。

早期资料的重新分析具有类似作用。它不只是用更强计算机把旧数字算得更快，而是重新检查哪些效应可以从有限观测中分离。如果温度与视差项在样本中高度相关，增加一个改正参数未必立刻解决问题，反而可能暴露原资料对两者区分不足。诚实的结果有时是扩大不确定范围，而非给出更多小数。

参照体系还需要时间延续。同一颗参照星如果在多年中有可测自行，就不能把最初位置永远当作固定值；但若任由每年零点自由变化，又可能吸收目标真正的周期信号。研究者必

须同时约束周期项、长期趋势和仪器零点。若观测只覆盖一年中的相似季节，长期趋势与周期局部片段还可能互相混淆。延长年限因此只有结合适当季节分布才真正增加辨别能力。年数本身不是信息充分的保证，观测在周期中占据哪些位置同样重要。

24.4 从角到距离

一旦年度视差得到可靠分离，几何关系就可以把角度转变为距离。在现代单位定义下，若视差以角秒表示，距离以秒差距表示，二者近似满足倒数关系。这一简洁写法依赖小角近似和规定的基线。它是成熟单位体系的表达，不能直接当作早期研究者未经转换的原始记录。

从角到距离意味着恒星获得了一种新的可比较属性。此前，两颗星的表观亮度差可能来自发光能力不同，也可能来自距离不同；知道距离以后，便可以逐步区分这两者。本章停留于几何测距本身，但它已经打开通向恒星本征光度的问题。测距不是给星表添上一列孤立数字，而是改变其他属性的解释可能。

这种改变具有本体论意义上的具体内容。肉眼天空中的星点主要由方向、亮度与颜色组织；几何距离进入以后，星点被重新组织为具有空间深度、相对运动和不同发光能力的对象。科学世界的结构发生变化，并非只是观察者在同一张平面图上知道得更多。新的空间属性允许新的操作与新的推断。

然而，角度倒数变成距离时，误差也会以非线性方式转换。视差很小时，同样大小的角度不确定性可能造成很大的距离不确定性。若测得的视差与零难以区分，不能因为可以按计算器倒数就宣称得到了精确远距离。精密测量与恰当统计处理必须共同限定结果的用途。

同样，未测出视差也可以有信息意义。如果仪器足以排除某一较大的角位移，就能够对目标距离给出下限。但这种下限与成功测得一个正视差不同。将“不比某个范围近”写成“已经知道它有多远”，会夸大证据。科学史应当保留这种较弱结论的价值，而不是只记录最后的成功数值。

早期视差研究还说明证据并不总按概念难度排序。几何推理在很早就可以提出，可靠测量却等待了很久。理论简单与实践容易之间没有必然联系。仪器制造、长期观测、误差模型和目标选择的协调，决定了一项已经可以理解的操作何时能够真正实现。

ESA 关于早期天鹅座六十一星观测的回顾，将贝塞尔的方法与后来更大规模的视差测量联系起来。[24-3] 这种连续性应理解为问题结构的延续，而非仪器条件始终相同。现代空间天体测量已经改变了采样、参考框架和误差处理，但仍须回答相似的根本问题：怎样把方向变化中的几何信号与其他变化分开。

因此，“地球会动”的证据并不是一块最终放上桌面的石头。它由不同操作链不断获得更清楚的形式。光行差使速度与接收方向的关系显现，视差使位置基线与恒星距离的关系显现。两者相互支持，也各自保留局限。真正的统一来自这种可区别的多重联系，而不来自把所有变化都叫作同一证明。

天空的深度就在这些微小角差中逐渐成为公共知识。它不再只是推测恒星“非常遥远”，而是能够说明某一距离判断由哪些观测、参照和几何关系支持。新的深度世界既扩展了人的视野，也使人更加依赖精密协作：眼睛、仪器、数学与标准共同构成了远方能够被测量的条件。

本章参考文献

[24-1] Bradley, J. A Letter ... giving an Account of a new discovered Motion of the fix'd Stars. Philosophical Transactions 35, 1728. https://royalsocietypublishing.org/rstl/article/35/406/637/25243/IV-A-letter-from-the-Reverend-Mr-James-Bradley

[24-2] Reid, M. J.; Menten, K. M. The First Stellar Parallaxes Revisited, 2020；包含贝塞尔、斯特鲁维、亨德森原观测的重新分析。https://arxiv.org/abs/2009.11913

[24-3] ESA Gaia. Image of the Week, 15 January 2025，天鹅座六十一星早期视差资料说明。https://www.cosmos.esa.int/web/gaia/iow_20250115

第 25 章 标准天球、时间与测量误差的制度化

25.1 一个“位置值”的复杂来历

星表中的一个位置值看起来只是两项数字。它们排列整齐，可以被抄录、计算和比较，仿佛恒星天生就带着赤经与赤纬。然而，这个位置属于哪个时刻、哪一种参考框架、怎样的观测条件，却决定了数字是否能够正确使用。删去这些条件，数值的简洁反而会掩盖意义。

观测者首先在某地、某时接收到光。望远镜的指向依赖仪器安装与刻度，光的方向受到大气折射影响，时间由实际时钟读出，地球本身又在转动和绕日运动。若要把这一具体观测转变为可与别处资料比较的位置，就必须说明哪些变化被保留为研究对象，哪些变化被换算到共同条件。

例如，恒星经过子午线的时刻可以帮助确定赤经，也可以用已知恒星来检查时钟。这种相互使用很有力量，却也带来循环风险。若完全依靠某星表校准时钟，再以同一校准后的观测证明该星表无误，结论就没有获得独立支持。相互约束的系统需要多种资料和明确的求解安排，而不是让各项误差相互掩护。

位置的“历元”与参考方向的规定也需要区分。恒星会有自行，必须知道其位置对应哪个时刻；坐标轴本身的方向也会随采用的标准而不同。在历史星表中，相关约定并不总以今天相同的术语出现，因此研究者必须阅读表的说明，不能把现代习惯直接投射过去。

贝塞尔一八一八年出版《天文学基础》，以布拉德雷在格林尼治多年观测为材料进行系统处理。[25-1] 书名与材料来历已经提示一个重要事实：观测结束多年以后，旧资料仍可因新的归算工作取得新的公共价值。保存数字只是第一步，重建它们的仪器条件和比较标准，才能使它们成为长期研究的基础。

这种知识生产不同于发现一颗醒目的新天体。它往往表现为旧数值被改正、资料被重新组合、同一批星获得更一致的位置。若只寻找思想史中的戏剧性时刻，这类工作容易被当作技术整理。但若没有它，恒星自行、微小视差和轨道残差的许多判断就失去了可靠背景。

本书将这种标准天球理解为一个共同操作世界。它不是悬挂在宇宙外面的绝对网格，而是通过观测、计算和约定形成的可比较结构。这个结构必须受到天体与仪器的实际行为限制，不能由约定任意改变；同时，它又需要人们明确选择共同表达方式，才可能跨地点使用。

因此，客观性并不意味着测量者消失。测量者要把自己的位置、仪器与操作方式纳入可说明的过程，使别人能够理解和复查。主体的介入并非总是污染，也可以通过规范化成为可靠性的条件。关键在于介入是否被公开、量化并受到检查，而不是是否假装它不存在。

25.2 改正程序怎样形成

进动、章动、光行差和大气折射常被共同列为位置改正，但它们的来源不同。进动和章动涉及地球自转轴方向变化，光行差涉及观察者运动与光传播，大气折射则属于光通过地球大气时的传播问题。把它们放在同一计算流程中，不等于它们通过同一种证据成立。

布拉德雷关于章动的研究依赖长期观测中不同恒星方向变化的比较。其公开书信说明了如何通过近乎相对方向的恒星之间的变化，判断某种共同效应是否可能来自地轴。[25-2] 这里的思想力量在于，不只追问一颗星为什么动，也追问许多星的变化是否具有参考轴运动所应产生的共同形式。

这是把单个对象的问题改写为系统问题。如果若干恒星同时呈现一种有规则的整体变化，逐一给每颗星增加独立运动未必是最好的解释；检查坐标轴和观察者状态，可能更有效。但也不能反过来把一切共同变化都归于参考系统。需要比较不同方向和时间的预期关系，判断哪种解释真正受到约束。

大气折射则要求注意恒星高度及观测环境。接近地平线时，光线穿过大气的路径条件更复杂；同一仪器在不同温度和气压下也不能无条件采用完全相同改正。在较粗精度下可以忽略的差异，到精密测量中会成为重要限制。一个通用改正表的便利，始终伴随其有效范围。

“改正”这个词容易使人误以为只有原始读数是真实的，后面的计算只是人为修饰。实际上，原始读数已经属于仪器转换后的结果；而改正也不能自动获得真实性。两者都需要解释其与研究量的关系。可靠工作不是坚持未经处理的数字，也不是相信所有处理，而是让处理步骤具有可检查的依据。

现代说明中，可以把位置观测写成由天体方向、地球状态、传播条件和仪器响应共同生成的测量关系。归算是在已知或估计这些条件时，求得指定标准下的位置。这个表达是教学性的整体描述，不是历史星表所用的同一公式。它突出一个事实：不同改正并非任意加减的一串数，而必须在共同坐标和时刻定义下才有意义。

若各项改正分别采用不一致的常数，形式上每一步似乎正确，组合起来却可能产生偏差。若一个项已经包含了某种效应，后续又重复扣除，就会过度改正。因而，标准化不仅要求提高每项精度，还要求保持整个计算体系的内在一致性。和谐性在这里具有明确的数学责任，而非审美上的整齐。

新的改正规则还能改变旧资料。一个原先被看成恒星运动的长期趋势，在改用更合适的参考框架后可能减弱；反之，去除共同效应以后，某颗星的特殊运动也可能更清楚。S 的变化促使 E 重组，E 重组再改变 D 中可辨的差异。这样的回写能够生产新发现，也能够撤销过去自以为稳定的发现。

但应避免用同一资料既确定改正又独立验证改正。若一组恒星用于求出某个系统偏差，那么它们归算后的平均一致，部分是求解方式保证的。要评价规则能否推广，还需要其他恒星、其他时段或其他观测链。这是统计上的问题，也是科学史中评价“标准获得成功”的重要条件。

25.3 跨台站标准化

两个天文台测量同一颗星，数值不同并不必然意味着一方粗心。它们可能使用不同星表、不同仪器零点、不同时间系统，甚至采用不同的观测者反应习惯。若这些差别不被辨认，把数值简单平均，可能只得到一个没有明确物理意义的折中结果。

格林尼治关于金尼布鲁克观测的历史记录，是人的反应进入精密天文学的著名案例。马斯基林发现助手记录的过中天时刻与自己存在系统差别，并在报告中作出评价。[25-3] 本章不替历史人物虚构动机，也不把后来形成的全部“人差”理论倒置到当时。重要的是，人与仪器共同产生读数，而人与人之间的稳定差别可能达到科学任务不能忽略的程度。

一种早期观测方式需要眼睛判断星像经过标线，同时耳朵听取钟的节拍，再估计时刻。视觉与听觉的协调并非对所有人完全相同。如果只把差别当作不认真，就可能错过系统研究；若承认它是可测的操作差异，就可以设计对比、求取相对人差，并改进记录方式。观测者在这里由外部裁判转变成测量系统的一部分。

但把人差纳入校正，也不能保证每个人的差别永久固定。亮度、疲劳、观测方法和仪器变化都可能影响反应。固定常数有时是有效近似，有时会掩盖变化。良好的制度既需要可执行的简化，也需要在条件改变时重新打开简化。这与任何成熟工具的封装相似：使用者可以依赖稳定程序，但系统必须保留发现失效的能力。

跨台站时间传递使共同观测获得新的可能。格林尼治在十九世纪中叶通过电报网络传递时间信号，相关原始年度报告记录了时钟、电气设备和时间球的配合。[25-4] 时间从本地天文观测结果，转变成能够被远处同时使用的公共基础设施。航海、铁路和城市生活也因此与天文学的精密操作发生更紧密联系。

时间信号仍然有传递过程。发送、线路和接收并非没有延迟，设备故障与钟差也可能影响结果。把各地时钟调到共同显示，并不等于完成了任何精度要求下的同步。标准传播得越广，越需要说明它的准确程度与监测办法，否则同一个错误也可能更广泛地被复制。

星表之间的比较具有类似问题。两份目录如果都使用同一基础星表或同一改正常数，它们的符合不能全部算作独立证据。共享来源可以帮助统一，也可以传播共享误差。研究者需要分辨一致来自独立观察，还是来自相同归算规则。数量上的多份文件，并不自动等于认识上的多条独立支撑。

十九世纪末的基本星表与天文常数协调，显示这种工作已经具有国际组织层面。纽康的基本星表和相关出版物，把不同历元、长期观测与统一系统的问题放在明确的计算计划中。[25-5] 共同标准需要选择，但选择之后仍应允许根据新资料修订。制度的稳定性与科学的可修正性必须共同维持。

从三原理看，跨台站的差异首先造成张力；统一规则使差异能够被比较；后来的更高精度又暴露规则残留的问题，推动再组织。这里没有一次永久完成的和谐。和谐更接近一种能力：不同台站可以保留各自仪器与观测特色，同时让结果在明确条件下接合，并能够定位不一致究竟出现在哪一环。

共享误差可以用一个完全假设的数值例子显示。设三个台站的独立随机误差各约为一个单位，但它们都采用了带有两个单位偏差的同一参考表。三者之间可能彼此相当一致，平均以后随机部分也会缩小，可是共同的两个单位偏差不会因此消失。若只按台站数量计算可信度，就会把复制同一标准当成新增独立证据。

要发现这类问题，需要改变某个共同环节。例如使用不同基础资料、不同观测方法，或寻找不依赖同一参考表的校准。方法多样性在这里不是为了展示丰富，而是为了使错误不至于完全相关。一个精密系统可能恰恰需要保留较不同的测量链，即使其中某条单次精度暂时较低。

跨表归算还需要区分整体旋转与个别恒星运动。如果一个旧表的坐标方向相对新表有微小转动，许多星都会呈现有规律的位置差。若逐星解释为自行，就会制造一组假运动；若过度用整体旋转吸收差异，又可能抹掉真实的共同运动。解决需要足够广的天区分布和对模型自由度的约束，不能只在局部少量星上决定。

星表的历元差也不是越大越好。较长跨度使微小自行累积成较大位移，但旧表中的系统误差也可能更难重建。如果旧数据的参考框架不明，长时间并不能自动变成高质量信息。时间基线的价值要与资料可追溯性相乘，而不是单独按年数评价。

一种良好的归算说明应当使后来者知道如何恢复计算链：原始观测时间怎样换算，仪器零点怎样估计，哪些恒星承担基础参照，各项常数采用哪个版本，以及哪些观测因何被排除。这些说明不要求把每一步都塞进最终星表，但必须在相关记录中可查。否则，目录只能被信任，不能被有效检验。

对误差的组织还会影响观测分工。如果某一类误差主要来自仪器方向，那么改变观测位置或反转装置可能有帮助；如果主要来自时间尺度，增加相同仪器的星数未必解决；如果来自人差，则需要交叉观测和操作比较。不同限制需要不同动作，不能用统一的“增加数据”替代诊断。

这种诊断能力是精密科学的重要成果。它让失败不再只是一个令人失望的数字，而成为关于系统某环节的信息。与此同时，诊断也必须容许暂时无法定位原因。当多种解释都与资料相容时，最负责任的标准化不是强行消除差异，而是保留不确定性及所需下一步检查。

同一目录中的误差还可能依天区而变化。若某些偏南目标总是在较低高度观测，折射改正的限制就可能集中影响这一组，而非平均分散到全表。把全天误差只报告成一个数，会掩盖这种结构。检查残差随高度、赤纬、季节或观测者的变化，能够把一个笼统的精度问题拆成可调查的来源。跨台站比较的价值也由此更具体：不同纬度使同一目标具有不同过中天高度，某一台站难以分开的目标性质与低高度效应，便可能在另一台站获得不同配置。地点差异在这里成为检验资源，而不只是必须消除的麻烦。比较时仍须统一时标和目标标识。

25.4 精度的历史回写

提高精度常被想象成在同一个数后面增加小数位数，但天文学的精度史更深刻。原先可以忽略的系统效应会进入对象定义，原先被当作稳定的参考会成为研究目标，原先被宣布为特殊天体运动的变化可能被重新解释。精度改变的不只是数值细密程度，还包括什么能够被当作事实。

例如，当地轴变化被更准确地处理，星表中的长期位置差就可能获得不同解释；当仪器误差得到重新估计，旧的微小视差主张可能不再可靠；当时间传递得到改进，跨地点事件的先后关系可以被更严格地约束。这些变化并非纯粹外加修饰，而是研究对象的可判定属性发生了重组。

因此，标准有一种双重力量。它使共同研究成为可能，也容易把自己的前提隐藏起来。日常使用者看到的是经过处理的表值，很少看到它们依赖哪些原始观测、哪些常数和哪些取舍。当异常出现时，若标准被视为不可检查的地基，研究就可能把全部责任错误地推给天体。

保留版本与来历是抵抗这种风险的方法。一个星表的位置值应能追溯到所用历元和归算体系；一个新的修正应说明与旧版有何差别；一项跨表比较应指出是否共享原始资料。版本记录看似属于文献管理，实际上影响科学推断能否分辨独立支持与重复使用。

这里还涉及一种重要的时间结构。早期观测可能精度较低，却拥有长时间基线价值；现代观测可能单次精度高，却无法替代过去一个世纪的位置。资料价值并不只按仪器新旧排序。把旧数据放入合适框架，能够使长期变化显露；但这种重新使用必须保留旧数据的不确定性，不能因新算法而假装旧仪器拥有现代精度。

从 SDE 的角度，标准天球不是一个完成后静止不动的 E。它由过去的显露与差异运行形成，又规定后来的测量怎样被接纳。新发现也能反过来修改它。例如，某些原本适合作为稳

定参照的对象，若出现特殊运动，就需要重新评价其参考角色。对象与标准之间存在持续的相互规定。

这一分析与普通的“社会建构”论述也有需要澄清的地方。参考框架的约定性，不意味着恒星位置可以任意设定；制度选择的作用，不意味着数据只反映权力。共同标准在一定意义上由人设定，但它必须使不同观测能在物理关系中一致运作。其好坏可以通过预测、交叉比较和误差表现受到约束。

反过来，承认自然约束也不能删除制度。若缺乏长期资助、数据保存、统一出版和人员训练，许多细小效应根本不会成为公共研究对象。不是自然缺少变化，而是复合研究系统尚未具备让这种变化稳定显露的能力。天文学世界的发生因此同时涉及对象关系与实践组织，但两者的因果责任必须分别说明。

本章所呈现的统一性，是多种观测能够在共同定义下比较；多样性，是不同地点、仪器与方法保留独立检验的价值；和谐性，是一致与差异都能在可追溯的程序中被解释。如果只有统一而没有独立差异，系统可能共同犯错；如果只有差异而没有共同标准，资料又难以累积。二者的平衡，是标准化真正的科学成就。

从轨道动力学到标准天球，天空已经成为一个可以预测、定位和比较微小差别的世界。接下来的转变不再只问星在哪里，而要问星光包含什么。方向测量所建立的校准纪律不会因此失效，它将以新的形式进入光谱：读数的稳定、装置的响应、参考的来历，仍然决定远方能够显露哪些性质。

本章参考文献

[25-1] Bessel, F. W. Fundamenta astronomiae pro anno MDCCLV, 1818，ETH 数字原件。https://www.e-rara.ch/zut/content/titleinfo/265250

[25-2] Bradley, J. 关于恒星视运动和章动的历史书信，MacTutor 原文转录。https://mathshistory.st-andrews.ac.uk/Extras/Bradley_letter/

[25-3] Maskelyne's analysis of Kinnebrook as an observer and a statement of the observing method in use at Greenwich in 1796，历史报告摘录。https://www.royalobservatorygreenwich.org/articles.php?article=1131

[25-4] Extracts from the Reports of the Astronomer Royal relating to clocks, chronometers, the time-ball and longitude determinations (1836–1868). https://www.royalobservatorygreenwich.org/articles.php?article=1357

[25-5] U.S. Naval Observatory. Astronomical Papers，纽康基本星表等原始出版物目录。https://www.cnmoc.usff.navy.mil/Our-Commands/United-States-Naval-Observatory/The-James-M-Gilliss-Library/Publications-of-the-US-Naval-Observatory/Astronomical-Papers-of-the-US-Naval-Observatory/

第 06 编

光谱学与恒星物理的诞生

第 26 章 太阳光谱暗线：星光怎样变成可测特征

第 27 章 实验室与恒星之间：光谱化学的归属方法

第 28 章 哈佛摄影与恒星光谱分类

第 29 章 赫罗图：两种测量怎样生成恒星演化问题

第 30 章 恒星为什么发光：从辐射到核反应

第 26 章 太阳光谱暗线：星光怎样变成可测特征

26.1 颜色不能解决什么

太阳光经过棱镜形成颜色带，这一经验并不直接告诉人们太阳由什么组成。颜色带中由红到紫的连续变化，首先提出的是光自身与折射的关系。若要把这种现象用于精密研究，研究者还要回答：怎样指认其中同一个位置？“偏黄一些”“接近绿色”可以交流大致印象，却难以支撑不同玻璃、不同装置之间的细微比较。

沃拉斯顿在一八〇二年的折射与色散研究中报告了太阳光谱中的暗线。[26-1] 早期发现并没有立即建立完整的元素分析。线条可以被看到，却未必已经具有稳定的测量身份；研究者仍需判断它们是否来自光源、仪器，或颜色区域之间的边界。发现某种差异与理解其机制之间，可以存在很长的历史距离。

夫琅禾费的研究来自精密光学制造中的实际问题。要改进消色差望远镜，必须比较不同玻璃对不同颜色的折射。玻璃并不是对所有颜色都具有相同折射率；不同玻璃的色散关系也不只是按一个固定比例整体伸缩。仅用一个宽泛颜色名称作为比较点，难以满足精密制作的要求。

在这种任务中，太阳光谱暗线具有特殊用途。它们在连续颜色带中提供可复认的细小标记，使研究者能够在不同玻璃与仪器条件下指向对应光线。夫琅禾费一八一七年的论文把玻璃折射与色散测量、太阳谱线研究联系起来。[26-2] 光学制造没有只是等待纯理论给出答案，它自身的精度要求参与生成了新的观测单位。

这一转变可以从“颜色”与“位置”的差别理解。颜色经验通常把一定波段的综合刺激作为整体，谱线位置则要求在展开后的光谱中识别局部结构。两束整体颜色相近的光，细部谱线可能差别很大；一种物质在不同条件下发出的光，也可能使整体颜色发生变化。颜色没有因此失去价值，但它不能承担全部辨别任务。

仪器使新的差异得以显露，却并非任意造出它的内容。狭缝、透镜与棱镜把入射光转化为可观察的空间分布。若没有适当条件，细线可能被展宽、重叠或背景淹没；改善装置以后，原来未能区分的结构才成为稳定对象。这里既有光的约束，也有操作条件的构成作用。

本书将这视为一种天文学世界的改变：星光不再只是天空中某一点的亮度与颜色，它可以展开为内部有次序、有细节、可以逐项比较的差异系统。远方的光在新的视觉 SDE 中形成了新的特征聚合体。这个聚合体不是原子理论的现成图解，因为其化学与物理解释仍有待建立。

26.2 暗线的测量身份

狭缝在光谱仪中承担一个容易被忽略的角色。宽的入射开口会使不同位置的光谱相互覆盖，窄线因而不易分开；缩窄狭缝可以提高某些细节的可辨程度，却减少进入的光。这不是越窄越好的单项优化，而是亮度、成像质量和分辨能力之间的协调。

棱镜把不同波长的光偏转到不同方向，但其色散不是天然均匀的刻度。光谱图上两段相同的空间距离，不一定对应相同波长差。换一种玻璃或改变装置几何，图中的间距也会变化。因此，暗线在纸面上的位置与它所代表的物理光学量之间，必须建立转换关系。

对玻璃折射测量来说，关键是每次比较是否指向相同谱线。研究者可以用较强、容易复认的线作为固定标记，再测量其通过不同材料时的偏转。此时，谱线的用途不需要先知道元素名称。它首先提供了稳定比较的机会，让光学材料的差异不再依赖观察者对颜色边界的粗略判断。

夫琅禾费对太阳谱线的标记与绘图使这些细节能够被传递。字母标签的作用，并不是解释暗线原因，而是让不同观察者讨论同一位置。命名在这里具有操作意义：它降低指认歧义，同时把可复核的次序保留下来。若只把字母当成后来元素符号的前身，就会误解它们最初承担的任务。

量化还要求区分线的位置、宽度与深度。位置比较可以用于折射或后来波长标定；线宽可能受装置分辨率和物理条件共同影响；线深则依赖光强、背景及响应。早期视觉图谱不一定能以相同精度记录这三类属性。一个画得很清楚的位置，不能自动承担精确强度测量的证据责任。

衍射光栅提供了另一种把光谱与尺度联系起来的方法。以现代说明式表达，在适当入射与远场条件下，光栅间距、衍射级次、角度和波长之间存在确定关系。它使机械长度与角度测量可以共同约束波长。这种说明有助于理解从棱镜位置到波长尺度的转变，但不能把各时期装置的误差与可用分辨率全部抹平。

棱镜和光栅的对比还揭示一种有力的检验：如果某个结构只能在一台仪器中出现，而换装置就消失，需要优先检查装置效应；如果在不同光学安排中保持对应关系，其光源身份获得更强支持。所谓稳定并不意味着纸面位置原封不动，而是经过各自正确的转换后仍能对应。

这就是 E 作为特征纠缠的一个具体例子。光源结构、仪器色散和刻度关系共同决定读数。线位置的稳定依赖这些关系匹配，而不是依赖某一元素孤立地“发出标签”。S 一旦获得稳定性，又使 D 转向更精细的逐线比较；新的比较要求则推动仪器与标定改变。谱线成为工具的过程，同时也是光谱世界继续发生的过程。

棱镜测量的几何关系能够更清楚地说明标定的作用。在现代表示中，已知棱镜顶角并测得某条线的最小偏向角，可以求得该波长对应的折射率。这里同一块玻璃不是只有一个对所有颜色通用的折射率；不同谱线对应不同值。精密光学需要的是一组彼此对应的量，而不是一个笼统的“折光能力”。

若两种玻璃在某两条谱线之间具有适当关系，组合透镜可以减少不同颜色成像位置的差别。但多条谱线显示的色散关系可能不完全按同一比例变化，因此消除两处色差，并不保证整个颜色范围没有残余。这一困难解释了为什么增加稳定谱线标记具有制造价值：它让原先被宽泛颜色描述隐藏的差异能够被比较。

图谱中的线条因此同时具有两种身份。作为太阳光的结构，它是需要解释的现象；作为可复认的标记，它又成为研究玻璃的工具。工具与对象之间并非永久分工。同一条线在一个问题中提供标准，在另一个问题中却成为未知。知识的发展常常发生于这种角色转换，而不是只靠增加新对象。

当谱线被当作标准使用时，还需要保持某些比较条件。如果狭缝倾斜、成像失焦或光路不稳定，线的测得位置和宽度可能变化。增加色散能把邻线拉开，却也可能降低单位面积亮度。若目标太弱，理论上更高的分辨能力未必能在实际记录中实现。设计装置需要在光量与区分能力之间取得适合任务的平衡。

这种平衡也使太阳与恒星观测不对称。太阳光充足，可以支持较复杂光路和较细比较；遥远恒星通常更微弱，观察者必须接受不同的曝光或分辨率。把太阳图谱中的每一细节都当成恒星观测应立即可见的东西，会忽略信号条件。共同原理并不保证共同资料质量。

夫琅禾费研究的思想意义因此在于形成了一种可迁移的方法，而非保证所有光源都能以同等精度测量。方法可以迁移，具体参数与限制仍要重新确定。正是这种有条件的迁移，后来使光谱成为连接光学制造、实验物理与天文学的共同工作领域。

26.3 从图谱到新问题

太阳图谱出现以后，一个新的比较空间被打开：实验室火焰、太阳和恒星是否具有对应结构？这不是把三幅图放在一起看起来像不像那么简单。不同光源的亮度、连续背景、线条明暗与仪器响应可能不同，研究者必须创造共同的比较条件。

夫琅禾费已经注意到人工光源亮线与太阳暗线之间某些位置的对应。后来的基尔霍夫研究把这一问题推进到吸收与发射之间的关系。[26-3] 对本章而言，关键是保留历史上的先后层次：最初稳定的是线的位置及相互次序，后来才逐步建立这些位置为何能够支持物质判断的理论与实验。

太阳暗线还并非全部都可以直接归于太阳。光在到达地面仪器前穿过地球大气，传播路径本身能够留下吸收结构。如果同一条线随太阳高度或大气条件改变，其来源判断就需要重新审查。比较不同时刻、不同方向或不同条件的光谱，可以帮助区分源端与传播端的贡献。

这种来源分解为天文学增加了新的对象层次。研究者看到的不只是“太阳的光谱”，而是太阳辐射、大气传播和仪器响应共同形成的记录。要研究太阳，就必须研究如何把其他作用从记录中辨认出来。中介不再只是通道，它也成为可以测量和建模的组成部分。

图谱的制作又带来一个选择问题。较强线容易进入图表，较弱线可能被忽略；相邻线在分辨率不足时可能被画成一条，装置改进后才分开。因而，不同图谱之间的差异未必意味着太阳变化，也可能意味着记录粒度改变。若不保留装置和绘图条件，后来的比较会误读这些差别。

这种选择不能简单称作缺陷。任何图谱都必须在可见性与表达能力范围内组织资料。问题在于，使用者是否知道图谱保留了什么。对于测量折射，几条强线可能已经足够；对于元素识别，更多细部对应可能必要；对于丰度研究，仅有位置图又远远不够。同一材料在不同任务中具有不同承重能力。

光谱图因此不是一张越详细越好的装饰画，而是根据研究任务组织差异的工具。它的精细化产生新问题：哪些线属于同一种物质，哪些是重叠，哪些随条件变化，哪些在其他恒星中消失或增强。新的 S 没有使问题结束，而是创造了更有区分力的 D。

这种变化也解释了为何天文学能够逐步突破“遥远天体不可取样”的限制。光不是带回一块恒星物质，却可以携带受物质与状态约束的结构。如果这些约束在地面实验中得到研究，远方光谱就可能支持超越位置与亮度的判断。不过，能够提出这种可能，与已经可靠知道恒星成分，仍然相隔一整套论证。

线的位置还可以承担仪器复现的任务。研究者更换玻璃、调整棱镜或比较两台仪器时，可以检查同一组线之间的间隔如何改变，而不必只依赖肉眼对某种蓝色是否相同的判断。然而，若仪器分辨能力不同，旧图中的一条线可能在新图中分成两条或更多；这时旧名称所指的测量标记就需要重新说明。谱图的持续使用既依靠标记的稳定，也依靠允许标记被细分。档案图谱因而不是越新越能使旧图无用：旧图仍记录当时仪器能分出的层次，新图则解释旧标记为何具有那样的形状与位置。记录狭缝宽度因此有助于后来的图谱比较。

26.4 早期解释不可后见

今天我们可以用原子能级和跃迁解释许多谱线，但不能把这一知识直接安放在十九世纪早期观察者的头脑中。夫琅禾费测量线条，并不等于他已经理解量子态；沃拉斯顿看见暗

线，也不意味着当时已经形成成熟恒星大气理论。历史上可靠特征的建立，可以先于后来解释它的理论。

这种先后关系对理解科学创新很重要。研究者不必先知道全部原因，才有资格稳定地测量一个现象。可复认的差异能够成为下一阶段解释的基础。反过来，后来出现的理论也可能改变早期特征的分组：原先看似一种线型，后来可能由不同物理过程产生；原先分开的表现，又可能被解释为同一元素在不同状态下的结果。

因而，不能把谱线史写成原子理论一步步揭开预先贴好的元素条码。线条的可见性依赖物理状态，观测记录依赖仪器，化学归属依赖多线匹配与实验支持。若把所有中介删去，所谓“星光告诉我们成分”就会变成一句遮蔽方法的比喻。

本章所说明的第一步，是一种新的差异单位如何成立。它必须能够被指认、重复测量和跨装置对应；其稳定性并非只属于观察者，也非只属于光源，而属于两者通过特定操作形成的复合关系。这个整体性判断不削弱光源的物理约束，反而要求明确约束怎样到达最后的读数。

还可以从失败情形检验这一解释。如果线的位置随更换装置而无规则改变，首先不能把它当作稳定光源特征；如果只有在某种大气条件下出现，就需要检查传播条件；如果不同观察者无法在图谱中确定对应位置，公共知识尚未充分建立。这些失败并非无价值，它们说明稳定特征需要哪些条件。

从六路径看，先拿着元素名称寻找谱线，容易把符合当作确认；从谱线如何被测得进入，则会先检查它是否是真实可复认的结构。后一条路径并不反对化学解释，而是为它建立合适的起点。若测量身份尚未稳定，赋予物理意义就可能过早。

精密光学、谱线标记和后来的化学分析具有不同责任，却在同一个发展过程中相互连接。制造望远镜的任务推动更精密的颜色比较，暗线提供新的比较标准，新标准又使遥远光源与实验光源的对照成为可能。统一性发生在这些环节能够接续，多样性则保留在每个环节不同的对象与方法中。

到这里，太阳光已经不只是照亮事物的媒介，它自身成为了具有内部结构的研究对象。天文学家获得的不是一幅更绚丽的彩虹，而是一种可操作的差异秩序。正是这一秩序，使下一步关于物质组成的判断能够不再停留于遥远天体的想象，而开始进入实验与观测共同约束的领域。

本章参考文献

[26-1] Wollaston, W. H. A Method of Examining Refractive and Dispersive Powers, by Prismatic Reflection, 1802；皇家学会原稿馆藏。https://makingscience.royalsociety.org/items/l-and-p_12_30/

paper-method-of-examining-refractive-and-dispersive-powers-by-prismatic-reflection-by-william-hyde-wollaston

[26-2] Fraunhofer, J. Bestimmung des Brechungs- und Farbenzerstreuungs-Vermögens verschiedener Glasarten, in Bezug auf die Vervollkommnung achromatischer Fernröhre, 1817. https://www.e-rara.ch/zut/doi/10.3931/e-rara-3769

[26-3] Kirchhoff, G. Ueber die Fraunhofer'schen Linien, 1860，原文转录。https://dingler.bbaw.de/articles/ar156012.html

第 27 章 实验室与恒星之间：光谱化学的归属方法

27.1 实验室光源的作用

实验室里的盐与天空中的太阳之间，相隔的并不只是一段距离。盐可以拿取、溶解、加热和比较，太阳大气却无法在研究者手中按同样方式改变。光谱分析的突破，是在这两种不同实践之间建立一种受约束的对应，使遥远光线能够参与物质判断。它依赖实验，却不把恒星变成一只可以任意调节的烧杯。

基尔霍夫与本生一八六〇年的光谱分析论文，讨论了如何用发光气体的谱线辨别物质，并以钠、锂等实例显示这种方法的敏感性。[27-1] 其装置把火焰光引入光谱观察，使狭缝、色散和比较尺度共同工作。与只看火焰总体颜色相比，谱线使不同成分在重叠的颜色印象中获得可辨位置。

火焰的实验作用首先是提供可以控制的材料变化。把不同盐加入火焰，比较所得谱线；在同一元素的不同化合物之间比较，检查哪些特征可以保持；对混合物进行观察，检查某种成分的特征能否仍被辨认。这样的操作使元素归属不只是一次相似判断，而是经过不同材料安排的重复支持。

实验同时揭示一个困难：极少量杂质也可能产生显著线条。谱线检测的敏感性是一种优点，却使污染控制成为必要条件。如果器皿、试剂或环境引入微量钠，观察者可能把钠线错误地归给待研究物。能够看见很少的成分，并不等于能够自动知道它来自哪里。

因此，“一条谱线对应一种元素”只能作为经过限制的简写。实际识别应考虑多条线的位置、相对行为和实验条件。若只在一处重合，而其他预期线完全缺失，就需要继续解释；若多条线按相同位移共同对应，归属更有说服力。一个线条的形状与位置也可能由邻近线重叠影响，不能把每条图上痕迹都视为单一过程。

还应区分元素身份与化合物状态。某些光谱来自原子或离子，另一些特征属于分子。加热可能使原先化合物分解；不同激发方式也可能产生不同谱形。实验所支持的归属，需要说明究竟指向何种粒子状态，而不是把最初放进火焰的物质名称直接贴在每一条线旁。

在 SDE 框架中，实验光源不是用来给已经确定的天体成分做装饰性验证。它参与建立成分判断的可操作含义：同样的物质在怎样条件下产生哪些特征，改变条件以后哪些特征保持、哪些改变。所谓元素的光谱性质，正是在这种差异运行中获得能够迁移的稳定性。

这并不意味着元素由实验者创造。实验者改变的是使某些特征能够显露的条件，物质的反应又限制这些改变能产生什么。稳定归属出现在主动操作与非任意响应的共同结构中。把这一结构删掉，只剩“光是物质的指纹”，容易使人以为指纹从来不受状态和仪器影响。

27.2 吸收与发射的理论关系

太阳光谱中主要被注意的是暗线，实验火焰中却可以出现亮线。若它们性质相反，为什么位置对应还能支持同一种物质？这一问题不能靠几何重合单独回答，必须建立辐射与物质相互作用的关系。基尔霍夫关于夫琅禾费线的研究与相关辐射理论，正是这一桥接的重要环节。[27-2]

一个简化实验可以说明关键：较冷的吸收气体置于更亮、具有连续光谱的背景前，某些位置会出现相对减弱；同一种气体在另一种观测安排中，又可显示相应亮线。亮与暗的区别并不意味着元素身份相反，而与背景、气体状态及沿光路的辐射交换有关。

不过，不能把这种简化图景无限扩大为所有恒星光谱的完整模型。真实恒星大气存在温度与密度分层，光可能在不同深度产生或被吸收，同一谱线中心与翼部也可能对应不同有效层次。实验室中的一层气体提供可理解的机制，但从它迁移到恒星，需要说明哪些条件保留，哪些条件必须扩展。

以现代教学语言表达，探测器记录的光谱可看作源端辐射经过传播和仪器响应后形成的结果。成分影响吸收与发射，温度和密度影响粒子状态，速度结构改变位置与形状，仪器再对这些结构进行展宽和采样。这个整体关系不是十九世纪论文中的原式，也不代替辐射传输方程；它使读者看见，谱线强度不可能只由元素总量一项决定。

如果两颗星的某条线深浅不同，至少存在几种解释：相关元素含量不同，能参与这一跃迁的粒子比例不同，连续背景不同，或观测响应不同。单凭线深不能唯一决定哪一种成立。归属与丰度必须分开：知道一种元素存在，并不等于知道它占多少；知道某条线很弱，也不等于知道这种元素稀少。

电离进一步使这种分离变得清楚。同一元素在中性与不同电离状态下具有不同线系。温度提高可以让某些中性原子减少，使其中性线变弱，同时使离子线变强。若只按原来一组线判断，会把状态转换误看成元素消失。元素总量保持而可见谱形改变，是反对“线强直读丰度”的一个基本对照。

激发状态也同样重要。某个可见跃迁需要粒子处于适当能级；总体数量很多，并不保证相应能级上的粒子很多。在不同温度范围，这种比例能够发生显著变化。对谱线的解释于是必须连接物质身份与状态分布，而不是先把两者拆成互不相关的标签，再在最后随意相加。

这里的特征纠缠具有明确的可反驳内容。若控制组成不变而改变温度，模型应允许某些线强改变；若模型宣称线强只随丰度变化，就会被这种情况限制。SDE 并不因此替代光谱物理，它要求把这些物理依赖真正纳入对象定义，避免在哲学层面又把可见特征还原为孤立的“内在标签”。

27.3 太阳与恒星的物质问题

将实验谱线与太阳暗线对照，使天文学能够提出一种此前难以用相同方式回答的问题：遥远天体中是否存在地面熟悉的元素？这里的推断依赖自然规律可以在两处迁移的假设，同时依赖多重位置对应和物理解释。它既不是触摸到了太阳物质，也不是毫无经验约束的类比。

佩恩在一九二五年《恒星大气》开篇直接指出这种张力：恒星为研究提供了地面难以达到的物质量与状态，但恒星主要可供观察，不能像实验室材料那样任意操作。[27-3] 她的研究因此从地面物理出发，却持续检查其用于恒星条件时的限制。物理定律的迁移，必须与天文观测之间的总体一致性结合。

她的论证尤其显示，成熟的光谱学不能停留于元素名称与线位置的对应。温度尺度、电离状态、电子压强、谱线强度和分类序列共同进入分析。谱型不同不必首先解释为成分根本不同；许多差别可以由状态变化产生。这样，恒星分类的表面多样性，与物质组成的一定共同性，才有机会在同一解释中协调。

佩恩在第七章比较了谱线“边缘出现”与强度最大位置的用途。所谓边缘出现，是一条线在某类恒星中刚刚可见或即将不可见。这个位置受到元素丰度影响，也受到谱线宽度、连续背景、邻线和仪器色散影响。强度最大位置在一定条件下更适合温度比较，但也有自身限制。[27-3]

这段分析不是细枝末节。它说明同一份谱线资料可以被不同操作定义使用，而定义的选择改变推断责任。如果把“第一次看见”当作纯物理阈值，就可能把仪器可见性误当作原子状态的唯一边界。反过来，若能保持底片和色散条件可比，边缘出现又可以成为估计相对丰度的有用材料。

在第十三章，佩恩明确讨论了相对丰度推断的若干假设，包括线条在边缘出现时未饱和，以及不同元素的有效吸收能力如何处理。她并没有把这些假设写成无条件真理，而是指出其中较强、需要谨慎使用的部分。[27-3] 这种公开条件，使后来的研究能够判断哪些结论应当保留，哪些数值需要修订。

更值得注意的是，同一章节推得氢与氦异常高的丰度，却对相应数值作出保留。我们今天可以知道这些轻元素在恒星中的重要地位，但不能因此把原文写成一项毫无犹疑的现代结论。历史价值恰恰在于：数据、计算结果和当时可接受的物理解释之间出现了张力，而这种张力并没有在论文发表时全部解决。

如果把这一情形只写成“天才被权威压制”，虽然可能触及部分社会史问题，却会遮蔽论文内部真正的推断结构。反之，如果只强调假设不完善，又会低估其新结果。较完整的写

法需要同时保留：方法使异常丰度得以算出，作者对模型外推有所顾虑，后来的证据与理论改变了对结果的接受方式。

三方程在这里可以落实为一次具体往返。恒星成分的显露由谱线比较与状态模型共同支持；异常结果推动研究者重新检查实验、分类和丰度估算；后来的改进又改变恒星大气的整体解释。S 不是一张永远固定的答案表，D 也不只是把更多谱线抄进去，E 更不是无关紧要的背景，它们在同一问题中相互约束。

佩恩所用的“边缘出现”方法，可以进一步拆成一个条件推理。假设一条线刚好达到可见阈值时，需要一定数量处于合适状态的吸收粒子；再利用电离与激发模型估算这种状态占该元素总粒子的比例，就可以反推总量。若合适状态的比例很小，而线仍然可见，推得的元素总量就可能很大。这个逻辑说明异常丰度不是单纯因为某条线画得特别深。

但推理中的“相同可见阈值”并不天然等于“相同粒子数”。不同跃迁的吸收效率可能不同，底片对不同波长的反应也可能不同，线条附近的连续背景和混合程度还会改变可见性。佩恩在原文中把不同元素吸收效率的处理列为重要假设，并限定其工作性用途。[27-3] 这意味着数值比较需要与假设强度一起评价。

她还区分谱线最大与边缘出现时的有效形成层次。在较强吸收下，光线可能主要反映更高层的条件；随着适合吸收的粒子减少，能够贡献的深度改变，线与连续背景的对比也随之变化。由此，线强不是一个固定薄层中粒子数的直接读表值。恒星大气的垂直结构参与构成最终特征。

这一细节使“同一元素”的稳定身份与“同一谱线表现”的变化能够同时成立。元素身份可以保持，形成区域和状态比例却改变。若把一条线看作完全独立的对象，就很难说明为何它与温度、压力和背景相互变化；把它放回辐射过程，才知道哪些变化是在预期之内，哪些变化真正构成异常。

原文还强调，相关相对丰度估计主要依靠互相一致的资料，而不宣称每个边缘出现位置都具有绝对、无装置依赖的意义。相同色散、相近底片密度和良好成像，目的是尽可能把比较限定在共同的可见性条件中。这是一个务实选择：当绝对校准困难时，先争取可信的相对关系，但不把相对结果冒充绝对精确值。

氢与氦的情形因此有不同层次。首先，相关谱线在很宽条件范围中保持可见，成为待解释事实；其次，特定模型由此推得很大的相对丰度；再次，作者对这些数值是否可以按字面接受作出保留。三个层次可以同时存在。把其中任何一个删除，都会改变这项研究实际提出的问题。

尤其不能把后来知道的正确方向变成早期方法每个假设都正确的证明。一个推论可能抓住重要事实，同时采用某些需要改进的近似。现代科学常在修订模型以后保留结果的一部

分，也可能改变其数值。历史评价应当区分方向性突破、定量准确性和解释成熟度，而不是只给整项研究贴“正确”或“错误”标签。

还应注意，恒星大气的组成与恒星整个内部的组成不是同一观测对象。佩恩在第十三章结尾明确限制光谱推断直接涉及大气，并讨论由此推向内部时的额外考虑。[27-3] 这个区别对本书十分重要。看到外层某种元素，并不能无需结构与演化理论就断言内部各处比例相同。

由外层到内部的跨越，类似于由实验室到恒星的跨越：都需要说明连接条件。并非说不能迁移，而是迁移本身属于论证的一部分。若恒星内部存在分层、混合或核反应，外层记录与核心状态的关系就需要具体研究。一个可靠的局部结论，不应因其重要而被无限扩张。

这也使本章与恒星能源问题自然相接。光谱分析提供组成与状态线索，核物理需要知道可用燃料和反应条件，但两者之间还隔着内部结构。新世界的统一发生于这些不同知识能够在明确桥接下彼此限制，多样性则要求每项方法保留自己的测量对象和有效域。

温度序列还为检验归属提供了一种整体手段。若把一条线归给某元素的某种电离状态，这个归属就应当与该状态在不同温度下的出现趋势相容。单独一颗星中波长接近可能是偶合；在一系列恒星中，多条线的增强、减弱和消失具有相互关联的次序，能够增加判断的约束。不过，不同线也可能来自不同激发能级，它们不应被要求机械地同涨同落。研究者需要比较理论预期的具体关系，而不是把“趋势一致”当成一个万能标准。这样，分类目录提供的顺序便进入了物质识别：它不是先完成分类、再另外查成分的两张独立表，而是通过状态变化使某种归属能够接受新的检查。若不同电离阶段的线给出彼此冲突的丰度，温度、压强及谱线归属就都需要重新审查。[27-3]

27.4 光谱成分判断的限制

谱线混合是最直接的一种限制。两种不同成分的线如果相距很近，在较低分辨率下可能表现为一条结构。若把它全部归给其中一种元素，就会高估该元素的贡献。提高分辨率有时可以拆开，但也会牺牲光通量或增加观测困难；面对微弱恒星，装置能力决定哪些问题能够实际回答。

速度也影响位置判断。若天体相对观察者运动，多条谱线可能共同发生位移。先假定每条线都必须落在地面实验的原位置，会错过这种整体变化；但若允许任意位移，又可能制造虚假匹配。可靠方法要求同一速度解释多条相关线，并结合其他信息检查。共同位移成为一种约束，而不是任意调整按钮。

温度与压力影响线形，不同深度的辐射贡献又可能使简单一层模型不足。对于丰度分析，连续背景的确定尤其重要。吸收线是相对背景的减弱，若背景估计改变，线强度的量化

也会改变。密集谱线区域中，所谓连续谱甚至不容易直接找到。读取一条曲线必须先说明它相对于什么测量。

线条还可能饱和。若中心附近的光已经被强烈吸收，增加吸收粒子不一定使中心按比例继续变暗，变化可能更多出现在翼部或整体等效宽度中。因此，“线深加倍意味着含量加倍”通常不是可靠规则。需要知道所在物理区间，才能判断哪种测量与数量关系近似成立。

仪器响应则在最后一道关口影响全部推断。不同波长的探测效率不同，光学装置对谱线有有限展宽，摄影底片的黑度也不必与入射光强线性对应。若这些效应不被控制，两个看似不同的恒星光谱可能只是记录条件不同。分类和定量分析对校准的要求也不一样，不能用前者的成功自动保证后者准确。

佩恩为了提高内部可比性，强调用于相关估计的底片应具有相同色散、相近密度和良好清晰度。[27-3] 这里的“相同条件”并非形式规定，而是为了减少仪器可见阈值变化对物理判断的影响。选择资料本身就属于论证，不是等全部数据收集完毕后才开始思考。

由此，光谱化学有三项必须分别完成的工作：确定某个记录结构值得信赖，判断它与何种原子或分子过程对应，进一步估计物质数量及状态。三者可以互相支持，却不能通过一次线条重合同时完成。忽略分层，最容易在第二项取得成功时过早宣布第三项也已解决。

对 SDE 而言，这个案例迫使特征纠缠获得实质内容。恒星“含有某元素”与恒星“呈现某条线”不是同一个断言；两者之间通过状态、辐射传输和测量联系。元素属性的科学显露存在于这个整体之中，却不因此成为任意主观判断。整体越被清楚拆解与重建，归属越能接受检验。

六路径的有效转换也在这里显示。以已知元素为起点找线，有利于形成候选；从不符合预期的线强出发检查温度、电离与响应，则可能推翻原来的丰度直觉。路径改变应当导致具体判断改变，例如保留元素存在、撤回含量估计，而不是每次换路最后都无条件证明同一个答案。

这部思想史所说的和谐，不要求太阳与实验室处于相同状态，也不要求所有恒星具有完全相同组成。它要求不同条件下的结果能够在明确物理关系中相接：哪些相同指向共同物质，哪些不同指向状态变化，哪些剩余差异仍待解释。遥远星光因此成为可分析的物质证据，而它的可信度恰恰建立在这些不能省略的差别上。

本章参考文献

[27-1] Kirchhoff, G.; Bunsen, R. Chemical Analysis by Observation of Spectra, 1860，原始论文英文转录。

https://www.chemteam.info/Chem-History/Kirchhoff-Bunsen-1860.html

[27-2] Kirchhoff, G. Ueber die Fraunhofer'schen Linien, 1860. https://dingler.bbaw.de/articles/ar156012.html

[27-3] Payne, C. H. Stellar Atmospheres, 1925，第一、七、十三章及相关章节；公开全文转录。https://www.gutenberg.org/files/73996/73996-h/73996-h.htm

第 28 章 哈佛摄影与恒星光谱分类

28.1 摄影底片为何改变样本

一张底片使观察可以离开夜空继续进行。望远镜在夜间取得记录，研究者在白天比较、编号、重新检查；多年以后，其他人还能带着新的问题回到同一张底片。摄影并没有取消观察者，而是把接收星光与判断星光之间的时间关系改变了。一次夜观由此可以成为多次研究的材料。

哈佛天文台的光谱工作依靠大规模底片积累、仪器安排和持续分类形成。亨利·德雷珀纪念项目及其相关目录，把这种工作组织成长期计划。一八九〇年的《德雷珀恒星光谱目录》已经体现了大量摄影光谱的系统整理，后来坎农参与完成的《亨利·德雷珀星表》进一步扩展了分类的公共规模。[28-1][28-2]

物镜棱镜提供了一种适合广泛调查的方法。把棱镜放在望远镜物镜前，视场内许多恒星的光可以同时展开成短光谱。一张底片于是容纳多颗星的谱形，而不必逐一把每颗星放到单独狭缝中。这个安排提高了取得样本的效率，也形成了相应限制：光谱可能互相覆盖，色散有限，最弱细节不一定能够分辨。

底片曝光还把时间纳入信号积累。较长曝光能够使微弱目标留下记录，却也可能让亮星过度曝光，使线条形态失去合适对比。相同天空中不同亮度的星，不一定能在同一曝光下同时得到最佳记录。因此，目录所包含的对象与可判读的特征，始终受到曝光选择和记录材料的限制。

这并非摄影不如眼睛的简单比较。眼睛与底片具有不同响应，摄影使保存、积累和重复比较成为可能，也要求研究者学习怎样解释黑度、背景和清晰度。所谓数据规模扩大，不是把已有观察方式机械复制更多次，而是形成一种新的材料系统。

底片的标记、日志与保存条件因此具有认识价值。一张失去日期、仪器和曝光信息的照片，虽然仍可观看，却难以承担长期变化的严格分析。相反，保存良好的底片与日志可以让后来者知道某个差异是否可能来自记录条件。档案不是研究结束后收纳成果的仓库，它从一开始就在构成可延续的研究。

哈佛底片馆藏保留的人员与项目资料，使我们能够看到这个系统并非一个抽象机构自动运作。弗莱明、莫里、坎农以及许多参与记录、测量与整理的人，承担了不同任务。[28-2][28-3] 她们面对的是具体图像质量、谱形边界和资料不一致，而不只是把自然给出的答案抄进表格。

SDE 分析在此把样本理解为发生的结果。天空中的恒星总体，不等于实际进入目录的恒星集合；进入底片，也不等于每种属性都可可靠读取。仪器、曝光、选源和判读共同决定哪些差异能够成为公共分类的基础。规模带来的新世界，同时包含增长的能力与明确的盲区。

28.2 谱型从细节到序列

今天熟悉的 O、B、A、F、G、K、M 并不是按字母表顺序排列的名称。这个不规则顺序保留了分类形成与调整的历史。早期类别被合并、重排或另作处理，最后保留下来的主要序列，使大量恒星的相似谱形能够用共同语言表达。记住顺序有实用价值，理解它为何这样形成则属于思想史。

分类首先要求决定相似性。看氢线强弱，某些星可以靠近；改看金属线或线宽，另一些关系会显现。没有单一特征天然负责所有分类任务。对低色散、大量样本而言，整体谱形可能比精确测量某一弱线更可操作。对少量高质量谱图而言，更细的特征又可能有特殊价值。

莫里对亮星光谱所作的分类，重视较细的谱线形态差异，其一八九七年研究保留了不同于后来常用简化序列的组织方式。[28-4] 这种工作不能只按“是否成为最终通用字母表”评价。某些细节对于大规模目录不够便利，却可能为后来理解不同光度或物理状态提供线索。方法的价值随任务粒度改变。

坎农的工作则显示一种能够大规模稳定运行的分类能力。她将大量谱形放入可连续比较的序列，并使分类成为可以广泛使用的目录资料。[28-2] 这不是因为每颗星都天然属于边界清楚的抽屉，而是因为连续变化需要被组织为实际可辨的等级。类别之间存在过渡，分类者必须处理边界个体。

佩恩后来在《恒星大气》第十四章提醒，德雷珀分类原本以谱形相似为基础，并没有预先建立完整演化理论。她还指出，实际底片色散较短，多数弱线不能精确测量，因此不能用理想高分辨率条件要求整个目录。[28-5] 这段说明帮助我们公平理解分类者面对的任务，而不是拿后来仪器指责早期类别过粗。

一套有用分类需要兼顾三种尺度。太粗时，它无法区分有意义的物理差异；太细时，资料可能不足以支持稳定判定；太复杂时，又难以在大量样本中保持一致。分类者的判断力就在这些限制之间形成。类别数量不是越多越科学，重要的是分类步骤是否与实际资料的可辨程度相配。

边界个体还可以成为新知识的入口。如果某颗星的氢线像一类，而其他特征像另一类，研究者可以将其视为误差，也可以保留为特殊对象。何时应重拍底片、何时应标注异常、何时应修改类别，需要具体判断。若所有不适配都被强行塞入现有序列，分类的统一性就会压制新的差异。

相反，若每个不适配都立即获得独立类别，目录便难以积累共同知识。成熟分类必须允许一部分差异被视为暂时不确定，同时保留足够信息供未来重查。这是一种制度化的开放：大多数对象可以按标准处理，少数异常不被抹掉，新的证据能够改变其归属。

28.3 劳动组织与知识归属

底片分类经常被称为“计算”或“整理”，这些词容易让人误以为工作完全机械。事实上，判定谱形、比较标准、识别异常和保持长期一致，都需要训练与判断。研究者不是只在已经明确的数字之间执行算术，而是在不完全清晰的图像中决定哪些差别值得保留。

弗莱明的工作包括谱型整理和特殊天体辨认；莫里在谱线细节上提出不同组织；坎农使大规模分类取得持久的公共形式。哈佛档案中的个人材料与目录出版物，使这些不同贡献能够被分别追踪。[28-2][28-3][28-4] 用“哈佛女性计算员”一个群体名称概括她们，有助于恢复被忽略的劳动，却仍不能替代逐人区分研究内容。

知识归属还有多种层次。资助者使长期计划得以维持，台长协调仪器与出版，观测者取得底片，分类者建立序列，理论研究者进一步解释物理含义。这些贡献彼此需要，但并不相同。公平叙述既不能把全部成就归给机构负责人，也不能在恢复一类劳动时抹掉其他环节。

目录署名、报告的作者名与实际参与工作之间，可能存在历史制度造成的不对称。研究这种问题需要档案，而不应凭今天的直觉虚构每一次合作中的具体冲突。能够确认的是，长期重复劳动本身具有创造性结果：它使过去偶见的谱形差异变成可以比较其分布、连续性和异常的知识系统。

大规模计划也会使个体判断受到标准约束。同样一张底片，两个分类者可能对边界略有分歧；若没有标准谱与复查安排，目录就可能随着人员改变而产生系统漂移。训练并非让每个人失去判断，而是使判断能够与共同尺度协调。共同体需要解释如何处理分歧，而不只是宣布最终表值一致。

熟练判读有时可以很快完成，但速度不能作为唯一质量标准。某些对象清楚符合典型，另一些需要反复比较；某些底片质量好，另一些则不宜强行给出精细分类。把全部工作统一为同样时长，可能误解专家能力。真正的熟练包括知道何时可以迅速封装判断，何时必须重新打开分析。

摄影档案还使劳动具有跨代延续性。后来的研究者可以重新测量旧底片，提出早期分类者没有研究的问题。这个可能性依赖原记录和背景信息被保存。公开目录提供便利索引，底片与日志则保留比目录更丰富的材料。两者的关系如同压缩结果与可重新分析的来源，不能互相替代。

从 SDE 角度，分类能力属于人、底片、仪器、标准和组织形成的复合体。说它是一种复合能力，并不等于个人贡献不可辨；相反，只有说明不同人怎样改变了操作与标准，个人创造性才不会被抽象机构吞没。主体性不是脱离工具的独立天赋，而可以在长期材料互动中形成独特判断。

这也使知识的统一与多样有了现实形态。共同目录为世界各地提供统一索引，不同分类方案与异常记录则保留重新解释的可能。和谐不是所有研究者从来没有分歧，而是分歧能够通过可查材料与明确标准转化为更细的区分，而不被声望或流程简单消音。

若把一项底片分类任务展开，至少可以辨认几种不同判断。先要将底片上的光谱对应到正确恒星，避免密集星场中的归属错误；再检查曝光和清晰度是否足以判读；随后比较强线、线组和整体谱形；最后决定采用正式类别、较粗类别或特殊标记。这里描述的是理解目录所需的操作结构，不是声称所有历史分类者都遵循同一份固定流程。

每一步都可能改变最后结果。若对象对应错误，谱型本身即使判得准确，也会落到错误星名下；若底片质量不足，却强行给出细分类型，目录就会制造虚假精度；若异常标记在汇编时被删掉，后来者便难以知道哪些对象值得重查。知识可靠性因此不只属于最后一次分类判断。

连续谱型中的边界处理，还涉及标准星的角色。标准星提供实际可比较的样本，使类别不只是抽象文字定义。但标准星也应接受重新检查，尤其当后来发现其具有变光、双星或特殊光谱性质时。一个被用作稳定参照的对象，并不会因获得标准身份而免于自然变化。

目录中的分类也具有分辨率。一个粗类别可在较弱资料上稳定识别，细分类别则要求更多细节。二者不必彼此否定：广泛目录提供共同骨架，专题研究可以在某一范围重分类。佩恩对此作了清楚说明，强调更细研究的需要不等于原来大规模分类毫无价值。[28-5]

这种多层分类对资料使用者很重要。如果研究的问题只需区分大致温度范围，粗分类可能已经足够；如果要研究精细丰度或表面重力，同一目录则不能承担全部任务。资料的适用性取决于问题，不取决于它是否被冠以权威目录之名。权威的合理作用是提供可追溯起点，而非保证所有用途。

分类序列中的字母顺序还体现历史沉积。保留某些旧字母可以维持与既有文献和目录的联系，即使它们已不再按最初的字母顺序排列。完全重命名也许在形式上更整齐，却会增加跨版本对应成本。知识系统的更新需要在概念清晰与历史连续之间选择，最终形式往往保存这种选择的痕迹。

这种连续性不是单纯保守。旧编号和类别让研究者能够把新理论返回到既有材料。若每次理论改变都把全部档案名称废弃，长期积累将变得困难。但保留名称必须伴随解释更新，否则同一个词在不同时代的含义变化会造成误读。目录既要可继承，也要可重新理解。

底片中异常对象的保存还具有一种未来价值。早期研究者未必知道某种特殊谱形最终对应何种物理机制，只能准确描述其不寻常之处。若这种描述足够具体，后来的理论就能回来寻找对应；若为了整齐将其抹平，未来解释便失去材料。暂时没有机制的差异，仍可以是高质量知识。

因此，经验分类不是理论缺席时的低级替代品。它可以通过稳定比较形成理论必须面对的约束。理论若能解释主要序列，却不能解释一批重复出现的异常，就需要继续发展；分类若无法在不同底片上稳定复现，也需要修订。两者不是高低等级关系，而是具有不同责任的合作关系。

从这里看，哈佛工作的创造性也包含节奏控制。大规模目录需要持续产出，专题异常需要较慢深入；所有工作都改为深入个案会失去覆盖，所有工作都追求高速分类则可能错失新物理。统一性、多样性与和谐性的平衡，表现为不同任务能够并存，并且彼此保留回流通道。

底片识别还有一个经常被文字目录隐藏的步骤：谱迹必须同天空中的目标配对。物镜棱镜同时记录许多对象，每条光谱沿某一方向展开，亮星与邻近暗星的谱迹可能重叠。如果把一个谱迹的特征记到另一个星号下，之后再精细的物理解释也会建立在错误归属上。位置索引、视场信息和再次查片因此属于分类结果的组成条件。一个星号并不是可以脱离原始图像永远自明的标签。

不同摄影材料还会改变哪些谱段最容易看见。某些波段记录灵敏，另一些波段记录很弱，最终留下的“特征丰富”便部分依赖材料响应。曝光增加可以使弱谱迹出现，也可能使强处饱和；曝光不足又可能只留下少数最突出的线。对比对象时，不能只看两个图像上都有一条黑带，还要看它们是否处在可比较的记录条件中。佩恩强调同类材料的比较，正是因为分类和测量必须在这些限制下建立连续性。[28-5]

这些实际困难也改变了熟练的含义。熟练者不仅记得哪些线属于哪一类，还能分辨哪些细节不足以判断，哪些材料值得重新观察，哪些差别可能只是底片质量。可靠目录既包含肯定分类，也需要保存疑问和特殊情况。若管理只奖励每小时完成的条目数，而让不确定性无处记录，目录表面可能更整齐，后续研究却更难找回判断依据。评价这类劳动，必须把速度与可追溯性放在同一工作过程内考虑。原片索引也是成果的一部分。

28.4 分类的再解释

光谱序列建立以后，物理学提出了新的问题：这些类别究竟主要改变了什么？如果谱线不同都表示元素含量不同，那么恒星之间的物质差异会非常复杂。电离与激发理论则提供另

一种可能：在一定组成条件下，温度和压力变化可以系统改变谱形。分类因此从相似性组织逐步转向物理状态解释。

这个转变没有使原目录失效。相反，正因为大量谱形已经被稳定排序，理论才有广泛材料检验自己是否能够解释序列。经验分类与物理理论不是简单的先后替代关系。前者提供需要解释的秩序，后者解释其中部分结构，又指出哪些类别内部还包含重要差异。

佩恩认为德雷珀序列与热电离程度存在重要联系，同时强调同谱型恒星仍可能在其他物理条件上差异很大。[28-5] 因而，字母类别不能直接代表恒星全部性质。两颗谱型相近的星，光度、半径和表面重力可能不同。单一分类轴有效，并不意味着宇宙只有单一变化维度。

这里尤其要避免把谱型序列当作年龄队列。沿着字母顺序排列，并不意味着每颗星都必须依次经过全部类别；不同质量的恒星在不同阶段可以处于相近谱型。一张分类表描述样本中的相似关系，演化路线则需要时间机制和额外证据。把序列直接看作生命历程，是从静态秩序跃迁到动态解释时常见的误读。

同一个标签也可能在历史上取得新的含义。早期研究者说某星属于某类，主要指其谱形与标准相似；后来读者可能立即联想到温度范围。两种用法有关，却不完全相同。若不注意这一变化，就会把后来物理知识倒灌入早期目录，误以为分类最初已经由成熟理论推导出来。

这种语义回写并非坏事。科学常常保留旧术语，同时深化它的解释。但深化应当允许重新划分：特殊星、不同光度类别或新发现对象可能超出现有单轴。若一套分类只能继续维持而不能接受拆分，它就会从研究工具变成认知障碍。

从三方程看，分类 S 由底片比较 D 与判读条件 E 共同稳定；分类又安排新资料怎样被检查；物理解释进入以后，E 中的相关特征关系被重新组织。某些旧差异被理解为温度效应，某些原来不显眼的线宽差异获得新重要性。这是有内容的回写，而不仅是给字母附上一段新的解释文字。

如果换一条研究路径，从理论温度出发寻找分类对应，容易强调序列的整齐；从历史底片出发检查实际判读，则会看到模糊、噪声和异常。两者需要同时保留。整齐显示理论的解释力量，异常说明理论与分类都尚有边界。只写整齐，会把科学史变成命中史；只写混乱，又无法解释目录为何如此有用。

哈佛光谱工作的长远贡献，因而不仅是给许多恒星贴上标签。它形成了可以跨对象比较、跨研究者交接、跨时代重新解释的材料与标准体系。星空从一颗颗分别描述的光点，进一步成为具有可比较物理差异的群体。这个群体世界为赫罗图提供了必要条件，却仍保留一个等待解决的问题：相似谱形的星，为何可以拥有极不相同的发光能力？

本章参考文献

[28-1] Harvard College Observatory. The Draper Catalogue of Stellar Spectra, Annals 27, 1890. https://books.google.com/books/about/The_Draper_Catalogue_of_Stellar_Spectra.html?hl=en&id=gzwhAQAAMAAJ&output=html_text

[28-2] Harvard Plate Stacks. Annie Jump Cannon，档案与原始目录索引。https://platestacks.cfa.harvard.edu/women-at-hco/annie-jump-cannon

[28-3] Harvard Plate Stacks. Williamina Fleming，档案索引。https://platestacks.cfa.harvard.edu/women-at-hco/williamina-fleming

[28-4] Maury, A. C. Spectra of Bright Stars Photographed with the 11-inch Draper Telescope, Annals 28, 1897. https://scixplorer.org/abs/1897AnHar..28....1M/abstract

[28-5] Payne, C. H. Stellar Atmospheres, Chapter XIV, 1925. https://www.gutenberg.org/files/73996/73996-h/73996-h.htm

第 29 章 赫罗图：两种测量怎样生成恒星演化问题

29.1 二维图表的资料代价

把恒星放在一张二维图上，操作似乎极其简单：横轴写谱型或温度，纵轴写绝对星等或光度，每颗星画一个点。然而，这两条轴的取得方式并不相同。光谱分类可以从光的结构入手，本征发光能力却需要距离和光度校准。图的简洁，是许多研究环节完成以后才获得的简洁。

表观亮度不能直接代表恒星自身发光多少。一颗遥远而明亮的巨星，可能与一颗近处较暗的星在地球上具有相近亮度。若把它们直接按表观亮度排列，空间距离与天体性质就混在一起。赫罗图的力量首先来自把这种混合部分拆开，使光度与谱型之间的关系能够显露。

在忽略消光、假设辐射近似各向同性的现代说明中，接收到的通量与光度之间满足平方反比关系。距离估计增加一倍，同一通量所对应的光度增加四倍。由此可见，光度轴不是单靠亮度计决定的；距离的系统误差会以确定方式进入图中位置。改进视差可以使恒星在图上移动，即使恒星本身没有发生任何物理变化。

谱型与有效温度也不是完全相同的观测量。谱型属于按谱形建立的分类，温度通常依赖物理解释与校准；颜色指数则是不同波段测光之间的比较，还受到消光及仪器响应影响。历史上和今天统称为赫罗图的图像，实际可能采用不同横轴。使用时应当说明是直接观测分类，还是经过模型转换的物理量。

罗素一九一四年讨论光谱与恒星其他特征的关系，展示了这种交叉比较的重要性；赫茨普龙此前也以相关资料研究恒星的光度区别和星团分布。[29-1] 两人的工作不应被写成某张现代教科书图像突然完整出现。不同资料来源和表示方式逐步形成了今天熟悉的共同图景。

星团提供了一个特别有用的安排。若成员在空间上相对集中，那么它们的距离差相对于整体距离较小，表观亮度差就能够在一定近似下反映本征亮度差。这使研究者不必先为每颗星取得独立高精度视差，就能观察群体中颜色与亮度的关系。但成员资格、星团深度和前景背景污染仍需检查。

同样，近邻恒星样本具有视差优势，却偏向某些对象；按表观亮度选出的样本，则更容易包含本征亮度高、可以从远处看见的星。不同选样可以使同一图的点数分布明显不同。图上某一类星多，不一定意味着宇宙中它们占同样比例。样本的进入机制必须随图一起理解。

因此，赫罗图不是把天然已经在一张图中的恒星照相保存。它将来自不同操作的属性整合到共同空间，使原先分离的资料发生新的联系。新的特征纠缠不仅包括恒星的物理属性，也包括使这些属性获得可比数值的校准关系。若缺少这种联接，两个轴只是一对并列标签。

距离校准的作用可以通过一个现代教学例子看清。设两颗星在所用波段接收到的通量相同，其中一颗距离为另一颗的十倍。只要暂时忽略沿途吸收，远星在这个波段的本征光度就是近星的一百倍。它们在表观亮度目录中可以相邻，在按绝对亮度排列的图上却必须分开。增加距离资料并不是给原有分类增加一个附注，而是重新规定比较所比较的东西。

星等使用对数标度，使这一转换还有容易误读的方向。较小的星等数值表示较亮；相差五等对应一百倍亮度比。因此，把同样表观星等的两颗星分别置于十倍不同的距离后，其绝对星等相差五等，而不是相差十倍。图的纵轴如果把更明亮对象放在上方，数值就通常向上减小。读图者必须知道这种约定，否则会把序列斜率甚至异常方向理解反。

这里还涉及波段。可见波段绝对星等不是全部电磁辐射的总光度。温度不同的恒星，辐射分配到可见光的比例不同。将某一波段亮度换算成总光度，需要相应修正；将颜色换成温度，也需要考虑组成、重力与消光。于是，观测的颜色—星等图可以比推导的温度—总光度图更接近原始材料，但后者更便于直接同结构模型比较。这不是选择其中一张为真、另一张为假的问题，而是每张图承担的转换不同。

星团距离估计的变化则具有另一种几何效果：若所有成员都采用同一距离，并且只修改这个共同距离，所有绝对星等便一起移动。图的整体纵向位置改变，内部成员间的亮度差却保持。若某一部分星的颜色或亮度相对其他成员发生不同变化，就不能只用共同距离调整解释。整体平移与内部形状改变的区别，帮助研究者识别校准问题和群体性质问题。

29.2 主序并不天然预设

当许多恒星进入同一图，分布并没有均匀占据全部区域。大量对象沿着一条主要序列排列，而另一些对象在相近谱型下具有明显不同的光度。这种形态要求解释：为什么某些温度与光度组合常见，另一些组合稀少？一个关于个别恒星的描述问题，转变成了关于群体可能状态的问题。

主序不能被视为绘图前已经规定的规则。若研究者强制所有点落在一条预期曲线上，就无法发现巨星与矮星的区别。图的知识作用在于让资料有机会违背简单期待。谱型相近而光度差别很大，说明表面光谱状态不足以单独决定总发光能力，还需要其他物理维度。

现代辐射关系提供一个直观说明。对于用有效温度描述的恒星，光度与表面积及温度四次方相联系。在相近有效温度下，更高光度意味着更大辐射面积，从而指向更大半径。这个说明不是说早期所有参数都已精确知道，而是解释图中分离为何能够转化为恒星尺度的新问题。

反过来，若两颗星光度相近而温度不同，半径也必须相应不同。二维分布于是能够帮助识别一个单轴分类看不见的关系：热并不等于总是更亮，冷也不等于总是更暗。对象的总发光与单位面积发光必须分开。图像迫使人们把这些原来容易混用的概念明确化。

罗素讨论恒星群体的平均特征和选样问题，说明早期研究已经意识到资料集合的来历重要。[29-1] 今天回看这些图，不应只欣赏它们与现代图像相似的部分，也要留意样本量、距离资料和特殊对象的限制。早期分布足以提出重大问题，却未必具有现代全天调查的完整统计含义。

图上的空白也需要谨慎解释。某处没有点，可能表示相应状态难以维持或停留时间很短，也可能表示对象太暗、距离测不准或分类资料不足。物理解释与选择效应之间必须竞争。不能把图像的每一个缺口都立即写成自然禁止的区域。

这种竞争使赫罗图成为新型研究工具。它并不只总结已经知道的恒星性质，还可以指导寻找缺少的对象、重测异常距离、检查分类是否错误。图作为 S 形成以后，改变 D 的安排；新的资料进入，又可能重画原来分支。图的稳定来自持续修订，而不是一次绘制后永久固定。

在本书的发生学解释中，恒星由此取得新的群体身份。一颗星不再只属于某个光谱字母类，也处于与其他星共同组成的物理分布之中。其异常性可能来自相对位置，而非来自某个孤立特征特别奇怪。分布本身生成了新的研究对象，这是从单体属性转向关系结构的重要变化。

选择效应也可以做成一个清楚的数量比较。仍用理想化的均匀空间分布，假设调查有固定通量下限，而且没有吸收及其他筛选。光度高一百倍的星，可以从十倍远处进入同一调查；半径十倍的球所包围的体积，是原来的千倍。即使这种亮星在单位空间内少得多，亮度受限的目录中仍可能相当醒目。图上点数不能脱离进入图中的空间范围来解释。

这个例子不是说真实银河系均匀，也不是说只需除以一个体积就能完成全部统计。银河结构、尘埃、双星和不同天区的观测完整性都会增加条件。它说明的是一个更基础的事实：常见于目录与常见于空间不是同一判断。若比较主序星和巨星的数量而不检查这层区别，就可能把可见距离优势误认为一种演化阶段长期占据优势。

29.3 离群星与理论重排

巨星与白矮星使简单的温度—亮度直觉受到挑战。巨星可以在较低表面温度下具有很高总光度，要求较大半径；白矮星则可以具有较高温度而总光度很低，要求较小辐射面积。如果再结合双星轨道提供的质量约束，就会出现密度问题。不同观测链在同一对象上会合，才使它成为理论难题。

以天狼星伴星这类对象为例，伴星的质量信息可以从轨道关系获得，而表面状态与发光量来自光谱和测光。这些资料不是同一台仪器重复给出同一个答案，它们约束不同量。若共同指向很小体积中包含可观质量，那么研究者必须检查普通物质状态的模型是否仍适用。

爱丁顿在《恒星内部结构》中把恒星结构、辐射输运和特殊高密度对象放在统一研究中。[29-2] 这里的历史张力不能被一句“后来有简并压所以解决了”迅速抹平。早期模型面对的是多条证据共同施加的压力，而解决需要新的物理概念、状态关系与进一步观测，不只是把图上的离群点换一个名称。

离群点也可能来自资料错误。距离估计不准、未分辨双星、错误的谱型或消光处理，都可以使对象偏离主序。成熟研究不能每见离群就宣布新物理，也不能为了维护主序而删掉所有离群。需要沿着每个坐标的来历逐步检查，分辨数据问题与真实的不同状态。

未分辨双星提供一个容易理解的例子。若两个恒星的光被记录成一个对象，总亮度会增加，谱形却可能仍近似其中主要成分。它在图上可能显得比单星更亮。此时，偏离不是一种新的单星结构，而是观察单位与物理单位没有一一对应。对象“一个点”的身份需要重新打开。

这一点特别适合检验 SDE 的整体性。若把图上每个点当成天然独立的恒星实体，就可能在分析开始时已经误设研究单位。仪器分辨率决定一个读数是否合并多个对象，合并又改变光度和谱形。相互规定不仅发生在几个抽象变量之间，也发生在“我们究竟测了一个什么东西”的层面。

离群研究还会反过来改变分类。若某些线宽或线强差异与光度关联，光谱本身就可能增加区分巨星与矮星的能力。原来在大规模分类中次要的细节，由于二维图提出的新问题而获得重要性。图不是简单消费谱型目录，它会回写目录的判读任务。

现代对毕星团的精密研究仍展示类似结构：更好的视差有助于形成更清楚的分布，但把颜色与绝对星等转换为有效温度和光度时，模型与系统误差仍然重要。[29-3] 这说明问题并不随着仪器先进而消失。新的精度会把限制推向新的环节，使从观测图到物理图的转换继续成为研究对象。

图上的多样性因此具有积极作用。主序提供可以理解的共同结构，巨星和白矮星要求扩大物理解释，双星和误差案例提醒研究单位与数据过程可能不同。一个有生命力的理论需要同时处理这些不同来源，不能用单一叙事把它们全部同化。

再把半径推断展开，就能够看出离群星为何要求多个尺度共同参与。设两颗恒星有效温度相同，而总光度相差一万倍。按光度等于表面积乘单位面积辐射的关系，半径比为一百，而不是一万。若暂时再设质量相同，平均密度便相差一百万倍。这里每一个额外条件都有作

用：没有温度比较不能从光度取得面积，没有质量约束不能从大小取得密度。这是一组用于说明的现代比例，并非对某两颗历史样本的实际测量。

白矮星的压力正是沿这样的关系累积起来。高温、小光度与轨道质量合在一起，才将小体积和高密度变成无法轻易回避的结果。若只看到暗弱，可以解释为普通低质量星；若只有高温，可以解释为发光很强但距离遥远；多项独立资料相接以后，这些简单退路才逐步受限。异常之所以有分量，来自可替代解释空间被压缩，而不是来自对象名字中的“奇异”。

未分辨双星同样有可计算的图形后果。在理想化的同温、同亮双星中，两颗星合起来使通量增加一倍，星等约减少零点七五等，颜色却可以近似不变。因此它们可以在主序上方形成相关分布，不能全部解释为单星正在离开主序。不过，这个固定偏移仅适用于近似相等的成员；亮度差很大的组合，偏移与合成颜色都不同。用一个简明例子认识机制之后，仍需回到各对象的实际组合。

29.4 静态分布与动态演化

一张赫罗图通常同时容纳许多不同恒星在某个时期的状态。它不是一颗恒星从出生到死亡的连续录像。把图上的点连接成箭头，需要额外理论：恒星怎样产生能量，内部成分怎样变化，结构如何响应，以及某一阶段持续多久。没有这些条件，邻近只是坐标邻近，不是时间先后。

早期研究曾尝试根据温度与光度分布构造恒星演化路线。它们具有探索价值，但不同能源机制和内部结构假设可以导向不同路线。后来核物理与恒星结构的发展改变了这些解释。图的观测秩序可以相对稳定，而对它的时间解释发生重要修订。

星团提供连接群体与时间的一种途径。若成员具有相近形成时期和相对一致的初始成分，那么不同质量恒星的状态差异能够帮助检验演化模型。较大质量与较小质量恒星不会以相同速度消耗可用燃料；在一定年龄下，某些对象已经离开主序，而另一些仍然停留。图上主序转折因此具有年龄信息，但这种信息依赖模型。

这里也需要保留近似。星团可能含有双星、不同旋转速度、成员污染或更复杂的形成历史；消光也可能在视场中不同。不能因为“共同出生”方便解释，就把它当成绝不需要验证的事实。星团研究的力量在于减少某些未知差异，而不是消除所有差异。

静态分布中的点数还受到阶段持续时间影响。如果某个演化阶段很短，即使许多星都会经过，某一时刻看到的数量也可能不多。反之，停留很久的阶段可以积累大量对象。因此，分布密度与演化速度之间可能存在联系，但还必须结合形成率和选样。单凭点多点少，不能直接读出生命周期。

可以用现代概念区分两类图：一类把观测对象按实际得到的属性放入平面，另一类把模型计算的某颗星随时间路径画在同一坐标中。它们重叠时能够相互检验，却不是同一种证据。模型轨迹穿过观测密集区域是支持之一；若其预测的停留时间或其他属性不符，仍然可能需要修改。

年龄比较还可以在不先知道精确年份时提出。若两个星团的成分、消光和成员资料足以比较，而一个星团的主序在较热、较亮的位置仍有成员，另一个较早结束主序并出现相应演化成员，那么模型可以把这种形态差异解释为不同年龄。但转换成多少年，要加入质量—寿命关系和组成等条件。形态上的先后约束与绝对年龄的数字，并不具有完全相同的依赖结构。

星团整体光度校准的改进因而未必推翻所有年龄比较，却会改变精确结果；反过来，若模型改变了核心混合或燃料利用程度，即使观测图完全没有改动，推导的年龄也可能变化。这种可追踪的变化很有价值。它使“观测更新了”与“物理换算更新了”能够分别被说明，避免把图像的持久存在误当成每个解释数字也必须永远不动。

同一图中也可以同时画出固定质量的演化轨迹和固定年龄的理论序列，但二者的含义不同。前者让一颗模型恒星沿时间变化，后者比较在同一年龄具有不同初始质量的模型恒星。观测一个近似同龄星团时，后一种比较尤其直接。若误把固定年龄序列当成一颗星的旅行路线，就会把不同质量对象之间的差异误读成先后阶段。两个理论图形可以使用同样坐标，却通过固定不同条件提出不同问题。区分这两种画法，也有助于理解为什么图上最靠近某个点的一段曲线，并不自动给出该星完整的过去和未来。若要追踪个体历史，还必须限定质量、组成和质量损失等条件，不能只凭当前图上位置。

三方程在这里不应被误用为恒星演化的替代方程。恒星内部变化由物理机制研究，本书没有用 S=F(D,E)代替能量和流体方程。发生学所追踪的是，何种观测分布使演化问题显露，何种新问题重排资料搜集，以及模型与校准关系怎样被重新组织。知识发生与天体物理发生相互连接，却需要不同证据。

六路径的转换提供一个具体检查办法。从成熟演化轨迹出发看图，容易把每个分支当成必然阶段；从原始测量、距离和样本选择出发，则会先问这个分支是否稳健。前者帮助解释，后者限制过度解释。它们的协调，使理论不会只在自己画出的轨迹中寻找自己。

赫罗图的伟大之处，因而不只是图形直观。它把不同观测传统的成果放到共同关系中，使恒星的尺度、结构与生命史成为相互约束的问题。它既组织已经取得的知识，也暴露已有知识不足。新的世界不是一张装满答案的图，而是一种能够持续生成问题并判断答案的关系空间。

当恒星既有确定光度，又有需要解释的长期寿命时，能源问题便无法回避。恒星为什么能够持续发光？一种机制不仅要提供足够能量，还要与温度、质量、结构和时间相容。赫罗图没有独自回答这个问题，却让任何答案都必须面对一群具体、彼此不同而又有规律联系的恒星。

本章参考文献

[29-1] Russell, H. N. Relations between the Spectra and other Characteristics of the Stars, Nature 93, 1914，分次刊载。https://www.nature.com/articles/093227b0 ；https://www.nature.com/articles/093252a0 ；https://www.nature.com/articles/093281a0

[29-2] Eddington, A. S. The Internal Constitution of the Stars, 1926，公开原件。https://commons.wikimedia.org/wiki/File:Eddington_-_The_Internal_Constitution_of_the_Stars,_1926.pdf

[29-3] de Bruijne, J. H. J.; Hoogerwerf, R.; de Zeeuw, P. T. A Hipparcos study of the Hyades open cluster: Improved colour-absolute magnitude and Hertzsprung-Russell diagrams, 2000/2001. https://arxiv.org/abs/astro-ph/0011565

第 30 章 恒星为什么发光：从辐射到核反应

30.1 旧能源解释的有效域

太阳的光与热不是一个只需命名便能解释的事实。任何能源理论都必须同时回答两件事：每秒释放多少能量，以及能够维持多久。一个机制可以在短时间内提供足够功率，却没有足够总储备；也可以拥有巨大潜在储备，却缺少在太阳条件下以适当速度释放的过程。能量与速率不能互相代替。

化学燃烧提供了日常经验中最熟悉的发光方式，但把太阳当作一个巨大火炉，会遇到能量储备与长期辐射之间的限制。引力收缩则具有更强物理根据：物质向更紧密状态聚集，重力势能可以转化为热和辐射。它并非一个后来被核能彻底作废的错误，恒星形成等阶段仍然离不开引力能的释放。

困难在于，它能否作为长寿恒星的主要能源。以现代数量级说明，质量为 M、半径为 R 的自引力球体，其可用引力能具有 GM^2/R 的量级，具体系数依赖密度结构。再除以光度 L，便得到一个相应时间尺度。这个估算不能代替完整恒星模型，但足以说明总能量、大小与发光率必须放在一起比较。

若用太阳的量级代入，引力收缩提供的时间远短于现代所确定的太阳年龄。历史上关于地质与生物演化时间的讨论，也使这种限制变得突出。这里不应把所有早期估计说成同一个精确数字，因为假设、观测量与地质解释曾经变化。真正持续的张力，是某些能源模型难以容纳越来越长的时间要求。

这一张力不是天文学内部一项孤立误差。地球历史、物理能量学和太阳辐射测量在同一问题上彼此施加约束。一个领域不能只凭自身习惯宣布其他领域的时间全部错误；反过来，长时间需求也不能凭愿望制造一个新的物理能源。必须寻找能够同时满足这些限制的具体机制。

爱丁顿在一九二〇年的恒星内部研究中，讨论了原子内部能量及氢转变为氦的可能意义。[30-1] 这一步扩大了能源候选，却尚未等于确定了今天所知的详细核反应链。质量亏损显示可释放能量的潜力，恒星条件下是否能够以合适速率进行，则是下一层问题。

在 SDE 解释中，恒星寿命的显露由能源预算、辐射与时间证据共同规定。一种模型可以在有限任务中有效，却在新的跨度下不够用。新张力并不要求把旧机制全部删除，而是要求重新安排其有效域：引力收缩保留在适合的阶段，长期主序发光则需要其他主要机制。修正可以表现为角色重分配，而不只是简单胜负。

核能预算可以用另一个现代估算展开。氢转变成氦时，转化为能量的质量比例约为千分之七；但不能把整颗恒星的全部质量都当作能够立即参加反应的氢。可用燃料还受到初始氢比例、反应所在区域及混合过程限制。以参加燃烧的质量乘这一转换比例，再乘光速平方，才得到相应能量量级。将其除以光度，才得到与观察寿命可以比较的时间。

因此，“质能关系可以提供足够能量”仍然只是预算上的一关。若把参与燃烧的比例多算一倍，寿命估计也随之多算一倍；若恒星光度随演化增加，用今天光度直接除全部储备所得时间就只是粗估。一个可信的演化计算必须逐步累计能量释放和燃料消耗，而不是拿一份固定储备与一个不变功率相除后便宣布全部生命史。

这个区别也说明为什么大质量恒星虽有更多燃料，却不必活得更久。若光度随质量上升得比燃料储备更快，它们就会更快消耗相应储备。可用质量与消耗速率必须一起比较。具体质量范围的关系需要模型，不能把一条幂律推广到所有星体；不过，预算逻辑已经足以纠正“体积大所以必然长寿”的日常想象。

30.2 太阳内部为何难以直接看见

可见光主要告诉我们太阳外层的辐射情况，不能像透明玻璃那样直接展示核心。要研究内部，必须让表面光度、半径、质量与内部物理方程形成约束。模型不仅需要一个能源名称，还需要说明星体怎样支撑自身、能量怎样向外输运，以及不同位置的温度和密度如何联系。

静力平衡提供一项基本关系：向内的重力趋势需要由压强梯度平衡。质量连续性说明某一半径内包含多少物质，能量生成与输运关系则控制光度和温度变化。具体状态方程把压强、密度、温度与组成联系起来。这里的方程组属于恒星物理，并不是本书三方程的另一个写法。

这些关系显示，核心温度不是一个可以为了匹配反应速率任意选择的参数。如果提高温度，压强、密度与结构都会受到影响；改变组成，也会改变粒子数、状态方程和不透明度。一个局部调整可能改善能源预算，却破坏其他观测约束。恒星模型的力量来自这种整体限制，而不是来自自由参数很多。

能量输运也不能省略。辐射在不透明介质中不断与物质相互作用，能量并非从核心沿直线立即跑到表面。某些区域还可能由对流承担重要输运。于是，太阳表面今天发出的光，不应被简单看作核心此刻产生的同一批光子直接抵达。不同信号所代表的内部时间关系，需要分别理解。

这也是为什么新增观测能够改变研究能力。光谱分析约束外层状态，轨道运动约束质量，后来发展的日震与中微子研究从不同角度约束内部。它们不是把同一窗口做得更亮，而

是给不同物理量增加独立检查。多个窗口之间的一致或不一致，能够帮助判断模型的哪一部分需要重开。

为了避免混淆，可以区分“反演得到的内部”与“任意想象的内部”。前者受明确方程、边界条件和观测限制，具有可检验后果；后者只要讲得通就可以无限变化。太阳内部虽不能被普通视觉直接看见，仍然可以因多条约束而成为严格研究对象。不可直接触及并不等于不可知识化。

E 在这里具有最清楚的相互规定：发光率、重力支撑、温度、组成与反应速率不能分别任意指定。把它们列成背景因素会漏掉关键。若某一能源机制要求的温度与能够支撑观测半径的结构不相容，机制就面临问题；若它在一个质量范围适用，在另一个范围不够，就应当限定而非强行推广。

这种分析使恒星从“会发光的气体球”变成具有内部耦合过程的对象。新的科学世界不仅增加一个核心图像，还增加了能同时计算与检查多个属性的能力。其稳定性不来自画面逼真，而来自不同可观察后果之间的相互约束。

30.3 核反应路径的建立

质量与能量关系告诉人们少量质量差可以对应大量能量，却不能单独给出恒星中的反应步骤。带正电的原子核相互排斥，怎样在恒星温度下发生足够数量的反应，需要核物理与量子理论。反应截面、粒子速度分布和隧穿概率共同决定速率，不能只说“核心很热”就结束解释。

贝特与克里奇菲尔德一九三八年研究质子结合形成氘的过程，为质子—质子反应的定量处理作出贡献。[30-2] 贝特一九三九年关于恒星能源的论文，则系统讨论了以碳、氮等核素参与的循环及其天体物理后果。[30-3] 两项工作承担不同环节，不能把它们压成某一瞬间一次性发现了全部现代太阳模型。

以现代概括表达，质子—质子链使氢逐步转化为氦，并释放能量及中微子等产物。最初步骤涉及弱相互作用，其缓慢有助于解释恒星为何不是瞬间消耗所有燃料。这里的“慢”不是缺陷，而是长寿稳定发光所需的重要性质。拥有巨大能量储备与以合适速度释放，终于能够在同一物理机制中相接。

碳氮氧循环则利用较重原子核作为循环中的参与者，使总结果同样实现氢向氦的转变。不同反应通道对温度的依赖不同，因此在不同恒星质量和核心条件下，其相对重要性也不同。统一的核物理没有要求所有恒星通过完全相同路径发光，而是提供一个能够解释多样性的机制族。

阅读贝特一九三九年的原始摘要，还能看到历史模型与今天结论之间的距离。论文对太阳条件下碳氮反应的重要性作出了当时的判断，并讨论不同亮度恒星的差别。[30-3] 今天不能把现代“太阳主要由质子—质子链供能”的完整认识，毫无说明地当成该论文每一项原始结论。科学的伟大成果同样包含后来被修订的部分。

反应网络的建立需要实验与天文相互作用。实验测量相关反应，理论将其外推到恒星内部条件，模型计算总能量产生，天文观测再检查光度与结构。某些恒星能量区域难以在地面直接达到足够统计，外推就需要清楚的不确定性。实验室不必复制整颗恒星，仍能提供关键约束；但局部反应数据进入整体模型时也不是无缝拼接。

一个反应能够释放能量，不代表它能成为主要能源。若原料过少、速率过低或路径很快终止，它就只能提供有限贡献。若反应速率对温度极端敏感，还必须研究恒星结构是否会调节或失稳。能源理论因此同时涉及核素数量、反应网络和整体响应，而不仅是核物理中的一条反应式。

恒星的自调节提供一种重要联系。内部变化可能引起膨胀或收缩，进而改变温度和反应速率。这里不能用一个普遍“负反馈”标签覆盖所有恒星阶段，因为不同状态与物质性质下，响应可能不同。对于适用的稳定阶段，结构与能源的相互作用帮助解释为何恒星可以持续维持，而不只是偶然暂时发光。

核反应解释还把恒星的时间写进组成。燃料逐步转化，核心的物质与结构随之改变，演化不再只是表面温度沿分类表移动。赫罗图中的路径于是能够与内部变化联系。群体分布、能源机制和寿命获得新的共同约束，但每一条具体演化轨迹仍需模型与资料检验。

反应链的净结果与其中最慢步骤的区别，能够进一步说明核能源计算为何不是记账而已。把若干中间反应相加，可以消去在网络中产生又消耗的中间核素，得到总的物质转换与能量差。但这个相加过程消去了中间物，并没有证明每一步都足够快。若其中一个必经步骤很慢，后面的通道即使迅速，也不能越过它凭空获得原料。

碳氮氧循环中的催化角色也应这样理解。参与循环的重核并不是不发生变化，而是在多个反应中改变核素身份，再在闭合路径中得到恢复。它们的含量仍会影响整个网络的工作速率。把净反应写成氢变成氦，便于检查能量与物质收支；保留中间步骤，才便于计算何种温度和组成下能够达到所需功率。两种写法服务不同任务，不能相互取消。

30.4 新模型的新负担

核能源解释获得成功以后，太阳中微子成为新的检验窗口。中微子与物质作用很弱，可以携带核心反应的信息到达地球；难点是它们也很难被探测。戴维斯、哈默与霍夫曼一九六八年的太阳中微子搜索报告，将这种间接研究推进为具体实验结果。[30-4]

早期氯探测实验不是测量全部太阳中微子的无差别计数器。它通过特定核反应响应特定能区，并需要从极大背景和极少产物中提取信号。因此，将实验计数转换成对太阳的判断，必须结合反应截面、能量阈值、探测效率和模型预测。计数偏低首先意味着整条预测—探测关系存在不一致，不能立刻判定唯一责任方。

尤其不能从“太阳的主要能源来自质子—质子链”，直接推出早期氯实验应当主要计数该链最初反应产生的中微子。不同反应产生的中微子能谱不同，探测反应又有阈值和随能量变化的截面。低于阈值的部分，即使数量很多，也不能通过那一探测通道留下相应产物。实验所敏感的可以是数目较少、能量较高而又对核心条件更敏感的分支。

比较预测与计数时，研究者实际上需要把产生谱、到达地球后的味组成、探测截面与效率结合起来。只更改其中一个量，所得事件率就可能变化。新的探测器若增加一个对不同味具有不同响应的通道，就不仅增加统计精度，还增加分辨这些因素的能力。这正是后来多通道结果能够改变问题结构的原因。[30-4][30-5]

太阳模型、实验系统与中微子传播性质都是需要检查的方向。若预先认定恒星模型必定正确，就可能不公正地怀疑实验；若认定实验结果只有一个解释，又可能过早宣布太阳没有核聚变。成熟研究的困难，在于各环节都有支持，却共同产生了张力，需要寻找能够区分分支的新观测。

萨德伯里中微子天文台利用不同反应通道，区分电子型中微子成分与所有活跃味中微子的相应总通量。一项二〇〇二年结果为太阳中微子味转换提供了强证据。[30-5] 它的重要性不是简单“测得更多”，而是改变了探测所区分的物理量，使原来混在一起的太阳产生与传播转化问题能够被拆开。

这是一种非常具体的路径转换。从太阳光度和核模型出发预测电子中微子计数，存在多环节依赖；从不同探测通道的相互比较出发，可以较直接地限制味组成。新的 D 不是沿原路线无限延长，而是改变可比较的差异类型。结果支持保留核能源解释，同时修改关于传播与粒子性质的原先简单理解。

本章因此把太阳中微子问题作为一个反例，而不只是成功后的庆祝。新模型会生成新的可检验负担；新负担出现不一致时，修正未必发生在最初被怀疑的层次。知识的和谐并非各数据永远一致，而是能够设计更有区分力的操作，使不同环节的责任逐步清楚。

多通道检验并没有免除实验校准。不同通道的背景、响应和能量选择仍需分别估计，合并分析也要保留相关误差。其进步在于增加独立约束，使一个单一计数中的多种解释不再具有同样自由度，而不是产生一个完全没有模型与仪器条件的读数。

恒星核能源还带来其他任务：不同核素丰度如何形成，哪些阶段能够生产更重元素，质量损失怎样改变后续演化，反应速率的不确定性如何传播到年龄估计。贝特早期论文的范围

与后来完整核合成理论不能混同。一个重大机制解释了核心问题，同时也使过去没有同等明确形式的问题成为必答题。

SDE 三方程可以帮助整理这种往返，但不能替代任何一次核反应计算。恒星能源的知识显露由多种实验与观测差异、物理关系共同支持；显露一旦稳定，就改变测量目标；新的测量又可能改写模型中的关系。这里最值得保留的是修正的层次性：旧机制可以保留，某项传播假设需要改变，新的粒子知识再反过来增强恒星研究。

到这一阶段，恒星的科学身份已经远超一个发光点。它拥有可测的位置与距离、可比较的谱型、受状态约束的成分判断、群体分布中的位置，以及能够接受内部检验的能源过程。这些属性并不是后来逐个贴上的孤立标签；它们通过共同测量与物理关系相互规定，构成一个持续生成新问题的恒星世界。

这种世界既统一又多样。统一来自共同物理约束，多样来自质量、组成、阶段与环境的不同。和谐则表现为这些差异能够在明确条件下被共同解释，并在解释失败时仍有修正通道。天文学的深化并没有使未知缩成一个最后角落，而是让未知获得更清楚、更能够行动的形式。

本章参考文献

[30-1] Eddington, A. S. The Internal Constitution of the Stars, Nature 106, 1920. https://www.nature.com/articles/106014a0

[30-2] Bethe, H. A.; Critchfield, C. L. The Formation of Deuterons by Proton Combination. Physical Review 54, 248–254, 1938. https://journals.aps.org/pr/abstract/10.1103/PhysRev.54.248

[30-3] Bethe, H. A. Energy Production in Stars. Physical Review 55, 434–456, 1939. https://journals.aps.org/pr/abstract/10.1103/PhysRev.55.434

[30-4] Davis, R. Jr.; Harmer, D. S.; Hoffman, K. C. Search for Neutrinos from the Sun. Physical Review Letters 20, 1205–1209, 1968. https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.20.1205

[30-5] Ahmad, Q. R. et al. Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory. Physical Review Letters 89, 011301, 2002. https://arxiv.org/abs/nucl-ex/0204008

第 07 编

宇宙距离阶梯与星系身份

第 31 章 距离尺度：校准链、独立锚点与可迁移性

第 32 章 勒维特：周期—光度联系的建立

第 33 章 “岛宇宙”争论：星云究竟是什么

第 34 章 从速度与距离到宇宙扩张关系

第 35 章 距离再标定：为什么一个旧数值的修正改变整个宇宙

第 31 章 距离尺度：校准链、独立锚点与可迁移性

31.1 几何视差的能力边界与其他锚点

把一颗星放在多远的地方，并不是为一个已经完全确定的对象补上地址。距离参与规定它的光度、大小以及能够产生这些性质的物理过程。天文学中许多看似关于对象本身的争论，实际上首先遭遇的是距离不能确定的问题。测距困难也有一种特殊结构：研究者无法把同一把尺子从地球铺到恒星，再铺到星系；他们必须使不同的测量关系在相互重叠的范围内接合。

周年视差提供了这种接合的一类起点。地球沿轨道运动时，近邻恒星相对于较远背景的位置出现周期性改变。采用相应的角度和距离单位，理想几何关系可写为距离等于视差的倒数。这是本章用来解释方法的现代简写，实际解算还要区分恒星自行、观测者运动、参考框架和仪器误差。能够在图上画出三角形，不等于已经得到三角形所要求的可靠角度。

视差随距离增大而减小，而仪器误差不会自动以相同速度减小。于是，同样大小的位置测量误差，在近星那里可能只造成轻微的距离不确定性，在远星那里却可能支配结果。特别是视差误差接近视差本身时，直接取倒数会产生高度不对称的距离分布。带误差的负视差也不能解释成天体位于负距离；它是位置解算产生的估计值，需要在统计模型中理解。[31-1]

这里已经出现一个贯穿本编的区别：几何公式的确定性，与从数据估计几何量的可靠性，并不是同一个性质。若把公式的严格性直接转给观测结果，误差就被藏进了一个看似朴素的倒数。反过来，承认估计需要统计前提，也不意味着视差失去几何基础。测量的任务恰恰是在仪器、样本与概率表达之间，把基础关系落实为可以承担误差责任的数值。

其他几何锚点使测距网络不必依靠单一支柱。NGC 4258 核区的水脉泽系统便提供了不同于恒星周年视差的路线：空间位置、谱线运动和随时间的变化，在盘运动模型中共同限制几何距离。里德等人的研究以这一距离校准后续标尺，并与其他几何锚点比较。[31-2]称其为几何测距，不等于宣布它没有模型；盘的构型和运动解释仍是推断的组成部分。

因此，“距离阶梯”应当理解为有交叉支撑的网络，而不是人人只能逐级攀登的单行楼梯。某种方法可以跨过通常的中间层，另一种方法可能只在较窄范围内建立可靠连接。真正需要保存的是从何种观测进入、采用哪些约束、哪些误差与其他方法共享。跳过一个环节并不自动更好；若新路线以更强的物理假设换取更远的覆盖，它只是把责任移到了别处。

在 SDE 分析中，距离的显露 S 不是脱离方法的孤立数字。位置比较或运动跟踪构成 D，而参考框架、几何形状、误差模型之间的相互规定构成 E。这里的纠缠有明确含义：改变盘

的倾角解释，会同时改变运动分量与距离估计；改变位置零点，会影响整组恒星的视差，而非只影响一次读数。正是这种不能任意拆开使用的限定，使一种方法取得自己的测距对象。

31.2 光度类比的条件

几何锚点覆盖不了全部研究范围，光度关系因此成为重要桥梁。一盏灯看起来暗，既可能因为它距离远，也可能因为它原来就不亮。天文学家要从接收到的光推断距离，必须获得关于内禀发光能力的额外约束。这个约束可以来自对象类别、光变性质、光谱或经过标定的经验联系，但不能来自“它看起来像另一颗星”这一未经检验的印象。

在忽略宇宙学效应的近距离情形，光通量随距离平方反比衰减。写成星等语言，现代距离模数关系为 $m-M=5\log_{10}(d/10\mathrm{pc})$，若有消光，还须加入相应修正。公式中 m 是观测视星等，M 是绝对星等。M 的取得正是困难所在：若先假定距离算出 M，再拿同一 M 声称独立测定了距离，形式上完成了计算，证据上却没有前进。

所谓标准烛光，通常不是宇宙替观察者准备的一组完全相同的灯。更多情况下，它是一类经过条件限定后具有可预测光度的对象。预测可以存在内禀散布，且散布可能随环境改变。造父变星的周期提供了比外观更稳定的比较入口，但类型、化学组成、波段、尘埃及混叠仍会影响实际关系。标准化因此是一项有内容的研究，而不是贴在天体名称上的荣誉标签。

这也解释了为什么相似光谱与相似绝对亮度不能直接画等号。两个对象可以在部分谱线或颜色上接近，却处于不同的演化阶段，具有不同的半径和总辐射功率。分类所保留的特征是否足够支撑光度推断，需要由标定样本回答。若分类过粗，原来混合的对象在近距离时尚能分开，转到遥远而拥挤的图像上就可能被误作同一群体。

可迁移性由此成为测距关系的核心问题。某项关系在近邻样本中成立，不等于它在所有星系、所有波段和所有观测阈值下保持同样斜率与零点。要把它送往远方，研究者必须追踪目标样本与校准样本之间哪些性质相同，哪些性质不同，差异是否已经被测量或校正。忽略差异不等于消除了差异，只是让它以系统偏差的方式进入结果。

哈勃空间望远镜关键项目把造父星校准与多种更远距离指示方法连接起来，展示了这一组织结构。[31-3]其思想史意义并不限于最后给出一个宇宙学数值：它把共同标定、不同方法和系统误差作为一个整体问题处理。科学合作的规模在这里有方法论原因，因为同一测距链跨越了不同对象、仪器和分析传统，任何一个环节都不能仅凭自己的内部精度为全链担保。

可以设想一个简单的迁移检验。用近处样本确定某类变星的周期与绝对亮度关系，再把这些对象的图像降到远处观测的分辨率和信噪条件，检验同一处理程序能否恢复原来的关

系。若恢复失败，问题尚不能归给恒星物理变化；观测操作本身已经足以制造差别。这一现代方法例说明，所谓研究条件不是从对象外面附着上去的噪声清单，而是参与规定“被选中的同类对象”究竟是哪一群。

在本书的语言中，E 尤其体现在类型与标尺的互相限定上。对象被判为某类，才可以使用那条光度关系；而关系的离散与系统残差又反过来检验分类是否恰当。二者不是随意循环，因为外部锚点和独立特征可以打断自我验证。没有这些约束，分类与光度可以共同迁就数据，使任何距离都显得合理。有了这些约束，联合规定才具有被反驳的条件。

31.3 误差如何沿阶梯传播

一条距离链最容易制造的错觉，是每一层看起来都已经经过验证，于是顶端结果必然可靠。但校准链会传递可信信息，也会传递共同误差。如果一组距离全部使用同一个偏移的零点，它们彼此之间仍可能非常整齐。增加对象数量可以压低随机散布，却未必触及这项整齐的错误。越是漂亮的直线，越需要问它的坐标是否共用了一把偏尺。

设想两个研究组分别测量不同星系，都使用同一组校准恒星。即使图像获取和拟合程序相互独立，它们的绝对距离仍共享校准零点的不确定性。若把这两项结果当作完全独立的测量相乘，合并误差就会被低估。统计意义上的独立性必须沿数据与假设追踪，不能由论文作者名单不同或望远镜不同来判断。

消光展示了另一种相互牵连。尘埃通常使天体变暗，也改变其颜色。颜色因而有助于识别消光，但恒星本身的温度或化学组成也会改变颜色。若研究者把全部颜色差都解释为尘埃，就可能过度修正；若完全不修正，又可能把遮蔽造成的暗误认为距离增加。多波段观测可以提供额外约束，却仍需要检查采用的消光关系是否适用于目标区域。

样本选择则决定哪些对象有机会进入图。一个亮度受限的调查在远处更容易看见同类对象中偏亮的成员；如果用近处完整样本的平均光度直接解释这批远处样本，估计便可能发生系统偏移。这里并没有哪颗星测错，错误来自把不同筛选条件下的集合当成可直接比较的集合。测距关系的成立对象，因此既包括单颗恒星，也包括由观测规则组成的统计总体。

同样重要的是未被选中的对象。某颗变星若因周期不确定而被剔除，剔除规则可能与亮度、观测时段或背景拥挤程度相关。删去离群点也可能收紧一条关系，但关系收紧的原因可能是准确排除了污染，也可能是把反证删出了样本。研究记录必须说明何时确定标准、是否看过最终结果以及替代标准会怎样影响结论。可靠性不仅由保留下来的点提供，也由排除过程是否受约束提供。

若用三方程来组织这项问题，S=F(D,E)要求说明距离结果怎样由观测序列和校准体系形成；D=G(S,E)要求检查初步距离是否改变后续选样，例如使研究者专门寻找更亮、更远的对

象；E=H(S,D)则检查新的残差是否促成对零点、类型或消光模型的修订。三个方向没有保证每次修订都正确，但它们迫使分析者追踪误差如何进入下一轮研究，而不是只在末尾附上一项误差条。

从已经发表的距离数值逆查这条链，常能找到由起点看不出的依赖。两个末端结果的一致，可能来自真正不同的测量结构，也可能来自相同的零点；两个结果的不一致，可能来自一种方法失败，也可能因为它们测量的对象群体并不等价。六路径在这里的用处不是把研究史强行排成箭头，而是改变审查入口，使“结果一致”这一粗粒度判断分解成不同的证据责任。

独立锚点的价值正是在此显现。它不只是多提供一个数字，还提供打断共同偏差的机会。若两种方法在同一对象上相遇，可以减少对象差异造成的歧义；若在不同对象上相遇，则需要另加跨对象的迁移论证。测距网络越复杂，越不能以“很多方法都同意”代替依赖关系分析。真正有力量的是说明哪些错误必须同时发生，才能制造全部观察到的一致性。

31.4 距离为何改变身份

一团模糊光斑在天空中所占的角度，可以保持不变，而它在物理世界中的解释却随距离而剧烈改变。在可以忽略宇宙学效应且采用小角近似的近邻情形，角尺度乘上距离形成线尺度，接收通量与距离平方共同给出光度；更远处则须分别使用角直径距离与光度距离。因此，同一幅影像可能对应一团邻近气体，也可能对应大量恒星组成的远方系统。仅凭形态认出对象，往往已经偷偷借用了一个距离范围。

距离不是任意改变身份的魔法。一个更远的估计若导致荒谬的光度或大小，研究者应当同时审查测距与分类，而不能只宣布诞生了奇异对象。反过来，不能因为旧分类容纳不了新距离，就自动否定测量。身份判定需要在多种约束之间寻找能够共同成立的组合。其关键不在于哪一种信息永远优先，而在于竞争组合能够产生哪些不同的可检验预期。

这使“发现河外星系”具有超过重新贴标签的意义。新距离把原先可见的模糊结构组织成与银河可以比较的系统；系统的内部对象、尺度与总光度取得新的关系。研究者随后可以讨论星系间的分布、运动和演化，这些问题在旧的对象单位中并不能以同样方式提出。新世界并非在某个公告时刻凭空生成星体，而是新的显露、差异操作和属性纠缠形成了可持续开展研究的聚合体。

如果把这个过程只写成“人类终于知道了真实距离”，就会漏掉新对象体系如何改变研究活动；如果把它写成“距离完全由人制造”，又会抹掉各种不相容估计受到观测约束的事实。SDE 的解释任务是保存这两个方向：对象的科学身份在实践中形成，而这种形成承受不

能由意愿取消的约束。距离链是其中最清楚的场所，因为每一个数值都必须携带其转换关系。

本章由此为后面的历史提供一项判断准则：当一条新关系出现时，应当分别问它是否揭示了相对差异、是否取得绝对零点、是否能跨对象迁移。三者常在不同时间由不同研究完成。把它们合称为“一把宇宙尺子”，适合快速说明，却不适合解释思想如何发生。下一项关键成果正来自这样一种尚未具备全部条件、却已经改变了后续研究可能性的关系。

还有一种容易被忽略的尺度差别：某项测量可以可靠地区分两个对象谁远谁近，却暂时不能准确确定它们各自离我们多少。相对距离与绝对距离承担不同任务。前者可能已经足以检验群体结构，后者才允许计算绝对光度或宇宙膨胀率。若要求一切发现必须等到绝对标定完成才获得价值，就无法解释许多关系为何在早期便能引导研究；若把相对关系直接当成完整标尺，则又高估了它能够支持的结论。

因此，测距的统一性不是把所有方法化成同一操作，而是使不同操作的输出能够以明确的转换关系相接。方法多样性也不是结果越多越好，而是不同误差结构提供真正的交叉检验。两者形成和谐，需要对适用范围保持克制：视差在其能力范围内承担几何锚定，光度关系承担迁移，独立几何路线检验传递，统计模型记录共享的不确定性。没有一种方法必须包办全部尺度，科学整体仍然可以由各自有限而相互制约的能力组成。这正是距离问题给予整个天文学思想史的组织启示。

本章参考文献

[31-1] Bailer-Jones, C. A. L. Estimating distances from parallaxes. 2015. https://arxiv.org/abs/1507.02105

[31-2] Reid, M. J., Pesce, D. W., Riess, A. G. An Improved Distance to NGC 4258 and its Implications for the Hubble Constant. 2019. https://arxiv.org/abs/1908.05625

[31-3] Freedman, W. L., et al. Final Results from the Hubble Space Telescope Key Project to Measure the Hubble Constant. 2001. https://arxiv.org/abs/astro-ph/0012376

第 32 章 勒维特：周期—光度联系的建立

32.1 共同距离为什么关键

如果把天空中所有变星放进同一张亮度表，最亮者不一定具有最大的发光能力，最暗者也未必发光最弱。距离与光度纠缠在一个观测数值里，单独增加条目不会自动把它们分开。勒维特对小麦哲伦云变星的研究具有特殊力量，正因为这个样本使其中一项差异可以在一定近似下暂时压低：它们处于同一远方系统中，彼此距离差相对于整个系统的距离较小。

1912 年哈佛天文台第 173 号通报公布了二十五颗小麦哲伦云变星的周期与亮度材料。正文说明这项陈述由勒维特准备，文末署有皮克林的名字，反映了当时机构发表研究的形式。表中列出恒星编号、极大与极小亮度、历元及周期，图则比较亮度和周期，并进一步采用周期的对数表示。[32-1]这些有限而具体的材料，比后来“发现了宇宙距离尺”的一句总结更能揭示发现的结构。

共同距离并不是说二十五颗星真的处于一张没有厚度的薄纸上。它是一种需要误差尺度支持的近似：系统内部深度对视亮度造成的差别，相对于研究所考察的总体变化是否足够小？如果答案大致肯定，亮度排序就更有可能反映内禀发光能力，而不只是远近排序。由此得到的不是所有因素都已消失的纯粹关系，而是在特定条件下让一种关系获得辨认机会。

这一操作与实验中的控制变量相似，却不能完全等同。研究者没有把恒星搬到同样距离，也没有操纵其周期。她从天体分布中找到一个能够提供比较约束的群体，再借观测和记录把它组织成研究对象。自然分布、测量实践与数学表示在这里协同工作。样本的价值不只是数量多，而是它的空间联系使一个本来不可分离的问题部分变得可分离。

用现代星等关系说明，若同一群体内的距离模数近似相同，两个对象的视星等差便接近绝对星等差，但还须注意消光等差异。这样的减法可以消去共同未知量，却不能确定这个未知量本身。知道两栋楼的高度差，不等于知道它们的海拔；知道变星之间的相对光度关系，也不等于已经取得每一颗星的绝对光度。这个数学区别正是勒维特成果与后来绝对标定工作的边界。

倘若所有对象的共同距离被整体改变，表中测得的周期和视星等并不会随之改变；被改变的是赋予这条关系的绝对零点。周期与亮度联系因而可以先于绝对距离知识取得稳健意义。思想史在这里不能只按照成品工具倒推发现：早期成果虽然尚未完成全部功能，已经改变了下一步值得测什么、应当寻找哪种恒星以及如何比较星群。

在本书的 SDE 解释中，所谓新显露不是孤立的“周期与亮度相关”一句话，而是这个联系在特定样本、底片测量和比较方式中的成立。E 也不是把“小麦哲伦云”“摄影

术”和“数学”并排列出来。共同距离使视亮度差获得光度意义，摄影比较使不同时刻的变化可以归属于同一个对象，而周期表示又重新组织这些变化。每项特征的研究身份都借助其他特征获得限定。

这一点也规定了本书解释的可检验边界。假如研究者换成跨越极大距离范围的混合样本，仍用视亮度直接检验同一关系，预期离散就会显著改变；假如已知距离并做了修正，关系又可能恢复。这里能够产生差别的不是有没有使用 SDE 名词，而是有没有明确区分共同条件被消去、被保留和被重新引入的环节。没有这种区别，发生学只能复述相关性的发现。

还有一项值得明确的反事实判断：倘若样本中距离差与周期恰好相关，即较长周期对象又系统地更近，那么观察到的亮度趋势可能被几何效应增强。共同星群并不是逻辑上绝无这种可能，而是给排除它提供较有力的空间约束。后来在不同系统、不同方法下的复查，能够帮助判断关系是否只是最初样本的偶然组合。一个发现的可靠性由此随着可独立检查的场合增加而增长，不能单靠最初图形的清晰度来决定。

32.2 光变周期如何确定

周期不是刻在恒星表面的数字。望远镜在某些夜晚取得影像，天气、季节、曝光和任务安排使观测时间分布不均。研究者面对的是一串带日期的亮度记录，要从这些记录中辨认重复结构。观察到两次亮度极大，也还需要判断它们之间究竟经过一个周期还是若干周期；这要求更多观测限制不同的候选解释。

光变曲线使不同时刻的读数取得整体意义。若选择一个候选周期，把各次观测按相位排列，理想情况下重复的变化会重合为一条较稳定的曲线。但当时间取样稀疏或存在规律性空隙，不同周期也可能给出相近排列。现代研究称某些情形为混叠。本章用这一概念说明普遍测量困难，并不把现代算法或术语归给勒维特本人。

对于摄影底片，亮度比较还涉及影像强弱与记录介质响应。不同曝光条件、底片性质和背景都会改变星像的表现。一个可靠的变量序列，要求把这些技术差异与恒星本身的变化区分开来。若某颗星的亮暗仅随底片整体深浅变化，就不足以成为变星证据；若同一张底片中参考对象相对稳定，而目标反复变化，解释才受到更强约束。

所谓重复观测，因此不是机械地拍摄更多影像。新影像必须提供能够区分候选周期或校准误差的信息。一个长时间跨度内只在相近相位取到的样本，未必比覆盖完整周期的较少观测更有力。研究活动中“多”与“好”的区别，在此表现为观测序列能否切开原来的歧义。D 作为差异序列的意义，也正在于不同观测之间形成了什么辨别，而非仅有多少次动作。

勒维特的研究来自更广泛的变星调查。哈佛底片档案介绍了她 1908 年关于麦哲伦云变星的工作，并说明 1912 年新增观测使周期与亮度联系得到进一步确认。[32-2]这一连续过程

应当保留，因为它表明二十五个对象不是凭灵感突然挑出的任意集合。关系的形成依赖更大范围的发现、编号、比较和复查劳动，其结果在简短通报中被高度压缩。

然而，我们不能把档案劳动的规模本身当成关系成立的理由。大量劳动可能形成错误分类，也可能错过关键联系。真正使劳动具有证明能力的，是资料之间能否建立公开的对应：哪些读数支持周期，极大与极小如何确定，个别对象有何不确定性，新的观测能否迫使旧周期改变。承认劳动的重要性，与检查论证的有效性，是相辅相成而非相互替代的要求。

周期还有一个对测距极有吸引力的性质：在这里讨论的近邻与河外应用范围内，研究者可以从重复的时间变化中得到它，而不需要先知道绝对距离。这使它能够成为亮度推断的独立入口。若用来预测光度的量本身必须先知道同一距离，所谓新标尺便会陷入循环。周期关系的创新力量，恰在于它把时间结构与发光能力联系起来，为几何信息不足的场所提供另一种约束。

这种独立性仍有条件。周期必须属于正确的对象，观察到的变化必须没有被其他星的光严重混合，采用的变星类别必须适合关系。遥远星系中一个难以分解的光斑，可能包含多颗星；其中一颗变化时，总光的变化幅度会被其他光稀释。周期可能仍可识别，平均亮度却已经偏亮。于是，同一个对象的两个测量特征会以不同方式受拥挤影响，不能笼统归入“测得不够准”。

SDE 分析需要把这种不对称性保留下来。时间序列中的稳定周期可以给对象身份以支持，但不能替平均光度的准确性担保。一个分量稳定，不意味着整个复合测量已经稳定。只有将周期、光度、空间分解和分类的相互限制写清，才可能解释为什么一个方法在某些距离上非常有效，在另外一些观测条件下需要重做。这也是“工具形成”区别于“发现一句定律”的地方。

32.3 原始表格与图的论证

阅读 1912 年的表与图，应当先把它们当成当时能够提出的证据，而不是现代教科书公式的插图。原表同时保留极大和极小星等，图中比较相应的两组亮度，采用对数周期后呈现出便于辨认的近线性趋势。[32-1]本章不把现代的平均星等定义、消光处理或标准零点填回原图，也不将今天的拟合精度转赠给当时的样本。

一张表能够做的第一件事，是使读者检查结论有没有最直接的数据基础。如果较长周期的对象系统地更亮，相关性就不再完全依赖作者的口头宣布。第二件事是保留偏离：对象并不因为总体关系清楚，就失去自己的误差和差异。第三件事是允许重新表示，同样数据在周期轴与对数周期轴上，会把关系形态展示得不同。图像的选择参与论证，却不应掩盖原始量。

这里有一个需要谨慎处理的历史事实与现代解释之间的距离。原文能够支持“周期与亮度之间存在可利用的规律”，但不能仅凭这张图确定恒星内部的全部物理机制。关系的经验稳定性与其动力学解释可以在不同阶段成熟。若等到机制完全清楚才允许使用，许多观测工具都无法发展；若因为可用就认为机制已经证明，同样会夸大证据。

用现代说明式可把这一结构写为 $m \approx a \log P + b$，其中 m 代表特定定义下的视星等，P 是周期，a 与 b 是描述关系的系数。这不是原论文全部论证的替代品，更不是跨种群普适公式。对一个近似共同距离的样本，b 包含绝对光度关系的零点与共同距离模数。仅拟合 a、b，并不能把这两个来源自动拆开。

这一不可辨识性可以用一个思想实验看得更清楚。把所有绝对光度假定得更高，同时把群体放得更远，接收亮度可以保持不变；只要周期关系的形状仍然匹配，原表本身无法选择其中唯一方案。因此，增加同一未知距离群体内的对象，主要能够改善关系的形状和离散判断，却不能无限逼近不存在于资料中的绝对零点。数据量无法替代缺少的约束类型。

相反，如果引入一批具有独立距离估计的同类变星，便有机会把零点固定下来。但“同类”必须受到检验，不能为了完成标定而预先宣布。若两个样本实际包含不同的变星种群，统一拟合可能得到一条整齐而错误的关系。历史上后来的距离修订提醒我们，经验联系既能够成为桥梁，也可能在未经检验的迁移中传递错误。关系发现的成功不为每次应用提供免检资格。

当研究者从这张原表进入问题，与从“标准烛光”这个成品概念进入问题时，得到的判断会不同。成品入口容易让人直接问距离是多少；原表入口首先迫使人问共同距离近似是否足够、周期是否可靠、亮度差是否可比较、零点在哪里。这种路径转换并非增加一套装饰性术语，而是改变了缺项能够被看见的位置。

同样，不能因图中能够画出近似直线，就把线性看成宇宙必须服从的审美规范。线性表示往往有操作优势，便于标定与外推，但它的适用范围需要更多样本检验。若扩大周期范围或改变波段后出现曲率，应当修改关系表达，而非优先判定对象异常。科学中的和谐不等于强迫所有差异服从最简单形状，而是让表达复杂度与已经掌握的差异相称。

本书从这里提出一个更一般的解释：新知识的发生有时不是增添一种孤立事实，而是重新分配已有差异的角色。亮度差原来混合了距离与发光能力，如今在共同距离的条件下成为光度差的线索；时间变化原来使对象不稳定，如今成为辨认和测量它的条件。原来妨碍分类的“变化”，反而成为使远距离对象可以比较的尺度来源。这种角色重组是本章希望解释的创造性所在。

同样应当区分测量精度的改进与问题结构的改变。把星等读得更精细，可以改善已经成立的比较；但若没有意识到共同距离的重要性，再精细的混合样本也可能看不清光度联系。

创造性有时发生在测量之前的研究组织，有时发生在记录之后的重新表示。勒维特的案例提醒我们，比较并非完全中性的工具，比较对象如何组成，决定了哪些特征有机会稳定显露。

绝对尺度尚未确定时，关系仍然能够承担相对测量；这项能力后来容易被成功的绝对标定遮蔽。若两个相近周期的对象在控制其他条件后具有不同的视亮度，研究者已经可以对它们的相对距离提出问题。零点的缺失限制某种答案，并不抹去所有信息。因此，知识发生并非等到整条推断链完成后才突然从无到有；不同关系可以先后取得不同程度的可用性，而每次应用都应清楚说明自己用了哪一层。

32.4 从规律到距离尺的漫长回写

一条经验关系离开原来的星群以后，会遇到此前被共同条件暂时压低的问题。距离不再相同，尘埃不再相近，摄影响应可能不同，目标又可能处于更拥挤的背景。它要成为距离尺，就必须把这些差异转变成可以测量、限制或纳入误差的条件。后续工作的价值因此并非仅仅“应用勒维特定律”，还包括重新打造使该关系在新场所成立的互动系统。

独立几何锚点是这项改造的重要部分。NGC 4258 的脉泽距离研究为光度测距提供一条外部校准路线；有关该星系造父星的研究同时考察消光规律与金属丰度等系统影响。[32-3][32-4]在这里，几何和光变两种方法并不相互取消，而是在同一星系中相遇。其价值是让不同依赖结构互相约束，从而检验由本地关系向更远距离传递的可靠性。

回写也会改变观察对象的选择。知道周期与亮度的联系之后，研究者能够把周期性变星作为远方星系的重点目标，安排时间观测来覆盖其变化，而不只拍一张尽可能深的影像。原来发现的 S 改变了后续 D；新仪器和分析则扩大可识别对象的范围，并重新限制 E。一个结果的历史力量，在于它使新的研究序列成为合理投入，而不只在书上多占一行。

但回写并非只能带来进步。过度依赖某类对象，也可能让其他测距路线被忽略；用既有关系筛选变星，可能偏好那些更符合关系的候选者。若再拿筛选后的紧密关系证明原关系无误，就形成了自我增强的封闭。解决办法不是拒绝使用已知知识，而是在关键处保留不同路线、不同样本及事先明确的筛选规则，使知识既能引导研究，又能受到研究的反向制约。

对勒维特工作的评价也因此应当具有层次。她没有在 1912 年完成现代距离阶梯的全部绝对标定，却建立了一种改变后续标定可能性的关系。强调这一边界不会削弱贡献，反而使贡献更准确：一个在特定群体中成立的比较结果，可以在其他人的工作中被转换为更广的测量能力。个人的创造与共同体的后续劳动，具有不同但可以连接的证据身份。

这种连接还要求保存劳动归属。观测、底片制作、测量、关系辨认、绝对标定、远方应用并不是同一件事。把全部成果归给最后使用工具的人，会抹掉工具的发生；把全部后续结果归给最初发现关系的人，也会抹掉迁移与校准的创造。更合适的历史书写，是追踪每个环

节新增了什么约束、消除了什么歧义、开启了什么操作，而不只为一个总成就寻找单一姓名。

从一般的统计解释看，勒维特工作已可被说明为控制距离混杂因素之后发现相关关系。SDE 不能因为自己使用新的术语，就否认这一解释的充分部分。本书的补充在于继续追问：相关关系怎样改变研究对象的组织，怎样引出绝对标定的缺项，怎样在迁移时重组属性的相互规定。如果这些分析不能产生不同的证据审查，本书就没有理由宣称取得了额外解释力。

由此形成的天文学世界具有新的构成能力。一颗不断变亮变暗的星，不再只是不稳定的例外；它可以成为连接时间结构、恒星物理和宇宙尺度的节点。新的对象身份来自这些属性在互动中的聚合，而不是从变星名称中自动展开。也正因为如此，这个世界可以继续修订：周期关系的稳定部分被保留，种群与零点的错误被打开，远方星系由此取得更可靠的距离。

本章参考文献

[32-1] Leavitt, H. S.; Pickering, E. C. Periods of 25 Variable Stars in the Small Magellanic Cloud. Harvard College Observatory Circular 173, 1912. 已核对原件三页与两图。 https://articles.adsabs.harvard.edu/pdf/1912HarCi.173....1L

[32-2] Harvard Plate Stacks. Variable Stars: Henrietta Swan Leavitt. 机构档案说明。 https://platestacks.cfa.harvard.edu/henrietta-swan-leavitt/variable-stars

[32-3] Humphreys, E. M. L., et al. Toward a New Geometric Distance to the Active Galaxy NGC 4258. III. Final Results and the Hubble Constant. 2013. https://arxiv.org/abs/1307.6031

[32-4] Fausnaugh, M. M., et al. The Cepheid distance to the maser-host galaxy NGC 4258: Studying systematics with the Large Binocular Telescope. 2014 预印本。 https://arxiv.org/abs/1412.2138

第 33 章 “岛宇宙”争论：星云究竟是什么

33.1 银河大小如何测

在望远镜中看见一个旋涡形状，并不能立即知道它是银河内部的一团物质，还是另一个可与银河相比的恒星系统。两种解释都可以容纳某些外观特征，却把对象放进完全不同的尺度。二十世纪初关于星云的争论，因而不是一方有想象力、另一方缺乏想象力的简单对立。它关联着银河有多大、太阳在哪里、恒星光能传播多远，以及哪些亮度比较可以跨系统使用。

1920 年的沙普利—柯蒂斯讨论以“宇宙的尺度”为题，后来发表的文本展示了双方对多个子问题的论证。[33-1]阅读这些材料首先应当拆开问题，而不是提前安排胜者。一个研究者可能对太阳偏离银河中心的判断具有重要洞见，同时对旋涡星云的距离持错误意见；另一个研究者可能正确支持外部星系，却低估银河的尺度。这并不矛盾，因为几项判断使用的证据并不完全相同。

确定银河大小面临一个从内部观察系统的困难。我们不能退到银河之外拍一张完整照片，只能从不同方向的恒星计数、星团分布和距离估计重建结构。但看不见某处的星，既可能因为那里缺少恒星，也可能因为星光被吸收，或目标低于观测限度。空间分布与可见性相互牵连，使直观的“向哪个方向看得更多”不能直接等同真实的物质分布。

球状星团提供了一种不同的取样。若能够估计它们的距离和空间分布，就有机会推断银河相关结构的中心与范围。但这一方法也要回答：星团是否适合标识所要研究的银河结构，距离标尺是否可靠，空间取样是否完整？从一种对象分布推断另一个整体的几何，需要物理与统计的桥接，而不是只把星团位置连成一个轮廓。

以现代分析语言说，银河“边界”本来就不是只有一个定义。明亮恒星盘、稀薄晕、气体和引力影响范围可能给出不同尺度。历史参与者未必使用今天的分量划分，但我们在比较旧估计时更应避免混合。两项数值相差很大，有时是标定错误，有时还包含所指对象不同。思想史必须先确定当时量词的实际含义，才能判断争议究竟发生在哪里。

SDE 分析在这里把银河大小看成一个由观察与对象构成相接而成的问题。被测量的整体 S 不是先由眼睛完整交付，随后才加上数字；星团距离、星数分布、消光解释与结构模型共同组织这个整体。E 的相互规定表现为：若遮蔽改变可见星分布，对中心位置的推断就必须改变；若中心与尺度改变，原来用来判断星云是否“在内部”的范围也随之改变。尺度判断会进入身份判断，不能独立存放。

因此，大银河和岛宇宙不是纯逻辑上的互斥命题。银河可以比某些研究者估计得大，而旋涡星云仍可以位于更远处；银河小也不会自动证明全部星云都是外部恒星系统。只有具体距离与结构证据，才能把这些命题连接或分开。这个朴素的逻辑区别，正是防止历史被压缩成一次二选一辩论的关键。

33.2 可分解恒星与星云结构

把星云分解为一个个恒星，似乎提供最直接的身份判据。但“没有分解出来”和“其中没有恒星”之间存在巨大距离。分辨能力、曝光深度和目标距离共同决定影像中能显示什么。一个由大量恒星组成的遥远系统，完全可能表现为连续光斑。若把仪器的未分辨状态当成对象的无结构状态，就把观测能力边界错误地写进了天体本体。

反过来，看到若干星点也不能立即证明它们全属于星云。前景恒星会投影在远方结构上，背景对象也可能混入。成员身份需要由空间分布、光变、光谱或其他约束建立。影像中的重叠是一个二维关系，物理系统中的共同成员则是三维乃至动力学关系。天文学家必须寻找能够在这些层次之间架桥的信息。

哈勃 1925 年对 NGC 6822 的研究，以及 1926 年对 M33 的研究，把可辨认的恒星与变星带入远方系统的距离判断。[33-2][33-3]这些论文提供的转折不是简单看见更多亮点，而是把某些亮点组织为可测量、可分类、能够承载距离关系的对象。只有当内部对象的性质与整个系统的空间身份相联系，分辨能力才转化为尺度证据。

把内部变星变成整个系统的距离，还需要一个成员群体。单个亮点可能有前景身份问题，若若干变星在同一片结构中呈现彼此相容的周期与亮度关系，共同距离便成为可以比较的解释。这并非把数量当作真理，而是让误认假说承担更具体的代价：它需要说明为什么多个彼此独立观测的变星，会同时表现出适合某个共同系统的关系。

但“多个”也须经过选择检查。如果观察者只保留那些恰好符合预期关系的对象，再以紧密关系证明它们同属远方系统，就可能由筛选制造一致。较可靠的做法，是把候选如何进入样本、哪些周期不确定、哪些亮点可能混杂等信息保留下来，让共同距离的支持程度能由他人评估。一个新对象世界的稳定性，依赖这种允许局部不整齐的记录方式。

新星亮度的比较则提示另一种风险。同一名称下若混有物理上不同的爆发现象，把某个异常明亮事件当成全部同类的代表，就会扭曲距离推断。历史争论的重要性之一，正在于促使研究者重新问“这一类对象真的可比吗”。距离标尺的使用与类别划分由此相互牵动，而不是先有一个永不变化的分类目录，测量只负责在目录后填写数字。

这也说明，仪器进步不会自动完成理论转向。更大的望远镜可以让新对象进入可见范围，但对象意味着什么，仍取决于如何跟踪和比较。拍出一颗变星的亮点，与取得足够光变

记录以确定其周期，是不同任务。若没有后者，影像可能只扩大目录；有了后者，它才可能把原来无法确定的距离转为有约束的推断。

旋涡形状本身同样存在多义性。形态是一种可比较的显露，但相似显露未必意味着相同的构成机制与尺度。在尚缺少可靠距离时，研究者可能借太阳系形成、气体运动或恒星系统类比来解释旋涡。历史上这些类比提供了探索方向，也可能制造过度迁移。判断一种类比是否有效，不能只看图像像不像，而要看它是否给出能够被其他观测检验的差别。

这里可以提出一个独立于具体历史人物的检验：若某旋涡系统被放在非常远的距离，而又声称看到了显著的整体角运动，那么两项判断可能合在一起要求极高的线速度。这种冲突应当促使人复核角运动、距离和运动模型，而不能只靠支持其中一个结论的权威解决。证据之间互相限制，正是对象身份逐步稳定的过程。

这一类冲突也提醒我们，不应把当时曾被重视、后来却失效的测量仅写成笑话。对历史参与者而言，具有专业来源的结果会进入可信度判断；错误的难点可能在于它与其他知识结合得过于顺畅。只有恢复当时可用的检验渠道，才能解释为什么某项结果获得力量，又为什么后来必须退出。后见之明如果只留下错误标签，就无法教人识别下一次相似错误。

在 SDE 的框架中，分辨率改变的是特征能够以何种粒度显露，而对象类别的改变还要求这些特征与距离、光度和共同成员关系形成新的纠缠。因此，从“星云”到“星系”的过渡既不是纯视觉升级，也不是纯概念替换。它是不同层次的特征重新聚合，使原来的模糊对象取得新的内部单位和外部比较尺度。

33.3 争论者的不同问题

辩论最容易被写成两套完整世界观的正面碰撞，但实际证据常常穿过阵营边界。沙普利和柯蒂斯所讨论的银河尺度、太阳位置、旋涡星云性质等问题，既互相关联，又不具有统一的胜负条件。[33-1]如果用后来形成的“银河只是众多星系之一”宣布所有相关争议已经由同一方获胜，就会抹掉真实认识发展的复杂性。

更细致的读法，是把每个主张与其承重证据配对。关于太阳位置的判断，可以依赖某类天体的空间分布；关于旋涡星云距离的判断，可以依赖其中新星的亮度或其他可识别对象；关于银河大小的判断，则涉及不同距离尺度的校准。证据可能在一处可靠、在另一处失效，不能因为同一研究者提出就被捆成一个不可分割的包。

这种拆分不是削弱理论整体性，而是为整体性提供可修订结构。一个体系若只有全部接受与全部拒绝两种状态，就很难吸收局部新证据。把相对独立的命题分开，研究者才可以保留成功的距离关系，修订对象类型，或重新估计某种消光影响。科学世界的变化往往发生在这些接合处，而非每次都从零开始重造一切。

文献的身份也参与我们今天怎样重建争论。现存常被称作“大辩论记录”的公开文本，发表于 1921 年；它不等于把 1920 年会场中的每一句话原样抄下。[33-1]发表稿允许作者安排论证、补充说明并面对书面读者，因此更适合用来研究公开主张与证据组织，不能未经其他史料便用于恢复现场的语气、即时反应或谁使谁当场信服。

这项区分与天文学中的对象辨认有一种有限的类似：资料的名称并不能保证它承担所有读者希望它承担的证明。原始文献确实比传闻更接近研究实践，但不同文献具有不同生成条件。认真保留这些条件，能够避免在赞扬科学证据纪律的同时，对科学史自己的证据放松要求。

“宇宙”一词本身也应受到语境限制。在一种讨论里，它可能指全部恒星系统，在另一种讨论里，它指所有能够设想的天体存在。银河尺度的扩大，与全部物理宇宙范围的扩大，不是同一推断。历史叙事若不断把不同含义替换，读者就会得到一种虚假的戏剧效果：每次望远镜改进都仿佛直接改写了存在的总边界。

本书关注的世界构成有更具体的尺度。原来被组织为银河内部云状物的对象，若取得独立恒星系统身份，研究者便获得了新的比较单位；银河也相应成为可以与其他系统对比的一员。改变的不只是一个距离值，而是“整体与局部”“系统与成员”的关系。新的科学世界能够容纳星系群、星系分布和系统间运动，这些概念需要新的对象单位才能稳定运行。

然而，不是任何分类修改都具有同样深度。某个对象由一种星云亚类移到另一亚类，可能只是目录层面的调整；若对象单位、测量方法和总体关系同时改变，才形成更深的重组。把所有修订都称为革命，会使革命失去区分力。SDE 分析应当以改变了哪些属性的成立条件来判断，而不是以使用了多少宏大词语判断。

这种方法也有助于公平呈现错误。某人判断错了旋涡星云的总体性质，不意味着他对所有资料的质疑都没有价值。他指出的距离误差、样本差异或方法循环，可能仍是后来研究需要解决的问题。正确结论也可能暂时建立在不充分理由上，随后才获得新的证据支持。史学评价因此必须同时考察结论和理由，不能把今日答案作为唯一评分表。

统一性、多样性与和谐性在这一章有非常具体的含义。统一性来自所有命题都要承受可追踪的证据限制；多样性来自不同测距和分类路线保有自己的能力；和谐性则来自它们能够在共同对象上相互校验，而不是被迫说同一句话。一个成熟的研究共同体并不需要每位参与者先接受相同宇宙图景，才能累积可共同使用的观察。

33.4 从辩论到可操作检验

让争论发生实质变化的，是能够把竞争解释转成不同观测责任的实践。若旋涡星云位于银河之外，其中可识别的变星应当给出与远距离相容的亮度关系；若它位于银河内部，就需

要说明这些对象为何表现出另一套尺度。新的测距不只在两种意见之间投票，而是改变双方必须共同回答的问题。

哈勃对远方恒星系统的论文表明，距离可以通过内部对象的持续研究获得支撑，而不只来自整片模糊光的总体类比。[33-2][33-3]本章不把当时的距离数值当成今天的精确结果。后来标尺修订会改变数值，但某些对象属于银河外恒星系统的结论仍可保持。这正好说明身份判断与精确参数估计具有不同的稳健性层次。

一个结论可以在大范围参数修订中保持，是因为支持它的关键差别大于修订幅度，或者存在多种相互制约的证据。反过来，若两个候选身份只在非常窄的距离界限上区分，同样的校准修订就可能翻转结论。思想史应当说明稳定性的来源，而不能笼统宣称“科学总在纠错”，仿佛任何改动都自动证明整体可靠。

从六路径的角度，由“河外星系”这一已成结果反查资料，会让我们看见哪些证据被后来叙事保留，哪些已失效的论据被隐藏；由当时的测量过程出发，则会看见不同结论为什么都曾具有吸引力。两条路线的交叉，有助于避免把历史人物写成知道答案却故意走弯路的人，也有助于防止把过去的不确定性夸大为任何解释都同样合理。

这段历史最终留下的，不只是宇宙比原先想象得大。距离、内部结构与类别形成了新的联合规定：星系成为一个能够跨对象比较的单位，银河的特殊地位相应改变。新的显露又引导研究者去测量这些系统的光谱与运动，从而使空间分布问题进入动力学关系。至此，尺度链开始与另一条独立发展的测量链相遇，宇宙扩张的经验问题才获得了实际入口。

本章参考文献

[33-1] Shapley, H.; Curtis, H. D. The Scale of the Universe. Bulletin of the National Research Council, 1921. NASA 转载发表文本，并非现场逐字记录。 https://science.nasa.gov/apod/nrc-transcripts-of-the-great-debate/

[33-2] Hubble, E. N.G.C. 6822, a Remote Stellar System. Astrophysical Journal 62, 409, 1925. https://articles.adsabs.harvard.edu/pdf/1925ApJ....62..409H

[33-3] Hubble, E. A Spiral Nebula as a Stellar System. Astrophysical Journal 63, 236, 1926. https://articles.adsabs.harvard.edu/pdf/1926ApJ....63..236H

第 34 章 从速度与距离到宇宙扩张关系

34.1 勒维特与距离条件

一张横轴是距离、纵轴是速度的图，容易让人忘记它的两条轴来自不同的历史。距离需要从星光、类别与校准关系中推断；速度则依赖谱线辨认、波长比较和运动解释。只有当两种研究能够指向同一批对象，并把误差带进同一分析，它们才可能形成宇宙尺度上的经验关系。宇宙扩张问题的发生，首先是一场测量体系的汇合。

在 1920 年代，这种汇合已有重要条件，却远未具有今天的精度。造父星关系使某些近邻星系能够获得距离估计，其他对象则需要依赖更间接的光度准则。哈勃 1929 年的论文明确把距离的数值基础与造父星周期—光度关系的零点联系起来，同时区分不同距离估计的可靠程度。[34-1]因此，他的图不能被想象成二十四项同质、等精度的直接测量。

把这种差别保留下来，才能理解后来的距离修订为什么改变斜率，却没有简单抹去整个关系。某些系统误差可以近似整体伸缩横轴：如果距离普遍偏小，相同速度所对应的斜率便偏大。修正标尺会改动膨胀率的数值，却未必消除远处对象总体具有更大谱移的趋势。另一些对象依赖的测距误差不相同，则会改变散布乃至关系形状。两类情况不能混称为“数据有误”。

原表内部的层次能够把这一点说得更具体。哈勃把二十四个对象的距离分成可靠程度不同的几部分：前七个较有根据，接下来的十三个更多依赖所能辨认的亮星之光度上限，末四个则按室女座星系团成员给予共同距离。[34-1]从造父星到最亮星再到星系团归属，距离不是沿同一种操作均匀延长的。距离越远，个别恒星越难辨认，研究者便更依赖群体性质；而群体性质是否可以跨对象迁移，恰好又需要较近对象提供检验。

这形成一种很实际的证据不对称。近处对象的距离往往更有把握，但局部运动在总径向速度中所占比例可能较大；远处对象更适于显示总体趋势，却更依赖间接距离准则。扩大研究范围并不意味着简单地把所有误差一同减小。研究者必须把两种困难交换着处理，并检查这项交换是否真的提高了对总体关系的分辨能力。

共同的星系团距离还引入另一项限制。四颗成员星系画成四个点，能够提供四次速度观测，却未必提供四次独立的距离校准。若星系团的距离有共同偏差，四个横坐标便一起移动；若该团相对总体流动具有共同运动，四个纵坐标也可能一起偏移。图上点数所显示的丰富程度，不一定等于独立信息的丰富程度。后来以协方差语言处理的问题，在这里已经有了可以直接辨认的对象结构。

距离首先还承担身份判定。若对象不是外部恒星系统，或者同名对象的不同成分被混用，其光度、光谱和运动就未必可以组成同一条宇宙关系。星系身份的稳定化使距离表与速度表能够较可靠地联结，这是上一阶段研究为本阶段提供的条件。新的整体关系没有凭空绕过对象分类，而是使用了分类已经建立的可比单位。

但对象单位建立以后，也不能立刻把星系看成静止的试验粒子。星系具有局部运动，群体内的引力相互作用会影响其径向速度。近处对象的这些运动可能与总体膨胀贡献相当，甚至占据优势。对于关系的初步检验，研究者因此需要处理散布，而非期待所有点严格落在一条直线上。离群点既可能提醒测量问题，也可能反映对象真实的局部动力学。

这里的 E 是一套相互规定的关系：造父星零点限定距离轴，星系成员身份限定如何把不同观测配对，太阳运动与局部运动解释限定速度轴。它们共同决定“距离—速度关系”究竟指向什么。只列出望远镜、研究人员和时代背景，尚不能显示这种纠缠；必须指出某个条件改变时，哪个推断跟着改变，哪一部分仍保持。

从这一入口观察，勒维特的贡献并不因她没有直接提出宇宙膨胀而缩小。她的关系改变了远距离测量的可能性，而绝对标定与星系应用又增加了新的条件。贡献的联系可以非常紧密，证据身份仍须分开。把后来形成的宇宙图景全部倒写成早期发现的内涵，会使历史失去真正的生成过程。

34.2 斯里弗速度如何测得

在距离轴逐渐成形的同时，星云光谱提供了另一条路线。谱线的位置可以与实验室或其他参考谱线比较，从波长差推断沿视线方向的运动。弱光目标要求足够长的曝光和稳定的仪器，谱线识别又要承受噪声及线宽的限制。这里测到的不是星系在空间中移动的录像，而是一组需要解释的波长位移。

斯里弗 1913 年发表的仙女座星云径向速度研究报告了负值，即按其运动解释指向接近而非远离。[34-2]这一事实很重要：历史不能为了讲述膨胀而把蓝移对象删去。总体关系与局部接近并不必然矛盾，尤其在邻近系统中；但这种相容性需要动力学说明，不能只是事后为理论免责。

我们用现代低红移近似 $v \approx cz$ 说明波长位移与速度的联系，其中 z 由观测波长相对于参考波长的比例变化定义。这个近似有适用条件，不能把它无限延伸到任意红移，更不能把每个红移直接视为普通欧几里得空间中的退行速度。宇宙学解释涉及时空模型，本章先保留经验资料与其动力学含义之间的层次。

速度测量还必须选择参考系。地球的运动、太阳相对于对象群的运动，会在天空不同方向产生不同投影。如果不同目标的数据没有一致处理，方向性变化可能混入所研究的距离趋

势。对近邻样本而言，这些改正尤其不能忽略。所谓一张统一的速度表，实际上包含了一套把不同观测位置和时刻转成可比较量的操作。

当谱移资料开始积累时，红移的普遍性提出了问题，但尚未自动回答原因。星云总体运动、引力解释以及相对论模型等都可能进入讨论。若没有独立距离，仅有速度分布仍难以形成可靠的速度—距离规律；若只有距离而缺少速度，也无法直接判断宇宙是否动态。两条链各自取得成功，尚需在共同对象上发生交叉，新的整体显露才可能成立。

这种交叉强调了知识协作的实际内容。一个研究者可以在论文中使用另一位研究者的速度资料，再加入自己的距离判断和统计处理。最终关系的署名不应遮蔽测量来源，也不能因此否认综合工作本身的创造。新成果常常正发生在资料原来没有连接的位置：不同量一旦进入共同坐标，便提出了各自研究中未必显现的规律性问题。

如果从 SDE 路径看，先由谱移异常进入，会优先追问运动来源及参考系；先由距离标尺进入，会优先追问分类和零点；先由理论解进入，则会寻找允许这些量共同变化的整体模型。三条入口并不保证同样快到达结论，却能暴露不同的缺项。它们的和谐并不是消除差异，而是让各自的限制在同一推断中保持可见。

34.3 勒梅特 1927 与哈勃 1929

勒梅特 1927 年的论文把相对论宇宙模型与当时的星云观测联系起来，提出随时间增长的尺度，并估计与观测相关的比例参数。哈勃 1929 年的论文则以距离与径向速度的经验关系为明确研究对象。[34-1][34-3]这两项工作不能合并成一件发生在一个瞬间的发现，也不宜通过夺走其中一人的全部贡献来恢复另一人的贡献。

两者的研究入口不同，证据组织也不同。由模型进入，首先面对方程允许什么宇宙，以及那些解如何对应可观测量；由资料关系进入，首先面对距离估计、速度资料和相关性的可信度。一个入口可能使解释先于高质量数据，另一个入口可能使经验规律先于明确承诺其最终机制。思想史应当比较这些工作的具体增量，而不是只寻找谁最先说过某个后来熟悉的词。

勒梅特的理论联接意味着红移与距离能够获得宇宙整体演化的解释；哈勃的经验工作增强了研究者必须认真面对该关系的理由。二者具有相互促进的可能，但不能假定全部参与者立刻接受同一套本体说明。经验关系可以在争议中继续使用，动力学解释也可以等待更强的观测支持。科学共同体的统一常常先发生在问题和可比较资料上，而非立即发生在所有解释上。

1931 年英文译本删去了 1927 年原文的部分数值讨论，曾引出关于归属的猜测。利维奥 2011 年查阅档案后指出，相关书信支持由勒梅特本人作出删节的解释。[34-4]这项史料发现

使“有人蓄意抹去成果”的叙事失去原先依据。它也说明科学史不能依靠文献差异直接推断动机，版本比较之后还需要书信与出版过程证据。

关于该事件，本章只保留已得到档案研究支持的判断，不以现代心理想象补写勒梅特的内心。一个作者认为某段暂时性的讨论无需重印，不等于他否认自己的早期贡献；同样，某段内容未出现在流传较广的版本中，也可能改变后来读者对历史的认识。内容成立与内容传播是两件相互影响但应分别研究的事情。

由此可以看到另一类回写：后来的教科书和名称会反过来组织人们阅读早期文献的方式。若读者已经确信某关系只有一个发现者，他可能在阅读时只寻找符合这一框架的段落；当新的档案出现，人物归属和问题发生过程才重新打开。知识传播不仅携带结果，也携带压缩方式，而压缩本身可能成为需要修订的研究对象。

这并不意味着命名没有价值。简短名称可以帮助交流，但不能承担全部历史说明。无论如何称呼距离—速度关系，都应保存其观测来源、距离标定、数学解释与版本传播的多条线索。名称是索引，不是证据链的替代物。对学习者而言，知道这条关系怎样形成，比只记住一个姓名更有助于理解后来为何还能修订其数值与适用范围。

贡献归属还需要区分“提供了必要条件”与“已提出后来结论”。没有可靠测距，速度资料难以形成尺度关系；但提供测距条件，并不意味着同时完成了扩张解释。没有光谱资料，理论也难以受到相应检验；但首先获得一批谱移，同样不等于已经确立某个宇宙模型。承认这些区别，才能既防止综合者遮蔽上游劳动，也防止把全部上游劳动自动提升为同一发现。

这种区别不只关乎公平，还影响我们解释发现为何在某时发生。若每一个先行环节都被写成已包含最终答案，最后的汇合便似乎无需任何新的创造。实际上，把来自不同目的的资料放到同一对象表中，判断其可比性，再选择一个能够揭示关系的数学形式，都可能增加此前没有的研究内容。综合工作的价值并不建立在抹去来源之上，来源越清楚，其真正新增之处才越容易辨认。

同样重要的是没有被立即统一的部分。一个研究者可能接受近似经验关系，却对其最终解释保持犹疑；另一个研究者可以较早接受动态模型，却无法提供足够精确的数值。思想史不必为这种不同步感到不耐烦。它们显示结论的各个层次有不同的成熟时间，也解释了为什么后来的共同知识会由当时并不完全一致的研究共同促成。

从发生学角度，这一事件尤其说明新世界的出现可以具有分布式结构。有人改进标尺，有人测得谱移，有人建立动态模型，有人把资料组织成公共关系。不能因为各环节分散，就否认整体创造；也不能因为整体最终形成，就把每个环节写成早已朝同一目标执行。世界构成在这里包含真实的汇合，而汇合之前存在不同的问题意识与失败可能。

34.4 有限样本的数理结论

哈勃 1929 年论文开头区分了已有径向速度的四十六个河外星云与具有个别距离估计的二十四个对象，并同时讨论逐个对象与分组处理。[34-1]这些数字不是装饰性的历史细节，它们规定了结论能够承受的范围。速度样本数量不能直接当成距离—速度配对样本数量；分组后的点也不能再被当作新增的独立对象。

如果把一批点分组取平均，局部随机运动可能被部分平滑，关系看起来也可能更清楚。但分组规则可能影响结果，且组内共享的距离或环境信息不会因此消失。一个图上散布减小，究竟是统计效率提高，还是处理方式隐藏了重要差异，需要比较原始对象与分组结果。统计图形提供的是可检查的论证，不是免于检查的视觉裁决。

原文的拟合还同时处理了太阳运动。用现代排版转写其形式，可以写成 $rK+X\cos\alpha\cos\delta+Y\sin\alpha\cos\delta+Z\sin\delta=v$，其中 r 是距离，K 为比例系数，后三项表示太阳运动在目标方向上的投影，α、δ 指定目标在天空的位置。[34-1]公式使两类有序变化处在同一论证中：一类随距离改变，另一类随方向改变。若不把方向项分开，天空覆盖的不均匀就可能侵入对距离项的判断。

这也解释了为什么“先修正速度，再画直线”并不总是一套彼此完全独立的步骤。当修正参数也从这批对象求得时，距离分布、方向分布和运动模型共同参与拟合。某一类对象恰好集中在某个方向，便可能使不同参数互相补偿。思想史的细读应当把公式当作研究责任的说明：它明确承认了哪些变化，哪些变化仍留在残差里，而不只是把 K 当作后来常数的早期数值。

哈勃比较了二十四个对象逐个处理与合并为九组的处理，并意识到最远处的室女座成员会对斜率产生较强影响。[34-1]用今天的语言说，最远端的点拥有较大的影响力，因为它们在横轴上更能决定直线应当有多陡。因此，一项最有说服力的后续工作应是在更远距离增加独立对象，而不是只在原来近邻范围内把点画得更密。新观测的价值由它能消除哪种不确定性决定，不能仅由数量决定。

原文还利用没有个别距离估计的速度对象，进行关于平均光度或距离分布的比较。这里需要特别辨认推断方向。如果先使用待检验的距离—速度关系从速度估算距离，再发现所得绝对光度分布并不荒谬，这属于相容性检查；它可以暴露明显矛盾，却不能被计作一套完全独立的测距验证。真正独立的检查应当让距离尽量由另一条不借用该关系的路线取得。两类检查都有作用，但承担的证明责任不同。

这一点不要求对早期工作作苛刻的事后指责。初步关系刚出现时，研究者有理由利用它探索更广泛的对象，看看会产生什么结果。问题在于后来的历史叙述是否把探索所得全部改写为最初结论的独立前提。若把时间方向打乱，一个先由关系推出、再用于检查的量，就会

看似从一开始已经证明了关系，论证于是被写得比实际更封闭。保留先后次序，才能看见假设如何试用、接受约束并获得进一步研究资格。

有限距离范围内的一阶关系也不等于完整的宇宙演化史。许多光滑函数在足够小的范围都可以近似为直线。要判断更远处的曲率、宇宙减速或加速，必须取得更广的观测并引入适当的模型表达。把早期近线性关系读成已经测出了整个宇宙从起点到未来的历史，是把后来的理论结构倒灌给最初资料。

对这段历史最有价值的现代复查，不是用今天的参数重画一张更漂亮的图后宣布过去幼稚，而是区分三个层次：原来是否存在值得解释的经验趋势；趋势的数值斜率受何种校准影响；趋势在何种模型下支持宇宙扩张。三个问题可以获得不同强度的答案。保留它们的差别，既避免夸大神奇的一次发现，也避免因旧数值不准而否认研究的创造性。

这里还存在一个重要的可识别性问题。距离轴若整体伸缩，斜率随之改变；若速度改正带有方向依赖，样本在天空的分布也可能影响结果。于是，提升关系的可信度不能只靠继续收集相似对象，还应扩大方向、距离与方法范围。新的 S 把 D 推向更广的调查，同时也提出 E 必须重新校验的要求。这种回写具有具体任务，而非抽象地说“科学继续前进”。

在本书的解释中，宇宙扩张关系形成了一种新的天文学世界：星系不仅作为远方恒星系统存在于目录中，还以共同尺度和运动关系进入整体研究。这个世界不是一张旧地图增加一条箭头，而是对象之间的可比较关系改变了。它的成立仍然依靠实际数据的约束，任何本体论都不能替代测距、测谱和模型检验。也正因为约束具体，新世界能够被继续修正，而不必在每一次参数改变时整体坍塌。

本章参考文献

[34-1] Hubble, E. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. PNAS 15, 168–173, 1929. NASA 原文转载。 https://science.nasa.gov/apod/a-relation-between-distance-and-radial-velocity-among-extra-galactic-nebulae/

[34-2] Slipher, V. M. The Radial Velocity of the Andromeda Nebula. Lowell Observatory Bulletin 58, 1913. https://www.roe.ac.uk/~jap/slipher/slipher_1913.pdf

[34-3] Lemaître, G. Un univers homogène de masse constante et de rayon croissant, rendant compte de la vitesse radiale des nébuleuses extragalactiques. 1927. 本章年代与内容参照原文标题、研究机构对该文的说明，未声称完成法文全文校勘。 https://www.pas.va/en/publications/acta/acta24pas/hertog.html

[34-4] Livio, M. Mystery of the Missing Text Solved. Nature 479, 171–173, 2011. 原始档案研究。 https://www.nature.com/articles/479171a

第 35 章 距离再标定：为什么一个旧数值的修正改变整个宇宙

35.1 “一个常数”背后的标尺

宇宙膨胀率常以一个数值进入公共叙述，仿佛它只是等待望远镜读出的刻度。然而，速度与距离的比例一端连接光谱，另一端连接整个距离网络。当光谱资料大体不变而距离尺度被修订，比例参数也会改变。因此，一个常数的历史不仅是越来越精确地测量同一东西，还包含标尺自身被重新认识的历史。

以最简单的近距离关系 $v \approx H_0 d$ 说明，如果全部距离增大为原来的两倍，相同速度对应的 H_0 就约减半。这个代数结果不涉及宇宙突然改变运动状态，而是同一资料在修订标尺下取得新的参数。若把参数变化直接称为物理宇宙发生了变化，就混淆了知识修订与物理演化；若因此认为只剩符号变动，又会忽略它对光度、尺度和年龄推断的实际影响。

这里至少有三类修订需要分别处理。零点修订改变同一关系的绝对位置；类型重分改变哪些对象可以共用一条关系；适用域修订则限制关系在何种波段、环境或样本范围内有效。第一类有时接近整体伸缩，第二类可能要求拆成多条关系，第三类则可能保留核心公式却更换其使用条件。把三者都称为“测得更准”，会掩盖不同的思想变化。

巴德关于造父星周期—光度关系的讨论，以及桑德奇对距离尺度问题的研究，是这些修订的重要历史材料。[35-1][35-2]本章不把复杂过程压成某一天突然发现一个简单错误，而关注其机制：对象种群、标准关系与距离应用之间原来默认的对应，被更细的观测迫使重新安排。一个成熟工具重新打开，往往不是因为完全失效，而是因为其精细使用暴露了原来没有区分的条件。

从 SDE 看，既有距离 S 已经被广泛使用，它会进入后续对象光度、尺度与宇宙参数的计算，形成回写。正因为应用广，校准 E 的改变才会传播到许多领域。这里的力量来自依赖网络，并非一个数字本身具有魔法。追踪哪些结论共享同一标尺，是判断一次修订究竟改变多深、传播多远的必要步骤。

35.2 不同造父星群与消光

如果两类变星在外观上都有周期性的亮度变化，却在相同周期下具有不同的典型光度，那么把它们混入同一标定关系就会产生系统错误。问题不在于周期测量一定失败，而在于从

周期到光度的转换省略了类型条件。对象内部的差异原来被封装进“造父星”这个名称，后来必须重新显露。

巴德的 1956 年论述回顾了这类距离尺度问题。[35-1]桑德奇作为相关研究的参与者，在回忆中指出旧的经典造父星标定需要显著调整，并把它与新的颜色—星等资料联系起来。[35-3]这些材料支持一个具体认识：重新校准不仅需要更多同样的点，也需要能区分种群和演化状态的新特征。分类精细化能够改变一条关系的物理含义。

但种群区分不应被写成万能解释。实际测距还面临消光、拥挤、光度零点以及筛选差异。若某一对象比预测暗，研究者不能立即说它属于另一种群；尘埃和其他影响也可能造成相似显露。同一残差具有多个候选来源，必须用不同波段、光谱、位置或独立距离来分解。发生学分析的任务是说明这些候选如何被限制，而不是为残差快速指定一个名字。

消光尤其显示了属性的纠缠性。颜色既是恒星自身性质的线索，也是尘埃影响的线索；亮度既依赖光度与距离，也依赖沿线遮蔽。若把颜色先全部归入某个原因，再用修正后的亮度验证该原因，就可能形成循环。额外波段提供的信息可以减轻退化，但是否真正分开了各项影响，需要看参数之间仍剩下多大相关性。

拥挤提供了与消光方向相反的另一种偏差。若有限角分辨率把目标变星与邻近恒星的光混在一起，目标便可能显得过亮；把这些光全归给变星，在其绝对光度固定时就会低估距离。距离更远的星系在同样分辨率下更容易把较大的物理区域压进一个图像单元，因此不能只假定近处与远处混光程度一样。

这里需要检验的不是“照片是否清晰”这一笼统印象，而是目标周围额外光的分布及其对光变振幅、平均亮度和筛选的影响。更高分辨率的观测可以拆开部分来源，在图像中加入已知亮度的人工目标则可以检查处理程序如何回收信号。后一种方法是对测量链的实验，不能替代真实天体物理，却能够把原先隐藏在软件与背景中的偏差变成可量化的对象。

对象类型与环境也不能被简单等同。某类变星常见于特定环境，环境可以帮助判别，但不能代替对对象的测量。否则，一个落在不典型位置的真实成员可能被删去，分类又会因为删去例外而显得更加整齐。成熟的测距系统应当允许不典型对象暴露模型边界，同时对错误辨认保持警惕。

这类修订并非把原来的勒维特关系宣布为错误。原关系在特定样本中的成立，与后续使用者把它转移到其他对象上的正确性，是不同问题。若迁移条件被误判，应当修订的是这条桥梁。科学历史中大量重要成果具有这种结构：局部发现相当稳健，广泛应用却需要更复杂的分类和校准。承认这一点，才能在纠错时保存真正已经取得的知识。

35.3 再标定如何改变宇宙量

距离变化会以不同次方进入不同推断。由同一角尺度得到的线尺度通常与距离成正比；由同一接收通量得到的光度在近距离近似下与距离平方成正比。于是，整体距离伸缩不会只让地图变大，还会改变对象的发光能力评价。某些原来被视为普通亮星的对象，可能因此需要重新考虑其种类或是否为混合结构。

星等制可以更精确地说明零点修订的传播。在已经处理消光的近距离情形，距离模数写作 $\mu = m - M = 5\log_{10}(d/10\mathrm{pc})$。保持视星等 m 不变，若给对象赋予更亮的绝对星等，便要把它放得更远才能解释同样的接收亮度。约一点五个星等的模数增加，对应距离约增大一倍。这是数学换算，不应被写成每一次历史修订都严格相同或同时完成。

桑德奇的参与者回顾以约一点五个星等的零点改动说明巴德修订的尺度。[35-3]但这个数值只有嵌入具体标尺才有意义。如果某个星系的距离不是通过受影响的关系推得，就不能无条件照搬同一倍数；如果其多个测距指标混用了不同标定，修订还需要追踪各指标在最终估计中的权重。历史上一个广为传播的“翻倍”概括，可以揭示结果的量级，却不足以代替逐对象的重算。

动力学质量与光度的传播甚至具有不同次方。仍以近圆运动的简化关系为例，保持观测速度和角半径不变，线半径随距离增加，所推断的动力学质量也近似按距离增加；同一通量所对应的光度则按距离平方增加。因此质量光度比会随距离变化，而不是自动保持不变。这一例子说明，一项距离修订可能改变对天体物质组成的判断，即使光谱与影像本身都没有改动。

这类计算还提醒我们，不同学科给出的“独立检验”可能共享上游尺度。恒星群光度、动力学质量和某些天体年龄估计，看似属于不同研究传统，却可能在转换观测量时共同采用一个距离。判断它们是否真正互证，不能只看研究者使用不同方程，还必须看方程里的距离是否来自同一来源。

宇宙时间尺度同样受到影响。H_0 的倒数具有时间量纲，但它不是对任何宇宙模型都直接等于宇宙年龄。实际年龄还依赖膨胀历史、物质成分和其他模型条件。因此，距离修订可以缓解某种年龄矛盾，却不能单靠倒数运算证明所有年代问题已经解决。必须同时说明采用的演化模型，才能判断数值改变带来的物理意义。

这种传播具有方向性。若距离零点改变导致多个星系光度一起变化，不能再把这些光度变化当作独立支持零点改变的新证据，否则同一个修订被重复计算。只有新增、且不完全依赖原零点的观测，才可能提供外部检验。推断链越长，越容易出现“许多结论都一致改变”的表象，而它们实际上只共享一个上游变化。

一个清楚的审查方式，是先固定观测记录，明确哪些量由何种关系重新计算。谱线波长可能不变，角大小不变，视亮度不变，绝对光度和线尺度改变。再检查这些新推断与独立的恒星物理、动力学或其他测距路线是否相容。这样可以把直接观测、模型转换和外部验证分开，使修订过程本身成为可检查的研究。

距离再标定由此既有建设性，也有破坏性。它可能让此前不相容的年龄估计重新协调，也可能使某类对象的光度超出原模型范围，迫使新的物理解释进入。不能预先保证每一次校准修订都会让宇宙图景更和谐。真正的和谐来自经过检验的相容，而不是把全部不一致通过同一个可调参数抹平。

在三方程的关联中，新的分类 E 会改变标定序列 D，并产生新的尺度 S；新的尺度又使研究者重新检查对象物理与样本选择。这里确实存在循环，但它可以是开放循环：每轮加入新的约束，并保存旧记录以供比较。若只有参数在同一资料里反复调整，而没有新检验，循环就可能变成自我适应。两者的区别不在是否反馈，而在反馈是否增加了失败的可能性。

35.4 稳定部分也必须保留

校准修订的叙述若只强调“整个宇宙扩大了”，容易造成科学知识随时全部翻转的印象。实际上，某些判断可以跨越明显的数值改动继续成立。例如，某一远方系统的河外身份可以保持，两个对象的相对远近可以保持，某个周期范围内的光度趋势也可以保持。改变绝对尺度并不必然取消全部关系结构。

这类保持并非因为研究者拒绝纠错，而是不同结论依赖不同精度。判断一片森林是否在河对岸，不要求知道每棵树到观察者的距离；精确比较树高则需要更细的尺度。天文学同样存在由粗到细的判断层级。把各种结论都要求到同一精度，既会低估早期成果，也会掩盖精密研究的新任务。

本章因此建议以变化的对象来评估修订深度。如果只是同一测量体系中零点变更，应当说明参数传播；如果分类依据和对象单位被重新组织，则应说明新的类别怎样成立；如果基本测量关系或模型适用范围改变，则应进一步检查原有推断是否仍有局部有效域。这样的区分比统一宣称“旧世界被推翻”更有解释力。

SDE 的世界构成主张也应受这项纪律约束。科学世界不是单个数值的集合，而是显露、差异操作与特征纠缠形成的可运行整体。零点变化可能在这个整体内部被吸收，也可能触发更深的结构调整。哪一种情况实际发生，需要由研究实践证明，不能由本体论先行宣布。

距离尺度史最终呈现的，是可靠性如何在可修订中增长。一条标尺可以因为不断打开封装而更可靠，同时保留跨阶段的比较能力。统一的宇宙图景并不依赖所有旧数值永远正确，

而依赖研究者能够说明新旧结果怎样转换、哪些判断继承下来、哪些必须放弃。这样形成的连续性，才使下一阶段对动态宇宙模型的讨论具有稳固基础。

本章参考文献

[35-1] Baade, W. The Period-Luminosity Relation of the Cepheids. PASP 68, 5–16, 1956. https://adsabs.harvard.edu/pdf/1956PASP...68....5B

[35-2] Sandage, A. Current Problems in the Extragalactic Distance Scale. Astrophysical Journal 127, 513–526, 1958. https://articles.adsabs.harvard.edu/pdf/1958ApJ...127..513S

[35-3] Sandage, A. The First 50 Years at Palomar: Baade's Factor of Two in the Distance Scale. 参与者回顾。 https://ned.ipac.caltech.edu/level5/Sept03/Sandage/Sandage2_2.html

第 08 编

膨胀宇宙与精密宇宙学

第 36 章 动态宇宙：场方程、假设与观测约束

第 37 章 宇宙微波背景：噪声怎样成为宇宙证据

第 38 章 轻元素丰度与早期宇宙：两条证据链怎样交叉

第 39 章 不可见质量：引力证据怎样改变物质分布图景

第 40 章 宇宙加速与参数拟合：理论何时需要再打开

第 36 章 动态宇宙：场方程、假设与观测约束

36.1 静态宇宙为什么吸引研究者

把广义相对论写成一组方程，并不等于宇宙已经有了唯一的形状和历史。方程规定几何与物质能量之间的联系，具体宇宙模型还要说明怎样组织物质、采用哪些对称性以及允许何种时间行为。二十世纪初的关键转折，正发生在理论定律与附加假设之间：被视为自然的静态限制，后来显示为可以放开的选择。

爱因斯坦 1917 年将广义相对论用于宇宙整体，提出静态模型并引入宇宙学常数。有关这篇论文的原始文本研究强调，其中既有物理与数学动机，也有当时的天文学背景，不能压缩为“伟大科学家害怕变化”的性格故事。[36-1]科学家在研究局部运动时可以熟悉变化，同时仍认为足够大尺度的整体大体不变；这两个判断并不在直觉上自动冲突。

静态还有一种方法上的吸引力。若整体结构不随时间改变，许多边界与长期行为问题似乎更容易处理；宇宙可以作为一个满足平衡条件的整体来研究。但容易设定不等于物理上正确，更不等于稳定。一个模型能精确满足静态方程，只说明存在某种平衡配置；若微小扰动使它偏离之后不能恢复，静态状态便未必能够长期实现。

平衡与稳定的区别值得在思想史中保留。把球放在山顶，可以在理想条件下处于受力平衡，但不能据此断言它会抵抗扰动。宇宙模型当然不是这个机械玩具，类比仅用于说明逻辑结构。一个解的存在、它对扰动的反应，以及它与观察的相容，是三个不同问题。数学上找到解，只完成了其中第一步。

用后文的现代方程回看一种简化静态模型，可以看见限制的强度。若忽略压力，要求尺度的一阶与二阶时间导数都为零，加速度方程便要求 $\Lambda c^2 = 4\pi G\rho$；再代入第一条方程，曲率项也要满足相应平衡。这里的 Λ 不能在任何密度与几何下随意取值。静态不是少考虑一个运动量那么简单，而是把原来可以变化的成分压进特定配合关系。

在这套简化条件下，若尺度稍微增大，物质密度随膨胀降低，而常数项保持不变，原来的平衡就会改变。这个现代说明展示了为什么不能把精确存在的平衡解自动当作能够抵抗扰动的长期状态。它并非声称所有静态模型都由同一个简单论证排除，而是说明稳定性必须重新计算，不能从“已经找到一个解”直接继承。

这个例子也使理论修改的意义更清楚。引入一个常数项，可以扩大方程所允许的状态，但不保证所追求的状态在物理上自然实现；放弃静态限制，则可以在保留方程的同时改变所寻找的解。两种做法位于不同层次。历史叙述若把它们统称为“改变理论”，就会漏掉研究者究竟修改了定律形式、对称假设，还是解的选择。

从 SDE 分析看，静态世界的形成并不只由一种主观偏好决定。可用观测、理论关系、物质假设及允许的几何结构共同限定它。E 的关键不是这些因素同时在场，而是它们互相规定：选择静态，便限制了物质与几何的关系；改变物质或附加项，可能使静态解消失或取得不同性质。所谓本体假设，正藏在这些对可允许状态的限制中。

因此，后来动态宇宙的出现，并非把全部旧理论抛弃。它可以保留基本场方程，同时放宽某些限制，由此展现原来被排除的解族。这种变化比一般意义上的“修正参数”更深，因为时间演化从需要排除的状态变成能够研究的对象；又比彻底推倒重来更有连续性，因为新的研究仍利用原理论已经建立的数学结构。

36.2 宇宙解的自由度

弗里德曼 1922 年的工作表明，相对论场方程允许随时间改变的宇宙尺度。莱顿保存的手稿与通信档案可以追踪这一研究，其说明明确区分了当年的宇宙曲率论文与另一份几何手稿。[36-2]这个区分虽看似书目细节，却重要：相近题名不能替代内容辨认，更不能把任意一份关于弯曲空间的文件当作动态宇宙论文。

用现代形式说明，若假定大尺度空间均匀且各向同性，可以用尺度因子 a(t)表达整体几何的时间变化。通常写出的弗里德曼方程包含物质密度、空间曲率及宇宙学常数。以 ρ 表示质量密度型的能量变量，第一式可写为 $H^2=8\pi G\rho/3-kc^2/a^2+\Lambda c^2/3$，其中 $H=\dot{a}/a$。这里的记号是现代说明，不声称复现 1922 年原式的符号布局。

这条式子本身还不决定完整历史。需要知道 ρ 怎样随 a 变化，而这又取决于物质的状态关系与守恒条件。非相对论物质与辐射在膨胀中具有不同的密度变化；附加成分可能又有不同表现。初始条件或某一时刻的参数也要给定。所谓“由方程推出宇宙”，因而通常意味着由方程加上一组明确假设推出一个候选历史。

均匀与各向同性也要分别理解。均匀说明在所讨论的空间尺度上，不同位置没有被预先指定为本质不同；各向同性说明从给定位置看，各方向没有特殊方向。它们不是“天空看起来很平滑”的同一句话。只从一个位置观察到近似各向同性，逻辑上并不足以无条件保证所有地方都均匀。把我们的观测位置不视为宇宙中特殊例外，是连接二者时使用的另一项原则。原则能使研究前进，但必须保留为可以陈述的前提。

这些假设也不宣称局部世界没有结构。星系、星系团、空洞的存在，要求模型说明在什么尺度进行平均，以及平均几何怎样与局部扰动联系。把一个星系团内部的不均匀直接当作大尺度模型的反证，会忽略模型适用尺度；把所有不均匀都预先说成无关紧要，又会使尺度要求失去检验。真正的问题是：采用的平均是否足以准确描述所研究的观测，以及剩余结构是否需要单独计算。

尺度因子的数值还包含记号约定。研究者常把今天的尺度因子定为一，这不是测得今天宇宙大小等于某个绝对单位，而是选定比较基准。真正有可观测意义的是适当条件下的比值和变化率。空间坐标也可以随约定重新归一化，与曲率参数的记号配合。因此，读取不同时代的公式时，不能仅凭 a 或 k 长得相同就断定它们采用同一量纲。数学形式相似，必须经由定义才能成为物理比较。

用现代守恒关系进一步说明物质假设，可以写成 $\dot{\rho}+3H(\rho+p/c^2)=0$。它与前面的方程一起告诉我们，密度变化不仅来自体积增加，还与压力所对应的能量变化有关。对压力可以忽略的物质，得到 ρ 随 a 的三次方反比变化；对辐射，则为四次方反比变化。多出的一次方对应膨胀过程中单个光子能量的变化。这里并不是给历史人物补上一页现代教科书，而是明确“选择宇宙物质内容”究竟怎样改变候选历史。

由此可见，同一时刻具有相同总密度的两个模型，仍可能在之后呈现不同演化，因为它们的压力与密度关系不同。只报告一个密度数字不能代替成分说明。反过来，理论研究可以问：怎样的观测能够区分这些眼下相似、未来或过去却不同的模型？模型前提于是变成安排观测的依据，方程不再只是总结结果，也参与生成可检验的问题。

第二类关系涉及加速度。采用相同约定，可写为 $\ddot{a}/a=-4\pi G(\rho+3p/c^2)/3+\Lambda c^2/3$，其中 p 为压力。这个现代说明式指出，在相对论宇宙学中，压力也参与引力效果。它并不意味着任何增加压力的过程都能用日常容器里的直觉来解释。整体模型中的压力与能量关系，必须在所用理论内理解。

这一组式子的思想力量，在于把宇宙整体从单一图像变成受条件控制的可能性空间。不同物质组成、曲率和附加项允许不同的膨胀或收缩行为。数学家可以先研究哪些行为可能，再问哪一种与观测相符。可能性并非现实性，但没有可能性空间，研究者容易把一个熟悉解误认为理论唯一允许的世界。

另一方面，解空间也不能无限扩张而不付代价。假如为了拟合每项新数据都随意增加成分或自由函数，模型就可能失去区分力。自由度既是探索资源，也是解释负担。一个新增项应当有明确的物理含义、与其他观测的关系，以及允许失败的条件。否则，从静态限制中获得的开放性会变成没有边界的适配。

一项常见误解是把空间曲率的符号直接当作宇宙命运的名称。在只含某些成分且采用特定限制的模型中，曲率与是否回缩可以有紧密联系；但允许宇宙学常数或其他成分之后，不能仅凭“空间闭合”就无条件断言最终回缩，也不能凭“空间开放”就替全部时间行为作结。几何性质与动力学命运由方程相连，仍是不同的问题。

这种区别可以用同一套方程中的项来看。曲率项随尺度变化的方式不同于物质、辐射及常数项；随着 a 变化，起主导作用的项也可能更替。某一时期占优势的成分，不必控制全部

历史。因此，一个只从今天的总体外观推断永远命运的图景，通常省略了成分及其演化条件。动态模型真正引入的，是不同阶段之间能够转换的因果结构。

初始条件还需要与初始原因区分。给定某一时刻的尺度、变化率和物质状态，可能足以在适用范围内求出之后的演化，却没有因此解释这些条件为什么出现。理论具有预测能力，并不意味着它已经提供全部起源说明。若向更早时期推演进入现有理论失去可靠性的范围，正确的边界是承认模型在此需要新的物理，而不是把数学外推的端点直接等同一份已经观察到的宇宙出生记录。

从静态到动态的思想转变因此同时增加了解释力和问题数量。静态模型可以把许多历史问题暂时排除，动态模型则必须面对不同阶段的联系、状态转换和初始条件。开放这些问题不是理论退步，而是把原先藏在“不变”假设背后的责任重新显露出来。只有把新增责任交给明确的计算与观察，开放性才会成为实际进展。

这里有一项 SDE 方法上的关键转换：先从既有宇宙图景 S 出发，研究者可能只寻找保持它的参数；先从方程和限制 E 出发，则会发现 S 只是解族中的一个成员。再从观测序列 D 返回，可以淘汰其中部分候选。路径变化的增量在于识别哪些“宇宙性质”其实来自先加的限制，而不是由证据唯一指定。

36.3 观测红移怎样进入

理论宇宙的尺度因子不是望远镜直接读出的量。望远镜记录光谱、角位置、通量和时间变化；将这些记录与 a(t)联系，需要光在特定时空几何中传播的理论。对理想的共动发射者与观察者，在均匀各向同性模型中，宇宙学红移满足 $1+z=a(t_0)/a(t_e)$。其中 t_e 是发射时刻，t_0 是接收时刻。[36-3]

这个关系说明红移如何成为宇宙演化的线索，也显示它的前提。局部运动、引力势差以及观测者运动都会影响实际谱移；它们不能一律由尺度因子代替。特别是近邻对象的接近运动，并不直接否定整体膨胀。相反，若研究者无论看到什么谱移都归因于整体膨胀，模型便失去处理局部差异的能力。

距离的含义也随动态几何变得更精细。光度距离由通量与光度的关系定义，角直径距离由物理尺寸与角大小的关系定义，共动距离则服务于特定几何描述。在近距离极限下，它们可以近似一致；在较大红移处，不能不加说明地互换。[36-3]使用同一个“距离”词并不能保证使用的是同一个量。

这并非数学家为了复杂而增加名词。宇宙在光传播期间发生变化，发射时的系统与接收时的系统不是同一时刻的截面；光子的能量与到达率也受到整体演化影响。不同距离定义把

不同观测操作与模型量连接起来。若把它们混用，可能在公式上得到整齐结果，却把实际比较对象换掉了。

由红移资料进入模型，与由模型进入红移预测，承担不同任务。前一路线首先检查观测是否可靠、局部影响是否处理、数据覆盖是否足够；后一路线首先检查给定假设会产生怎样的距离—红移关系。二者相遇才形成模型检验。一个漂亮的理论曲线没有数据约束便只是可能性，一组数据没有明确比较关系也不能自动说出全部宇宙历史。

距离之间还有可展示其关系的现代公式。在度规理论中采用通常的光传播，并满足光子数守恒等条件时，光度距离与角直径距离满足 $D_L=(1+z)^2D_A$。[36-3]这个关系没有把两个定义取消为一个，而是说明不同观测操作在特定物理条件下应当怎样对应。若比较出现偏离，首先需要检查测光、尺寸判定和选择效应，之后才谈是否有传播条件或理论前提需要修改。

这个例子使“模型依赖”具有比赞成或反对更精确的含义。一个距离结果依赖某些条件，并不意味着它可以任意选择；条件越明确，越能寻找其他操作检验它。相反，声称某项距离完全没有模型，往往只是把从通量到距离所用的关系藏起来。科学上的可靠性并不来自消除一切前提，而来自把前提组织成可以跨观测检查的关系。

还应区分宇宙年龄与回溯时间。一个远方对象的光传播了多长时间，涉及从发射到接收的区间；宇宙模型的年龄则涉及模型所容许的更早历史。红移不是一只无条件的钟，从红移得到时间，需要采用膨胀历史。于是，先选定模型求得远方天体的“当时年龄”，再把这个年龄当成完全独立的数据证明同一个模型，就有重复使用前提的风险。天体内部演化给出的年龄限制若能通过不同物理取得，才构成更有区分力的交叉约束。

“空间膨胀”的说法因此需要恰当使用。它不是一张照片上能够看到的空间物质向外拉伸，也不是所有局部束缚系统都以相同比例变大。它是对特定整体几何演化的描述。恒星、行星系统和其他局部结构是否跟随，取决于具体动力学，不能由一个宇宙尺度因子替代全部层次的运动分析。

这种层级区分保护了本体论讨论的严肃性。SDE 可以指出整体宇宙对象如何由测量与模型共同构成，却不能凭该框架推导光子的红移或局部束缚系统的行为。这些问题需要相应物理理论。发生学解释与物理解释可以紧密相接，但各自必须承担自己能够检验的命题，不能靠跨层借词来制造证明已经完成的印象。

36.4 模型自由度与观测约束

动态模型成为研究对象以后，宇宙学不再只问是否膨胀，还要问膨胀历史、组成、曲率与结构如何相互限制。但每个新问题都不必由同一类资料回答。距离—红移关系约束整体几

何，背景辐射涉及更早阶段的热史，物质分布与增长又提供不同的限制。多通道的价值在于它们对参数组合的敏感方向不相同。

参数退化说明了为何一个观测关系通常不够。两套不同参数可能在某个距离范围内给出近似相同的曲线；增大样本数量会缩小随机误差，却未必打破这种结构性相似。研究者需要找到对两套模型作出不同预测的观测。这与距离阶梯缺少绝对零点时的困难具有共同逻辑：数据可以很多，但缺少的可能是另一种信息，而非更多同类信息。

可识别性还取决于观测跨度。假设两条膨胀历史在低红移处具有相同的一阶行为，而在更远处才分开，那么再精密的极近邻测量也可能主要限制共同的一阶系数。它并非无用，只是无法承担它不敏感的历史差别。扩展红移范围、增加另一种距离定义或加入结构演化信息，能够改变问题的敏感方向。观测设计因而应从候选模型的差异出发，而不是只追求每一种测量都更精密。

理论比较还需要保存模型失败的具体含义。如果某一组合在全部合理参数范围内都无法解释资料，压力落在模型结构上；如果仅某组参数不合适，则首先淘汰的是该组参数。两者之间还有观测误差与输入物理需要核查。区分这些层次，能够避免每次参数更新都被夸大为宇宙观翻转，也避免真正的结构性问题被无限解释成参数尚未调好。

同时，联合检验必须承认共享假设。两个探针使用不同仪器，不意味着它们在理论层面完全独立；它们可能共同采用相同的均匀性假设、引力理论或早期宇宙物理。若这些共同条件不写清，联合结果会显得比实际更无条件。反过来，共享某些理论也不取消其他层面的独立性，关键是说明独立发生在哪里。

这里的和谐并非把所有数据逼入同一个参数点，而是在统一模型中让不同观测各自保有误差与适用范围。若出现不一致，先定位它可能来自观测、数据处理、物理输入还是模型结构，再决定如何修改。发现张力是一项研究成果，但张力本身不会自动指定新理论。错误校准与新物理都可能表现为残差，只有增加可区分的检验才能推进判断。

与静态模型相比，动态宇宙真正打开的是整体历史的研究资格。时间不再只是观测者记录资料的顺序，也成为模型对象本身的维度。新的宇宙世界由尺度演化、物质组成和可观测关系共同构成，它使背景温度、轻元素与大尺度结构能够进入同一历史框架。这个框架的力量来自可以把不同时间层次的证据联系起来，也来自它可能在这些联系处遭遇失败。

因此，本章所强调的创造性不是“方程告诉人类宇宙一定怎样”，而是研究者通过改变限制，创造出可以公开推演、比较和否定的宇宙候选。随后观测反过来压缩候选空间。三方程所要求追踪的相互规定，在这里形成清楚分工：显露必须由观测与模型共同支持，研究序列受初步结果引导，理论限制又必须因新比较而接受重审。动态宇宙既成为物理研究的对象，也成为检验知识如何持续发生的一处重要场所。

本章参考文献

[36-1] O'Raifeartaigh, C.; O'Keeffe, M.; Nahm, W.; Mitton, S. Einstein's 1917 Static Model of the Universe: A Centennial Review. 原始文本研究，2017. https://arxiv.org/abs/1701.07261

[36-2] Instituut-Lorentz, Leiden University. Friedmann papers: 1922 手稿及通信档案。https://lorentz.leidenuniv.nl/history/Friedmann_archive/

[36-3] Hogg, D. W. Distance Measures in Cosmology. 1999，2000 修订。现代公式说明依据。https://arxiv.org/abs/astro-ph/9905116

第 37 章 宇宙微波背景：噪声怎样成为宇宙证据

37.1 异常温度如何被排除

一个接收器发现比预期更高的噪声，最初并不意味着宇宙学发生了转折。它可能来自设备、地面、大气，也可能来自尚未正确计入的天区辐射。科学发现的困难就在于，异常最早往往以“测量没有做干净”的形式出现。若每次异常都立即命名为新宇宙现象，研究便会充满伪发现；若一律把异常删去，新的观测对象又无从成立。

彭齐亚斯和威尔逊 1965 年报告，在 4080 兆周每秒的频率上，天线有效噪声温度比预期约高出 3.5 开尔文，并指出同一期迪克等人的文章提供了一种可能解释。[37-1]这是一项具体频段中的超额测量，不是已经完成整个微波背景谱的绘图。后来的知识越丰富，越应保留当时论文实际承担的有限命题。

天线温度是一种把辐射强度折算为温度单位的表达，不能不加区别地理解成天线金属本身的温度。接收到的功率需要经过天线方向响应、接收系统增益和噪声特性等关系，才能成为可比较的辐射量。一个残差能否稳定存在，取决于这些转换是否经得起检查。没有仪器模型，残差甚至不能被明确地定义为“比什么多出来”。

排错过程的认识价值也在这里。研究者不是只问某一个原因是否可能，而要比较不同原因应当怎样随方向、时间或设备状态改变。如果残差来自地面拾取，可能表现出与天线指向或环境有关的变化；如果来自某种接收器故障，更换或校准环节应当影响它。具体项目的检验范围需要依据记录，不能用后来流行的故事替代。本章关注的是排除必须形成可比较预期这一方法结构。

即使已知的本地来源得到排除，也不能仅凭排除法唯一推出宇宙早期遗迹。逻辑上仍可能存在其他天体来源。肯定身份还需要正面关系：天空分布、频谱行为以及与宇宙热史的联系。异常从“设备问题”转成“天体候选”是一层变化，从“天体候选”转成“宇宙背景”又是另一层变化。把它们压缩成一个瞬间，会隐藏发现真正经历的证据升级。

还应区分稳定的残差与充分确定的来源。若重复测量总得到相近的剩余，说明某种现象或某项系统偏差具有重复性，并不自动表明已经摆脱仪器。一个稳定的校准错误也可以非常稳定。只有改变那些据说无关、但若误差解释成立就应当有关的条件，重复才逐渐成为对来源的检验。重复测得一个数，与重复排除其候选来源，是两项相接而不相同的工作。

因此，不能用一个漂亮的偶然发现故事替代测量责任。科研叙事常喜欢把转折压缩到一次通话或一个恍然大悟的时刻；这样的交流可能促成解释相遇，却不能独自使信号成为宇宙

证据。使它获得公共地位的是能够记录和复核的结果，以及解释将继续接受哪些测试。知识的传播事件可以成为故事的场景，知识成立的条件却必须仍留在测量与推断中。

在 SDE 语言中，原来作为观测障碍的剩余量，后来成为新的 S。促成转变的不是给它换名字，而是 D 中的排错与复查重新限制了 E：仪器响应、前景贡献与宇宙模型之间的关系开始允许一种新的对象身份。这里的发生性尤其鲜明，因为同一信号在不同研究系统中可以先是噪声，后是需要保留的宇宙信息。

37.2 理论预期是否独立

宇宙学解释并非在 1965 年才第一次出现高温过去的思想。阿尔弗与赫尔曼 1948 年的研究已经讨论膨胀宇宙中的热史及残余辐射问题。[37-2]但有过理论预期，不等于后来观测者只是按一张完整寻宝图找到目标。理论工作的传播、参数估计与实际探测计划之间，仍存在具体的历史距离。

迪克、皮布尔斯、罗尔和威尔金森 1965 年的论文将微波辐射与宇宙热史联系起来。[37-3]它与观测报告在同一期发表，使读者能够分别看见测量结果和物理解释。这样的文本分工本身具有方法价值：解释不应吞没测量的实际内容，测量也不必假装自己没有理论关联。两者在可以核查的地方接合，才能形成强于各自独立陈述的证据。

检验理论预期是否独立，需要问它在知道新数据之前已经限定了什么。预先指出可能存在低温背景，与事后选择一个参数恰好拟合观测，具有不同的证据力量；但前者若数值范围很宽，也不能被夸大为精确命中。历史上预言的概念内容、数量精度与传播影响应分别评价。用“早就预言了”概括全部问题，容易把真正的开放性删掉。

热宇宙解释的优势，不只在于给一项超额温度找到名字，而在于把它联系到更广的物理关系。一个早期处于热状态的辐射场，随宇宙演化可以留下有特定频谱性质的背景。于是，解释同时要求研究者去检验更多频率与更细的天空分布。能产生新测试的解释，比只能适配一个数字的解释具有更大的研究内容。

用现代物理关系说明，均匀绝热膨胀中的自由传播会使辐射波长随尺度因子增长，同时使光子数密度下降；一个已经具有黑体分布的辐射场可以在这种演化中保持其谱形，而温度按尺度因子的反比变化。这解释了为什么今天的低温辐射能够与高温过去相接，却不是说任意初始辐射都会仅凭膨胀自动变成黑体。初始谱形如何形成，仍是另一项物理责任。

这个区别限制了一种容易出现的倒推。今天观测到黑体性质，为早期热状态提供重要支持；但若要断言某个精确时刻、某个独特机制或完整初始状态已经确定，还需加入额外模型和观测。一个强有力的宇宙遗迹并不是整部宇宙史的逐页记录。它保存哪些信息、抹去了哪些信息，同样由相互作用和演化决定。

黑体的理论特征也使后续测量具有明确目标。不同温度会改变整条谱线的强度分布，而某些能量注入或传播过程可能产生不同形状的偏离。因此，寻找偏离并不是背离原来的发现，恰好是在原有对象已经稳定后提高问题分辨率。科学研究可以同时确认背景的总体性质，并主动寻找它没有被简单模型描述的细节。

稳恒态相关解释构成了真实的竞争压力。争论不应被写成背景一发现，所有其他意见就立刻在逻辑上不可能。替代方案可以尝试说明弥散辐射的来源与处理过程，但它们必须同时承担强度、谱形、方向分布以及与其他观测相容的责任。后来更精细的数据提高了这种责任，而这些数据不能提前赋给 1965 年的参与者。

本书从这里提炼的判断是：独立性具有层次。仪器异常与某项宇宙学预期可以来自不同工作链，二者的相遇因而有重要价值；但解释阶段一旦共同依赖某些物理关系，就需要记录这种依赖。科学互证既不是纯粹互不相干的结果偶然相同，也不是同一个假设换几个名字后的重复。它发生在部分独立、部分共享的结构之中。

37.3 从发现到精密谱形

单一频率上的辐射强度，不能唯一确定整个谱形。许多不同来源都可以在一个频点具有相近强度。黑体谱的主张则更强：它要求一系列频率上的强度能够由同一温度的特定分布解释。因此，从发现超额微波信号到确认近乎黑体的频谱，增加的不只是精度，而是约束类型。

COBE 上的 FIRAS 通过天空辐射与精密黑体参考的比较，检验宇宙微波背景对黑体谱的偏离。Fixsen 等人 1996 年的研究给出了严格的偏离限制。[37-4]本章不把后来改进的温度数值写成 1965 年的结果，也不把“符合黑体”写成绝对没有任何可能畸变。实验报告的含义是，在指定频段、误差和分析条件下，偏离被限制到很小范围。

绝对谱测量与方向差分测量承担的任务也不同。前者要知道天空在各频率上的总强度，因而特别依赖绝对参考；后者比较两个方向的微小差别，有机会消去部分共同偏差，却不能单凭差值恢复全部平均强度。一台很擅长测差异的设备，不会因此自动成为同等可靠的绝对温度计。背景研究的不同结果需要各自适当的校准链。

由此也能理解为何不能把后来的精密各向异性图当作对早期谱形测量的简单重复。它们可以共同支持宇宙模型，却检验不同的可观测关系。一幅角分布图主要回答天空各处相差怎样，一条频谱主要回答辐射强度怎样随频率分配；二者对仪器、前景与理论的依赖交叉但不相同。只有保留这种差异，联合证据才不会被错误地当成同一测量做了两次。

黑体参考也需要被测量系统证明可靠。温度读数、发射性质与仪器响应如果产生共同偏差，比较结果可能受影响。精密实验因而不仅比较天空与理论曲线，还检验参考源和仪器自

身。所谓“理论与观测吻合”，常常经过多层校准；把这些层次展开，不会削弱结论，反而使结论的可靠性有可追踪来源。

银河前景又提供了另一重分离问题。天空不同方向的辐射既包含背景，也包含银河成分。FIRAS 数据产品说明中明确包含对背景单极、偶极及银河辐射成分的处理，并保存残差与不确定性。[37-5]我们不应把处理后的背景谱想象成仪器单独收到的一份纯净信号。它是通过不同空间和频谱特征加以分离后获得的量。

这种分离之所以不是任意构造，是因为各成分受到不同约束，并可以通过替代模型、天区选择和校准检验比较。若一种分离方法必须把任意不合意信号都放进前景，便可能失去反驳机会；若在多种合理处理下都得到相似背景谱，稳健性才增强。精密观测的认识论重点，不在是否使用模型，而在模型能否被不同资料限制。

前景分离还涉及天区选择的代价。避开银河辐射较强的方向，可能减轻某些污染，却会减少可用天空并改变对整体量的估计；使用更大天区，则要求更充分的前景描述。没有一种选择只带来好处。分析应检查结果是否随着合理选择发生可理解的变化，尤其要警惕只有某个特定范围才得到预想结果的情况。

频率资料与方向资料在此相互补足。某成分在天空中的位置变化，帮助判断它是否与银河结构有关；其频谱又帮助判断不同频点是否在观察同一来源。只沿其中一个维度分析，有时会留下难以分开的混合；联合使用两者，才能使拆分获得额外限制。这是多通道观测的具体认识增量，不只是把许多颜色叠在一幅更吸引人的天空图上。

因此，黑体谱的建立提高了宇宙解释的门槛。一个竞争方案若只能说明某频率的背景强度，却无法解释广频段形状，就需要增加新的物理过程。这些过程本身又必须与其他观测相容。证据不是通过“主流同意”单独增长，而是通过替代解释要同时满足的具体条件不断增加而增长。

37.4 背景的回写效应

背景辐射一旦成为稳定对象，研究重点就可以从“有没有”转向“是否完全均匀”“有哪些微小结构”“这些结构怎样关联宇宙组成”。原来需要压低的仪器差异，此时又成为限制可探测结构的关键因素。新的显露改变测量序列，测量序列要求新的校准与模型，这是回写在科研计划中的具体形式。

均匀与差异在这里不是非此即彼。背景的大体均匀性支持一种整体描述，而微小差异能够承载结构形成的信息。若为了保持统一而把所有差异当噪声，就会错失新的研究对象；若只强调差异而忽视背景整体性质，又会失去解释它们的共同参照。宇宙学的进展依赖在不同粒度上组织稳定与变化。

微小结构还需要与观察者的状态分开。相对于背景的运动可以产生有方向性的温度模式，这类信号首先反映观察者与辐射场的关系，不能一律归为早期宇宙的原始不均匀。去除或建模这些成分之后，剩余结构才进入相应的宇宙学解释。这里所谓“剩余”并非不重要，而是再次改变了对象身份：一次发现中的主要信号，在更精细研究里可以变为必须先处理的成分。

这形成了一种有层次的测量历史。早期研究努力确认一个大体无方向差别的背景；后续研究利用特定方向模式获得另一类信息，再继续寻找更细的结构。不同阶段并不是相互否定。它们逐渐改变需要达到的精度，以及哪些特征被当作研究目标。一个已确立对象能够分解出新的问题，正是稳定知识为创造留下空间的方式。

这也改变了“看见宇宙”的含义。微波背景不是把光学星系照片拍得更深，而是以另一组辐射特征构成新的天空。它拥有自己的前景、校准、图像和可比较量，又能与星系分布及物质组成相接。新的天文学世界因此不是旧世界的附加颜色，而是具有不同对象单位和推断关系的显露系统。

SDE 解释需要进一步承担边界责任：背景身份的成立，不能凭本体论替代热史计算；近乎黑体的频谱，也不能单独证明某个完整早期宇宙理论的每一细节。它对一类模型提供强约束，同时仍给初始条件和其他机制留下研究空间。知识的力量来自约束范围清楚，而不是把局部成功扩张为全局免检。

从一个无法解释的剩余温度到精密宇宙研究对象，最深的转变在于研究者改变了什么值得保留。噪声与信息并非永远固定的标签，但也不是可以任意交换的词。只有排错、正面预测和持续复查共同改变信号的身份，新的世界才获得稳定性。宇宙微波背景的历史因此说明，科学创造常从认真对待一个暂时不合适的剩余开始。

本章参考文献

[37-1] Penzias, A. A.; Wilson, R. W. A Measurement of Excess Antenna Temperature at 4080 Mc/s. Astrophysical Journal 142, 419, 1965. https://scixplorer.org/abs/1965ApJ...142..419P/abstract

[37-2] Alpher, R. A.; Herman, R. Evolution of the Universe. Nature 162, 774–775, 1948. https://www.nature.com/articles/162774b0

[37-3] Dicke, R. H.; Peebles, P. J. E.; Roll, P. G.; Wilkinson, D. T. Cosmic Black-Body Radiation. Astrophysical Journal 142, 414, 1965. https://articles.adsabs.harvard.edu/pdf/1965ApJ...142..414D

[37-4] Fixsen, D. J., et al. The Cosmic Microwave Background Spectrum from the Full COBE/FIRAS Data Set. 1996. https://arxiv.org/abs/astro-ph/9605054

[37-5] NASA LAMBDA. COBE FIRAS: CMB Monopole Spectrum. 数据产品与校准说明。
https://lambda.gsfc.nasa.gov/product/cobe/firas_monopole_spect.html

第 38 章 轻元素丰度与早期宇宙：两条证据链怎样交叉

38.1 模型的输入是什么

宇宙早期的核反应并没有留下一段可直接播放的录像。研究者从今天或遥远天体的物质组成出发，结合实验室核物理与宇宙演化模型，推断轻元素曾怎样形成。这条研究路线把极小尺度的反应与极大尺度的历史联系起来，其力量也来自这种联系所承受的多重限制：不能只调整宇宙故事而无视核数据，也不能只罗列核反应而忽略温度和密度随时间的变化。

阿尔弗与赫尔曼 1948 年关于膨胀宇宙热核反应的论文，是这条早期研究链的一部分。[38-1]后来的标准核合成计算更加精细地组织了反应网络、热史和各种不确定性。Cyburt 等人的研究既给出更新的轻核丰度计算，也讨论核反应输入、中子寿命和宇宙学参数如何影响结果。[38-2]本章不把几十年发展的全部形式倒写成早期论文已经拥有的成品。

“输入”一词需要细分。核反应速率可以受到地面实验约束，膨胀速率则通过引力理论和宇宙能量成分与温度历史相接；重子含量影响粒子碰撞和反应进行的条件。若其中一个输入改变，多种产物可能同时改变，而且变化方向与幅度并不相同。轻元素丰度因此不是四个彼此独立、可以任意调节的旋钮。

反应能否充分进行，还取决于反应时间尺度与宇宙冷却、膨胀时间尺度之间的竞争。即使某条反应在实验室条件下可以发生，也不代表宇宙会一直等待它完成。温度降低、粒子密度减小，会改变反应机会。这种“有何过程”与“有多少时间运行”的联合规定，是核合成模型真正的动力内容。

氘的形成可以具体显示这种时间竞争。质子与中子结合能够生成氘，但高温辐射中的高能光子也能把它破坏。核结合能并不单独决定氘何时能够大量保存，因为光子相对于重子的数量也参与破坏机会。通常所谓“氘瓶颈”，正是生成与光致破坏在膨胀冷却过程中改变相对强弱的结果。[38-2]单问一条反应是否在能量上允许，无法获得完整历史。

这个例子使早期宇宙热史区别于一只温度恒定、无限等待的容器。假如冷却时间改变，即使核反应规律完全相同，能够进入后续反应的物质也会不同；假如重子含量改变，反应物相遇的机会会改变，却未必以相同比例改变每种最终产物。模型的可检验性恰好来自这些不相同的响应，而不是来自“早期很热，所以会发生核反应”这一宽泛判断。

计算中的“冻结”也不意味着反应突然违反物理定律或在某一瞬间绝对归零。它通常表示某种过程已经慢到不足以维持原来的平衡关系，之后的数量便开始保留过去演化的痕迹。

对思想史而言，这是一种关键的时间观：今天的比例可以是过程未能继续充分运行的遗存。遗存不是未经解释的化石标签，而是一套速率关系留下的可计算后果。

用 SDE 解释时，不能把差异序列 D 随意写成“宇宙变化”，也不能把 E 写成宽泛背景。这里 D 包含物理模型中的反应与演化序列，E 包含速率、密度、温度和时间尺度之间的相互限定；S 是模型计算出的丰度及其可观测后果。若我们转而研究知识形成，D 又是实验测量、谱线分析和模型比较。两种层次可以连接，但必须说明何时正在讲物理过程，何时正在讲研究过程。

两者相接之处，是模型输出能否通过观测成为有约束的判断。一个计算即使精确，也只说明在给定输入下将产生什么；一个丰度测量即使可靠，也未必代表原初丰度。物理模型与天体历史之间的桥梁，正是下一层困难。

38.2 恒星丰度并非原初丰度

今天看到的元素比例经过漫长历史。恒星可以生成某些元素、消耗另一些元素，并把处理后的物质送回周围环境。测到一颗恒星表面的丰度，也不等于直接测到它形成时整团物质的丰度，因为内部混合、扩散和其他过程可能改变表面表现。因此，“古老天体”是寻找早期痕迹的重要入口，却不是未经加工的宇宙样品。

研究者需要选择加工较少或能够校正加工影响的场所。贫金属气体与古老贫金属恒星之所以重要，是因为它们提供不同于普通近邻恒星的历史线索。但贫金属并不逻辑上保证所有轻元素都完全保持原初状态。每种元素经历的生成与破坏过程不同，不能把一个总体标签当成所有同位素的共同通行证。

Cooke、Pettini 和 Steidel 对低金属丰度吸收系统的氘研究，展示了如何从多条吸收线和均一分析中约束氘氢比。[38-3]其方法价值在于，不只提出一个“古老气体”的故事，而是在具体谱线、吸收结构和误差中建立丰度估计。把某个系统称为接近原初，必须由可测的组成与处理历史支持。

谱线分析本身也不是直接数原子。线强与柱密度、温度、速度结构、电离状态等有关，不同谱线还可能混合或饱和。若一个弱特征既可能是氘，也可能是另一速度分量的氢，研究者需要其他谱线或结构约束来区分。更多观测的价值在于让候选解释作出不同预期，而非只让同一个模糊特征变得更亮。

氦的观测又要求另一种处理。研究者可以在低金属丰度的电离气体中比较氢与氦发射线，再处理温度、电离和吸收等影响，并研究丰度如何随金属含量变化。[38-2]这里通向原初值的路线不是找到一份自宇宙早期从未改变过的气体，而是利用加工程度不同的对象，约束加工趋于很小时的值。这个推断把对象之间的差异转成历史信息。

外推因此需要特别注意范围。若全部样本都离目标条件很远，原初截距便可能由所选函数形式强烈决定；若扩展了更接近目标条件的样本，外推负担才可能减少。大量精密测量集中在同一个狭窄区间，并不自动解决这一困难。设计新观测时，选择什么样的对象有时比在原对象上增加曝光更重要。

同一个“丰度”还可以指不同定义。按原子核数目报告某同位素相对于氢的比值，与报告某成分在总重子质量中所占份额，不是同一数值。两者能够换算，但要明确组成与质量权重。把不同定义的比例直接排在一张表中，很容易制造表面的差别。精细的宇宙图景首先依赖这些并不壮观的计量约定，历史叙述也应保留其位置。

由此可见，丰度的 E 包含不可任意拆开的规定：对象的速度结构影响谱线归属，谱线归属影响元素比例，而元素比例又影响对象是否适合作为原初样本的判断。若全部步骤只凭同一个结论互相担保，便可能循环；若利用不同跃迁、不同对象与独立物理输入，就能让这种联合推断受到外部限制。

本书因而把“原初”视为一种需要建立的研究身份。它不是由对象年龄自动赠送，也不是研究者任意赋予。原初丰度是在模型、测量与天体加工历史的相互约束中形成的可检验推断。这一点既保护了物理史的真实约束，也让知识发生的构成条件保持可见。

38.3 多种元素能否互证

多个元素的丰度能够同时检验一个早期宇宙模型，但它们的作用并不相同。某种丰度对重子含量较敏感，另一种丰度可能更受膨胀速率或其他输入影响。它们的误差来源也不同：气体吸收与恒星大气测量具有不同系统困难。正是这种差异，使联合检验比重复同一类测量有潜在优势。

然而，潜在优势只有在依赖结构明确时才能兑现。两项预测可能共用同一个反应速率，或同一个宇宙学参数；两项观测也可能采用相同的原子数据与分析程序。若忽略共享误差，联合限制会过紧。反过来，共享某些输入并不取消全部互证，只要求研究者准确说明哪些误差可以独立压低，哪些必须共同保留。

与微波背景比较提供了更长距离的交叉。背景辐射与轻元素核合成对应宇宙演化中的不同阶段，观测通道也不同，但在一个模型内可以共同限制重子含量等量。Cooke 等人的 2018 年研究比较了氘推断与当时背景数据的结果，报告在其分析条件下相容。[38-3]这应当作为特定研究的结论表述，而不是没有日期和条件的永恒精度声明。

比较所共享的量时，还必须分清重子与全部物质。核合成主要追踪参与这套核反应的普通核子成分，不能把其推得的重子含量直接替换成宇宙中所有引力成分的总量。背景数据在

模型中也可以对不同成分具有不同响应。若两条链对重子含量相容，而动力学需要更多总物质，这两项判断在逻辑上可以同时成立；它们连接的是物质组成问题，并不相互抵消。

重子与光子之比也不是任意时代都无条件相同的口号。将早期核合成的比值与较晚背景推断连接起来，需要说明两个时期之间是否有改变相关数目或熵的过程。标准演化框架提供了一套连接，其他模型若改变它，就必须承担对两个时期各自观测的后果。跨时期相容所检验的因而不只是一项数字，也包括把两个时代联结起来的物理历史。

这项联系还能反过来限制某些新解释。若为解决一个局部丰度问题引入晚期能量注入，不能只计算该元素如何变化，还要检查光子背景及其他产物是否随之改变。一个新过程越处于多条证据链的交会处，解释责任就越广。复杂模型不是因复杂本身被排除，而是不能只享受增加自由度的好处，却忽略它带来的共同预测。

这种相容性强在哪里？如果两条链的主要误差来源不同，却在共享物理模型中给出相近结果，那么一个仅为某项资料调节的偶然解释就更难同时满足两者。但它仍不能独立证明共享框架的每一前提。联合成功通常支持的是一组相互联系的假设，进一步分辨各项假设还需要更多观测或实验。

一个很有用的现代复查是暂时撤去其中一类资料，用其余资料预测被撤去的量，再看预测与实际观察是否相符。这样的留出比较能减少“先用全部数据调好，再宣布全部数据吻合”的循环风险。它也让不同元素的独特角色显露出来：哪一种主要确定参数，哪一种承担外部检验，哪一种仍带来冲突。

另一项值得保留的差异，是预测误差与观测误差的交换。早期某个丰度测量很粗时，改善光谱可能带来最大进展；当观测已经足够精密，实验室反应速率的不确定性就可能成为主要限制。此时继续增加同类天体，不如测好一项相关核反应更能缩小模型比较的误差。望远镜之外的实验于是成为宇宙学的承重工作。

这种交换说明研究领域之间的联系并非仅是理论上的统一愿望。天文学结果可以对核物理实验提出具体的精度要求，核实验的新结果又会改变天文学参数与未解差异的评价。若一项反应的修订使氘更相容，却使另一项丰度更不相容，研究者不能只报告有利的一边。共同反应网络的约束力，正在于一个改动必须承担跨产物的后果。

从 SDE 六路径来看，由统一模型 E 进入可以展示多种丰度如何共同发生；由某一种观测 S 反查则会暴露样本选择与天体加工；由实验与谱线序列 D 进入又会显示不同输入的精度边界。路径转换的价值就在这些不对称之处。若无论从哪里进入都只得到“万物相互联系”，分析就没有达到应有的区分力。

38.4 部分成功与未解张力

锂问题使本章不能以全部证据圆满汇合结束。斯皮特夫妇1982年报告了古老贫金属矮星中的锂丰度，并讨论其宇宙学意义。[38-4]后来当其他资料更严格地限制模型输入时，标准计算与某些恒星锂丰度之间的不一致成为持续研究问题。Cyburt等人的计算明确保留了这一困难，而没有因为氘和氦表现较好就把它删去。[38-2]

这个案例首先要求区分“预测的原初锂”与“如今测得的恒星表面锂”。如果恒星加工使表面锂减少，两者不同不一定意味着早期核反应计算错误；如果加工不足以解释差异，就需要检查其他输入或理论。任何一种解释都必须给出定量后果，不能仅说“可能有复杂过程”便宣布问题解决。

同样，局部不一致也不能直接推出全部热宇宙框架失效。一个框架可能在若干独立量上表现良好，同时在一个重要量上存在尚未解释的困难。合理回应既不是以整体成功压制异常，也不是以单项异常消除全部成功，而是精确定位哪条推断链承受压力，并设计能够区分候选原因的研究。

对于锂，研究程序也可以按照责任分层。先检查谱线与恒星大气转换是否稳健，再检查恒星从形成到今天能否改变表面含量，然后检查核反应输入与早期模型。这个顺序不意味着前一层永远比后一层更可能出错，而是要求每个候选把改动位置说清。若一种解释同时改变多层，便需要说明哪项新增观测能够识别它。

这也说明共同体不必在问题未解时选择一种单一措辞。有的研究工作主要压缩核反应误差，有的测量不同恒星状态，有的探索额外物理；只要它们对同一差异给出可比较的后果，多条路线就能并行。保留路线多样性并不是平均分配可信度，而是让不同风险与检验方式继续存在，直至证据能够更清楚地区分。

这里的张力具有真实的创造潜能，但潜能不等于成果。它可能推动更精细的恒星模型、核实验与光谱分析，也可能长期未能闭合。SDE的三原理若用于此处，必须允许“张力存在而新的可靠解释尚未生成”。若每逢不一致就写成第三维已经诞生，发生学便成为预设成功的修辞。

轻元素研究最终给予天文学思想史的是一种成熟的整体性：不同物理时期、不同对象与不同测量通道可以共同约束宇宙历史，而不必在每一个局部都已完全一致。这个整体之所以有力量，恰恰因为它能够保存未解决的差异，并让差异继续决定研究方向。和谐不是把锂问题藏起来，而是使成功与未决各有清楚的位置，互不冒充。

本章参考文献

[38-1] Alpher, R. A.; Herman, R. Thermonuclear Reactions in the Expanding Universe. Physical Review 74, 1198, 1948. https://journals.aps.org/pr/abstract/10.1103/PhysRev.74.1198.2

[38-2] Cyburt, R. H.; Fields, B. D.; Olive, K. A.; Yeh, T.-H. Big Bang Nucleosynthesis: 2015. 2015 预印本, 2016 发表于 Reviews of Modern Physics。 https://arxiv.org/abs/1505.01076

[38-3] Cooke, R. J.; Pettini, M.; Steidel, C. C. One Percent Determination of the Primordial Deuterium Abundance. 2018. https://arxiv.org/abs/1710.11129

[38-4] Spite, M.; Spite, F. Lithium Abundance at the Formation of the Galaxy. Nature 297, 483–485, 1982. 本轮核读公开摘要。 https://www.nature.com/articles/297483a0

第 39 章 不可见质量：引力证据怎样改变物质分布图景

39.1 动力学为何可能看见不可发光成分

光告诉研究者哪些成分在特定波段发出辐射，运动则可以限制维持某种状态所需要的引力作用。当两者给出的图景不相称，问题并非简单的“眼睛没有看见”。它首先是不同测量关系之间的差异：根据发光成分估计的质量，是否足以解释对象的运动？这个问题允许多种回答，却要求每种回答面对同一组可检查的关系。

兹威基 1933 年关于河外星云的研究把星系团运动与质量问题联系起来，是不可见质量讨论的重要历史节点。[39-1]后来鲁宾、福特和索纳德对一组旋涡星系旋转性质的研究，进一步提供了星系内部运动分布的系统资料。[39-2]两条历史线索不能合并成同一个实验：星系团与单个星系具有不同的几何、动力状态和误差来源。

用现代简化例说明，在球对称质量分布与近圆轨道假设下，圆周运动满足 $v^2 \approx GM(<r)/r$。若全部相关质量已经集中在较小半径内，外侧速度应随半径以特定方式下降；若在某段范围内速度近似不变，则这一简化模型要求包围质量继续增加。实际盘星系需要更精细的几何处理，不能把球对称说明式直接当作全部数据分析。

星系团的推断与圆轨道例子又有区别。研究者看到的是许多成员星系的径向速度分布，通常不能逐一追踪完整三维轨道。在近似平衡等条件下，可以用速度弥散与系统尺度估计维持这套运动所需的质量，量级关系为 M 约与 $\sigma^2 R/G$ 成正比，其中 σ 表示速度弥散、R 表示相应尺度。这是对动力学方法的现代简化说明，比例系数与适用性还取决于密度分布和轨道结构。

一个速度异常大的对象究竟是真正成员，还是投影在同一方向上的前景或背景星系，会直接影响估计；一个正在合并、尚未达到平衡的系统，也不能无条件套用稳定系统的解释。由此可见，星系团目录的成员边界不是纯粹文书问题。它参与规定哪一组速度共同属于一个动力学对象，进而决定质量推断是否有意义。

但这些困难没有让研究退回只能数发光星系的阶段。它们反而规定后续应获得什么：更多成员的速度、系统的空间结构，以及能够不依赖同一平衡假设的质量指标。当一种方法的前提受到怀疑时，最有价值的进展常是找到改变前提的新测量，而不是只在原公式中选一个更合意的数。

这里推断的是引力质量分布，而不是已经辨认出某一种微观粒子。一个系统需要额外引力贡献，与贡献究竟来自何种物质，是不同层次的问题。更不能从某条旋转曲线直接推出所有暗物质候选的性质。把层次分开，才能既承认动力学证据的力量，又保留未知的真实范围。

不可见并不是没有任何显露。若某种成分通过引力影响运动、光线偏折或结构演化，它便通过这些互动进入研究。它缺少的是在某些电磁观测中的直接亮度对应，而非绝对退出一切经验。SDE 在这里能够提出清楚的对象观：对象身份由具体可重复的作用关系逐步建立，不能把视觉发光当成存在的唯一资格。

不过，这个对象观也不能替代物理竞争。引力推断需要既定定律与模型，如果定律在所考察范围需要修订，额外质量的估计也会改变。因此“看不见仍可以研究”只是允许问题成立，不是保证某一解释必然正确。真正的论证要进入不同探针能够施加的限制。

39.2 质量估计的不唯一性

从星系图像与谱线得到旋转曲线，需要几何。盘面相对于视线的倾角决定实际运动如何投影为径向速度，距离决定角半径怎样变成线半径。非圆运动、气体扰动和其他结构也可能影响解释。一个质量估计若没有这些条件，就只是把观测量代入不适合的公式。

发光质量本身也需要模型。相同亮度可以来自不同年龄和组成的恒星群，不同群体具有不同质量光度比。气体的贡献也不应忽略，且不同气体状态可能需要不同波段估计。若只拿可见照片的明暗当成全部普通物质分布，就会把没有在该波段显露的普通成分误算为更特殊的未知物质。

然而，指出这些不确定性并不意味着任何质量缺口都能被任意消除。可调参数受到其他观察限制。倾角不能为了让曲线符合预想而随意改变，恒星质量光度比也需要与光谱及群体模型相容。一个解释若在每个星系中都必须采用与独立资料冲突的参数，就失去说服力。自由度是需要证明合理的资源，而不是免费修补工具。

质量分解还存在一类结构性不唯一。同一条总旋转曲线，在一定范围内可以由不同的恒星盘质量与外延质量分布组合产生。总引力效果比较明确，不等于每个成分的贡献都同样明确。若只用旋转曲线同时确定所有成分，再宣称这些成分分别得到了旋转曲线的独立证明，就把一个联合拟合拆成了并不存在的多项证据。

不同半径的资料能够提供部分区分。较靠内的曲线可能更敏感于发光盘和核球，延伸到更大半径的运动则能限制更外部的引力分布。但“外部”并非自动等于“没有普通物质”，仍需估计气体和低亮度成分。方法上的进步不是先为每一段曲线指定一个名称，而是使不同成分的空间形状与其他可观测量共同接受比较。

距离误差也有具体后果。角半径转成物理半径需要距离，但谱线提供的径向速度并不会以相同比例随距离改变。因此，改变距离会改变由运动估计的质量尺度，也改变从通量推得的光度尺度，两者一般不同步。距离再标定史由此进入质量问题；它不是独立于暗物质研究的一段已经结束的前史，而是仍然嵌在其参数责任之中。

米尔格罗姆 1983 年提出修正牛顿动力学作为隐藏质量假说的替代路线。[39-3]其重要性在于把质量差异问题转向动力学规律的适用范围，而不只是再添一类发光不足的成分。公平比较必须承认这是一条具有可检验内容的研究路线，不能因为它没有采用主流对象名称就先宣布失去科学资格。

同样，替代路线也不能只凭成功描述某些旋转关系，就免除对星系团、透镜与宇宙学的责任。一个针对非相对论运动的方案，若要处理光传播和整体宇宙，还需要相应的完整理论形式。质量补充与引力修订的竞争，不能只选双方各自最擅长的一张图比较，而应考察跨尺度解释需要增加哪些条件。

SDE 的分析在这里应当保持中立的检验标准。由暗物质这个已成名称进入，容易把问题直接理解成“寻找它是哪种粒子”；由观测残差进入，则还会看见普通物质估计、几何和引力解释的多条责任。路径转换扩大了候选空间，但扩大之后必须用更具体的观测重新限制，不能停留在“任何说法都有可能”。

39.3 多通道约束的独立性

引力透镜提供了不同于恒星和气体运动的入口。背景光经过前景引力结构时，其图像会发生变化；在相应理论下，这些变化可以限制投影质量分布。透镜不需要把每个质量成分都当成已达到动力平衡的运动示踪者，但它仍依赖几何距离、背景源性质和引力理论。它独立于某些动力学假设，并不独立于全部物理前提。

Clowe 等人 2006 年对子弹星系团的透镜研究，把引力信号与可见星系、热气体分布进行比较，报告了支持不可见质量的重要空间分离证据。[39-4]这里关键的不只是“总质量太多”，还包括不同物质示踪与引力分布之间的位置关系。多个通道共同组成了一幅不能只从光学明暗获得的物质图景。

空间分离之所以有更强的区分力，是因为碰撞可以使不同成分表现出不同运动响应。热气体能够相互作用并发生明显的空间重排，星系中的恒星成分则在这一尺度上呈现不同的通过行为。Clowe 等人把这一系统中的透镜重建与气体、星系位置相比较，发现引力分布并不简单跟随主要发出 X 射线的气体分布。[39-4]问题因而由“质量还少多少”推进为“解释所需的引力分布应当在哪里”。

仅仅修改一个把普通质量总量换算为引力强度的系数，很难自动回答这种空间问题。一个完整的替代模型必须同时生成相关位置、形状及传播效应；若需要新增不发光成分，也应明确承认这一条件。这里不必预先裁决所有可能理论，便能够提出比单一总质量比较更严格的共同考题。

透镜图也需要谨慎阅读。它不是望远镜直接拍到的暗物质照片，而是从背景源形状等资料，在指定关系与重建方法下获得的投影引力信息。不同色彩叠加在一张图上，容易产生所有通道都以相同方式成像的印象；实际上，X 射线亮度、可见星系位置和透镜质量重建分别承担不同操作。将三者对齐，正是论证的重要工作，不是论证开始之前已经完成的视觉事实。

这项结果的论文标题采用了非常强的表述，但本书仍区分证据层次：它不是在实验室中捕获某种暗物质粒子，也不是自动排除所有可能的引力理论。其力量来自一个具体系统对解释施加的多重约束。任何替代模型都必须说明为什么相关分布会出现这种关系，不能只重申一般性的“引力也许不同”。

宇宙微波背景与大尺度结构又在更大范围提出限制。组成不同的宇宙会具有不同的扰动演化与结构形成行为，在给定模型下可以与数据比较。这里的质量成分不能只为某一个星系单独调整，而要在统一框架中同时承担多个时期和尺度的后果。这种跨尺度责任，是物质解释获得力量的重要来源。

但联合证据仍需检查共享模型。如果旋转、透镜和背景分析都采用同一个引力框架，它们的一致不能被表述为完全不依赖引力理论的证明。相反，它们可以说明：在该框架及各自可检验的观测条件下，额外质量的解释具有广泛相容性。这样的表述并不软弱，因为明确条件后，竞争者才知道自己必须在哪些地方提供更好的说明。

跨尺度比较还不能只比较“都有缺口”这一句话。一个物质模型需要给出从较早状态形成今日分布的动力学途径，说明这种成分如何聚集、与普通物质如何相互影响，以及哪些尺度上应留下不同结构。若一个模型在星系团总质量上可行，却在形成过程上无法产生相应分布，它仍未完成全部任务。对象分布与对象来源是相接的两个问题。

同理，普通物质过程可能改变星系中心的运动和密度结构。这样的改变不能被事先禁止，也不能被当作无限可调的解释。若使用恒星形成或气体流动说明某种质量分布，就要检查能量、时间与物质交换是否足以产生所需效果。把复杂天体过程纳入比较，会增加计算困难，也增加能够由不同观测限制的物理责任。

这使模拟结果的地位需要明确。计算机能够在给定初始条件、物理规则与分辨率下产生一个可能的演化世界，但与某一幅观测图相似，不是全部验证。需要比较它使用了哪些参

数、哪些资料参与了调节，以及哪些结果属于未被直接用于校准的预测。模拟的解释价值来自受约束的生成能力，不来自图像本身看起来足够真实。

一个真正独立的新探针，应当尽可能改变已有的误差结构或作用机制。例如，实验室相互作用搜索与天体引力推断承担不同任务；即使某类实验没有发现信号，也只限制与其实验灵敏度和模型相应的候选，不会自动抹去全部引力证据。同样，天体证据强也不能把任何候选粒子自动变成已发现实体。不同层次要通过具体模型相接。

39.4 暂定对象身份

暗物质的科学身份包含一种有意义的不对称：对某些引力效应的认识相当强，对其微观组成的认识仍有限。这并不要求在“已经完全知道”与“根本没有任何知识”之间二选一。科学对象可以先由稳定的作用约束获得身份，再由新的互动方式逐渐细化。天文学中大量对象都经过类似的分层确立。

但是，暂定身份不能成为无限延期检验的借口。具体候选应当给出可探测范围、相互作用方式或其他后果；模型若不断退到不可检验区域，就应当降低其解释地位。把多个候选统称暗物质，适合表述共同问题，却不能让它们相互借用成功。一个候选失败时，应准确说明被排除的是哪部分参数或机制。

还可以把对象身份分成几项逐步收紧的问题。第一项是引力作用是否超出了已计入发光与气体成分所能说明的范围；第二项是额外作用在空间上怎样分布、怎样随时间演化；第三项是若以新物质解释，它通过哪些相互作用参与世界；第四项才是某个具体微观候选能否承担前面全部要求。后一步应当继承前面的约束，却不能倒过来把前面证据自动记入自己名下。

实验室无信号的意义也因此十分具体。探测器只对某种质量范围、耦合强度和相互作用形式具有灵敏度，天体环境与局部候选分布还可能参与预期信号。一个排除结果可以非常重要，即使它没有回答全部质量问题；它压缩的是此前具有可行性的候选区域。若每次无信号都只说“暗物质仍然存在，所以没有影响”，便抹去了实验真正提供的信息；若直接说“暗物质已被否定”，又把局部限制扩大成了没有根据的总判决。

对科学史而言，这类结果值得与成功探测一样记录。它们改变研究者愿意投入的候选、仪器设计和理论范围，使暂定身份获得更明确的否定边界。一个对象可以通过被排除的相互作用而变得更受限制，但这种限制仍不能替代对剩余候选的正面识别。

同理，物质与引力修订之间的竞争也不是简单的实体数目竞赛。一个方案看似少加物质，却可能需要更复杂的动力结构；另一个方案保留引力定律，却需要新的物质种类与形成

历史。比较必须同时考察拟合能力、统一范围、自由度和独立预测，不能只用“更简单”一词替代这些具体问题。

本书的本体论解释在此保留一个开放结构。新的物质分布世界由光学、气体、运动和透镜等特征纠缠而成，无法再由发光图单独承担。这个世界的对象单位与证据关系已经改变，即使微观机制仍未完全确定。承认这一构成不等于预先选择所有物理细节，而是说明天文学怎样在看不见的地方形成可持续研究的对象。

从残差到对象、从对象到新探针、从新探针到身份修订，形成了一条可能继续发展的回路。它可以推进知识，也可能在某些候选上走入死路。引力分布的稳定推断不等于微观粒子已经被辨认，某一粒子候选受到限制也不等于全部质量差异消失。保留这两道界线，研究者才能既利用已有证据推进具体搜索，又让新的相互作用测量有能力改变对象身份。暗物质问题的下一步，应当发生在能够区分候选的观测与实验中。

本章参考文献

[39-1] Zwicky, F. Die Rotverschiebung von extragalaktischen Nebeln. 1933. 加州理工原文再版入口。
https://authors.library.caltech.edu/records/858yp-34954

[39-2] Rubin, V. C.; Ford, W. K.; Thonnard, N. Rotational Properties of 21 Sc Galaxies with a Large Range of Luminosities and Radii. Astrophysical Journal 238, 471–487, 1980.
https://articles.adsabs.harvard.edu/pdf/1980ApJ...238..471R

[39-3] Milgrom, M. A Modification of the Newtonian Dynamics as a Possible Alternative to the Hidden Mass Hypothesis. Astrophysical Journal 270, 365–370, 1983.
https://scixplorer.org/abs/1983ApJ...270..365M/abstract

[39-4] Clowe, D., et al. A Direct Empirical Proof of the Existence of Dark Matter. 2006.
https://arxiv.org/abs/astro-ph/0608407

第 40 章 宇宙加速与参数拟合：理论何时需要再打开

40.1 标准烛光的标准化过程

远方超新星比某个模型预期更暗，可能意味着距离更远，也可能意味着发光能力、尘埃或观测处理发生了差异。宇宙加速的证据正是在这种多义性中建立的。它不是望远镜直接看到星系运动越来越快，而是通过不同红移上的距离关系，比较宇宙演化模型。要理解这一转折，必须先理解超新星怎样成为可以比较的距离指示物。

Ia 型超新星并不是未经处理就具有完全相同峰值亮度的灯。研究者利用光变形状、颜色及其他信息进行标准化，再比较近处和远处样本。珀尔马特等人的研究明确使用光变宽度与光度的联系，并将远处超新星与近处样本联合分析。[40-1]标准化的必要性没有取消其测距价值，而是把可靠性要求从“每颗都相同”转向“差异能否被可检验地处理”。

获取光变曲线也意味着研究不能只依靠发现时的一张影像。需要追踪变亮变暗的过程，确定相应阶段，并把观测波段与对象静止参考系中的辐射联系起来。距离越远，观测频段所对应的对象辐射部分越需要注意。测得的是某一波段中的光，推断的则是经过一致定义的可比较亮度，中间存在明确的转换责任。

珀尔马特团队采用的“伸展”方法，可以说明标准化怎样在真实资料中工作。先以模板描述光变曲线，再允许其时间轴按伸展参数改变，以表示不同超新星内禀变化快慢的差别；峰值亮度的校正随之与该参数联系。[40-1]这里有两个容易混淆的时间因素。宇宙学红移使观测到的时间间隔相对于发射者参考系变长；内禀伸展则描述把红移效应处理之后，不同超新星本身仍存在的光变宽度差异。两项变换作用于同一份时间序列，却不是同一项物理解释。

若把内禀宽度差别全部误当成宇宙时间伸长，就会错误处理对象之间的差异；若把宇宙时间伸长全部吸收到自由模板里，也会使一个有明确预测的关系失去独立可检查性。因此，标准化程序必须保留每项变换的定义和适用位置。看似技术性的操作顺序，实际决定资料中的哪些变化被解释为天体性质，哪些变化被解释为传播与宇宙几何。

标准化的斜率也可以与宇宙学参数共同拟合，而不必先被当作毫无误差的固定常数。这样做承认对象模型和宇宙模型之间会共享不确定性。若预先固定斜率时遗漏其误差，最终宇宙参数的范围可能过窄；若允许过于自由的校正，则可能把本来应由宇宙模型解释的变化吸收到对象处理之中。合理的自由度应当由数据覆盖、近远样本的可比性及外部检验共同限制。

近处样本通常更明亮，资料也更完整，但这些优势不能原封不动地变成比较优势。在原研究的一致分析中，团队限制近处光变曲线所使用的时间范围，使其与远处样本能提供的范围相对应。[40-1]这种主动舍弃局部额外信息的处理并不是不重视好数据，而是避免一端用充分采样确定峰值，另一端却依赖不同程度的外推，然后把处理差别误认为宇宙差别。

这使“同一方法”获得了较严格的含义：并非只在程序开头调用同名算法，还要检查输入的时间覆盖、波段、信噪比和删选规则。两批对象若经历不同的可见性限制，即使使用同一个拟合函数，也未必形成同等的测量操作。标准化的目标因而不是抹除所有差别，而是知道剩下的差别是否会沿红移方向系统变化。

宿主星系背景同样参与问题。一个超新星叠加在星系上，研究者要分离爆发光与宿主光；尘埃会改变颜色与亮度，分类不纯则可能把不同爆发机制混进样本。若近处与远处对象的筛选标准不同，即使每个对象的测光都很精确，也可能在总体比较中产生偏差。因此，仪器精度、对象分类与样本选择应当共同进入标准化分析。

里斯等人 1998 年的论文报告了支持当前加速膨胀的超新星证据，并讨论演化、消光、选择、局部流动、透镜和污染等系统影响。[40-2]这说明承重论证并非“点落在曲线上”那么简单，而包含对替代误差来源的检查。本章仅将这些结论归于该论文的资料与模型条件，不把当时的显著性直接作为所有后续资料的共同数字。

E 的纠缠性在此具有可操作含义。对象被判为合适类型，才能使用标准化关系；标准化残差又帮助检查分类与环境影响；红移决定波段转换，转换结果又进入距离估计。每个量都不独立完成自己的任务。若只把这些写成一串背景因素，就会漏掉为什么一个环节的变化能在整个距离—红移关系中产生有方向的偏移。

40.2 加速结论的模型条件

膨胀与加速并不是同义词。尺度因子随时间增加，只说明宇宙在膨胀；其二阶时间导数为正，才对应这里讨论的加速。一个宇宙可以在膨胀同时减速，也可以在某些阶段减速、另一些阶段加速。不同历史会改变给定红移所对应的光度距离，因此可以通过标准化超新星比较。

距离—红移关系不是对加速度的直接局部读数，而涉及光传播所经历的整体历史。模型参数通过积分关系影响观测曲线，几项参数可能在有限范围内互相补偿。因此，拟合得到的不是一幅无需假设的宇宙录像，而是哪些历史在给定框架与误差下更符合资料的比较。

还要分清距离曲线的形状与绝对尺度。用星等表达，在简化情况下有 $m = M + 5\log_{10} D_L +$ 常数。若模型把光度距离整体按 H_0 的倒数缩放，而超新星绝对星等 M 没有外部标定，那么改变 H_0 的作用可以由改变 M 补偿。仅凭这种相对距离资料，不能同时无条件

确定两个量。这项退化却不等于无法研究曲线如何随红移弯曲，因为一个共同零点只能整体平移，不能随意改变其形状。

因此，加速研究并不必须先把整个绝对距离阶梯的每一级都定到最高精度。它可以利用近远样本之间的相对关系，对不同膨胀历史进行比较。与此同时，如果近远样本的零点并非真正共同，而存在随红移变化的校准偏差，这种偏差就会进入形状，成为更直接的威胁。区分整体偏移与红移依赖的偏移，能够使系统误差检查集中到真正影响加速判断的位置。

另一项现代说明可以从压力入手。在把宇宙学常数等效写成一种具有密度 ρ 和压力 p 的成分时，若定义 $w=p/(\rho c^2)$，它对加速度方程的贡献包含 $\rho(1+3w)$。对正密度而言，w 小于负三分之一可以带来相应的加速倾向，但总宇宙是否加速还取决于其他成分。这是特定理论中由方程得出的条件，不是通过超新星直接测得某种物质的压力，也不是任意一种负压物质已经得到微观识别。

此处的参数化具有双重价值。它允许研究者在尚不知道微观机制时比较一类动力学行为，也会引入新的退化与先验问题。把 w 固定为常数，与允许它随时间改变，并不要求同样多的信息；有限红移范围对复杂时间变化的约束，可能主要来自所选函数形式。新增参数应当对应资料能够回答的问题，否则看似更开放的模型反而可能使结论更依赖人为设定。

在均匀各向同性的相对论模型中，物质密度、曲率和宇宙学常数等成分共同决定距离关系。若加入平直性之类的先验限制，参数允许范围可能明显缩小；若放开限制，约束也会改变。科学报告必须说明这种变化，而不能把带有较强先验的结果表述成完全来自同一组超新星。

两支团队在 1990 年代末的研究提供了彼此重要的复查，其资料、分析与系统误差检验并不完全相同。[40-1][40-2]团队不同并不等于所有条件独立，因为它们仍可能共享超新星物理与宇宙模型。但部分独立已经具有价值：若不同处理和样本都指向相似结论，单一操作失误的解释就受到限制。

一个具体的思想实验可以显出证据的承重点。假设远处超新星在标准化后仍比近处同类对象内禀更暗，而研究者却继续赋予它们相同的标准亮度。由同样的观测通量反推，距离就会被估得过大。这正可能朝支持加速的方向移动距离关系。因此，演化检查不能只问远处对象是否“看起来差不多”，而应问尚未排除的差异有多大、符号如何、是否足以产生所见偏离。

里斯团队围绕这类问题比较光变和光谱信息，并讨论宿主环境、尘埃和样本选择。[40-2]这些检查不是对所有可能演化作一次性的逻辑否定。它们建立的是有条件的限制：某些能够显著改变亮度的机制，同时也应造成颜色、光谱或时间行为上的变化；如果这些伴随变化

没有出现，该机制作为主要解释的空间就变小。只调整峰值亮度而不交代其他可观测量的替代说法，不能与具有完整预测的模型等量齐观。

消光也显示出同样的论证结构。常见尘埃会使对象变暗并改变颜色，因此颜色提供另一个把手；若提出近于灰色的消光，使颜色变化很弱，就需要进一步解释粒子性质、空间分布和其他观测后果。替代解释可以增加复杂性，但新增条件会增加新的检验责任。不能因为某一种普通尘埃模型受限，就宣布一切消光解释已被纯逻辑排除；也不能因为能说出“特殊尘埃”，就认为已经同样解释了整套资料。

选择效应则可能让被发现的远处对象偏向较亮者，效应方向不一定与待解释的变暗相同。方向判断很重要：不是列出一个潜在误差名称，就已经削弱了结论。研究者要说明它在给定发现阈值、光变采样与筛选规则下产生什么偏移，偏移多大，与其他校正如何相关。一个完整的系统误差讨论应当把可能机制转成对数据的预期，而非把不确定性积成一张无法判断的清单。

将结论表述为“在已检查的对象模型与宇宙框架中，加速模型更能解释距离关系”，因此不是退让性的修辞，而是对推断形式的准确描述。后续研究可以通过更宽红移范围检验曲线形状，通过不同波段约束尘埃，通过更好的宿主资料检查种群差异。每增加一种能够区分解释的观测，就把最初的亮度偏离变成约束更密的证据网络。

加速膨胀与暗能量微观身份还必须分开。一个宇宙学常数项能够在模型中描述加速行为，不意味着其微观来源已经被确定。更一般的暗能量形式、引力修改或其他候选，也需要通过各自完整预测比较。参数符号代表模型中的作用方式，不能自动获得某一种具体实体的全部物理性质。

这种区分不是把可靠结论稀释成什么都不知道。研究者完全可以对某类加速行为具有较强证据，同时对其深层机制保持开放。对象的可确认特征与对象的全部本质不是同一个知识目标。SDE 分析在这里应保存分层可靠性，使已经形成的显露能够继续引导研究，又不把未知部分用一个名称封住。

从一个暗能量成品概念进入，容易先问它占多少；从测光与标准化进入，首先问资料如何支持距离；从动力学进入，则问何种成分能产生所需历史。三条路径共同工作，才能区分测量事实、参数推断和机制候选。它们不能互相替代，却可以互相限制。

40.3 独立探针怎样构成限制

若宇宙加速只依赖一种天体的亮度比较，就必须持续面对这种天体是否演化的疑问。增加不同类型探针的价值，是改变对系统误差的依赖。重子声学振荡提供一种来自物质分布的

标准尺度路线，微波背景则提供更早阶段的几何与物理约束，透镜及结构增长还可以限制物质如何随时间聚集。

Eisenstein 等人 2005 年在 SDSS 亮红星系的相关函数中报告声学峰，并将其作为宇宙距离的约束。[40-3]这里的标准尺不来自某颗对象的固定亮度，而来自分布中的统计特征。它因此与超新星具有不同的观测误差结构。不过，标准尺度的物理解释仍连接早期宇宙模型，不能把它描述为完全没有理论前提的尺子。

联合探针的力量可以用参数平面理解。一项观测可能允许一条狭长区域，另一项观测允许另一方向的区域；若二者交叠，联合限制就变得更小。这个几何图像只是说明统计原理，实际计算还需包括误差分布与相关性。若两项资料共用校准或宇宙样本，简单相乘可能重复计算信息。

共享天空区域也会带来相关。某些探针使用同一批星系、同一套红移或相近的前景模型，即使测量名称不同，误差也可能共同变化。宇宙样本自身的有限性还会使不同统计量受到相同大尺度结构影响。因此，联合分析不仅是“多收几种数据”，还需要建立这些数据如何共同发生的协方差描述。

统一模型也承担额外压力。它需要使几何关系、背景信息与结构增长相互相容，而不能为每个探针单独安排一套无关参数。如果一个模型能够拟合超新星，却要求与背景或增长资料冲突的条件，就必须进一步修订。反过来，某项张力出现时，也不能自动将问题归给整个模型，因为某个探针的校准或选择效应仍可能承担责任。

这种结构展示了统一性与多样性的积极关系。不同探针越具有真实差异，统一解释面临的检验就越强；如果全部探针只重复相同假设，表面上的多样并不会增加多少证据。和谐则是不同约束经过公开比较后形成的相容状态，不是研究者事先要求所有曲线必须交于一点。

本章由此把“精密宇宙学”理解为约束关系的精密化。精度不只意味着最后多报几位小数，还意味着知道那些小数来自哪些测量、哪些共同条件以及哪些尚未拆开的退化。一个误差条很小但依赖结构不清楚的参数，不比一个误差较大而责任明确的参数更成熟。

40.4 张力如何组织未来研究

加速证据建立之后，理论并没有因此永远封闭。新的资料可以提高对时间演化的敏感度，新的标准化可以改变系统误差，新探针可以打破旧退化。继续研究的合理目标，不是不断寻找能够宣布旧理论死亡的标题，而是明确哪些测量能够区分当前仍相容的候选。

盲分析提供一种方法上的约束。在关键选择完成之前暂时隐藏某些最终参数信息，可以减少研究者因为喜欢或不喜欢某个结果而调整处理的风险。它不是让研究者不理解资料，而

是把诊断工具、质量规则与最终结论的可见性作适当分离。具体项目如何实施需要依其程序，这里讨论的是避免反馈变成结果迎合的一般原则。

独立校准则针对另一类风险。若一个微小偏差同时影响许多对象，增加样本无法消除它；改变校准路线、比较不同仪器或使用不同对象，才有机会暴露共同偏差。前面距离尺度史已经说明，某些修订会保持总体关系却改变参数。因此，新观测既要检验加速的稳健性，也要明确它改变的是哪一层结论。

SDE 所强调的回写在这里具有双刃性。一个成功模型会组织望远镜时间、调查设计、模拟和分析程序，使研究变得高效；同时，它也可能让研究者较少关注模型没有预期的特征。成熟的回写应当保留失败通道，例如预先规定的异常检查、不同模型下的重分析与公开数据。模型可以引导注意，却不能垄断什么才算值得看见。

本章不展开另一部分快速发展的参数张力史，也不以未经核验的最新数字为加速发现补写结局。这里已经足以确定一个重要的思想变化：宇宙整体的组成开始通过多个时期、多个观测通道共同限制，而某个拟合项的微观机制仍可以开放。这使“知道宇宙怎样演化”与“知道全部组成究竟是什么”既相互关联，又不能混为一谈。

从距离标尺到动态宇宙，再到背景、元素与加速，本编形成了一种新的整体。它不是由一项万能测量建立，而是由各自有限的研究能力相互接合。这个整体能够精密运行，也能够在接合处重新打开。其创造性不在于终于不再有问题，而在于把过去无法提出的问题转成可以分工、比较和持续修订的研究对象。

本章参考文献

[40-1] Perlmutter, S., et al. Measurements of Omega and Lambda from 42 High-Redshift Supernovae. Astrophysical Journal 517, 565–586, 1999. https://arxiv.org/abs/astro-ph/9812133

[40-2] Riess, A. G., et al. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. Astronomical Journal 116, 1009–1038, 1998. https://arxiv.org/abs/astro-ph/9805201

[40-3] Eisenstein, D. J., et al. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. Astrophysical Journal 633, 560–574, 2005. https://arxiv.org/abs/astro-ph/0501171

第 09 编

多波段、多信使与不可见宇宙

第 41 章 无线电天空：新波段如何生成新的天体目录

第 42 章 X 射线与伽马射线：突破大气遮挡的观测

第 43 章 脉冲星与中子星：时间序列如何确证对象

第 44 章 引力波：非电磁信号怎样成为一类公共事件

第 45 章 EHT 黑洞图像：从干涉数据到稳健特征

第 41 章 无线电天空：新波段如何生成新的天体目录

41.1 从“噪声”到方向性信号

在无线电通信工程中，噪声通常意味着妨碍接收的东西。它可以来自雷暴、设备，也可以来自尚未辨明的方向。这样的名称首先表达一项工作的目的：希望获得的通信内容是信号，妨碍它的波动是噪声。天文学的一个新世界，恰恰从这种目的划分的边缘开始形成。当天线反复记录到一种不能随意消去、又具有方向规律的嘶声时，研究者面对的问题便不再只是怎样让接收更清楚，而是这种重复从何而来。

央斯基在 1933 年的报告中讨论来自太阳系外的无线电波，把到达方向随时间的变化当作判断线索。[41-1]这项工作的思想史意义不在于某次听见了奇怪的声音。声音可以被想象为无数来源，单次异常没有足够约束。重要的是将天线方向、地球转动和信号强弱组织成一段可以反复比较的差异序列。若变化跟随民用钟表所规定的一天，仍可能涉及太阳加热或人类活动；若其重复更接近恒星日，则需要转向相对恒星背景固定的方向解释。

这里的“来自天空”不是感觉强度的属性。接收机没有在噪声旁边打印宇宙来源的标签。方向性必须通过仪器响应辨认，周期必须通过跨日记录辨认，排除地面来源则依赖对另一组可产生同样读数的过程进行比较。三个判断相互约束：一个没有方向分辨能力的装置，即使记录到周期，也难以建立天区；一个有方向性的装置，如果只记录很短时间，也不足以区别天球运动和局部干扰。

由此可以看见，E 作为特征纠缠并非“工程环境加天文学知识”的清单。所谓固定方向，是在天线方向图、地球自转和信号重复共同规定下成立的特征；所谓天文周期，又必须借助方向与历时记录才能解释。抽去其中一个约束，剩下的并不是性质完全相同、只是可信度略低的对象，有时连研究对象的单位都会改变：它可能退回为无法定位的一段接收机波动。

把这一过程写成 S=F(D,E)，S 应是可辨认的天区射电显露，D 是方向与时间的差异比较，E 是这些特征互相赋予意义的组合。方程不承担电磁辐射的数值计算，也不代替排错记录；它提出的是一个构成问题：为什么在此之前被当作需要消除的波动，现在能够成为独立研究的天空？答案包括接收目的的改变，却不能只归因于目的。研究者愿意关注它以后，仍须让它经受可重复性与替代来源的约束。

为了理解方向证据的作用，可以设想两种不同检查。第一种只观察嘶声最强时天线指向哪里，第二种连续记录强弱随天线旋转的完整变化。第一种容易被偶然峰值诱导；第二种则

能够与方向图比较，判断信号是否具有稳定的天空分布。完整序列保留了反对最初猜测的机会，也使“最大值”不再脱离周围结构孤立存在。

另一个必要区分是重复性与来源唯一性。地面设备也可能每天规律运行，某个传播过程也可能使干扰随昼夜变化。恒星日对应之所以有辨别力，是因为它与太阳日、工作时刻等其他周期并不完全相同，跨更长时段观察能够让候选分开。这是一种依靠差异逐渐积累而形成同一性的过程，并非一次记录就拥有最终身份。

从这个角度看，偶然发现不是没有准备的反面。最初任务可以不是天文学，但仪器必须保存足够信息，研究者必须允许异常改变原有问题。若通信系统只输出“可用／不可用”两个状态，而不保存方向和时间结构，后续重解释的空间就会很小。知识发生的一项条件，正是已有任务没有把所有非目标差异提前抹掉。

41.2 天线如何成为望远镜

无线电天线成为天文望远镜，并不是把装置朝向星空后就自动完成。它必须能够说明自己对天空各方向的不同响应，能够区分天体变化和系统变化，并能够把不同时间得到的测量置于可比较的尺度。天文学研究所需要的不是一台“更灵敏的收音机”，而是一套使亮度、位置和频率可相互限定的操作系统。

波长越长，在口径相同时获得的角分辨能力通常越差。以现代近似式 $\theta \sim \lambda/B$ 说明，λ 是观测波长，B 是有效口径或干涉基线，θ 表示可分辨角尺度。这只是尺度关系，实际成像还受采样、加权与噪声约束。它解释了为什么射电研究不能简单照搬光学望远镜的尺寸：若想在长波段区分相邻结构，需要大口径或分散天线共同工作。

干涉测量使“一个望远镜”的含义发生改变。两处天线接收到的电场记录经延迟校正后比较，得到与天空亮度分布有关的相关量。阵列不是若干单独照片的拼接；基线长度和方向规定了能够采集的空间信息。增加一处站点，既可能增添灵敏度，也可能增添此前缺少的结构约束。地球自转使基线对天空的投影变化，时间序列因此进入成像过程。

这套操作改变了对象边界。一团在宽波束里连成整体的辐射，可能在更高分辨率下分成核心和延展结构；两个光学亮点，也可能处于同一射电结构之内。不能据此断言早期观测全错，因为早期资料支持的是相应分辨尺度上的显露。但也不能把后来的细分理解为仅仅给原有名词增加细节：当研究从“一个源有多亮”转向“核心、喷流和延展区域怎样联动”，解释单位已经重新组织。

目录在此承担双重作用。它保存结果，也决定下一轮观测要指向哪里。先以射电峰值建立目录，再寻找光学对应，会保留一些光学上不显眼的候选；先以光学星表挑选目标，再测它们的射电辐射，则可能系统性遗漏另一群对象。两条路径各有明确用途，却不是等价抽

样。六路径的意义就在这里：入口改变会影响哪些差异被保留下来，而不只是改变讲述顺序。

新显露还能回写仪器要求。位置误差过大导致光学对应含混，研究者便有理由提高定位能力；结构变化提示单一测量频率不足，观测策略便可能转向频谱与时变联合。这个回写须有实际项目或操作变化才能认定，不能因为新发现和新设备先后出现就断言前者造成后者。射电天文学的持续发展，正适合沿着这些具体要求追踪，而非把一切都压缩为“技术推动科学”。

阵列还带来一个容易被忽视的反例：分辨率更高并不保证对所有结构都更敏感。很长基线擅长约束紧致结构，却可能对大尺度平滑辐射缺少合适采样。如果一幅高分辨率图看不到某片扩展辐射，不能直接说它不存在。需要比较短基线或单天线资料，判断是否是空间尺度选择造成的“失踪”。

因此，“进步”必须说明所服务的对象尺度。对寻找紧致源而言，长基线可能提供决定性帮助；对测量扩展气体总量而言，覆盖大尺度信息同样重要。一个仪器体系的成熟，不是所有装置沿同一指标竞赛，而是不同尺度的测量能够互补。统一性必须建立在尺度接口上，而不能要求每台仪器都输出同一种对象。

目录中的源数也受分辨率影响。低分辨率时，几个近邻结构可能合为一个源；更高分辨率时，一个系统又可能分成多个峰值。若不记录源识别规则，比较两份目录的数量便可能把算法分割差异误当作宇宙族群变化。源的基本单位不是完全任意的，但必须由分辨、阈值与物理关联共同说明。

41.3 21 厘米谱线的新地图

连续射电辐射使天空有了新的亮度分布，中性氢 21 厘米谱线则进一步提供可与气体运动相联系的频率结构。尤恩与珀塞尔于 1951 年实现相关探测，现存装置档案以及尤恩的回顾保存了这一研究的技术背景。[41-2][41-3]这条谱线的重要性不只是穿透某些妨碍可见光观测的区域，还在于它把方向与速度信息放进同一资料体系。

一个天区的谱线并不总是单个尖峰。不同速度的气体沿同一视线贡献辐射，接收结果可以表现为复杂线形。于是，天空地图不再只是横纵两个角坐标上的亮度图，而可以成为方向与频率组成的数据集合。每个频率通道揭示一部分速度结构，但不能把每个通道直接当成一个确定距离的气体薄层。

原因在于径向速度到距离的转换需要银河运动模型。相同速度可能在不同位置出现，非圆周运动也可能使简单映射失效。如果直接把频率差等同于空间深度，就会把动力学假设悄

悄转化为“看到”的事实。严格做法是先说明方向与谱线是怎样取得的，再说明使用何种旋转关系推断空间分布，并标出歧义所在。

这一区分不削弱地图的科学价值。许多重要科学图像都是受约束的重建。问题在于能否保留从图像返回测量和模型的通道。对同一区域使用独立距离线索，或比较不同动力学假设产生的空间图景，能够揭示哪些结构稳健、哪些结构依赖特定转换。图像越直观，越需要让这种返回成为可能，否则颜色和形状会使读者忘记推断层次。

中性氢地图也给“创造新的天文学世界”提供具体含义。在旧的光学观察系统里，恒星位置和亮度可能是主要对象；在新的射电系统里，扩展气体、速度分量和连续结构成为可比较、可追踪的基本单位。不是人类让银河气体开始存在，而是新的 SDE 聚合体使气体以另一种具有操作规定性的世界成立。它有自己的对象、差异和关联，不能全部翻译成“原来那张天空照片看得更清楚”。

谱线强度与中性氢数量之间也不是无条件一一对应。在特定光学薄近似下，积分亮度可以与柱密度建立较简洁联系；若吸收与自吸收重要，就需要考虑激发和辐射转移。这里不应把近似式推广为任何天区都成立的测量定义。气体的“多少”在数据层面须经过物理模型才能确定。

相同视线上的多种成分还有另一种歧义：谱线分开不必表示空间上彼此孤立，谱线重叠也不必表示气体彼此接触。速度空间与位置空间是不同组织方式。研究者借助模型把它们联系，所形成的结构具有解释力量，但关系需要独立材料约束。这种差异说明所谓“新地图”不只换了一套坐标颜色，而是建立了不同属性之间的新转换制度。

如果把地图只当作一种主观表象，就无法解释为什么某些转换会失败、为什么另一些预测能被验证；如果把地图当作无需解释的实物副本，又无法说明模型改动为何重排结构。SDE 要求把地图放回互动整体中：它既受数据约束，又能指导观测和产生新预测，其真实性在可检验的作用链中表现出来。

41.4 跨波段身份确认

射电源与光学源是否属于同一天体，是需要建立的判断，不应是登记目录时的默认前提。位置接近只能提供初步支持。误差区域越大，其中偶然出现一个光学对象的可能性通常越高；背景源越密集，同样角距离所代表的证据力度也越弱。匹配必须将定位精度与候选分布一起考虑。

3C273 是这一问题的经典节点。施密特在 1963 年的论文中辨认其强烈红移的氢谱线，使一个外观近似恒星的光学对应体进入新的解释结构。[41-4]“像恒星”在此说明成像尺度

上的形态，不能独自决定物理类别。跨波段研究迫使形态、光谱和射电身份发生碰撞：原本足以完成初步分类的外观，不再足以承担最终归属。

这里应避免倒置历史。后来的类星体知识不能预先放进最初那张照片，也不能把精确定位、谱线辨认和距离解释合并成一次顿悟。每一环节增加的约束不同。定位让某个光学目标值得研究；谱线序列支持特定红移辨认；物理与宇宙学解释使红移获得距离及能量意义。一个环节的成功并不免除其他环节的验证。

跨波段同一性也不意味着所有结构必须重合。不同辐射机制可能突出同一系统的不同区域，中心与延展部分的相对亮度也可能不同。因此，合理匹配既要防止把无关源合并，也要防止因为形态不同而把同一系统拆散。可用证据包括定位、结构连接、变化关联以及符合物理机制的谱分布，不能只依靠最醒目的颜色块。

本章所讨论的新天空，是一套既能独立建立源集合、又能与其他天空交叉约束的世界。它的独立性保证异常不会全部被旧分类吞没，它的交叉性保证新目录不成为自我封闭的命名系统。统一性由可核查的对应关系实现，多样性由不同波段保存的对象单位实现，而和谐性体现为这些单位在共同研究中能够互相修正，既不强行同一，也不彼此隔绝。

一个实际可行的身份判断，应当允许“不匹配”成为有价值结果。若一个可靠射电源在现有光学资料中没有对应，它可能提示遮挡、光学极弱或资料深度不足，也可能暴露射电定位问题。将所有无对应源都删去，会使目录在建立之初就被光学先验筛选；把它们一律称为全新天体，也同样越过证据。

因此，对应工作的产物最好不仅有一个名称，还保留对应概率、误差范围和未解决原因。这样的信息让后续更深观测可以重新打开判断。身份的稳定不靠遗忘争议实现，而靠争议逐渐被更好的约束解决。未解决部分能够被准确记录，也是公共知识的一种成熟形态。

射电天文学说明，同一宇宙可以生成不同而可相连的科学世界。其差别并不只是人的心情或修辞，而体现为不同仪器互动下可成立的基本对象与属性。它们相互比较时，既可以建立共同天体，也可能发现旧同一性已经不够用。科学发展的创造性就在这种可约束的重组中展开。

这种重组还包含频率选择的作用。一个源在某个频段很强，在另一个频段可能很弱；以固定频段和阈值建立的目录，自然突出特定辐射性质。比较两个目录的源群时，必须先问它们观察了什么频率、达到什么灵敏度，不能把数量差异立即归为天体演化。多频资料能够限制谱形，但谱形拟合也要考虑非同时观测造成的时变混合。

因此，射电源的身份可以逐步增加维度：先有位置与亮度，再有谱形与结构，之后有时间变化和光学对应。这不是每个源必须经历的固定阶梯，而是可以根据问题选择的研究顺

序。身份越丰富，越需要保留各个属性取得的时刻和方法，避免把不同时期的状态拼成一个从未实际出现过的静态对象。

目录的可用性还取决于单位和基准是否一致。不同频率的流量、不同角分辨率下的峰值和积分量，不能未经转换就直接比较。保留这些元数据，能让一次发现从孤立记录成为后续研究可以接续的对象。

本章参考文献

[41-1] Jansky, K. G. Radio Waves from Outside the Solar System. Nature, 132, 66, 1933. https://www.nature.com/articles/132066a0

[41-2] NRAO. Ewen-Purcell Horn. https://www.gb.nrao.edu/~fghigo/fgdocs/HI21cm/ephorn.html

[41-3] Ewen, H. Doc Ewen: The Horn, HI, and Other Events in US Radio Astronomy. 2001 回顾，NRAO 档案。https://www.nrao.edu/archives/static/Ewen/ewen_HI.shtml

[41-4] Schmidt, M. 3C 273: A Star-Like Object with Large Red-Shift. Nature, 197, 1040, 1963. https://www.nature.com/articles/1971040a0

第 42 章 X 射线与伽马射线：突破大气遮挡的观测

42.1 为何必须离开大气

地球大气既使生命能够在地面活动，也使地面天文学只能取得部分来自宇宙的电磁信息。某些波段可以通过，某些波段被显著吸收。于是“天空没有发出什么”与“信号没有抵达仪器”成为必须区分的判断。可见光天空已经非常丰富，并不意味着可见光提供了天体全部重要活动的公平样本。

对大部分宇宙 X 射线，研究者需要将探测器带到吸收层上方。早期火箭实验提供短暂观测机会，卫星平台则使持续跟踪和大范围调查成为可能。NASA 保存的历史资料将 1962 年的火箭飞行列为成功探测太阳系外宇宙 X 射线源的关键事件，参与者包括贾科尼、古尔斯基、保利尼和罗西。[42-1]这里的突破首先发生在取得资料的条件上：此前无法形成的计数差异，终于能够被记录下来。

但“离开大气”不是一项完成后即可遗忘的背景条件。轨道、遮挡、探测器温度、粒子环境和有效观测时间都会改变资料性质。把同样的仪器放到不同轨道，不只是增加或减少数据量，也可能改变主要背景来源和可观察的时间尺度。一个突发过程如果发生在地球遮挡期间，未检出便不能解释为天体没有活动。

从 SDE 角度看，平台改变首先重构 D 的可能范围。研究者不再只能比较地面可取得的波段，而能比较高能计数随方向、时间和能量的分布。E 则体现在这些差异如何相互限定：某个方向的显著性依赖该方向的曝光，某个能段的强弱依赖探测效率，某种短时变化又依赖仪器在相应时间尺度上的稳定性。平台不是挂在科学之外的支架，而进入观测属性的构成。

这并不意味着所有高能天文学都要在太空完成。极高能伽马射线也可通过其在大气中引发的次级过程间接探测。这样，大气由遮挡者转为探测介质的一部分。两类方法使用不同观测量，误差结构也不同。思想史中的“突破”因而不能只画成从地面向太空的单线进步，而应理解为研究者不断重组哪些自然过程可以承担测量功能。

大气窗口还会改变研究者最初以为“典型”的天体。若某类过程主要在高能波段释放可探测辐射，它在光学目录中可能并不突出；若一种普通恒星在光学上很亮，也不代表它在 X 射线中占据同样地位。新波段因此重新排序天空的重要性。旧目录既可提供定位线索，也可能使研究过度围绕已有显著对象展开。

火箭与卫星的差别也不能只用观测时间长短表示。短时飞行必须在有限窗口内完成扫描，重复同一目标的机会较少；长期轨道平台可以积累变化序列，却要面对周期性遮挡与仪

器随时间的演变。两种资料对瞬变和稳定源的约束不同。为了说明某种变化真实存在，研究者需要先说明平台怎样取样时间。

如果一次高能事件在两个观测窗口之间发生并结束，仪器没有记录到它，是观察制度的事实而不是天体过程的否定。时间覆盖在瞬变研究中起到与空间分辨率类似的作用：它决定哪些结构可被区分。把遗漏事件纳入发生率估计，需要描述未观测时段，而不只是公布成功触发的清单。

42.2 事件计数不是照片

在许多高能观测中，仪器首先记录一个个候选事件：什么时候产生响应，响应落在何处，沉积了多少能量。读者最终看到的彩色图像，通常已是筛选、校准、分箱和背景处理后的表现形式。事件表与图像之间有一连串操作，不能因为最终产品适合视觉阅读，就把它视为普通相机直接拍下的天空。

一个区域记录到更多事件，可能因为它有更强的天体辐射，也可能因为曝光更长、探测效率更高，或者背景更强。最简单的现代说明模型可以写成预期计数等于源通量经过仪器响应的积分，再加上背景贡献。这个关系不是历史发现时所有研究者使用的统一原式，而是说明计数与物理源之间的必要中介。若响应随能量变化，不能用一个常数把所有计数换成能量。

背景估计尤其不能当作清理杂质的最后一步。若研究对象很弱，源与背景怎样分开本身就是核心问题。邻近区域可以提供背景线索，但邻近区域未必具有相同仪器响应；其他时段可以作对照，但粒子环境可能变化。好的分析需要检验这些替代背景是否足以改变结论，并说明模型无法约束的部分。

一个简单例子能够说明显著性与亮度的区别。假设两个方向记录相同事件数，第一处只观测很短时间，第二处观测很久。在相似背景下，它们代表的通量不同；若第二处还具有很高背景，则同样总数甚至可能不支持存在一个独立源。数字本身没有离开操作条件而成立的天文意义。所谓“强源”和“可靠源”是不同评价，不能相互替换。

在这里，E 并不是把背景、曝光和效率并排写在表格里。它要求看到三者的相互规定：曝光决定背景应积累多少，效率决定何种能量能进入事件表，背景模型又影响对低能或高能源谱的判断。如果错误地增加一个背景分量，源谱可能变硬或变软；如果误判效率下降，天体可能被误写为正在变暗。变化可以发生在天体，也可以发生在构成测量的系统中。

因此，公开一张图还不等于公开一项可复查的观测。事件筛选规则、响应文件、曝光信息和背景方案决定别人能否重建结论。公共性不是人人都看见相同图片，而是不同研究者能够进入足够明确的操作关系，判断同一物理主张是否经得起另一种合理处理。

计数统计可以用一个简化例子澄清。假设目标区域总计记录一百个事件，背景模型预计九十个，不能把一百个都算作源贡献。即使背景期望知道得相当准确，计数涨落仍使十个净事件的意义需要评估；若背景估计本身还不确定，结论就更依赖模型。这里的数值只是说明问题，实际资料应使用适当的统计方法而非机械套用正态近似。

能谱重建还涉及能量再分配。入射光子某一真实能量，并不必定落到探测器一个固定通道；探测响应可使它以一定概率出现在若干测量通道。因而一个通道的事件数不等于该能量处宇宙光子数。常用思路是让候选物理谱通过仪器响应，预测各通道计数，再与数据比较，而不是把原始通道曲线直接解释成天体本征谱。

如果一个看似突出的窄特征刚好位于响应变化较强的区域，就需要特别检查校准。另一台具有不同响应的仪器能够提供重要帮助，因为两者若在相同物理能量上显示相容特征，部分仪器解释会受限制。反之，仅在同一仪器多个处理版本中重复出现，不足以排除共同校准问题。

曝光地图也具有模型含义。它记录不同方向获得的有效观测条件，常随能量而变化。对扩展源，亮度分布与空间响应需要联合处理；若直接按原计数上色，图中的暗区可能只是曝光不足。读者看到的颜色越像真实结构，研究者越有责任说明响应校正是否可能生成或抹去它。

42.3 从定位到归属

探测到宇宙高能辐射之后，还有一个更艰难的问题：它来自什么？早期仪器的定位能力有限，一个误差区域内可能包含许多光学与射电对象。若只因某个对象最醒目就认定对应体，研究会把旧天空的显著性冒充新天空的身份证明。位置缩小、时间吻合和物理相容需要逐项建立。

伽马射线暴的历史将这种困难放大了。Vela 卫星原本服务于核试验监测，研究者却发现不能按原有任务来源解释的短暂辐射事件，相关宇宙伽马暴报告于 1973 年发表。[42-2][42-3]这种转向并非把军事任务名称改成天文任务名称那么简单。必须排除地面和太阳等来源，并利用多卫星的到达时间及响应寻找方向约束，突发事件才能获得宇宙归属。

“来自宇宙”仍然没有回答距离。一个附近较弱的爆发和一个远方极强的爆发，都可能给探测器相近的通量。没有距离，就不能可靠决定释放能量的尺度。天空分布能够排除或削弱某些空间模型，却不自动给每一次事件提供宿主。高能天文学因此长期承受一种张力：事件本身非常醒目，关于事件所在世界的知识却很不充分。

后来的余辉跟踪与更精确定位，使爆发能够与其他波段资料衔接。重要变化在于观测时间也成为匹配的一部分。稳定星表中的对象允许研究者从容复测，快速衰减的余辉却要求及

时响应；一次错过不能轻易由下个月补拍弥补。发现系统、告警传递、望远镜排程和后续光谱由此形成新的协作结构。它们并非纯行政附属，而决定哪些天体属性能够被接续地获得。

跨波段吻合仍须避免过度推断。同一误差框中出现瞬变，支持关联，但若时间窗放得过宽、候选范围过大，偶然对应概率也会增加。若研究者先看见一个吸引人的对应体，再回头选择有利时间窗，关联强度可能被夸大。可靠做法是让检索条件和背景事件率成为可讨论对象，而非只展示成功匹配的画面。

本章的路径转换因此具有实际作用：从一个已命名的高能源出发，容易沿着现有类别解释其谱；从事件背景、定位和时间筛选出发，则可能发现类别判断所依赖的弱点。若第二条路径不能动摇身份，原结论会增强；若它暴露显著的偶然匹配，就应将对象降回候选。这是研究能力，不是对发现的敌意。

从伽马暴天空分布推断来源，展示了统计对象与个体身份的区别。若事件在天空中近似各向同性，可以削弱某些明显集中在银河盘的模型；但近似均匀并不独自给出每个事件的距离。统计分布对模型族形成约束，个体宿主与红移则进一步提供具体尺度。二者解决的问题相连而不相同。

NASA 的高能天文学史保留了从早期事件到后续调查的技术脉络。[42-4]这些记录说明，同一个“伽马暴”名称之下，知识内容可以逐渐增加：先确立突发信号并非预期地面来源，再研究天空分布，继而寻求余辉与宿主，最后讨论物理机制。把最终解释写进第一份报告，会让读者误以为对象一出现便带着完整身份。

及时定位也使失败获得新的含义。一次未找到余辉，可能因为源确实很暗，也可能因为搜索太晚或范围太大。若不记录开始观测的时刻、覆盖区域和灵敏度，就无法把未检出用于模型比较。成功案例和失败案例必须共享足够清楚的记录标准，才能避免目录只保存幸运结果。

42.4 新对象的物理解释

高能辐射为研究高温气体、强引力环境和粒子加速过程打开了窗口。但观测到高能光子，并不表示它们只有一种生成机制。热等离子体、非热粒子和物质吸积可以产生不同却部分重叠的谱形。判断通常需要把能谱、时变、空间结构和其他波段结果联合起来。

X 射线双星体现了一种尺度压缩带来的新问题。可见光可能主要显示普通伴星，而 X 射线突出紧致天体附近的吸积活动。同一系统在不同波段具有不同的主要显露。若坚持每个波段都必须呈现同一种“恒星形象”，便无法理解系统的分工；若将每个波段各自看成完全独立天体，又会丢失轨道与能量交换的联系。对象统一性由物理关联建立，而不是由外观相似保证。

星系团热气体提供另一种反差。光学研究容易把星系团首先写成许多星系的集合，高能观测却使弥散气体成为不可忽视的组成。此时新增的不只是一个背景颜色，而是质量、热状态和动力学研究的新对象。推断总质量仍需说明物理假设，例如气体是否近似满足相应平衡条件；扰动明显的系统不能无条件套用同一模型。

SDE 解释在此应承担明确任务：说明何种属性组合使一个对象从旧类别中分化，及这种分化怎样改变后续测量。它不能凭借“纠缠”一词替代辐射转移、等离子体或引力理论。具体物理解释负责计算哪些机制能够产生这些数据，发生学分析则追踪这些计算、仪器条件与对象分类怎样共同成为新的研究世界。

新世界也可能保留长期未解区域。一个伽马源即使位置已经可靠，仍可能没有充分证据确定类别；一个变化谱即使统计显著，仍可能允许多种物理模型。成熟的目录应能够容纳这种分层状态。若每个新信号都必须立刻获得一个完整故事，研究就会通过命名制造虚假的确定性。

从大气遮挡到事件计数，从定位到机制，高能天文学显示“看见”可以是一种精密而公开的间接关系。它的可信度来自关系链上的交叉约束，以及允许身份判断在新证据下改变。新的天空没有消除解释的责任，而是使这种责任进入更细的时间、能量和空间尺度。

关于紧致源的尺度，快速变化可以提供一种有条件的限制。若一个区域的变化需要内部信息在有限速度下协调，则观察到的变化时间与其可能尺寸相关。但相对论运动、几何和辐射过程能够改变简单解释，不能把观测时间乘光速一律当作物体真实半径。数量级论证很有力量，前提是条件被明确保留。

能谱与时变的联合也能分开一部分退化。两种机制可能在有限能段内产生相近平均谱，但对变化过程的反应不同；反过来，相近时间曲线若在不同能段具有不同延迟，也可能约束辐射区关系。这样的差异比再给源增加一个形容词更有解释力，因为它能够指向可检验的物理过程。

高能研究的一个长期要求因此是保存事件层资料。最终图像与平均谱服务于特定问题，未来新的问题可能需要重新按时间、能量或位置划分。若只留下当前最满意的汇总图，知识平台就失去重新组织差异的能力；保留可重分析资料，则使旧观测仍能参与新的对象构成。

高能事件筛选还需考虑探测器的处理能力。强源可能使记录链条无法逐个保存所有到达事件，多个近乎同时发生的响应也可能被合并，改变测得的计数或能量分布。此时“更亮”不一定在原始记录中表现为简单同比增加。研究者必须使用适合的响应与质量控制，而不能把探测器饱和后的曲线当作天体本征行为。

这也使强源与弱源具有不同测量困难。弱源受背景限制，强源可能受记录非线性限制。同一套筛选对一类对象有效，不保证对另一类同样可靠。理解这些差异，才能让目录横跨大范围亮度而不制造虚假物理规律。

本章参考文献

[42-1] NASA Imagine the Universe. X-ray Astronomy.

https://imagine.gsfc.nasa.gov/science/toolbox/history_xray2.html

[42-2] NASA. NASA Looks Back at 50 Years of Gamma-Ray Burst Science. 2023.

https://www.nasa.gov/universe/nasa-looks-back-at-50-years-of-gamma-ray-burst-science/

[42-3] Bonnell, J. T. A Brief History of the Discovery of Cosmic Gamma-Ray Bursts. NASA 资料。

https://apod.nasa.gov/htmltest/jbonnell/www/grbhist.html

[42-4] NASA HEASARC. A History of Gamma-Ray Astronomy Including Related Discoveries.

https://heasarc.gsfc.nasa.gov/docs/history/

第 43 章 脉冲星与中子星：时间序列如何确证对象

43.1 周期为何可信

对于记录天空的装置，过于规则的信号有时比杂乱信号更令人怀疑。自然界可以有周期，电子系统也可以产生周期；地面发射机尤其容易制造规律。1967 年进入研究视野、1968 年正式报告的快速射电脉冲，把这种困难推到突出位置。休伊什、贝尔等人的论文报告一类具有极规则快速脉动的射电源，但其物理解释不是由“规则”二字直接给出的。[43-1]

首先需要确定的是，重复到底属于谁。若信号只在某台装置、某种设置下出现，它可能来自设备；若其出现时间随恒星日移动，并能定位到天空某方向，则天文解释获得支持。跨频率比较还可以检验传播效应，因为通过星际等离子体的射电脉冲，其不同频率成分不必同时抵达。这些检查把周期从一个容易误认的形式特征，转变为与位置、传播和重复观测相联系的天体属性。

周期的准确度也不是从单个峰值读出的。每个脉冲的形状和强弱可以变化，研究者常需要折叠多次记录，获得平均轮廓，再确定到达时刻。折叠需要一个试探周期，试探周期又通过跨时段相位能否保持来检验。只要原始记录、搜索范围和判据保留，这种往返就不是循环论证；它是一种参数估计。若仅在选中的短区间内成立，却无法预测随后记录，周期模型便受到限制。

这个例子尤其适合区分差异稳定与绝对不变。天体脉冲的可信周期并不要求每次观测完全一样，而要求变动具有可以解释和比较的结构。传播延迟、观测者运动和仪器时钟的影响可以使原始到达间隔变化。校正后能够形成相干序列，才使“天体时钟”成为有用说法。若把所有变化都视为误差并预先删去，就可能删掉真正重要的轨道信号。

SDE 中的 S 在这里首先是稳定可辨认的脉冲序列，不是中子星内部物质的直观呈现。D 包括搜索、重复、折叠和不同改正方案的比较，E 包括位置、传播规律和计时基准的相互规定。同一段原始记录放在错误时标下，得到的周期及其导数可能不同。因此，“周期属于天体”是经过共同约束成立的判断，而非仪器天然给出的无条件标签。

射电脉冲的频率依赖给来源判断增加了一种物理尺度。在冷等离子体近似下，传播延迟与频率平方的倒数相关，而比例系数涉及沿途电子柱密度。这里引用的是现代说明关系，不表示仅靠一条关系就能完成全部来源判定。它的作用在于把不同频率的到达差异组织成具有物理含义的结构，而不是把延迟当作无序散乱。

如果不先改正这种色散，不同频段的脉冲相加会变宽，微弱周期可能被掩盖。搜索因此需要尝试不同色散量，使脉冲在合理条件下对齐。但试探值越多，统计上偶然出现某种峰值的机会也会增加。确认时必须考虑搜索范围，不能只展示最漂亮的那一次折叠。

平均轮廓与单脉冲之间也应区分。一个对象可以具有相当稳定的平均脉冲形状，同时单次发射波动很大。研究若只保存平均轮廓，会失去部分辐射机制信息；若只盯单脉冲不规则性，又可能看不到长期稳定周期。不同时间粒度构成不同对象，选择哪一种取决于要解决的是身份、计时还是辐射物理问题。

观测者运动需要另作处理。地球绕太阳运行改变传播距离与到达时刻，把地面记录转到适当参考系有助于比较不同季节。若改正模型错误，残差可能出现与地球运动相关的周期。因而一个年尺度信号在被归为远方伴体之前，必须接受参考系检查。周期越接近测量系统已有节律，越应谨慎。

43.2 模型竞争

当信号的天文来源获得支持，问题从“是否真实”转向“什么能这样运行”。1968 年的最初论文仍讨论白矮星或中子星等可能性，[43-1]这提醒我们不要把后来的标准解释塞回发现瞬间。快速周期要求一个具有相应动力学尺度的系统，但周期自身不足以证明特定内部结构。

现代数量级分析可以说明约束怎样形成。对于自引力天体，特征动力学时间通常与平均密度的平方根倒数相关。这里使用的是解释尺度的近似关系，不是把某个单一公式当作历史争论的全部。较低密度的大天体难以维持非常短的整体运动时间，较致密对象则具备不同可能性。因而极短周期可以排除一部分普通恒星解释，但排除一部分模型不等于已经唯一选中某个模型。

旋转中子星模型把重复脉冲与方向性辐射联系起来：如果辐射束随旋转扫过观测方向，接收者就会看到脉冲。这个图景能够将极短周期、紧致尺度和长期计时联系起来。但灯塔比喻只提供几何直观，不能取代磁层辐射的具体物理。灯塔灯泡怎样发光与脉冲星粒子怎样辐射并不是同一个机制。

自转减速使模型获得更强约束。若周期缓慢增长，可以与能量损失联系起来；若辐射及周围结构所需能量与旋转能变化相容，解释便形成新的交叉支持。这里要谨慎区别模型依赖的推断量与直接计时量。由周期及其变化估计某种磁场强度或年龄，通常需要特定转矩规律和初始条件；它们不是独立于模型的测量。

同样，“看到了中子星”不等于已经直接辨认其内部所有物态。不同内部状态方程可能允许相近的外部计时表现。质量、半径、冷却和转动异常等资料能够进一步约束内部模型，

但从一条脉冲曲线跳到完整内部图景会抹去证据距离。本书强调科学世界构成，正要求保留这种层级，而不是把构成理解成语言可以补足一切缺失。

模型之间的比较应问：何种新观察能使解释分开？继续提高已经很准的周期，有时不如增加另一频段或获得伴星轨道资料有效。从“中子星应当是什么”出发，可能优先寻找理论预测的细节；从“这个信号还可能是什么”出发，则会优先寻找排他性证据。两种路径并行，可以避免理论使候选过早封闭，也避免排错永远停留在没有组织的怀疑。

对于旋转解释，能量预算提供超出几何比喻的限制。在简单刚体近似中，旋转能与转动惯量和角速度平方相关；周期缓慢增长意味着角速度降低。若把测得的周期导数用于估算能量损失，还必须知道或假定转动惯量，并明确观察到的周期变化包含哪些非本征贡献。这个推断链与直接测量周期不同。

所谓特征年龄也有类似问题。由周期与周期导数构造的时间尺度，只有在特定制动规律与初始条件下才接近真实年龄。若历史转矩变化或初始周期不可忽略，它就不能被无条件当作出生日期。这个案例提醒读者，名字中的“年龄”可能是模型导出的尺度，而非天体档案直接记载的时间。

脉冲突然发生的变化同样不能被简单视为计时失败。若排除仪器与传播后仍存在真实跳变，它可能约束天体内部与转动耦合。不过，从外部变化到内部机制仍有模型竞争。观测能使某些解释受限，却不意味着每次异常都直接揭开了中子星深处的结构。

因此，稳定性与异常并不是互相排斥的知识资源。没有稳定基线，异常很难辨认；没有保存异常，稳定模型又容易变成自我保护的封闭系统。脉冲星之所以成为精密工具，正因为研究同时记录规则与偏离，而不是为了让时钟看起来完美就删去所有不合格时段。

43.3 双星脉冲星计时与引力检验

赫尔斯与泰勒在 1975 年发表的双星脉冲星发现论文，为周期变化赋予了新的作用。论文描述测量周期时发现的变化，并将其放进双星轨道解释。[43-2]在这样的系统里，脉冲星既是被研究的天体，也成为探测轨道运动的计时器。原本可能妨碍“稳定周期”的变化，反而成为最有信息的差异。

脉冲星朝向或远离观察者运动时，脉冲到达间隔会发生多普勒变化。其在轨道不同位置到观测者的距离改变，还带来光行时效应。长期计时必须把这些影响与地球运动、星际传播和时钟转换分开。一个轨道模型如果正确，不只应贴合已见数据，还应预言随后一段时间的到达变化。

更精密的记录使相对论效应进入检验。近星点进动、时间延迟等可以在适用模型下对系统质量和轨道参数形成多重约束。研究的力量来自约束之间能否相容，而非把每一个拟合参数都称作独立证明。若两个量共享同一数据与模型，它们的相关性必须保留；若独立可测的后开普勒参数多于确定系统所需参数，剩余关系才提供额外一致性检查。

轨道周期的长期变化是另一个重要层次。通过引力波带走轨道能量的理论预言，能够与计时所得的轨道衰减比较。但实际比较还要考虑系统整体运动等影响，不能只把观测曲线与理论曲线重合的图当作全部论证。泰勒的诺贝尔演讲将长期计时及引力检验置于完整研究过程中，显示发现与精密检验之间的距离。[43-3]

这类证据与后来地面干涉仪直接记录引力波不同。双星脉冲星提供的是轨道动力学变化对辐射损失的检验，干涉仪提供的是经过地球的应变信号。两者相互支持，却不能在发现史中互相替代。也应区别“双星脉冲星”和“双脉冲星”：前者指含脉冲星的双星系统，后者还涉及两个成员都表现为可观测脉冲星的特殊情况。名称相近不意味着证据结构相同。

由此形成一种深刻回写。第一次脉冲星研究借助既有计时技术确定新对象；新对象稳定之后，又成为检验时空理论的工具。S 转而组织新的 D，新的 D 要求更严格的 E。这不是主体退场，而是人的判断、仪器改正、天体规律和理论约束组成新的研究主体能力。人不需要每次重新发明时钟，却必须知道何时封装的计时判断不再适用。

计时的“相位连接”可以用一个简单例子理解。假设两个观测时段之间缺了一段资料，研究者必须判断其间究竟转过多少圈。如果整周数不确定，多个周期解可能暂时相容。后续观测若能排除错误整周数，长期模型才真正连接起来。所谓多年连续计时，并不意味着望远镜每秒都盯着对象，而是模型能够在观测间隔之间保持可检验的一致计数。

轨道模型也是如此。普通开普勒参数描述主要周期结构，相对论修正则在更精细层面增加约束。若用过多自由参数吸收残差，可能使某些物理检验被提前消解；若模型过于简单，系统误差又可能被误认为新效应。需要让拟合自由度与待检验预言之间的关系保持清楚。

在质量平面上的约束比较，有助于展示这种清楚。一个已测后开普勒量，在指定引力理论下对应一组允许质量组合；另一个量给出另一组。若它们交于共同范围，理论获得一致性支持；若第三项独立约束也通过相同范围，则检验进一步增强。这里“独立”仍需考察参数估计相关性，不能只数图上画了几条线。[43-3]

长期轨道衰减尤其需要考虑系统相对太阳系的运动与银河引力环境。记录到的周期变化可能含有观测几何贡献，比较辐射预言时要将其分离。若改正量本身不确定，理论检验的精度也相应受限。这是一项重要纪律：不能只选择最符合理论的原始数字，而忽略使比较成立的条件。

43.4 研究条件的回写

脉冲星的后续研究逐渐形成长期监测的制度。单次看见与连续计时服务于不同问题：前者确认源的存在，后者积累缓慢变化和罕见异常。稳定的望远镜使用、时标、频率覆盖和数据保存，因此成为知识本身得以增长的条件。若只统计新发现了多少颗脉冲星，会忽略同一个对象几十年连续记录的独特价值。

计时残差说明这种制度为何不可省略。残差是记录与模型预言的差，不是“尚未清理的垃圾”的同义词。它可能来自仪器、星际介质、天体自转、轨道模型或更广泛的物理过程。每一类来源具有不同频率依赖、时间关联或空间关联。通过改变入口，研究者可以判断异常应写回仪器校准，还是写回天体模型。

如果多个对象同时出现类似残差，需要考虑共同的时钟或参考系问题；若只有一个对象变化，就要更多检查局部传播与天体过程。这个逻辑展示了 E 的整体约束：某颗脉冲星的性质并非仅由它自身资料决定，还可借助其他对象的共同或不共同变化辨认。但不能把所有共同变化立即归为宇宙新信号，因为共享设备和处理也能制造共同性。

计时阵列因此不是将很多独立“好时钟”简单相加。它要求研究不同天空方向的相关结构，并对共享误差作明确建模。数组的价值在于提供单颗脉冲星没有的区分能力。多样性在这里直接服务于统一检验：对象分布与传播条件不同，反而有助于定位共同因素。

脉冲星思想史最重要的转变之一，是将时间本身提升为对象身份的基本维度。一个天体不再只由天空位置和光谱被识别，还可以由相位连续性、周期演化和轨道调制被识别。新的天文学世界拥有自己的可见性：它未必表现为更漂亮的图像，却表现为能够在多年之后仍被追踪、比较和反驳的时间结构。

多频计时为残差分类提供具体办法。若某项变化明显随频率平方倒数变化，传播解释会得到支持；若变化在频率上近乎共同，则需要考虑不同来源。这个判据并非万能，因为频率相关的仪器效应也可能混入，但它为安排观测提供明确方向：扩大频率覆盖可以增加来源区分，而不只是提高同频数据量。

不同天区的脉冲星还可以帮助区分共同时间基准问题与方向相关参考系问题。前者可能在各对象中产生类似时间偏差，后者可能随天空方向呈现特定结构。阵列的力量来自这种可比较差异，不能由“加入更多脉冲星就平均掉一切噪声”概括。相关噪声不会像独立噪声那样简单消失。

研究的后续问题也因此很具体：哪一组残差需要更多频段，哪一颗对象需要更长观测弧，哪一种共同变化需要检查时钟或历表。脉冲星所创造的时间世界，并非远方放着一批天

然完美钟表，而是人们能够把不同误差结构逐项拆开，逐渐使某些极微弱的物理关系获得稳定显露。

计时精度还应区分单次到达时刻误差与长期模型准确度。某次测量很精确，不代表参考系或传播改正没有缓慢偏差；许多次较粗测量如果跨度足够长，也可能对某种长期趋势形成约束。研究者需要按目标时间尺度设计观测，而不是只追求单次误差最小。

观测频率与脉冲轮廓之间的联系也会影响时刻定义。不同频段的平均轮廓可能不同，若模板选择不当，得到的“到达时刻”会包含系统偏移。多频资料因此既帮助分离传播，也要求处理轮廓演化。所谓同一个脉冲抵达何时，在精密层次上必须有可复查的操作约定。

这种约定的价值，不在于消除一切人为选择，而在于使选择能够被他人采用、比较和修改。没有约定，各组高精度数字可能无法相接；约定若被当作不可更改，又可能掩盖真实轮廓差异。

若只提高单次精度而没有稳定的长期档案，某些缓慢变化仍无法判别。对于这类问题，持续观测的时间跨度本身就是测量资源，不能由一次更灵敏的短观测完全替代。

本章参考文献

[43-1] Hewish, A., Bell, S. J., et al. Observation of a Rapidly Pulsating Radio Source. Nature, 217, 709–713, 1968. https://www.nature.com/articles/217709a0

[43-2] Hulse, R. A., Taylor, J. H. Discovery of a Pulsar in a Binary System. Astrophysical Journal, 195, L51–L53, 1975. https://hosting.astro.cornell.edu/~shami/psrintro/papers/HT75.binarypsr.pdf

[43-3] Taylor, J. H. Binary Pulsars and Relativistic Gravity. Nobel Lecture, 1993. https://www.nobelprize.org/prizes/physics/1993/taylor/lecture/

第 44 章 引力波：非电磁信号怎样成为一类公共事件

44.1 事件如何越过噪声门槛

2015 年 9 月 14 日，LIGO 的两台探测器记录到后来命名为 GW150914 的短暂信号。2016 年发表的发现论文将它解释为双黑洞旋近、并合及余振的引力波，报告了两台站共同观测及统计显著性。[44-1]这次事件的特殊性在于：天文学不再只通过接收电磁辐射研究该系统，而是通过极微小的应变变化获得新的事件证据。

“极微小”并不等于“不可测”。测量的核心是让目标效应与其他变化在结构上可以区分。激光干涉仪比较不同臂方向上的光程变化，仪器输出经过校准才成为应变时间序列。地面振动、热运动、激光噪声和其他扰动都可能进入记录。因而一条形状漂亮的曲线，若没有关于测量系统和背景的说明，仍不足以承担天文学结论。

两座相距甚远的台站提供重要区分。某个局部干扰通常不会以适当的时间差和形态同时出现在另一台站，而传播中的天体信号需要满足两地间的光行时限制。但“双台站同时出现”也不是万能判据：还需考虑共同环境扰动、仪器状态和信号是否符合两台探测器的方向响应。独立性必须落实到误差来源，而不能仅凭地理距离宣布。

统计背景回答的不是“噪声会不会存在”，而是在既定搜索规则下，噪声产生同样或更极端候选的频率。合理改变台站时间对应关系，可以构造不再包含真实天体同步信号的背景估计，进而比较原候选的异常程度。这一方法的适用性依赖数据状态和噪声性质，需要明确搜索时段、筛选条件与统计量。若候选出现后才不断调整规则，使其看起来最显著，就会损害判据的公正性。

检出置信与源参数精度应分开理解。事件可以非常确定地属于引力波，而距离、自旋等参数仍有较宽不确定区间。因为确认“不是典型背景”与辨认“具体是哪一种源参数组合”是两个不同问题。混淆二者，会使公众以为任何参数变化都意味着发现被推翻，也会让研究者低估模型不确定性的重要性。

SDE 的关键不在于给两条波形分别贴上 S 和 D，而在于说明“事件”怎样由相互规定的特征构成。跨台站时差只在已知地理布局和时钟条件下有意义；波形一致只在响应与校准约束下成立；显著性又只能相对于明定搜索与背景来表达。E 正是这些约束的共同作用。脱离它们，保留下来的曲线仍可以观赏，却不再完整表达科学事件。

原始时间序列通常不会像宣传图中那样清楚地露出一段啁啾。不同频率的噪声强弱不同，特定仪器频率还可能形成狭窄谱线。将资料作适当频带选择或白化处理，可以让具有特定时频结构的信号更容易比较。LIGO 与 Virgo 的数据分析指南专门说明窗函数、频域转

换、噪声谱估计和似然计算等步骤。[44-5]这些步骤应当被解释为测量分析，而不是偷偷修饰图像。

白化的直观含义，是根据噪声在不同频率的强度重新加权。一个在噪声很大的频段出现的变化，与在噪声很小频段出现的同等变化，不能给予完全相同证据权重。重加权有助于比较，但输出曲线的视觉形态依赖处理条件。若只截取处理后的几百毫秒图像，不说明所用滤波，就难以让读者判断哪些形状来自资料，哪些来自显示方式。

滤波可能引入边缘效应，因此时间窗口也必须谨慎选择。对突然截断的数据作频域处理，可能把边界不连续扩散到其他频率。适当窗函数可以减弱这类问题，但也改变有效资料权重。科学展示最好保留足够前后时段与处理说明，使异常不是在一个人为挑出的狭小窗口里才显得特殊。

环境监测通道提供另一类证据。若候选时刻出现地面磁场、声学或其他扰动，应检查它能否耦合到主测量通道。辅助通道没有异常并不证明一切未知干扰都不存在，但可以排除已测且耦合方式明确的来源。排错的强度取决于监测覆盖与耦合知识，而不是一句“工程师已经检查过”。

一个事件在两台站幅度不同也不自动构成矛盾。探测器对天空方向和偏振的响应不同，同一波到达后可以产生不同投影。比较时应使用物理响应关系，而不是要求两条曲线逐点完全相等。相反，若两台站出现超出允许响应与时间差的结构，就需要重新评估来源。相容性是有模型内容的关系。

44.2 波形模板的帮助与限制

双黑洞并合的波形不是任意一段振动。广义相对论给出不同质量、自旋与轨道条件下的波形关系，数值计算和近似方法共同提供可供搜索的模板。匹配滤波利用这种结构，在噪声中寻找与模板相容的信号。它不是让理论替数据发言，而是明确提出：如果这个模型成立，数据应包含怎样的时间与频率关系。

理论参与测量，并不自动构成循环证明。循环危险出现在研究者只允许一种解释，又把找到该解释的最佳拟合当作排除一切替代。避免这一点，需要比较不同参数、不同波形方法、非建模搜索，以及扣除候选信号后剩余资料是否具有异常结构。合理的模板能够提高灵敏度，同时也规定了搜索最擅长发现什么。

例如，如果真实信号与模板族差异过大，匹配程度可能下降，源也可能被漏检。由此可见“没有找到某类事件”必须结合搜索覆盖范围理解。只对一部分质量、自旋或形态敏感的搜索，不能直接宣称宇宙中不存在其他可能。选择效应并非大规模目录才出现，它已经内在于第一项发现的技术形式。

参数之间的纠缠同样不可忽略。某些源参数组合可以产生相近波形，探测器方向响应也会使距离与取向关联。若把每个参数的单独误差条读作相互独立的测量，就会丢失联合推断结构。公开后验分布及其先验条件，有助于别人辨认哪些约束来自数据，哪些来自模型与范围选择。

在本书的六路径语言中，从理论条件出发形成模板、搜索数据并确认事件，是一种有效入口；从异常波形出发检查背景，再问哪些理论可解释，则提供互补入口。二者不会自动给出相同敏感度，也不必完全同等精确。它们的价值在于对不同失误敏感：前者擅长积累特定结构的微弱证据，后者有助于暴露模板预设过强的问题。

对 GW150914 的天体物理解释进一步提出双黑洞系统怎样形成、是否能在宇宙年龄内并合等问题。相关论文把观测与不同形成途径联系起来，但一次事件并不足以唯一选定所有形成机制。[44-2]这正是 S 向 D 回写的具体方向：新事件不只给已有理论盖章，也使研究者重新安排种群模拟、事件率估计和后续搜索。

若没有这种回写，发现可能停留在一张纪念性的曲线上。回写使它成为可继续发生知识的条件：之后的不同质量与不同环境事件可以检验最初解释的适用范围。这里的发展并非每次都淘汰旧理论，也可能是把原先宽泛的一组可能性逐步分化成可区分的模型。

啁啾形态为什么能够限制质量，可以从轨道能量变化理解。双星旋近时，轨道频率上升，引力辐射频率也随之变化。在适用近似中，频率变化速率与质量的特定组合相关，这个组合常称啁啾质量。它不是两个质量的简单算术平均，而是波形动力学所特别敏感的组合。观察到频率怎样加快，可以比单看某一瞬间振幅更有效地约束它。

但一个被精确约束的组合，不表示两个成员质量都同等精确。不同质量比可能在有限精度下产生相近演化，自旋也会与部分参数关联。参数推断应保留联合分布，才能看出哪些方向窄、哪些方向宽。把每一列参数写到相同小数位，容易制造不存在的同等精度。

波形的不同阶段还提供不同信息。旋近阶段追踪较长的轨道演化，并合附近涉及强非线性过程，余振则与最终对象的扰动响应有关。真实数据是否足够清楚地区分各阶段，取决于信噪比与仪器频带。不能因为理论模板包含完整三阶段，就宣称每个事件都独立测到了每一阶段的全部细节。

一个有用的一致性检验，是让不同频段或阶段对最终状态给出约束，再比较相容性。但这种拆分会减少各自可用信息，而且可能共享模型近似。它提供的是特定理论与分析框架中的检验，而不是完全免模型的证明。准确说明这种限制，才能避免把内部一致性夸大为对所有可能引力理论的全面裁决。

信号注入则检验搜索系统能否找回已知输入。指南介绍软件模拟注入与通过设备动作形成的硬件注入两类测试。[44-5]前者可以大量探索参数空间，后者触及更多仪器链条；二者

各有覆盖范围，不能互相替代。注入成功证明系统在相应条件下具有发现能力，并不单独证明某次未知候选一定来自天空。

44.3 独立团队和公开检验

一项巨大合作完成的发现，如何成为合作之外也能讨论的知识？仅仅公布作者名单或声明经过严格审核是不够的。GW150914 的公开资料页面提供数据与处理教程，使外部研究者能够接触构成结论的部分记录。[44-3]这种公开不会让每个人瞬间具有仪器专家的能力，却改变了质疑与复核的条件。

必须区分几种“独立”。两台探测器提供不同地点的测量，但共享某些设计思想；同一资料由不同算法分析，可以检验算法依赖，却没有增加新的原始观测；不同团队若使用相同校准文件，也可能共享校准偏差。独立性是分维度的，不能把团队数量简单换算成证据次数。

假设三组分析者都使用同一份错误标定，得到相近源振幅。这种一致不能排除标定错误。若另一种标定方法或独立仪器对同一物理量提供相容结果，证据结构才真正增加新的方向。因此，评估合作内部复核时，应问它改变了哪一项可能共同出错的条件，而不是只问多少人同意。

公开资料也不等于所有复现都具有同样深度。有人可能只复现经过滤波后的图形，有人重新做信号搜索，有人研究噪声分布，有人检验波形模型。每种工作可以回答不同问题。复现图形说明绘图和所选处理可以运行，却不自动复现完整发现概率；完全重复原代码，则较难检出原代码与假设共同持有的偏差。

严肃的公共检验需要把结论分解成层次：仪器中存在什么异常，异常是否具有天体关联，哪种源模型得到支持，参数估计依赖哪些前提。这样，一项局部修正不必被夸大为整体崩溃；反过来，发现总体可靠也不能成为拒绝修正局部参数的理由。公共性表现为能够精确说明争论发生在哪一层。

这种证据制度与 SDE 本体论之间存在可以认真讨论的联系。引力波事件作为公共科学对象，不是一个人私有的感受，也不是一串脱离操作的数值。它在可共同进入、可转译、可反驳的差异系统中稳定下来。其公共身份既要求仪器物理的可靠性，也要求记录与程序能够被跨主体接续。两种要求不能互相替代。

然而，这种分析并不能从“依赖共同程序”推导出物理事件由投票创造。远方并合的物理过程、地球上的测量过程和公共对象的确立，具有不同成立条件。新的天文学世界是在这些过程相互约束时形成的研究整体；它的真实性表现为能够承受不受研究者意愿控制的资料，而不是人人使用了相同词语。

公开代码与公开数据也有不同作用。数据允许研究者改变模型或统计方法，代码允许检查原分析怎样实现；校准说明与质量标记则告诉使用者哪些时段适合什么问题。如果缺少后两者，外部分析即使能够运行，也可能误用不适合的资料。开放不是把一个压缩包放上网就自然完成，而是使操作条件足以被理解。

对外部质疑，应当回答它具体改变了哪项条件。有的质疑指出滤波方式可能影响图形，有的质疑涉及背景估计，有的则提出另一种物理模型。前两者不必然否定后者的事件解释，但如果它们影响关键统计，必须重新评估。把所有异议都当成不懂仪器，或把任何处理争议都说成发现造假，都会阻碍有效检验。

同行复核也不是一次性的仪式。资料版本修订、噪声模型改进和新事件积累，可以使过去结果获得新的比较环境。一个早期极罕见的事件，随着同类波形增加，可以被放入更明确的种群；也可能暴露最初参数先验的限制。这种更新应以版本与来源记录为基础，而不是悄悄替换数字后让读者以为从来没有变化。

事件目录的可靠性又不同于单事件的显著性。把多个较弱候选合在一起估计发生率，需要处理天体概率与选择效应，不能把每个候选都硬分成绝对真实或绝对虚假。一个在个体层面尚不够确证的候选，可以在恰当统计处理中提供有限信息，但必须保留其不确定性，不可在目录总数中被无条件计为发现。

44.4 GW170817 的历史转折

2017 年的 GW170817 将另一种证据组合带入历史。联合观测论文报告引力波事件、随后出现的伽马暴，以及在相关天区寻找并追踪到的光学等波段对应体。[44-4]各通道不是把同一个故事重复播报，而是提供不同物理过程与时间阶段的约束。引力波限制致密双星的动力学，电磁信号揭示并合后的物质和辐射活动。

事件统一性首先要经受位置与时间的检验。两个信号接近，并不自动证明同源，必须考虑各自定位范围、事件背景率和时间关联。确认关联之后，才可以把它们的参数放进同一物理模型。先假定同源，再把所有相容特征当作证明，会隐藏推断的关键步骤。

多信使并不意味着每种信使都必然被看到。联合论文同时报告了某些通道的未检出。[44-4]未检出可以提供上限或限制，却不能被随意写成没有相关物理过程，因为探测灵敏度、方向和观测时段都参与解释。一部只保留成功亮点的思想史，会遗漏多信使证据的另一半：何处仍没有足够约束。

这次联合研究使“同一天体事件”成为由多种显露相互限定的对象。某个光学瞬变单独看有自己的分类问题，引力波单独看有自己的参数退化；相互联系后，一方提供的线索可以

改变另一方的解释范围。E 在这里不是把引力波、伽马射线和可见光并排摆放，而是使其时间、位置、距离和物理机制形成相互制约。

也不能将这种成功倒灌到 GW150914。后者的发现论证不以获得类似的明确电磁对应为必要条件。双黑洞系统与含中子星的并合具有不同物质环境，适合的验证路径也不同。统一方法不等于每个案例必须拥有完全相同证据，否则方法会把真实多样性裁剪成单一成功模板。

这类联合观测还有一个操作上的难题：时间紧迫时，研究者不能等待所有参数都达到最终精度才发出通知。初步定位与候选等级必须足够快地传递，同时标明不确定性，让其他望远镜知道该搜索多大区域、优先哪些目标。之后资料若改变，通知也应更新。迅速与可靠之间的平衡，不靠省略误差实现，而靠可修订的信息结构实现。

一次成功后随还不能证明告警制度已经对所有事件同样有效。目标在天空的位置、天气、季节、亮度衰减和仪器覆盖，都可能改变结果。无对应体的事件也需要记录搜索覆盖与深度，才能判断是源性质不同，还是没有充分观察。只有成功与未成功后随都进入档案，系统才有条件评估自身真正能力。

从长期脉冲星计时，到地面干涉仪的短暂应变，再到跨信使事件，研究者获得了不同时间尺度和不同载体上的天体世界。接下来的问题已经可以具体提出：哪些波形参数值得扩大频带去测，哪些定位误差值得增加台站去缩小，哪些电磁过程要求更快响应？新的天文学对象由此成为一组能够指导实验设计的问题，而不只是一次发现的纪念名称。

引力波的听觉展示也应被准确理解。将应变时间序列转换为音频，可以帮助人听出频率上升的结构，但探测器并没有在真空中收集空气声波。声音是资料的表现方式，天体信号则是引力波引起的测量响应。把二者分开不会削弱展示的感染力，反而说明数学与感官转译如何让不可直接听见的过程进入人的比较能力。

应变作为相对长度变化，也不同于某块物体在普通空间中平移了多远。仪器需要通过控制与校准把输出转换成相应物理量。若校准振幅存在不确定性，会传播到距离等参数；若相位校准存在偏差，则可能影响跨台站与波形比较。参数估计因此应纳入校准误差，而不是先把校准结果视为无限精确。

台站网络的作用还能由定位几何理解。两个台站的到达时间差可以限制一组可能方向，却通常不能独自把源缩到唯一一点；增加方向不同的台站和响应信息，可以进一步缩小区域。这里台站数量之外，空间布局也重要。若所有新增台站提供高度相似的几何条件，定位增益可能不同于在新的方位增加约束。

灵敏度提高也不只意味着同样事件变得更清楚。它可以扩大可探测距离，使进入目录的源群改变。若要比较不同观测期的事件数，必须考虑观测时长、有效灵敏度与搜索条件。某

期发现更多事件，可能主要因为仪器可进入的体积增加，而非宇宙并合率突然增长。这一点把单事件测量与种群推断连接起来。

不同质量系统在探测器频带中停留的时间和频率范围不同，搜索效率因此也不同。目录里的质量分布不能直接当作自然总体分布。推断形成渠道时，应让模型经过检测选择，再与样本比较；否则最容易被仪器看见的一群系统，可能被误写成宇宙最偏爱的形成结果。

非建模搜索同样不是“没有任何假设”。它可能不预设完整并合模板，却仍会使用时频局部性、跨台站相干性和一定信号形态条件。准确的比较应说减少了哪一类模型依赖，而非宣布进入纯数据世界。每种方法都有其擅长与遗漏，互补路线的价值正在于这些差异可以被说明。

最后，发现论文里的错误报警率不能直接解释为“这次事件为假的概率”。前者描述背景过程在指定搜索中产生极端候选的频率，后者还涉及天体事件率与假说比较。二者相关，却不是同一个统计对象。把频率语言与个体概率混用，会在公共传播中制造看似精确而实际错误的解释。

信号能否被公众听见或看见，是传播的成功；它能否在不同分析条件下保持天体解释，是证据的成功。两者可以相互促进，却不应相互替代。一段音频足以唤起兴趣，完整判断仍须回到校准、背景、波形与跨台站关系。

这一分工也避免把复杂技术全部转化为权威门槛。普通读者未必能够亲自复算，却可以理解哪一层结论依赖哪一类检查，并据此提出有针对性的问题。专业共同体的责任，是让这种分层可以被说明，而不是只要求外界相信专家数量。

对于研究者，真正值得保留的不是一种永远不会错的自我形象，而是当校准、模型或资料更新时，可以具体指出哪些结论保留、哪些需要改变的能力。事件因此才能越过一次新闻发布，进入长期可接续的天文学知识。

公开资料的版本号应与分析结果对应，避免不同研究实际上使用不同校准数据，却把差异误认成理论冲突。

本章参考文献

[44-1] LIGO Scientific Collaboration and Virgo Collaboration. Observation of Gravitational Waves from a Binary Black Hole Merger. Physical Review Letters, 116, 061102, 2016. https://arxiv.org/abs/1602.03837

[44-2] LIGO Scientific Collaboration and Virgo Collaboration. Astrophysical Implications of the Binary Black-Hole Merger GW150914. 2016. https://arxiv.org/abs/1602.03846

[44-3] Gravitational Wave Open Science Center. Data release for event GW150914. https://gwosc.org/events/GW150914/

[44-4] Abbott, B. P., et al. Multi-messenger Observations of a Binary Neutron Star Merger. Astrophysical Journal Letters, 848, L12, 2017. https://arxiv.org/abs/1710.05833

[44-5] LIGO Scientific Collaboration and Virgo Collaboration. A Guide to LIGO–Virgo Detector Noise and Extraction of Transient Gravitational-Wave Signals. 2019 预印本，2020 发表。https://arxiv.org/abs/1908.11170

第 45 章 EHT 黑洞图像：从干涉数据到稳健特征

45.1 原始数据是什么

2019 年公布的 M87*图像，让一个带有亮环和暗中心的结构进入公共视觉。它具有照片的直观力量，但其形成方式与日常相机显著不同。事件视界望远镜利用分布各地的射电设施组成甚长基线干涉阵列，对毫米波信号进行相关处理。2017 年观测、资料校准、成像与物理解释被分别写入系列论文，[45-1][45-2][45-3]这种分工本身就提醒读者：图像是研究链中的一个层次。

单个站点没有各自拍出一个黑洞局部，然后拼成圆环。相关器比较不同站点记录的信号，经过时间和频率校正获得基线上的可见度信息。在简化的远场成像描述中，可见度与天空亮度分布的傅里叶分量相关。不同基线提供不同空间频率的样本，地球转动改变投影基线，逐渐增加可用约束。

这意味着原始资料并不是一张布满未显影像素的照片。有限站点只覆盖部分空间频率，某些信息从未被直接采样。由有限可见度返回图像，是一个存在不唯一性的逆问题。很多亮度分布可能在测量误差范围内产生相近资料，重建任务必须判断哪些分布既符合数据，又符合明确提出的约束。

毫米波甚长基线观测还受到快速大气相位变化、站点差异和宽带记录等因素影响。EHT 资料处理论文讨论多套相位校准与条纹搜索流程，并通过质量检查比较一致性。[45-1]“地球大小的望远镜”因此是有用比喻，但不能误解为地球表面铺满了一面连续镜子。阵列的稀疏性是真实的，不能被宣传语抹去。

闭合量提供一种重要思路。在适当误差模型下，对多条基线的相位或振幅作组合，可以减弱某些站点误差影响。这并非消除所有误差，而是利用误差的共同结构构造更稳健的观测量。若某种系统误差不满足假定形式，闭合量也不能自动豁免。稳健性需要说明针对哪类扰动成立。

从 SDE 出发，图像 S 并非可见度 D 外面再加一个解释标签。某种空间结构能否显露，取决于哪些差异被阵列采到，以及采样、校准和重建约束如何共同作用。E 不是站点名单，而是这些特征的耦合：失去某种基线覆盖会改变可辨认结构，校准自由度改变会影响允许图像集合，重建约束又决定怎样在资料不足处作选择。

可见度为什么能够约束环，而不只是提供“有亮度”的证据？可以先考察一个完全理想化的教学例子，不把它当作 M87*的实际拟合。设天空上有一个均匀细圆环，圆心位于坐标

原点。沿相反方向的亮度贡献，在某些空间频率处会发生相位抵消，使相关振幅下降；在另一些空间频率处又可以恢复。环半径改变时，这种下降发生的位置也相应移动。

如果换成中心最亮、向外平滑降低的高斯形分布，其理想傅里叶振幅具有不同变化形式，不会简单复制细圆环的抵消模式。因而即使没有完成视觉成像，基线上的振幅怎样随空间频率变化，已经可以限制允许形态。真实目标并非完美细环，噪声和不对称也会改变图样，但这个例子说明：环形判断不是只由人眼看见一张圆图后产生。

EHT 校准论文报告了 M87 资料中相关流量的零陷结构。[45-1]对于思想史，关键不在于把每个基线数字背下来，而在于理解零陷所承担的证据工作：它使一部分无结构的紧致源模型受到限制，却还不足以唯一确定每个像素。一个整体尺度可以较早被约束，图像的局部形态则仍需要更多信息。

相位信息又补充振幅无法区分的部分。两个具有相同径向亮度分布但方向不同的不对称结构，可能在某些振幅比较中相近；多方向基线与相位关联可以帮助识别不对称。若只保留振幅而忽略相位条件，得到的结构集合通常更宽。因此“原始资料是什么”必须说到测量量的类型，不能只用数据量多少概括。

45.2 图像如何重建

逆问题并不因“不唯一”而失去科学价值。医学成像、地震反演和许多天文测量都需要从有限数据推断结构。关键问题是如何限制解，以及如何判断结论对限制条件是否敏感。EHT 成像论文明确研究图像特征对成像假设的依赖，而不是仅展示一张最吸引人的结果。[45-2]

一种重建方法可以偏好某些平滑性，另一种可以采用不同稀疏结构或正则化。选择这些约束时，不应将“符合先验”冒充“被资料独立证明”。但是，完全拒绝任何约束也未必更客观，因为有限数据无法唯一决定无限自由度。合理做法是公开约束，改变其强度和形式，观察哪些结构持续存在。

可以用一个思想实验理解这种检验。设有两种重建，一种亮环较厚，另一种较薄，两者都能拟合可见度。如果环的整体直径变化很小，而局部亮点位置变化很大，那么数据对整体尺度的支持强于对细小亮点的支持。把两种图像说成“完全一致”过于粗糙，把它们说成“互相否定”同样失真。科学判断需要按特征逐项比较。

模拟资料也能提供检验。研究者可以先选择已知结构，模拟阵列在相同采样和误差条件下会取得什么记录，再看重建能否恢复关键特征。测试对象不能只包含预期中的环形，否则算法擅长找环就可能被误当作无偏能力。不同形态的测试可以检验它是否会把非环结构也重建成环。

模拟成功仍有边界。真实天空可能包含模拟库未覆盖的变化，真实系统误差也可能比注入模型复杂。因而模拟不能与观测交叉检查互相替代。它回答的是“在这些假设内，方法能做到什么”，真实资料的稳健性分析则进一步问“改变合理处理后，关于这个目标的判断是否保留”。

图像的颜色也是需要说明的表达选择。射电强度被映射为视觉色彩，橙色并不表示人眼若接近 M87*就必然看到同样颜色。颜色有助于呈现亮度差异，但不应让读者把展示方式当成独立物理测量。图像并非因此虚假；准确的解释应说明视觉表面承载的是哪种观测量。

在科学世界构成问题上，EHT 具有特别清晰的示范意义。它让原本不能直接以日常视觉尺度观察的结构成为可共同查看的显露，但这种显露必须保留返回可见度、误差和模型的道路。图像不只是装饰知识的插图，而是新的比较界面；同样，它也不能脱离产生它的操作而独立承担所有结论。

重建问题还可以用一个有限维例子说明。假设待估计图像有一百个独立像素，而只有三十项相互独立的有效约束。即便每项约束都精确无误，一般也不可能唯一求出一百个自由数。非负性、有限视场或平滑约束缩小了解集合，却不能被误写成增加了七十次新的观测。这些是假设所承担的信息，不是探测器取得的信息。

实际图像像素彼此相关，测量也不是如此简单的线性无噪声系统，但这个教学例子能定位常见误解：高分辨率输出文件拥有很多像素，并不代表每个像素都有同样多的独立资料支持。为了显示和计算而细分网格，可以比有效分辨能力更精细。把图像放大，并不会使新增的小方格获得新的天文证据。

应当检验的是特征函数而非每个显示像素。例如，把图像转为环径、中心与环部亮度比、整体不对称方向等量，再比较它们在不同重建中的分布。某些特征变化很小，另一些变化很大，这样的结果比简单展示四张相似彩图更清楚。特征稳定还应在适当的测量误差与模型集合下定义，不能挑选一个有利的狭窄参数范围。

一个值得警惕的反例是，所有重建都被要求具有相同固定总流量、相似紧致程度和非负亮度。在这种条件下，整体亮区相近可能部分来自共同限制。如果数据还独立支持特定零陷与闭合量结构，论证便更强。验证的任务不是排除任何先验，而是拆清哪些结论即使放松先验仍然保留。

45.3 为什么需要独立团队

如果所有研究者一起观看同一个初步结果，再共同调整流程，最终一致可能包含相互暗示的影响。让团队在某些阶段分别处理资料，有助于减少这种人为趋同。不同成像方法和参

数扫描还可以揭示某个结构是否依赖单一路线。[45-2]团队差异在这里是一种认识资源，而不是必须尽早消除的管理成本。

但是，四组团队使用同一阵列、同一批资料，不会使原始观测数量变成四倍。它们对算法偏差的检验有价值，却可能共享采样缺口与某些校准问题。真正严谨的统一性，应当明确一致发生在哪一层：是图像整体环径相容，还是所有细节相同？是不同算法得出相近结果，还是独立天文观测也给出相容质量？不同回答具有不同证据重量。

EHT 的系列论文为这类区分提供了材料。资料校准、图像重建和物理模型拟合分别开展，使读者能够检查结论是否完全依赖成图这一环。若某个几何尺度还可通过可见度域的模型拟合得到相容限制，就增加了一条不必逐像素重建的分析途径。[45-3]这种途径仍共享原始资料，因而应称互补分析，而非完全独立发现。

移除站点或改变校准条件，是另一种有针对性的检验。它可以显示哪些结构由少数特定数据强力支配。但不能机械要求删除任何资料后结果都完全不变。某些关键基线本来就提供不可替代的信息，去除后不确定性增加是合理结果。重要的是变化是否符合信息损失的预期，是否出现此前未被承认的脆弱性。

这一点对 SDE 的“和谐性”有实际启发。和谐不是各种方法输出相同图片，而是差异能够被解释并用于界定结论范围。平滑方法弱化细节、另一方法保留更多局部结构，并不必然说明一方错误；如果它们在数据最有约束力的尺度上相容，就可以形成分层共识。反之，即使视觉上相似，若共同依赖未经检验的假设，也不能以和谐之名提前结束争论。

研究者还需要区分性能选择与结果选择。根据独立模拟和预先说明的指标选参数，与挑选最像预期黑洞图的参数不同。前者给方法表现提供外部尺度，后者可能把结论偷偷写进选择过程。公开选择流程，能让读者判断最终图像为何被保留，其他允许结果又为何被舍弃。

本章提出的三原理检验，因而可以落在一个具体张力上：显露要求直观明确，资料却稀疏不完备。解决不应靠抹掉不确定性，而应改变验证结构，让图像特征经过多路线和多条件检查。新的验证结构又可回写后续观测设计，告诉阵列哪些缺失信息最值得补充。发展发生在约束能力的增加，而不是展示图越来越漂亮。

可以进一步设计一项概念性的盲测：向分析团队提供三组具有相同噪声水平与采样的模拟资料，一组来自环，一组来自双点源，一组来自弥散不对称结构，但不告诉团队来源。若某条流水线把三组都变成圆环，它的形态选择就值得怀疑；若它能区分不同结构，又对噪声和校准扰动保持适当稳定，就提供较强的性能证据。这里描述的是检验逻辑，而不是声称历史上恰好采用这三种特定输入。

盲测完成后还要检查损失函数的选择。只评估总流量误差，不能证明形态可靠；只评估视觉相似，也可能忽略物理尺度偏差。若科学问题是测量环径，评价标准就应直接考察环径

恢复、偏差与覆盖率。如果科学问题改为判断细小亮斑，原先环径测试的成功不能直接迁移。这是一种按结论粒度配置验证粒度的原则。

多样方法的平衡由此不再抽象。某方法容易恢复紧致结构，某方法更稳健地保留弥散成分，某方法对校准误差更敏感。把这些差异保存下来，能够帮助设计下一轮观测。若为了统一发布图像而提前抹平差异，研究会失去知道自己哪里最弱的机会。

45.4 物理结论的分层

EHT 关于 M87*的结论至少需要分几个层次理解。首先是资料支持一个具有中心亮度下降的紧致环状结构；其次是其特征角尺度；再往后才是把这些结构与强引力下的发射和黑洞阴影模型联系，并结合距离推断质量。系列论文第六篇讨论几何模型、相对论磁流体模型与资料的比较，[45-3]不是从一个黑色圆洞直接称出黑洞质量。

暗中心也不等于拍到了事件视界内部。事件视界是因果边界，观测图像涉及周围发射、光线弯曲和捕获的共同作用。阴影的表观尺度与事件视界本身的几何尺寸不能简单画等号。若把亮环称为“黑洞表面”，还会引入不存在的普通固体边界想象。

环径与质量的联系包含距离信息。一个角尺度本身只说明在天空上看起来多大；要推断对应物理尺度，需要距离；要把物理尺度对应到引力尺度，需要模型。这并不使质量推断任意，反而说明不同来源的资料怎样共同缩小允许范围。独立动力学估计可与之比较，检验不同方法是否相容。

对广义相对论的检验也应避免无限升级。一幅与某类黑洞模型相容的图像，是对强引力描述的重要支持，却不等于一次性排除所有替代理论或测完所有时空参数。允许模型集合、发射物理和测量精度共同决定能够检验什么。若不说明范围，“验证爱因斯坦”会变成含义不确定的宣传口号。

从图像 S 倒查 E 与 D，可以把大众直觉逐层拆开：颜色来自显示映射，细节来自重建与采样，环尺度来自相对稳健的约束，物理质量来自角尺度、距离与模型的联合。反过来，从资料和条件正推，则可以说明为什么某些核心判断不会因图像细节改变而消失。这两条路径相互补充，使怀疑能够针对真正依赖项，而不必在全信与全否之间选择。

EHT 的新世界因此既是视觉的，又不是裸眼世界的简单扩大。它由跨大陆仪器、时间同步、波形相关、数学反演和物理模型共同构成，最终形成可公共讨论的空间显露。其创造性不在于任意画出黑洞，而在于建立一种能够抵抗任意图像的新型可见性。科学上的美也由此获得较严格含义：众多不同操作在有限目标上形成协调，同时把不能协调的部分保留为继续研究的边界。

距离依赖还可以用不涉及实际 M87 参数的计算示例明确表达。假设角尺度和相对论映射条件固定，采用的距离上调百分之十，则对应线性尺度也上调百分之十，依该映射推得的质量尺度会随之改变。这个变化不要求图像像素有任何变化。它说明质量结果可能因外部距离资料更新而修订，同时环形显露本身仍然稳定。

反过来，若距离固定而发射模型改变，观测环与引力尺度之间的映射也可能产生系统变化。因此最终误差不能只用图像重建的随机波动表示，还要反映物理映射条件。将统计误差与系统误差分开报告，目的正是避免一个看似精确的数字掩盖不同来源的不确定性。

这样的分层还有教学意义。读者可以同时接受三句话：图像揭示重要的稳健结构，个别像素并非逐点直接拍摄，黑洞参数仍有模型与距离依赖。这三句话彼此相容。科学可见性不必以零中介、零模型、零误差为标准，它需要的是中介被说明、模型被检验、误差不被隐藏。

图像的时间性质也是物理解释的一项限制。一次联合观测并非瞬间冻结整个目标，资料可能跨越多个时段。若目标结构在积累资料期间变化，静态重建便是在特定时间组织下获得的有效描述。不能把它的每一处亮度不对称都当作永久不动的地形。比较不同观测日和不同时间分组，有助于识别整体尺度稳定与局部变化之间的区别。

这种时间问题与阵列采样相互纠缠。地球转动帮助增加空间频率覆盖，同时也意味着不同基线信息不完全来自同一瞬间。当源变化不能忽略时，更多时间不必然等于更接近某一静态真图，而可能混合不同状态。研究必须在成像分辨率、信噪比和时间稳定之间作出有依据的选择。

还可以明确一个反事实检验：若未来新增基线使环径明显改变，而原不确定区间未覆盖这种变化，应重新检查过去的系统误差与模型约束；若新增资料主要改进局部亮度，却保留整体环尺度，则支持过去分层判断的合理性。这样的预期使稳健性成为可被后续研究检验的命题，不只是发布时的形容词。

读者最终得到的并非一个无需再问问题的黑洞图标，而是一组能够继续增长的约束。资料能够限制结构，结构能够约束物理，而新的物理问题又能具体指出下一代阵列需要增加什么。图像真正的力量正在这里：它使一个不可随意接近的强引力系统，成为能够通过明确操作逐步缩小未知的研究对象。

对图像的评价还应区分可视化清晰度与证据强度。一个经过适当平滑的展示图可能更容易阅读，却不会因此增加原始约束；一个不够漂亮的残差图，反而可能暴露重要模型问题。研究出版应让两者各司其职，不能只用最美的成图代表全部分析。

本章参考文献

[45-1] Event Horizon Telescope Collaboration. First M87 Event Horizon Telescope Results. III. Data Processing and Calibration. Astrophysical Journal Letters, 875, L3, 2019. https://arxiv.org/abs/1906.11240

[45-2] Event Horizon Telescope Collaboration. First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole. Astrophysical Journal Letters, 875, L4, 2019. https://arxiv.org/abs/1906.11241

[45-3] Event Horizon Telescope Collaboration. First M87 Event Horizon Telescope Results. VI. The Shadow and Mass of the Central Black Hole. Astrophysical Journal Letters, 875, L6, 2019. https://arxiv.org/abs/1906.11243

第 10 编

行星系统、系外行星与生命问题

第 46 章 太阳系探测：从远观到到场测量

第 47 章 行星、矮行星与类别边界：定义会怎样回写探索

第 48 章 第一批系外行星：径向速度怎样支持看不见的伴星

第 49 章 凌日、成像与系外行星人口统计

第 50 章 生命迹象与天文可居住性：证据距离有多远

第 46 章 太阳系探测：从远观到到场测量

46.1 “看见表面”的不同含义

一颗行星进入望远镜视野时，它可能首先是一片亮暗相间的圆面。探测器抵近之后，圆面分成地形；着陆器到达之后，地形又分成岩层、颗粒和化学样本。这里发生的不只是分辨率连续提高。不同工具能够与行星发生的互动改变了，因而可以成立的属性也改变了。远处的一块暗斑与钻取粉末中的矿物，并不是同一个观察单位的简单放大。

照片提供经过光照和反射共同作用的形态，雷达提供与电磁散射和传播有关的资料，光谱揭示对特定波长的响应，现场化学检测则让物质经历另一套操作。四者都可以称为研究表面，却不能将其读数当作同一种性质的直接副本。一个雷达明亮区域未必在可见光上同样明亮，一种光谱特征也不必唯一对应一种岩石成因。

“到场”尤其容易引发误解：既然仪器就在行星上，是否可以不再依靠模型？实际上，载荷仍需要校准，采样仍会改变物质，温度与粒度会影响响应，远程指令还需考虑通信延迟和能源限制。亲近对象减少了一些距离问题，却增加了接触、处理与局部代表性的问题。没有一种观测位置能自动免除解释责任。

以好奇号在盖尔撞击坑的研究为例，NASA 对任务目标的说明强调评价过去环境是否可能支持微生物生命，而不是宣布已经发现生命。[46-1]地面研究者通过车载仪器接续完成观察、取样和分析。这里的研究主体不是一位科学家缩小后坐在火星岩石旁边，而是由人员、软件、载荷、行星物质和通信关系构成的复合活动系统。

这种复合体的能力既有扩张，也有严格边界。它能在另一颗行星上实施精细操作，却不能随时走到任意地点。一个钻孔附近的判断较有直接支持，整个火星的推断则需要额外比较。若把局部可靠性无条件放大为全球可靠性，便会将到场测量的优势变成新的过度自信。

样本处理给“亲见”增加了一个不容易注意的层次。钻取、研磨、筛分或加热可以使仪器更容易检测成分，却也改变样本状态。加热后释放的气体，不必等于原先自由存在的气体；它可能来自矿物分解或样本内部反应。正确解释需要把处理步骤纳入化学模型，而不是把出现在检测器中的分子直接搬回原环境。

即使没有化学处理，图像中的颜色也依赖照明与相机响应。为了比较不同地点，研究者可能作白平衡或其他校正；为了展示微弱差异，也可能增强色彩。这些处理有明确用途，但公众应知道展示的是近似人眼外观，还是强调材料差别的科学图像。两者都可以真实服务研究，却回答不同问题。

原位测量的优势是能够增加互动方式。例如，先观察外表，再清理覆盖层，或在允许条件下取得内部材料，可以比较表面风化与内部组成。这样的操作比单一远观多了一种可干预差异，但依然不能随意重复：耗材、钻头状态和任务安全都约束操作。研究者需要把最有区分力的问题优先安排进去。

46.2 探测任务为何改变提问

远观阶段的一个自然问题是：那里有没有水？抵近研究之后，这个问题会分化为不同时间、位置与状态的问题：哪些岩层经历过液态水，水是否长期存在，化学条件如何，后来又发生了什么改变？新的资料不只是回答旧问题，也改变了问题的分辨率。

好奇号任务在黄刀湾等地点取得的岩石与化学资料，被用于讨论古代湖泊环境及其潜在宜居性。相关研究将沉积特征、元素和化学条件放在同一解释中，[46-2]其力量来自多种限制共同作用，而非某一个“像湖底”的照片。照片可以提出沉积假说，粒度和层理可以约束搬运过程，矿物和化学还可以检验环境性质。

思想史写作应当保留证据形成的次序，却不能把次序本身当因果机制。研究者先看见某种岩层，随后取样，再提出解释；这只是操作顺序。真正的论证还要说明为什么所见结构更符合某类过程，为什么替代成因受到限制，以及哪项新测量改变了判断。只有这样，探测史才不会退化为航天器走到哪里、拍了什么的旅行日记。

选址便是一个具体的回写环节。已有影像使某些地区被认为具有值得读取的历史记录，任务因此选择抵近或取样；新结果又改变后续路线。这个过程可以使研究越来越集中，也可能强化初始偏见。如果所有后续选择都服务于证明最初湖泊假说，别的地质过程便可能被忽视。因此需要保留对比地点，并让不符合预期的资料进入解释。

“有记录的地方”与“最典型的地方”也未必相同。研究者往往选择暴露良好、可抵达、操作安全的地点，但这些地点可能并不代表行星平均状态。工程约束与科学目标共同形成样本。E 由此不能只写成火星环境与人类技术，它应具体体现：可到达性、地层暴露、载荷作用方式和所求属性相互限定，改变任一条件都可能改变被选中的研究对象。

新的提问还可以跨越学科。某种层状结构使地质时间成为问题，化学能量使宜居环境成为问题，当前大气则使历史保存与侵蚀成为问题。行星不再只是绕太阳运动的球体，也成为相互关联的过程系统。这个世界通过探测活动构成，但其内部过程必须由物理证据限制，不能因理论偏好而任意填充。

沉积解释可以从颗粒入手。粒径、磨圆程度、分选与层理组合，能够限制搬运方式和沉积条件。但单一圆滑颗粒并不能独自证明长久河流，单一层状形态也可能由不同过程形成。

比较的力量来自一组特征是否共同支持同一种过程，以及替代过程是否必须附加过多不受支持的条件。

岩层的空间关系同样重要。上覆与下伏、切割与覆盖、坡面与层面之间的关系，可以建立相对先后。这样的先后关系与绝对年龄不同：看出某结构晚于另一结构，并不等于知道它发生于多少年前。若年代依据来自另一种测量或区域模型，必须说明连接方式，不能让一张地层剖面承担它没有提供的时间精度。

任务路线也不只是“从容易到困难”。先研究较易到达的岩层，可能帮助理解远处目标；但若路线选择持续绕开不符合预期的地形，得到的历史就可能偏斜。有限资源下无法遍历一切，关键在于记录为什么选择某点、未选择的区域有什么已知差别，以及这一选择会限制哪类外推。

一个合理的对照策略，是比较同一地层不同位置，或比较相似形态但不同矿物响应的区域。前者帮助区分局部改造与广泛环境，后者帮助判断外观相似是否对应相同过程。对照不是额外装饰，它使某一地点的显露能够放进差异序列中，避免孤立样本支持过多故事。

46.3 不同航天器数据如何衔接

轨道图像提供大范围背景，着陆器提供局部细节，两者的结合很有力量，也很容易产生不恰当的对应。轨道光谱的一个像元可能混合多种物质，车载样本却来自厘米乃至更小尺度。若局部检测没有发现轨道推断的矿物，不一定意味着某方错误；它可能没有取到该像元内的相关成分，也可能存在覆盖层或响应差异。

因此，衔接需要共同的空间框架。目标位置、地形和尺度必须明确，才能说明一个样本在大图中的代表性。时间也不能省略：尘暴、季节变化和光照几何都会影响不同观测的可比性。两幅颜色不同的图像，不足以直接证明表面发生化学变化，除非已经处理相应的观测条件。

长期记录尤其依赖标定。仪器老化、探测器响应变化或处理版本更新，都可能制造趋势。对于所谓长期变化，研究者应同时检验物理变化模型与测量变化模型。有时另一载荷或另一航天器能够提供独立支持，有时只能保留不确定范围。对比并非把两份图像并排放置，而是建立足够相同的比较条件。

NASA 关于盖尔撞击坑沉积与侵蚀的资料，将河流、湖泊和后期风蚀解释组织成多阶段图景。[46-3]这种图景必须明确区别已测地层与重建过程。地层是留下的结果，过去过程不能直接重新演一遍；研究依靠结构关系、物质性质与可行机制，判断怎样的历史能够产生这些结果。

这也说明三方程并非只适用于“从原因到结果”的单向故事。研究者从显露 S 出发，反推可能的差异过程 D 及约束 E；又通过模拟或地球对照检验这些过程是否真的能形成相应显露；结果若不相容，就要修正过程或条件。反推通常不唯一，所以必须比较多个历史，不能把最流畅的叙述当作唯一真实。

在统一性与多样性之间，资料接口起到核心作用。统一坐标、单位和质量标记，使不同任务能够对话；保留各载荷的原始响应与局限，使差异不被过度标准化。如果只保存一张统一地图而丢掉形成过程，后来的研究者就难以判断某处冲突来自天体还是来自拼接规则。可接续的档案因此是知识能够再发生的基础。

尺度转换可以用一个假设例子说明。假设轨道像元覆盖一平方千米，而车载分析来自一小片岩面。若这片岩面富含某种矿物，不能直接赋予整个像元同样比例；若像元平均响应显示该矿物，也不能保证任意一处岩面都会检测到。两者需要通过地质单元与混合比例联系，分辨率越不同，越要明确代表性。

同样，两个着陆点没有发现同样物质，不足以说明仪器互相冲突。行星本来可能不均匀，局部历史也不同。跨地点比较应尽可能选择相似测量方式和可比地质条件，或者建立差异模型。把不同地点测得不同结果一律叫作误差，会抹去行星真正的多样性；把所有差异一律归为地质，又可能遗漏校准问题。

时间变化的证据需要一组相容条件。假设某坡面在两次影像中出现新的暗带，应先检查太阳高度、观察角度、分辨率与处理方式，再考虑物理变化。即使确认形态改变，仍要比较尘埃移动、坡面过程或其他可能来源，不能一步跳到最引人注目的解释。证据链上的每次升级都应对应新的区分能力。

历史模型可以通过前向模拟受到约束。若某种沉积与侵蚀过程被认为形成了现存层序，就应问它是否能够产生相应几何、粒度与空间分布。一个能讲通局部照片、却不能解释周围结构的模型，需要修正。反推获得候选，前推检查可行性，再用其他地点比较范围，这种往返比单线叙事更能保护发生解释的严谨性。

46.4 行星作为过程系统

当行星被当作过程系统，其“现在是什么”与“怎样成为现在”开始紧密关联。但两者仍需区分。今日干燥的表面可以保存古代水活动记录，过去某处适宜生命不意味着今天同样适宜，更不意味着那里曾有生命。宜居性是条件判断，生命存在是另一层命题，不能因为故事连贯便跨越证据间隙。

盖尔撞击坑的研究还提示地质记录并非透明档案。后来流体和化学过程可以改变原有矿物，某些历史线索可能被擦除或重写。NASA 关于岩石记录改变的报道介绍了相关矿物研

究，[46-4]使“找到古老沉积物”与“读到未经改变的古老环境”之间出现必须讨论的距离。

这个距离为 SDE 回写提供了自然科学中的具体案例，但应避免混淆层次。行星上的后期过程改变岩石，是物理历史中的改造；研究者因新检测修正古环境解释，是知识过程中的改写。两者相关，却不能用一个词跳过中介。明确哪一种变化发生在物质、哪一种发生在研究系统，才能把发生学真正落到实处。

过程视角也改变探测的价值判断。一个没有发现生命的任务，仍可能通过排除某类环境、识别保存条件或建立可靠对照，显著推进研究。若把任务成败只绑定于最吸引公众的终极问题，研究者容易压缩不确定性，忽视那些改善未来判断能力的成果。知识增长可以表现为找到对象，也可以表现为知道哪里不该作强断言。

从远观到到场，人类并没有摆脱主客互动而获得“无中介的行星”。相反，互动的层次与能力变得更丰富：轨道观测、移动、接触、钻取和化学操作共同生成新的显露。行星研究的成熟，体现为能够让这些不同显露相互校验，同时承认局部样本、有限行程和历史重建的边界。这使新的行星世界既有可进入的细节，也有不能任意越过的尺度。

地球类比提供机制资源，也带来迁移风险。相似层理或矿物可以启发解释，但火星重力、大气、温度与化学环境不同，相同外观未必具有完全相同成因。类比应该提出可检验假说，而不是通过相似照片直接宣布同一过程。若模型能说明环境差异如何改变结果，类比才真正成为跨行星推理。

“适合微生物”的判断也需要多条件组合。液态水只是其中一项，化学可用性、能量来源和环境持续性同样重要。Grotzinger 等人的研究之所以不仅是发现水的另一种说法，是因为其讨论的条件组合更具体。[46-2]但即便组合相容，也只能说明一个潜在生境，不能替代生命痕迹本身的检测。

保存条件使这一问题再分叉。一处环境曾经宜居，未必保存了容易辨认的生命痕迹；一处保存较好，也不必最适合生命。后续任务选址需要平衡曾经可能有生命、记录能够保存以及仪器能够辨认三项条件。它们若被压成一个“最宜居地点”的排名，实际研究目标就会变得含混。

到场测量最终提供的，是把这些分叉变成具体操作的机会：在何处观察层序，在哪种材料中取样，选哪一种处理检验来源。一个有价值的探测计划，应当预先说明出现不同结果时怎样改道。若任何结果都被写成支持原假说，航天器即使抵达遥远行星，研究仍可能困在最初的概念之中。

取样位置的记录具有与测量值同等的重要性。一份化学分析若失去其地层、朝向和邻近结构信息，仍然是有用成分数据，却难以承担同样强的历史解释。材料属性与空间关系共同限定它能说明哪个过程，样本不是可以任意搬进任何故事的碎片。

对于未来复查，处理版本和操作历史也应随样本资料保存。某次背景改正或新的矿物库可能改变解释，旧结果若保留完整链条，就能重新计算。任务结束不必意味着资料失去发生能力，但这种延续依赖当时已经建立的可追踪记录。

在有限任务寿命内，保存一次操作失败的条件同样有用。它能告诉后来者某种材料或测量方式为何不可行，避免把未取得资料误写成目标不存在。

本章参考文献

[46-1] NASA. NASA’s Mars Curiosity Rover Marks First Martian Year with Mission Successes. 2014. https://www.nasa.gov/news-release/nasas-mars-curiosity-rover-marks-first-martian-year-with-mission-successes/

[46-2] Grotzinger, J. P., et al. A Habitable Fluvio-Lacustrine Environment at Yellowknife Bay, Gale Crater, Mars. Science, 2014. NASA 原论文存档：https://ntrs.nasa.gov/api/citations/20150008374/downloads/20150008374.pdf

[46-3] NASA. Sedimentation and Erosion in Gale Crater, Mars. https://science.nasa.gov/resource/sedimentation-and-erosion-in-gale-crater-mars/

[46-4] NASA. NASA’s Curiosity Rover Finds Patches of Rock Record Erased, Revealing Clues. 2021. https://www.nasa.gov/solar-system/nasas-curiosity-rover-finds-patches-of-rock-record-erased-revealing-clues/

第 47 章 行星、矮行星与类别边界：定义会怎样回写探索

47.1 行星定义历史如何变化

“太阳系有多少颗行星”看似一道只需要数数的问题，实际先要求回答哪些对象应当进入被数的集合。发现新天体以后，旧集合可以扩张，也可以重划边界。若不说明归类规则，数字的变化就容易被误解为自然界突然增加或减少了某个成员。

行星一词原本与天空中的游走现象相关。这样的识别依赖观察者所在位置以及星空背景中的相对运动。后来日心体系、引力理论和新的探测能力，使行星分类不能继续只依赖可见运动。一个名称在长时段内保留下来，不代表其判断条件一直不变。思想史需要追踪词语下方的对象构成，而不能只追踪词语本身。

冥王星与外太阳系新发现使这一点变得尖锐。若传统九大行星只是列举，新的相似对象出现后就必须回答是否继续加入；若规定某个大小阈值，又需要说明阈值为何具有科学意义；若以轨道动力学为主，则同样大小的天体在不同环境中可能有不同类别。不同方案不是关于同一条现成自然分界的简单测量误差，它们还涉及分类要突出哪种关系。

分类目的并不意味着分类可以任意。动力学分类必须能够表达系统组织，地球物理分类必须能够表达内部结构或表面过程，历史名称也必须保证沟通可行。一个定义若既不能形成清晰对象集合，也不能支持有效比较，就难以服务研究。科学分类同时受自然差异与研究用途制约。

SDE 分析在这里应特别避免把 S 视为孤立固有属性。所谓“行星身份”并不只由对象的质量单独决定，它在特定规则下与形状、轨道位置和周围对象关系共同成立。E 不是外加背景，而是类别判断不可抽去的相互规定。移走动力学条件，得到的不是相同定义的简化版，而是另一个分类系统。

分类可以在逻辑上区分几个层次。一个规则首先要说明适用对象域，例如是否包括卫星、是否只讨论太阳系；其次选择判据，例如形状与动力学；最后规定如何处理测量不足的边界对象。如果不先限定对象域，即使判据本身清楚，也可能在迁移到另一个领域时产生不相容结果。

这是“行星数量”容易引起争论的一个原因。不同讨论者可能并不是对同一对象作出相反测量，而是使用了不同对象域。有的人把具有一定内部结构的卫星纳入类行星比较，有的

人只统计直接绕太阳运行的主要成员。若争论不澄清这一点，双方会不断用不同问题的答案互相否定。

一个名称能否承担多个分类目的，是可以讨论的问题。日常语言允许同一词具有较宽范围，专业目录则通常需要操作精确。若同一个词用于不同任务，最好明确限定语；若使用多个词，则要提供对应关系。名称并非越少越科学，关键在于是否有助于保留真实差异、避免错误推断。

行星定义也不能仅以方便学生记忆多少个名字为最高标准。教育需要简洁，但研究分类还必须能容纳新对象。若一个定义的主要功能是把数量固定在熟悉范围，新发现会不断迫使其例外化。反过来，无限制扩张一个名称若不能增加解释力，也未必服务研究。数量大小不是单独判据。

47.2 2006 年方案不是一稿定案

2006 年 8 月 16 日的 IAU 草案与 8 月 24 日通过的定义必须分开阅读。草案主要围绕自引力使对象近乎球形、绕恒星运行且不属于恒星或卫星等条件展开，并讨论由此扩大的行星集合。[47-1]它与最终文本的差异，使我们能够看到共同体怎样在争议中选择分类重点。

正式决议对太阳系行星采用绕太阳运行、足够自引力形成近乎球形，以及清空轨道邻域三方面条件；矮行星则具有相应形状条件而没有满足轨道邻域条件，且不是卫星。[47-2]这里的“清空”不应被理解为轨道附近绝无任何小天体，否则会把动力学主导关系误读成空间必须彻底空无。

草案与定稿的比较至少揭示两个重要选择。一是定义适用域的表达不同，最终决议明确针对太阳系；二是动力学关系获得更强地位。由此，球形不再足以决定正式行星身份。一个对象可以拥有复杂内部与表面，同时在该分类下被列为矮行星。名称层级不能直接换算成科学价值层级。

美国天文学会行星科学分部在当年说明中记录争议及教学影响，并指出并非所有行星科学家都同意决定或其程序。[47-3]这一材料使我们不能把最终表决写成所有研究问题都获得解决。表决可以建立组织采用的规范，却不能取消关于分类目的、边界判据和适用范围的讨论。

同样，不能以存在争议为由宣称定义完全没有科学内容。形状与动力学条件都与物理过程相关，争议在于哪类过程应作为最高分类原则，以及如何将其操作化。若把辩论压缩为“科学家喜欢或不喜欢冥王星”，会遮蔽真正值得研究的判断差异。

这个案例也限制了历史因果的轻率解释。外太阳系新对象推动分类问题重新进入公共议程，有清楚的材料支持；但具体条款怎样形成，需要草案、会议记录和正式文件分别说明。不能只知道最终结果，就为每位参与者虚构同一种动机。统一结论可以来自不同理由，并不要求参与者拥有完全同一的理论立场。

正式条款中的“近乎球形”也需要物理解释。自引力、材料强度、转动与历史状态共同影响形状，不能简单规定任何超过某个固定直径的对象都必然满足条件。草案说明中提及的大致尺度不应被误读成最终规则唯一的硬阈值。[47-1]对边界对象仍需具体资料，而不是拿一把直尺完成所有分类。

动力学主导也不是一个脱离系统的固有数。轨道区域大小、周围质量分布和演化时间会影响一个天体如何组织邻域。将地球放到一个极远的假设轨道上，再问它是否保留相同动力学身份，是检验分类性质的思想实验；但它并不自动反驳实际太阳系中的分类，因为该规则本来就包含关系条件。真正的争论在于，关系条件是否应该定义这个名称。

这种反事实具有价值，前提是双方承认它测试什么。如果主张行星必须只由自身物理属性决定，那么位置改变不应改变身份；如果主张行星表达系统内的动力学角色，那么位置改变带来身份变化并非逻辑矛盾。两种定义可以各自一致，却服务不同目的。需要比较它们解释与组织研究的能力，而不只是寻找一句听起来反常的话。

决议通过还包含程序效力与知识效力的差别。一个组织可以决定在自己的正式使用中采用某个术语，这提供协调标准；该术语是否最适合所有行星科学问题，仍可继续研究。将组织规范尊重为规范，与将它视为不可讨论的自然定理，是不同态度。专业合作需要前者，不必要求后者。

47.3 物理对象与共同体分类

表决没有改变冥王星的轨道，这句话正确，但若据此认为定义只是无关紧要的标签，又遗漏了科学世界中的实际变化。目录集合、比较对象、教学结构和研究问题都可能因分类改变。需要区分物理过程没有被投票改变，与公共研究对象的关系被规范重组，两者能够同时成立。

一个思想实验可以说明这种关系。假设我们以内部地质活动为主要问题，某些大型卫星与冥王星可能比它们与小型不规则天体更适合比较；若研究轨道组织，是否绕主行星运行、是否在太阳轨道上占据主导地位又变得重要。同一对象进入不同比较集合，是因为问题不同，并不意味着其中一个集合必然伪造自然。

真正的困难在于是否把某一分类目的隐藏起来，冒充唯一可能的对象本体。动力学定义能够有效组织某类问题，却不应该因此禁止地质学依据自身需要建立另一个比较系统。反过

来，地质复杂性也不能直接推翻动力学分类，因为复杂表面并非后者唯一关心的特征。争论必须比较双方实际判据，而不是让一方用自己的问题给另一方判零分。

本书所谓科学世界构成，正包括这种对象单位的形成。天体自身的物理历史不会由命名瞬间改写，但“行星世界”作为有边界、有典型对象、有比较规则的研究整体确实可以变化。这个变化有实在后果，又不等于任意创造物质。把它只归为语言问题与把它夸张为投票改变宇宙，都会错过中间这一层。

六路径在分类问题中体现为两个入口。从正式名称 S 进入，研究者可以追问它所依赖的 E，再检查 D 是否真的支持界限；从新发现的差异 D 进入，则可问哪些分类条件需要改变，最后形成新的 S。前者防止规则漂浮，后者防止旧规则压制新差异。它们并不承诺每次都达成同一分类，却能清楚说明分歧来自哪里。

还应区分边界模糊与定义失效。某个对象的质量、形状或轨道关系尚不充分，并不立即证明定义无用；它可能只是缺少资料。但如果定义在一类明确对象上产生持续不适用的结果，就需要修订。不能用“更多数据以后就会解决”无限推迟真正的概念问题，也不能以一处测量不确定为由否定整个分类。

一个分类的价值可以通过它允许怎样推理来检验。若把两种对象放在同类，研究者通常期待某些性质可以比较、某些机制具有共同性。但同类并不保证全部性质相同。八颗正式行星中存在显著组成与结构差别，名称不能代替详细物理模型；矮行星内部也不必同质。粗分类应为进一步分化提供入口，而不是终止分化。

这也解释为什么层级体系和多维标签有时需要并存。单一树状分类要求每个对象放进一个主要分支，多维描述则允许同时记录轨道角色、组成、形状和活动。对数据库而言，保留多维属性往往使未来重分类更容易。若只保存最终名称而丢失判据，当定义更新时就难以重建对象归属。

在 SDE 语言中，分类 S 应该能够返回其差异 D 与特征纠缠 E。若对象名称已经封装成不容追问的本质词，研究者就容易从名称推出未经支持的属性。例如，因为某对象被称为矮行星，就断言它没有复杂地质；或因为是正式行星，就假定具有某种大气。这些推断不由定义保证，需要独立证据。

评价分类争论时，也要给不同方案公平的失败条件。动力学方案若无法清楚操作其关键关系，需要补充指标或范围；地球物理方案若让重要结构差异被同一名称遮蔽，也需要更细分层。任何方案都不应先把自己的目标设为唯一目标，再以另一方案没有完成该目标为由宣布胜利。这是分类研究能够前进的基本对话条件。

47.4 新类别怎样塑造研究

定义可能回写研究，但回写必须有具体证据。最容易核实的是教学与目录：采用正式定义的材料重新组织对象集合，学生学习的太阳系结构也随之改变。更强的命题，例如某次任务因分类改变才获批准，不能只靠时间先后推断。

新视野号提供一个有力的约束例子。它于 2006 年 1 月发射，早于同年 8 月的分类决议，2015 年飞掠冥王星。[47-4]因此不能把该任务的启动归因于矮行星定义。任务继续研究冥王星，也说明科学价值并未随着正式类别变化而消失。这个反例能阻止发生学解释把所有后续活动都算成规范的回写成果。

更细的影响则需要查具体研究文件。某种新类别可能改变比较样本、会议分组或教材章节，也可能只是提供新的总称而没有改变观测计划。前者具有较明确的操作效应，后者主要是命名协调。两者都可以研究，但应给予不同强度的因果判断。没有发生操作改变时，应承认回写证据不足。

定义也可以帮助提出此前不突出的问题。例如，当对象被放进较大的外太阳系族群，研究者可以更多比较轨道分布、组成与形成环境。不过这种研究兴趣也可能早于正式定义。规范可能整理已有研究，而非单向创造研究。双向关系需要逐项追踪，不能用一句“分类塑造科学”替代历史证据。

在系外行星领域，2006 年太阳系决议不能原封不动扩展为统一法律。恒星、棕矮星、自由漂浮天体及探测限制带来新的边界问题，需要另行说明使用哪个工作定义及版本。若只把“绕太阳”换成“绕任何恒星”，就以为全部哲学与技术困难已经消失，便忽略了分类适用域自身的发生。

教学中的变化可以采取比“九变八”更有解释力的方式。让学生先按天空运动分组，再按绕行关系分组，再按形状与内部过程分组，他们便能看见同一批天体如何形成不同集合。最终学习正式定义时，理解的将不只是一条需要记忆的命令，而是某种公共分类为什么采用这些判据。

研究数据库也可以采用类似透明方式。名称、采用版本、支撑参数和不确定状态分开保存，可以防止未来使用者把暂定身份当作永恒结论。若某个边界对象后来获得新测量，系统应能说明是物理参数修订导致归类改变，还是定义版本改变导致归类改变。这两种更新的知识意义不同。

跨文化的公共传播则需要注意名称引起的价值联想。“降级”容易让人以为研究价值或对象复杂性下降，而正式分类讨论的并不是荣誉等级。更准确的说法是按某一规则改变归

属，同时继续研究其物理性质。名称若带来情感反应，可以理解，但不能用情感反应替代判据讨论。

这场争论留下的具体任务，是让名称与用途之间的关系保持可查。编制一份动力学样本时，应说明轨道判据；比较地质活动时，应说明为何纳入某些卫星或矮行星；讨论系外伴体时，应注明所用工作定义。冥王星没有因为名称改变而退出研究，真正需要改变的是一种习惯：不能再用背熟一张名单，代替说明这张名单为什么成立。

“清空邻域”的日常措辞还需要避免制造错误图景。一个动力学占优势的行星周围仍可能存在小天体，某些对象还可以处于稳定共振关系。将规则理解为轨道上所有其他物质都被彻底清除，就会用不属于原意的绝对条件反驳它。反过来，若用“占优势”替代原文，也要进一步说明怎样在具体研究中判断优势，而不能靠直觉完成。

这意味着正式定义与实际分类指标之间可能还有一层工作。定义给出原则，研究者需要把质量、轨道与演化条件转化为可操作检验。不同指标若在边界对象上给出不同结果，应该比较其假设与用途。测量越多，不一定自动消除规范问题，因为争议可能恰恰在于哪个指标应该优先。

分类稳定性可以通过新对象检验：一个规则面对从未用于制定它的对象时，能否给出可说明的结果，还是需要不断增添例外？这样的检验比只用现有八颗行星证明定义合适更严格。若定义只是在事后重述熟悉名单，它的预测与组织能力就有限。

若一个新天体仅因资料不足而暂时无法归类，目录应保留未定状态。强行选择名称虽然使表格整齐，却会让之后的使用者忽略缺失参数。相反，明确列出尚缺的形状或动力学证据，可以把分类争论转化为观测任务。这样的暂缓是一种精确判断，不是研究者没有能力回答。

分类的操作记录因此应与名称一起保存，让使用者能够区别测量改变与规范改变。

本章参考文献

[47-1] IAU. The IAU Draft Definition of “Planet” and “Plutons”. 16 August 2006.

https://www.iau.org/IAU/Iau/News/PR2006/iau-draft-planet-pluton-definition.aspx

[47-2] IAU. IAU 2006 General Assembly: Result of the IAU Resolution Votes. 24 August 2006.

https://www.iau.org/IAU/Iau/News/PR2006/iau-2006-general-assembly-resolution-votes.aspx

[47-3] AAS Division for Planetary Sciences. What Is a Planet? 11 October 2006. 含正式决议转录。

https://dps.aas.org/education/what_is_a_planet/

[47-4] NASA. New Horizons. https://science.nasa.gov/mission/new-horizons/

第 48 章 第一批系外行星：径向速度怎样支持看不见的伴星

48.1 光谱线周期变化的意义

寻找另一颗恒星的行星，不一定从拍到一个暗弱光点开始。恒星与伴体围绕共同质心运动，恒星光谱因此可能出现周期性的多普勒变化。研究者测量的是恒星辐射，不是行星表面的影像，却能由运动关系限制伴体存在。这种间接性并不使证据次等，关键在于关系是否足够稳定、是否能够排除其他来源。

细小谱线位移的解释面临至少三类竞争。仪器波长尺度漂移可以移动谱线，恒星表面活动可以改变谱线形状与重心，轨道运动也可以产生周期速度。三者在某些资料上能够相似，不能仅因拟合出一条正弦曲线就决定来源。需要比较不同谱线、活动指标、测量时段和独立仪器，使候选解释面对不同约束。

质量与倾角的退化是径向速度法的基本限制。以现代二体近似式说明，速度半振幅 K 与伴体质量、轨道周期、偏心率、恒星质量以及轨道倾角共同相关。在伴体质量远小于恒星质量时，测量通常约束 m sin i，而不是单独的真实质量 m。视线接近轨道轴时，即使伴体较重，径向运动分量也可能较小。

这个关系不是一句技术附注，而是“发现了什么”的组成部分。若只说测得某颗行星质量，却省略倾角条件，就把联合约束误写成独立属性。反之，知道质量存在退化也不能直接否定行星解释，因为可以利用额外观测和合理模型缩小允许范围。科学判断通常在有限信息下分层增强，不是在完全直接与完全未知之间二选一。

E 在此体现为相互规定的几何与动力学特征。速度振幅脱离倾角不能唯一规定质量，倾角信息也只有结合轨道与恒星性质才转化为质量限制。恒星活动并非单独放在最后的“噪声项”，它可能与自转周期、谱线形成和测量方式相关。将这些关系明确化，才能说明哪一个新观测真正增加了辨别能力。

把这一关系写得更明确，现代近似下径向速度半振幅可表达为

$$K=(2\pi G/P)^{(1/3)}\cdot m\sin i/(M+m)^{(2/3)}\cdot 1/\sqrt{(1-e^2)}。$$

这里 M 是恒星质量，m 是伴体质量，P 是周期，e 是偏心率，i 是轨道倾角。这是用于解释观测依赖的现代标准关系，不是对 1995 年论文推导顺序的复原。当 m 远小于 M 时，分母主要受恒星质量决定；但即使恒星质量完全已知，m 与 sin i 仍不能仅由 K 分开。

可以用一个明确的假设例子显示其后果。假设数据限制 m sin i 为半个木星质量。若轨道几乎侧对观察者，sin i 接近一，真实质量接近这个下限；若 sin i 为二分之一，真实质量就接近一个木星质量。两组参数可以给出相同径向振幅，却不是同一真实质量。因此“最小质量”不是研究者出于谨慎随意加上的修辞，而是观测几何留下的数学退化。

倾角先验可以评估不同质量的相对可能，却不能替代特定系统的倾角测量。随机取向假设是关于总体的条件，不应被说成已经知道这颗伴体的姿态。如果后来获得凌日或天体测量轨道，才有可能在新的数据条件下进一步收缩质量范围。引入先验与增加观测属于不同证据动作。

测量谱线的方式也会影响“速度”的含义。真实恒星光谱包含大量不同深度和形成条件的谱线，观测可以通过综合比较提高精度。若所有谱线近似整体移动，轨道运动解释较自然；若某些线的形状或深度随相位变化，则需要检查恒星表面与大气。把整段光谱压缩为一个速度值能提高效率，却可能丢掉辨认替代机制所需的信息，因而原谱保存十分重要。

还有一种来自时间采样的退化。若总在某些固定时刻观测，真实周期与采样周期可能组合出另一个显著周期。增加同类观测次数未必完全解决问题；改变观测时刻、增加不同地点资料或扩大季节覆盖，才可能拆开候选周期。这里需要的是差异设计，而不只是数量增加。

48.2 51 Pegasi 的关键异常

梅耶与奎洛兹 1995 年的论文报告了 51 Pegasi 的周期径向速度变化，并据此提出一个靠近太阳型恒星的木星质量量级伴体。[48-1]其震动作用并不只来自“太阳系外也有行星”。一个巨行星处在非常短周期轨道上，使以太阳系布局为直觉模板的形成想象面临新的压力。

必须区分违反直觉与违反物理定律。热木星并不因为靠近恒星就违反引力规律；它首先挑战某些关于形成位置和后续演化的预期。如果发现可靠，理论便要解释巨行星怎样到达这样的轨道，或者是否存在此前低估的形成渠道。林等人的轨道迁移研究提供了这一回应的一种方向。[48-2]

但“需要迁移解释”与“已经证明唯一迁移历史”不是同一句话。一个现今轨道可以由不同历史产生，模型还需要解释停止机制、系统结构及其他统计特征。SDE 的发生学要求追问过去过程，却不能因为重视发生便虚构唯一过程。生成解释的价值，在于提出能被资料区分的路径，而不是给现状配上一个流畅起源故事。

发现对搜索本身也形成回写。若研究者只期待类似木星的长周期轨道，就需要漫长监测才能看到完整变化；短周期伴体则可能在较短时间里积累多次重复。由此，什么值得搜、用多密的时间采样、如何安排观测，都会受到新对象影响。新世界的构成同时改变了研究对象和发现对象的能力。

不过，短周期巨行星易于被发现，也可能让人误以为这种系统最典型。速度振幅较大、周期较短的对象通常更容易进入早期资料集。第一个令人震惊的样本，并不自然代表总体。若在思想史中把“先发现”直接写成“最普遍”，便会把仪器选择转成宇宙事实。

51 Pegasi 之所以适合 SDE 分析，正在于它把三种变化联在一起：恒星谱线的细小移动成为新显露，观测时序与排错方式形成新的差异运行，行星形成与搜索预期被重新组织。三者不是同时自动发生，也不是一条固定历史路线。它们需要通过论文、后续观测和方法变化分别建立联系。

近轨巨行星为什么比长周期木星类似物容易首先进入径向速度样本，可以直接从上式观察。保持质量、恒星质量和偏心率不变，周期缩短会增大 K；同时，在同样几年观测期内，短周期对象能够反复完成轨道。振幅与重复次数的双重优势，使它们更容易被确认。这是测量过程的偏好，不是形成过程必然偏好的证明。

不过，不能把这一优势写成任何短周期都容易识别。若周期与恒星自转或仪器运行节律相近，替代解释可能更加棘手；若观测间隔太稀，还可能错过变化。探测能力是振幅、周期、噪声结构和采样共同决定的，不是单一参数排序。

理论面对异常时有两种不够充分的反应。第一种因为太阳系里没有这种布局，就先宣布资料不可能正确；第二种一旦发现可能可靠，就宣布所有原有形成理论已被推翻。前者把既有样本当成宇宙边界，后者把一个新增约束当成对所有机制的全面否决。更有效的做法是指出哪些环节受到挑战：形成地点、迁移过程、轨道稳定，还是观测偏好。

迁移解释也必须面对停止问题。若一种机制能把巨行星向内移动，它为何没有使所有对象最终落入恒星？林等人的论文讨论了可能使迁移停止的物理条件。[48-2]思想史应保留这种问题扩展：新解释解决一个矛盾，同时产生新的可检验要求。提出迁移不是故事结束，而是开始把轨道分布与盘、潮汐及其他过程联系起来。

若以后样本显示各种不同轨道，就需要允许多条形成与演化路径。单一案例启发的机制不必垄断全部系统。多样性在这里首先是自然样本给理论的约束，研究者不能为了保持模型整齐而把不同系统的历史全部压成一个版本。

48.3 从“伴体候选”到行星

51 Pegasi 的行星解释并未因为发表就免于挑战。谱线形状是否发生变化，是检验恒星脉动等解释的重要方向。1997 年的质疑与 1998 年的后续高精度研究，使这项发现经历了针对关键替代的检查。后续论文报告未发现所需的系统性线形变化，从而增强了行星解释。[48-3][48-4]

这一过程的价值不是证明怀疑者“阻碍科学”，而是把可能混淆的机制具体化。若恒星表面运动导致谱线形状变化而非整体移动，就会在更细的谱线结构中留下线索；如果不同谱线保持相容位移、而相关形态变化不足，轨道解释获得优势。双方争论围绕可检测差异，才使后续观测能够改变判断。

独立重复也需要细分。另一台光谱仪测得相近周期，有助于排除某些仪器问题；但恒星活动属于共同天体来源，两台仪器都可能看到它。因此“独立仪器也看见了”不能单独排除所有非行星解释。需要使独立性与假说相匹配：仪器替代针对仪器漂移，线形检查针对表面活动，其他动力学资料则针对伴体质量与轨道几何。

更长时间的稳定性可以增加约束，但不是简单认为时间越长就必然越真。长期活动也可能有周期，采样窗口还会制造混叠。研究者要比较相位是否持续、不同季节结果是否一致、周期是否接近观测节律，并检验缺测如何影响拟合。统计上的显著峰值必须回到实际观测安排中理解。

若有多颗伴体，问题又会复杂化。一个偏心轨道或几个较简单轨道可能在有限资料上相近，加入更多参数总能改善部分拟合，却可能只是追随噪声。模型选择需要考虑参数自由度、预测能力和物理可行性。最小残差不自动等于最佳天体解释，尤其在数据少而模型复杂时。

这说明对象确认是一项逐步收缩可能性的工作。“候选”不是对研究价值的贬低，而是当前证据状态的精确名称。把候选升级为确认，需要说明哪项竞争已被显著限制；若后来出现更好的活动解释，也应允许降级。科学世界的稳定性来自这种可修正机制，而非永久冻结第一次命名。

从 S 入手的路径容易问怎样完善这颗行星的参数，从 E 入手则会重新检查恒星与仪器可能生成什么信号。前者提高研究效率，后者避免身份封装失效后继续累积细节。成熟的判断需要按风险轮换：当信号稳定且替代已受约束，可以使用封装结论；当新异常出现，就应重新打开封装。

谱线质疑提供了一项值得细读的证据程序。若恒星表面某些区域向我们运动、另一些区域背离，且其亮度权重随时间变化，总谱线中心可能偏移，但线的两翼未必对称改变。所谓线形或线平分线检查，便试图抓住这种差别。轨道造成的整体多普勒平移，在理想情况下则更接近把整个轮廓一起搬动。实际恒星与仪器更复杂，所以这一判别需要足够分辨率和信噪比。

1998 年的两篇研究不只是再画一次相同速度曲线，而把注意力放在替代解释要求的谱线变化上。[48-3][48-4]这一点十分重要：最有效的复核往往不是重复最初最显眼的证据，而是

寻找竞争假说独有的附带后果。若附带后果没有出现，某种替代的可信度下降；但下降幅度仍取决于观测到底有能力看到多大的变化。

没有观察到线形变化，也不等于数学上证明所有活动机制都绝无可能。它意味着在相应精度与时段内，没有找到该竞争解释所要求的支持。科学论证常以这种有边界的排除前进。若把每一项排错写成绝对消灭全部替代，叙述表面更强，实际反而更脆弱。

模型比较还要注意资料的独立使用。若先用所有数据选择一个周期，再用同一数据宣称这个周期被独立预测，就重复计算了支持。可以用后续时段检验相位和周期，也可以预留部分数据检验预测表现。对于长时段天文观测，后续自然产生的数据尤其重要，因为它们不能被最初分析者在事后挑选来拟合。

一颗候选被接受以后，新的观测仍可能改变其参数，甚至揭示其他伴体。参数更新与身份撤销应分开记录。若后来的质量估计改变，但仍支持行星质量范围，就不必将其叙述为发现失败；若替代机制能够解释全部信号，才需要重新评价身份。档案中的版本信息，使这种差别能够被追踪，而不是由公众记忆中一句“发现过”覆盖。

恒星活动与行星信号还可能叠加，而非只能二选一。一个真实伴体存在时，恒星仍然会有自转、斑点和其他变化。若把任何活动相关性都当成行星不存在的充分证据，就会走向另一个极端。分析应比较包含轨道与活动的联合模型，检查加入活动描述后，伴体周期与振幅是否仍然得到支持。

联合模型也面临自由度问题。如果活动模型足够灵活，可能吸收真实轨道信号；如果过于简单，又会把活动残差归为行星。最好的模型并不必然具有最多可调参数，而应在独立时段和多种指标上保持合理预测。恒星活动指标的用途，是增加与物理来源有关的信息，而不只是为速度曲线增加拟合空间。

小振幅搜索使这一矛盾更尖锐。测量精度提高后，过去可以忽略的恒星变化会进入目标尺度。技术进步因此不只是让同一问题更容易解决，有时反而把研究带入新的噪声结构。仪器误差下降并不保证天体物理误差同步下降，二者需要不同办法。新的行星世界会要求新的恒星世界作为其测量条件。

一个有用的观测设计，是同时追踪不同物理敏感度的谱线或活动指标，再比较它们与候选周期的关系。若某种变化随活动强弱显著改变，而候选轨道信号长期保持，可以增加分离机会；若二者紧密同相，则要更加谨慎。不存在脱离具体恒星和资料质量的通用一招，判断必须在多项约束中形成。

48.4 历史归属与两种首例

“第一颗系外行星”若不说明范围，容易把不同历史突破混为一谈。沃尔什钱与弗雷尔1992 年报告围绕毫秒脉冲星 PSR1257+12 的行星系统，依据精密计时中的周期变化；1995 年的 51 Pegasi 发现则指向太阳型恒星附近的行星。[48-5]两者都重要，但所用信号、宿主性质和形成问题不同。

脉冲星计时把天体当作高稳定时钟，其到达时刻变化可以揭示伴体引起的运动。光学径向速度则依据恒星谱线变化。两种方法共享引力导致反射运动的逻辑，却使用不同载体和误差结构。若只用“看见恒星晃动”概括两者，就会失去时间传播与光谱形成的差别。

脉冲星行星还提出特别的发生问题：经历致密天体形成与演化的系统，为何拥有行星质量伴体？这与在普通恒星原行星盘中讨论巨行星形成并不完全相同。把所有行星都塞进一个太阳系起源模板，会忽略宿主历史；反过来，也不能因为宿主特殊，就取消这些对象对行星存在史的贡献。

“首例”是一种研究归属，它依赖确认标准和对象范围。最早提出候选、最早获得后来被接受的证据、最早在特定宿主类别中发现，并不总是同一事件。严谨的思想史应该把这些范围写出，而不是寻找一个可以覆盖所有维度的英雄名次。

在脉冲星系统中，行星间相互摄动还能够提供超出简单多个独立周期的检验。若多个伴体只是为拟合残差而添加的正弦项，它们未必满足相互作用的动力学关系；若后续计时出现模型预言的摄动，则为真实系统增加支持。Rasio 等人 1992 年提出的观测检验，正把注意力转向这种可区分后果。[48-6]

这一思路说明，精确拟合不是最后目标。一个模型若真实地描述多个质量体，就应在它们的相互影响中产生额外预言。这样的预言往往比继续改善原有拟合更有辨别力，因为它不完全由最初用于设定模型的周期项决定。发生学所说的系统整体，在这里必须兑现为相互作用的可检验结果。

两种首例也分别改变了“适合寻找行星的宿主”这一判断。太阳型恒星提供与太阳系较容易比较的背景，脉冲星则说明研究不能只在最像太阳的对象中寻找。扩展目标范围有潜在收益，也会带来新的形成问题与误差结构。搜索多样性应由方法能力支持，不能仅以想象中的可能性替代可观测性。

历史归属还必须尊重失败候选的作用。曾被提出而未获维持的周期解释，可以帮助研究者发现计时改正、观测节律或活动模型的薄弱环节。但本章没有逐一核验全部早期候选，因此不以未经核实的撤回故事制造戏剧性。可以肯定的是，确认标准需要容纳撤销与修正，否则“首例”就会退化为最早宣布者的永久产权。

对后续研究而言，保留最初论文的参数与解释范围很重要。后来的精确质量、更多伴体或新类别不能全部归给首次发现。最初证据解决了什么，后续又增加了什么，应分别给予历史位置。这样既不缩小突破，也不会让一个英雄时刻吞没长期共同劳动。

两个首例共同改变了天文学的对象分布。行星不再仅以太阳系内可逐个列举的天体出现，还可以作为遥远系统中的动力学成员，通过其对恒星或脉冲星的影响被确认。这个变化不是简单把已有名单延长，而是建立新的测量、排错和比较制度，使原本看不见的伴体进入公共研究。

与此同时，间接探测留下的问题促成方法互补。径向速度限制质量组合，凌日可以限制半径比与倾角，直接成像突出另一类距离与亮度条件。下一项最值得寻找的证据，取决于当前候选究竟缺什么：若不清楚信号来自恒星还是伴体，应优先排错；若伴体身份已稳健而质量仍退化，就要寻找几何约束。把这两类任务分开，才能使更多观测真正增加知识，而不只是延长同一条速度曲线。

伴体身份还包含质量与形成概念之间的张力。径向速度支持一个行星质量量级的伴体，并不自动揭示它以哪一种机制形成。若分类主要按质量范围，研究者应说明质量不确定性；若进一步按形成历史区分，则需要超出当前轨道的证据。不能在一篇发现报告中不加说明地来回更换判据。

这类张力在极端系统中特别明显。一个伴体的动力学量可能相当清楚，形成路径却仍有多种候选。将其称作行星可以是工作分类，但不应把名称本身当作某一形成模型已经获证的理由。否则研究就会从分类推导起源，再用推导出的起源证明分类，形成隐蔽循环。

另外，恒星质量估计也会传播到伴体质量。上式显示恒星质量不是无关输入；若宿主参数更新，最小伴体质量可能随之调整。不同论文参数表不一致时，应先检查恒星模型和资料版本，再讨论是否真的存在观测冲突。数值差异有时来自共同输入更新，而非某组观测失效。

这些具体依赖为后续复查提供了清单式的方向，但研究本身不能只照清单机械运行。对于活跃宿主，应重视活动解释；对于几何不确定但运动稳定的系统，应重视倾角；对于多伴体候选，应重视相互摄动和长期稳定性。问题不同，最有价值的新差异也不同。

发现史最后仍要回到预测：候选周期能否在新时段保持，活动解释能否预言线形变化，伴体模型能否与其他观测相容？当这些问题逐项得到回答，一个看不见的伴体便获得越来越具体的公共身份。下一步研究应追随仍未分开的可能性，而不是继续为已经稳健的部分反复增加相同证据。

特别是最小质量与真实质量，必须在参数表和正文中使用相同口径，不能在摘要中省略关键几何条件。

本章参考文献

[48-1] Mayor, M., Queloz, D. A Jupiter-mass Companion to a Solar-type Star. Nature, 378, 355–359, 1995. https://wwelsh.sdsu.edu/CLASSES/ASTR510/PAPERS/Mayor-Queloz_51Peg.pdf

[48-2] Lin, D. N. C., Bodenheimer, P., Richardson, D. C. Orbital Migration of the Planetary Companion of 51 Pegasi to Its Present Location. Nature, 380, 606–607, 1996. https://www.nature.com/articles/380606a0

[48-3] Gray, D. F. A Planetary Companion for 51 Pegasi Implied by Absence of Pulsations in the Stellar Spectra. Nature, 391, 153–154, 1998. https://www.nature.com/articles/34365

[48-4] Hatzes, A. P., Cochran, W. D., Bakker, E. J. Further Evidence for the Planet around 51 Pegasi. Nature, 391, 154–156, 1998. https://www.nature.com/articles/34369

[48-5] Wolszczan, A., Frail, D. A. A Planetary System around the Millisecond Pulsar PSR1257+12. Nature, 355, 145–147, 1992. https://www.nature.com/articles/355145a0

[48-6] Rasio, F. A., Nicholson, P. D., Shapiro, S. L., Teukolsky, S. A. An Observational Test for the Existence of a Planetary System Orbiting PSR1257+12. Nature, 355, 325–326, 1992. https://www.nature.com/articles/355325a0

第 49 章 凌日、成像与系外行星人口统计

49.1 一条光变曲线能推出哪些量

行星经过恒星与观察者之间时，可以遮去一小部分星光。研究者记录的不是一个黑色圆盘逐像素走过恒星，而常常是总亮度随时间的细微下降。这样的曲线把几何、轨道和恒星表面亮度结构联系起来。它能够提供丰富信息，但每种信息都依赖明确条件。

在均匀亮度、忽略混合光等简化条件下，凌日深度大致等于行星与恒星半径比的平方。实际恒星存在临边昏暗，轨道可以偏心，行星未必从圆面中心穿过，其他天体的光也可能稀释下降深度。现代简式适合解释主要尺度，却不能当作所有资料的精确处理。尤其需要注意：光变首先限制半径比，绝对行星半径还依赖恒星半径。

重复下降之间的时间限制轨道周期，进入与退出的形状帮助约束路径几何和半径比，持续时间则与轨道速度和恒星尺度相关。若只观测到一次下降，周期通常不能像多次凌日那样确定。若资料缺口恰好遮住关键阶段，允许几何范围也会增大。参数表中的一个数值，不应掩盖形成它的资料覆盖。

HD209458 的凌日探测展示了方法互补的力量。此前径向速度已经指向近轨伴体，测光又记录到凌日，使轨道倾角与行星尺度获得额外限制。[49-1]不是两种方法各自测出完全相同的信息，而是它们的退化方向不同，联合之后能够得到单一路径难以确定的性质。

这也是 E 的具体含义。若恒星半径重新标定，行星半径可以随之改变；若混合光比例被修正，凌日深度对半径比的解释也会改变。行星属性并非与恒星研究无关的独立标签。对象虽然在空间上分离，其测量属性却在共同光路和模型中相互规定。SDE 的整体视角只有把这种依赖写清，才比笼统说“万物有联系”更有解释力。

一个数值例子可以揭示恒星参数为何重要。设某次理想化凌日下降百分之一，忽略临边昏暗与混合光，则半径比约为十分之一。若后来发现恒星实际半径比原估计大百分之二十，行星绝对半径也要相应上调，而测光曲线本身无需改变。对单颗对象，这只是参数修订；对大量对象，如果恒星半径存在共同偏差，整个行星尺寸分布都会移动。

混合光具有另一种作用。假设恒星旁还有一颗未分辨恒星，提供额外恒定光通量，行星遮去的目标星光在总光中所占比例就被稀释。若不知额外光存在，研究者可能低估半径比。高分辨率成像在这里不是为了直接拍到行星，而是为了确认照明与混合条件。不同观测方法因此可以在一个看似纯测光的问题上承担关键工作。

凌日持续时间也不能独自充当恒星密度的无条件测量。其解释与轨道偏心率、凌日位置等相关。若把圆轨道当作固定前提，而真实轨道明显偏心，某些推得量可能偏移。这并不要

求每次分析都让所有参数完全自由，而要求报告所采用的条件，并在重要结论上检查放松条件后是否仍然成立。

49.2 排除食双星等伪阳性

亮度下降不唯一指向行星。恒星伴星的掩食、背景食双星的混合光，以及仪器或数据处理异常，都可能产生类似信号。发现流程因此需要从“存在下降”进入“哪些机制能产生这种下降”。如果把算法输出的候选直接列入确认行星名单，就跳过了最关键的身份判断。

不同排错手段针对不同替代。比较奇偶次下降可帮助识别某些食双星结构，搜索次食可以增加约束，多色资料能够检查深度是否具有不符合简单行星遮挡的色依赖，高分辨率成像则寻找邻近污染源。没有一种检查能够保证单独排除一切伪阳性，组合证据必须按具体目标组织。

统计验证与动力学确认也不是同一种工作。前者比较行星模型与各种假阳性模型的相对可能性，在特定假设下认为替代概率足够低；后者可能取得由伴体引力产生的径向速度或其他动力学证据。统计验证不是没有价值，但其结论依赖候选模型集合与先验，不能写成已经直接测得伴体质量。

NASA 系外行星档案对假阳性概率表的说明，使这些状态能够在目录中分别表达。[49-2]状态字段看似行政细节，实际上承载证据等级。若汇总研究把不同状态混在一起，结果可能同时包含已确认对象、可靠候选和仍有重要争议的信号。一个宏大的“发现总数”可以因此失去精确含义。

直接成像也有自己的排错问题。HR8799 多行星成像研究展示了从强恒星光旁取得暗弱伴体信号的可能，[49-3]但图像中的点状结构仍需与背景对象、仪器散斑及处理伪影区别。重复观测中的共同运动、轨道变化和谱性质可以提供约束。看起来直接，并不意味着不需要经过推断。

直接成像与凌日偏好的对象也不同。较宽分离、较亮的年轻伴体有利于某些成像搜索，而短周期与合适轨道方向有利于凌日。把两种目录直接拼起来，不能自动形成公平的行星总体。方法的多样性扩大可进入世界，却要求统计时保留各自的选择条件。

直接成像中的一个困难，是恒星光经过不完美光学系统形成的散斑可能看起来像暗弱点源。改变观测角度、时间或波段，能够帮助区分天体位置与仪器相关结构；但是资料处理本身也可能削弱真实伴体信号。用已知人工伴体检验处理后的流量损失，是理解灵敏度的一种办法。看见某个残差点之后，仍要说明它怎样通过这些检验。

共同自行也不是所有轨道参数的替代。若暗弱源随恒星一起运动，背景恒星解释会受限制，但完整轨道通常还需要更长时间弧段。只看到一小段运动时，允许轨道可以很宽。成像给人的空间直观越强，越应提醒读者实际取样只是漫长轨道的一部分。

亮度到质量的转换还依赖年龄与冷却演化。年轻伴体可能仍然发出较强热辐射，同样质量在不同年龄下可以有不同亮度。若年龄估计不确定，仅凭亮度给出的质量也带有相应条件。径向速度的质量—倾角退化与成像的亮度—年龄—模型依赖不同，因此方法互补不是“直接的一定胜过间接的”，而是各自限制不同可能性。

在统计验证中，背景食双星等情景的先验数量也会影响结果。若目标附近恒星密度被低估，假阳性概率可能被低估；若某种重要替代根本没有进入模型，精确计算也不能自动补上遗漏。因此概率应当伴随候选集合与输入资料说明，而不是作为脱离模型的信誉分数。

49.3 人口统计为何需要选择函数

如果一项调查找到一百个候选，不能立刻把它除以观察恒星数就宣称得到真实出现率。首先，只有一部分轨道方向会发生可见凌日；其次，即使凌日发生，信号也可能太弱、观测次数太少或落入缺测区间；最后，被程序选中的候选还可能包含假警报。总体推断必须同时处理几何概率、检测完备度与可靠性。

Kepler 的 DR25 目录以统一自动流程和可评估的完备度、可靠性为重要设计目标。[49-4]这样的目录与“尽可能收集所有有趣对象”的目录服务不同任务。前者为了统计，需要知道选入和排除规则；后者为了个体研究，可以采用更复杂的人工判断。两者没有天然高下，但不能不加说明地混用。

注入恢复测试是估计完备度的一种重要方法。将已知参数的模拟信号注入资料，运行同一搜索和审核流程，观察哪些被找回。它回答的是在相应资料与流程下，这类信号有多大概率进入目录。只在干净人工噪声中测试，往往不足以代表真实恒星变化和仪器系统误差；因此测试放在哪一层、覆盖哪些目标，也属于结论的一部分。

可靠性处理另一方向的问题：目录中被标为候选的对象，有多少可能是误认。完备度低意味着漏掉真实信号，可靠性低意味着纳入虚假信号。二者不能相互抵消，也不能用同一个百分比概括。提高筛选门槛可能提升可靠性，同时降低完备度，这是一项需要根据研究问题平衡的选择。

Bryson 等人的研究显示，在特定小半径、较长周期区域，对可靠性的处理可以显著改变出现率估计；恒星参数版本变化也会影响结果。[49-5]这一事实说明所谓“类地行星有多普遍”并非从候选数量直接读出，而是数据、恒星性质和检测过程的联合推断。本章不将某个版本的数值提升为永久宇宙常数。

SDE 的路径差异在此可以严格检验。从已发现样本出发归纳行星分布，会突出易被发现的对象；从检测函数出发模拟一组潜在总体，再看哪组能够产生观测目录，可以显露被漏掉的区域。如果两条路径未经校正就给出不同结论，差异首先可能来自观察制度，而非宇宙在不同研究者面前改变物理性质。

可以用一项完全假设的调查算清这些因素。设观察一万颗适当恒星，某类行星若存在，其轨道方向造成可见凌日的概率为百分之五，已发生凌日后被流程找到的概率为百分之八十。如果每颗恒星平均有零点二颗该类行星，那么在忽略其他复杂性的教学模型里，预期检出数为一万乘零点二乘百分之五乘百分之八十，即八十个。

若直接把八十除以一万，得到的百分之零点八只是进入目录的比例，远低于模型设定的每星零点二颗。反过来，若八十个候选中有一成是错误信号，则真实检出应按可靠性修正，再推回总体。这里的数字纯为说明推断结构，不是 Kepler 某一真实子样本的结果。

现实资料中的检测概率不是一个统一百分比。不同恒星亮度与活动不同，不同行星半径和周期不同，观测缺口也不同。因此一般需要逐目标或分层计算有效敏感度。将一个在短周期大行星区测得的高完备度，直接用于长周期小行星，会产生系统性偏差。最需要估计的稀有区域，也常常是检测最困难的区域。

注入恢复本身也有边界。如果人工信号只覆盖特定形状和参数，恢复率就只适用于这些条件。若流程对真实多行星系统的相互影响、恒星活动或特殊凌日形态反应不同，测试还需扩展。一个高恢复比例不能被推广为“算法找到了所有行星”，它表达的是在明定模拟分布上的表现。

对于样本中没有检出的区域，结论也应与灵敏度相连。若对某类对象几乎没有探测能力，零检出提供的限制很弱；若本应容易发现而仍没有检出，才可以给较有力的上限。空白地图不是单一含义，既可能是对象少，也可能是研究尚未有效进入。这种区分决定下一项任务应增加观测能力，还是修改形成模型。

可靠性还可以通过反转或打乱时间序列等适当的对照资料来研究，但对照是否保留真实噪声特征必须审查。若对照处理破坏了相关噪声，得到的假警报率可能过于乐观；若保留了真实凌日结构，又可能把真信号混进背景。统计实验的设计与天文实验一样，需要说明哪些结构被保留、哪些被破坏。

目录审核中的人工介入也影响统计。专家可以识别算法漏掉的异常，但若只对最有趣目标进行深入复查，而其他目标按自动规则保留，样本就不再具有完全相同的选择过程。解决方法不是禁止专家判断，而是记录额外观察和改变状态的原因，并为总体估计定义可追踪子样本。用于个体发现的最好目录，未必是用于总体推断的最简单目录。

例如，某项团队为候选目标补做高分辨率成像，发现部分宿主存在伴星。如果只修订这些候选的半径，而其余被观察恒星的伴星情况仍未知，分子与分母采用的恒星信息可能不对称。发生率研究需要审查这种不对称，而不只更新候选表。Thompson 等人的目录设计对统一处理的强调，正与这种问题有关。[49-4]

另一个容易误解的量，是每颗恒星平均行星数与拥有行星的恒星比例。一个系统可以有多颗行星，两种量不相同。若一百颗恒星中二十颗各有两颗某类行星，平均每星数量是零点四，而宿主比例是百分之二十。研究标题只说“行星出现率”时，应明确采用哪一种定义，尤其在多行星系统丰富的样本中。

跨方法合并还需要处理重复对象。同一行星可能既有凌日又有径向速度记录，它们提供两种证据，但不能被当作两颗行星计数。若两种调查选择了不同恒星群，直接合并也会混淆宿主性质。有效联合必须先建立对象对应和调查选择模型，再讨论总体，而不是把不同数据库数字相加。

49.4 行星类型反过来改写形成模型

当样本达到足够规模和精度，研究不再只问某颗行星是否存在，而开始问半径、周期和系统结构如何分布。富尔顿等人的研究利用改进的参数讨论小行星半径分布中的间隙，[49-6]这种结构把个体发现提升为种群约束。它不只是目录上的凹槽，还要求解释形成与演化怎样共同塑造对象集合。

不过，一个分布特征的出现需要先检查测量误差。若半径误差较大，真实间隙可能被抹平；若恒星参数具有系统偏差，也可能改变间隙位置。样本选择与完备度若随半径急剧变化，更可能制造表面结构。因此，从直方图到发生机制之间仍有一段工作，不能看到双峰就直接宣布两种本体被证明。

不同机制还可能产生相近分布。例如，行星形成时的组成差异与后期大气损失，都可能影响半径。要分辨它们，需要考察轨道周期、恒星性质和年龄等额外关系。SDE 发生学应鼓励这样的交叉约束，而不是只把分布命名为“新特征显露”就结束分析。

一旦某种稳健结构成立，它会回写观测计划。研究者可能优先测量间隙两侧对象的质量，或比较不同恒星环境，以寻找机制差别。这里真正改变的是问题与样本选择，不只是文章标题。回写也可能产生偏差：如果所有资源都投向已知特征附近，其他类型可能更难被发现。因此需要在目标跟踪与开放调查之间保持平衡。

行星人口统计最终呈现一个由多方法共同构成的世界。它既包含个体轨道与大气，也包含分布、缺口和关联。对于半径分布中的某一道谷，下一步不能只增加一幅更平滑的直方图。需要测量谷两侧对象的质量，比较不同辐照与恒星年龄条件，并重新评估参数误差是否

足以移动边界。只有这些区别开始指向不同发生机制，目录中的形状才真正转化为行星演化的证据。

这些区别使行星人口研究逐渐形成一种新的严谨性：不只发表发现，还发表没有发现时的灵敏度，不只公布候选，还公布审核与误认条件。一个对象未进入目录的原因，开始与进入目录的对象同样重要。关于“有多少”的知识，由此依赖一份关于“怎样才会被看见”的知识。

这种方法不会抹去发现的惊喜。相反，当选择效应被充分理解后，一个仍然异常的系统或分布才更有力量挑战理论。搜索规则能够解释的集中，不宜过早赋予宇宙意义；搜索规则解释不了的结构，则值得进一步投入。统计与个体探索在这一点上相互支持，而不是一种负责冷却想象，另一种负责制造故事。

小样本区域的不确定性还具有离散性质。只发现一两个候选时，新增或撤销一个对象就会明显改变估计，不宜用过多小数位制造稳定感。若结果依赖少数低可靠候选，应明确展示去除这些对象后的变化。这样的敏感性分析比只给一个最优数值更能帮助读者判断证据重量。

恒星子样本的限定也必须与结论一致。在某类矮星中估得的出现率，不应未经论证推广到所有恒星；在某个周期范围内得到的结果，也不能直接外推到更远轨道。总体名称越大，越需要检查实际数据覆盖了多少范围。可靠的人口统计始于准确说明自己统计的是哪一群对象。

同一份目录可以支持不同问题，但每次使用都应重建相应分母。研究平均行星数、宿主比例或多行星系统结构，所需统计对象并不相同。把对象域与选择条件写清，才使不同研究的结果真正可比较。

对于低完备度区域，外推的模型依赖尤其需要单独报告，不能隐藏在一个全局平均误差中。

本章参考文献

[49-1] Charbonneau, D., Brown, T. M., Latham, D. W., Mayor, M. Detection of Planetary Transits Across a Sun-like Star. 1999 预印本，2000 发表。https://arxiv.org/abs/astro-ph/9911436

[49-2] NASA Exoplanet Archive. Kepler False Positive Probabilities Table Description. https://exoplanetarchive.ipac.caltech.edu/docs/fpp_TableDescription.html

[49-3] Marois, C., et al. Direct Imaging of Multiple Planets Orbiting the Star HR 8799. Science, 322, 1348–1352, 2008. https://arxiv.org/abs/0811.2606

[49-4] Thompson, S. E., et al. Planetary Candidates Observed by Kepler. VIII. A Fully Automated Catalog with Measured Completeness and Reliability Based on Data Release 25. Astrophysical Journal Supplement Series, 235, 38, 2018. https://escholarship.org/uc/item/5br413nb

[49-5] Bryson, S., et al. A Probabilistic Approach to Kepler Completeness and Reliability for Exoplanet Occurrence Rates. https://arxiv.org/html/1906.03575v4

[49-6] Fulton, B. J., et al. The California-Kepler Survey. III. A Gap in the Radius Distribution of Small Planets. 2017. https://arxiv.org/abs/1703.10375

第 50 章 生命迹象与天文可居住性：证据距离有多远

50.1 “宜居带”描述的是什么

当一颗系外行星被称为处在宜居带内，公众很容易把它理解成另一个地球。实际上，宜居带通常是在特定大气和气候假设下，讨论行星表面可能维持液态水的轨道区域。它是一项条件化模型判断，不是对海洋、生态系统乃至居民的观测报告。科帕拉普等人的工作更新宜居带估计，并讨论其对恒星性质及行星质量等条件的依赖。[50-1][50-6]

相同入射能量未必产生相同表面环境。大气组成、云、反照率、自转、热输运和水的存量都会影响气候。一个模型可用较少变量描绘主要范围，但不能因此消除其他变量。将宜居带当作筛选工具是合理的，将它当作生命存在的充分条件则改变了原命题。

时间尺度还带来第二层限制。某颗行星今天处在某一辐照范围内，不代表它过去一直如此。恒星演化、行星大气逃逸和地质过程会改变环境。若早期已经失去大部分水，后来进入适当辐照区也未必恢复类似地球的条件。宜居性因此需要历史，不能仅由当下轨道距离决定。

“宜居”还依赖研究生命的范围。以已知地球生命和液态水为参照，有其经验基础，但不能假装已经穷尽宇宙所有可能生命。合理做法是说明研究对象为哪类环境与化学条件，把可操作的搜索范围与关于生命本质的更广泛问题区分。没有经验支持的无限想象，不能直接用于确认一个天文信号。

SDE 在此的贡献，是迫使我们检查一个价值浓厚的名称怎样成立。宜居属性不是轨道独有的内在标签，也不是观察者单方面的愿望，而由辐照、介质、演化与所指生命条件共同限定。E 是这些特征的相互规定；改变其中一项，可能需要重新计算或重新定义判断，而非只在原结论后添加脚注。

宜居带边界不是一圈画在宇宙中的硬墙。模型可以给出内外边界，但边界随采用的温室条件、云处理和其他假设而变化。处在边界附近的对象尤其不适合被简单分成绝对宜居与绝对不宜居。更有信息的做法，是报告其在一组合理条件下允许怎样的环境，并指出最影响判断的未知量。

平衡温度同样不是地表温度的同义词。它依据吸收与辐射的简化能量平衡得到，可以帮助比较辐照尺度，但实际表面还受温室效应和热输运等影响。若一个新闻数字给出行星温度而不说明是哪种温度，读者可能误以为已经拥有当地气象记录。数值精确不等于定义清楚。

还要区分表面宜居与地下可能环境。传统宜居带主要服务于表面液态水的遥感搜索，不能据此断言带外任何位置都绝无适合生命的局部环境。另一方面，地下环境即使在物理上可

能，也未必能够留下远距离可识别迹象。可居住性与可探测性是不同轴线，搜索计划需要同时考虑，不能让一个概念包办另一个。

50.2 远距离大气光谱如何生成

凌日时，部分恒星光穿过行星大气，不同波长受到不同影响，可以形成透射光谱。次食、相位变化和直接成像等方法则提供其他观测入口。这些方法取得的不是从远方装瓶运回的空气，而是混合光中的细小差别，需要通过辐射与仪器模型解释。

从一个谱形到某种分子，已经包含多重判断。首先要确认差异超出噪声与仪器系统变化；其次比较分子的吸收结构是否与资料相容；然后考虑云、温度、压力和其他分子是否能产生类似特征。检索模型输出一个丰度范围，并不意味着丰度已经脱离这些条件被直接测出。

恒星本身尤其不能视作完全均匀且不变的灯泡。星斑和亮斑使行星遮挡区域的光谱与整个恒星平均光谱不同，可能在推得的透射谱中留下结构。Rackham 等人的研究将这一问题明确化，说明恒星不均匀性可以制造或改变被归到行星大气的特征。[50-2]误差来源不只在探测器，也在照明系统本身。

这个例子显示对象属性如何在互动中成立。行星大气当然具有物理过程，但被报告的“大气谱”需要从恒星、行星和观测系统的共同资料中分离。若分离条件改变，某个分子信号的强度或可信度也会改变。不能把这种依赖说成大气只存在于模型，也不能把它当作无关紧要的技术清理。

多次观测与多波段覆盖可以增加区分能力。某个分子若有多个应当同时出现的特征，观察其一致性比依赖单个起伏更有力量。不同时间的恒星活动变化也可以帮助检验污染。但重复使用同一种处理流程可能保留同一系统偏差，所以复测之外还需要处理与模型比较。

大气光谱生成之后，尚未进入生命确认。确认某种气体、限制其丰度、说明其化学来源，是三个不同层次。把它们连接起来需要更多资料，而不能用“这种气体在地球上与生命有关”直接跨越。思想史在此必须保存推断链的长度，因为公众最容易在最吸引人的结论处忘记前提。

透射信号的强弱与大气尺度高度有关。在简化等温模型中，尺度高度与温度成正比，与平均分子质量和表面重力成反比。较轻、较热的大气可以在更大高度范围内改变遮挡面积，较重分子组成的大气则可能产生更小信号。这只是解释趋势的近似，并不把复杂大气压缩成单一温度。

因此，容易探测到大气的行星，不必最接近地球。观测便利性可能偏向更大、更热或具有膨胀大气的对象。若样本主要来自这些条件，不能直接把其大气统计推广到温和岩石行星。和行星发现一样，大气世界也受到测量选择的组织。

光谱平坦同样有多种解释。可能缺少可检测大气，也可能云或雾遮蔽了特征，还可能仪器精度不足。仅凭没有清楚吸收峰就宣布“没有大气”，需要额外条件。相反，存在一个起伏也不能立即确定某种分子，因为不同吸收源在有限波段里可能重叠。

检索中的丰度、温度和参考压力可能存在退化。若增加某种分子丰度与改变大气结构能产生相近谱形，后验范围就应反映这种不唯一性。过窄先验可以让输出看起来非常精确，但精确来自限制而非新增数据。比较不同合理先验和模型族，是检查这一点的具体办法。

多波段联合也要处理不同观测时刻。若恒星表面状态改变，把不同季节的片段拼成一条高精度光谱，可能形成并不存在于任何一次观测中的组合特征。联合分析需要估计这种变化，而不是只按波长无缝连接。大气信号的细微性，使时间协调成为化学解释的一部分。

50.3 非生命化学的竞争

氧常被视作重要生命迹象，是因为地球生物圈使其具有特别意义。但某种气体在地球上的主要来源，不自动等于它在所有行星上的唯一来源。光化学、地质过程和大气逃逸可能形成不同化学状态，宿主恒星的光谱也会改变反应路径。Harman 等人的研究专门分析不同恒星条件下非生物氧累积的可能性。[50-3]

Luger 与 Barnes 的模型进一步讨论某些行星在早期失水后积累非生物氧的情景。[50-4]这类模型不是宣布每个含氧行星都没有生命，而是提出必须排除的一组替代。它的价值在于告诉观测者哪些环境信息值得补充，例如水、压力、恒星历史和与氧相关的其他化学约束。

单一分子与分子组合的证据力也不同。若两种气体难以在给定条件下长期共存，它们的同时存在可能提示持续补给或化学失衡。然而“不平衡”仍不必然等于生命，因为非生物过程也可以持续驱动系统。需要比较产生速率、消耗途径和环境能量，而不是把化学失衡当作神奇签名。

这里可以用两条研究路径相互制约。第一条从生命假说出发，推算其可能产生哪些光谱与时间变化；第二条从无生命的反应网络出发，询问在允许环境中能否产生相同资料。若只走第一条，容易把一切相容现象视为支持；若只走第二条而无限发明未受约束机制，又可能使生命假说永远不可能被接受。公平比较必须要求双方使用同样严格的物理和观测条件。

证据强弱依赖候选假说的预测差异。如果某个信号在有生命和无生命条件下都很常见，它的区分能力有限；若一个由多项独立资料支持的组合很难被现有非生物机制解释，则生命解释的相对支持会增强。但“现有模型尚不能解释”不是绝对证明，因为模型集合仍可能不完整。

对这种开放性，最有用的回应不是放弃判断，而是明确下一项最能改变判断的观测。如果扩大波段能区别两种分子，如果更好的恒星活动资料能排除污染，如果独立质量与半径限制能够改变大气结构，就应把资源投入这些分叉点。三方程在这里表现为候选 S 反过来组织 D，并使 E 的关键不确定项受到新的约束。

化学来源可以用收支关系理解。某气体的大气含量取决于生成、消耗、输运和逃逸等过程。相同丰度可以来自较快生成与较快消耗，也可以来自两者都很慢。只测到含量，未必知道通量。若希望用丰度推断生物活动，就需要对寿命和汇过程施加额外限制。

这使环境背景成为生命判断的实质组成，而非文章末尾的一段背景介绍。恒星紫外谱改变光化学速率，海洋和地表可能吸收或释放气体，火山和岩石反应也会提供来源。Harman 等人的模型所讨论的，正是这些条件如何改变氧的非生物解释空间。[50-3]对它们不确定时，生命解释的区间也应变宽。

“难以由非生命产生”还必须说明难在哪里。是现有模型需要不合理通量，还是违反已测温度和组成，抑或只是尚未尝试某个反应网络？前两类限制具有较强排他性，最后一种更多反映研究未完成。将缺少研究混同于不可能，会使生命判据过早封闭。

同一原则也适用于假阴性。存在生命不保证形成强烈可遥测信号，生物圈可能很弱，气体可能被迅速消耗，或者生命主要处于难以影响大气的环境。因此没有生物标志不能直接等于没有生命。生命搜索的结果需要同时讨论检测概率和标志产生条件，而不是只讨论仪器噪声。

这不会使任何结论都无法成立。它要求结论适当限定：在这些环境和信号产生假设下，资料支持或限制了哪一类生物圈。这样的命题比“这里有／没有任何生命”窄，却更接近实际可验证内容。知识可以通过许多有限而可靠的命题逐步增长，不必等待一次包办所有可能的终极宣布。

50.4 证据等级与公共传播

关于生命的消息具有强烈的公共吸引力，也最容易在传播中越级。“宜居带内的行星”“可能具有大气”“某种分子候选”“潜在生物标志”和“生命确证”分别表达不同命题。若标题将它们压成“发现外星生命”，即使正文后来保留谨慎措辞，也已经改变公众最先接收的对象。

Schwieterman 等人对可遥测生命迹象的综述，将气体、表面及时间变化等不同线索放在更广框架中，[50-5]说明搜索并不只有寻找某一条谱线。多种线索的协同是机会，也是责任：若它们共享同一恒星模型或相同处理偏差，不能简单当作几份独立证据相乘。

研究报告应让读者知道结论的升级条件。确认分子意味着哪些观测限制已经满足，确认其生物来源还需排除哪些机制，哪一项新结果会使当前解释下降。这样的表述不是削弱研究，而是使发现成为可继续检验的公共对象。一个能够清楚说明何时会错的结论，比没有边界的宏大宣告更有科学生命力。

同样，后续未能复现某个候选，不应被简化为“科学什么都不知道”。它可能说明信号过弱、处理方案不稳健，或某个替代机制此前被低估。撤回与修正构成研究的必要环节。但也不能用“科学本来会修正”免除最初过度宣传的责任，因为传播造成的误解未必会随论文更正自动消失。

SDE 的主体、互动与客体在这个问题上都不能缺席。客体侧有气候和化学约束，互动侧有观测与传播过程，主体侧有人类寻找同伴的期待与判断。承认期待存在，有助于辨认它如何影响选题和解释，却不能让期待替代证据。一个公平的生命搜索，应允许令人失望的无生命解释获得与令人兴奋的解释同等严格的检验。

从遥远光谱到生命之间的距离，并不仅是空间距离，更是尚需补足的发生机制与排他性证据。天文学已经能够构成越来越精细的行星环境世界，但环境知识与生命知识仍需分别成立。保持这种区分，既保护探索的创造力，也保护结论的可信度：未知不是任意填充的空白，而是下一轮互动应当具体进入的位置。

一个简化的贝叶斯思想实验能够显示为什么罕见性与区分力必须分开。某个谱形即使在无生命模型下比较少见，若在有生命模型下也同样少见，它未必强烈偏向生命。相反，某组资料如果明显更符合生命模型，才提供有方向的比较支持；最终判断仍与先验及模型集合有关。这里并不指定任何实际行星的生命概率，因为现有输入未必足以给出可信单值。

这种概率纪律与公开传播可以相容。研究者可以明确说一个信号值得优先复测，同时不把它称为生命证据已经完结。优先级是关于下一步资源安排的判断，确证是关于当前证据状态的判断，二者不必使用同一门槛。一个候选足够有趣值得投入，并不意味着它已经足够可靠可以写进教科书。

后续观测如果否定了原候选，仍能改进搜索方法：哪类恒星污染容易混淆，哪个波段需要更好校准，哪些模型先验过强，都可以回写研究系统。只有当这些变化真正进入操作与记录，失败才转化为能力；若只是把旧新闻撤下，再对下一个候选重复同样宣传，就没有完成知识上的回写。

生命假说的检验还需要防止事后灵活改写。若一种气体没有出现就说生命不产生它，出现后又说它证明生命，假说便失去区分力。研究者应尽可能预先说明在特定生物与环境模型下期待哪些组合，以及哪些结果会使该模型受挫。生命可能多样，并不意味着每一种解释都可以免于失败。

公共传播也可采用明确的条件句：在排除某类恒星污染且化学模型适用的前提下，资料支持某种气体。这比把所有条件塞进附注更诚实，也让公众理解后续修正为何发生。条件改变时，结论调整不是任意摇摆，而是沿原本公开的依赖关系进行。

同一个研究领域可以同时保持强烈探索意愿与较低当前确证程度。对未知生命有兴趣，不要求对每个候选表现出高度确信；坚持证据门槛，也不要求贬低探索的价值。真正可积累的进展，往往发生在两者之间：把最值得追问的问题，转成能够区分解释的下一次观测。

如果未来出现强得多的候选生命证据，判断框架也不应只靠一个固定分数自动升级。必须检查新增证据是否解决了此前关键替代，是否来自新的独立条件，是否能被另一团队重现。更多同类数据可能提升精度，却未必解除同一个系统偏差；更有区分力的数据，才能缩短从分子到生命的证据距离。

这要求观测申请明确写出竞争模型会给出什么不同预期，使下一次测量有机会真正改变判断，而非只重复一个吸引人的候选名称。

本章参考文献

[50-1] Kopparapu, R. K., et al. Habitable Zones around Main-Sequence Stars: New Estimates. 2013. https://arxiv.org/abs/1301.6674

[50-2] Rackham, B. V., Apai, D., Giampapa, M. S. The Transit Light Source Effect: False Spectral Features and Incorrect Densities for M-dwarf Transiting Planets. 2018. https://arxiv.org/abs/1711.05691

[50-3] Harman, C. E., et al. Abiotic O2 Levels on Planets around F, G, K, and M Stars: Possible False Positives for Life? 2015. https://arxiv.org/abs/1509.07863

[50-4] Luger, R., Barnes, R. Extreme Water Loss and Abiotic O2 Buildup on Planets Throughout the Habitable Zones of M Dwarfs. 2015. https://arxiv.org/abs/1411.7412

[50-5] Schwieterman, E. W., et al. Exoplanet Biosignatures: A Review of Remotely Detectable Signs of Life. Astrobiology, 2018. https://arxiv.org/abs/1705.05791

[50-6] Kopparapu, R. K., et al. Habitable Zones around Main-Sequence Stars: Dependence on Planetary Mass. 2014. https://arxiv.org/abs/1404.5292

第 11 编

大规模巡天、AI 与证据责任

第 51 章　巡天不只是更多数据：选择函数重写对象集合

第 52 章　Gaia：百亿观测怎样成为视差、星表与银河结构

第 53 章　模拟宇宙与统计模型：模型结果能成为何种证据

第 54 章　AI 巡天候选：发现、分类还是制造标签

第 55 章　宇宙学张力：共享误差与理论重开的判据

第 51 章　巡天不只是更多数据：选择函数重写对象集合

51.1　天空覆盖与探测深度的权衡

一份包含百万行记录的天体目录，很容易使人产生一种直观：天空的总体已经进入计算机，剩下的任务只是寻找其中的规律。但目录的第一项科学内容，恰恰是哪些东西没有进入目录。没有被望远镜覆盖的方向、没有跨过探测阈值的光源、被当成伪影排除的信号，以及被挤在亮源周围而无法分离的暗源，都参与决定最后出现的分布。它们不是结果之后附加的瑕疵说明，而是这个统计对象的构成条件。

巡天首先分配有限时间。对一个方向停留更久，通常有利于积累弱信号，却会减少在同一时期内能够访问的方向；更频繁地重访，可以识别短时变化，但也可能压缩空间覆盖。这不是技术尚未完善时才有的暂时困难。任何具体观测系统都必须在角度、波段、深度和时间之间作出安排。所谓“覆盖天空”，从来不是一个不需要补语的操作：它总是以某种灵敏度、某种采样节律和某些质量条件覆盖天空。

因此，选择函数不是给目录套上的数学装饰。它表示具有一定属性的源进入特定分析样本的机会。这里的属性可能包含方向、亮度、颜色、大小、拥挤程度和变化时间尺度。不同研究问题需要不同的选择函数；为寻找亮类星体准备的目标规则，不能未经重新评估便当成所有星系的抽样规则。一个记录“足够多”的目录，也可能对问题所需的关键类别几乎没有辨别力。

SDSS 关于光谱样本完备性的说明，把成像探测、目标选择、观测铺排和光谱成功分类分别列出。这种区分说明，目录成员资格是多重条件的交集，而不是望远镜看见一次就自然成立的身份。[51-1] 一个源即使出现在图像里，也未必被选为光谱目标；即使获得光谱，也未必能可靠确定红移。因此，图像样本、目标样本与可靠红移样本，不能在统计推理中悄悄互换。

可以用一个纯粹示意的例子理解这一点。设想同一区域内有两类总光度相近的星系，一类光集中在紧凑中心，另一类光分散在较大面积上。若探测算法依赖每个像素或局部区域的对比，后一类可能更难进入目录。于是样本中紧凑星系占优势，并不立即意味着宇宙本身优先制造紧凑星系。必须先比较两类源在实际噪声、背景和分离算法下的检出机会。这个例子不是某次巡天的实测结果，而是对推断结构的说明。

在 SDE 分析中，显露 S 是可查询的样本及其分布，差异序列 D 是逐场观测、阈值筛选和质量判断，特征纠缠 E 则表现为灵敏度、形态可辨认性、背景估计与成员资格的相互规定。E 不能简写为“望远镜加制度”。例如，改变背景估计会改变扩展源的亮度与面积，二者又共同影响是否达到检测条件；被认为可靠的源集合随后可能被用于重新估计背景或校准。这才是需要展开的纠缠机制。

从巡天设计进入，能够说明目录如何形成；从目录的一条漂亮规律反向进入，则会追问它在未被纳入的区域是否仍然成立。两种入口的差异具有实际后果：前者可能展示计划执行得很成功，后者却可能发现该成功不足以支持某个总体结论。选择函数把这种差异变成可以讨论和检验的研究问题。

51.2 目录建立的匹配难题

跨目录研究常被描述为把资料连接起来。但连接发生以前，研究者必须判断两条记录是否属于同一天体。这一步不是纯粹的数据管理。假如一条射电记录和一条光学记录被错误地合并，系统便可能创造出一个拥有不相容属性的“复合天体”；如果同一快速移动天体在两个时代被当成不同对象，目录又会把一个持续对象切成两个。这些错误直接改变研究对象的数量、类型和物理解释。

最简单的匹配办法是比较位置，把距离最近的记录视作同一对象。它在位置精确、天空稀疏且历元接近时很有用，但没有普遍充分性。位置误差区域内可能有多个候选；不同波段的辐射中心也可能不同；时间间隔较长时，自行和其他运动必须进入比较。于是“最近”只有在明确误差模型和对象假设后，才成为有意义的证据。

Budavári 与 Szalay 提出的概率交叉识别框架，将多个观测属于同一源的假说与其他假说比较，并允许把位置以外的属性纳入判断。[51-2] 这一工作的重要性不在于给匹配盖上概率印章，而在于使对应关系的前提能够写出。匹配概率依赖怎样的源密度、怎样的误差和怎样的物理相容性，研究者因而有机会检查结果，而不是只接受数据库里已经连好的行。

但引入颜色或红移也不是无条件改善。假如研究目标正是寻找颜色异常的对象，而匹配阶段先强烈偏好颜色相近的候选，异常可能在正式分析前就被排除。这里出现了一个发生学上的闭环：已有类别决定匹配条件，匹配条件决定能够显示的多波段对象，多波段对象又被用来说明已有类别的稳定。这种循环可能提高普通对象的处理效率，也可能缩小对新对象的敏感性。评价它必须同时考察正确匹配率与异常保留能力。

多对一关系使问题更加明显。低分辨率图像中的一个亮斑，在高分辨率图像里可能分成若干源。强行要求每条记录只对应一条记录，会把真实的观测差异误当成数据冲突。反过

来，无限制地允许多重对应，又会使属性归属失去确定性。合理的处理往往需要保留关系结构及其不确定性，而不是在最早阶段便把所有疑问压成一个唯一编号。

这意味着天体身份至少有两个层次。一个层次是某次观测内的检测身份，即算法把哪些像素或时间序列片段识别为一个源；另一个层次是跨观测的持续身份，即不同检测为何被认为涉及同一研究对象。两者都与自然过程有关，却都不能仅凭名称自动成立。目录中的统一标识符方便调用，并不消除它所封装的对应判断。

对思想史而言，现代数据库因此延续了古代星表的一项根本困难：不同时间和条件下的显露怎样被认作同一个对象。现代方法把问题的规模和精度推到新的层次，也把过去由少数观察者承担的判断分散到算法、格式、校准和用户查询之中。认识主体不再只是看见一颗星的人，而是能够追踪这条身份链、检查其适用范围的复合研究系统。

一种可行的检验是保留多个匹配候选，再观察某项群体结论是否依赖其中少量含糊关联。如果删去歧义项就使结果消失，结果便需要以更弱的形式报告。若不同合理匹配方案产生相近结论，则获得的是关于该结论的稳健性支持。这里能够被确认的不是“所有匹配都正确”，而是当前的不确定性不足以推翻某个有限主张。

51.3 由样本到总体分布

样本中的数量并不直接等于总体中的比例。一个亮度受限的观测，在较远处往往更容易保留本征明亮的源。如果把近处和远处的样本直接放在一起比较，就可能把可见性差异当成演化。类似问题也出现在时间域：持续较长的事件更容易与离散访问相遇，极短事件可能在两次观测之间发生并消失。数量统计必须同时讨论空间可见范围和时间可见窗口。

作为现代说明式，可以把观测分布写为真实属性分布经过测量误差与选择过程后的结果。设真实属性为 x，记录属性为 y，误差核为 $K(y|x)$，纳入分析的概率为 $q(y,x)$，则样本分布与对 x 积分的 qKp 有关。这个表达式只是展示依赖结构；具体巡天的似然还需要考虑计数过程、背景污染、相关误差和缺失机制。它提醒我们，同一观测分布可能由不同真实分布与选择组合产生。

由此可见，完备度与污染率承担不同任务。完备度关注应当出现的目标有多少被保留，污染率关注保留的记录中有多少不属于所研究类别。提高门槛往往可以减少污染，却可能同时降低完备度；放宽门槛能够保留更多微弱信号，也可能使伪影大量进入。不存在脱离研究目标的“最干净样本”。对于稀有事件，少量误报也可能支配候选集合；对于总体分布，过严筛选又可能造成更严重的缺失。

人为注入模拟源并测量回收率，是研究检出能力的一种办法。但注入源的形态、颜色和变化模型必须涵盖研究对象的实际范围。若只注入熟悉的点源，回收率无法直接说明复杂扩

展源的完备性。验证工具也具有自己的选择结构。严密分析不是无限追加一句“可能有偏差”，而是把某种偏差具体转化为可测试的输入变化，观察最终结论如何响应。

统计权重可以补偿部分不均匀抽样，但不能创造完全不存在的信息。当某一区域或某类对象的纳入概率为零，给已经看到的对象更大权重，无法凭空恢复被完全遮蔽群体的性质。此时必须引入其他观测或明确限制结论的总体范围。把“在我们能够观测的域内成立”写成“宇宙普遍如此”，是从样本到世界时最容易发生的越界。

误差还有相关性。若同一校准偏移同时影响大量源，增加源数可以减少部分随机波动，却不会让共同偏移按同样速度消失。样本越大，统计误差条可能越窄，系统偏差反而越显眼。一个精密的错误结论，可能比一个粗略但诚实的结论更难修正，因为它使不相容的新结果看起来像别人工作的失败。大数据的思想挑战，因而不是简单地从少量证据进入大量证据，而是重新区分可平均的误差与不可平均的依赖。

从 S=F(D,E)出发，本章真正增加的要求是：讨论分布 S 时，要让产生它的筛选 D 与相互制约 E 留在论证中。这个要求并不取代现代抽样理论；选择函数与误差模型已经给出专业分析工具。SDE 能够尝试贡献的，是把抽样结果进一步连接到对象类别、研究注意力与后续观测的改变，并追查这些改变是否反过来修订选择条件。若没有这种新增追踪，只把统计公式重命名为 SDE，便没有获得独立的解释增量。

51.4 公开数据库的回写

数据库公开改变了研究的起点。研究者不必先拥有望远镜时段，便可以从现成样本提出问题，重新组合波段，或者检查已发表结论。SDSS DR17 提供图像、光谱、数据立方体和目录参数等不同产品，也保留数据模型与访问方式。[51-3] 这些产品承担的证据角色不同：一列红移可用于统计，但若要检查某次分类失败，可能必须回到光谱和处理标记。

公开因此应当理解为一组有层次的可追踪能力。只公开最终参数，能够支持某些重复分析；公开质量信息，能够支持筛选敏感性检查；公开处理版本和算法，才能进一步解释不同版本为何变化。下载便利并不自动意味着所有人都能够重做全部观测链，但它能够使原来集中在一个团队内部的部分判断转化为共同检验对象。

这种开放也会改变研究注意力。容易查询、格式整齐、文档充分的数据，更容易进入教学、代码和论文。它们随后成为新的比较标准，更多研究围绕其覆盖域提出问题。这是生产性回写：可靠公共资源降低重复劳动，扩大研究参与。但同一机制也可能让难以标准化的对象退出常规分析。数据的可访问性与样本的代表性应当分别评价，不能因为一个数据库使用广泛，便认为它覆盖了所有重要现象。

目录版本变化进一步说明科学对象具有历史。某次重新处理可能修改源的分离、亮度或分类；同一个标识所对应的参数因此需要版本说明。严肃的复现不能只写“使用某巡天数据”，而要记录发布版本、查询条件、质量切割和关联规则。保存这些内容，不是向读者展示技术繁琐，而是保留使研究对象再次成立的路径。

在这里，三方程的后两个方向获得了具体问题。目录 S 改变 D，表现为后续观测根据候选集合安排重访、研究者根据已见分布选择新的统计量；S 与 D 改变 E，则要求更强证据，例如异常关联促成新的匹配规则，或公开复核导致质量说明与校准标准修订。不能因为一篇论文使用了数据库，就声称证据制度已经改变。回写需要可指认的规范或操作差异。

进一步说，选择函数还会随着研究者自己的操作变化。原始目录可能有清楚的覆盖说明，但使用者又要求每个源同时具有若干波段、某个质量值和可靠距离，最终分析样本便形成新的选择。原发布团队没有义务替每一种查询预先计算完备性，使用者也不能把原目录的质量承诺直接转移给任意子样本。科学责任在数据交接时并未消失，而是改变了承担位置。

一个具体的审查办法，是把最终结论沿查询过程反向拆开：先看去掉某项质量切割后关系是否改变，再看缺失某个波段的对象是否集中于特定方向、亮度或形态。这里并不主张盲目保留低质量数据，而是要求理解删去的对象有什么共同性。如果删除只减少噪声而不改变相关结构，较严格的样本可能合理；如果删除恰好拿走某一类物理状态，则需要重新限定总体主张。

同样，天空上的空白需要分类解释。没有安排观测、安排了却因天气失败、成功曝光但没有检出、检出后被质量筛选排除，是不同意义的空白。它们在最终数据库里都可能表现为没有一行记录，但对总体推断的意义完全不同。发生学分析的价值，就在于阻止这些差异在最终表格中被抹平，使“没有”也保留自己的形成过程。

这也给开放数据提出一个可检验标准：一个陌生团队能否凭文档恢复主要筛选步骤，并判断结果对这些步骤的敏感性？若只能重复下载结果，不能重建对象纳入的理由，开放仍然是局部的。局部开放可以很有价值，但必须以实际可复核范围衡量，不能把接口存在本身当成全部证据链已经透明的证明。

于是，巡天形成的新天文学世界，不只是旧天空多添若干记录。它包含可跨时间识别的对象、可检验的群体分布和可共同调用的证据链。这种世界的成立受到自然信号约束，也受到筛选与匹配过程的规定。它的可靠性不来自抹去这些条件，而来自使条件能够被看见、比较和修订。一个真正成熟的目录，必须让使用者知道它允许问什么，也知道哪些问题尚不能由它回答。

本章参考文献

[51-1] SDSS. Tiling: determining survey completeness. DR7 历史技术说明。https://classic.sdss.org/dr7/products/tiling/index.php

[51-2] Budavári, T.; Szalay, A. S. Probabilistic Cross-Identification of Astronomical Sources. 2008. https://arxiv.org/abs/0707.1611

[51-3] SDSS. Data Release 17. https://www.sdss4.org/dr17/

第 52 章　Gaia：百亿观测怎样成为视差、星表与银河结构

52.1　全局解算为何是联合问题

在星表中，恒星的位置、视差和自行各占几个整齐的栏位。这样的形式容易让人把它们想象成三把尺分别测出的读数：先量位置，再量距离，最后量运动。但对 Gaia 而言，精密参数的成立依靠一个联合问题。星像经过焦平面时留下的观测，需要结合航天器在该时刻的朝向、仪器几何和参考框架解释；这些条件本身又不能完全独立于恒星观测预先确定。星表的整齐，是经过联合求解才获得的整齐。

Lindegren 等人关于 Gaia 天体测量核心解算的论文，把源参数、姿态参数、校准参数和全局参数组织为相互联系的处理块。[52-1] 这里的“全局”并不表示每一颗星都必须同时出现在一次曝光中，而表示许多局部观测通过共享参数连接起来。某一时段的姿态影响该时段观测到的多个源，同一个源又在不同时间被扫描，从而把不同时段联系起来。精密参考系统是在这类交叠约束中建立的。

用现代示意式表示，一次观测可以写为 $y_k=h_k(s_i,a_k,c_k,g)+\varepsilon_k$。这里 s_i 代表源参数，a_k 代表相关时刻的姿态，c_k 代表仪器校准，g 代表全局参数，ε_k 表示未被模型解释的误差。这不是原论文完整观测方程，也不承担具体精度计算；它只揭示一个关键事实：某个残差不能在未作其他判断以前，自动归给一颗星。

假设一组同一时段的观测出现同方向偏移。如果把姿态固定为完全正确，求解可能把偏差分配给源参数；如果允许姿态变化，共同偏差又可能主要由姿态项吸收。哪一种归属合理，要靠其他时间、其他方向以及模型约束加以区分。恒星属性与仪器状态并不是任意混合，但也不是在数据进入分析之前就已经完全分离。能够分离到何种程度，本身是观测几何与约束结构的成果。

进一步考虑一个更抽象的退化例子。若仅有两组相对角度，整体旋转参考系而同步旋转所有源的位置，某些观测差异可以保持不变。数据本身便不能确定这类整体自由度。需要参考框架条件，才能把内部一致的解放到具有公共意义的坐标系统中。消除这种自由度不是替自然添加任意性质，而是明确哪些数值差别属于表示选择，哪些差别才对应可检验的天体状态。

官方处理说明指出，核心全局解算使用一组适当的源估计共同参数，然后能够对其他源进行相应的源参数求解。[52-2] 这也纠正了另一种直观：联合问题并不要求把所有参数不加

区分地塞入一个不可分割的计算。分块处理能够提高可操作性，但各块在整个迭代中仍保持依赖。计算上暂时固定一个块，不等于本体上证明它与其他块无关。

SDE 在这个案例中可以作出精确而有限的说明。S 是带误差与质量说明的源参数，D 是重复扫描产生的差异以及残差更新，E 是使这些差异能够共同约束源、姿态与校准的结构。E 不是卫星、人员和电脑的清单，而是“某类偏移能够怎样被归属”的相互规定。删除这种结构，只剩孤立读数，视差作为可比较属性便不能以同样精度成立。

然而，联合最小二乘、可辨识性与参考框架理论已经解释了这里的重要数学机制。不能因为这些机制可以写成 SDE 语言，就声称 SDE 发明了它们。本书希望追问的是更长的链条：联合参数怎样进入银河研究，使用中暴露的问题怎样促使质量模型更新，以及新的公共星表怎样改变后续研究能够提出的问题。只有把这些环节具体连接起来，分析才超出术语替换。

理解相互约束，还需要区分精度与可辨识性。某一参数的形式误差很小，表示在既定模型和误差假设下，数据对它具有较强约束；但如果模型遗漏了能够产生相似信号的过程，这个小误差并不自动涵盖遗漏造成的偏移。于是，检查参数之间的相关性、改变合理模型、引入不同几何的观测，承担的是不同于增加重复次数的任务。

设想所有观测都集中在很短时间内。即使每次测量非常精确，某些随时间展开的运动分量也可能难以区分。延长时间基线所提供的新信息，不是把同一个数字再测几遍，而是改变不同效应在数据中分离的条件。这类思想实验解释了为何观测节律属于 E 的构成部分：时间安排规定差异序列能够让哪些属性显露，而已经显露的属性又会影响后续重访如何安排。

还要注意，模型中的源并非永远能够被当作理想点源。若光心相对质心发生运动，仪器记录的是光的分布所规定的位置。把光心运动解释为质心运动，需要额外物理模型。这个区别提示，所谓一个天体的“位置”也有操作层次。它不是语言上的相对主义，而是要求说明哪个位置量被估计、以何种模型连接到研究问题。

52.2 质量标记的研究意义

一份科学星表不能只提供数值。它还需要告诉使用者，这个数值是在怎样的模型中得到的，观测对它有多强约束，拟合是否留下异常，以及不同参数之间有何相关性。否则，用户容易把有条件成立的估计当成无条件的对象事实。质量标记的存在，不是星表不够可靠的证明，而是其可靠性能够被使用的必要条件。

Gaia 数据处理中常见的 RUWE，中文可理解为经过重新归一化的单位权重误差。[52-6]它帮助使用者评估所采用天体测量模型与观测的一致程度。但较高值本身不等于某一种特

定物理原因。未解析的复合源、源结构、拥挤环境、模型不足或其他观测问题，都可能影响拟合表现。把它简单译成“坏星”标签，会把应当研究的差异直接删除。

同样，一个较低的拟合指标也不能证明所有偏差都已消失。共同校准偏差如果在模型中被一致地吸收，单源残差可以看起来并不异常。内部自洽与外部准确必须分别检验。用同一处理模型产生的残差验证同一处理模型，能够发现一些错误，但对某些共同模式缺乏辨别力；这正是外部参照和不同类别源有价值的原因。

EDR3 视差零点研究发现，偏移与星等、颜色和天空位置存在复杂联系，并且不同参数解类型需要不同处理；论文提供的是有适用条件的暂行校正表达，而非对全部源统一有效的常数。[52-3] 这项事实十分适合说明“平均偏差”与“个体修正”的区别。某个总体平均值可以提示问题存在，却不意味着对任何子样本直接减去同一个数就能消除偏差。

零点校正的困难还在于参照样本也有覆盖边界。某类源适合检验某一亮度或颜色范围，不自动代表其他范围。扩展校正往往需要另一些具有不同约束的样本。这里的稳健性来自多种参照之间的交叉约束，同时也要求说明它们共享哪些假设。若所有参照最终依赖同一有偏标尺，数量增加并不等于独立性增加。

选择质量阈值时，研究目标也必须明确。研究单星运动的工作可能合理地排除不适合单星模型的源；研究双星或异常运动的工作却可能恰恰关注这些对象。相同的质量标记，在不同问题中承担不同角色。阈值既改变噪声水平，也改变科学对象集合，不能只把它当成技术清洗。清洗记录应当保留，否则后来的使用者难以判断被删去的是随机错误还是某类特殊天体。

版本尤其重要。一次数据发布的校准函数和质量建议，不应不加检查地施加到另一版本。EDR3 的天体测量解算有专门论文说明其处理与数据性质。[52-5]数据处理更新可能改变源关联、观测时间范围和模型结构；同名栏目也需要核对说明。引用星表版本是对估计成立条件的指认，不是文献格式上的礼节。所谓“Gaia 给出的距离”，若没有说明数据版本、零点处理与距离推断办法，实际上还不是一个充分确定的科学陈述。

这里可以区分三种不同的变化。第一种是固定模型内参数的数值改进；第二种是模型适用范围和质量规则的修订；第三种是因新类型对象或新问题而改变研究程序。它们都可能有价值，但只有后两者在适当证据支持下，才涉及较强的条件重构。把每次迭代都称为新的认识革命，会掩盖真正改变对象定义或验证方式的转折。

因此，从星表数值 S 转向残差 D 和校准结构 E，具有可操作的意义：它把“这颗星很奇怪”的判断，打开为“天体性质、模型适用性和共同系统误差中的哪一项需要改变”。这一转向并不保证问题立刻解决，却阻止某一种解释由于数据库已经输出数值而自动取得优先地位。

外部验证也不能被理解成寻找一个永远正确的最终裁判。不同方法具有不同误差结构，比较它们是为了利用约束差异，而不是把其中之一神圣化。若两个方法不一致，应当检查误差预算、样本选择和共同假设；若两个方法一致，还需要判断这种一致是否部分来自共享校准。独立性是一项需要说明的关系，不能只从团队名称不同推断出来。

52.3 逆问题中的先验

视差与距离的关系在理想几何中非常简单：选定适当单位后，距离与视差互为倒数。困难不在这个几何关系本身，而在观测得到的是带误差的视差估计。非线性变换会改变概率分布的形状，尤其当误差相对于视差很大时，用一个带噪声的小数直接取倒数，往往不能产生可靠距离。数学公式正确，不等于对随机估计进行同样代入就获得正确推断。

负视差最能揭示两者的区别。它不是恒星位于负的物理距离，也不说明整个测量失去意义。在真实视差很小、噪声又不可忽略时，无约束估计可能分布在零点两侧。若只因为负值“不像物理量”便全部删除，样本会被有方向地截断。保留原始估计及其误差结构，反而使后续推断能够更诚实地处理弱信息。

Bailer-Jones 等人针对 Gaia DR2 的距离估计工作，明确区分视差反演与统计推断，并用具有空间变化的弱距离先验处理大量源。[52-4] 这里的先验不是让研究者随意决定恒星在哪里，而是把观测无法单独确定的范围限制和群体知识写入概率模型。它所提供的信息必须公开，并允许研究者检查结果对其选择的敏感程度。

作为说明，可以写出后验关系 $p(r|\hat{\varpi}) \propto p(\hat{\varpi}|r)p(r)$，其中 r 是正距离，$\hat{\varpi}$是观测视差，似然描述观测误差，先验描述距离的可接受分布。若有零点偏差，似然的中心也必须相应处理。这个公式不是某个发布产品的完整模型；它展示的是距离结论同时依赖数据和推断条件。任何单一距离数值都压缩了这种依赖，使用时需要恢复其不确定性。

高信噪比时，似然可能强烈集中，合理先验之间的差别对结果影响较小；低信噪比时，先验可能发挥更明显作用。这种差异应当被报告，而不是被统一的目录格式遮蔽。两个源都拥有一个距离栏目，并不意味着它们的距离由同等强度的几何信息支撑。对于群体研究，忽视这种区别，可能把先验结构误认为新发现的银河结构。

这还涉及个体估计与总体推断的关系。若每颗星先以同一个群体模型获得距离，再用这些距离重新“发现”该群体模型，论证可能部分循环。更合适的办法可能是直接在总体模型中联合处理观测和选择，或明确检验不同先验下结论是否稳定。具体方法取决于研究目标，但原则相同：不能在第一阶段注入的信息，在第二阶段又被当作独立证据计算一次。

先验的必要性并不消除自然约束。强数据可以使许多先验受到限制，新的观测也可以否定旧群体模型。反过来，拒绝明确写出先验并不意味着没有预设；直接截断负值、只取高信

噪比源或者规定固定距离上限，也会改变结果。区别在于这些判断是否可见、可比较和可修订。透明的假设比隐藏的假设更适合接受检验。

从 SDE 看，距离 S 并不是从仪器流入表格的现成属性。差异序列 D 通过误差模型、正距离条件和群体结构 E 被组织成估计；估计又可能指导下一次对象选择或模型修订。这里讲的是科学属性的成立过程，并不意味着恒星的物理位置由统计程序创造。它意味着我们实际拥有的可使用距离，是一种包含几何、观测与推断关系的复合成果。

这一区分也让“新的天文学世界”获得具体内容。旧研究能够稳定比较的可能只是天球位置；新的联合系统使位置、运动、距离与群体关系进入同一分析空间。对象获得了可共同约束的属性联系，研究者因而能够提出先前无法以同等方式提出的问题。新世界的发生不需要诉诸凭空制造天体，而体现为新的可辨认整体与可检验操作的成立。

52.4 后续科学如何回写

星表发布不是研究链条的终点。用户把参数用于星团成员识别、银河运动研究、恒星性质推断和距离标定，也会发现某些对象或区域表现异常。异常有时指向真实的天体结构，有时指向选择和校准问题。它们能够成为新一轮研究的入口，但必须保留两种可能，不能把与预期不同的结果立即解释为自然新现象，也不能把所有异常都清洗掉。

以星团为例，成员可能在位置与运动上形成关联。若先根据某种运动模型严格筛选成员，再以所选成员证明该运动模型，便需要注意循环。另一方面，若完全不使用任何群体信息，个体噪声又可能妨碍识别。实际任务是把成员概率、观测误差和群体模型的依赖关系明确化，检验不同假设下群体结构是否稳定。星表提供了新的材料，但不替研究者完成全部判断。

标准烛光校准也如此。较好的视差可以改善某些附近标尺的绝对尺度，但最终对遥远对象的推断还要经过类型、环境和其他观测条件。某个近端误差减小，不意味着整条距离链所有误差同比缩小。新的精度可能恰好暴露旧时被噪声掩盖的差异。研究发展因此不仅表现为误差条变短，也表现为误差来源重新排序。

这里应当严格区分技术迭代与历史回写。在固定观测模型内，当前源参数参与下一轮残差计算，属于算法的正常循环。若外部科学应用促成新的质量建议、校准模型或对象处理策略，才涉及研究制度和知识条件的改变。前一循环可以通过算法文档说明，后一改变则需要发布说明、验证论文或实际规范修订的证据。不能从“算法在循环”直接推出“共同体必然在自我革新”。

同样，E=H(S,D)并不是允许每一次困难都靠调整条件来消除。若校准模型不断增加自由度，只为了让结果贴合预期，它可能吸收真实信号。条件修订必须由独立材料约束，或者对

未参与调整的数据作出更好的判断。能够降低残差只是必要线索之一；能否保留真实变化、能否改善外部一致性和能否经受不同样本检查，同样重要。

Gaia 因而呈现一种富有张力的认识形态：源参数需要共同系统才能高精度成立，共同系统又要借助适当的源和外部参照接受检验。相互依赖并不必然造成恶性循环，因为不同观测具有不同约束方向，新的资料能够排除某些共同错误。但若所有检验都沿同一个退化方向变化，内部一致就不能提供足够保证。可靠性来自约束的互补，而不只是关联数量增加。

可以从这一点理解统一性、多样性和和谐性的具体含义。统一性表现为公共参考框架与可比较参数；多样性表现为不同源、不同观测条件和不同验证材料保留各自信息；和谐性表现为这些差异能够相互约束，而非被强行修整成一致。没有统一框架，资料难以合并；抹去多样性，系统可能失去识别偏差的能力；没有恰当的相互约束，统一也可能只是把共同错误扩散到更多对象。

将这些要求带入教学，会改变学习星表的方式。学生不仅应学会查询一列视差和画出距离分布，还应比较不同质量切割、不同距离估计和不同零点处理所产生的变化。这样形成的能力不是怀疑所有数据，而是把一个结论拆成可以检验的依赖。科学知识的可靠性，也不再被误解为永远无需重开的封闭成品。

对研究共同体而言，最有价值的回写未必是提出新天体类别。一次校准修订、一项可靠的质量说明或者一份保留歧义的关联表，也可能扩大后续研究的真实能力。这些工作有时缺乏戏剧性，却使更宏大的银河叙述获得承重基础。思想史若只记录最终的新图景，而忽视这些条件的重建，就会把科学创新描写得过于轻巧。

这样的世界也具有使用边界。若使用者需要研究某个特殊源，就应检查该源是否属于标准模型能够良好描述的范围；若研究的是群体结构，则应检查选择与共同误差。参数已经公开，并不意味着后续责任可以全部交回数据生产者。可靠知识必须在每一次新的应用中重新说明其成立条件。

本章支持的结论因此有明确边界。Gaia 的天体测量展示了精密属性如何在观测与联合约束中成立，并展示了参数产品为何必须携带质量、版本与推断条件。它没有单独证明一种普遍本体论，也没有保证所有后续科学都正确。它为 SDE 提供的是一个可以细查的案例：一个新的可操作天文学世界，怎样由源、仪器、差异运行和参考结构共同形成，又怎样保留继续被修订的入口。

本章参考文献

[52-1] Lindegren, L. et al. The astrometric core solution for the Gaia mission: Overview of models, algorithms and software implementation. 2012. https://arxiv.org/abs/1112.4139

[52-2] Gaia Collaboration. Gaia DR2 Documentation, Astrometry: Overview. https://gea.esac.esa.int/archive/documentation/GDR2/Data_processing/chap_cu3ast/sec_cu3ast_intro/ssec_cu3ast_intro_overview.html

[52-3] Lindegren, L. et al. Gaia Early Data Release 3: Parallax bias versus magnitude, colour, and position. 2021. https://arxiv.org/abs/2012.01742

[52-4] Bailer-Jones, C. A. L. et al. Estimating distances from parallaxes IV: Distances to 1.33 billion stars in Gaia Data Release 2. 2018. https://arxiv.org/abs/1804.10121

[52-5] Lindegren, L. et al. Gaia Early Data Release 3: The astrometric solution. 2021. https://arxiv.org/abs/2012.03380

[52-6] Gaia Collaboration. Gaia EDR3 data model: gaia_source, ruwe. https://gea.esac.esa.int/archive/documentation/GEDR3/Gaia_archive/chap_datamodel/sec_dm_main_tables/ssec_dm_gaia_source.html

第 53 章　模拟宇宙与统计模型：模型结果能成为何种证据

53.1　从物理方程到数值粒子

计算机屏幕上出现细丝、空洞和聚集的亮点时，观看者很容易把它理解为宇宙的缩影。模拟图像的视觉说服力，来自它能够把庞大的时空演化压缩为一段可以观看的过程。但图像首先是某个模型在数值条件下的结果。它要成为关于宇宙的证据，还必须经过一道不能省略的转换：哪些计算量对应观测量，哪些相似只是显示方式造成，哪些差别足以挑战模型?

N 体模拟中的粒子通常是数值表示单位，不应直接理解为一颗恒星或一个基本粒子。它们承载质量与运动信息，用有限计算资源近似某种连续或高维的物理分布。分辨率决定哪些尺度上的结构能够被解析，体积决定哪些大尺度模式能够被纳入。增大体积与提高细节之间存在成本权衡，因此模拟并不是一个无条件包含全部宇宙过程的微型世界。

数值方法也有自己的误差。时间步长、力计算、边界条件与初始条件都会影响结果。一个结构在屏幕上看起来稳定，并不意味着它已经摆脱数值设置的影响。研究者需要改变分辨率或计算方案，考察目标统计量是否趋于稳定。这里的收敛不是所有细节逐点相同；复杂系统中微小初始差异可能造成局部差别。关键是所宣称的群体性质或物理关系，是否在允许的数值变化下保持可比较性。

EAGLE 项目把引力、流体演化和未解析过程的参数化连接起来，研究星系及其环境的形成。[53-1] 这一安排说明，模拟输出不是单一方程未经中介的直接展开。它包括理论选择、数值实现以及有限尺度上的近似。认识这些中介并不会贬低模拟，反而帮助明确它为何有能力回答某类问题，同时无力回答另一些问题。

可以用一个思想实验说明分辨率的作用。假设两个模拟使用相同的大尺度初始分布，却以不同细度表示气体。若某种局部高密度过程只在较细模拟中得到解析，二者就不能仅凭“采用相同物理”被视为完全等价。粗模拟可能需要用有效规则补偿遗漏过程，而该规则的参数不必在提高分辨率后仍然保持同样意义。模型名称相同，并不保证实际运算对象相同。

因此，SDE 中的显露 S 在这里应指明是计算输出，而不是已经确认的自然结构。D 是数值演化与参数变化形成的差异序列，E 是方程、解析尺度和表示规则的相互规定。先区分计算世界与观测世界，才能讨论二者怎样建立有效对应。如果直接把模拟中的结构称作宇宙真实结构，后面的比较便会在开始之前失去检验意义。

模拟最重要的认识贡献之一，是让一些隐含前提能够被共同运行。一个关于反馈、冷却或质量聚集的设想，在口头上可能与其他设想都相容；进入同一计算系统后，它会对多个结果产生关联影响。这个关联使模型承担更多约束。模拟因而可以比单独叙述更严格，但前提是研究者不把每一种不符都交给新的自由参数处理。

53.2 亚网格模型为何必要

星系研究涉及跨度巨大的尺度。一套覆盖广阔宇宙体积的计算，通常无法同时细致解析每个恒星形成区域、每次爆发现象及其与周围气体的全部作用。因此，亚网格模型用可计算的有效规则表示那些没有直接解析的过程。它不是对无知的任意粉饰，而是有限分辨率下不可避免的建模选择；它的科学性取决于依据、约束和可检验范围。

但“物理上有依据”不等于“参数已经唯一确定”。不同有效规则可能都具有合理动机，也可能在某些观测量上产生相近结果。Crain 等人对 EAGLE 亚网格校准与模型变体的研究说明，仅匹配低红移恒星质量函数，并不足以保证得到合适的星系；其他性质能提供额外限制。[53-2] 这提醒我们，一个成功指标能够约束模型，却不必完全识别机制。

设想两种反馈方案都得到相近的恒星总质量。一种可能较早抑制气体形成恒星，另一种可能允许早期形成较多恒星而在后期改变气体供应。总量相近并不意味着历史相同，也不保证空间分布、化学组成和随时间的变化相同。若只考察最终总量，两种机制便难以区分；加入不同时期或不同性质的观测，才可能削减这种退化。

这里需要区分校准目标与外部检验目标。某个观测量已经用于选择参数，随后模拟与该量相符，主要说明校准实现了预定目标；它不能以同样权重再算作独立预测成功。未参与校准的关系若也被合理重现，则提供更强的外部支持。但“未参与”必须按实际研究过程判断，不能只根据论文最后呈现的章节顺序判断。

参数变化研究也有两种不同用途。一种是寻找更好的拟合，另一种是分析结果对机制的敏感性。前者回答“怎样使输出更接近资料”，后者回答“哪一个过程控制什么变化”。两者可以合作，却不能混同。一个经过多次试调的模型，可能预测表现很好，但研究者仍需解释其参数是否具备稳定的物理意义，以及多个参数之间是否相互补偿。

从发生学角度看，E 的纠缠在这里表现得很具体：分辨率改变有效过程的作用范围，校准目标改变可接受参数，参数又改变输出中可被当作星系的结构。若这些变化重新规定了研究对象，就不能只把它们记为后台技术设置。模拟中的星系身份需要明确识别算法、边界和属性提取规则，才可能与观测中的星系建立公平比较。

这也意味着不同模拟之间的争论不能只看图像美观或单一拟合优度。研究者应说明它们共享哪些物理假设、使用哪些不同有效规则，以及哪些观测量真正有能力区分它们。多种模

型并存可以是认识尚未充分约束的表现，也可以是检验稳健性的资源。强行要求所有模型立即汇入一种表示，并不必然增加可靠性。

53.3 模拟与观测公平比较

模拟知道自己内部的全部物理变量，观测只能通过有限信号推断属性。若直接把模拟里的“真实恒星质量”与观测依据光谱和颜色推断的恒星质量比较，就可能把估计方法的差异误当成物理模型的差异。公平比较要求模拟也经过类似的观测过程，或者明确建模两种量之间的转换。

模拟观测的基本思想，是先把模型生成的对象转化为可被仪器记录的信号，再施加相应的误差、分辨率、天空覆盖和选择规则，最后使用与真实数据相容的分析办法。这样得到的模拟目录，才与观测目录处在更接近的证据层次。这里没有要求复制所有技术细节，而是要求保留对当前结论有重要影响的环节。

例如，若研究某类低表面亮度结构，观测的背景扣除与检测阈值可能十分关键。模拟内存在这类结构，并不意味着观测一定能够识别它；观测没有报告，也不能直接证明模拟过量产生。必须问：如果这些结构按照模型存在，经过实际检测链以后，应该有多少被保留？这个问题把抽象的“看起来不一样”转化为可以进行概率判断的差异。

EAGLE 公开数据库提供从不同体积、分辨率和物理模型的模拟提取的星系与晕属性，使研究者能够比较这些条件下的输出。[53-3] 公开结果扩大了复查机会，但目录中的方便栏目仍然需要定义。不同孔径内的质量、不同边界内的气体、不同时间平均下的形成率，都可能具有不同数值。比较以前应当先比较量的含义。

统计摘要也是一种选择。两份目录可能有相近的一维分布，却具有不同的相关结构；也可能在某个相关函数上相似，而在稀有尾部明显不同。选择哪种摘要，取决于要检验的物理主张。若模型只是为某一摘要调参成功，不能把它扩展为“整个星系形成理论得到证明”。相反，一个未匹配的摘要也不能无条件推翻所有相关物理，需要定位差异来自哪一层。

作为现代示意，可以把比较链写成物理参数产生模拟状态，模拟状态经过观测算子产生合成数据，合成数据再经过与观测一致的统计程序产生可比较结果。每一步都有潜在不确定性。前一步的真实变量与后一步的估计量不应同名而不加区分，否则方法误差容易在命名中消失。

从漂亮输出 S 进入分析，容易问“它像不像宇宙”；从观测过程 D 进入，则会问“若模型为真，我们实际会得到怎样的记录”。后一个问题更能暴露选择效应，也更适合设置反证。这是六路径比较在本章的有限增量：改变入口，使视觉相似不再主导判断，而让预测记录与实际记录之间的差异承担主要检验责任。

但是，观测算子也不能成为保护模型的万能缓冲层。如果每次不符都被归给未精确建模的探测过程，而这些调整又没有独立约束，模型就会获得不合理的免检地位。应当事先说明哪些观测条件已知、哪些可以变化以及允许变化的范围。公平比较意味着双方都接受约束，不是替其中一方寻找无限解释空间。

53.4 负结果的意义

模拟与数据不一致时，最有价值的问题不是谁输谁赢，而是哪种主张受到限制。代码实现有错误、分辨率不足、有效模型不适用、初始条件不同以及基本物理假设有问题，是不同层级的可能原因。它们不能仅靠一张差异图自动区分。负结果只有与诊断过程结合，才能推动知识改变。

一种有力的诊断是保持主要条件不变，只改变具有明确物理意义的过程，观察目标差异是否随之改变。另一种是使用未参与调参的数据，检查修改能否同时解释新旧现象。如果改善一个指标必然恶化另一个具有独立支持的指标，研究者便需要面对真实的模型张力。把所有指标都放进校准目标，有时可以获得更好的总体拟合，却也可能减少独立检验的空间。

研究记录必须保留失败版本。只发布最终表现最好的运行，会使读者看不到参数搜索范围，也难以评估结果是否经过大量尝试才获得。保留哪些修改没有帮助、哪些修改引入新问题，能够让后续研究避免重复，也能防止把探索中的偶然成功误写成事先预测。负结果不是论文正文之外的废料，而是识别机制边界的材料。

在三原理的讨论中，模拟差异可以构成张力，但张力不会自动产生有效创新。它可能促成新的数值方法，也可能促成更合理的观测比较；还可能只是让团队增加参数而未增加区分力。只有当改变使独立检验更有约束，或明确排除了一个原先可行的解释时，才有理由把它称为生产性发展。单纯让图像更像预期，不足以承担这一判断。

本章也提供一个对 SDE 自身的限制。若 SDE 只说模拟结果由运行和条件共同产生，它与一般建模常识没有明显区别。要获得额外价值，必须具体追踪某项输出怎样改变后续比较、怎样修订模型条件，以及这些修订在哪些独立材料上得到支持。若传统数值分析和模型选择已经完成这些工作，应承认 SDE 主要贡献了跨环节整理，而非新的物理机制。

还应区分数值验证与物理验证。数值验证问程序是否正确实现了指定方程，并在适当测试中达到预期精度；物理验证问这些方程与有效规则是否足以描述所研究现象。程序可以正确求解一个不适用的模型，也可以因为数值误差偶然补偿物理缺陷而得到看似良好的结果。把两种验证分开，才能避免把运行成功直接等同于理论成功。

模拟中的因果干预也有明确边界。关闭某项反馈，观察星系性质改变，能够说明在该模型结构中这一过程发挥怎样的作用。但它不能单独证明真实宇宙中的同名过程具有相同效

应。模型中的干预结果需要通过观测后果与其他约束取得现实支持。模拟提供的是受控可能性空间，其说服力来自这个空间与实际证据的连接强度。

另一方面，模型并不因为不完整就失去价值。一个有意忽略复杂细节的简化模拟，可能更清楚地揭示某种大尺度机制；一个包含更多过程的模拟，则可能适合检验多种观测量的联合相容性。复杂度与解释力没有简单的单调关系。选择模型时，应当先说明要回答的问题，再决定需要保留哪些过程，而不是把计算规模当成唯一质量指标。

开放模拟数据也会形成新的回写。研究者能够提出原设计者未预设的问题，对同一输出采取不同识别算法，或将它送入新的模拟观测流程。若这些研究发现某类统计结论强烈依赖对象边界，下一轮模拟发布就可能需要同时提供多种定义。此时变化的不是自然方程，而是共同体认识模拟对象的方式及其公共条件。这样的具体修订，才是可以追踪的历史回写。

最后，负结果需要保留尺度范围。某一解析尺度内的失败，不必否定模型在更大尺度上的有效性；某一时期的成功，也不能直接延伸到全部宇宙历史。适用域越清楚，失败越能提供定位信息。没有边界的宏大主张容易被一切支持，也容易被一切反驳，反而难以成为有效研究的对象。

模拟宇宙因而具有双重地位。它是研究者实际构成的、能够运行和修改的理念与技术复合体；它又必须凭借观测接受关于现实宇宙的约束。这两种地位不能互相替代。模拟的自由使我们能够探索未曾观测的可能，观测的抵抗则使这种探索不至于变成任意想象。新的天文学世界就在这类可运行设想与可反驳记录之间逐步形成。

本章参考文献

[53-1] Schaye, J. et al. The EAGLE project: simulating the evolution and assembly of galaxies and their environments. 2015. https://academic.oup.com/mnras/article/446/1/521/1316115

[53-2] Crain, R. A. et al. The EAGLE simulations of galaxy formation: calibration of subgrid physics and model variations. 2015. https://arxiv.org/abs/1501.01311

[53-3] McAlpine, S. et al. The EAGLE simulations of galaxy formation: public release of halo and galaxy catalogues. 2016. https://arxiv.org/abs/1510.01320

第 54 章　AI 巡天候选：发现、分类还是制造标签

54.1　候选标签从哪里来

一个天文警报到达计算系统时，首先不是一个已经命名的宇宙事件。它可能是一组亮度变化、一幅差分图像、一段位置记录，以及若干关于观测质量的说明。分类器在这些材料上给出分数或标签，帮助研究者决定接下来做什么。这个标签有时极其有用，却不能独自承担“新天体已被确认”的全部意义。

训练标签来自过去的工作。光谱确认、人工审查、历史目录和既有分类规则，都可能成为模型学习的依据。因此，AI 并不是从毫无条件的天空中自行提取所有类别；它继承了观测共同体已经形成的判断，也继承这些判断的覆盖边界。容易取得光谱的亮源、常见类型和受到重点关注的事件，更可能拥有可靠标签。训练集并不天然代表未来进入系统的全部对象。

ALeRCE 实时图像分类器使用科学图像、参考图像、差分图像及警报元数据，进行包括超新星候选等类别的快速识别。其研究同时安排专家选择与后续光谱确认环节。[54-1] 这个实际流程显示，机器分类与科学确认可以相互协作；快速标签的价值是为稀缺后续观测争取时间，而不是消除后续观测的必要性。

“制造标签”也应作两层理解。所有分类程序都会生成数据库标签，这是一项正常操作；有问题的是把标签的生成误认成对象物理身份已经成立。模型可能极其稳定地把某种图像模式标成某类事件，但如果模式与该类事件的联系在新条件下失效，标签仍可稳定地产生。稳定输出本身不能证明稳定的自然归属。

标签还可能隐藏类别的层级。一个源被判断为“真实变化”与被判断为“某类超新星”，需要不同证据；被判断为“值得优先后随”与被判断为“物理上罕见”，也不是同一结论。若系统把这些层次混在一个分数中，使用者就难以理解它究竟在优化什么。高分可以意味着容易识别，也可以意味着值得研究，两者的差异会影响资源分配。

Fink 文档把原始警报信息、交叉关联与机器学习产生的附加信息区分开来，并在缺乏额外判断时保留未知状态。[54-2] “未知”在这里具有积极意义。它使系统不必为每个源强行分配一个熟悉名字，也让未来的新资料能够继续改变解释。一个不能表达不知道的分类器，往往会把新现象压进旧类别，或者把它们当成无关异常排除。

从 SDE 看，S 是暂时成立的候选身份及其分数，D 是图像比较、光变累积和跨目录核验，E 是类别规则、训练材料、误差特征与后随资源的相互规定。这种定位要求我们问：标签以后怎样改变观测？如果它只是一个无人使用的输出，回写很弱；如果它决定望远镜转向和训练样本更新，标签就进入了一个能够改变知识生产的循环。

这个循环也改变主体能力的构成。研究者不再逐一目视全部警报，而通过模型组织注意力。其判断力并未简单转移为一个网络的准确率；它分布在标签定义、训练选择、阈值设定、审计和后随解释之中。只有这些环节能够互相校验，快速机器分类才会成为可靠的复合研究能力。

54.2 在线发现与漂移

离线测试通常把已经收集的资料划成训练集与测试集，衡量模型对未参与训练样本的表现。这是必要检验，但时间域巡天面对的是持续变化的资料流。仪器状态、观测条件、处理版本乃至进入研究视野的对象类型都可能变化。过去分布上的良好表现，不自动成为未来分布上的保证。

分布偏移有不同形式。输入图像的噪声或分辨率改变，可能使模型不再识别原来的模式；各类事件的出现比例改变，可能使相同评分对应不同的实际可靠性；训练中没有充分出现的类型进入系统，模型可能仍给出熟悉类别的高分。它们都不能仅靠总体准确率解释。研究者需要区分输入变化、类别比例变化与类别边界变化。

一个简化的数值思想实验能够说明基率的影响。假设某类事件在一批资料中极少，即使分类器对普通样本只有很低的误报率，误报总数仍可能超过真实目标数。因此，在平衡测试集上获得良好准确率，并不直接说明实际警报流中高分候选多数为真。这个例子不用于估计某个真实系统的性能，而是说明评价指标必须与部署环境连接。

在线系统还面临时间信息逐步展开的问题。第一次检测时，只能依据有限图像与历史记录给出判断；数日以后，光变形状可能提供更强约束。早期标签与晚期标签不一致，并不一定表示早期算法失败，可能只是可用信息增加。合理的系统应当记录判断时点与资料范围，使人能够区分“当时证据下的合理候选”与“已被充分确认的类别”。

这种时间结构对训练也很重要。如果模型训练时使用完整光变，而实际部署时要求在最早时刻决策，测试便可能高估可用能力。未来信息不能悄悄进入过去的判断。对早期发现系统的评价，应当模拟实际可得数据，并把延迟、缺测和观测节律纳入。否则，模型看似提前识别，其实只是利用了部署时尚不存在的信息。

异常检测与分类又有不同目标。分类器倾向于把对象归入已有类别，异常检测则寻找与已知分布不相容的模式。但异常不等于新物理。数据损坏、处理伪影和罕见观测条件，都能产生异常。真正有价值的流程需要把数据质量检查与物理新颖性评估连接起来，同时保留人工能够检查的证据，而不是用一个不可解释的异常分数结束研究。

Fink 提出以模块化方式处理警报，结合交叉匹配、分类和后随选择，说明在线研究本来就是多个操作的联接。[54-3] 本章据此提出的进一步要求，是在系统更新时记录分布与性能

变化。它是一项研究设计建议，而不是对所有现有平台已经实现完备审计的断言。是否有效，需要实际部署资料检验。

因此，在线可靠性不能只在模型发布之日测一次。它需要持续比较不同时间、亮度、天空区域和对象类型的表现。总体稳定可能掩盖某个小类别的退化；某类性能下降也不必要求全部模型废弃。精确定位变化，才能决定重新训练、改动阈值、增加人工复查还是调整观测策略。

还需要区别排序能力与概率校准。模型能够把多数真实目标排在伪影前面，并不意味着其输出的某个数值可以直接解释成确认概率。一个系统可能排序良好但过度自信，也可能总体校准不错却在少见类别上失准。若资源调度依据分数作精细权衡，使用者就需要知道它的统计含义，而不能只因输出落在零与一之间便把它当成客观概率。

这种问题在跨平台使用时更加突出。某分类器在一个巡天训练，再应用到另一个巡天，输入栏位即使名称相同，也可能对应不同滤光片、误差定义或图像处理。格式兼容不等于物理可比。正确迁移需要检查信号形成过程，而不是只把数据整理成模型接受的数组。E 的改变在这里不是外部背景轻微变化，而可能重新规定输入特征与类别之间的联系。

需要特别保留的是标签修正的方向。如果一类对象总是在获得更高质量资料后从普通类别转为特殊类别，模型可能正在系统性压低该类的早期可见性；如果大量高分特殊候选最终回到普通类别，则可能存在某种稳定的混淆。两种情况要求不同干预。只报告总错误数，会丢失这种能够改变下一轮研究的信息。

这种分方向审计还能帮助决定何时让模型拒绝分类。拒绝不是效率损失的同义词；当误分类会误导昂贵后随或污染关键样本时，保留未定状态可能更有价值。但拒绝范围过大，又会把重要对象留在系统外。因此，拒绝机制也必须有公开评价，说明它减少了什么错误、遗漏了什么机会。

54.3 后随观测的资源分配

警报可以大量产生，望远镜时段却有限。候选评分于是进入资源分配：哪些对象值得立即获得光谱，哪些应继续积累测光，哪些暂时搁置。这个决定既包含科学价值，也包含时间窗口、可见性和仪器适配。一个分数不可能无条件代表所有目标的优先次序。

例如，容易确认的常见事件有利于建立高质量统计样本，难以分类的异常则可能更有探索价值。若只追求单位时段确认数量，资源会倾向前者；若只追求最大不确定性，又可能花费过多时间处理伪影或无法解释的微弱源。有效调度必须明确当前任务是完善群体统计、捕捉早期阶段，还是探索未知类别，并允许这些任务之间存在合理分工。

后随选择还会改变未来的训练材料。高分对象获得更多光谱，因而拥有更多可靠标签；低分对象缺乏确认，继续留在训练集之外。下一代模型便更熟悉高分区域。这样形成的循环能够提高已知类型的效率，也可能强化最初的盲区。其危险不在于循环本身，而在于系统把自己选择出来的样本误当成天空的自然总体。

可以把这一机制概括为候选显露 S 改变观测 D，后随结果与标签积累再改变训练条件 E。三方程在这里不是三句平行口号，而是一个可以记录的依赖链。若要判断回写是否改善研究，需要检查新训练集究竟增加了哪些先前缺少的信息，而不能只看标签数量增长。重复确认熟悉区域，可能让模型更精确，却没有让它更开放。

一种可能的改进是在常规高分选择之外，预留少量随机审计或分层探索样本。其目的不是假定低分对象更重要，而是估计被忽略区域中还包含什么。另一种办法是主动学习，根据模型不确定性与观测成本选择最能改善分类边界的对象。Fink 过滤文档列有针对特定分类分数区域的主动学习选择，表明这种思路已有具体工具入口。[54-4]

但主动学习也需要验证。如果不确定性主要来自不可消除的噪声，反复观测这些源可能收益很低；如果模型对陌生对象表现为错误的高置信度，单靠不确定性选择又可能错过它们。因此，随机审计、物理规则、多个模型分歧和专家判断可以承担互补作用。没有一种入口能够自动包办全部未知。

公平比较还应包括简单规则与人工基线。一个复杂模型若只在离线指标上更好，却没有提高后随资源的实际产出，或使稀有类型更难被保留，其部署价值需要重新评估。这里的产出也不能只有发现数量，还应考虑确认速度、可复核性、群体偏差和新问题的产生能力。优化目标的选择，本身构成研究方向的一部分。

更深的变化在于，候选集合不再是观测之后被动形成的列表。它反过来参与决定天空接下来怎样被观察。研究对象与研究能力因此共同发展：某类对象越容易被确认，相关模型和仪器程序就越成熟；另一些对象可能长期停留在没有足够证据的边缘。思想史必须记录这种不对称，才能解释为何技术增长未必让所有问题同步前进。

资源选择还可以用信息价值来理解。一次光谱观测未必确认一个最稀有事件，却可能排除某个导致大量误报的模式，从而改善后面许多决策。相反，重复确认一个已经非常熟悉的高分候选，可能增加样本数量但几乎不改变分类边界。因此，单次观测的价值有直接确认与系统学习两个方面，二者需要在项目目标中明确，而不能以一个统一分数掩盖。

不过，信息价值本身也是模型估计。系统可能认为某个对象能帮助区分类别，实际观测却因信噪比不足没有提供有效信息。因而调度模型也应接受事后评估：预期的信息增益是否实现，哪些成本被低估，哪些类别长期得不到机会。这使后随策略从不可见的经验决定，转变为能够被比较和修订的研究对象。

没有完成确认的候选也应保留适当状态。它可能是假阳性，也可能是由于天气、可见性或资源限制而失去验证窗口。把所有未确认对象都当作错误标签，会把观测失败与分类失败混在一起；把它们都当作潜在真发现，又会夸大成果。对原因进行分层记录，才能计算有意义的性能指标，并理解知识为什么没有在某些方向继续发生。

54.4 证据责任

一个机器给出的类别，可以由另一机器再检查，但只有当第二次检查提供了不同约束，才具有相应的独立性。把同一个模型重复运行，或让另一个模型读取同样的标签后复述结论，不构成新的天文证据。独立性来自数据、测量通道、模型假设或误差结构的实质差异，而不是程序数量。

证据责任首先要求保留来源。候选的原始警报、处理版本、模型版本、评分时间和后随记录，应当能够连接起来。若后来标签改变，需要知道改变来自新资料还是算法修订。这样既能避免把合理的早期不确定性当成错误，也能识别系统性误报。撤回或改类不是应被隐藏的丑闻，而是知识状态变化的正常记录。

人员与机器的分工也应具体。自动系统可以承担大量筛查、关联和排序，人类专家则需要对确认标准、异常处理和公开主张负责。但这不意味着每条记录都必须由人逐像素审查。责任可以通过抽样验证、独立测试和有针对性的复核实现。关键是能否说明整个复合程序如何捕捉重要错误，而不是仪式性地要求人最后点击一次确认。

公开候选尤其需要措辞分级。被算法列为超新星候选、具有支持某类超新星的光变、获得光谱分类，是不同证据状态。读者和后随团队需要知道目前处在哪一层。若传播时把它们都压成“AI 发现某天体”，公众会误以为命名、检测和确认发生在同一步，也会低估大量不具戏剧性的验证工作。

同样，训练数据的改进应当保留版本关系。若新确认结果推翻了旧标签，更新训练集是合理的；但评价历史模型时仍需知道当时使用了什么资料。否则，今天修订后的标签会被倒置成过去已经拥有的知识，使模型改进看起来比实际更顺畅。版本史不仅服务软件管理，也服务科学史对创新和纠错的解释。

本书据此提出一个可以实际检验的程序：在部署前固定主要评价指标，在运行中保留分层审计样本，在后随后分别统计真实确认、无法判断和明确排除，再检查模型更新是否改善预定指标。这个方案目前是方法建议，不能被写成已经验证的成绩。若与简单基线相比没有改善，就应缩小关于 SDE 分析带来增益的主张。

这项限制对于 AI 研究特别重要。一个理论框架可以帮助提出正确问题，却不因此自动提供更好的分类器；一个分类器可以提高效率，却不因此证明该框架的全部哲学命题。能力

发生的不同层次需要分别建立证据。只有把方法建议、工程实现和科学成果的支持程度分开，跨层合作才能在保持创造性的同时维持清楚的责任。

本章的核心结论是，AI 可以改变天文学对象发生的速度、规模与注意力分布，但它的能力必须在完整的确认链中评价。标签 S 由差异 D 和训练条件 E 形成，随后又改变观测与知识条件。这个循环可能扩大未知的入口，也可能把既有偏好自动化。判断两种结果的依据，应是独立确认、遗漏审计与可追踪回写，而不是“用了 AI”这一事实。

因此，统一性应由公共证据标准保障，多样性应由不同分类入口与探索样本保留，和谐性则体现在速度、可靠性和新颖性之间可以解释的配置。三者没有一个固定最优比例。研究系统需要根据实际任务持续调整，并让调整后的结果接受检验。能够承认标签只是暂时判断，反而是 AI 参与科学发现的成熟条件。

本章参考文献

[54-1] Carrasco-Davis, R. et al. Alert Classification for the ALeRCE Broker System: The Real-time Stamp Classifier. 2021. https://authors.library.caltech.edu/records/wghve-frq51

[54-2] Fink. Alert classification documentation. https://doc.ztf.fink-broker.org/broker/classification/

[54-3] Möller, A. et al. Fink, a new generation of broker for the LSST community. 2021. https://academic.oup.com/mnras/article/501/3/3272/5992334

[54-4] Fink. Filters documentation. https://doc.ztf.fink-broker.org/broker/filters/

第 55 章　宇宙学张力：共享误差与理论重开的判据

55.1　“张力”的统计条件

宇宙学中的两个数值不同，并不立即构成理论危机。首先要问，它们估计的是否是同一个量，采用怎样的模型，误差区间表示什么，以及是否共享资料或校准。只有这些条件得到说明，数值差异才具有明确的统计意义。否则，“不一致”可能只是把不同问题的答案放在同一行比较。

哈勃常数提供了一个重要例子。基于近邻距离标尺和超新星的估计，与根据宇宙微波背景及指定宇宙模型推断的估计，虽然都以同一参数名称报告，却通过不同依赖链成立。后一方法不是直接在早期宇宙放置一把测速尺，而是由早期信号约束模型，再推断今天的参数。模型条件因此必须跟随数值一起出现。

Planck 2018 参数论文在基准平直 ΛCDM 模型下给出哈勃常数约为 67.4、标准差约为 0.5 千米每秒每百万秒差距；SH0ES 团队 2022 年的综合距离阶梯测量给出更高的局部估计。[55-1][55-2] 这些固定版本的结果构成重要研究背景。本章以截至 2026 年 10 月 16 日所核对的具名论文为讨论材料，不把其中任何一个数值宣称为全部当代研究的最终裁决。

通常所说的若干标准差张力，需要附带统计模型。若两项估计近似正态且独立，差值的方差可以由两者方差相加得到；若存在相关性，还必须包含协方差。这个现代说明式揭示，误差条不是可以脱离产生过程随意拼接的装饰。同一数据在不同分析中多次出现，会使简单的独立比较产生误导。

参数空间的维度也重要。两个模型对一个参数给出相近结果，不保证它们对其他参数或联合分布相容；某个一维边缘分布出现差异，也可能需要在完整退化方向中理解。显著性不是从图上两条误差棒是否相交得到的唯一答案。比较方法应当与实际主张一致，并说明是否预先选择了观察到差异最大的指标。

张力的科学意义不取决于是否立刻推翻模型。它首先表示，现有测量、误差说明与理论关系之间存在需要解释的困难。困难可能来自尚未充分处理的系统误差，也可能指向模型不完整，还可能包含统计波动。承认这些分支，不是消解问题，而是把“令人兴奋的差异”转化为可以工作的研究程序。

从 SDE 看，S 是带条件的参数估计，D 是比较、残差分析与重算，E 是参数定义、校准关系和模型前提。若 E 尚未对齐，D 中出现的差异不能直接被称为自然反常。反过来，如果经过不同路径复查仍保留差异，研究者就有理由扩大模型检验范围。张力是否具有生产力，要看它是否增加了可区分问题，而不是是否赢得一个更响亮的名称。

差值显著性还受误差估计是否完整的影响。统计误差通常来自有限样本和测量噪声，系统误差则涉及校准、模型近似及选择。二者并不是天然可以完全分离的实体：某个校准参数可以作为随机变量进入联合模型，某种样本效应也可能随数据增加而暴露。关键是误差预算是否覆盖主张所依赖的实际变化，而不是它被放在哪一个名称下面。

统计比较还应清楚说明异常是在分析之前指定，还是在许多结果中事后挑出的最大差异。两种情形的证据强度并不相同。探索发现可以引导后续研究，却需要独立资料或预先规定的检验来提高支持程度。忽略这一点，容易让偶然波动获得过大的理论负担。

因此，扩大样本之后张力增强，有两种不同解释。一种是随机误差缩小，使真实差异更清楚；另一种是共同偏差仍然存在，而误差模型没有充分表示它。仅凭显著性上升无法区分二者。新的数据应尽量改变约束方向，例如使用不同锚点、不同天体类型或不同物理通道，而不是只沿同一链条积累更多记录。

55.2 共享系统误差如何识别

多个团队、多个望远镜与多篇论文，并不自动等于多个独立证据。不同研究可能使用相同的几何锚点、相同的超新星样本或相同的校准关系。一个共同偏移可以沿这些路径进入不同结果，使它们相互接近；某个局部处理差异也可能让使用相近资料的结果发生分歧。理解这些关系，比简单统计支持某个数值的论文数量更重要。

2024 年发布、后有修订的 CCHP 研究以 JWST 观测比较红巨星分支顶端、JAGB 星和造父变星等距离方法，并利用 NGC 4258 几何距离设置零点。[55-3] 这里的“多方法”确实提供不同天体物理约束，但共同锚点也意味着某些误差可能相关。不同方法的价值与共享依赖可以同时存在，不能因为发现共享项就否定全部互证，也不能因为方法名称不同就忽略共同项。

Riess 等人的另一项 JWST 比较强调，样本选择和共同超新星的处理会影响不同局部哈勃常数估计的比较，并检查 HST 与 JWST 距离的一致性。[55-4] 两项研究的存在说明，争论不能压缩为“新望远镜支持哪一方”。望远镜改善了某些观测条件，但最终参数仍经过样本、标尺和统计程序。技术更新没有取消推断结构。

共享误差的识别可以从依赖图开始。每个最终参数回溯到使用的校准源、观测样本、标准化关系与处理版本，再标明哪些节点共同。图中不同的路径如果在较早节点汇合，就不能被当成完全独立的整条链。真正新增的约束来自那些拥有不同误差来源的环节，它们能够检查共同链条的某一部分，而不必独立验证全部。

选择函数也进入这里。较近、较亮或较容易取得高质量资料的宿主样本，可能与完整样本具有不同性质。比较两个子样本时，需要估计这种选择造成的预期变化。样本较小带来的

误差增大，可能使张力的显著性下降，但这不同于张力的中心差异已被物理解释。不能把“目前无法拒绝一致”改写成“已经证明完全一致”。

系统误差也不是任何不舒服差异的通用名称。提出某项误差，需要说明其机制、方向、大小及可检验后果。若怀疑拥挤影响测光，就应利用分辨率变化或环境比较检查相应趋势；若怀疑零点，就应寻找不同锚点或独立校准。只有能够产生具体预测的误差解释，才与新物理解释处于可以公平比较的位置。

这一程序体现了从 S→D→E 到 E→D→S 的路径转换。前一路径由参数差异引出调查，后一路径先审查共同条件再估计差异。二者可能给出不同判断：原来看似独立的支持被降级，或者原来被怀疑的误差在专门检验后受到限制。路径转换的价值在于这些明确变化，而不是每种箭头都出现了一次。

55.3 新物理假说如何可证伪

如果已知误差不足以解释差异，扩展宇宙学模型是合理方向。但“可以调到目标数值”只是起点。一个新增成分或新的演化规律，会同时改变其他可观测量；真正有力的假说必须承受这些关联后果。若只在一个参数上改善拟合，却破坏其他独立观测的描述，它的总体地位未必提高。

新增自由度通常能够扩大可接受范围。这种扩大可能反映真实物理可能性，也可能只是降低模型的约束性。因此，比较时应当关注新增参数的代价、先验范围及对未参与调整资料的预测能力。把误差区间放大到与另一结果重叠，不能直接叫作解释张力。解释应说明为什么原来会出现那个方向和规模的差异。

一个可工作的假说，应当提出当前资料之外的辨别条件。例如，它若改变早期演化，就需要说明对相关早期信号和后续结构有哪些关联影响；若修改晚期动力学，则应接受不同红移范围内距离与增长信息的联合约束。这里不预先选定某一种方案，而强调新物理不能只在提出它的那一个数字附近活动。

也需要保留“部分缓解”的地位。某方案可能减少一项差异，却没有成为所有资料下的最佳描述；某个分析可能显示在特定先验下可行，换用其他合理条件后则不稳定。这些结果仍有研究价值。清楚标明它们的条件，比在“解决”与“失败”之间作过早二选一，更有利于后续积累。

三原理在这里可作为动力问题的组织方式。已有参数显露与模型约束之间的张力，可能推动新的比较程序，也可能推动新的理论设定。但并非任何模型改动都构成生产性回写。如果增加自由度只是使所有结果都可以被容纳，却减少可反驳预测，改变反而削弱了研究的辨别力。动力的方向需要用结果检验，不能由“发生了改变”自动赋予正面价值。

公平的竞争应当允许多个分支同时接受约束。系统误差方案不能永远以“还有未知偏差”保护旧模型，新物理方案也不能永远以“参数还没调好”保护新模型。双方都应提出在什么观测下自己的解释会受损。科学进展有时来自选择一个胜者，也有时来自发现原有二分不够准确，例如某个样本问题与某种理论局限同时存在。

对 SDE 而言，同样的要求必须适用。不能把任何张力出现、任何模型修订或任何长期未决，都解释成框架必然正确。如果框架声称某类路径转换能够识别遗漏的依赖，就应展示实际识别了什么，并与已有统计和科学方法比较。若没有额外收获，应承认此处主要发挥整理作用。这种限制使方法有机会从一般语言进入可评估实践。

假说检验还需要避免事后挑选。在看见某个数据组合最有利于新模型后，才将它确定为主要证据，会高估支持力度。探索性分析当然允许发现有趣模式，但应把它与预先指定的验证区分。理想情况下，研究者在模型形成后提出可区分预测，再等待或寻找没有参与构思的资料检验它。

如果新资料不支持预测，修订仍然可以合理。问题在于修订是否保留了失败记录，是否增加新的独立约束，以及是否承认原先承诺没有实现。一个不断改变说法却不记录旧版本的模型，会让读者无法判断它到底预测了什么。理论的开放性不是可以永远撤销检验标准，而是允许以可追踪方式修改标准和主张。

同一要求也适用于误差解释。假如某项偏差原先被认为足以消除张力，专门观测却显示它太小，研究者应降低这项解释的权重。不能只因仍可能存在别的误差，就把已受限制的具体机制继续当作充分答案。对新物理保持谨慎，与对旧解释施加同等严格的检验，可以同时成立。

55.4 科学传播的责任

宇宙学争论具有强烈传播吸引力。两个数值不一致，可以被写成宇宙年龄错了、标准理论崩溃了，或者新望远镜解决了一切。这些句式把条件性的研究结果改写成整体性结论，常常跨过了最重要的论证步骤。公众因而看到的是连续革命，却难以理解每次变化究竟修改了什么。

较准确的表达应当保留三个层次：某组资料在某种分析下显示差异；若干误差解释已受到怎样的检验；哪些理论扩展仍然有待区分。这样的表达并不必然枯燥。真正值得说明的，是为什么不同观测链可以如此精确地约束同一个宇宙，又为什么小的残余差异足以驱动新的研究。问题的深度不需要靠宣布全部知识失效来表现。

截止日期与版本也属于传播责任。当一个研究领域持续更新，引用某年论文应说明这是当时的结果，而不是默认它永久代表最新共识。本章因此以具名来源讨论推断结构，不给出

一个假装涵盖全部 2026 年成果的总裁决。未来若新资料改变某个数值，其方法论意义仍需按独立性、选择和预测约束重新评价。

开放资料与公开代码可以降低复查障碍，但复现并不只是重跑一次得到相同输出。更重要的是理解哪些假设使输出成立，并检验合理变化会怎样影响结论。严格的复现能够确认计算链，也能够发现方法敏感性；两者都应公开。一个结果可以在计算上完全可复现，同时在模型解释上仍存在争议。

张力还具有社会时间。研究者需要在不完整资料下决定观测计划、发表阶段结果和分配资源。过早宣布解决，可能压缩其他路线；过度渲染危机，也可能把注意力从基础校准转向更有吸引力的理论故事。成熟共同体需要允许稳健的维护工作与大胆的替代方案同时存在，并以可比较证据评估它们。

天文学思想史由此不应把未决问题当成尚未写完的胜利章节。长期争论可能表明约束还不够，也可能表明不同方法正在逐渐分离此前混合的误差。它的价值在于问题结构变得更清楚：哪些量共享祖先，哪些资料提供独立方向，哪些理论付出额外代价。这样的清楚，本身就是知识的发展，但它必须与最终解决区别开来。

可以把整个讨论理解为三种责任的结合。测量团队负责说明资料如何形成和校准，统计分析负责说明不确定性怎样传播，理论解释负责说明参数为何能在同一模型中连接。任何一方都无法单独承担全部结论。把争论归结为某一个团队是否可信，会把方法问题人格化，也会遮蔽各环节之间真正可检查的依赖。

这也是为什么数据公开、方法复核与理论比较应当共同推进。

这种分工并不削弱综合判断。恰恰相反，只有每一层主张都清楚，研究者才能判断差异应该在哪一层处理。一个新的测量可能修正参数而保留理论，一个新的理论也可能重释相同记录而不改变原始信号。思想史需要记录变化发生在哪里，才能解释何种意义上的世界改变了，何种层次仍然连续。

未决状态还可能具有建设性秩序。共同体可以暂时保留多个估计与模型，并围绕最有区分力的观测合作。这种秩序既不同于所有观点同样有效，也不同于立刻选出唯一胜者；它按现有证据给不同主张不同权重，同时保留修订入口。在复杂科学中，这往往比仓促一致更接近真正的和谐。

新的天文学世界并不要求所有参数立即和谐一致。它要求差异能够在公共规则下被保留、定位和检验。统一性是共同讨论同一问题的条件，多样性是揭露共同盲区的资源，和谐性则是让不一致转化为有方向的调查，而不是把它消音。张力因此既不是失败的羞耻，也不是创新已经完成的证明，而是下一轮严格研究需要承担的任务。

本章参考文献

[55-1] Planck Collaboration. Planck 2018 results. VI. Cosmological parameters. https://arxiv.org/abs/1807.06209

[55-2] Riess, A. G. et al. A Comprehensive Measurement of the Local Value of the Hubble Constant with 1 km/s/Mpc Uncertainty from the Hubble Space Telescope and the SH0ES Team. 2022. https://arxiv.org/abs/2112.04510

[55-3] Freedman, W. L. et al. Status Report on the Chicago-Carnegie Hubble Program: Three Independent Astrophysical Determinations of the Hubble Constant Using the James Webb Space Telescope. 2024 预印本及 2025 修订。https://arxiv.org/abs/2408.06153

[55-4] Riess, A. G. et al. JWST Validates HST Distance Measurements: Selection of Supernova Subsample Explains Differences in JWST Estimates of Local H0. 2024. https://arxiv.org/abs/2408.11770

第 12 编

天文学思想发生的跨时代检验

第 56 章　天体身份的四种发生：观测、推断、确认、再分类

第 57 章　宇宙距离与时间：两种不可直接取得的尺度

第 58 章　用三方程比较五次天文转折

第 59 章　六路径与三原理：创新为何有时发生、有时受阻

第 60 章　新的天文学世界：已成能力、盲区与开放条件

第 56 章　天体身份的四种发生：观测、推断、确认、再分类

56.1　对象标识与自然事件区分

天文学史常用“发现”把性质不同的工作连接起来：记录一个新亮点，推断一个不可见质量，确认某类事件，或者重新定义一个既有对象。这种简洁在叙事中方便，却会遮蔽创新究竟发生在哪里。本章提出四种身份操作的区分：观测使某种信号成为可重复追踪的对象，推断使一组信号与物理模型建立联系，确认使竞争解释受到足够限制，再分类则改变对象在知识体系中的位置。

四种操作不是必须依次经过的固定阶段。一个对象可以先由理论推断而后获得观测，一个已经确认的对象也可以重新分类；某种新观测可能同时改变分类与确认标准。区分的目的不是制作统一流水线，而是避免把不同证据责任混成一句“人类终于看见了”。科学身份的形成，常常跨越多个操作并包含反复修订。

自然事件与科学身份必须分层。一次恒星爆发现象的物理过程，不以研究者是否及时命名为条件；而“这是某类爆发”的公共判断，需要信号、比较标准和排除程序。两者有关，却不是同一个发生。SDE 若讨论新的天文学世界，应当说明由何种显露与属性联系构成了新的可研究整体，不能把知识身份的成立直接移作天体物理形成的原因。

对象标识也不是物理类别。一个坐标编号能够帮助不同研究者指向同一记录，却不保证该记录的解释已经确定。随着资料增加，同一个源可以从未定状态进入某类别，也可能因分辨率提高被拆成多个源。名称的连续与对象构成的连续，因此需要分别考察。数据库保存旧名，并不能证明其对应的科学单位从未变化。

这一点在现代巡天特别清楚。一次检测身份由图像分离或时间序列分析形成，跨观测身份由匹配建立，物理身份则可能需要光谱与模型。三者往往在成品目录中合并为一行，使使用者误以为它们天然一致。思想史的任务是把这些被封装的操作重新展开，识别创新是增加了记录，改进了对应，还是重构了类别。

因此，本章采用一个强弱有别的世界构成判据。只改变名字而不改变可检验属性与操作，是较弱变化；改变对象的比较条件和推断联系，是较强变化；进一步使新的观测组合、问题和公共规则持续成立，则形成更广泛的系统重构。这个判据是本书的分析建议，实际案例需要分别检验，不能因为词语宏大便自动归入最强等级。

四种操作还可通过它们最容易失败的位置加以区分。观测身份可能因伪影或错误关联而失败；推断身份可能因多个物理模型给出相近信号而不唯一；确认可能因所谓独立证据其实共享错误而被削弱；再分类则可能因为任务不清或边界规则不适用而失去效力。失败位置不同，所需补救也不同。增加一次曝光未必解决类别定义争议，修改名称也不能补足一个信号的统计显著性。

这个区分使同一事件可以得到分项评价。研究者可以肯定检测可靠，同时保留物理机制未决；可以接受某个机制具有很强支持，同时认为类别名称仍有讨论空间。分项评价避免了全盘接受或全盘拒绝的两极化，也允许不同专业在自己能够提供约束的层次开展合作。

56.2 从图像到间接波形

“直接看见”在日常语言中似乎最可靠，但科学图像同样经过信号形成、仪器响应和重建。EHT 的 M87 成像研究利用不同成像办法与测试检查稳定结构。[56-1] 因而图像中的稳健特征与某一种重建所呈现的细节，不应获得完全相同的证据地位。可见性不是无需推断，而是推断以一种便于感知的形式组织起来。

引力波事件提供另一种显露。探测器记录的不是日常意义的天体照片，而是与时空扰动有关的时间序列；解释依赖仪器验证、噪声分析及物理波形比较。2016 年发表的首次双黑洞并合引力波观测论文，将信号检出与源参数解释连接起来。[56-2] 这种对象身份并不因为缺少普通图像而天然较弱，它需要按自身证据链评价。

图像与波形的差异，也不能简单译为一个直接、一个间接。图像具有空间布局优势，却可能受采样覆盖和重建先验限制；波形具有时间演化信息，却可能存在参数退化。若二者指向不同属性，比较谁“更真实”便失去具体意义。真正需要问的是，每个通道独立约束了什么，哪些结论仍然依赖共享模型。

跨通道身份更进一步。多个信号被认为来自同一事件，需要时间、方向及物理相容性的支持。它们并列出现，并不自动构成共同来源；而成功关联以后，单一通道难以确定的某些属性可能获得更强约束。新对象并非几份记录的机械相加，而是这些记录之间建立了能够相互限制的联系。

这使 E 的含义得到清楚展示。E 不是“有多个仪器”的背景，而是不同通道的误差区域、时间关系和物理模型如何共同规定归属。若位置不相容，即使时间接近也可能不应关联；若一条通道缺失，也要考虑其检测能力，而不能立刻否定事件。这些条件共同决定“同一事件”何时具有足够支持。

确认的程度因问题而异。证据可能足以确认存在一个天体事件，却不足以唯一决定其所有内部机制；可能支持一种源类别，却不能精确测出每个参数。把确认理解为所有疑问同时结束，会使科学主张在过强与过弱之间摇摆。更精确的做法，是把确定性分配给具体命题。

从这一比较得到的新综合，不是所有观测最终都可以归为同一流程，而是确认依赖的约束方向具有多样性。新的天文学世界扩展了什么能够作为显露，随之改变哪些属性能够相互联系。其统一性来自可共享的检验规则，而非要求所有对象都以同一种感官形态出现。

一个有用的对比是“同名异构”。两套目录都把某类记录称作源，但其中一套的源由空间亮斑分离，另一套由时间变化检测形成。名称相同，纳入条件与边界却不同。反过来，不同历史名称也可能在特定操作中指向相近对象。对象连续性需要检查标识与属性对应，而不能从词语连续直接推断。

对于综合史研究，这意味着比较表的每一行都要固定对象单位。不能在一例中把 S 设为一个脉冲信号，在另一例中设为整个制度，而后宣称它们具有完全相同的发生模式。可以先比较局部身份形成，再研究它如何扩展为共同体变化；每次扩大尺度都需要新增证据。层级纪律使宏观论断获得可追踪的基础。

56.3 分类标准何时产生回写

2006 年 IAU 关于太阳系行星的决议，把绕太阳运行、近似流体静力平衡形状和轨道邻域条件纳入定义，并据此区分行星、矮行星等类别。[56-3] 这是一项明确的分类规范改变。它没有在决议通过时改变相关天体的质量或轨道，却改变了研究与教育中哪些属性被用来组织类别。

这个例子说明，科学分类既受物理差异约束，也包含研究目标和命名规则。若主要关心内部结构，某些对象可以形成一组；若关心轨道动力学地位，分组边界可能不同。不同分类未必都在竞争同一个任务。争论的第一步，应当说清分类希望表达什么联系，而不是只争夺一个词的唯一合法使用。

与之相比，快速脉动射电源的发现涉及新的信号类别。1968 年的原始报告展示了迅速而规则的脉冲现象，并讨论可能的物理解释。[56-4] 信号识别与具体机制解释并未在逻辑上合为一步。这里的分类形成不是先由规则投票决定一种自然记录应当存在，而是异常信号迫使研究者建立新的比较对象，再寻找能够解释其属性联系的模型。

两类变化都可能产生回写，但路径不同。规范分类可以改变目录、教学和比较分组；新信号类别可以促成新的搜索节律、探测程序和理论问题。不能因为都出现新名称，便把它们归为同一种革命。回写的强度需要检查哪些后续操作确实改变，以及改变是否有持续记录。

某些分类调整的科学收益可能有限。一个更方便的名称能够改善交流，却不一定提高预测能力；一次对象重分可以使目录更整齐，却未必产生新的测量。这样的变化仍有价值，但应按实际功能评价。反过来，一个没有引起公众注意的质量分类，若明显改变了参数可靠性，也可能比一个醒目名称更深刻地影响研究。

在 SDE 框架中，这一区分阻止了 S 的泛化。若 S 有时指名称、有时指物理参数、有时又指整套研究系统，那么任何改变都能被叫作显露创新。严格比较需要固定本次研究的 S：讨论命名，就以分类规则及使用证据为对象；讨论新属性，就以观测与推断联系为对象。跨层变化可以存在，但桥接必须另外说明。

分类的生产性还取决于能否容纳边界对象。若规则要求每个对象立即进入唯一类别，而实际资料不足，错误的确定性就会扩散。允许候选、混合或未定状态，可以保留后续研究空间。分类体系的成熟，并不表现为没有剩余，而表现为知道剩余为什么存在，以及需要什么资料才能减少它。

判断复合通道是否产生较强世界构成，还可以问一个反事实问题：移去其中一个通道后，哪些属性关系不再能够同样成立？如果移去只是降低精度，而主要类别与操作保持不变，改变可能属于增强；如果移去使事件归属或某个关键物理联系无法建立，则组合具有更强的构成意义。这个判据不要求所有通道地位相同，而是识别它们各自不可替代的约束。

同样，确认标准的改变需要检验是否实际改变了对象集合。若新标准只把旧做法写得更清楚，它主要改善透明度；若它使过去接受的候选受到排除，或使过去无法确认的类别获得可靠路径，就改变了知识的边界。规范的发布与规范的实际作用仍然是两件事，需要使用记录把它们连接起来。

56.4 最强反例

对世界构成论最有力的限制之一，是稳定物理性质与持续分类争议可以同时存在。人们能够对某天体的轨道、形状或光谱取得较强一致，却仍争论它应属于哪一类别。这说明类别统一并不是全部科学知识成立的必要条件，也说明分类争议不能被直接推断成自然对象没有稳定约束。

另一类限制来自普通的增量发现。某个新源完全符合已有类别与验证程序，其加入目录增加了样本，却没有改变研究机制。若把每条新记录都称作一个全新世界，概念就失去区分力。更合理的说法是，既有世界的对象集合获得扩展，而构成它的主要规则暂时保持连续。

还有一种反例是更名而无新证据。共同体可以出于交流便利修改术语，相关观测、推断和预测并未改变。这类事件支持知识具有制度维度，却不能单独支持强意义的科学世界重构。承认这一边界，反而使真正改变属性联系的案例更清楚。

还存在一种重要的不支持情形：新仪器提高了信号质量，却没有改变既有身份判断或后续研究程序。它证明技术能力改善，不能单凭这一点证明强意义的世界构成。相反，某次看似微小的证据变化若迫使共同体重新区分两类对象，其理论意义可能更大。把技术规模与概念变化分别评价，有助于避免以设备壮观程度代替思想史判断。这个限制也适用于本书自己的案例选择，不能只挑最著名、最容易呈现为革命的事件。

因此，四种身份操作之间没有统一的价值排序。一次观测可能极具创新，也可能只是重复；一次再分类可能深刻改变问题，也可能只调整词汇；一次理论推断可以开启新领域，也可能长期缺乏确认。评价还应保留明确的比较任务。评价应当落在新增约束、可复查程度和后续操作改变上，而不是落在“发现”“理论”或“命名”这些名称上。

本章暂时支持的综合判断是：天体科学身份具有多种发生方式，它们共同依赖自然信号与研究操作，却并不共享一套不可变顺序。SDE 的适用价值在于追踪显露、差异和属性联系如何相互规定；其边界在于不能从这种相互规定直接推出所有实体由研究者任意创造。保留这个边界，才能让新的天文学世界既具有发生性，又承受来自实际观测的限制。

本章参考文献

[56-1] Event Horizon Telescope Collaboration. First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole. 2019. https://arxiv.org/abs/1906.11241

[56-2] LIGO Scientific Collaboration and Virgo Collaboration. Observation of Gravitational Waves from a Binary Black Hole Merger. 2016. https://arxiv.org/abs/1602.03837

[56-3] IAU. IAU 2006 General Assembly: Result of the IAU Resolution votes. https://www.iau.org/IAU/Iau/News/PR2006/iau-2006-general-assembly-resolution-votes.aspx

[56-4] Hewish, A.; Bell, S. J.; Pilkington, J. D. H.; Scott, P. F.; Collins, R. A. Observation of a Rapidly Pulsating Radio Source. Nature 217, 709–713, 1968. https://www.nature.com/articles/217709a0

第 57 章　宇宙距离与时间：两种不可直接取得的尺度

57.1　从角度到几何距离

天文学的尺度并不是一把逐渐加长的尺。附近恒星的视差、变星的光度关系、遥远星系的红移和宇宙年龄，分别通过不同的操作成立。它们能够连接，是因为研究者建立了校准、类型对应与理论关系；它们不能任意互换，也正因为各自依赖的条件不同。本章不再重述每一种方法的发现，而比较尺度怎样获得跨层迁移能力。

几何视差具有一个看似直接的基础：观测位置改变时，近源相对于远方参照的方向发生变化。已知基线与角度关系能够约束距离。但实际使用还需要分离自行、参考框架变化及仪器误差。所谓几何方法，不意味着不需要模型，而表示其核心距离联系以几何约束为主，较少依赖某些天体内禀光度假设。

第 52 章讨论的 Gaia 联合解算表明，视差参数必须与姿态和校准共同成立。[57-1] 因此，一组恒星具有各自的误差条，并不保证这些误差彼此独立。若共享零点偏移，它会同时影响许多后续标尺。距离阶梯的依赖分析应当从这一点开始：个体数值进入下一级校准时，哪些误差可以平均，哪些误差以共同项继续传播？

这里可以用校准祖先概念组织比较。某项遥远距离估计的祖先，包括用于设定绝对尺度的几何对象、处理这些对象的观测与校准，以及连接到后续对象的关系。两个最终估计即使使用不同远端样本，只要共享关键祖先，就存在相关依赖。祖先不是历史荣誉名单，而是误差与尺度能够沿之传递的结构。

这种结构不必是一条单线。多个几何锚点可以共同限制一个标准关系，某个关系也可以服务多个远端样本。校准网络比“阶梯”隐喻更能表现这种交叉。它的可靠性不仅取决于每条边的精度，还取决于是否存在真正不同的约束方向。许多通向同一祖先的支路，未必比少量具有独立来源的路径更有检验力。

以 SDE 语言表达，某级距离 S 成为下一级研究的条件 E，同时规定新的差异比较 D。原来作为待估结果的量，后来被当作校准输入使用。这个角色转换是尺度扩展的关键，却也容易隐藏不确定性。一个结果进入标准工具以后，不能因为使用频繁就失去原有误差与适用范围。封装应当保留必要的接口，使下游研究可以重新打开关键依赖。

依赖网络还包含方向。某个几何距离用于校准光度关系，不等于光度关系已经独立验证了这个几何距离。如果随后再用同一批资料反推几何结果，必须检查是否只是把输入重新取回。循环可以成为联合估计的一部分，但其信息量不能按照经过几次运算就增加几次。信息来源与计算路径应当分开计算。

一张校准图也不只是误差传播图。它同时记录对象的适用域：某个标尺对哪些类型有效，某种关系在哪个波段建立，某次修正依赖哪些环境信息。若新对象越出适用域，即使各级数值误差很小，整条推断仍可能失效。准确计算一个不适用的关系，不能产生可靠尺度。

57.2 从光变到宇宙尺度

周期—光度联系能够把容易测量的时间变化与内禀发光能力连接起来，但它首先是一种需要标定和适用类型限定的关系。勒维特 1912 年的材料展示共同系统中周期与亮度的联系；绝对距离尺度需要另外的校准。[57-2] 因此，“发现联系”与“得到完整标准尺”之间存在逻辑距离，不能由后来的用途倒置填平。

现代应用还需处理对象类型、消光、观测波段和环境差异。这里的重要问题不是列出一串修正项，而是判断哪些修正与待估尺度发生退化。例如，观测变暗可以与距离增加有关，也可能与消光有关；若缺乏额外信息，二者难以完全分离。更多波段或其他约束的价值，在于改变这种可辨识性，而不只是增加资料数量。

几何锚点能够限制绝对零点，但无法单独保证一个关系跨所有环境不变。把在某一群体中校准的联系迁移到另一群体，需要检验类型与物理条件的相容性。相同名称之下可能存在不同子类；相同平均关系也可能具有与环境相关的残差。尺度的可传递性因此是一项可以失败的主张，不是校准完成后的自动福利。

NGC 4258 的几何距离研究，为跨方法标定提供了具有明确物理模型的锚点。[57-3] 它的价值在于提供不同于某些传统标尺的约束，但使用它的多个后续方法仍共享这一零点。若比较这些方法的一致性，应当区分共同尺度的支持与各自天体物理关系的支持。它们可以检查彼此的局部环节，却不能把共同祖先的误差自动消除。

一个思想实验有助于说明：假设两种标准烛光都用同一个偏高的距离锚点校准。两种方法在遥远处可能给出相近结果，这种一致支持某些相对关系，却不能单独排除共同尺度偏移。若再加入一个独立几何锚点，结果发生一致移动，就说明此前的相互支持需要重新解释。重点不是一致性没有价值，而是价值应当对应实际检验的环节。

距离网络中最容易被忽略的，是标准化本身。某类对象并非每个都有完全相同的光度，而是通过若干可测属性形成可用关系。标准化参数如何确定、样本怎样选择、散布怎样处理，会影响最终尺度。一个看似自然恒定的标尺，可能实际是一套持续接受修订的复合操作。

从总体数量看，远端资料可以十分丰富，近端锚点却相对有限。这造成一种重要的不对称：远端统计误差能够很小，但整体尺度仍受少数共同条件限制。天文学的扩张并不总是由

最远处缺资料阻碍，有时受限于近端标尺的准确性。新的大望远镜与细致的基础校准，因此不是宏大探索与琐碎维护的对立，而是同一尺度系统的不同承重环节。

这张网络可以按四层阅读。最底层是原始方向、通量或时间记录；第二层是经校准的观测量；第三层是对象模型产生的距离或光度估计；第四层是把多个对象连接起来的宇宙结构推断。相邻层之间都有转换，但并非每个最终结论都以相同方式使用全部层。识别具体路线，才能判断一项新资料到底约束了哪里。

若某项远端结果发生变化，反向检查也应沿这些层进行。先确认观测量是否改变，再检查对象类型与标准化，然后检查共同尺度和宇宙模型。这个次序不是不可变的算法，而是避免跳跃归因的纪律。不同线索可能要求从不同层先入手，但最终应说明为什么某一层足以解释差异，而其他层暂时保留。

57.3 不同宇宙距离的定义

在宇宙学范围，“距离”需要说明定义。由亮度与通量关系定义的光度距离，由实际横向尺度与角大小关系定义的角径距离，以及用于描述共同膨胀坐标分离的共动距离，并不具有无条件相同的数值。Hogg 关于宇宙学距离的技术说明集中呈现了这些量及其联系。[57-4] 使用同一个日常词，不能取消它们不同的操作意义。

在相应几何和光传播假设下，光度距离与角径距离存在标准联系，其中红移因子不可省略。这个联系使不同观测能够互相约束，也提供检验假设的可能。但如果比较时把一个方法报告的光度距离与另一方法报告的角径距离直接相减，得到的差异首先是定义错误，而不是宇宙反常。量的名称、单位与模型条件必须一起核对。

回望时间同样不等于今天的空间分离简单除以光速。信号传播期间宇宙尺度因子可以变化，不同距离概念对应不同计算。红移与时间之间的映射依赖膨胀历史；对膨胀历史作不同假设，可能改变同一红移所对应的回望时间。我们从记录进入宇宙历史，需要一套连接事件与时空结构的模型。

这并不使宇宙年龄成为纯粹约定。模型受到多类观测限制，彼此不相容的时间尺度可以成为检验线索。但年龄数字必须与估计方法共同理解。某个恒星年龄、一个星系的恒星形成历史与宇宙整体年龄，并不是同一种钟。它们涉及不同物理过程和不确定性，能够互相约束，却不能在未作转换以前被视为同一量。

可以把空间与时间尺度的关系理解为一组互相校验的映射。角度、通量、周期和红移是不同显露，几何、发光机制与宇宙模型为它们建立联系。某条映射失效时，研究者需要定位是测量、对象类型还是理论关系出了问题。把全部映射压成“一把宇宙尺”，会使错误归属更加困难。

这种区分也改变了宏观图像的意义。绘制一幅宇宙三维分布图时，每个点的位置可能已经包含红移转换、选择函数与距离模型。图像不是不真实，而是其真实性具有可说明的条件。若模型更新导致位置变化，并不意味着望远镜当时接收到的信号被改写；改变的是信号与空间结构之间的推断关系。

SDE 在这里需要同时保留连续与发生。光子记录及某些测量可能保持连续，空间分布 S 却因 D 的处理与 E 的模型联系改变而重构。若要主张新的科学世界发生，必须指出哪一组属性联系与研究操作随之改变，而不只说数字换了。反过来，也不能把一切归为“同一世界的不同描述”，从而忽视新尺度实际上开放了不同的可检验问题。

另一个需要分开的概念是传播时间与对象年龄。看到很远的星系，意味着收到较早时期发出的信号；但这个星系中某一恒星群的年龄，还需要根据自身演化线索推断。较大的回望时间不等于我们看见了一个同样古老的恒星群。事件的发生时刻、信号的发出时刻与接收时刻，在宇宙学叙述中必须清楚定位。

这种时间纪律能避免把观测窗口误写成全历史。某个红移范围中看见的对象，是在特定选择下保留下来的群体，不是那个时期所有对象的完整名册。用它重建形成历史，需要同时处理未被看见的部分和模型之间的退化。距离与时间从来没有脱离样本选择而独立工作。

57.4 尺度的历史回写

尺度一旦成为公共工具，就会改变研究对象的组织。具有距离的光源可以被赋予绝对光度，运动的角速度可以在条件满足时连接到实际速度，群体的空间分布可以进入结构研究。距离并非在完整对象上追加的一个无关属性；它会重新规定其他属性之间的关系。正因为如此，尺度修订能够在多个领域产生连锁变化。

这种回写有强弱之分。一个零点微调可能使若干参数整体移动，却保留主要理论结构；一次新的独立标尺则可能暴露旧链条的共同偏差，改变验证规则。判断变化的深度，应检查哪些推断因此改变、哪些操作需要重做、哪些假说受到新的限制，而不是只看数字变化的百分比。

引力波标准汽笛提供一种具有不同依赖结构的距离研究路径。GW170817 的标准汽笛哈勃常数研究把引力波距离信息与宿主相关红移信息结合起来。[57-5] 它的意义之一，是不沿用完全相同的传统光度标定链；但它仍有自身的模型、方向与观测不确定性。新的路径可以增加独立约束，却不是无条件免校准的终极尺。

因此，未来尺度改进的重点不只在延长可测范围，也在改变误差相关结构。若一种新方法的精度暂时较低，却能检查旧方法共同无法识别的偏差，它仍可能具有重要价值。反之，

一个精度很高但依赖完全相同祖先的结果，主要增强已有链条，未必提供同等强度的独立检验。

这也说明，尺度网络的质量不能只用节点数量衡量。增加一个独立锚点，可能比增加许多共享同一零点的远端对象更能改变结论。反之，若研究的是某类对象的内部散布，大样本又可能比新锚点更关键。资源选择应当由当前最主要的不确定性决定。认识自己的依赖结构，能够使有限观测投入到真正具有辨别力的位置。

尺度的扩展因此同时是一种测量能力与依赖识别能力的增长，二者需要共同维持。

本章的综合结论是，宇宙尺度的可信性来自可迁移关系与可追溯依赖共同成立。统一尺度使不同观测进入共同问题，多样路径帮助发现共同盲点，和谐连接则要求共享误差不被重复计算。只有保留这张依赖网络，天文学才能在向遥远处扩展的同时，知道自己的每一个宏大数字究竟站在哪里。

本章参考文献

[57-1] Lindegren, L. et al. The astrometric core solution for the Gaia mission. 2012. https://arxiv.org/abs/1112.4139

[57-2] Leavitt, H. S.; Pickering, E. C. Periods of 25 Variable Stars in the Small Magellanic Cloud. 1912. https://articles.adsabs.harvard.edu/pdf/1912HarCi.173....1L

[57-3] Humphreys, E. M. L. et al. Toward a New Geometric Distance to the Active Galaxy NGC 4258. III. Final Results and the Hubble Constant. 2013. https://arxiv.org/abs/1307.6031

[57-4] Hogg, D. W. Distance Measures in Cosmology. 1999. https://ned.ipac.caltech.edu/level5/Hogg/Hogg_contents.html

[57-5] LIGO Scientific Collaboration et al. A gravitational-wave standard siren measurement of the Hubble constant. 2017. https://arxiv.org/abs/1710.05835

第 58 章　用三方程比较五次天文转折

58.1　五个案例为何可比

把托勒密、开普勒、光谱化学、勒维特和 Gaia 放在同一章，并不是说它们采用了相同技术，也不是说现代卫星可以替古代几何证明一套普遍发展规律。可比较之处，在于五例都涉及一个共同问题：某种天文关系如何在有限资料与约束结构中成立，又怎样改变后续研究。比较对象是关系的成立与使用，不是五个人或五套装置的整体相似。

本章固定第 8、18、27、32、52 章为案例来源。托勒密一例关注位置计算与运动规范，开普勒一例关注轨道构造与残差，光谱一例关注实验与恒星成分归属，勒维特一例关注周期与亮度联系，Gaia 一例关注源参数与联合校准。它们的结果 S 不同，差异序列 D 不同，特征纠缠 E 也不同。统一比较必须保留这些差别，而不能把所有结果都改写成一个空泛的“新认识”。

首先需要固定研究层级。这里讨论的是天文学关系与操作的形成，不直接把恒星物理形成、仪器读数与共同体制度放在同一个因果层面。物理过程对信号形成的作用，可以成为模型前提；制度对资料可得性的作用，需要历史材料说明。它们能够进入同一研究，但不能因为都被叫作条件，就获得相同证据地位。

其次需要区分三类联系。构成依赖说明某个量在当前系统中如何被定义和估计；推断支持说明哪些证据使一个判断比竞争判断更可信；历史因果则说明某项成果实际上促成了后来的研究变化。三者可以同时存在，却不能互相替代。一个模型在逻辑上允许某种用途，不等于历史研究者已经采用了该用途。

这一区分使三方程承担不同任务。S=F(D,E)追问显露如何成立；D=G(S,E)追问已有显露如何规定可进行的比较与操作；E=H(S,D)追问结果和差异运行如何重组约束关系。它们在本书中是研究方向，不是已经给出函数形式和参数的自然定律。只有把每一式落实到实际材料，它们才有可评价内容。

公平基线同样必要。传统科学史已经研究观测、数学、仪器与制度的相互作用，现代逆问题理论已经分析参数耦合与可辨识性。SDE 不能把这些成果重新命名后据为独有发现。本章只把能够改变比较判断的部分算作增量，例如识别把关系发现误当绝对校准、把内部迭代误当历史回写，或把预测有效性误当物理解释充分性。

由此形成的比较，不以五例都完美符合三式为成功。某一方向材料不足，应当保留不足；某一式只能提供整理便利，也应如实评价。跨时代统一性来自相同的检验纪律，多样性来自不同机制，和谐性则来自让这些机制相互校正，而不是为得到整齐答案压缩差异。

这里还有一个数学上必须澄清的问题。写出 S=F(D,E)，并不能在一般条件下保证唯一写出 D=G(S,E)或 E=H(S,D)。一个结果可能由多组差异与条件产生，映射可能不可逆。三方程作为方法方向，允许研究者分别追问三者；若要把它们作为严格函数系统，就必须额外说明变量空间、可辨识性与唯一性条件。本书不以符号对称性代替这些证明。

勒维特案例特别适合展示这一点。同一相对周期—亮度关系可以对应不同绝对零点，仅凭关系 S 与部分条件不能唯一恢复距离。Gaia 的参考框架自由度也说明，相同内部观测关系可以容许某些表示变换。两例共同提示，相互规定不等于每个方向都有唯一逆函数。承认多解与退化，才能避免三式在最需要分析的地方预先宣布问题已经解决。

58.2 S＝F(D,E)的研究任务

托勒密体系的 S 不是一张静止宇宙图，而是能够给出特定时刻天体位置的计算关系。《至大论》的几何结构与数表传统，使记录可以进入有组织的预测。[58-1] 其 D 包括时间与位置的比较，E 包括几何构造、参数和运动要求。这里的重要区分是，一套方法可以在位置预测方面有实际效力，同时留下物理解释的代价。单一“正确或错误”判断会丢失这种双重性。

三方程复查在这一例中增加的，不是宣称模型具有条件，而是要求说明成功发生在哪个任务。若角位置预测得到支持，不能自动推出距离变化也同样可靠；若某项运动规范受到批评，也不能抹去计算用途。S 必须按具体量和允许误差限定。只有这样，旧体系的持续使用与改造压力才不再显得互相矛盾。

开普勒案例则把残差与构造限制放到中心。《新天文学》以火星研究展示对既有轨道安排的重作。[58-2] 在综合比较中，最重要的不是把创新压缩为某一次误差触发灵感，而是理解同一残差在不同误差判断与几何要求下具有不同地位。被当作可容忍噪声时，它不要求改变轨道；被认为超过合理误差并具有结构时，它才对模型构成压力。

第 18 章进一步区分经度拟合与距离几何：某种替代构造在有限经度任务上可以很好工作，加入其他约束以后却暴露困难。这个区分使残差不再被误写成一个圆对所有资料都统一偏离的数字。轨迹、时间规则与观察者位置必须共同检验，才能解释为什么局部成功不足以保护完整构造。

因此，新轨道关系不是从数据单独读出的图形，也不是不受资料约束的数学选择。D 与 E 共同限制候选 S。这个判断可由反事实深化：若观测精度不足以区分两种构造，材料便不能承担同样强的选择；若只调整局部参数已经解释全部相关资料，改变轨道形式的必要性也会不同。这里的分析要求保留候选之间真实的区分条件。

光谱化学案例提供另一种机制。实验室材料变化使谱线归属获得约束，发射与吸收的联系又使这些约束能够用于恒星研究。[58-3] S 是经过操作支持的成分判断，而非光谱图上所有线条的直接翻译。D 包含不同材料与状态的比较，E 包含辐射机制、波长尺度和仪器响应。若只把谱线强度等同于丰度，就越过了状态依赖的桥梁。

这一例中，关系迁移比单纯相关更重要。地上与天上不能进行完全相同的操作，研究者必须说明哪些特征在条件改变后仍可用。SDE 复查因而追问稳定性是在哪种变化中建立的，哪些变化超出实验支持。它把“物质指纹”这一方便隐喻重新展开为有适用域的归属方法。

勒维特案例又不同。共同系统中的距离近似相同，使视亮度差更适合表现内禀关系；周期表示将离散光变组织成可比较量。[58-4] S 是周期与亮度联系，E 的关键在于一个共同未知量可以在相对比较中部分消去。但消去共同距离不等于测出了共同距离。此处三方程分析必须明确关系成立与绝对零点成立的不同责任。

Gaia 将相互规定推进到联合参数层次。源、姿态和校准通过大量交叠观测共同估计。[58-5] S 的可信程度取决于误差与约束结构，而非孤立读数数量。其重要机制是不同参数可能解释相似观测变化，因此需要参考条件和互补观测限制退化。这不是古代案例的现代翻版，而是一种具有明确数学形式的耦合成立。

五例对比由此得到一个有限共同点：新关系都不是把预先完备的对象属性简单搬进语言，而是在差异操作与约束联系中获得可使用形式。但这一共同点本身仍很一般。真正有解释力的是五种不同机制：任务分化、残差辨别、跨条件归属、共同量消去和联合估计。若这五种机制被抹平，统一框架反而降低了思想史的分辨率。

对五例还可以分别设置移除条件的比较。移除托勒密模型的时间与参数规定，就无法获得同样的具体预测；降低开普勒案例中的误差辨别能力，候选构造之间的约束会减弱；移除光谱实验的材料控制，线条归属更容易受污染或状态混淆；移除勒维特样本的共同距离近似，视亮度趋势不再具有同等解释；移除 Gaia 的参考与校准约束，某些参数便不能同样稳定地分离。这些反事实不是历史上实际做过的全部实验，而是由各例机制提出的分析检验。

这种比较的收益，是把不可替代条件与可替换条件分开。有些材料可以由另一种等效观测补足，有些条件则需要改变整个问题的可辨识性。若不同技术实现仍然保留相同约束，结果的稳定性可能跨工具迁移；若关键约束丢失，即使保留所有熟悉名词，也不能认为方法原封不动。方法的同一性因此应当落在约束功能，而不仅落在设备或术语上。

58.3 D、E 式的严格核验

第一式较容易在案例中找到对应，因为任何科学结果都依赖资料和条件。真正严格的检验集中在后两个方向。一个结果是否改变了后续差异运行，必须指出改变了哪些问题、选择

或操作；是否重组了 E，则要说明哪些约束、校准或公共规则发生变化。不能仅凭后来的研究时间上排在前面之后，就称为回写。

托勒密模型使计算者能够在特定历元和参数下生成位置，因而为后续比较提供基准。这里 D 受 S 规定的证据，主要存在于模型与表格的使用。至于某次改写是否改变制度和训练，则需要另外的传本、教学或观测记录。模型可传播性与实际传播史是不同命题，不能由逻辑上的可复制性自动推导出历史影响范围。

开普勒的新构造改变了怎样计算轨道位置，也改变了哪些残差值得继续分析。但它在不同研究环境中如何被接受，需要独立材料。发现者完成一种关系，不意味着整个共同体立即采用同一规范。综合叙述应当保留接受迟延、用途分化和部分采用，而不是让科学史在论文发表时突然翻页。

光谱归属建立以后，观测者可以围绕成分和状态提出新的问题。这是研究空间扩展，但要判断某种仪器或分类规范是否因之改变，仍需具体技术史支持。一个常见误区是把后来全部天体物理学回溯为某篇论文的直接后果。较准确的因果叙述需要保留中间模型、仪器改善与新的实验约束。

勒维特的联系为距离研究提供了可用方向，但绝对校准并非该联系的内含结论。D 式在这里要求追踪后来的标定操作，E 式则要求检查标定怎样改变联系的使用范围。不能把全部距离阶梯的成熟反过来写入 1912 年结果。早期成果的力量，恰恰在于它尚未完成所有功能，却能够改变后续研究的目标。

Gaia 则容易发生另一种混淆。算法迭代中当前参数影响残差和下一步更新，是明确的技术循环；科学应用促成质量标准或校准模型修订，则属于历史发展。两者都可以与三式相关，但证据不同。前者有算法定义，后者要查发布与验证记录。如果把数值循环直接当成共同体自我更新的证明，就发生了层级越界。

由此可把五例的支持程度作出文字判别。对 S 在具体 D/E 中成立，五例都有可以讨论的机制；对 S 改变后续操作，模型使用、标尺应用与数据处理提供不同强度的支持；对 E 发生持续重组，则应逐一补足历史材料，不能概括为五例同样充分。这个不对称不是模型失败的羞耻，而是材料与主张需要匹配的结果。

更值得关注的是回写的选择性。一项成果可能改变计算，却不改变基本宇宙观；可能改变分类，却不改变观测设备；可能改变校准规范，却不产生新的自然类别。E 不是一个同时整体翻转的背景。把其内部变化分开，能够解释为什么科学转折既有连续性，又有局部的不连续。

这也提供一项实际的史学增量：在每个案例中区分“成果的成立条件”和“成果后来重组的条件”。两者可能有重叠，但不必相同。实验室使某种谱线归属成立，后来受影响的可

能是天文台的观测计划；同一系统中的共同距离使关系显露，后来被重组的可能是远端尺度校准。追踪这种条件迁移，比笼统说一切互相影响更有解释力。

还必须区分改变方向与改变速度。一个结果可能让既有计算更快，却不改变问题结构；也可能使一种过去无法进行的比较成为可能。两者都属于能力增长，但后者更接近研究空间的扩展。跨时代比较若只统计成果数量，会把效率提升与对象构成混在一起，难以解释某些看似小的概念改动为何具有长远作用。

以光谱归属和 Gaia 作对照，前者强调不同实验条件与远方信号之间的物理桥接，后者强调同一观测网络内部的联合约束。二者都让过去无法取得的属性关系进入操作，但桥梁的性质不同。一个框架若真正尊重多样性，应能解释这种区别为何导致不同的验证方法，而不是只把它们归结为 E 发生变化。

58.4 反例怎样进入理论修改

五个案例都是经过历史筛选的重要成果，本身带有成功者偏差。仅凭它们能够用三方程组织，不能估计框架在所有研究中的有效性。需要加入失败、停滞与无显著变化的案例，检验相似张力为什么没有产生相似结果。否则，框架只是在已知终点后重新画路。

一个直接反例是数据增加而判断不变。若新观测落在原有误差范围内，既有模型继续有效，三式可以描述其稳定运行，却不应被写成创新机制的成功。另一个反例是新显露没有进入后续实践：成果可能正确，却由于资料不足、工具限制或传播条件而没有形成强回写。发生与持续发展需要不同解释。

还有一种需要收缩的命题是“相互依赖必然不可分”。某些研究任务在给定精度下可以合理近似分块处理。Gaia 的全局依赖不意味着每项后续应用都必须重做整套卫星解算；光谱状态依赖也不意味着任何元素识别都需要无穷复杂模型。SDE 若把整体性理解成拒绝有效封装，就会妨碍科学操作。关键是知道封装保留什么、牺牲什么，以及何时需要重新打开。

框架还应容纳零增量情况。如果专业史已准确区分模型任务、传播环节与证据边界，三方程再整理没有改变判断，那么其作用可以是教学与索引，而非新增解释。这并不使框架毫无价值，却限制了原创性主张。方法评价应当区分便于记忆、发现遗漏和产生新预测三种不同收益。

五例还展示不同的误差角色。托勒密问题中的误差限定预测任务的有效域，开普勒问题中的残差能够成为构造修改的压力，光谱中的污染可能破坏归属，勒维特样本中的离散要求检查共同条件，Gaia 中的相关误差则约束参数的公共使用。它们不能被一个抽象的“噪声”概念抹平，因为每种误差所指向的补救不同。

如果跨案例分析只发现都有误差，它没有增加多少理解；若能区分误差何时只是降低精度、何时改变身份、何时暴露模型退化，就获得更具体的研究价值。由此也能说明，新的天文学世界不是在完全消除误差后才出现。它是在能够以更清楚方式组织误差、使某些属性可靠成立时发生，并始终保留进一步修订的余地。

这一比较为后续检验留下一个明确任务：选取不以成功著称的案例，检查相同误差类型是否引出了不同响应，并追查差异来自哪里。若新材料表明所谓共同机制只是成功叙事的选择结果，本章就需要收缩对应概括。统一框架的成熟应当体现为能够据此修改自身。

本章据此提出一个更谨慎的阶段结论：三方程能够作为跨案例检查相互规定与回写的组织工具，五例展示了若干值得继续追踪的机制；但它们尚未构成具有独立函数形式和普遍预测能力的历史定律。其进一步价值，取决于能否在新案例中事先识别重要依赖，或在竞争解释遗漏之处改变实际判断。

阶段性的支持需要随材料质量而调整；它既不等于整体证明，也不等于放弃综合，而是让综合结论具有可修改的明确边界。

统一性、多样性和和谐性在这里获得同一项要求：共同问题不能抹平不同机制，不同机制也不能阻止严格比较。五例最有启发的联系，是它们都迫使研究者说明一个成果怎样取得可靠性，又怎样进入下一轮工作。保留这些具体路径，才能让宏观思想史成为有证据支撑的综合，而不是在时代之间重复同一句哲学判断。

本章参考文献

[58-1] Library of Congress. Mathematics—Rome Reborn: The Vatican Library & Renaissance Culture. https://www.loc.gov/exhibits/vatican/math.html

[58-2] Kepler, J. Astronomia nova. 1609，Smithsonian 数字原件目录；本章综合第 18 章所用原典及重构材料。https://library.si.edu/digital-library/book/astronomianovaa00kepl

[58-3] Kirchhoff, G.; Bunsen, R. Chemical Analysis by Observation of Spectra. 1860，英文转录。https://www.chemteam.info/Chem-History/Kirchhoff-Bunsen-1860.html

[58-4] Leavitt, H. S.; Pickering, E. C. Periods of 25 Variable Stars in the Small Magellanic Cloud. 1912. https://articles.adsabs.harvard.edu/pdf/1912HarCi.173....1L

[58-5] Lindegren, L. et al. The astrometric core solution for the Gaia mission. 2012. https://arxiv.org/abs/1112.4139

第 59 章　六路径与三原理：创新为何有时发生、有时受阻

59.1　六种入口的经验分布

从一个已经成功的发现出发，历史常常显得道路清楚：异常出现，研究者改变设想，新对象得到确认。但在结果尚未确定时，异常可以指向仪器、计算、类别或物理机制，研究者也可能没有足够资料区分这些方向。六路径的价值，应当在这种未定局面中检验，而不是在终点已知以后把经过重新排成整齐箭头。

S、D、E 的六种排列，表达不同研究入口及其连接顺序。从显露进入，可以追问它经过什么差异操作、依赖什么条件；从差异进入，可以比较哪些约束使结果可辨；从条件进入，可以考察现有系统允许什么显露。排列给出的是提问空间，不是六条在历史上同样常见、同样有效的因果道路。

因此，本章不虚构一个覆盖全史的路径频率表。现有案例经过主题选择，古代与现代资料的保存程度不同，研究者的实际思考顺序也不总能从最终论文恢复。把每章标定一个主导路径，最多提供工作性编码；要统计经验分布，需要预先定义样本、编码规则，并由不同研究者检查一致性。没有这些步骤，百分比只会制造量化外观。

海王星案例可以说明历史路径与今日复查的区别。历史研究从天王星运动问题走向扰动解释和可搜索位置，最终观测报告提供新的约束。[59-1] 今天从已经确认的海王星反向检查，可以追问早期预测哪些参数准确、哪些并不充分，以及当时有哪些替代解释。反向路径不会改变当年的事件，却可能纠正后来把预测写成无误推演的叙事。

EHT 案例则适合从图像转向采样与重建。公众先看到环状显露，复查者再问哪些特征受到数据稳健约束、哪些细节依赖算法。[59-2] 如果改从干涉资料与校准开始，研究者可能先得到一个允许图像的集合，再判断共同特征。两种入口的预期差异，不在于一方有图像、一方没有，而在于是否把单幅成品的所有细节都赋予同等可信度。

AI 候选提供第三种路径变化。从高分列表开始，容易评估已选对象的确认率；从训练和调度条件开始，则会注意低分对象没有获得后随而造成的盲区。第 54 章提出的分层审计，正是把入口差异转成可执行检查。其有效性仍须由实际部署结果评价，不能因为提问更全面就宣布性能已经提高。

路径比较的最小成果，应是一项明确改变的判断。某个结论被限制到更小适用域，某个所谓独立证据被识别为共享校准，或某个被排除候选被重新列入调查，都可以记录。若两条

路径没有改变任何实质判断，就应记为零增量。这样，六路径才不会沦为必须全部出现的写作装饰。

六种路径还可能具有不同成本。从既有结果反查条件，常常能够利用现成文献，成本较低；从缺失的观测条件出发设计新调查，则可能需要新的仪器或长期数据。路径在逻辑上对称，现实中的可行性却不对称。若只看到研究者偏爱某个入口，便把它归因于认知习惯，会忽视资源与时间对路线选择的影响。

同一路径也可能承担不同任务。S→D→E 可以用于发现错误，也可以用于合理化既有结果。区别在于分析者是否允许 D 暴露不相容资料，是否允许 E 真的被修订。因此，箭头本身不具有正面价值；它必须与公开的比较规则连接。没有这种规则，六路径既可能扩大视野，也可能成为六种替既定结论寻找理由的方式。

59.2 三相位不能强行等量

本书采用的三原理相位，将 D 与 E 的张力引出 S 改变记为 P1，将 S 与 E 的张力引出 D 改变记为 P2，将 S 与 D 的张力引出 E 改变记为 P3，并要求检查改变之后是否回写。它们不是三种自然力，也不是每次创新必经的三个时间阶段。它们帮助区分受压力的关系与实际被改变的环节。

P1 关注已有差异无法在原约束下形成合适显露。轨道残差与某种几何限制不相容时，新构造可能出现；但也可能通过参数修正或误差重估维持原结构。只有记录实际候选与比较，才能判断改变发生在哪里。不能先把结果指定为新 S，再把所有前史都组织成必然趋向它的张力。

P2 则强调已有显露与解释条件不协调，研究者改变差异运行。例如，一个异常结果暂时不适合现有模型时，可以重新安排观测、寻找不同波段或改变比较样本。这里的创新可能首先是方法与操作，而非新理论。若新操作发现原结果属于伪影，P2 仍然可以产生知识收益；创新不必以增加一个新实体为终点。

P3 关注已有结果与持续差异之间的不相容，促使研究者修改校准、分类或验证结构。Gaia 质量规则、图像稳健性比较和 AI 训练样本审计，都提供可以调查的方向。但只有同一结构内的参数更新，未必足以称作 E 重组。必须说明改变了什么约束，以及为什么原有约束不足。

三种相位在现实研究中可以交织。一个新校准改变参数显露，新的参数又改变观测选择；一个候选模型促成专门测试，测试结果再迫使分类调整。分析者可以分段研究，但不能因为形式上能够切成三段，就宣称历史过程本来如此整齐。切分是一种模型，需要接受材料检验。

宇宙微波背景案例尤其提醒我们，信号身份升级需要不同层次的支持。1965 年的超额天线温度报告与同年的宇宙学解释承担不同任务，后来频谱测量又增加了约束。[59-3][59-4] 不能把排除本地噪声、提出宇宙解释与形成完整验证全部挤入同一瞬间，再给它贴上单一相位标签。相位分析应当帮助展开过程，而不是再次压缩它。

同样，不同相位无需等量分布。现代大型项目保存大量校准文档，P3 更容易得到材料支持；古代传世著作可能主要保存最终构造，使后续条件修订难以恢复。这种差异可能来自档案结构，而不直接反映历史动力的真实频率。研究者需要区分“没有发生”与“目前没有材料证明”。

三原理最有用的地方，是迫使我们把张力与响应分开。同一张力可以产生多种响应，响应也可以失败。D—E 不相容不是新 S 必然出现的充分条件；资源、表示工具和可用假说都可能影响结果。若理论把任何结局都称作三原理实现，就不再具有辨别力。

张力还需要与普通差异区分。两个结果不同，只有在共同目标或约束下不能同时满足某项要求时，才构成这里所说的张力。不同波段显示不同形态，可能完全符合物理预期；两个模型服务不同尺度，也可能没有直接冲突。如果把所有多样性都写成待消除的张力，就会错误地把科学的正常分工作为缺陷。

因此，相位判断应当先写出不能同时满足的条件。例如，某种校准既要保留一组可靠参照，又要解释另一组持续偏差，却在当前参数空间内无法做到；这才形成可以定位的压力。随后还要列出可能响应，包括保留未决、增加资料或缩小适用域。改变并不总意味着扩大理论，有时合理收缩才是知识进步。

59.3 失败案例不可删除

海王星与火神星的对照，要求把“异常意味着新天体”这一成功经验限制在具体证据内。某些残差可以通过新增实体获得有效解释，另一些则可能要求不同的理论改变。水星近日点问题与广义相对论解释提供不同分支的历史材料。[59-5] 本章不把两例压成一成一败的英雄对照，而关心为什么相似的推断形式不能保证同样结果。

失败首先可以是辨别力不足。两个模型在当时误差范围内产生相近预测，更多争论未必能够替代新的资料。此时坚持未定，并不是缺乏创造性，而是对证据结构的恰当响应。若历史叙述把后来正确的一方写成当时已经显然正确，就会误判那些没有接受它的人，也无法解释新观测为何重要。

失败也可以表现为错误稳定。一个分类或校准方案不断选择与自身相容的资料，排除不相容部分，最终得到高度一致的系统。这种一致不是毫无价值，但可能缺少外部约束。AI 高

分候选优先后随、再进入训练的循环，是当代可研究的例子；它说明回写并不总是正向能力增长，也可能把盲区制度化。

第三种失败是局部改善被误写成整体解决。模拟调参改善一个统计量，却使其他量更差；图像算法稳定一个尺度，却不能确定所有细节；新距离方法降低一项误差，却仍共享旧零点。它们不应被简单记为成功或失败，而应标明哪些部分获得支持。局部判断能够积累，宏大二分往往阻断积累。

第四种失败来自传播和使用。某个关系已经具备合理支持，却没有进入可持续操作，可能因为资料难以取得、计算成本过高或缺少公共规范。此时成果的真值与社会效果不能混同。SDE 若强调回写，就需要解释为什么一些结果没有回写，而不是只记录形成制度的成功案例。

第五种失败是新条件制造新的歧义。更高灵敏度可以带来更多候选，也可能增加污染；更多波段可以改善归属，也可能带来复杂交叉匹配；更强模型可以拟合更多细节，也可能扩大退化空间。技术进步并不保证每个中间判断都更简单。创新的收益需要在新的错误结构中重新评价。

这些失败之间不能相互代换。辨别力不足需要新增约束，错误稳定需要打断选择循环，局部改善需要扩大外部检验，传播受阻需要改变使用条件，新歧义则需要重新设计表示与判断。三原理若能够帮助定位应当改变的环节，便具有方法价值；若只把所有失败写成“E 还不够好”，就没有完成解释。

失败档案还应记录没有发生的预期改变。研究者原本预计某项校准会消除偏差，结果没有；预计新波段能够唯一分类，结果仍然退化；预计公开资料会带来独立复核，实际却只有重复调用。这样的记录能够约束关于动力与回写的主张，比事后只保留成功路径更重要。

因此，历史中的创造性不能简单等同于不停改变。某些情况下，坚持可靠约束能够防止伪发现；另一些情况下，过度保护既有条件会压制新关系。判断取决于替代方案和证据，而不是对“变化”或“稳定”赋予固定价值。统一、多样与和谐需要的是可辨别的调节，而非永远增加新解释。

第 30 章所涉及的恒星能源问题也可以从这种角度理解：一种机制能够解释短期发光或提供部分能量，并不保证它满足所需的长期尺度。此处真正的张力要同时列明能量来源、释放速率与持续时间；若只问“能否发光”，就可能过早接受局部答案。本章不替代该章的具体计算，而指出跨案例比较应保留约束目标，否则成功与失败的判准会在叙述中移动。

EHT 的稳健性比较则说明，失败有时属于输出的某个细节。多种重建可以共同支持较粗结构，却对较细亮度分布不完全一致。合理结论应当保留获得支持的部分，并限制细节主

张。若为了得到一幅完全统一图像而隐藏差异，就失去了利用多样性识别不确定性的机会；若因细节不一致否定全部结构，又浪费了已有约束。

AI 案例中的探索审计也不是保证发现新类别的机器。预留低分样本可能发现此前遗漏，也可能主要得到普通对象或噪声。后一个结果同样有价值，因为它限制关于盲区规模的猜测。评价审计不能只奖励戏剧性发现，否则审计本身也会形成新的选择偏差。科学方法需要允许一次严格调查得到“没有发现预期增量”。[59-6]

59.4 有用的方法如何被证伪

评价六路径与三原理，可以从有限研究试验开始。选取一组尚未完成解释的案例，在知道结论以前固定主要问题，让不同分析者分别使用常规方法与 SDE 复查，再比较新增依赖、错误识别和可验证预测。这不是本书已经完成的实验，而是由现有分析提出的后续检验方案。明确这一点，能够把理论愿望与经验成绩分开。

比较指标需要具体。发现更多可能因素并不一定更好，可能只是增加无法检验的复杂性；提出更多解释也不一定提高理解，可能降低区分力。更有意义的结果是识别此前遗漏且实际影响结论的条件，提出后来得到资料支持的辨别观察，或者减少某种明确错误。方法的收益应当对应这些可追踪变化。

盲态编码可以降低成功者偏差。分析者先不看最终结论，只根据当时材料判断哪些方向值得调查，再与后续结果比较。历史案例不能完全消除已有知识污染，但可以尽量选择不熟悉材料，保留原始判断。若只有知道答案以后才能把路径安排得漂亮，就应降低框架的预测性主张。

反证也有层级。某一次路径没有产生增量，不足以否定整个组织框架，却限制其在该任务中的必要性；多次在明确条件下未优于简单基线，则应收缩普遍优势主张；若核心概念无法被不同研究者稳定区分，就需要修改定义。不能把所有反例解释为使用者尚未真正理解理论，那会使方法逃离检验。

同样，P1、P2、P3 的编码需要允许不确定。某次变化究竟主要改变 D 还是 E，可能取决于研究尺度。若不能通过预先定义解决，应保留双重或未定编码，并说明原因。形式整齐不是高质量的替代物。能够承认边界模糊，反而使后续定义改进有真实材料。

方法的和谐性最后体现为与专业工具合作。统计模型负责量化不确定性，仪器知识负责信号机制，科学史负责材料与因果，SDE 则可以组织这些层次之间的相互规定。若它试图以几句总公式替代所有专业工作，便会失去承重基础；若它能够让专业工作之间的接口更清楚，则可能增加综合能力。

实际检验还需要预先规定何为停止。如果已有证据足以支持一个有限命题，继续无止境增加路径并不一定改善结论；如果关键条件仍无法取得，也应允许暂时保留未决，而不是以更多术语填补。方法的效率包含知道何时需要扩展，也包含知道何时应当收束。统一、多样与和谐的平衡由此进入研究时间本身。

对历史材料的不足，亦应采用同样纪律。没有记录的思考顺序不能被想象补足，已经丢失的失败尝试不能被编成完整历程。综合分析可以提出可能机制，但必须说明它与直接史证的距离。这样的限制让思想史保持开放，也防止方法模型替历史人物决定他们实际做过什么。

方法比较还应记录分析成本。若一个框架确实找到更多遗漏，却需要远超任务价值的时间，它可能适合重大争议而不适合每次常规处理。反之，一种简洁核查虽不能产生新理论，却能稳定减少常见误用，也有明确价值。工具的好坏应当联系问题粒度与风险，不能要求每次查询都进行整部思想史的重建。

这也解释为什么方法可以被部分封装。常见的版本检查、选择函数审查和来源追踪能够成为固定程序，让研究者把注意力转向真正不确定的环节。但封装必须保留重新打开的条件，例如遇到持续残差、跨方法冲突或新对象类型。成熟能力不是永远从零开始，也不是永远沿旧程序运行，而是在证据要求时改变判断层级。

若六路径与三原理能帮助识别这种重开时机，其价值将比形式覆盖更具体。这个主张仍然需要检验，尤其要比较没有使用这些名称但实行相同核查的团队。方法的有效部分可以共享，不应因为术语归属而被夸大或贬低。

本章因此把创新理解为一种有条件的响应能力：面对张力，研究系统能够识别应当改变什么，形成可操作方案，并让结果接受独立检验。六路径扩展入口，三原理追踪改变与回写，但二者都不能保证成功。保留失败与零增量，是这一方法继续发生的条件，而不是需要从历史中清除的障碍。

本章参考文献

[59-1] Account of the Discovery of the Planet of Le Verrier at Berlin. Monthly Notices of the Royal Astronomical Society 7, 153, 1846. https://academic.oup.com/mnras/article/7/9/153/983971

[59-2] Event Horizon Telescope Collaboration. First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole. 2019. https://arxiv.org/abs/1906.11241

[59-3] Dicke, R. H. et al. Cosmic Black-Body Radiation. Astrophysical Journal 142, 414, 1965. https://articles.adsabs.harvard.edu/pdf/1965ApJ...142..414D

[59-4] Fixsen, D. J. et al. The Cosmic Microwave Background Spectrum from the Full COBE/FIRAS Data Set. 1996. https://arxiv.org/abs/astro-ph/9605054

[59-5] Einstein, A. Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie. 1915，CERN 原文与译文馆藏。https://cds.cern.ch/record/632319

[59-6] Möller, A. et al. Fink, a new generation of broker for the LSST community. 2021. https://academic.oup.com/mnras/article/501/3/3272/5992334

第 60 章　新的天文学世界：已成能力、盲区与开放条件

60.1　下一代信号有哪些新增辨别力

走到天文学思想史的末端，最容易写下的是一串更大、更远和更精确的愿景。但新的天文学世界并不由设备尺度自动保证。真正需要说明的是：新信号能够区分哪些过去混在一起的可能，哪些属性因此获得新的联系，哪些研究操作能够稳定开展。没有这种辨别力说明，宏大技术清单仍只是愿望。

多信使观测已经提供一个可以分析的方向。GW170817 相关研究把引力波与电磁观测联系起来，使不同通道对同一事件提供互补约束。[60-1] 其意义不是“信息越多越好”的一般命题，而是不同信号与源参数之间具有不同依赖。恰当组合可以缩小某些退化，也可以检验单一通道难以检验的联系。

未来研究应当据此提出具体任务。例如，若一种新观测主要改善天空定位，它的价值在于增加可靠关联与后随机会；若改善时间采样，则可能分离过去混合的演化阶段；若增加波段覆盖，则可能更好限制状态或传播影响。不同改进对应不同问题，不能用一个总体灵敏度数字包办全部科学价值。

时间域尤其改变对象的研究单位。一次静态图像中的亮点，可以在持续记录中成为具有上升、衰减、周期或突变结构的事件。新的对象不只是原来位置上增加一个时间栏，而是某些属性只有在序列中才获得定义。采样节律、检测延迟和缺测模式因此直接参与对象构成。过于稀疏的记录可能把不同过程压成同一个形状。

多波段联合也要求警惕虚假的完整性。不同设施未必在同一时间观测，角分辨率不同，检测阈值不同。把结果放在同一张图上，并不意味着已经得到同一个物理状态的全部切片。可靠联合需要说明时间对应、位置归属与模型联系。差异只有在这些条件下才能转化为互补，而不是新的混淆。

因此，开放议程应当从当前缺失的约束出发。先说明某个问题为何无法区分，再讨论什么新观测能够改变这种状况。技术愿景由此接受可检验目标：如果新资料到来后仍不能分离候选，就需要重新评价原先的能力主张。辨别力需要在资料到来后实际检查，并在不满足预期时修订项目的解释目标。未来不是对现有理论的自动兑现，而是新的约束可能改变现有问题。

新增辨别力还应考虑没有探测到信号的结果。若仪器在某个范围内具有明确灵敏度，未检出可以限制某类模型；若覆盖不足或时机错过，未检出则可能信息很弱。未来研究计划需要事先说明什么结果会支持、削弱或不能判断一个假说。这样，设备即使没有带来预期发现，也能够产生可用知识，而不是只有成功图像才算成果。

联合观测还可以通过降低共享误差来增加价值。两个精度一般但误差来源不同的通道，有时比同一技术的大量重复更能检验关键关系。应当在项目设计阶段比较约束方向，而不是在设备建成后才寻找适合它的问题。技术与问题可以在实践中相互促进，但不能由技术愿景单向决定全部研究价值。

60.2 AI 应承担什么

AI 在天文学中的价值，可以首先放在资料规模与判断结构之间理解。大量警报、图像和文献超出人工逐项处理的能力，机器能够帮助筛查、关联和发现模式。但研究任务并不是把更多记录贴上标签，而是形成具有可追踪证据的对象与关系。效率增长要与确认能力一起评价。

Fink 等警报处理系统显示，机器分类可以嵌入交叉关联和后随工作流。[60-2] 在这种复合系统里，AI 承担哪些环节，应由可验证性能决定。它可以迅速提出候选，也可以帮助检查已知错误模式；某些高风险解释则仍需独立资料。责任分配不是按“人一定高明、机器只能辅助”的抽象等级，而是按具体任务的可靠性与可复核性安排。

一个值得推进的方向是让系统不仅输出答案，还保存答案的证据位置。某次分类使用了哪些观测、排除了哪些候选、对哪个输入最敏感，应当能够回查。这不要求所有模型都提供一段自称解释的自然语言。语言解释若没有连接实际计算与资料，也可能只是事后叙述。真正重要的是可以重做或反驳的依据。

AI 还可以帮助建立知识发生档案：保留早期候选、后续确认、分类修改与模型版本。这样，研究者能够分析哪些判断持续可靠，哪些错误反复出现，以及什么新资料真正改变了结论。档案的价值不只在记住成功，也在让失败成为下一轮学习条件。未经核验的机器总结不能替代原始记录，但可以作为进入记录的索引。

对于新奇发现，系统需要保留不熟悉的部分。一个只优化既有标签准确率的模型，可能非常有效地排除难以分类的对象。探索能力则要求某种拒绝分类、异常保留或审计机制。它们会消耗资源，因此必须检验收益。AI 的创造力不能只用输出新名称的数量衡量，而应看是否促成了可确认的新关系或更清楚的反例。

这也改变人的判断任务。研究者可能不再检查每一条原始记录，但仍需要判断模型适用域、评价指标、训练资料与后随策略。能力并没有因为局部封装而消失，而是转移到新的层

级。若这些高层判断也被封装，系统就需要新的验证机制。复合主体的可靠性来自层级之间可以重新打开的接口，而不是假定任何一层永久无误。

SDE 在此可以提出一项明确问题：新的显露是否真的改变了差异运行，并形成更好的共同条件？若 AI 只是加快旧流程，所得是重要的效率提升；若它让过去无法组织的序列、关联与异常进入持续研究，则可能产生更强的世界构成。两种价值都应保留，但不能混称。

60.3 国际协作与开放证据

现代天文学的许多对象，需要跨设施、跨团队和跨处理系统才能稳定成立。国际协作因此不仅是把更多人放在同一个项目名称下，而是协调时间标准、坐标、数据格式、质量说明与确认规则。没有这些公共接口，资料即使分别优质，也可能难以形成可共同检验的整体。

开放数据的第一层价值，是让不参与原观测的研究者能够提出新的问题。第二层价值，是让原结论能够接受独立复查。两者都需要文档与来源信息。只有一组最终数值而没有处理历史，可能足以开展某些应用，却不足以检查某类争议。公开应当按实际可追踪范围评价，而不是简单分成开放或关闭两个标签。

国际虚拟天文台联盟的来源信息模型，尝试描述数据实体、活动及相关责任者之间的关系。[60-3] 这类标准的重要性，在于使“这个数值从哪里来”具有可交换的表示。来源记录不能保证处理正确，却让错误有机会被定位。标准的价值因此是为检验提供结构，而不是替科学判断发出合格证。

公共数据还会改变参与门槛。拥有计算与分析能力的研究者，可以使用大型设施留下的资料，开展原团队未预设的研究。但获取不等于平等使用。计算资源、专业训练、文档语言和长期维护仍会影响谁能够提出有效问题。知识共同体若只提供下载入口而忽视这些条件，开放的潜力可能无法充分实现。

同时，开放不应以抹去劳动为代价。仪器建设、校准、资料清洗和软件维护都是科学成果的承重部分。清楚的引用与贡献记录，可以使复用者知道依赖谁的工作，也帮助共同体维持公共资源。一个不断产生论文却无人维护校准的体系，很难长期保证对象的可靠性。

差异也需要公共容器。不同团队可以使用不同算法、先验或分类，只要它们说明条件并能比较结果，就可能共同增加稳健性。强行统一到一种处理方式，可能提高效率，却也可能使共同错误失去对照。和谐协作不是没有分歧，而是把分歧组织成能够互相约束的研究关系。

因此，新世界的公共性包含两个方面：人们可以在同一标准下分享对象，也可以在标准不足时提出有证据的修订。前者提供稳定性，后者保留发生能力。如果公共规则只容许调

用、不容许质疑，它会逐渐从协作条件转成认知边界。开放证据的目标，是使这种边界能够被清楚看见并合理改变。

长期维护也包含保存旧版本。当新的处理改进某些结果时，旧数据产品不应无说明地消失，否则已经发表的工作难以复查。合理做法是保留版本、标明已知问题，并使用户能够找到更新关系。科学世界的持续性，需要这种把历史判断与当前判断连接起来的基础设施。

对跨学科使用者，标准还承担翻译责任。统计人员、仪器专家与物理解释者使用同一栏目时，可能赋予它不同含义。清楚的单位、模型和质量定义能够减少这种错位，但不能完全消除。合作中的疑问不应被视为沟通效率的障碍；它们有时揭示了对象尚未真正共享的边界。

60.4 本书如何允许被改写

一部发生学思想史若把自身写成不可修改的最后解释，就违背了它所强调的知识发展。本书应当允许三类材料改变结论：新的原典改变对历史过程的理解，新的科学资料改变案例的证据地位，新的竞争解释表明某些 SDE 概括没有额外价值。三类改变分别作用于史实、科学判断和方法评价，不能混在一次笼统修订中。

对于原典，重要的是回到具体主张。若新的传本表明某种方法早已存在，创新归属与传播叙述就应修改；若材料显示研究者没有采用本书假定的中介步骤，就应撤回相应因果解释。后来的模型可以帮助理解历史结构，却不能代替历史行动的证据。一个优美解释如果不符合材料，仍然需要改变。

对于现代案例，版本记录应当保留。新的 Gaia 发布可能改变误差说明，新的距离研究可能改变张力评估，新的警报系统可能改变分类流程。修订应标明哪些结论被加强、哪些被削弱、哪些只是数值更新。这样，读者能够看见知识的实际演变，而不是每一版都呈现为从来如此。

对于 SDE 方法，最重要的开放问题是增量。三方程是否帮助识别此前遗漏的相互规定？六路径是否在新问题中改变了判断？三原理是否能够区分生产性改变、停滞与错误稳定？这些问题不能由理论自己的语言循环回答，需要案例、对照与明确失败条件。方法也必须进入它要求别人进入的检验过程。

修订也需要区分核心主张与修辞。若新材料只改变一个日期，不应宣称整个框架受到证明；若它否定一个承重机制，也不能只改几处措辞便保留原结论。修改的幅度应与证据作用匹配。对于本书，这意味着每个跨时代概括都应能够回到具体章节的材料与推理，而不是依靠反复出现获得权威。

开放议程还包括未被本书充分覆盖的传统和对象。新的史料可能改变现有章节的重心，使原来边缘的实践成为解释关键。发生学史不应把既定目录当成世界本身的结构。目录是当前研究能力形成的组织方式，它应当服务材料，而不是要求所有新材料迁就它。

本书提出的新的天文学世界，因而具有有限而具体的含义：新的显露形式、差异运行与属性纠缠形成可持续研究的复合整体，使对象身份、尺度和证据关系获得新的组织。这个主张不等于天体物理过程由文字制造，也不等于所有旧知识都被取消。它强调科学活动实际形成了新的可感、可算、可操作和可共同检验的存在联系。

在这些联系中，人、仪器、数学、算法与公共规则共同承担认识能力。任何单一环节都不能独占全部创造，也不能免除自身责任。未来天文学的开放程度，将取决于这个复合系统是否既能稳定使用已成能力，又能保留异常、反例和新的合作方式。统一性给出共同问题，多样性提供新的辨别方向，和谐性使它们在冲突中继续形成可工作的整体。

天文学思想史因此不是一条从黑暗走向全知的直线。它记录的是，人类怎样一次次改变能够显露什么、怎样比较差异、怎样让属性联系成为公共世界。每一次成功都打开新的领域，也留下新的盲区。真正延续这段历史的方式，是把已成能力变成继续提问的条件，并让下一次可信的差异有机会改变我们现在的答案。

本章参考文献

[60-1] LIGO Scientific Collaboration et al. Multi-messenger Observations of a Binary Neutron Star Merger. 2017. https://arxiv.org/abs/1710.05833

[60-2] Möller, A. et al. Fink, a new generation of broker for the LSST community. 2021. https://academic.oup.com/mnras/article/501/3/3272/5992334

[60-3] IVOA. Provenance Data Model. Recommendation, 2020. https://www.ivoa.net/documents/ProvenanceDM/

结语　已支持、应收缩与开放问题

结.1　科学世界的形成留下什么可辨认的变化

一部天文学思想史的结束，不应把宇宙写成已经完成的知识地图。星空没有因为章节结束而停止提出问题，历史材料也没有因为解释连贯而失去重新排列的可能。本书真正需要交代的是：从复认天象到联合解算，从位置记录到多信使关联，我们究竟获得了什么足以检验的理解，又有哪些希望超过了材料能够承担的范围。

首先需要保留的判断，是天文学对象具有可以追踪的研究身份。天体不以人的确认作为存在条件，但“同一个可持续测量、可跨仪器比较、可进入推理的对象”，需要一套能够运作的辨认关系。天空中的亮点与目录中的记录既不能直接等同，也不能被说成全无联系。它们之间的联系，要经过位置、时间、响应特征和误差范围等条件的共同限定。研究身份成立以后，它又反过来组织新的观测。

这一判断的分量，在于它要求历史叙述说明具体变化。某种光变怎样成为可重复辨认的周期，某种谱线怎样成为可比较的物理指标，某个图像特征怎样在处理方法改变后仍然保留，都是可以追问材料与操作的问题。若只说人类“终于认识”了对象，构成研究能力的许多条件便从叙述中消失了。若只说对象是人类“建构”出来的，信号对任意解释的抵抗又容易被遮蔽。科学世界的形成必须同时说明这两方面。[结-1][结-2]

第二个需要保留的判断，是差异的研究作用取决于差异怎样进入序列。一个孤立的不符，可能来自仪器、识别、模型，也可能提示新的物理关系。只有当比较对象、误差限度和允许改变的条件逐渐明确时，不符才获得更确定的含义。历史上的创造性并不只是提出大胆答案，还包括把一种模糊的不舒服转化为能够分别检验的困难。新的差异序列使某些回答变得可能，也使另一些回答更难维持。

第三个判断涉及特征之间的相互规定。测量方法并不是面对一个已经完全规定的属性清单，而是在处理信号时确定哪些属性能够共同估计、哪些属性仍然混合。反过来，对属性关系的理解又会改变观测安排和仪器要求。这种依赖可以在一次联合解算中出现，也可以在较长的研究史中形成。二者不能直接等同，却可以分别说明为什么对象、操作和解释很少各自独立地前进。[结-2]

这三个判断共同支持一种有条件的科学世界构成论：新的天文学世界，可以体现为对象身份、可行比较和属性关系形成了新的稳定组合，进而产生过去不能可靠开展的研究行动。“世界”在这里不是修辞性地放大任何小进步。它要求变化具有一定的关联范围和可迁

移性，能够影响不止一次孤立判断。某一结果获得更小的误差，与一种新的研究关系取得公共可用性，仍然应当区别。

这个解释也为历史连续性留下位置。新体系未必抛弃旧体系的所有成果，新技术也常常依赖较早建立的记录、坐标与分类。变化有时集中在某个关键关系上，其他部分继续有效。把历史写成一连串整体替换，无法说明旧资料为什么还能被重新使用；把历史写成单纯累积，又无法说明同一份资料为什么会取得不同的推断位置。具体的保留与改变，比预先规定革命或连续更有解释力。

这些判断并不自动证明 SDE 比所有其他框架更优。精细的仪器史、模型史与统计方法研究，本来就能揭示许多相同事实。本书的解释主张，主要在于把显露、差异序列与特征纠缠置于同一分析中，要求三者的关系能够落实为操作与判别。它的价值需要在具体案例里衡量，不能从覆盖题材的广泛性直接推出。[结-3]

结.2 哪些命题必须收缩

最先需要收缩的是把“发生”当作普遍历史因果的冲动。我们可以说明一项结果在逻辑上依赖什么，却尚未因此说明历史人物为何走向它。一份新星表可能使某种比较成为可能，但研究者是否真的使用了它，还需要传播、阅读和操作的证据。一个模型存在后来看来明显的困难，也不代表这一困难当时已经成为改变模型的直接动力。构成依赖与历史推动不能互相代替。

其次，三方程只能在本书的使用层次上承担关系表达与问题组织的责任。它们不是具有给定量纲和边界条件的天体物理方程，也没有凭借符号形式提供独立的数值预言。若要把某个案例表达为函数关系，必须另外说明变量如何确定、映射如何识别，以及什么结果能够否定这种表达。否则，任何历史变化都可以被重新命名为一次回写，框架就失去了区分能力。[结-3]

对六种进入路径也应作同样限制。从显露进入、从差异进入或从已有关系进入，确实可能改变研究者首先检查的条件；但是，路径的排列并不等于已经发现六种普遍历史规律。它们尤其不能替代实际发生次序。一个现代读者从某个结果回溯条件，属于分析策略；一位历史人物先作什么、后作什么，属于需要材料支持的过程判断。两者可以相互帮助，却不能混写。

第三种需要收缩的命题，是任何反馈都能促进知识进步。校准可以改善结果，也可能在共同假设下强化同一种偏差；更严格的样本选择可以提高纯度，也可能使另一类对象更难进入研究；自动分类可以扩大处理能力，也可能使后续观测持续优先选择既有类别。反馈的存

在不是成功证据。需要问的是反馈究竟改变了什么，是否引入了新的辨别条件，是否让原先被排除的可能性重新接受检验。

同样不能把科学世界的构成无限推广为全部自然存在的生成。恒星形成、核反应、星系演化等物理过程有各自的证据与模型。人的研究关系会改变这些过程怎样被识别和说明，却不能凭借思想史论证断言人的认识使它们在物理上发生。本书讨论的不同层次具有联系，联系需要桥梁；取消层次之间的区别，并不会使理论更深刻，只会使判断更难检验。

还有一种更安静的过度：为了保证全书统一，把每个案例都写成同一套成功故事。古代纪时、精密位置、光谱反演与巡天统计，面对的资料形态和可靠性问题并不相同。如果三种概念在所有章节中都只起到替换名词的作用，那么统一只是语言重复。一个有效框架应允许某些章节主要说明一种关系，允许某些材料暂时无法纳入，也允许某些地方并没有比已有解释增加更多认识。

这些收缩不是为理论预先安排免受批评的退路。相反，收缩必须使理论承担更清楚的责任：若某个所谓关键纠缠在历史材料中并不存在，应撤回相应解释；若竞争框架已经同样充分地说明过程，应承认增量有限；若不同路径不产生可辨认的判断差别，就不应宣称路径分析已获得支持。理论的自我修订应留下可检查的改动，而不是仅以开放态度作答。[结-4]

结.3　能力增长为什么仍然留下盲区

天文学的研究能力常以更远、更暗、更快、更精确来描述。这些表述有明确意义，但仍然只是能力的一面。另一面是，新的探测范围总要通过具体选择与处理程序成为可用材料。仪器能够记录的信号，不等于项目实际保留的信号；项目保留的资料，不等于研究者已经能够可靠解释的资料。能力增长首先扩大可能性，可能性转化为知识仍需经历具体工作。

一个研究共同体因此也由它能够辨认的失败构成。如果只能报告成功匹配，不能保留未解释的残差；如果只能输出最优参数，不能交代参数之间的相关；如果只能发布分类结果，不能追踪样本怎样进入分类，那么能力的公共可检验性便受到限制。失败记录、误差结构和数据出处看似处于结论外围，实际上决定其他人能否在新条件下重新使用结果。

联合分析使这种责任更加明显。当多个观测通道共同支持一个对象时，需要确定它们提供了怎样的独立性。多个结果来自不同仪器，并不必然意味着所有误差都独立；共享的校准或模型假设，可能构成共同约束。反过来，方法之间存在联系，也不意味着它们毫无互相检验的力量。关键是把联系说明白，使一致性与独立支持不再仅凭数量判断。

对图像而言，公开最终图像有重要价值，但还不足以说明某个结构在怎样的处理改变下能够保持。对目录而言，公开大量记录有重要价值，但还需要理解质量标记、筛选条件和缺失来源。对模型而言，提供程序有重要价值，但程序能够运行，不代表其输入假设已经取得

经验支持。开放的科学工作不是一次上传动作，而是尽可能使别人具备重新判断的条件。[结-1][结-2]

人工智能加入这一过程后，同样需要区分产出速度与证据力量。它可以帮助寻找候选、整理文献、比较表述和发现不一致；这些帮助是否提高研究质量，取决于错误能否被发现、来源能否被追踪，以及关键判断由谁核验。一个语言上完整的解释，不会因为篇幅更长而变得更可信。本书自身采用协作写作，也必须接受相同尺度：术语统一不能代替史料核对，章节齐全不能代替论证完成。[结-4]

研究能力的社会条件也不宜只被放入背景。谁能使用仪器、谁能接触资料、哪些工作获得承认，会影响问题怎样被选择和延续。但指出这种影响，仍然需要具体材料，而不能直接从社会位置推导某项天文学结论的真假。思想史应同时保留证据判断与劳动组织两个问题，使它们发生联系而不互相吞没。

统一性、多样性与和谐性的平衡，最终也可以在这里得到检验。统一性保障结果之间能够交流，多样性保留不同信号、方法和历史处境提供的辨别力，和谐性则表现为这些差异能够相互制约并共同工作。和谐不要求消除所有冲突。某些尚未解决的不一致，恰恰可能阻止共同体过早关闭问题。能够保存分歧并使分歧继续可研究，是成熟能力的一部分。

结.4 把开放问题交还研究

本书留下的首要问题，是怎样更严格地识别一次科学世界重构。不能只依据发现的名气、仪器的规模或后来的影响作判断。更可行的方向，是比较变化之前与之后的研究行动：哪些对象获得稳定身份，哪些比较可以重复，哪些属性关系能够迁移到新的资料，哪些失败变得可以定位。这样的比较既需要个案细读，也需要避免事后只保留成功指标。

第二个问题，是如何在跨文明研究中使用共同概念而不抹平差异。某些社会以历法、占验或行政记录组织天空知识，某些研究传统以几何模型或物理机制作为主要解释资源。共同分析并不意味着这些目标相同。下一步研究应继续深入具体文本、计算实践与传承关系，让比较的单位由材料逐步确定，避免用现代学科边界先行裁剪所有传统。

第三个问题，是如何追踪未被保留的选择。成功的理论和目录通常留下较清晰的文献，失败的拟合、被筛除的信号和未被采用的解释则可能较难恢复。仅凭最终发表结果，容易把历史写得比实际更直接。对于这些缺失，需要寻找工作记录与版本差异，也需要承认无法恢复的部分。未知不是等待漂亮叙事填满的空白，而是解释范围的一条边界。

第四个问题直接面向 SDE：它究竟在哪些场合增加了可检验的认识？今后的研究可以对同一材料分别使用精细模型史、实验与仪器研究、统计证据分析以及 SDE，比较各自提出了哪些不同问题，发现了哪些原先遗漏的关系。如果增加的只是术语，而没有新的区分，框架

就应缩小主张。如果确实揭示了可核对的相互规定，则应明确指出其证据，而不依赖理论声势。

这里还需要一种尺度上的耐心。单个案例可以反驳过强的普遍判断，却未必足以建立另一种普遍判断。跨案例的重复模式只有在对象选择、材料差异和反例范围得到交代以后，才可能承担更广的解释责任。历史比较的数量与比较的有效性，应当分别衡量。

因此，本书的终点是一组继续工作的责任，而不是一个能够替代所有天文学解释的总公式。让对象的身份接受追踪，让差异的作用接受比较，让关系的构成接受拆解；同时，让每一种解释保留被材料改变的可能。宇宙并不需要配合我们的叙述取得整齐。天文学思想的力量，正在于人能够在这种不配合中形成新的问题、公共的判断和仍然可以继续扩展的研究世界。

本章参考文献

[结-1] Event Horizon Telescope Collaboration. First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole. The Astrophysical Journal Letters 875, L4, 2019. DOI:10.3847/2041-8213/ab0e85。https://authors.library.caltech.edu/records/d6ns1-m4g65 。用于图像处理、方法比较和稳定结构的讨论；具体史事参见本书第 45 章。

[结-2] Lindegren, L. et al. The astrometric core solution for the Gaia mission. Astronomy & Astrophysics 538, A78, 2012. DOI:10.1051/0004-6361/201117905。https://arxiv.org/abs/1112.4139 。用于联合估计关系的说明；不以数值迭代直接证明历史反馈。

[结-3] 王德生：《SDE 三方程导论》第二章“三方程的标准形式与理论地位”。https://sdeuniverses.com/books/m/51/chapter-02/ 。概念来源与表达层次；天文学应用为本书解释主张。

[结-4] 王德生：《SDE 三方程导论》第三十九章“反例、批评与理论修订宪法”。https://sdeuniverses.com/books/m/51/chapter-39/ 。用于理论修订与 AI 协作的规范讨论，不作为天文学经验结论的独立证据。

阅读之后

一本思想史的完成，不是把所有天体塞进同一种解释。它应当使读者能够辨认：某个判断从哪里来，需要哪些操作维持，又会被什么证据改写。

本书的完整网页、学习包与智能体共用同一章序。你可以保存自己的初答、复答与异议，也可以把一条新的反例带回书中，重新检验它的解释范围。

阅读 · 学习 · 智能问对

sdeuniverses.com/books/astronomy-thought-history/

天体身份

一个亮点怎样成为可确认的对象

宇宙尺度

距离与时间怎样获得可比较的单位

证据制度

记录、仪器与共同体怎样形成判断

从古代观天、历法与几何宇宙，到光谱、距离阶梯、引力波与大规模巡天。十二编六十章，追问天文学世界如何发生，又怎样被新的证据重新打开。

阳涌 著 · 指导老师 王德生