AGI为何

永远不可能

工具、发生与智慧的无尽升级

王德生　著

德麦国际出版社

DEMAI INTERNATIONAL PRESS

出 版 信 息

书名	《AGI为何永远不可能——工具、发生与智慧的无尽升级》
著者	王德生
执笔	复合主体：王德生立题、口述承重命题与裁定，Claude 执笔与统稿
出版	德麦国际出版社（Demai International Press）· 新加坡
编号	德麦国际专著第 330 号
ISBN	979-8-90690-483-6
定价	US$20
开本	170 mm × 240 mm
字数	约 20 万字
版次	2026 年 10 月第 1 版

AI 协作声明：本书的判断、口径与承重命题（工具是已发生智慧的封装、使用即碰撞、错位定理、一界有工具三界有新生、创造力与智慧共生、人类水平是含着追赶者的过程、AGI 是理念界的一座塔等）由王德生口述给出；文字在其指令下由 AI（Claude）协作成稿，并经其审定。书中史实、年份、研究与人名，凡标〔待核〕处尚未逐项核对原始出处，汇总于参考书目末尾的「引文核验总表」。书中所举日常场景（如装修、洗衣机、一碗面条、写婚礼致辞）均为说明用的构造场景，不是某个人的真实经历。关于心理、修行与健康的章节只讲理解与习惯，不构成医学或宗教建议。

全书结构：出版信息、作者介绍、前言、导读、目录；导论与十编四十七章；结语、参考书目、附录四件。

作 者 介 绍

王德生，博士，SIO 本体论与 SDE 发生学的创立者，新加坡德麦国际（Demai International Pte. Ltd.）创始人、董事长，SDE Universes 平台主持人。

他是计算数学出身：湘潭大学数学系本科与硕士，中国科学院数学研究所博士，中国科学院计算数学研究所博士后；此后在英国威尔士大学斯旺西分校任高级研究员，在新加坡南洋理工大学数学系任教与研究十三年，任博士生导师，在国际期刊发表计算数学与图形学论文四十余篇，研究集中在网格生成、变分表面重建与波的模拟。

2015 年起他回国推广自己的方法论，2017 年 1 月以一篇《给你一双慧眼看天地》开始了「三视角」时期。此后他的思想走过四站：三视角（2017 年 1 月起）、321 智慧系统（2022 年 10 月起）、SIO 本体论（2024 年 7 月起）、SDE 本体论（2026 年 1 月起；同年 5 月起「结构」改读为「显露」）。本书是他对 AI 时代"通用智能"这一理念的发生学回答，也是他提出的"工具让渡律"与"主体经济学"思路中的一环。

前 言

它什么都能做的那个下午

◆ 一、一篇文章写完的时候

先设想一个场景。这是一个构造出来的场景，不是某个人的真实经历，但读者里大概有不少人经历过它的样子。

一个周日的下午，你的表哥要结婚，家里人推你写一段致辞，说你是"读书人"。你对着空白文档坐了半小时，一个字没动。后来你把几句零散的话丢给了AI：新郎姓什么，新娘做什么工作，两家人在哪个城市，要求"温暖、得体、三分钟左右"。

几十秒之后，一篇完整的致辞出现在屏幕上。有开场，有祝福，有一段关于"携手"的排比，结尾还很会收。你把杯子放下，往椅背上靠了一下。那一刻你心里冒出来的念头，大概也是许多人第一次用它时的念头：它什么都能做。

那种感觉很具体：不是惊讶，更像是松了一口气，又隐隐有点发凉。松的是，这件让你发愁了半小时的事，居然几十秒就有了交代；凉的是，你忽然不太确定，接下来还有哪些事，是只有自己才做得了的。

接下来几天，你又拿它试了别的事：把一份冗长的通知改成三句话，帮邻居的孩子检查英语作文，把一段含糊的合同条款翻译成人话。每一次它都做得不坏。于是那个念头越来越像一个结论：既然写作、翻译、总结、解题它都行，那么剩下的那些事，不过是时间问题。这样的东西，叫它"通用人工智能"，似乎并不过分。

然后事情出现了一点不对劲。

致辞你没有直接用。读第二遍的时候你发现，它写的是"一场婚礼"的致辞，不是"你表哥的婚礼"的致辞。里面没有他小时候把你的自行车骑进河沟的那件事，没有他追新娘时每个周末坐四个小时火车的那份笨拙。你动手改。改着改着，你开始

往里加东西：先加了河沟，再加了火车，加到后来，整篇都不是原来的样子了。更奇怪的是，在这个改的过程里，你第一次弄清楚了自己到底想对表哥说什么。那句话，在你对着空白文档发呆的半小时里，并不存在。

你当时可能没有多想。可是把这件小事反复回想，会越想越觉得，它里面藏着一个很大的问题。

◆ 二、缺的未必是信息

AI交给你的，是"婚礼致辞"这件事被做完的样子。而你真正需要的，是一个你之前没有的东西：一段只属于你和表哥的话。前者可以被封装进一个工具，后者要等你动手之后才会出现。工具越是把前者做得漂亮，你越是清楚地看见后者还在门外。

这个场景讲给身边的人听，反应往往出奇地一致：有人说"那是你提示词没写好"，有人说"过两年就不是这样了"，还有人说"其实我也有这种感觉，只是说不上来"。三种反应都默认了同一件事：缺的那一块，是可以补上的。补法也简单，要么多告诉它一些信息，要么等它再强一些。

可是缺的未必是信息。河沟和火车，你当然可以事先打字告诉它；但在告诉它之前，你并不知道这两件事就是整篇致辞的骨头。是写的过程，让它们从一堆记忆里站了出来。你要的那句话，不是先躺在脑子里、等着被取出来的东西，而是在动手的过程里才出现的东西。取出来，和长出来，是两回事。

换到别的场合也一样。你让AI替周末的饭桌列一份菜单，它列得头头是道。可是下周你妈妈要来，孩子刚出牙不能吃辣，冰箱里还剩半棵蔫了的白菜。哪几样要紧、哪几样可以放弃，是你站在冰箱前、一样一样掂量的时候，才弄明白的。菜单可以交给工具，那份掂量的结果，要在掂量的过程里才有。

这本书，就是从这样一个下午长出来的。

◆ 三、这本书要说什么

书名里有一个很重的词："永远"。《AGI为何永远不可能》，这样的书名容易被看成一句气话，或者一次对技术的不甘心。所以我必须在前言里把话说得更精确一些。

我要说的是一个有分寸的判断，分三层。

第一层：作为终局，AGI不可能。所谓终局，是指这样一种设想：某一天，有一个系统，它的智慧已经"装满"了，人类能想到的问题、将来会冒出来的问题，它都能处理，再没有哪个问题需要更高一级的智慧来对付。我认为这个设想在结构上就站不住。

第二层：作为阶段，类似AGI的能力很可能出现。我并不否认AI会在越来越多的领域里做得和人一样好，甚至更好。书里没有一句话是要贬低这件事的价值。

第三层：这种能力一出现，就不再"通用"。因为它一旦被造出来、被大量使用，使用它的人和它之间就会碰撞出新的问题，而这些问题在它被造出来的时候还不存在。通用的边界被推远了，但推远的同时，边界之外又长出了新的东西。

三层合在一起，才是本书的立场。只记住第一层，会把这本书读成一本唱衰的书；只记住第二层，又会读成一本唱好的书。我真正要走的路在第三层：能力可以越来越强，而智慧的升级，没有尽头。所以书名里的那个"永远"，说的不是机器会不会进步，而是升级这件事有没有终点。

这就是全书的总命题，我先在这里写下来，后面要用整整十编去论证它：

> **AGI不是智慧的终点，而是又一次封装；每一次封装，都是下一次发生的起点。**

这句话我会叫它"判词"。它在书里一共出现四次：前言一次，导读一次，导论一次，结语一次，一个字不改。

◆ 四、AGI感：看不见，不等于没有

既然如此，为什么这么多聪明人会觉得AGI已经近在眼前，甚至已经到来？本书的第二个任务，就是解释这一点。我把这种感觉叫作"AGI感"。

它是一种真实的感觉，有真实的原因。每一次新工具出现之后，都有一段时间，旧问题被它解决得干干净净，而新问题还没有显露。想一想你刚换上洗碗机的头几个星期：碗盘闪闪发亮，厨房里再没有一个脏碗，你会觉得，洗碗这件事，算是翻篇了。要等到几个月以后，你才会发现新的事情：哪些锅不能进、餐具怎么摆才洗得到、盐和亮碟剂什么时候补，以及全家洗碗的分工，悄悄变成了另一种样子。

在这段"封装已完成、新问题尚未显露"的等待期里，人们很容易把看不见当成没有。那不是愚蠢，也不是被广告蒙蔽，是一种很正常的认知反应。

这样的感觉并不新鲜。电力刚普及的时候，人们觉得它能解决一切；1965年，有人预言二十年内机器能做人能做的一切工作；信息高速公路的年代，也有人觉得信息一旦畅通，问题自然就消失了。这些感觉都是真诚的，没有一个成为事实。所以本书对AGI感的态度，是解释它，不是嘲笑它。把这种感觉当作"证据"，说"大家都觉得它通用了，它就是通用的"，是另一回事；这个问题留给第十编。

◆ 五、这不是一本反AI的书

我想把一个误会提前拦住。这本书不是写给想要"打击AI"的人的，也不是写给想要"安慰自己"的人的。如果读完之后，你觉得"哦，那就不用担心了"，那是我没写好。

恰恰相反，本书的论证有一个并不轻松的推论：AI越强大，它引发的新问题就越多、越深，而且这些问题多半不是AI自己能够解决的。核武器没有让人类变笨，它逼出了威慑理论和核伦理；汽车没有让人类变懒，它逼出了交通法、城市规划和

保险业。工具变强的时候，人面对的是一个更复杂的世界，需要更高一级的智慧。这并不是一个让人松口气的消息，倒更像是一份通知：升级的任务，一直都在。

我也不想拿"永远"去吓唬谁。一个普通人读到"升级没有尽头"，第一反应可能是疲惫：又要学，又要变。可是回到前言开头的那个下午，你会发现，所谓升级，并不是另外抽出时间去上一门课，它就发生在你动手改那篇致辞的半小时里。

我也要说明本书的位置。我早年学的是计算数学，博士阶段的工作是网格自动生成。那类程序是"自动化"的一个小缩影：计算机把过去由工程师一点一点画出来的网格，自动生成了。自动化吃掉的是一层操作，而工程师的注意力随之转到另一层：判断网格好不好、该不该这样划分、这个问题到底要算什么。〔待核：此处对工程师工作转移的描述是一般性概括，请作者按亲身经验确认或改写〕后来我用了很多年的时间思考这类现象，才慢慢把它整理成一套叫作SDE的看法。SDE在导读里会用大白话交代，这里先不展开。

◆ 六、怎么读，以及它允许被推翻

全书十编，四十七章，另有导读、导论、结语和四个附录。你可以从头读到尾，也可以先读导读，那里有一张"五层论证地图"，标出了每一编在整个论证里站的位置。

先记住一条线就够了：第一编立靶子，第二编搭原理，第三至六编用三界的工具史作佐证，第七、八编回答智慧与人类水平，第九编收成一个结构证明，第十编解释人们为什么会有AGI感。每章从一件日常的小事开场，比如洗衣机、导航、新手爸妈的第一个夜晚、一勺糖；凡是用到专门的说法，我会在它第一次出现的地方用日常的话解释一遍，不堆术语。

书里的几个核心想法，在邻近的学科里都有近亲：波普尔谈知识增长的不可预测，技术演化论谈技术从技术里长出来，创新经济学谈新工具带来的新岗位，组织学习理论谈"换一种想法"。每到相应的章节，我会写明本书和它们相同在哪里、不同在哪里。这不是客套：一个想法如果在隔壁早有人说过，读者有权知道，我也该知道自己到底多出了什么。

这本书还允许被推翻。第4章会专门列出全书的"可错条件"，也就是如果将来发生了哪几件事，我就认为本书的核心论断站不住。一个只会说"你看我多有道理"的论证，是不值得信任的；能说清楚"在什么情况下我是错的"，才有资格被认真对待。

◆ 七、这本书不做什么、写给谁

先说不做什么。它不做技术预测：不预言哪一年会出现什么，不评价任何一家公司或任何一款产品，也不碰当下的政治争议。书里涉及的历史事实，我尽量只用有把握的；记忆不够确切的地方，在句末标了〔待核〕，请读者和编辑朋友帮我核对。引用他人的话很少，且都只是片言只语，其余一律转述。对当前的大模型，我只谈它的结构性特征，并且标明"这是当前技术状态，不是原理必然"，因为技术状态变得很快，而我想写的是不会过时的部分。

再说写给谁。写给普通人：每天用着这些工具，心里偶尔嘀咕"它是不是什么都会了"的人；被"过两年就不一样了"和"永远不可能"两种说法来回拉扯的人；正在为自己的工作、孩子的学业或一家小店的将来犯愁的人。你不需要懂算法，也不需要读过哲学。你只需要有过那样一个下午：把事情交给工具，工具做得很漂亮，而你仍然觉得有什么东西，还在门外。

如果读到某一编觉得吃力，不妨先只带着导读里的"随身三问"，去翻翻各章的开头和末尾的"小账"。那里用最短的话，记着每一章封装了什么、逼出了什么。

最后，回到那篇致辞。设想你在婚礼上念的是改过的那一版，念到河沟那一段，表哥在台下笑得直不起腰。那个笑声是AI给不了你的，不是因为它写得不够好，而是因为那一段话在被写出来之前，你自己也不知道它会存在。

这就是我想在这本书里讲清楚的事：智慧不是一个可以装满的仓库，它是一个人每一次动手之后，多出来的那个自己。

王德生

导 读

五层论证地图

◆ 一、先给你一张地图

一本二十万字的书，读到第三编的时候，人很容易忘了自己是从哪里出发的。所以我先把整张地图摊开。你可以把这一节当作全书的目录之外的另一份目录：目录告诉你每一章叫什么，地图告诉你每一章为什么在那里。

全书的论证一共五层。前四层回答同一个问题：为什么AGI不可能是智慧的终点。其中第一、三、四层击中的是"装满了、不必再升级"的终局AGI；第二层转向更常见的说法，也就是"达到人类水平"，要说明连这个说法也没有一个停得住的终点。第五层回答另一个问题：既然不可能，为什么这么多人觉得它看起来可能。

前四层像四根柱子，从不同的方向撑住同一个结论：AGI不是智慧的终点。第五层像一座桥，把"不可能"和"人们的感受"连起来。少了前四层，书没有论证；少了第五层，书没有同情，也就解释不了为什么读者此刻会拿起它。

先用一张表把五层摆在一起，后面各节再一层一层展开：

层	它回答的问题	主场	它在什么情况下错
一　没有终局智能	工具会不会有一天把智慧用完	第二编	找到一件开放领域里的工具，大量使用之后不生任何封装之外的新问题
二　没有固定的人类水平	"达到人类水平"是不是一个确定的目标	第八编	AI追上某条横杆之后，公认的"人类水平"清单长期不再增补
三　没有可装满的智慧	智慧是不是一座可以装满的库存	第七编	某个开放领域里程碑一路靠存货达成，此后多年也不冒出存货之外的新问题类别
四　三界工具只会提升三界智慧	工具史上看，工具到底带来了什么	第三至六编	找到一件广泛使用的工具，几十年里的新用法在设计之初的文件里都已写过

层	它回答的问题	主场	它在什么情况下错
五　AGI感是时间差现象	既然不可能，为什么觉得可能	第十编	AGI感进入内卷期之后依然持续高涨，没有回落

整本书压成一句话，就是前言里写过的那句判词。它在导读里再出现一次：

> **AGI不是智慧的终点，而是又一次封装；每一次封装，都是下一次发生的起点。**

◆ 二、先用一个厨房讲清"封装"与"新问题"

在进入五层之前，我需要先借一个日常的例子，把后面反复出现的两个词讲明白：封装，和新问题。

想一想洗衣机。在洗衣机出现之前，洗衣服是一整套手上的功夫：搓、揉、捶、拧、晾，哪种布料怕热，哪种污渍怕碱，全靠长年累月的经验。洗衣机出现之后，这套经验被"封装"进了一个盒子：你按一个键，它替你搓、揉、拧。那些曾经要靠手感去判断的东西，变成了程序里的水位、转速和时间。

这是工具的第一个面目：它把已经有人弄明白的事，固定下来，让后来的人不必再弄一遍。

但故事没有停在这里。有了洗衣机，人们开始洗得更勤。衣服换得更勤，床单洗得更勤，家里的"干净标准"悄悄升高了。与此同时，一些前所未有的新问题出现了：洗衣机该买多大的？洗衣液里的成分对皮肤好不好？洗衣机坏了找谁修？这些问题，在盒子造出来的那一天，没有一个已经存在。它们是人们开始使用这个盒子之后，才从使用之中长出来的。

这就是本书的核心图像：工具封装已经发生的智慧；使用工具会碰撞出封装之外的新问题。封装的东西，是过去；新问题，是未来。过去被装进盒子的那一刻，未来还没有到来。盒子里装不下它，不是因为盒子做得不够大，而是因为装盒子的时候，它还不在。

我在书里把这个现象叫作"错位"。这是后面几乎每一编都会回到的一个词。

◆ 三、SDE：三个字母的大白话

书里会用到一套叫SDE的说法，我在这里先用最朴素的话把它交代一遍，免得你读到后面卡住。

S，代表"显露"。一个东西在你眼前露出来的、相对稳定的样子，就是S。洗衣机的按钮、菜谱上的步骤、一份合同的条款，都是S。我们平常说的"结构"，就是显露出来之后稳定下来的那个样子。

D，代表"差异"，也就是发生的过程。从一个样子变成另一个样子，从不会到会，从没有想法到有了想法，中间那个"变"的过程，就是D。

E，代表"纠缠"，可以想成土壤、场合、意义所在的那片地方。同样一句话，在婚礼上是祝福，在法庭上可能是证词。让一件事"是这件事"的那个背景网络，就是E。

SDE三个字母放在一起，说的是同一件事：任何一个看得见的东西（S），都是在某片土壤里（E），经过一个变化的过程（D），才显露出来的。工具，就是一个被封装起来的、稳定下来的S。而使用工具，会让它重新落进新的土壤，重新开始变化，于是又有新的S显露出来。

读到这里就够了。后面每一编再提到SDE，我都会再用一两句话提醒。

◆ 四、五层论证，一层一层看

第一层　没有终局智能（第二编）

这一层回答："工具会不会有一天把智慧用完？"

答案的结构很简单。前面说过，工具封装的是已经发生的智慧；使用工具会碰撞出封装之外的新东西。这样一来，任何工具，无论多强，只要有人用，就一定会在使用中生出它自己没有包含的新问题。所以，不存在这样一个时刻：工具把所有

智慧都装完了，再没有什么需要新发生的了。回应这些新问题的智慧，不在原来那件工具里，它比那件工具高一级；它一旦也被封装成工具，又会长出自己的外面。

第二编用五章把它搭起来：工具是封装（第一条承重命题）；使用即碰撞（第二条）；封装与新问题在时间上错位（第三条，叫错位定理）；二阶碰撞逼出下一级的升级；最后一章讲"工具让渡律"的四个阶段：扩大期、内卷期、新产业诞生、让渡。这是一条在历史上反复出现的节奏，你会在饭店老板和他的面条里看见它。

第二层　没有固定的人类水平（第八编）

这一层回答："那'达到人类水平'总是一个确定的目标吧？"

不是。因为"人类水平"这条横杆本身一直在升高。一百年前，会开车不算一项人类水平；今天，它算。更要紧的是，横杆里含着追赶者：今天的人类水平，已经把AI的存在计算在内了。当AI追上昨天的横杆，今天的横杆已经被AI自己的出现抬高。这一层要立的是：人类水平不是一个高度，而是一个含着追赶者的发生过程。判错的办法也很具体：如果AI追上某一条横杆之后，人们公认的"人类水平"清单长期不再增补，横杆停在了原处，这一层就站不住。

这一层里有一个我必须正面对待的反例，就是象棋。1997年深蓝战胜了世界冠军卡斯帕罗夫，此后顶尖棋手在正式较量里再没有占过明显上风。这是确凿的，我不绕开它。第八编会说明：封闭的领域，规则固定，边界清楚，被永久越过是可能的；开放的领域不行。而"通用"这个词，恰恰意味着开放。

第三层　没有可装满的智慧（第七编）

这一层回答："智慧是不是一个可以被装满的库存？"

这里有两种图景。一种是库存图景：智慧像仓库里的货，知道得越多，货越多，将来某一天仓库装满了，就是通用智能。另一种是共生图景：智慧和创造力是一起长出来的，没有创造，就没有智慧；每创造一次，创造者本人也被改变了，多了一个从前没有的自己。回到前言里的那篇致辞：你在改它的时候，同时也发生了

变化。那句你想对表哥说的话，是写的过程生出来的，不是写之前就躺在你脑子里等着被取出来的。

第七编用两个日常的场景把这件事讲透：新手父母的第一个夜晚，和番茄炒蛋里的第一勺糖。最后给出一个检验的判据，并老老实实说明它的边界。

第四层　三界工具只会提升三界智慧（第三、四、五、六编）

这是篇幅最大的一层，因为它是证据的主体。

我把世界分成三界：现实界，是眼前的东西、能摸到的东西；理念界，是概念、数学、理论这些看不见却起作用的东西；自我界，是人对自己的认识与塑造。AGI这个词本身就住在其中一界：第3章会说，它首先不是现实界里的一台机器，而是理念界里的一座塔，导论里会先带你走一遍。我在这里先说一个很自然的想法："理论也是工具"。一个好的数学方法，和一把好用的锄头一样，是把已经弄明白的事封装起来，让后来的人拿去用。

第三编讲现实界的工具：火、犁、时钟与铁路、蒸汽机、电力与洗衣机、计算机与网络。第四编讲理念界的工具：文字、零、几何与逻辑、微积分、概率论，最后讲到AI做数学。第五编讲自我界的工具：镜子、日记与书信、修行法门、心理学、自拍与社交主页。第六编讲三界如何交织，并回答两个最重的反驳：一，工具让人退化了吗？二，毁灭性的工具怎么办？

这一层的要点是，一个工具从它出生的那一界出发，很少只停在那一界。时钟本来是现实界的东西，可它改变了人对"准时""效率"的想法，也改变了人怎样安排自己的一生。这就是承重命题之一："一界有工具，三界有新生。"

第五层　AGI感是时间差现象（第十编）

这是整座桥。它回答："既然终局AGI不可能，为什么这么多人觉得它快来了，甚至已经来了？"

答案是：每一次工具出现，都有一段"封装已完成、新问题尚未显露"的时间差。在这段等待期里，旧问题被解决得很漂亮，新问题还没有露头，于是人们很自然地以为，问题已经被解决完了。第十编的定音句是：AGI不是一个终点，而是一段等待；在旧问题已经消失、新问题还未显露的时候，人们把看不见误当成了没有。

这一层也给出了实用的结论：既然AGI感是时间差，那么不同的人、不同的群体，等待的长短就不同。能缩短这段等待的人，是那些更早看见新问题的人。所以全书最后落在发生学哲学：它为什么应当成为每个人的主科。

◆ 五、第九编：把前面的材料收成一个证明

第九编不单算一层，它的角色有点特别：把前面的材料收成一个结构证明。设想有一个终局AGI，它必须能处理一切问题，包括由它自己的存在而引发的新问题，而这恰恰是它做不到的，因为这些新问题是在它被造出来之后才发生的。也就是说，终局AGI这个概念，一旦认真对待，就会把自己否定掉。

第九编还要处理两个绕不开的追问。第一个：基础问题自己也在发生，比如"什么能被计算"这样的底层问题，是在近代才被清楚提出的；而AI恰恰制造了它自己解决不了的基础问题，比如对齐、可解释性、机器意识。第二个：问题是对谁而言的问题？我把"问题总是对某个主体而言的问题"明确立为前提，并且在书里写明，这是一个前提，而不是被证明的结论。

◆ 六、一页纸看完全书的走向

如果你只有五分钟，看这张对照就够了：十条承重命题，各自落在哪一编、属于哪一层。

第一至三条（工具是已发生智慧的封装；使用即碰撞，碰撞必生新SDE；错位定理）立在第二编，属第一层，是整个证明的骨架。第四条（一界有工具，三界有新生）贯穿第三至六编，属第四层，是证据的主体。第五条（被封装能力的退场是让渡，不是退化）在第二编第9章立起，第六编回应"工具让人退化了吗"时再用。

第六条（创造力与智慧共生，智慧不是先在的库存）在第七编，属第三层。第七、八条（人类水平是过程，横杆里含着追赶者；封闭领域可被永久越过，开放领域不可）在第八编，属第二层。第九条（基础性本身在发生，AI制造了自己解决不了的基础问题）在第九编，把前几层收成结构证明。第十条（AGI感是维度未配平时期的时间差现象）在第十编，属第五层。

第一编不属于任何一层，它只做三件事：立靶子，交代方法，列出可错条件。

◆ 七、读法路线，与每章怎么写

这本书可以从头读到尾，也可以挑着读。想先看结论的人，读完导读和导论，直接翻到第4章，看方法和可错条件，再读第二编。想先看故事的人，可以从第三编的火和洗衣机读起，需要原理的时候再回第二编。想先弄明白"我为什么会觉得它快来了"的人，先读第十编，再回头读第二编。

每章的写法都一样：先从一件日常的小事开场；再把背后的道理用一句话说出来；然后请出历史上的例子作佐证；回扣到某一条承重命题；老老实实交代这一章的论证边界，以及它在什么情况下是错的；章末有一段"带走的话"，和一个两行的"小账"；最后一行是承接或预告。每一编的开头有一页"编序"，告诉你这一编有几章、各章做什么、站在哪一层论证上。

◆ 八、推翻办法：去哪里查

这本书允许被推翻，怎么推翻，写在两个地方。

第一处是第4章第七节"全书的可错条件"。那里有三条可以观察的预言：任何开放领域，在AI能力大幅跃升之后的数年之内，会出现由这次跃升本身引发的新职业或新的问题类别；社会上的AGI感，在扩大期最强，进入内卷期之后会明显下降；无论哪一次宣称"AGI已经实现"，此后被公认的前沿基础问题清单，只会变长，不会变短。第一条预言还划了一个窗口，从那次跃升算起三到八年，窗口过了还没有出现，就按落空算。那里还有四处结构论证自己的可错点：使用必然碰撞出

新结构这一前提，错位能否被预先模拟消除，问题对主体而言这一前提，以及开放与封闭的划分；也有几类事先写明"不算反驳"的情形。

第二处是附录C。它把三条可错预言连同判定办法收在一处，将来对账时直接翻到那里。

除此之外，每一章里都有一节"这个论证的边界在哪里"，其中都写着一个具体的找法：找一件什么样的工具、查什么样的记录，如果查出来是什么样子，这一章就错了。想自己动手检验的人，可以按这个顺序：先看第4章的可错条件，再翻附录C，最后看各章自己的那一条。

◆ 九、随身三问

读完全书记不住那么多，三个问题就够带走。以后看到任何一件新工具，包括下一个被叫作"通用"的东西，都可以问它们：

1. 封装了什么？——它把谁的哪一段"昨天"固定了下来？
2. 逼出了什么？——用起来哪里对不上？撞出了什么说明书里没有的东西？
3. 腾出的位置，谁占住了？——被它接手的活儿让出去之后，人转到了哪个更高的位置？如果位置只是空着，那就不是让渡，是退化。

每章末的"小账"，问的就是前两问；第三问要等读到第9章，才讲得完整。

◆ 十、给读者的三个提醒

第一，请记住：撑住全书的是证明，不是例子。证明是第二编的结构论证，由第九编收束；第三至六编的工具史、各章的日常故事，都是佐证。我在第4章里会专门讲这一点。历史上的所有例子，加起来也证明不了"永远"，因为"永远"是关于未来的话。真正撑住"永远"的，是封装与新问题之间那个时间上的错位。佐证让你觉得这个结构不是空想；证明让你知道它为什么必然如此。两者的主次，我不会颠倒。

第二，请记住：本书反驳的是终局AGI，不是所有关于"通用AI"的说法。第1章会专门区分三种AGI，你会看到，主流的定义其实并没有要求"解决所有问题"，所以我对阶段性的"人类水平通用AI"，另外用第八编的论证去回应。

第三，请记住：这本书讲的"永远"，是一个关于升级的永远，而不是关于停滞的永远。它不是在说"AI不会进步"，而是在说"进步这件事，没有终点"。对一个普通人来说，这既不是坏消息，也不是好消息。它只是一个事实：我们每一个人，仍然站在发生的现场。

好，先读导论，然后我们去看第一件事：我们到底在反驳什么。

目 录

作者介绍 · · · · · iii
前言　它什么都能做的那个下午 · · · · · iv
导读　五层论证地图 · · · · · x

导论　一座塔，和一座塔的工地 · · · · · 1

第一编　靶子：我们到底在反驳什么 · · · · · 5
第 1 章　三种AGI：终局AGI、人类水平AGI、AGI感 · · · · · 7
第 2 章　库存图景：主流AI想象背后的哲学底牌 · · · · · 14
第 3 章　发现学的AGI与发生学的反驳 · · · · · 21
第 4 章　本书的论证方法：历史归纳作佐证，结构论证作证明；全书的可错条件 · · · · · 32

第二编　原理：工具发生律 · · · · · 41
第 5 章　工具是智慧的封装 · · · · · 43
第 6 章　使用即碰撞 · · · · · 49
第 7 章　错位定理 · · · · · 55
第 8 章　二阶碰撞与升级律 · · · · · 62
第 9 章　工具让渡律的四阶段 · · · · · 69

第三编　现实界工具史 · · · · · 77
第 10 章　火：从烤肉到冶金，再到历法与水利 · · · · · 79
第 11 章　犁与农耕：粮食多了，问题更多了 · · · · · 86
第 12 章　时钟、铁路与电报：从时刻表到相对论 · · · · · 93
第 13 章　蒸汽机：工具先于科学，热力学是被逼出来的 · · · · · 99
第 14 章　电力与洗衣机：几十年的等待，以及"省下来的时间去哪了" · · · · · 105
第 15 章　计算机与网络：程序员、软件危机与网络安全 · · · · · 112

第四编　理念界工具史：理论也是工具 · · · · · 120
第 16 章　文字：契约、法律与"和自己对话" · · · · · 122
第 17 章　零与数字：复式记账、公司与家庭账本 · · · · · 129
第 18 章　几何与逻辑：从丈量土地到证明的观念 · · · · · 135
第 19 章　微积分：解决了运动，生出了无穷小危机，一路走到哥德尔 · · · · · 141
第 20 章　概率论：保险、统计与"按概率做决定的我" · · · · · 147
第 21 章　AI做数学：从"你解了哪些题"到"你发生了哪些新数学" · · · · · 154

第五编　自我界工具史 ····· **161**
第 22 章　镜子：自画像与"我的样子" ····· 163
第 23 章　日记、书信与自传：内心是怎样被写出来的 ····· 171
第 24 章　修行法门：认识自己的古老工具 ····· 178
第 25 章　心理学：潜意识，以及新的自我焦虑 ····· 185
第 26 章　自拍与社交主页：被展示的我 ····· 192

第六编　三界交织：一界有工具，三界有新生 ····· **199**
第 27 章　跨界碰撞的机制：为什么工具从不只停在出生的那一界 ····· 201
第 28 章　三界同步升级的结构证明 ····· 209
第 29 章　反驳一：工具让人退化了吗? ····· 216
第 30 章　反驳二：毁灭性工具 ····· 223

第七编　智慧不是库存 ····· **230**
第 31 章　两种图景：取货重组，还是一起发生 ····· 232
第 32 章　日常的共生：新手父母的第一个夜晚，番茄炒蛋里的第一勺糖 ····· 239
第 33 章　SDE的写法：S是显露的稳定态，智慧是创造沉淀下来的显露 ····· 246
第 34 章　一个判据：创造的时候，创造者自己有没有被改变 ····· 253

第八编　移动的横杆 ····· **261**
第 35 章　人类水平不是一个高度，而是一个过程 ····· 263
第 36 章　横杆里含着追赶者 ····· 271
第 37 章　象棋的反例 ····· 277
第 38 章　通用即开放 ····· 284

第九编　终局AGI为何自我否定 ····· **290**
第 39 章　结构证明：越被使用，越不通用 ····· 292
第 40 章　基础问题也在发生 ····· 300
第 41 章　AI制造了自己解决不了的基础问题 ····· 308
第 42 章　问题是对谁的问题 ····· 316

第十编　AGI感：等待期的时间差 ····· **324**
第 43 章　时间差的机制：封装是瞬间的，发生要等待 ····· 326
第 44 章　历史上的AGI感：电力、1965年的预言、信息高速公路 ····· 332
第 45 章　个体差异：AGI感的强弱与碰撞的深浅成反比 ····· 338
第 46 章　群体差异：同一时刻，不同群体在不同阶段 ····· 345
第 47 章　缩短等待：主动碰撞，以及为什么发生学哲学是每个人的主科 ····· 351

结语　没有终点的智慧 ····· **358**

参考书目 · · · · · **363**

附录 · · · · · **374**

附录A　三界工具史年表 · · · · · 374

附录B　十条承重命题汇总 · · · · · 378

附录C　可错预言 · · · · · 380

附录D　术语表 · · · · · 382

导 论

一座塔，和一座塔的工地

◆ 一、一个周日的相册

你的手机里存着八千多张照片，换手机之前，你决定拿一个周日把它们整理好。先按年份建文件夹，再按人、按旅行分。傍晚的时候终于弄完了，每一张照片都有地方可去。你很满意：这一次，一步到位。（这是一个设想的小场景，用来说明结构。）

一个星期之后，你发现整理本身带出了新问题。同一次旅行的照片，该进"旅行"还是进"孩子"？家里人想要的是"某一年的合照"，你的分类里偏偏没有"合照"这一类。外婆一口气发来三百张截图，要不要为她新建一个夹子？更微妙的是，你再拍照的时候，会先在心里过一遍"这张该放哪个夹子"，连拍法都被分类法改了。

整理并没有做错。错的是那个念头：以为"整理完"是一个可以到达的地方。分类法一旦用起来，它自己就成了一个新东西，在自己的外面长出了新的问题。

这本书讲的，就是这样一个很小的道理，在很大的地方的样子。

◆ 二、这本书的路线

书名里的"永远"，针对的是一种叫作"终局AGI"的东西：设想有一个系统，智慧已经装满，再没有哪个问题需要更高一级的智慧去对付。第1章会把AGI这个词拆成三种意思，本书击中的只是这一种。对另外两种，一种是"人类水平的通用AI"，一种是"用着它觉得通用已经到来"的那份感觉，本书各有各的回应。

路线很短，可以用三个词说完：封装、碰撞、错位。工具把已经发生的智慧封装起来（第5章）；它一被使用，就和活的处境碰撞，碰出封装里没有的新东西（第6章）；这些新东西在封装的那一刻还不存在，所以必在封装之外，这就是错位（第7章）。接下来，错位逼出更高一级的智慧（第8章），在产业层面走出扩

大、内卷、新产业诞生、让渡四个阶段（第9章）。这条主线是第二编的事，也是全书的证明。

证明有了，还要看它在真实的世界里是不是这个样子。所以我把世界分成三界：现实界，是眼前能摸到的东西；理念界，是概念、数学、理论这些看不见却起作用的东西；自我界，是人对自己的认识与塑造。第三、四、五编分别去三界里查工具的历史，第六编看三界怎样彼此牵连。再往后，第七编问智慧是不是一座可以装满的仓库，第八编问"人类水平"是不是一个固定的高度，第九编把这些收成一个结构证明，第十编解释人们为什么会觉得AGI已经来了。五层论证各自站在哪一编，导读里有一张图，这里不重复。

在这条路线上，我想先在门口提出一个新的看法，它是整本书的一个入口。

◆ 三、理念界的巴别塔

很多人听到"AGI"，脑子里浮现的是一台机器，一个越来越强的系统。我想换一个看法：AGI首先不是一台机器，而是理念界里的一座塔。

先想想巴别塔的故事。《创世记》第十一章说，那时天下人的口音言语都是一样的，他们商量着要建一座城、一座塔，塔顶通天。后来口音被变乱，人听不懂彼此，工程停了。这个故事有很多种读法，本书不参与争论，只取它的结构，三件事。第一，人们想要一个统一的、收口的东西：一种语言，一座塔。第二，这个东西的目标是"通天"，一个再没有更高处的顶点。第三，也最要紧：越是认真地建，建造的过程本身就越生出分歧。

把这三件事对到AGI上，对得相当整齐。"通用"对应那一种统一的语言：不再有一个个分开的专门系统，而是一个东西什么都能做。"解决一切问题"对应塔顶通天：到了那里，没有更高处，没有更难的问题。而建造中的分歧，对应新的SDE：越是认真地使用这件工具、把它推向"通用"，使用者与它、与别的行业、与彼此之间，就越是碰撞出此前不存在的新问题、新行业、新分工。

日常里也有这样的剧情。一家公司决定上一套"统一系统"，把销售、仓库、财务全接进同一个平台，从此一个系统管所有事。上线之后，每个部门都多出一个专门和系统打交道的人，多出"这个数该按谁的口径算"的争论，多出一批只有在这套系统里才听得懂的词。系统越想让所有人说同一种话，工地上越是长出新的方言。（这同样是设想，用来说明结构，不指哪一家公司。）

机器属于现实界，可以被造出来、被测试、被关掉，它有具体的能力和具体的边界。而"通用""终局""一切问题"这些词，属于理念界：它们是被构想出来的形状。现实界里造出来的，永远是某一个具体的系统；"那个可以收口的智能"，是理念界里发明出来的形状，然后被人们当成现实界里的施工目标。这个区别可以检验：机器的能力可以验收，给它一份考卷，答对多少，有个数；塔顶没法验收，因为"一切问题"这张考卷，只能由已经出现的问题来出，而造出这个系统、使用这个系统，本身就在往考卷之外添题。所以人们能做的，只是一次次宣布"这一层通了"，不能宣布"顶到了"。

塔每高一层，工地上就多出一批新的工种、新的争论、新的语言：巴别塔不是被推倒的，是在建造中自己长出了离心力。这个想法在第3章第九节有更完整的写法，也写明了它的边界：巴别塔是神话，不是证据，我只拿它当结构上的比方；说AGI是理念界的发明，也不是说它是假的或没有价值。恰恰相反，理念界的发明往往是最有力的工具，它给建造者一个方向。

◆ 四、塔，和塔的工地

塔和工地，是同一件事的两面。塔在理念界，是那个"一统""通天"的形状；工地在现实界和自我界，是每天真正发生的事：一批批新的系统被造出来，一批批人开始使用它们，一批批工种、争论和说法，从使用里长出来。

想想那些一心要"一步到位"的人，装修的，整理相册的，上统一系统的公司。他们的图纸往往画得很漂亮，错的不是图纸，是"一步到位"这个想法本身。塔越高，站在塔下的人越觉得天就在眼前，而他们看不见的，是塔身两侧正在长出去的那些新枝杈。AGI感，正是站在塔下的感觉。

工地上长出来的东西，也不都是麻烦。巴别塔的人们没有白忙，他们发明了怎样烧砖，怎样用沥青黏合；人类为了"通用智能"这座塔，也确实造出了一批前所未有的工具。塔的副产品，正是人类的又一次升级。这也是本书不悲观的原因：永远到不了塔顶，不等于工地上没有活，恰恰是因为到不了，活才永远有。

还有一点，请读者留意。塔是被人发明的，也是被人拿去施工的。"通用智能"这个理念，本身就封装着人们对"智能是什么"的昨天的理解，正在被大家使用，所以它也逃不出本书的路线：被使用，就会碰撞，就会长出它自己的外面。一座塔，被自己的工地一层一层地改写。

◆ 五、再说一遍这本书的主张

最后，请允许我把整本书的主张，用最短的话再写一遍。

> **AGI不是智慧的终点，而是又一次封装；每一次封装，都是下一次发生的起点。**

这句话我叫它"判词"。读完导论，你已经知道它的三个来路：它来自一个很小的道理（整理完不是一个能到达的地方），来自一条很短的路线（封装、碰撞、错位），也来自一个新的看法（终局AGI是理念界的一座塔）。接下来的十编，就是替它一层一层地作证。

这个看法需要读者自己对照后面各编的论证去检验，不该当成不证自明的定义。如果读到最后，你觉得塔顶其实是到得了的，那么请回到第4章，看看我替自己写下的那些"如果……就是我错了"。

好，现在去看第一件事：我们到底在反驳什么。

第 一 编

靶子：我们到底在反驳什么

编 序

第一编一共四章，不搭论证，只做三件事：把靶子立好，把方法交代清楚，把全书可能错在哪里提前写在纸上。读到后面几编觉得迷路的时候，回到这一编，就能找到出发的地方。

四章各做一件事。第1章把"AGI"这个词拆开，拆成终局AGI、人类水平AGI、AGI感三种，说清本书击中的是哪一种、对另外两种怎么办。第2章翻开主流AI想象背后的"库存图景"，看它的四张底牌，也看它在哪里是对的。第3章翻开第一张底牌，区分发现学与发生学，并说明终局AGI首先是理念界里的一座塔，不是现实界里的一台机器。第4章交代方法：结构论证作证明，历史归纳作佐证；补上"谁来升级"和象棋两道缝，列出全书的可错条件。

章	要回答的问题	一句话
第1章	"AGI"到底指什么	三种AGI，本书击中的是第一种
第2章	它为什么听起来顺理成章	库存图景的四张底牌
第3章	第一张底牌对不对	问题是发生出来的，不是等着被发现的
第4章	凭什么敢说"永远"	结构论证作证明，并写明可错条件

第一编不属于五层论证里的任何一层，它是论证之前的准备。但后面每一层，最后都要回到这里立下的靶子和写下的可错条件上来检验。

第 1 章

三种AGI：终局AGI、人类水平AGI、AGI感

◆ 一、一顿饭上的三场对话

周末聚餐，桌上坐着三个人，话题不知怎么转到了AI。

大哥放下筷子，语气笃定："通用人工智能？早到了。你让它写文章、写代码、翻译、讲题，哪样不会？还要怎样才叫通用？"

对面的表姐摇头："不可能的。人是有灵魂的，机器永远是机器，它不可能真正像人一样想事情。"

坐在角落里的小侄子在大学读计算机，慢慢开口："其实现在还差得远。它在很多事情上看着聪明，可是你给它一个没见过的新情况，它就露怯。真正的AGI，得等到它能自己学会任何事才行。"

三个人说得都很起劲，可我听了一会儿，发现他们根本不在一个问题上吵架。大哥说的"通用"，是指它在很多常见任务上都够用；表姐说的"不可能"，是指机器无法拥有某种内在的东西；小侄子说的"还差得远"，是指它离"学会任何事"还有距离。三个人用的是同一个词，指的是三件不同的事。

这一章要做的，就是把这三件事拆开。因为一本以"AGI为何永远不可能"为名的书，如果不先说清楚"AGI"指什么，就会像饭桌上那场争论一样，各说各话，谁也说服不了谁。

我在这里先把话说在前头：AGI这个词，至少指三件不同的事——终局AGI、人类水平AGI，以及AGI感。 本书对这三件事的态度各不相同，所用的论证也各不相同。

◆ 二、第一种：终局AGI

先说最强的那一种，我称它为终局AGI。

它的意思是：存在（或将会存在）这样一个系统，它的智慧已经"装满"了。凡是人类已经遇到的问题，它都能解决；凡是将来会出现的问题，它也都能解决；再也不会出现一个需要更高一级智慧才能对付的问题。用大白话说，它是智慧的终点站：到了这里，旅途结束，不用再往前走了。

这样的设想，在科幻作品里很常见。一个无所不知的超级大脑，一台能回答一切问题的机器，一个"最后的发明"。人们谈论"奇点"、谈论"超级智能"的时候，有时指的就是这一种。

终局AGI有两个特征。第一，它是关于"全部"的：所有问题，包括未来才出现的问题。第二，它是关于"终点"的：到了它这里，智慧不再需要升级。这两个特征合起来，就是我们这本书要反驳的东西。本书的全部结构论证，主要都是冲着这一种来的。

它也是最容易被拿来当稻草人的靶子。不少争论里，反方花很大力气证明“机器不可能解决所有问题”，正方一句“谁说AGI要解决所有问题了”就挡了回去。这句挡回去的话是对的，因为多数人说的AGI根本不是第一种。这引出下面最重要的一点。

◆ 三、第二种：人类水平AGI，以及"定义缝"

第二种是目前最常被讨论的版本：人类水平的通用人工智能。

它的常见说法大致是这样的：一个AI系统，在大多数（或绝大多数）有经济价值的、需要认知能力的任务上，达到或超过普通人，甚至超过专业人士的水平，并且能把这种能力迁移到新的任务上，不用为每个任务单独重造。（以上是我综合各家说法的转述，不对应某一家的原文，各机构与学者的具体表述并不相同。）

仔细看这个说法，你会发现里面有几个关键的词：大多数、达到人类水平、迁移到新任务。它并没有说"解决所有问题"，没有说"没有任何问题能难倒它"，也没有说"智慧已经装满"。它说的是：和人比，它在大多数事情上不输。

我把这个区别叫作"定义缝"，意思是：许多对AGI的反驳，都从这条缝里漏了过去。

打个比方。你家请了一位非常能干的家政阿姨，做饭、收拾、带孩子、辅导作业、照顾老人，样样拿得出手。你说她是"通用的家政人员"，没人觉得不对。可是，这位阿姨能解决所有的家庭问题吗？显然不能。孩子突然查出一种罕见的过敏，老人需要一个复杂的医疗决定，家里要不要搬到另一个城市去，这些问题不在她的能力里，也不在大多数人的能力里。我们并不因此说她"不通用"。"通用"的意思，从来不是"万能"，而是"在一大片常见的事情上，不需要为每件事单独去找人"。

把这个道理用到AI上：一个AI系统被称为"通用"，只要求它在广阔的一片任务上够用，不要求它没有盲区。如果有人拿一个"总有它做不到的事"来反驳这样的AGI，那就是把靶子立错了地方。这种反驳不仅无效，而且会让人觉得，反方是在回避真正的问题。

所以，我在这里要把话说得非常明确：第二种AGI的定义，并不要求"解决所有问题"；因此，仅仅指出"AI总有做不到的事"，既不能反驳它，也不是本书的论证。 本书击中的是第一种，即终局AGI；对第二种，我会另用一个论证，放在第八编。

那个论证的要点，先透一句：不是说人类水平的AI不会出现，而是说“人类水平”本身是一条会移动的横杆，而且横杆里已经含着AI自己。这要到第八编才展开。

◆ 四、第三种：AGI感

第三种和前两种性质不同，它不是一个关于AI的断言，而是一种关于人的感觉。

我把它称为AGI感：当一个人使用某种AI工具，发现它几乎什么都能做的时候，心里升起的那种"通用智能已经到来"的感觉。

AGI感是真实的。前言里我写到过，第一次用AI写完一篇文章时，我自己就有过这种感觉。它不是愚蠢，也不是被广告蒙蔽，而是一种正常的认知反应：当旧问题被一件新工具解决得干干净净，而你还没来得及遇到新问题的时候，眼前的世界就显得"没什么问题了"。

AGI感值得单独成一类，是因为它常常被当作一个"证据"。有人会说：你看，大家都觉得它已经通用了，那它就是通用的。可是感觉毕竟不等于事实。历史上有过不少类似的感觉：电力刚普及的时候，人们觉得它能解决一切；1965年，有人预言二十年内机器能做人能做的一切工作；信息高速公路的年代，也有人觉得信息一旦畅通，问题自然就消失了。这些感觉都是真诚的，却没有一个成为事实。

我把这类感觉叫作"时间差现象"：工具已经把旧问题封装好，新问题还没有显露，这一段空白的时间里，人们容易把"看不见"当成"没有"。这一点，是第十编的主题。

这里需要先说明一个关键的态度：本书对AGI感的态度，是解释它，不是嘲笑它。AGI感有它的道理，也有它的位置。我们不需要说"感觉到AGI的人是错的"，只需要说"这种感觉是一个阶段性的现象"，它来自一个可以被说清楚的结构。

◆ 五、三种AGI，三种态度

现在可以把三者放在一起看了。我用一个小表，把它们的区别列出来。

种类	它在说什么	本书的态度	主要论证位置
终局AGI	智慧装满、一切问题都能解决、不再需要升级	永远不可能	第二至七编、第九编
人类水平AGI	在大多数认知任务上达到或超过人类，并能迁移	阶段性可能，但"人类水平"本身会移动	第八编
AGI感	使用时觉得通用智能已经到来的感觉	真实存在，是时间差现象，可以解释	第十编

这张小表是本书的"导航图"之一。以后每当读者感觉论证有点散的时候，可以回到这张表，看看当前这一段在回答哪一行的问题。

你可能会问：为什么不干脆只反驳第二种？毕竟第二种才是人们真正在谈的。有两个原因。第一，两种说法在日常讨论里经常滑动：人们起初说“人类水平”，说着说着变成“什么都能做”，再说下去变成“终极智能”。这种滑动是语言的惯性，不是恶意；不先把第一种单独拎出来，滑动就一直在。第二，第一种恰恰最有哲学分量。它背后有一整套关于“智慧是什么”的想象，我在下一章叫它“库存图景”。要反驳它，不仅要说“事实上做不到”，还要说“结构上为何不可能”，这正是本书的主体。

◆ 六、一段必要的诚实交代：这个区分的边界在哪里

写到这里，要停一下，把这一章的论证边界交代清楚。

边界之一：三种AGI的区分，是我为了讨论的清晰而做的整理，它并不是学界公认的标准分类。业内的人对AGI有许多不同的定义，有的比我说的第二种更强，有的更弱。如果读者心中的AGI恰好介于第一种和第二种之间，比如"能够在几乎所有领域超过所有人类的系统"，那么本书的论证对它的覆盖是不完整的：它比第二种强，却未必达到第一种的"装满"。对于这类中间形态，我的立场是：只要它被设想为"一旦达到就不再需要升级"，它就落入第一种，被第二至七编、第九编的论证击中；如果它不作这个设想，它就落入第二种，由第八编处理。

边界之二：有人会批评说，这样划分等于给自己留后路，无论对方怎么定义，我都能说“你说的是第二种”。我的回答是：每一种都有各自的论证，也有各自的失败条件。第4章会把可错条件一条一条列出来，凡是本书断言“会发生”的，都写明了“如果没有发生，我就错了”。这个区分是责任的划分：对不同强度的主张，我承担不同强度的论证。

边界之三：我所说的"不可能"，针对的是第一种。对第二种，我说的是"可能，但不稳定"。这一点有时会被误读成"本书承认AGI会到来"，这并不准确。准确的说法是：本书承认类AGI的能力可能出现，却否认它是终点。

边界之四，也是这一章最容易被查证的地方：我说主流的AGI定义并不要求“解决所有问题”，这句话可以去核对。找十份被广泛引用的定义来读，如果其中多数明文要求“没有盲区”或“一劳永逸”，那么我说的定义缝就不存在，第二种该并入第一种来反驳，这一章的分家也就白做了。

◆ 七、回扣：本章落在哪个命题上

本章看似只是在做定义的整理，却已经触到了两条承重命题。

其一，第十条：AGI感是维度未配平时期的时间差现象。“维度未配平”的意思是：工具把一部分事情做完了，另一些部分还没跟上，整体看起来像“已经完成”，其实是“还没长出来”。第十编会展开。

其二，第三条：错位定理，新SDE必在封装之外，因为封装时它还不存在。第一种AGI所设想的"装满"，恰恰违反了这一条。

◆ 八、小结

这一章要说的话，其实可以浓缩成三句。

AGI这个词指三件事：终局AGI、人类水平AGI、AGI感。主流的AGI定义并不要求解决所有问题，所以"AI总有做不到的事"不是对它的有效反驳。本书所反驳的，是那个设想"智慧可以装满"的终局AGI；对人类水平，我将另用论证；对AGI感，我选择解释而不是嘲笑。

饭桌上那三个人，其实并没有谁错。大哥说的是第二种，表姐想的是别的东西，小侄子惦记的是接近第一种的那个远方。只要把词拆开，他们之间的争吵就会变成分工：每个人各负责一块。

> **带走的话**
>
> 吵"AGI"之前，先问一句：你说的是装满了的终点、不输给人的水平，还是用着顺手时的那个感觉？本书只对第一个说"永远"。

> **小账** 这一章立了什么？——把AGI拆成终局、人类水平、AGI感三种，说明本书击中的是第一种，对后两种另有论证。　　错在什么情况下？——找十份被广泛引用的定义，若多数明文要求"没有盲区"或"一劳永逸"，定义缝就不存在，分家白做。

下一章预告：第2章将追问，为什么第一种AGI听起来那么自然？因为它背后站着一整套关于智慧的哲学想象，叫作"库存图景"。

第 2 章

库存图景：主流AI想象背后的哲学底牌

◆ 一、背完两千个单词的孩子

邻居家的孩子，初二，英语老师要求背单词。他背得很认真，手机里的软件每天打卡，连续一百多天没断过，词汇量据说已经过了两千。他妈妈很高兴，觉得再背一阵，英语就"装满"了，到时候出国旅游、看原版书都不成问题。

可是有一次，家里来了一位外国朋友，孩子站在一旁，想打个招呼，嘴巴张了半天，一句话也没说出来。事后他很沮丧："单词我都认识，可是话到嘴边，就不知道该拿哪个。"

这个场景，很多学过外语的人都经历过。它背后藏着一个我们几乎从不怀疑的想法：能力是一种存量。背的单词越多，语言能力就越大；背到一定数量，就算"满"了。

这个想法在日常生活里太自然了，以至于我们很少停下来问一句：它对吗？在本书里，我给它起了一个名字，叫库存图景：把智慧想象成仓库里的存货，知道得越多、存得越满，智慧就越大；仓库装满了，就是通用智能。

这一章要做的事，是把库存图景从日常的直觉里拎出来，看看它背后究竟压着哪几张哲学底牌，又为什么说：正是这几张底牌，让"AGI终将到来"听起来像一件顺理成章的事。

◆ 二、库存图景长什么样

先说一个澄清：库存图景不是某一位科学家的学说，也不是某一家公司的立场。它更像是一种弥漫在许多人头脑里的背景想象，多数人并不会明说，但他们的判断往往以它为前提。这也是我要把它叫作"底牌"的原因：牌桌上没有人亮出来，可是每个人都在照着它出牌。

库存图景的基本画面是这样的：

世界上有一大堆问题。每个问题，都有（至少在原则上）一个答案，或者一种解法。智慧，就是掌握这些答案和解法的能力。掌握得越多，智慧越高。一个人的智慧是他脑子里的存货，一个文明的智慧是它图书馆里的存货，一个AI的智慧是它参数里的存货。随着时间的推移，存货只增不减；终有一天，存货多到涵盖了所有的问题，那一刻就是"通用"的到来。

这个画面里，有四张底牌。我把它们一张一张翻开。

第一张：问题的全集是先在的

库存图景默认，所有的问题是预先摆在那里的，就像一座大楼里所有的房间都早已存在，我们只是一间一间走进去。科学家的任务是"发现"，工程师的任务是"解决"，AI的任务是"更快地发现、更快地解决"。至于未来会不会冒出此刻想不到的新问题，这个图景没有给它留位置：就算有，也不过是"还没走到的房间"。

这一张底牌最关键，也最隐蔽。第3章会专门用它来引出发现学与发生学的分野，这里先点到为止。

第二张：智慧是可以存放的东西

第二张底牌说，智慧是一种可以被存放、被拷贝、被搬运的东西。你把一个人的本事写成书，另一个人读了书就有了本事；你把一个专家的经验灌进系统，系统就有了专家的水平。智慧像一桶水，可以从这个桶倒进那个桶。

这张底牌在计算机领域尤其有吸引力。因为在计算机里，一切东西——数据、程序、模型——确实都是可以复制的。久而久之，人们就把"可复制"这个属于信息的特征，悄悄地搬到了智慧身上。

可是回想那位背单词的孩子。他的词汇表可以复制到任何一个人的手机里，但是"开口"这件事，却无法复制。词汇是可存放的，开口的能力却不是一桶可以倒来倒去的水。它长在一个人与具体场合、具体对象的往来之中。

哲学里早有一对相近的区分：知道“是什么”，和知道“怎么做”。英国哲学家赖尔在1949年的《心灵的概念》里把它讲透了；波兰尼后来又说，我们知道的比我们说得出的多。本书和他们相同的地方是：能力不是能整桶倒出来的存货。不同的地方在后面：本书进一步说，“怎么做”是在做的过程里发生的，而且每发生一次，做的人本身就变了。这一步要到第七编才走。

第三张：智慧与主体可以分离

第三张底牌说，智慧可以脱离谁的智慧、为谁所用的问题。有"一个"智慧，放在哪里都一样；它是客观的，中立的，不依赖于任何具体的人。于是我们可以谈论"智能本身"，把它从人身上剥下来，装进一台机器。

这张牌的意思不是说"机器不能有智慧"，而是说：智慧可以和"某个主体面对某个处境"这件事无关。一个问题之所以是问题，是因为有人在某个处境里遇到了它；但库存图景往往把这一层略去，只剩下"问题"和"答案"。这一点，第九编还要重新拾起来。

第四张：进步是对上限的逼近

第四张底牌说，进步是一条向上限逼近的曲线。存货在增加，离"满"越来越近，就像进度条在爬。每一次模型变大、数据变多、算力增强，就是进度条往前爬了一格。所谓"通用人工智能还差多远"，就是在问进度条还剩多少。

这张牌在日常里也随处可见：学习进度条、游戏通关度、"你已经掌握了百分之七十的内容"。进度条的前提是，终点是已知的，路是有限的。可是智慧这条路，有没有一个已知的终点，恰恰是要问的问题。

◆ 三、库存图景在哪里是对的

我不想把库存图景当作一个笑话来讲。它之所以流行，是因为它在一大片地方是对的。

在封闭的领域里，库存图景几乎完全成立。象棋就是典型。棋盘的大小、规则的内容、胜负的判据，全部固定。局面再多，也是一个有限的、事先划定的集合。在这样的领域，"存货越多越强"是真的，"总有一天超过人"也是真的。深蓝在1997年战胜卡斯帕罗夫，就是这个领域里存货与算力一路增长的结果。这一点我不回避，第八编会专门作为"象棋反例"来处理。

在许多日常的查询类任务里，库存图景同样好用。查一条成语的出处，问一个城市的人口，算一道题的步骤，这些需要的就是存货，而且存货越足越好。一部好用的字典，一个大容量的知识库，确实是库存图景的胜利。

所以，我对库存图景的批评，并不是"它全错了"，而是"它被推广到了不该推广的地方"：从封闭领域推广到开放领域，从查询推广到创造，从工具推广到智慧本身。

◆ 四、历史上的佐证：库存图景一次次碰壁

库存图景并不是今天才有的。它在历史上多次出现，每一次都遇到了同样的麻烦。我挑三个例子，作为佐证。注意这是佐证，不是证明；证明要留到第4章与第九编。

第一个：希望把数学装满一个系统。　二十世纪初，不少数学家怀着一个宏大的愿望：把全部数学奠基在一套明确的、没有矛盾的公理系统之上，使得凡是真的数学命题，都能从这套公理推出来。这是库存图景在数学里的样子：把数学的全部真理，装进一个完备的仓库。可是罗素发现的悖论，先敲响了第一次警钟；到了1931年，哥德尔的不完备定理给出了更深的回答：在足够强、无矛盾、公理可以一条条列出来的形式系统里，总有一些命题，系统自己既证不了、也否不了。这不等于"数学不可知"，而是说：任何一个被固定下来的系统，都留下了系统外的东西。数学并没有因此枯萎，反而因此长出了新的分支。

第二个：二十世纪六十年代的乐观预言。1965年，赫伯特·西蒙预言，二十年之内，机器将能够做人所能做的任何工作。这句话背后，就是库存图景：人的工作，不过是一堆可以拆解、可以存进去的技能；把技能一项一项存进去，机器就等

于人。这个预言没有如期兑现。我不是说他"愚蠢"，而是说：他看到了存货的增长，没有看到存货之外的东西也在增长。

第三个：软件工程里的“危机”。1968年，北约在德国举办的一次软件工程会议上，“软件危机”这个词叫开了：程序越写越大，工期总超，质量总差。此后几十年，每一代工具吃掉上一代的困难，又带来这一代才有的新困难。库存图景说问题终会被解决完，现实是问题的类型一直在换。第7章还会回到这一例。

三个例子，三个领域，一个共同点：存货在增加，而存货之外的东西也在增加，这个"之外"并不因为存货的增长而缩小。

◆ 五、为什么库存图景让AGI显得顺理成章

现在可以回到AGI了。库存图景的四张底牌一翻开，终局AGI的画面几乎自动浮现：

既然问题的全集是先在的，那么存在一个"全集"，装满它就是终点；既然智慧是可存放的，那么可以把它存进机器；既然智慧与主体可以分离，那么机器的智慧和人的智慧在性质上没有差别；既然进步是对上限的逼近，那么只要把模型做得更大、数据喂得更多，就是在爬进度条。

四张牌合在一起，就得出了结论：AGI只是时间问题。

反过来说，如果你想真正质疑"AGI只是时间问题"，不是在结论上争论，而是要回到这四张底牌上去，看看哪一张可能是错的。这也就是本书的做法：第3章翻开第一张牌，第七编翻开第二张，第七编与第九编合力翻开第三张，第二编与第八编翻开第四张。

◆ 六、与库存图景相对的，是共生图景

那么不是库存图景，又该是什么？我先给出一个预告。

在本书里，与库存图景相对的叫共生图景。它的核心想法是：智慧不是先于创造而存在的存货，创造力与智慧是共生的。一个人在做事的过程中，同时获得了做

这件事的智慧；智慧是在做事的过程里长出来的，而且每长一次，做事的人本身也不同了。

回到那位背单词的孩子。在共生图景下，"开口"不是词汇存量的自然结果，而是在一次次真实对话中长出来的东西：他有一次讲错了，被人听懂了或没被听懂，从这次碰撞里他知道了点什么，下一次就不一样了。词汇是他的材料，开口才是他的发生。

共生图景要到第七编才系统地展开。这里先放一个基本的区分：库存图景问"你有多少"，共生图景问"你在变成什么"。

◆ 七、这一章的论证走到哪里为止

这一章的论证有五处边界。

第一，库存图景是我对一种背景想象的概括，不是对任何具体人物观点的复述。很多严肃的研究者并不持有这么朴素的看法，他们对"智慧不是单纯的存货"早有自己的理解。我批评的，是那种在公共讨论里默认成立的版本。如果某位读者认为我概括得不公道，请把我的概括看作"一种理想类型"，而不是对某人的指控。

第二，四张底牌不一定同时成立。有的人可能只持有其中两三张。我把它们并列，是因为它们在"AGI只是时间问题"这个结论里是互相支撑的。

第三，对库存图景的批评，不是对"大规模存货"的批评。大规模存货很有用，我也不反对继续扩大。我只是说：存货不等于智慧，更不等于智慧的终点。

第四，历史上的三个例子，只是佐证。数学的不完备、预言的落空、软件的危机，都不能直接证明"终局AGI不可能"。它们只是显示库存图景的老毛病：当人们相信存货会装满的时候，现实总是多长出一些存货之外的东西。把这一点升格为证明，是下面几章与第九编的任务。

第五，对库存图景的批评，本身可以被推翻。它需要的反例有一个具体的样子：某个开放领域，里程碑一路靠存货增长达成，此后很多年也没有冒出存货之外

的新问题类别。找到这样一个领域，我对库存图景的批评，在那个领域里就站不住。

◆ 八、小结

库存图景把智慧想象成可以装满的仓库，它背后有四张底牌：问题全集先在；智慧可存放；智慧与主体可分离；进步是对上限的逼近。在封闭领域，这幅图景几乎是对的；在开放领域，它一次次碰壁。AGI显得"顺理成章"，是因为它站在这四张底牌上；要质疑它，不是争结论，而是翻牌。

> 带走的话
>
> 背了两千个单词，不等于张得开口。"装满"听着像智慧，其实只是仓库；有人说AGI只是时间问题，先别争结论，回去翻那四张牌。

> **小账** 这一章立了什么？——库存图景及其四张底牌：问题全集先在、智慧可存放、智慧与主体可分离、进步是对上限的逼近。 错在什么情况下？——找到一个开放领域，里程碑一路靠存货达成，此后多年也不冒出存货之外的新问题类别，对库存图景的批评在那里就站不住。

下一章预告：第3章翻开第一张底牌。问题的全集究竟是先在的，等着被发现，还是不断发生的？这是发现学与发生学的分野。

第 3 章

发现学的AGI与发生学的反驳

◆ 一、拼图与带孩子

先请你想象两件事。

第一件事：拼一幅一千块的拼图。盒盖上印着完整的图案，每一块都有它唯一的位置，哪怕你现在不知道在哪里，它也一定在那里。你拼错了，是你还没找对；你拼完了，是你把所有该找的都找到了。拼图的整个过程，就是一个"发现"的过程：答案早就摆好了，你的任务是把它一块一块找出来。

第二件事：带一个新生的孩子。没有盒盖，没有完整的图案。孩子今晚为什么哭，是饿了、胀气了，还是只是想被抱着，没有人能事先告诉你。更要命的是，每过一个月，问题就换一批：刚学会了怎么哄睡，孩子开始出牙；刚把出牙应付过去，孩子会爬了，家里的插座都变成了问题。你不是在"找"事先存在的答案，而是在跟一个不断变化的处境一起往前走，问题本身在你面前一件一件冒出来。

这两件事，大概就是这一章要讲的两种世界观的日常缩影。

前一种，我称之为发现学：世界上的问题是先在的，一个完整的全集摆在那里，知识就是把它一个个发现出来。后一种，我称之为发生学：问题不是躺在那里等人，而是在主体与世界的往来之中，不断发生出来的。

上一章末尾，我们翻开了库存图景的四张底牌，第一张是"问题的全集是先在的"。这一章就来回答：这张牌，是谁的牌？它如果错了，会怎样？

◆ 二、发现学的世界

发现学的核心主张，可以用三句话概括：

第一，问题的全集先在。不管有没有人去问，那些问题都已经在那里了。

第二，知识是对先在之物的发现。科学家是探险者，数学家是勘探员，每一次进步都是把地图上的空白处填上一块。

第三，进步就是"发现得更多"。随着发现的积累，未知的部分越来越少，总有一天，空白填完了。

这幅图景是如此自然，以至于大多数受过教育的人都会觉得它就是"常识"。我们在学校里学到的科学史，也往往是这样的叙述：某人"发现"了某个定律，某人"发现"了某个元素，某人"发现"了某条规律。这个词用得太顺手了，让人没有注意到它预设了什么。

现在把这个图景放到AI身上，会得到一个相当清晰的推论：

如果问题的全集是先在的，那么存在一个有限的或至少可穷尽的"全集"；AI的任务就是尽快把它发现完；速度比人快、记忆比人好、不知疲倦的AI，迟早会走得比人更远；走到全集的尽头，就是AGI。

我把这种立场下的AGI，叫作发现学的AGI。它不一定是某个具体的人持有的明确观点，却是上一章那四张底牌自然推出来的结论。它的逻辑是严密的：只要前提成立，结论几乎无法逃避。

所以，要反驳它，最好的办法不是在推论上挑刺，而是追问前提：问题的全集，真的是先在的吗？

◆ 三、发生学的世界

发生学的回答是：不是。

发生学的核心主张，同样可以用三句话概括：

第一，问题是发生出来的。一个问题成为问题，需要某个主体，处在某个处境里，用某种方式遇到它。在这之前，它没有作为"问题"存在。

第二，新的问题随着新的结构而发生。每当一个新的东西显露出来——一件新工具、一个新制度、一种新的生活方式——围绕它就会生出一批此前不存在的问题。

第三，进步不是填满空白，而是发生新的空白。解决一个问题的同时，往往会露出下一批问题；而且这些新问题，在解决之前根本看不见。

这里需要用大白话交代一下SDE，因为发生学正是用它来说话的。S，是显露：一个东西露出来、稳定下来的样子。D，是差异：从一个样子变成另一个样子的过程。E，是纠缠：让这件事成为这件事的土壤和场合。发生学说，一切东西都是在某片土壤里（E），经过一个变化的过程（D），才显露出来（S）。所以，当新的S显露时，它必然带着它自己的新土壤、新过程，而在这片新土壤上，会生出此前没有的问题。

用日常的话再说一遍：家里添了一个孩子，家里就多出了一个"孩子"这个新结构；围绕这个新结构，一整批新问题——奶粉、睡眠、早教、学区——才第一次成为了问题。在孩子出生之前，它们不是"已经存在、只是没发现"的问题，它们是被这个孩子的出现发生出来的。

◆ 四、两个史例：铁路与"几点钟"，微积分的"后来者问题"

发生学听起来抽象，但历史上有大量的例子，说明新的问题是随着新的结构发生出来的。我挑一个最具体的：铁路与"几点钟"。

在铁路普及之前，每个城镇各自按太阳来定自己的时间。太阳到了头顶，就是正午。相邻两个镇子之间可能差几分钟，没人在意，因为马车要走大半天，没有哪个人会在一天之内从一个镇子赶到另一个镇子，还必须在两处的钟上对齐时间。"几点钟"这个问题，在那个年代，没有一个"全国统一"的版本。

铁路来了。火车按时刻表运行，一列车要经过十几个镇子，如果每个镇子的钟都不一样，时刻表就根本写不出来，更会有撞车的危险。于是，一个此前根本没人

提出的问题冒了出来：全国的钟，该以哪个为准？1883年11月，北美的铁路公司实行标准时区，这是对这个新问题的回答。

请注意这里的逻辑。"统一时间"这个问题，不是一直藏在某个角落里，等着聪明人去发现。它是在铁路这个新结构显露之后，才第一次成为问题的。在铁路之前，它不是"没被发现的问题"，而是"还不是问题"。

更有意思的是后续。科学史学者伽里森（Peter Galison）的研究指出，在十九世纪末二十世纪初，如何让远处的钟同步，是一个非常实际的技术问题，而这个问题和后来关于"同时性"的物理思考之间，有着深刻的牵连。〔待核：具体论断以其2003年出版的《爱因斯坦的时钟，庞加莱的地图》为准〕也就是说，铁路逼出来的"时间统一"，又进一步逼出了人们对"同时"这个概念本身的追问。一个工程问题，生出了一个哲学与物理问题，而后者在铁路之前，同样是不存在的。

发现学的叙述会说：同时性的问题早就在那里，爱因斯坦只是发现了它。发生学的叙述则说：同时性在那个时代成为问题，是因为时钟、铁路、电报这些新的结构，把它推到了人们面前。两种叙述都可以讲得通顺，但它们对"未来会不会再冒出此刻想不到的新问题"给出了完全不同的回答：发现学说，也许有，但那只是还没走到的房间；发生学说，一定有，因为新的结构还在不断出现。

第二个例子来自理念界，也就是数学。

十七世纪，牛顿和莱布尼茨各自独立地发展出了微积分。这套工具威力惊人：行星的轨道、物体的运动、曲线的面积，一下子都可以算了。它被封装进了一套可以反复使用的符号和规则里，后来的人拿去就用，很快就用出了许多成果。

可是，用了一百多年之后，一个此前没有人认真追问的问题浮现了：这套东西里反复出现的"无穷小"，究竟是什么？一个量，它既不是零，又小到可以忽略，这在逻辑上说得通吗？早在1734年，哲学家贝克莱就批评过这类说法的含混，把消失中的无穷小讥为“已逝量的幽灵”。到了十九世纪，柯西，以及后来的魏尔斯特拉斯等人，才把"极限""连续"这些概念重新严格化。

注意这个次序。严格化的问题，不是微积分诞生那一天就已经摆在桌面上的问题。它是在工具被广泛使用、碰撞出一批新情况（比如某些奇怪的函数、级数的收敛问题）之后，才成为数学界必须正面回答的问题。换一个说法：微积分这件工具，把一批老问题解决了，同时也发生出一批新问题；而这批新问题，在工具被发明时不可能被提出。

发现学会说：这些问题一直在那里。可是即便承认这一点，有一件事是无法绕开的：在牛顿的时代，没有任何一个人、任何一台机器，能够"发现"这批问题，因为产生它们的那些使用场景还没出现。这一点对我们论证AGI至关重要，我稍后回扣。

◆ 五、发现学的AGI，被发生学怎样反驳

现在可以把反驳的线索理顺。

发现学的AGI，建立在"问题全集先在、可被穷尽"之上。发生学的反驳，不是说"AI永远发现不完"，那是一个关于能力的断言，容易被技术进步轻易挡回。发生学的反驳是：这个全集不存在。

具体地说，三点：

第一，新的问题随着新的结构发生，而新的结构又来自工具的使用。这一点是整个论证的发动机。AI本身就是一种新的结构，它的使用必然生出新的问题。所以，一个正在被广泛使用的AI，一直在为自己制造"全集之外"的东西。

第二，这些新问题，在AI被造出来的那一刻，不可能被包含在AI里面，因为那时它们还不存在。这就是导读里提过、第7章要专门证明的"错位"：封装发生在先，新的问题发生在后。

第三，所以，不存在一个"装满了"的时刻。装满的前提是全集固定，而全集在不断长大。

用日常的话说，这就像一位带孩子的母亲，哪怕读完了市面上所有的育儿书，也不能说她"装满了"育儿的智慧。因为明天的孩子会遇到明天才有的问题，而她与孩子一起长出来的新的难处，书里不可能有。

这就是发生学对发现学的AGI的反驳。它一句话就能概括：问题不是等着被发现的，而是不断发生的；所以不存在一个可以被穷尽的问题全集，终局AGI也就失去了它的立足点。

◆ 六、两个必须正面回答的反驳

这个反驳听起来干脆，但它一定会遇到至少两种质疑，我在这里正面回答。

质疑一：数学真理不是先在的吗?

有人会说，数学里的真理难道不是一直就在那儿吗？牛顿的时代，极限的严格定义虽然没人提出，但它是不是一直潜在地成立?

我的回答是：本书并不需要对数学的本体论表态。你可以相信数学真理是先在的，也可以不相信。本书的论证只依赖一个更弱的前提：问题总是对某个主体而言的问题。　一个命题即便在抽象的意义上"早已为真"，它要成为某个主体面前的"问题"，需要这个主体处在一个能够提出它的处境里。而这样的处境，是随着工具与结构的出现才发生的。AGI要做的，是解决人与世界的各种"问题"，而不是持有一个抽象的真理清单。所以，即便抽象的真理是先在的，作为"问题"的那一部分，仍然是发生的。

这个前提——问题总是对某个主体而言的问题——是我明确立下的前提，不是被证明的结论。第九编会再回到它，并标明它的地位。

质疑二：新问题也可以被AI提前"预测"出来吧?

有人会说，AI那么强，能否在新结构出现之前，就把它可能带来的问题推演出来?

我的回答分两层。一层是：对于已知规律的外推，AI确实可以做，而且做得比人好。可是，外推出来的问题，是旧结构的延长线。真正的新问题，来自新结构与

具体处境的碰撞，那个碰撞的细节，在碰撞发生之前，是没有的。第二层：即便AI能预测一部分，它这一预测动作本身，又是一个新的结构，又会带来下一批预测不到的东西。所以，预测的能力不能替代发生的过程。

◆ 七、这一章可能错在哪里

这一章的边界有六条。

第一，发现学并非全错。在封闭领域，比如象棋，或一道有标准答案的习题，"发现"就是对的叙述。发现学是发生学的一个特例：当结构固定、土壤固定的时候，发生就退化成发现。本书反驳的，只是把这个特例推广成全部的世界观。

第二，发生学是一个哲学立场，不是一个可以靠一两个实验证明的科学结论。我用了铁路与微积分作为佐证，它们让这个立场显得合理，却不能把它"证实"。证明的责任，由后面的结构论证来承担。

第三，我所举的两个例子，都经过了后人的整理，叙述里难免带着"后见之明"。比如，说"统一时间是被铁路逼出来的问题"，是一种解释，也有历史学者对时区的推行细节有不同看法。读者可以在此处保留自己的判断。

第四，发生学并不意味着一切皆不可预测。工具使用中的很多问题是可以预见的，也应当预见。发生学只说：总有预见之外的，而且预见之外的那一部分，不会随着预见能力的提高而消失。

第五，这个反驳在邻近的领域里有近亲，我该交代。波普尔在《历史决定论的贫困》（1957年成书）里论证过：科学知识的未来增长，没法用科学方法事先预知，所以依赖它的历史进程也没法预测。考夫曼讲生物圈和技术圈时用过“相邻可能”的说法：每走一步，就打开一批此前不在选项里的新步子。〔待核：相邻可能的出处与表述〕本书和他们相同：未来的问题空间不是预先给定的。不同在于，本书把这个想法压在一个更窄也更硬的地方，也就是封装在先、使用在后的次序，并且拿它去判终局AGI；据我所知，他们没有这样用。

第六，这一章的判断在什么情况下是错的？如果有人在某件新工具大规模使用之前，写下一份它将带来的问题清单，事后逐条核对，大半命中，而且清单不是靠事后补充凑出来的，那么“新问题不可能预先装进去”就要打折扣。只猜中个别问题不算，比如有人事先想到火车会让时刻表变麻烦，这不稀奇。要预见的是一整类问题，包括使用者读到这份清单之后的反应。

◆ 八、一座理念的巴别塔：AGI首先是理念界的一项发明

前面八节问的是“问题全集是不是先在的”。这一节换一个问法：AGI这个东西，到底是哪一类东西？

先从装修说起。很多人装修房子时都动过一个念头：这次一定一步到位。收纳、动线、孩子长大以后的房间、老人将来要扶的把手，全部提前设计好，做一个“终极方案”。图纸画得很漂亮，什么都有了。住进去半年，生活就长出图纸里没有的东西：孩子迷上拼装模型，需要一张永远不能收的桌子；老人来住，储物间成了卧室；一场疫情，书房变成全家的会议室。图纸没有画错。错在“一步到位”这个想法本身：它假定生活是一件可以一次画完的东西。

人类很早就讲过这种想法的故事，就是巴别塔。《创世记》第十一章说，那时天下人的口音言语都是一样的，他们商量着要建一座城、一座塔，塔顶通天，为要传扬自己的名，免得分散在全地上。后来口音被变乱，人听不懂彼此，工程停了，人散往各处。

这个故事有很多种读法，有人读成对骄傲的惩罚，有人读成语言多样性的来历。本书不参与这场争论，只取它的结构，一共三件事。第一，人们想要一个统一的、收口的东西：一种语言，一座塔。第二，这个东西的目标是“通天”，一个再没有更高处的顶点。第三，也最要紧：越是认真地建，建造的过程本身就越生出分歧。 施工的人说不到一块儿，工程反而把人分散了。

日常里也有这样的剧情。一家公司决定上一套“统一系统”，把销售、仓库、财务全接进同一个平台，从此一个系统管所有事。上线之后，每个部门都多出一个专门和系统打交道的人，多出“这个数该按谁的口径算”的争论，多出一批只有在

这套系统里才听得懂的词。系统越想让所有人说同一种话，工地上越是长出新的方言。（这是设想，用来说明结构，不指哪一家公司。）

把这三件事对到AGI上，对得相当整齐。“通用”对应那一种统一的语言：不再有一个个分开的专门系统，而是一个东西什么都能做，这种整齐的美，和“天下人口音言语都是一样的”同源。“解决一切问题”对应塔顶通天：到了那里，没有更高处，没有更难的问题。而建造中的分歧，对应新的SDE：越是认真地使用这件工具、把它推向“通用”，使用者与它、与别的行业、与彼此之间，就越是碰撞出此前不存在的新问题、新行业、新分工。塔每高一层，工地上就多出一批新的工种、新的争论、新的语言。巴别塔不是被推倒的，是在建造中自己长出了离心力。

所以，AGI首先不是一台机器，而是理念界的一座塔。

机器属于现实界，可以被造出来、被测试、被关掉，它有具体的能力和具体的边界。“通用”“终局”“一切问题”这些词属于理念界：它们是被构想出来的形状。现实界里造出来的，永远是某一个具体的系统；“那个可以收口的智能”，是理念界里发明出来的形状，然后被人们当成现实界里的施工目标。人类在理念界里发明过许多这样的东西：绝对的公平，完美的语言，永动机。它们都有力量，因为整齐，因为美，因为让人愿意为它去建造。

这个区别可以检验。机器的能力可以验收：给它一份考卷，答对多少，有个数。塔顶没法验收。“一切问题”这张考卷，只能由已经出现的问题来出，而造出这个系统、使用这个系统，本身就在往考卷之外添题。验收单上的每一项都是昨天的问题；塔顶写的是“没有更高处”，需要一张永远写不完的单子。所以人们能做的，只是一次次宣布“这一层通了”，不能宣布“顶到了”。“塔顶通天”是一个定义，不是一个高度：你走近它，它就退后一步。

这也解释了第五层要讲的AGI感。塔越高，站在塔下的人越觉得天就在眼前，而他们看不见的，是塔身两侧正在长出去的那些新枝杈。

这里要防两个误会。第一，理念界不是意识界。　本书说的理念界，是被构想出来的形状、规则和目标所在的那一界，与现实界、自我界并列，不是“脑子里的

想法”的别称。第二，说AGI是理念界的发明，不是说它是假的或没有价值。“公平”也是理念界的发明，没有人因此说公平不重要。理念界的发明往往是最有力的工具：它给建造者一个方向，一个愿意为之付出的理由。巴别塔的人们没有白忙，他们发明了怎样烧砖，怎样用沥青黏合；人类为了“通用智能”这座塔，也确实造出了一批前所未有的工具。塔的副产品，正是人类的又一次升级。

这个想法有一位分量很重的近亲。康德把“无条件的整体”这类东西叫作理性的理念，说它们可以给探索指方向，可一旦被当成能够到手的对象，思想就会陷进幻象。把AGI看成一座塔，和康德的提醒同路：终局是一个指向，不是一个目的地。本书比他多走了一步。康德的理念停在那里不动，而“通用智能”这个理念会被人拿去施工，施工又会碰撞出新的SDE。理念自己也是理念界里的一件工具，封装着人们对“智能是什么”的昨天的理解，正在被大家使用；所以它不只是指方向，它还会生出它自己的外面。

这一节的边界有三条。第一，巴别塔是神话，不是证据。本书拿它当结构上的比方，帮助看清“一统的终局”与“建造中的分歧”的关系，真正承重的仍是前八节的结构论证。第二，这一节给“终局的通用智能”这个理念定了性，对任何具体机器的能力不下判断：一个具体系统可以强到在许多领域越过人，这与“终局是理念界的一座塔”并不矛盾。第三，把AGI说成理念界的发明，是本书的一个主张，需要读者对照后面各编的论证来检验，不该当成不证自明的定义。

这一节什么时候是错的？如果有人拿出一份对“通用”的验收标准，多年广泛使用之后，没有人提出增补新的项目，也没有出现因这套系统而起、需要新标准才说得清的争议，那么验收单是写得完的，塔顶是到得了的，我这一节就错了。（眼下能看到的一个小现象是：各种能力测验被答满之后，人们总会回头出新题。这是当前的情形，不构成证明。）

下一步的问题随之而来：既然塔是理念，那么是谁在发明它，又为了谁发明它？这个问题留到第九编第42章，那里会把“问题总是对某个主体而言的问题”明确立为一个前提。

◆ 九、回扣与小结

这一章落在第三条承重命题上：错位定理，新SDE必在封装之外，因为封装时它还不存在。本章从"问题全集是否先在"这个角度，给出了错位定理的哲学背景。

小结起来：发现学认为问题全集先在，AGI是把全集发现完；发生学认为问题随着新结构不断发生，不存在可穷尽的全集。铁路逼出了"全国几点钟"，微积分的使用逼出了"什么是极限"，两者都显示，新的问题是被工具及其使用发生出来的。所以，一个把世界想成可以穷尽的AGI，其前提本身就不成立。而终局AGI本身，首先是理念界里的一座塔，不是现实界里的一台机器。

> 带走的话
>
> 拼图的答案早印在盒盖上，带孩子却没有盒盖。问题若是发生出来的，就没有一个"全集"可以被发现完；终局AGI是理念界里一座没法验收塔顶的塔，不是现实界里的一台机器。

> **小账** 这一章立了什么？——问题随新结构而发生，不存在可穷尽的问题全集；终局AGI首先是理念界的一座塔。 错在什么情况下？——有人在新工具大规模使用前写下问题清单，事后大半命中且不靠事后补充；或"通用"的验收单用了多年，无人再增补。

下一章预告：第4章将回答一个最容易被人抓住的问题：你举了这么多历史例子，难道几个例子就能证明"永远"吗？答案是不能，所以要把"佐证"和"证明"分开。

第 4 章

本书的论证方法：历史归纳作佐证，结构论证作证明；全书的可错条件

◆ 一、那班从不晚点的地铁

有一位朋友，在城里上班，十年来每天坐同一条线路的地铁。他对我说，这十年里，这班车没有一次晚点超过两分钟。他得出结论："这条线永远准时。"

我没有反驳他。我只问了一个问题："你打算拿十年的记录，来保证明年也准时吗？"

他想了想，笑了："那倒不能保证。万一哪天信号系统出故障呢。"

这是一个再普通不过的对话，却包含着一个很古老的哲学难题：从过去的许多次，推不出永远。 十年的准点，让他对明天更有信心，但它不构成"永远准点"的证明。信心和证明，是两回事。

这个道理，用在这本书上，格外重要。因为本书的书名里有一个"永远"，而书里大部分篇幅都在讲历史：火、犁、时钟、电力、文字、微积分、镜子……如果有人说，"你列举了这么多历史案例，就想证明未来永远如此？"，这句质问是有分量的。我不能躲开它。

这一章要做两件事：第一，把本书的论证方法交代清楚——历史归纳是佐证，结构论证才是证明；第二，把全书的可错条件列出来——如果哪些事发生了，我就认为本书错了。这两件事，是一本书对读者的基本诚实。

◆ 二、归纳为什么不证"永远"

先把"归纳"这件事说透。

归纳，是从许多个别的例子，概括出一个普遍的规律。每天太阳都升起，所以明天太阳也会升起；每次工具出现都引发了新问题，所以下一次工具出现也会引发

新问题。这种推理在日常生活里无处不在，也确实很有用。没有归纳，我们连早饭吃什么都没法安排。

但归纳有一个绕不过去的弱点：它的结论总是比前提"多一步"。前提只说"到目前为止如此"，结论却说"以后也如此"。这多出来的一步，没有任何已有的例子能够担保。

十个例子、一百个例子、一万个例子，都只是把信心加厚，永远加不成必然。哲学里把这个问题叫作归纳问题，休谟在十八世纪就把它讲得很清楚。这不是说归纳没用，而是说，它的用处是提高可信度，不是提供必然性。

回到本书。如果我把全书的论证建立在"历史上每一次都这样，所以永远这样"之上，那么读者完全有理由说：好，历史上是这样，可是AI这一次不一样呢？这个质问，用归纳是回答不了的。因为"这一次不一样"，恰好是归纳无法排除的可能。

所以，本书必须有另外一种论证，不依靠"次数"，而依靠"结构"。

◆ 三、什么叫结构论证

结构论证，说的是：不去数例子，而是去看一件事的构造，指出它在构造上就必然如此。

举一个日常的类比。有人想知道，会不会有一个"最大的质数"。他的做法是：先查一亿以内的质数，没找到最大的；再查十亿以内，还是没有。这样查下去，是归纳：查得越多，越觉得"质数大概没有尽头"，可是永远不能保证下一个区间不会出现尽头。

但是，早在古希腊，欧几里得就给出了一个完全不同的论证：假设有一个最大的质数，把所有质数乘起来再加一，得到的新数，不能被任何已有的质数整除，于是它要么本身就是一个新质数，要么含有一个新质数因子。无论怎么设想"最大"，都会被这个构造推翻。

这就是结构论证的味道：它不问"迄今为止是什么样"，而问"这样的东西，在构造上能不能是那样"。一旦构造上不可能，就不是"迄今为止没有"，而是"永远不会有"。

本书的主体论证，就是这样一个结构论证。我现在把它的骨架摆出来，用最短的话。

前提一：工具是已发生的智慧的封装。　一件工具，总是把此前已经弄明白的事固定了下来。这是工具的定义，不是经验概括。

前提二：使用工具，是把它放进具体的处境里去碰撞。　工具被使用的时候，遇到的是使用者、使用场合、其他工具与其他人，这些东西与封装工具时的情形不完全相同。

前提三：这些新的处境，在封装的那一刻，尚未发生。　因为封装在先，使用在后；新的处境是使用本身产生的，它不可能先于使用而被封装进去。

结论：使用工具所碰撞出来的新结构、新问题，必然在封装之外。　这就是我在前面多次提到的"错位"：封装与新问题在时间上错位，新问题永远晚于封装。（第7章会把这几句压成一个三段论：前提一对应大前提，前提二、前提三合起来对应小前提。）

然后，终局AGI作为一个被封装的系统，它的使用就会碰撞出封装之外的新问题；要处理这批新问题，需要比它更高一级的智慧；于是它不是终局。这个推演，第二编会详细展开，第九编会收束成完整的证明。

现在请你注意这里的关键：这个论证里，没有一个地方是在数历史例子。它依靠的是"封装"和"使用"的时间次序，这是一个结构上的事实。它在今天成立，在一万年后也成立，只要"工具被封装在先、被使用在后"这件事还成立。

◆ 四、那历史例子是干什么用的

既然结构论证已经是证明了，为什么还要用整整三编去讲火、犁、文字、镜子？

因为证明和佐证，各有各的作用。

证明回答的是"必然不必然"。佐证回答的是"像不像真的"。一个结构论证，哪怕在逻辑上无懈可击，读者也会问：现实世界真的这样运转吗？它是不是只是纸面上的漂亮推演？这时候，历史案例就派上了用场：它们显示，这个结构不是我凭空编出来的，在火、在犁、在时钟、在微积分、在镜子里，我们都能看到它在运转。

打个比方：一位工程师算出了一座桥的受力，结构上是安全的，这是证明。但他还是会去看看同类的桥梁在历史上的表现，看有没有哪种情况是他算漏的。历史的桥梁不能证明他的桥是安全的，却能帮他发现自己的计算里有没有遗漏。

所以，历史案例在本书里有三个作用：

第一，增强可信度，让读者看见结构论证在真实世界里的样子。

第二，充当检验：如果某个历史案例明显违反了结构论证的预期，我就必须正视它，要么修正论证，要么解释清楚。

第三，帮助读者"看见"。很多时候，人们并不是被逻辑说服的，而是被一个看得见的画面说服的。洗衣机、导航、镜子，让那个抽象的结构变得有手感。

但是，主次不能颠倒。　　如果读者读完后，觉得"哦，因为历史上每一次都这样，所以永远这样"，那就说明我没有写好。正确的读法是：结构论证给出必然，历史案例给出手感与检验。

◆ 五、再加一道保险：谁来升级，也要说清楚

在方法论部分，还需要补上一道前面提到的"缝"：谁来升级的缝。

上面的结构论证，证明了一件事：终局AGI之外，永远还有新的问题，需要更高一级的智慧去对付。它没有证明的是：这个更高一级的智慧，必须是人。

这一点很容易被忽略，也很容易被对方抓住：你说需要更高一级的智慧，那它为什么不能是下一代AI?

我的回答是诚实的：从结构论证本身，推不出必须是人。　本书所证明的，只是"终局"不存在，也就是不存在一个升级的终点。至于升级的主体是谁，本书需要另一个前提：问题总是对某个主体而言的问题；E（意义的土壤）的源头，在主体。

这个前提的意思是：一个问题之所以是问题，是因为有一个存在着的、有处境的、有所在乎的主体，在面对它。没有这样的主体，一堆结构和差异只是物质的运动，不构成"问题"。至于AI能否成为这样的主体，那是另一个问题，本书不替它作答，只是指出：目前的论证，在这一点上依赖于这个前提。

请读者特别留意：这是一个前提，不是被证明的结论。第九编会专门讨论它，并明说它的地位。如果读者不接受这个前提，那么可以保留这样一个版本的结论：终局AGI不可能，而升级总会有人（或类似人的主体）来做。这个较弱的结论，仍然是本书的核心，我也愿意单独为它负责。

◆ 六、第三道缝：象棋这个反例

还有一道缝，我在第2章稍微触及过，这里要正式处理：象棋。

1997年，计算机深蓝战胜了当时的国际象棋世界冠军卡斯帕罗夫，此后，人类顶尖棋手在与机器的正式较量中，再没有占过明显上风。〔待核：此后人机对弈的具体情况以最新资料为准，此处为概括〕这是一个确凿的例子：在这个领域，AI永久地越过了人类。

这个例子，对本书的论证，似乎是一个反例：你说没有终局，可象棋在这里有一个终局：AI赢了，而且将一直赢下去。

我的处理是：承认它，并用它划出边界。

象棋是一个封闭领域：棋盘固定，规则固定，胜负的判据固定。在封闭领域，"封装"和"使用"之间不会碰撞出新的问题，因为这个领域的全部问题，早在规则被确定的那一刻就已经设定好了。对于这样的领域，永久越过是可能的。

可是，"通用"这个词，指的恰恰是开放领域：规则不固定、处境不固定、问题会随着使用而生长。在开放领域里，象棋式的"永久越过"不成立。

所以，象棋不是本书的敌人，而是本书的边界石。它告诉我们：本书的主张，不是"AI永远不能在任何事上超过人"，那是错的；本书的主张是"AI不能在开放领域里永久地、一劳永逸地超过人"。这两个主张之间的距离，就是象棋的位置。第八编会专门讨论。

◆ 七、全书的可错条件

现在是这一章最重要的部分。一个论证，只有当它说得出"在什么情况下我错了"，才配被认真对待。我把全书的可错条件列在下面。

第一类：可以被观察到的预言。　这三条是我对未来做出的判断，如果它们落空，本书的核心主张就受到严重损伤。

预言一：　任何开放领域，在AI能力出现大幅跃升之后的数年之内，会出现由这次跃升本身引发的新职业，或新的问题类别。

这里的"开放领域"，指规则和处境都不固定、问题会随使用而生长的领域，比如写作、教育、医疗、法律、城市管理，不包括象棋这类封闭领域。"新职业或新问题类别"，可以这样判定：出现了此前没有名称、现在有了专门岗位或专门学科的活动；或者出现了此前不被当作问题、现在被广泛讨论和投入资源的议题。如果在某个开放领域里，AI能力大幅跃升之后相当长的时间，没有出现任何这样的新东西，只有旧问题被消除，那么这条预言落空。

预言二：社会上的AGI感，在扩大期最强；进入内卷期之后，会明显下降。

"扩大期""内卷期"是工具让渡律里的两个阶段，第二编第九章会讲清楚。这条预言说的是一个可观察的节奏：新工具出现的初期，人们最容易觉得"什么都能做"；等到大家都在用、竞争加剧、红利递减的时候，这种感觉会下降，人们开始谈论新的难处。如果AGI感在内卷期依然持续高涨，没有任何回落，这条预言落空。

预言三： 无论哪一次宣称"AGI已经实现"，此后被公认的前沿基础问题清单，只会变长，不会变短。

这是第九编的预告。我的预期是：每一次宣称实现，都会随之带出新的、此前不被当作基础的问题。如果某一次宣称之后，基础问题的清单真的缩短了，也就是说，学界公认的前沿难题在这一次之后一件件被解决、没有新的难题补上，那么这条预言落空。

第二类：结构论证本身的可错点。 这几条，是我认为论证最薄弱、最该被检验的地方。

可错点一：前提二是否普遍成立？ 结构论证依赖"使用必然碰撞出封装之外的新结构"。如果有人能构造出一个开放领域里的工具，它被大量使用之后，不产生任何封装之外的新问题，那么这条前提就被推翻，本书的结构论证也就被击穿。

可错点二：封装与使用之间的错位，能否被"预先模拟"消除？ 如果有人证明，存在一种方法，能把一件工具在所有可能的使用场景下的新问题，在封装之前就完整地推演出来，而且这个推演本身不会引入新的使用场景，那么错位定理就不成立。我认为这不可能，因为推演的行为本身就是一次新的使用，但这一条我愿意写出来，让读者检验。

可错点三：主体前提。 如果有人令人信服地证明，在没有任何主体的情况下，"问题"仍然可以作为问题而存在，也就是E的源头不必在主体，那么我关于"问题对谁而言"的前提就要放弃，本书中依赖它的那一部分（主要在第九编）要重写。

可错点四：开放与封闭的划分。 我的论证依赖"通用即开放"。如果有人表明，现实中的某个"通用"任务集合，实际上是一个可以完全形式化的封闭系统，那么象棋式的永久越过就可能推广到它身上。

第三类：什么不算反驳。 为了避免将来无谓的争吵，我也事先写出几类不能算反驳本书的事情。

某个AI在某天做到了某件此前认为它做不到的事，这不是反驳。本书并不主张AI有一个固定的能力上限。

在封闭领域里AI永久越过了人，这不是反驳。这是象棋的情形，本书已经承认。

某个人宣称"AGI已经实现"，这本身不是反驳，也不是证实。要看他宣称之后发生了什么：基础问题清单是变长了还是变短了。

◆ 八、这个方法的边界在哪里

最后，检查一下这一章自己的方法。

第一，结构论证的力量，取决于它的前提。上面的三个前提，如果任何一个被推翻，论证就会受损。这是结构论证的代价：它比归纳更强，但也更脆；归纳是慢慢变薄，结构论证是一断就断。我选择它，是因为只有它才有资格谈"永远"。

第二，我把预言写成可观察的形式，但"数年""大幅跃升""新问题类别"这些词仍然有弹性。我尽量给出判定办法，但在边缘情形下，仍然会有争论。为了不让“数年”无限拉长，我给预言一划一个窗口：从那次跃升算起，三到八年；窗口过了还没有出现，就按落空算。其余的弹性，是预言类命题的天然困难，我不假装它不存在。

第三，可错条件的列出，并不保证本书是对的；只保证它是可检验的。

第四，“把可错条件写出来”不是我的发明，它来自波普尔在1934年《研究的逻辑》里立下的可证伪标准。我只是把它用到了一个哲学味很重的命题上，所以判定办法不得不写得啰嗦一些。

◆ 九、第一编小结，以及接下来的路

这一章，也是第一编的收束。

第1章把AGI拆成三种，指出本书击中的是终局AGI。第2章翻开库存图景的四张底牌。第3章翻开第一张，区分发现学与发生学，并说终局AGI首先是理念界的

一座塔，不是一台机器。第4章交代方法：结构论证作证明，历史归纳作佐证；谁来升级的缝、象棋的缝都已处理，全书的可错条件已列出。

靶子立好了，方法也交代了。接下来，我们进入第二编，把那个结构论证的骨架，一根一根搭起来。我们从最日常的东西开始：为什么说一台洗衣机，是被装进盒子里的智慧？

> 带走的话
>
> 十年准点，保证不了明年；十个例子，也保证不了"永远"。本书拿结构当证明，拿历史当佐证，还把自己可能错在哪里，提前写在了纸上。

> **小账** 这一章立了什么？——结构论证（封装在先、使用在后）作证明，历史归纳作佐证；补上谁来升级、象棋两道缝，列出可错条件。 错在什么情况下？——三条预言落空（窗口为跃升后三到八年），或使用必生新结构、错位不可预先模拟、主体前提、开放封闭划分，任一被推翻。

下一编预告：第二编"原理：工具发生律"。第5章，工具是智慧的封装。

第 二 编

原理：工具发生律

编 序

前一编把靶子立好了：终局的AGI，是一个"装满了全部智慧、从此不必再升级"的东西。这一编要做的事，是从最日常的地方出发，一步一步搭起一条结构论证，说明这样的东西为什么立不住。论证只有四个环节：工具是封装，使用是碰撞，新的东西必在封装之外，封装之外的东西逼出更高一级的智慧。四个环节环环相扣，读完这一编，全书的骨架就都在手里了。

第二编一共五章。第5章立起点：工具是已发生智慧的封装，用洗衣机和账本来讲。第6章讲使用：工具一进入真实的生活，就和一个它从没见过的世界碰撞。第7章把前两章压成一个三段论，证明新东西必在封装之外，并说清它在哪里失效。第8章看二阶碰撞：补丁补不平的时候，需要一种更高一级的智慧，并明说"问题总是对某个主体而言"这个前提。第9章把这一切放到产业层面，用一碗面条的账，走出工具让渡律的四个阶段。

章	一句话	对应的承重命题
第5章	工具封装的是已经发生的智慧	第一条：工具是已发生智慧的封装
第6章	一用就撞，撞出说明书里没有的东西	第二条：使用即碰撞，碰撞必生新SDE
第7章	新东西必在封装之外	第三条：错位定理
第8章	补丁补不平，就逼出更高一级	升级律，并明说主体前提

章	一句话	对应的承重命题
第9章	旧能力让渡，新位置被占住	第五条：被封装能力的退场是让渡，不是退化

本编是第一层论证"没有终局智能"的主场，也是全书的证明所在。这一编的例子以日常为主：洗衣机、导航、做菜、一碗面；历史案例只作佐证。第7章会说得很直白：这些故事可以全部抹掉，三段论依然站着，故事只是让它好懂。

第 5 章

工具是智慧的封装

◆ 一、一台洗衣机里藏着什么

先从一台洗衣机说起。

你按下"标准洗"，水进来，转筒转起来，过一会儿水排掉，再进水，漂洗，甩干，嘀嘀嘀响三声。整个过程你什么也不用想。可是这三声"嘀"背后，压着很长很长的一段历史。

人类洗衣服，最早大概就是在河边石头上拍打。后来有了皂，知道了油污遇到碱会变得好洗。后来有了搓衣板，知道了摩擦能带走污渍，却不伤衣服。再后来，人们慢慢摸出来：白衣服和深色衣服要分开洗，毛料不能用热水，丝绸不能拧，水温高一点污渍好洗但是会缩水，泡一泡比使劲搓更省力。这些东西，没有一样是一个人坐下来一次想出来的。它们是一代一代的人，在一件一件被洗坏的衣服上，试出来的。

洗衣机做的事情，就是把这些试出来的东西，固定在一个会转的桶和一块电路板里。水温设几度，转几分钟，正转几下反转几下，什么时候停，什么时候排水。每一个数字背后，都是一个当年的"原来如此"。

所以我们可以这样说：洗衣机不是一堆铁和塑料，它是一堆已经发生过的智慧，被装进了铁和塑料里。

这就是本章要立的第一句话，也是整个论证的起点：工具是已发生智慧的封装。

◆ 二、先把几个词说明白

这本书里会反复用到一个三字框架，叫SDE。在正式进入论证之前，用大白话交代一下，免得后面卡住。

S是结构，意思是"显露出来、稳定下来的样子"。一台洗衣机的程序，一本账簿的格式，一条交通规则，都是S。它们是看得见、摸得着、可以复制的东西。

D是差异，意思是"发生的过程"。一个人第一次发现毛衣用热水会缩，那一刻就是D：从"以为没事"走到"原来有事"，中间有一道差，有一次跨过去的动作。S是D留下的脚印，D是走路本身。

E是纠缠，意思是"土壤、场合、意义所在的那片天地"。同样一件毛衣，在寒冷的山区和在海边的小城，缩了不缩、要不要紧，意义完全不同。E就是这一团相互牵扯的背景：人、物、习惯、需要、环境，全搅在一起。

好，有了这三个字，就可以把上面那句话说得更精确一点：工具是把一些已经走完的D，沉淀成稳定的S，再打包带走。 被打包的，是智慧在某个时刻、某个场合、某群人身上的发生结果。

◆ 三、封装有四个特点

"封装"这个词是从计算机里借来的，但意思很朴素：把里面的东西包起来，只留外面一个简单的把手。工具的封装有四个特点，后面的论证会一个个用到，这里先摆出来。

第一，固化。 原本灵活的、因人而异的做法，被固定成一套。老师傅洗衣服，今天水凉一点就多泡一会儿，明天衣服脏一点就多搓两下，他心里有一本活的账。洗衣机把这本活账固定成十几个程序。固定意味着稳定，也意味着不会再自己变。

第二，可转让。 这是封装最大的好处。你不需要懂洗衣的全部学问，也能洗得不错。一个十岁的孩子按下按钮，他拿到的是几代人的经验。智慧因此从一个人的头脑里，搬到了千千万万人的手里。文明能够累积，靠的就是这个。

第三，隐去过程。 用的人看不见里面发生过什么。你不知道为什么棉质衣物的默认程序是四十分钟，不知道有人曾经为此毁掉过多少件衣服。过程被隐去，结

果留下来。这既是便利，也埋下了后面要讲的伏笔：当里面的东西不够用了，使用者往往不知道该从哪里改起。

怀特海在1911年说过一句同路的话：文明的进步，靠的是增加那些不必再去想就能做的操作。〔待核：出处为其《数学导论》，措辞以原书为准〕本书和他相同：封装让人不必再想。不同在于，本书盯住的是，不必再想的那一层，正是新处境最先撞上的地方。

第四，停在某个时刻。　　这一点最重要，也最容易被忽略。封装发生的那一天，智慧就被"拍了一张照片"。洗衣机出厂时装进去的，是出厂那一天为止人们知道的洗衣方法。它不会自己长出新的方法。十年以后，衣服的纤维变了，人们对洗涤剂的看法变了，对水的看法变了，那台十年前的机器还是照旧转它的四十分钟。

这四个特点，合起来说的是同一件事：工具封装的，是昨天的智慧。

◆ 四、不只是机器：账本也是一种封装

也许有人觉得，洗衣机太"硬"了，是机器才有的事。那么换一个不那么硬的例子：记账。

在意大利的商人们开始做大买卖的时候，一笔钱怎么记，是各人有各人的办法。有人记流水，有人记在小本子上，有人干脆靠脑子。买卖做得小，没有问题。买卖一大，合伙的人多了，钱从这里转到那里再转回来，就必然有人想知道：到底赚了没有，钱到哪里去了。

后来，一种"每一笔钱都同时记两处，一处记来源，一处记去向"的办法，在商人们中间流行开了。这就是复式记账。1494年，一位名叫帕乔利的数学家把它写进了一本书里，这本书后来流传很广。要说清楚：他并不是复式记账的发明者，他整理和传播的，是威尼斯商人已经在用的方法。

把这件事放在本章的框架里看，意思就很清楚了。复式记账是一个封装：许多商人的许多个"原来如此"，被压成了一套格式，谁学会了谁就能用。一个刚入行的

小伙计，不需要重新发明如何防止做假账，他把借和贷填在两边，账自己会告诉他对不对。账簿，是用纸和墨水做成的洗衣机。

热力学也有类似的故事。蒸汽机先在矿井和工厂里转了很多年，工匠们靠经验把它越改越好。1824年，卡诺写下了关于热机效率的思考，后来人们回头看，说那是热力学的开端。这件事的次序值得留意：先有机器里封装的经验，后有把经验变成理论的那一步。而理论本身，一旦写成公式，又成了新的封装：后来的工程师不必再重走一遍卡诺的思路，只要把数字代进去。

这里已经能看出一个很有意思的现象：封装是一层一层摞起来的。　经验封进机器，机器的行为又被总结成理论，理论又被封进软件，软件又被封进别的工具。第一层是智慧，第二层是封装了智慧的东西，第三层是封装了这个封装的东西。今天我们手边的任何一件工具，都是这样一摞又一摞的结果。

经济学家布赖恩·阿瑟在《技术的本质》（2009年）里，把技术看成从已有技术里组合出来、层层叠起的东西，这与上面说的摞法相合。本书和他相同在“技术生于技术”；不同在于，本书盯的不是技术怎样组合出新技术，而是每一层封装之外，使用怎样逼出封装里没有的东西。

◆ 五、为什么一定是"昨天"

读到这里，一个自然的疑问是：为什么说封装的一定是昨天的智慧？工具设计者难道不能把"明天会发生的事"也想到、也装进去吗？

这个疑问值得认真对待，因为它正好是后面几章要回答的。先给一个简单的回答，细的留给第7章。

道理是这样的：封装这个动作，本身就要求被封进去的东西"已经有了"。你不能把一件还没发生的事固定下来，就像你不能把明天的天气拍成今天的照片。设计洗衣机的工程师，可以非常聪明，把他能想到的各种衣服、各种水质、各种用户习惯都考虑进去，可是他考虑的，仍然是他所知道的、已经发生了的那些情形。明天

才会出现的新布料、新的生活方式、新的人对"干净"的新要求，不在他的清单上，因为它们还没有发生。

所以"昨天"在这里不是一个时间点，而是一条结构上的界线：封装的那一刻，把已发生的东西和未发生的东西，分在了两边。 工具只能站在界线的这一边。

◆ 六、回扣：这是第一命题

把以上几节收一收，本书十条承重命题里的第一条，现在可以完整地立起来了：

> **工具是已发生智慧的封装。**

它有三层意思。其一，工具里有智慧，不是死铁。其二，这些智慧是被"已经走完的发生"沉淀下来的，是S，不是D。其三，因为是沉淀，所以它停在某个时刻，带着那个时刻的E。

如果读者此刻心里想的是"那又怎样？工具装着昨天的智慧，也很有用啊"，这个想法完全对。工具当然有用，非常有用，文明靠它们才站得起来。本书从头到尾，没有一个字是在贬低工具。要说的只是：它们有用，而且它们是封装，这两件事同时为真。

◆ 七、这个说法的边界在哪里

先交代这一章的论证走到哪里为止。

第一，"封装智慧"是一个比喻，比喻有它好用的范围。有些工具的诞生，更像是偶然碰上的：一块石头碰巧是锋利的，一种霉菌碰巧杀了细菌。这些工具里封装的是什么智慧？严格地说，最初是"一次被注意到的发现"。我们仍然可以说，发现的那一刻是一次发生，之后的反复使用是把它固定下来，所以框架依然成立。但是读者应当知道，这个框架在这类情形里是被"拉过去"用的，不是天然合身。

第二，被封装的不一定是正确的智慧。历史上有大量工具封装的是后来被证明有偏差的做法。这并不破坏本章的论点，反而加强它：封装的就是封装当时的认识，对也好，偏也好，一起带走。

第三，也是最需要说清楚的：本章只立了"工具是封装"，并没有证明"封装之外一定有新东西"。　　那是第6章和第7章的事。如果有读者觉得到这里论证还没有力量，那是对的，它现在还只是一个定义，一个起点。

第四，这一章的判断什么时候是错的？如果找到一件被广泛使用的工具，几十年里人们对它的所有新用法，在设计之初的文件里都已经写过，那么“工具只封装昨天”就被推翻了。这样的例子可以去找；我没有找到。

◆ 八、小结

洗衣机是几代人试出来的洗衣智慧，被固定在一个会转的桶里；复式账是许多商人摸出来的防错办法，被固定在两栏表格里。固化、可转让、隐去过程、停在某个时刻，这四个特点决定了工具的长处和短处：它能把昨天的智慧送到千万人手里，却长不出明天的智慧。

可是工具造出来，是要被人用的。一旦被用，会发生什么？

> 带走的话
>
> 按下洗衣机的按钮，你拿到的是几代人试出来的"昨天"。它好用，所以更要记得：它停在了出厂那一天。

小账　这一章立了什么？——命题一：工具是已发生智慧的封装，有固化、可转让、隐去过程、停在某个时刻四个特点。　错在什么情况下？——找到一件广泛使用的工具，几十年里的所有新用法，在设计之初的文件里都已写过。

承接：下一章看"使用"。工具一进入真实的生活，就要和一个它从没见过的世界相遇。

第 6 章

使用即碰撞

◆ 一、洗衣机省下的时间去哪儿了

洗衣机进入家庭的时候，几乎所有人都觉得，家里的女人这下可以松口气了。再不用在水池边弯一下午腰，再不用拧到手酸，再不用晾得满院子都是。省下的时间，大家想，总该有别的用处。

可是研究家务史的美国学者鲁思·考恩，把这件事掰开来看，得出了一个让很多人意外的结论：二十世纪里，家用技术越来越多，家庭主妇花在家务上的时间，却并没有明显减少。她1983年的书《母亲的活更多了》，讲的就是这个反差。具体的时间数字各项调查口径不一，本书不引用，只取她的核心判断：技术并没有把家务时间省下来。

为什么？她和后来的研究者给出的解释，都指向同一类原因，只是各有侧重。

第一，衣服洗得更勤了。洗一次不再是一件要折腾半天的大事，于是"穿一天就洗"成了常态，衣橱里的衣服也越来越多。以前一件衬衫穿一个星期，现在一天一换。

第二，"干净"的标准升高了。有了洗衣机，白衬衫的领口就不能再有一点黄；孩子沾了泥的校服，第二天必须是平整洁净的。标准是社会慢慢抬起来的，没有人宣布过，但每个人都感觉到了。

第三，原来帮忙的人退场了。在洗衣机之前，洗衣常常是几个人的事，或者交给专门的洗衣工、洗衣店。机器一来，这一部分外包的劳动没有了，全回到了家里，回到了那位主妇的身上。〔待核：这一点在考恩及相关研究中的具体论述〕

你看，洗衣机本身没有任何问题，它的四十分钟程序运转得非常准。可是它一进入真实的家，就和一堆它从来没见过的东西撞在了一起：人的期待、邻居的眼

光、家里的分工、一年四季的衣橱。撞出来的结果，是一个工具出厂时没有写进去的新局面。

这就是本章要讲的事：使用即碰撞。

◆ 二、碰撞是什么意思

"碰撞"这个词，在本书里有一个具体含义，不是随便打的比方。

上一章说，工具是封装好的S，是稳定下来的结构。而使用的地方，是一个活的E：具体的人，具体的屋子，具体的需要和习惯，还有一直在变的周围环境。使用，就是把一个已经定了型的S，放进一个正在流动的E里。

这两样东西的"年龄"不同。S是昨天定下的型，E是今天正在发生的现场。两者相遇，一定会有对不上的地方。那些对不上的地方，不会消失，它们会变成一个又一个小的D：一个差异，一个别扭，一个"咦，怎么会这样"。人们为了处理这些别扭，会想办法、会调整、会发明新的做法。这个过程里长出来的东西，就是新的S、新的D、新的E，合起来叫新的SDE。

把这个过程写成一句话：使用把封装好的结构，放进活的土壤，对不上的地方就是新发生的起点。

注意，碰撞这个词里没有褒贬，不等于冲突，也不等于坏事：撞出来的可能是好东西，也可能是麻烦，多数时候两样都有。它只是说，只要用，就一定会发生封装里没有的事。

◆ 三、另一个例子：导航与那条小巷

洗衣机的例子，是慢慢发生的，要几十年才看得清。再举一个快的，你大概自己亲眼见过。

你开车去一个不熟悉的地方，打开导航。导航算了一下，说：这条路堵，我带你绕一下，走旁边那条小路。你照做了，果然省了十分钟。你很高兴。

但是你可能没有想到，同一时间，附近几百辆车的导航，都在对它们的主人说同一句话。于是那条平时只有零星几辆车经过的小巷，忽然涌进了几百辆车，开始堵，而且堵得比大路还严重。小巷两边的居民从此日夜听着喇叭声，原本安静的街区成了"导航热点"。〔待核：此类现象在多地媒体有报道，具体案例与数据未核〕

这里面发生了什么？

导航里封装的，是一个很漂亮的智慧：给定路况，求最快的路线。这个智慧在"只有你一个人用它"的世界里，是对的。它被封装的时候，那个世界就是它的E。可是等到几百万人同时使用它，E变了：现在路况不再是外面给定的，路况本身就是被这几百万个导航共同造出来的。封装里没有这一条，因为封装的时候，用的人还没有那么多。

导航算的是"最快的路"，结果制造出"所有人都去走同一条路"。而这个结果，反过来推翻了它算出的"最快"。这就是一次干净利落的碰撞：工具的S和使用者聚集起来的E撞到一起，新的D出来了，没有人事先写过它。

更有意思的是后面。小巷居民会投诉，城市管理者会在小巷口设限行、加路障、改单行。导航公司会开始把"这条小巷太挤"这样的信息也算进去。司机们会慢慢学会，高峰时不要太相信导航的"神仙小路"。你看，一个新的局面，引出了新的规矩、新的做法、新的信息。这些都是此前不存在的。

◆ 四、铁路与时间：一个更大的碰撞

第3章讲过铁路与“几点钟”，这里只取其中和碰撞有关的一层。铁路封装的，是“按时刻表运送大量人和货”的智慧。这套智慧一旦放进一个各地各用各的本地时间的世界，就撞出了1883年北美的标准时区：时间这个原来由太阳决定的东西，变成了全社会共用的约定。设计铁路的人，没有一个把“重新定义几点钟”写进过图纸。那是使用撞出来的。

◆ 五、碰撞为什么"必然"发生

读到这里，读者也许会想：这些都是好故事，可它们是不是挑出来的？有没有哪种工具，用了以后什么新东西也没发生？

先看碰撞为什么不能不发生，这是一个结构上的问题。

使用要发生，就要有一个使用者，和一个使用场合。这个使用者是一个具体的人（或一群人），带着具体的需要，处在具体的环境里。而封装的时候，设计者面对的是"一类人""一类场合"。"一类"和"这一个"之间，永远有差。　　这个差可以很小，小到看不见，但它不会是零，因为"这一个"总是带着此时此地的独一无二的东西：这一刻的天气，这个人今天的心情，这间屋子的格局。

差一旦存在，使用者要把工具用起来，就必须多做一点工具没替他做的事：调整、将就、变通、绕开。这"多做的一点"，就是一次小小的发生。绝大多数这样的小发生，会随着使用结束而消散，没有后果。但是当它们积累起来，被很多人重复，被写进新的习惯、新的规则、新的产品里，就成了看得见的新SDE。

所以我们不是说"每一次使用都会改变世界"，而是说：每一次使用都有碰撞，碰撞里都有发生的成分，只是大小不同，多数很小，少数很大。　洗衣机、导航、铁路，是碰撞大到连历史学家都看得见的案例。

◆ 六、回扣第二命题

于是本书的第二条承重命题可以立起来了：

> **使用即碰撞，碰撞必生新SDE。**

这里的"必"字，要说清楚它的分量。它说的是：只要工具被真实地使用，在使用的那个E里，就一定会有封装没有预先定好的东西发生。至于这个东西大还是小、好还是坏、是否被人注意到，这一章不保证，也保证不了。

◆ 七、碰撞说得过头的地方

第一，碰撞不等于进步。　　导航的例子里，碰撞的第一个结果是拥堵，是坏事。洗衣机的例子里，碰撞的结果之一是家务没有减少，很多女性的日常负担并没有因此变轻。本书要证明的是"新的东西必然发生"，不是"新的东西必然是好的"。后者是另一个问题，第六编会专门谈：工具能提升智慧，不等于没有代价。

第二，碰撞的大小，取决于E的差距。 如果一个工具用在一个和它的出生地非常相似的环境里，碰撞就小；如果被带到一个很不一样的环境，碰撞就大。一把菜刀在厨房里用，几乎不撞出什么新东西。一把菜刀被带到一个没有刀的地方，可能就是一场革命。

第三，也是本章必须坦率承认的：到目前为止，"必生新SDE"这个说法，一半靠例子，一半靠"一类与这一个之间有差"这个结构理由。 例子不能证明"必然"，结构理由才有可能。至于这个结构理由能不能撑起一个更强的结论，要等到下一章，用一个三步的推理来检验它。

第四，这一章的观察在邻近领域有近亲。社会学家默顿在1936年讲过“有目的的行动的意外后果”；经济学家杰文斯在1865年讲过一个煤的故事：蒸汽机更省煤了，全国用煤反而更多，后人叫它杰文斯悖论。洗衣机越省力、洗得越勤，是同一个形状。本书和它们相同：好工具的结果，常常在设计者的意图之外。不同在于，本书不满足于说“意外”，而要追问这个意外为什么在结构上必在封装之外，那是下一章的事。

第五，什么情况下“必生新SDE”是错的？找一件被几百万人用了十年以上的工具，查它的使用史。如果找不到任何一个说明书里没有、后来被写进习惯、规则或产品的新做法，这一命题就被推翻了。

◆ 八、小结

洗衣机省下的时间，被更高的标准和更勤的更换吃掉了；导航省下的十分钟，被几百辆车的涌入吃掉了；铁路送货的效率，撞出了一套统一的时间。共同点是：工具是昨天定型的结构，使用的地方是今天流动的土壤，两者相遇，对不上的地方便长出新的东西，而这些新东西不在说明书里。

可是"不在说明书里"，还只是一个观察。有没有办法，把它变成一个站得住的结论：不只是"碰巧不在"，而是"必定不在"？

> 带走的话
>
> 工具出厂时对着的是"一类人"，用的却是"这一个人"，中间那点对不上，就是新东西发生的地方。洗衣机省下的时间，就是这样不见的。

> **小账** 这一章立了什么？——命题二：使用即碰撞，碰撞必生新SDE；碰撞不等于进步，大小取决于土壤的差距。 错在什么情况下？——找一件被几百万人用了十年以上的工具，查不到任何说明书里没有、后来写进习惯、规则或产品的新做法。

承接：下一章用三句话搭一个三段论，把"新的东西一定在封装之外"证明出来，并说清它在哪里失效。

第 7 章

错位定理

◆ 一、一个看起来可以补上的漏洞

假如你是导航软件的工程师。你听说了小巷被堵死的事，很不高兴，决心把它修好。

办法并不难想。你让软件多做一步：不要只看当前的路况，还要估计"如果我把这条路推荐给所有此刻在附近的人，会发生什么"。你把几百万次这样的模拟装进去。新的版本上线，它不再把所有人导进同一条小巷，而是分流：一部分走这条，一部分走那条。

问题解决了。看起来，只要工程师足够聪明，所有碰撞出来的新东西，最终都能被事先想到，被封装进下一个版本。如果这样，那么"新的东西在封装之外"就只是暂时的，只要不断打补丁，补丁总有一天会补到头。这样一来，一个"什么都封装了"的终极工具，岂不就是可能的?

这正是终局AGI想象里最顺口的一条路：只要模型够大、数据够多，就能把碰撞也预测进去，把一切都封装进去。

本章要做的，就是回答这条路为什么走不通。方法是把前面两章的内容，整理成一个清清楚楚的三段论。

◆ 二、三段论

先把推理摆出来，再一句一句解释。

> 大前提： 封装只能封装封装之时已经发生的东西。
>
> 小前提： 使用所碰撞出的新SDE，是在使用之中才发生的，封装之时它还不存在。
>
> 结论： 因此，新SDE必在封装之外。

这就是本书所说的错位定理。"错位"二字的意思，是封装和新SDE在时间上差着一个位置：封装在先，新SDE在后；封装能装的在前一个位置，新SDE要出现在后一个位置。前后对不上，永远对不上。

下面一句一句看。

大前提，是第5章已经立住的内容。封装是把已经走完的发生固定下来。你可以固定一次已经发生的洗衣实验，不能固定一次尚未发生的。这是封装这个词本身的含义。

小前提，是第6章立的。注意它的措辞：新SDE是"在使用之中发生"的。为什么不能提前发生？因为它是工具的S和使用者的E碰撞出来的，而使用者的E，包括他此刻的需要、环境和与别人的关系，是在使用的当下才成形的。在工具造好的那一天，这个E还没有来到这个世界。

结论，不是"新的东西大概会在外面"，而是"必在外面"。区别在这里：这是一个由前提推出的结论，不是一个经验的估计。只要两个前提成立，结论就一定成立，不管工具有多大、多聪明。

◆ 三、回到那个补丁

现在回头看那位工程师的补丁。

他的新版本，把"所有人同时被导到小巷"的情形预测进去了。这确实是一次进步：他封装了一个此前不在封装里的东西。可是请注意，这个新东西是什么时候被

他封装的？是在小巷被堵死之后。他封装的，是已经发生的那一次碰撞的结果。这仍然是"昨天的智慧"。

新版本上线以后，会发生什么？

司机们发现导航现在会分流。有人发现，分流的规则大概是"高峰时间推荐远一点的路"，于是故意在导航推荐前先自己看一眼路况，或者干脆提前十分钟出门避开。有人发现导航分流之后，原来很空的第三条路变得忙起来。商家发现，导航不再把人导进某条街，于是那条街的生意变差了，商家会向平台抱怨，或者想办法买推荐位。管理者发现，原本想靠限行解决的问题，因为导航分流变了形。

这一堆都是新的。它们在新版本封装的那一天，同样不存在。

所以补丁没有补到头，它只是把旧的错位推后了一步，在新位置上，生出一个新的错位。　这个过程可以无限重复：每一次封装都是对上一次碰撞的回答，每一次回答都会引起一次新的碰撞。追补丁的人，永远在追一个比他快半步的东西，就像追自己的影子。

有人会说：那就把使用者的反应也都模拟进去。这个反驳要认真回应，下面单独谈。

◆ 四、为什么"模拟使用者"也不行

设想那个工程师做了一个极其精细的模拟：每一个司机的性格、习惯、对导航的信任程度都有一个数字模型，几百万个模型一起跑，预测出未来的交通。

这个办法的困难，不在计算量，在于一个更根本的地方：被模拟的人，会因为这个模拟的存在而变成另一个人。

一旦有人看到了"导航会这样分流"这件事，他的行为就会改变；而他的行为改变了，模拟里的他就不再是他了。模拟要把这一层也算进去，就得模拟"知道了这个模拟的人"，可是知道了这个新模拟的人，又变成了更新的人……这是一个自己追自己的圈。圈的每一环，都是一次新的发生。

更日常的说法是：你没法在镜子里提前看到自己照镜子之后的样子，因为你看了镜子，神态就变了。模拟和被模拟者之间，隔着"被看见"这件事。这个自指的圈，是错位定理最深的一个根：新SDE之所以必在封装之外，不仅因为时间上的先后，还因为封装本身，一旦进入使用，就成了被封装者的环境的一部分。

这个圈在邻近学科里有名字。经济学里有卢卡斯批判（1976年）：依据过去的数据预测政策效果，人们一旦知道了政策，行为就变，旧数据管不住新行为。波普尔也早说过，预测知识的增长，本身就是知识增长的一部分。本书和它们相同：被预测的人会因预测而变。不同在于，本书把这个“变”提到了封装的层面：不只是政策，任何一件工具，一进入使用就成了使用者环境的一部分。

◆ 五、把三段论放进更大的历史里

错位定理说的是一个结构，它在历史里有大量的证据。这些证据不是证明，是佐证，本书第4章已经交代过主次。

第一个证据，是索洛悖论。1987年，经济学家索洛说过一句很有名的话，大意是：计算机的时代到处都看得见，唯独在生产率的统计里看不见。为什么？一个常被引用的解释来自经济史学者保罗·戴维，他在1990年发表的论文里研究过电动机：电动机在十九世纪末就有了，可是工厂的生产率大幅提升，却等到几十年之后。原因是，早期的工厂只是把蒸汽机换成电动机，其他一切照旧；直到人们重新设计了厂房、流水线和工人的分工，电的好处才真正释放出来。

这个故事里的错位非常清楚：电动机被封装出来的那一天，"围绕电动机重新布置的工厂"还不存在，没有人能提前把它装进电动机里。它是在使用中，一点一点碰撞出来的。

第二个证据，是1968年的"软件危机"。这个词是在当年北约举办的一次软件工程会议上被叫开的，指的是软件越做越大，却越来越难按时、按预算、保质地做出来。在此之前，人们相信有了大型计算机和高级语言，写软件会像造房子一样越来越有章法。结果却是，每一次工具的提升，都带来了新规模的软件，新规模带来了

新的混乱。工具封装了"怎么写代码"的智慧，使用却撞出了"怎么管理几千万行代码"这类封装里没有的新问题。

这两个例子，都不能"证明"错位定理，它们只是说明：错位定理在历史里不是纸上谈兵，是你用放大镜在任何一个时期的任何一个工具身上，都能找到的纹路。

◆ 六、这个定理的边界

错位定理在哪里成立，在哪里不成立，没有证明什么，要说清楚。有六条。

第一，它依赖一个前提：发生不是预先存在的。

大前提和小前提里，都暗含着"新SDE是真正新的，不是早就藏在某个地方等着被取出来"。如果有人坚持一种彻底的决定论，认为宇宙在最初那一刻就把未来的一切都定好了，只是我们还没算出来，那么"新"只是看起来新，而一个足够强的计算者原则上可以提前全部算出来。这时，三段论的小前提就不成立了。

本书不假装这个前提无需辩护。它是一个哲学上的立场，也就是本书说的"发生学"立场；与之对立的"库存"立场，在第二章里有过讨论，后面第七编还会专门谈。读者可以不同意这个前提，但要看清楚：不同意的地方，在前提，不在推理。推理本身是严格的：只要承认发生是真发生，错位就跟着来。

第二，它在封闭领域里"错位"仍在，却不要紧。

象棋的规则，不会因为有人下棋而改变，棋盘的大小、棋子的走法，在使用中一概不变。在这样的封闭领域里，使用并不碰撞出改变问题本身的东西，错位虽然在形式上存在，却无关痛痒。所以，AI在象棋上的永久胜出，不违反错位定理，也不被它排斥。第八编会正面讨论这个反例。这里只需记住：错位定理主要在"使用会改变问题本身"的开放领域里起作用。

第三，它只证明"存在"，不证明"多大"。

结论说的是：新SDE必在封装之外。它没有说新SDE一定很重要，一定很大，一定值得我们操心。一个工具被使用以后，可能真的只是产生了一些微不足道的小

变化。定理保证的是"有"，不保证"大"。大不大，要看具体的工具、具体的E，这是靠历史和经验来回答的，也是本书把大量篇幅放在工具史上的原因。

第四，它没有证明"谁"来处理这个新SDE。

这一条与"谁来升级缝"直接相关，也最容易被误读。错位定理只说：封装之外必有新东西。至于这个新东西由谁来发现、由谁来回应，是人，是另一个机器，还是人和机器一起，这一章没有说，也说不出来。后面第8章会推进一步，但那一步需要一个明说的前提，到时候再交代。

第五，它有一个近亲，叫科林格里奇困境（1980年）。 一项技术的影响，要等它被广泛使用之后才看得清，可到那时它已经很难改了。他说的是信息不足的难。本书说的是更硬的一层：不是信息没到手，而是要预见的东西，在封装的时候还没有发生。

第六，什么情况下这条定理是错的？ 回到导航：如果分流版上线几年之后，城市里没有出现任何一类设计者当初没有列进清单的新争议，那这个补丁就真的补到了头，错位定理在这个例子上就落空。

◆ 七、回扣第三命题

本书的第三条承重命题，到这里完整了：

> **错位定理：新SDE必在封装之外，因为封装之时它还不存在。**

它是前面两章的自然推论，不是一个新的断言。这也是为什么说本书的核心是"证明"，不是"举例"：洗衣机、导航、铁路这些故事，可以全部抹掉，三段论依然站着。故事让它好懂，不是让它成立。

◆ 八、小结

工程师的补丁补的是昨天的洞，补上了，今天又冒出新洞；模拟使用者也不行，因为被模拟的人会因模拟而变，这是一个追自己影子的圈。三段论本身极简单：封装只装已发生的；碰撞出的新东西在使用中才发生；所以新东西必在封装之外。它依赖“发生是真发生”这个前提，只证明“有”而不证明“多大”，也没有说谁来处理。

那么，新东西既然总在封装之外，它出现以后，会逼出什么？

> 带走的话
>
> 补丁补的是昨天的洞，补上了，今天又冒新洞。你没法在镜子里提前看到自己照镜子之后的样子，封装也没法把使用之后的世界先装进去。

> **小账**　这一章立了什么？——命题三，错位定理：封装只装已发生的，新SDE在使用中才发生，所以必在封装之外。　错在什么情况下？——"发生是真发生"这一前提被推翻，或补丁上线几年后，城市里没有出现任何清单之外的新争议。

承接：下一章看"二阶碰撞"。当新SDE撞上了旧封装，想要处理它，就需要一种比原来那一级更高的智慧。

第 8 章

二阶碰撞与升级律

◆ 一、先从一个新手做饭的人说起

有一个人，刚开始学做饭。他买了一本菜谱，菜谱上写得很清楚：盐五克，油十毫升，大火三分钟。他照着做，第一次做出来的菜，竟然像模像样。

这本菜谱，就是一个封装。里面是很多厨师试出来的数字，固化、可转让、隐去过程，样样齐备。

第二次，他换了一口锅。同样五克盐，菜却咸了。他不明白：数字没变，怎么结果变了？一查才知道，新锅的水分蒸发得慢，味道留得更重。他于是学到一个菜谱上没写的东西：盐要看锅。

第三次，家里来了一位不吃辣的客人，他得把菜谱改一改。第四次，他发现冬天的菜和夏天的菜，口味的底子都不一样。慢慢地，他不再翻菜谱，他在"看着办"。

这个过程里有两层事情。

第一层：菜谱（封装）和他的厨房（土壤）相遇，对不上，于是他学到了"盐要看锅"。这是第6章说的碰撞，我们把它叫一阶碰撞：工具与使用的现场相撞，生出封装之外的新东西。

第二层：他把"盐要看锅"这个新东西，加进自己的做法里。这时他的做法，已经不是原来的菜谱了，是菜谱加上他的新发现。可是这个新做法，在下一次遇到别的情形（客人不吃辣）的时候，又会对不上。他于是又加了一条，再加一条。直到有一天，他意识到：不是菜谱缺条目，是菜谱这种"照着数字做"的方式本身有局限。　于是他开始理解"火候""味道的平衡""食材的脾气"这类菜谱写不出来的东西。他的做饭，升了一级。

这第二层，就是本章的主题：二阶碰撞。

◆ 二、二阶碰撞是什么

一阶碰撞，是工具的封装与使用现场相撞，产生新SDE。

二阶碰撞，是我们对一阶碰撞结果的回应（也就是新的补丁、新的规矩、新的封装），再与更广的现场相撞，产生下一轮新SDE。

换个说法：一阶碰撞暴露的是"这个封装不够用了"；二阶碰撞暴露的是"我们补封装的这种办法，也不够用了"。前者逼人补一条，后者逼人换一种想法。

用本书的话讲，二阶碰撞的特征，是被碰撞的对象从"工具"变成了"处理工具的办法"。第一轮，我们问："这个工具怎么改？"第二轮，我们开始问："我们改工具的这套思路，对不对？"

下面用本书的日常案例再走一遍，把二阶看得更清楚。

◆ 三、导航：从改软件到改城市

回到那个导航。

一阶碰撞：所有人被导进同一条小巷。工程师的回应：软件分流。这是对封装的修补。

二阶碰撞：分流之后，新问题冒出来了。司机学会了与导航博弈，商家抱怨流量被平台左右，居民发现原来被保护的安静小街，现在反而因为分流，也有了零星的车辆经过。更重要的是，城市的管理者发现：导航公司的算法，正在实际上决定着城市的交通。　一个原本由公共规划决定的东西，悄悄被私人的软件接手了一部分。

这时再继续打软件的补丁，已经不够。因为问题已经不是"软件不够聪明"，而是"谁有权决定车流怎么分配"这样的问题。　这是一个关于规则、权责、公平的问题。它需要城市管理者、导航平台、居民代表一起，坐下来商量。商量的结果，可能是一套新的交通数据共享规定，可能是对小街的保护规则，也可能是一种新的出行时间错峰制度。

请看这个变化：一阶的回应，是工程师写代码；二阶的回应，是制度。从程序到制度，不是"同一类事做得更多"，而是换了一个层次。 能够处理这一层的人，需要看懂的不再只是路和车，而是路、车、软件、居民、规则之间的关系。他的眼界，比原来写代码的人，高了一级。

◆ 四、洗衣机：从省力到重新分工

洗衣机的例子也有这一层。

一阶碰撞：省力的机器，没有省出时间。于是人们的回应是：发明更多的辅助家电，烘干机、熨烫机、洗碗机，一件一件补。

二阶碰撞：家电越补越多，家务并没有真的少。这时候，一部分人开始问另一类问题：家务到底该由谁做？为什么默认是同一个人做？ 这是一个关于家庭分工、性别观念、社会期待的问题，不是电器能回答的。它引出了新的讨论、新的家庭契约，也引出了专业家政和托育这类新的社会服务。

换句话说，当补丁一再补不平的时候，人们终于被逼得把目光从"工具"移到"工具所在的关系"上。 这是智慧的升级。它不是靠更精巧的洗衣机达成的，是靠对"家"这个整体的重新理解。

◆ 五、升级律

把几个例子的共同结构抽出来，可以得到本章的核心命题，我们把它叫升级律：

> 每一次二阶碰撞，都要求比原来那一级更高的智慧来回应。

这里的"更高一级"有一个具体的含义，不是随便说的"更聪明"：更高一级的智慧，是能把下一级的封装连同它的外部，一起当作对象来看的智慧。

做菜的人，原来的智慧是"照着菜谱做"，对象是食材和火候。升级以后，他的智慧是"知道菜谱为什么这样写，以及什么时候不该照着写"，对象变成了菜谱本身与厨房。工程师原来的智慧是"让软件算得更准"，对象是路径。升级以后，管理者的智慧是"让软件和城市怎样共处"，对象变成了软件与城市的关系。

组织学习的研究者阿吉里斯和舍恩（1978年）把只改做法的学习叫“单环”，把连目标和规则一起质疑的学习叫“双环”。本书和他们相同：补丁不灵的时候，要换到质疑规则的那一层。不同有两点：他们讲的是个人和组织可以选择的学习方式，本书讲的是被碰撞逼出来的升级；据我所知，他们也没有主张这条层级链永远走不到头，而本书说这个楼梯没有顶。

你可以把它想象成一个楼梯。每一级台阶，都是一种看世界的高度。站在一级台阶上，你能看见这一级的所有东西，却看不见这一级本身是怎么砌起来的。要看见它，你得上去一级。而每上一级，你又会遇到新的、在这一级上看不见的东西。升级律说的是：这个楼梯是没有顶的。

◆ 六、为什么没有顶

没有顶，听起来像一个大话。看看它凭什么。

回想第7章的三段论：任何封装，新SDE必在其外。现在假设有人造出一个"终极封装"，宣称它把一切都封进去了，包括所有未来的碰撞。

那么，请这个终极封装被使用。它一旦被使用，就进入了某个真实的现场，与使用者的E相撞。根据第6章，碰撞必生新SDE；根据第7章，这个新SDE必在封装之外。所以，这个"终极封装"并不终极，它也有它的外面。

于是，要处理它的外面，就需要一种把它当作对象的智慧。这种智慧，按照定义，比它高一级。而这个更高一级的智慧，如果也被封装成工具，又会重复同样的过程。

有人会说：让这个终极封装自己学、自己升级，它就不是静止的封装了。可一个会自己升级的系统，升级的依据仍然是它已经遇到、已经处理过的碰撞，每一步

都是把刚发生的东西封进自己，所以它仍然晚于碰撞半步。至于“什么算问题、值得升”，还要回到下面要说的主体前提。

这个论证的关键，可以用一句话说出来：任何封装，只要被使用，就是一级，就有它的外面；而"外面"需要更高一级的智慧去看；更高一级的智慧一旦封装，又是一级。　　这条链没有终点，不是因为它长得望不到头，而是因为每一环的结构里，都包含着下一环的入口。

这就是整个第二编的定音句：

> 工具封装的是昨天的智慧，逼出来的是明天的智慧。

◆ 七、一个必须明说的前提

读到这里，聪明的读者会提出一个尖锐的问题：你这个论证，证明了"需要更高一级的智慧"，可没有证明"这个智慧必须是人"。如果AI本身就能做这个升级，不就完了吗？

这个问题正好指向本书的"谁来升级缝"，必须正面回答。

诚实的回答分两步。

第一步：本章的论证，确实只证明了"永远需要更高一级的智慧"，没有证明"必须由人来提供"。如果AI可以提供，论证并不受损，它只说明智慧的升级可以有AI参加。

第二步：但是，有一个前提，本书明确把它立在这里，并说明它是一个前提，不是被证明了的结论：问题，总是对某个主体而言的问题。　　导航的拥堵之所以是"问题"，是因为有人不愿意被堵，有居民需要安静；洗衣机的"家务没少"之所以是问题，是因为有人觉得不公平。没有这些主体的不愿意、不满意、不甘心，拥堵只是一个数字，家务只是一个时间表，其中没有任何"问题"的成分。换句话说，E（那片纠缠的、有意义的土壤）的源头，在主体。

如果接受这个前提，那么"哪些新东西算问题、值得升级"这件事，始终要回到某个有处境、有需要的主体那里去。 AI可以帮助主体看见、计算、模拟，但它所面对的问题，终究是向着某个主体敞开的。这并不意味着人类对升级拥有垄断权，只意味着升级的"方向感"，离不开主体。

读者完全可以不接受这个前提。接受或不接受，会导致不同的结论：若不接受，升级的链条依然成立，只是链条上的行动者可以是任何有主体性的存在；若接受，链条上的升级就与人（以及与人相互纠缠的存在）深深绑在一起。本书采取后一种立场，并在第九编再展开。

◆ 八、升级律可能错在哪里

第一，"更高一级"不是一个可以拿尺子量的东西。 它是一个结构上的说法：能把下一级当对象。在具体事例里，哪一步算"升级"，哪一步只是"同级的扩展"，常常有争议。比如，导航公司增加一个新功能，是升级还是扩展？有时候要等过一段时间才看得清。本书并不假装这个界线是清晰的。

第二，二阶碰撞不总是被人们注意到。 历史上有很多次，本应触发升级的碰撞，被忽略、被压住、被延误了。升级律说的是"要求"，不是"一定做到"。文明可以不回应这个要求，代价是停滞，或者崩溃。所以，升级是必要的，不是自动的。这一点，在第六编谈个体层与文明层的区别时还会提到。

第三，这条链"没有顶"，不等于每个具体问题无解。 一个具体的拥堵问题，可以被解决；一个具体的数学题，可以被证明。"无尽"说的是智慧作为整体的结构，不是说你每天都在解一个解不了的题。恰恰相反，在每个具体的台阶上，问题常常是可以解的，只是解完一个，台阶上方又露出新的台阶。

第四，升级律什么时候是错的？如果某个二阶问题，也就是补丁一再补不平的那种，最后只靠同一层的更多补丁就平息了，没有引出新的规则、权责安排或看问题的框架，那它就不是二阶碰撞，升级律在这个案例上落空。比如导航的争议，若几年之内只靠再改几版算法，就没有人再争“谁有权分配车流”，我在这里的分析就是错的。

◆ 九、小结

学做菜，从照着菜谱，到明白菜谱为什么这样写；城市处理导航，从改软件到改规则；一个家处理洗衣机，从买更多电器到重新想谁来做家务。每一次，都是补丁补不平之后，眼界上了一级。推到一般：任何封装，使用就有碰撞；补丁再碰撞，便逼出更高一级的智慧；这一级再封装，又是一级。这个楼梯没有顶。论证没有说升级者必须是人，但把“问题总是对主体而言”立成了前提。

这个结构，在经济和产业的层面，有一个更宏大的样子：不只是一个人、一座城市，而是整整一个行业，都会经历同样的几个阶段。

> 带走的话
>
> 菜谱不灵了，先添一条；添了还不灵，就该问问为什么非要照着数字做。补丁补不平的那一刻，就是眼界该上一级台阶的时候。

> **小账**　这一章立了什么？——升级律：每一次二阶碰撞，都要求更高一级的智慧，楼梯没有顶；并明说"问题总是对某个主体而言"是前提。　错在什么情况下？——某个二阶问题只靠同一层的更多补丁就平息，没有引出新规则、新权责或新的看问题的框架。

承接：下一章讲工具让渡律的四阶段，一碗面条的账，会帮我们看清它。

第 9 章

工具让渡律的四阶段

◆ 一、一碗面条的账

街角有一家小饭店，老板姓周，招牌是一碗手擀面。十几年前，面是他自己揉的：天不亮起来和面，醒面，擀开，切条。一天最多做六七十碗，手腕常年酸痛。一碗面卖十二块，他觉得，这是一门手艺活。

后来，他买了一台和面机和压面机。这是个好东西。面和得又匀又快，压得又薄又齐，一天可以做二百碗，他的手腕也终于歇了。

这个故事在这里，就是前面几章讲的"封装"：无数面点师傅试出来的和面、压面的智慧，被装进了一台钢铁机器。接下来发生的事，就是本章的主题：一个行业，在一件好工具的冲击下，会经历四个阶段。　本章把它叫作工具让渡律。我们跟着周老板的面馆，把四个阶段一个个走一遍；然后再换到一个看起来离面条很远、结构却很像的场景：AI做数学。

需要提前说明：这四个阶段的整体框架，作者在另一部专著《新文明的诞生》里已有专门展开，本章是从"工具发生律"的角度，把它放回到这本书的论证链条里，两本书可以互相参照。本章里的面馆，是一个设想出来的小账，用来说明结构，不是对哪家店的实录，里面的价格也是设想。

◆ 二、第一阶段：扩大期

工具刚出现的时候，最直接的结果，是同样的人，能做得多得多。

周老板的面馆从一天六七十碗，变成一天二百碗。他雇了两个帮工，加了几张桌子，又开了一个窗口卖外带。街上的人都说：周记面馆红火了。因为和面机太好用了，附近别的面馆也纷纷买了机器，扩大了店面。

这个时期的特点是：价格没有太大变化，数量成倍增长，大家都能赚到钱。谁先用上新工具，谁就占了先手。人们谈起这项新技术，满口都是"效率"和"机会"。市场里充满了乐观的气氛。

这个阶段，恰恰就是本书后面要重点讲的"AGI感"最强的时候：因为工具刚到，封装已经完成，它的威力一眼可见，而它的外面，那些新SDE，尚未显露。人们看到的是一个"什么都能做"的工具，和一个"好像什么都不用愁"的前景。

◆ 三、第二阶段：内卷期

可是好事不会一直下去。

当和面机、压面机成了每家面馆的标配，一件事就变了：大家都能一天做二百碗。 二百碗不再是本事，是起点。

街上的面馆越来越多，每一家都能快速地出大量的面。顾客不会因为"你家的面更快出锅"而多吃一碗，胃是有限的。供给涨了，需求没涨，结果就是价格往下走。十二块的面，渐渐有人卖十块，再有人卖八块。

为了不被挤出去，周老板也降价。但降价意味着利润薄了，他要想办法省成本：面粉换便宜的，帮工少雇一个，营业时间拉长，从早上五点开到晚上十一点。他比以前更累，赚的却不比以前多，有时还不如以前。

这就是内卷期。它的特点是：每个人都比以前更努力，用的是同样的工具，做的是同样的事，结果谁也没有多得到什么，只是都更累了。 原本由新工具带来的好处，被竞争压平了，转化成了更低的价格，惠及了顾客，却没有留在做面的人手里。

内卷期与前一章讲的"补丁"也有关系。内卷，本质上是在一级台阶上反复补丁：更快、更便宜、更多，所有的竞争都发生在封装好的那个S之内。用同一个封装，比谁用得更狠。这条路越走越窄。

◆ 四、第三阶段：新产业诞生

内卷持续到某一点，会有人想：既然大家都在一件事上挤，那我换个方向。

周老板的邻居，一位姓林的年轻人，不再跟着降价。他想的是：能不能做一碗只有我这里才有的面？他开始问顾客：你喜欢硬一点还是软一点？汤想要清淡还是浓厚？他记下常客的口味，记下谁不吃葱、谁要多放醋。他甚至做了一个小本子，每位熟客一页。

他给自己的店起了个新名字，叫"我的面条"。顾客来了，他会说：您上次说想要汤清一点，我今天特意给您调了。同样一碗面，他卖三十块，而对面老周的面只卖十五块。奇怪的是，有人愿意付三十块。

为什么？因为这碗面里，多了一个东西："我"。 顾客买的不再只是一碗面，而是一碗"为我做的面"，一碗被记住、被认出来的面。和面机封装的，是"一般的面怎样做得好"；林老板新长出来的，是"这一位顾客需要什么样的面"。这个东西，在和面机被造出来的那一天，不存在，也不可能存在，因为它产生于一次又一次具体的相遇。

这就是新产业诞生。请把它和第7章的错位定理对照：工具封装之外的新SDE，终于长成了一个能养活人的产业。

这里"15元"和"30元"的数字是设想，不是调查。我们关心的是结构：当通用的做法的价格，被工具压向地板的时候，新价值会长在"不通用"的地方。 关于这一点，作者在《新文明的诞生》和另一部关于"我经济"的专著里，有更系统的论述。

邻近的经济学里，相近的说法不少。熊彼特讲“创造性破坏”：旧产业被新产业取代。阿西莫格鲁和雷斯特雷波讨论自动化时也强调，机器替掉一部分任务的同时，还会有新任务被创造出来。〔待核：二人相关论文的年份与表述〕本书和他们相同：旧位置被取代，新位置同时长出。不同在于，本书把次序讲死了：先有扩大，再有内卷，内卷把通用做法压到地板，新价值才长在“不通用”的地方；而让渡成不成立，要看腾出的位置有没有被占住。

◆ 五、第四阶段：让渡

这时候，还有最后一个动作，也是最容易被误会的一个：让渡。

随着新产业的长大，原来那些被封装的活儿，渐渐不再由人来做了。揉面、压面，几乎没有人再用手去做，连年轻的面点学徒也不再练这一门。他们改学如何与顾客交谈、如何设计一碗面的口味、如何讲述一碗面的来历。

有人会说：这不就是手艺失传了吗？不就是退化吗？

本书的第五条承重命题，要在这里说清楚：

> 被封装能力的退场是让渡，不是退化。

区别在哪里？退化，是一个能力丢了，没有任何东西来接替，人变得比原来更弱了。让渡，是一个能力被交给了工具，人腾出来的位置，被一个更高一级的能力占住了。揉面的手艺，交给了机器，人的手腕和精力，换成了识人、调味、讲故事的本事。后者是以前没有的，而且是以前没有地方发展的。

要特别注意"占住"这个词。如果腾出来的位置，没有被新的东西占住，那就真的是退化了，因为空出来的地方，只是空着。让渡的成立，依赖于新产业的诞生；没有第三阶段，第四阶段只是一个空洞。这一点，第六编里谈"导航与海马体"时还会回来。

于是，四个阶段合在一起，是这样一条线：

- 扩大期：工具到来，同样的人做得更多，皆大欢喜。
- 内卷期：人人都有工具，竞争压平好处，所有人都更累。
- 新产业诞生：有人在封装之外长出新的价值，一个新的行业成形。
- 让渡：被封装的旧能力退场，人转到新的位置上。

这四步，就是工具让渡律。它和前几章的关系很清楚：封装（第5章）→碰撞（第6章）→新SDE在封装之外（第7章）→逼出更高一级的智慧（第8章）→这个升级在产业层面表现为四阶段（本章）。

◆ 六、换一个场景：AI做数学

现在，把面条换成数学。

假设有一位数学教授，一年能解出二十个问题，发二十篇论文。在他的领域里，这已经是很高产的成绩了。

现在，AI出现了，而且特别擅长这类工作：给定一个明确的猜想，它能快速地尝试各种证明路径，常常找到人想不到的办法。假设有了这样的工具，教授一年可以解两百个问题，发两百篇论文。（这是设想，用于说明结构，不是对任何已发生事件的描述。）

按照四阶段，我们依次看：

扩大期：同样的教授，产出翻了十倍。期刊的稿子涌入，学术圈兴奋。人们说：数学要迎来黄金时代。这个阶段，不少人会觉得"AI数学家"已经来了。

内卷期：很快，每一个数学家都有同样的工具。"解二十个问题"不再是本事，"解两百个"也是起点。问题库被快速消耗，容易的问题很快被一扫而空。论文数量暴涨，评审的人看不过来，期刊的门槛被迫提高，评价体系开始失灵：以前以论文数量衡量一位学者，现在数量已经没有意义。大家更忙，却越来越难说清楚什么是"有价值的数学"。

新产业诞生：到这一步，问题会被换一种问法。不再问"你解了多少题"，因为解题已经是工具的事。问题变成："你发生了哪些新数学？" 也就是说：你提出了哪些以前没人想到的概念？你发现了哪些让旧问题换了面貌的新视角？你开辟了哪些此前不存在的研究方向？这类东西，恰恰是第7章说的"封装之外"的东西：解题的工具封装的是已有的数学，而新数学是在使用这个工具的过程中，碰撞出来的。

打个比方：在内卷期，数学家像是在给定的棋盘上比谁下得多；在新产业期，他们开始问：这个棋盘本身，能不能改?

让渡：证明里那些繁琐的、技术性的推演，让渡给机器，数学家把时间腾出来，用在提出概念、判断什么值得研究、把新的对象和旧的世界连起来上。这些是此前没有太多时间做的事。

和面条的故事比较，结构是一模一样的：通用的、可封装的部分，价格（或者说，价值）被压到地板；不通用的、"我"的部分，成为新的价值所在。

关于数学自身在工具史里的位置，第四编的最后一章会详细讲，那一章的主问题，正是"从你解了哪些题，到你发生了哪些新数学"。

◆ 七、把四阶段连回AGI感

有了这个四阶段框架，本书的第五层论证（AGI感是时间差现象）就有了一个落脚点。

在扩大期，工具威力最大，新SDE还没显露，人们只看到封装带来的成果，“这个工具已经通用”的感觉最强，这是AGI感的高峰。进入内卷期，工具人人都有，处境却没变好，惊喜转成焦虑，AGI感开始下降。到新产业诞生，人们看到工具并没有“做完一切”，反而催生了想不到的新东西，AGI感自然淡去。

这与我们说的"强弱与碰撞深浅成反比"（第十条承重命题）是一致的：工具越强，在扩大期的AGI感越强，但同时，使用的范围越广、碰撞越深，随后的新SDE也就越丰富。这一节只是一个预告，具体机制留到第十编。

第7章讲过的索洛悖论和戴维的电动机，在这里也能对上：滑过去的那段空档，就是从封装到新SDE长成产业所需要的时间。旧的世界已被打破，新的世界尚未成形，人们往往最困惑，也最容易作出“要么万能、要么无用”的极端判断。

◆ 八、四阶段不能当时间表用

第一，四阶段是一个典型的结构，不是时间表。 本书不预言任何一个行业会在哪一年进入哪一个阶段，也不认为每个行业都会走完四步。有的行业可能停在内卷期很久，有的可能很快跳到新产业。周老板的面馆，换成了别的行业，每一个阶段的长短、顺序，都会有差别。

第二，新产业诞生，不保证每个人都能搭上。 林老板找到了"我的面条"，周老板可能没有。让渡律说的是行业层面的结构，不是对每个从业者的承诺。个人是否能跨过去，要靠本人的碰撞和升级，这就是后面第六编要讲的"个体层与文明层的区分"。

第三，数学例子是一个设想，它并不等于预言AI一定会这样改变数学。 它用来说明：如果这个工具真的这样强，那么价值会向哪里移动。至于现实中会怎样，要由事实来回答。作者希望读者看清的是推理的结构，不是对数学界未来的断言。

第四，四阶段什么时候是错的？如果找到一个行业，新工具普及十几年后，通用做法的价格被压到地板，从业者收入长期下滑，却始终没有长出任何新岗位和新的价值来源，那它走的是退化，不是让渡。偶尔有这样的行业，不推翻本章；但如果它们成了常态而不是例外，四阶段就不是一条律，只是少数幸运者的故事。

◆ 九、小结

一碗面条走完了四步：和面机让手艺人做得更多（扩大期），人人有机器让大家更累（内卷期），有人长出了“我的面条”（新产业诞生），揉面的手艺让给了机器而人去做了更高一级的事（让渡）。AI做数学的设想走的是同一个结构：解题成了工具的事，数学家转向问“你发生了哪些新数学”。四阶段是封装、碰撞、错位、升级在产业层面的样子，合起来是一次工具的完整生命。

第二编到此结束。本书已经有了一条不依赖例子的论证主线：工具封装昨天的智慧；使用即碰撞；新SDE必在封装之外；二阶碰撞逼出更高一级的智慧；产业层

面表现为四阶段。接下来的三编，会拿这条主线去过一遍人类的工具史：先看现实界，再看理念界，再看自我界。

> 带走的话
>
> 手艺让给机器不要紧，要紧的是腾出来的位置有没有被占住。占住了，叫让渡；空着，才叫退化。

> **小账** 这一章立了什么？——工具让渡律四阶段：扩大期、内卷期、新产业诞生、让渡；通用的部分被压向地板，新价值长在"不通用"的地方。 错在什么情况下？——某行业通用价格被压到地板、收入长期下滑，却始终长不出新岗位或新价值来源，而且这成了常态而不是例外。

承接：第三编从火开始，看现实界里最古老的几件工具，怎样一次次撞出封装之外的世界。

第 三 编

现实界工具史

编 序

前面两编，我们把原理立了起来：工具是已发生智慧的封装；使用工具就是一次碰撞，碰撞必生封装之外的新东西；这个新东西在封装的时候还不存在，所以工具自己装不下它。原理讲得再顺，也只是讲。这一编开始，我们把原理放到历史里去验。验的方式很朴素：挑六件现实界里的大工具，一件一件看——它封装了什么，人们拿它和谁碰撞，碰出了什么当初没人想到的东西，最后逼出了什么更高的智慧。

先说一句边界。六个例子，不是证明。证明在第九编，是结构性的。这六个例子的分量是另一种：如果原理是对的，历史上就该看到这种形状；如果六件大工具里有一件是“装进去就完事、再无后话”的，原理就该打个问号。我们认真看，也认真留着那个问号。

再交代一句：这条思路并不是凭空来的。默顿早就讲过，有目的的行动会带来没打算要的后果；罗森堡讲过，一项技术的瓶颈和失衡，会指出下一项技术往哪里走；布雷斯纳汉和特拉伊滕伯格讲过，通用技术要等配套的发明跟上才发力；佩蕾丝把一次技术革命分成“导入”和“展开”两段。这些说法，本书都用得上。我们和它们相同的地方，是都从历史里看形状。不同的地方在于，它们多半把“没想到”当作经验上常见的现象，本书要问的是一个更硬的问题：这个“没想到”，能不能是原理上必然的？封装的时候，那个新东西还不存在，所以不可能在封装里。

这一步站不站得住，留给第九编去检验；这六章，只是先看看历史有没有留下反例。

这一编六章，每章一件工具：

章	工具	这一章看什么
10	火	从熟食走到金属、历法与水利
11	犁与农耕	粮食多了，账、边界、官和国家跟着来
12	钟表、铁路、电报	从时刻表一路问到“同时是什么”
13	蒸汽机	机器先到，热力学后到
14	电力与洗衣机	等了几十年，省下的时间去了别处
15	计算机与网络	软件危机与网络安全

放进全书五层论证里，这一编属于第四层“三界工具只会提升三界智慧”的现实界一块。它给的是历史佐证，不是证明；结构性的证明，要到第九编。

第 10 章

火：从烤肉到冶金，再到历法与水利

◆ 一、一口锅，一盆火

夏天夜里，路边烧烤摊。炭火红着，师傅翻着串，孩子蹲在旁边盯着火，大人在等。

这一幕里有一个很容易错过的事实：火，是我们这个物种最早的“工具”之一，而且是一件我们至今没有离手的工具。你在家里按下燃气灶的旋钮，蓝色火苗窜起来，你并没有觉得自己在使用一项几十万年前的发明。可它就在那里。

这一章要讲的，是从这堆火开始，一环扣一环生出来的一长串东西。它们大多不是点火的人想要的。

火封装了一种能力：把“生的”变成“熟的”，把“散的”变成“可控的”。可是人一旦围着火生活，生出来的事情，远远超出了“熟”和“可控”。

◆ 二、第一圈：火封装了什么

控制用火的最早证据，学界的说法并不统一。常被引用的是南非旺德沃克洞穴（Wonderwerk Cave）里约一百万年前的灰烬和烧过的骨头，由弗朗切斯科·贝尔纳等人2012年报告；也有学者认为，经常性地用火要到大约四十万年前之后，才在旧大陆普遍起来。年份不必纠缠，只需要一个共识：在智人出现之前，人类的祖先就已经在和火打交道。

火的第一层封装，是“别人已经想明白的事”。第一个把一块肉放在火边的人，发现它更香、更软、更不容易让人生病。这个发现一旦被记住、被教给孩子，就不再需要每个人自己重新发现。它变成了一套做法：怎么起火，怎么保火种，怎么控制火候。工具的定义，在这里看得很清楚：封装的不是那堆木头，是“控制火”这件事背后的一套已经发生过的理解。

关于烹饪的后果，有个很有名也很有争议的假说：熟食好消化，肠子于是可以变短，省下来的能量去养了更大的脑子。这个说法的“熟食假说”一支，主要由灵长类学者兰厄姆（Richard Wrangham）提出；“昂贵组织假说”则出自艾洛（Leslie Aiello）与惠勒（Peter Wheeler）。学界至今有人支持、有人反对〔待核〕。我们不依赖它，下面的事不管它成不成立，都是确定的。

◆ 三、第二圈：火和谁碰撞

火遇到了三样东西。

第一样，是黑夜。白天的时间属于觅食和干活，夜晚原本是一片只能躺着熬过去的黑。火把夜晚变成了可以用的时间。人类学家波莉·魏斯纳（Polly Wiessner）2014年在非洲南部朱/霍安西人（Ju/'hoansi）中记录过：白天的交谈多是生计、抱怨和闲话，讲故事只占很小一部分；夜里围着火堆，讲故事的比例猛升到八成上下，讲远方的人、过去的事、别的群体的事。这是对今天采集社会的观察，不能直接当成几十万年前的记录，但它指出了一种可能：火不只照亮了夜，还让夜里多出一种白天没有的交谈。故事、想象、对不在场的人和事的共同记忆，由此有了土壤。

火是现实界的工具，一个晚上就跨进了自我界（“我们是谁，我们从哪里来”）和理念界（“故事”这种可以传下去的东西）。这是承重命题4最朴素的一个例子：工具不会停在出生的那一界。

第二样，是泥土。火边的泥土被烧硬了，成了陶。陶的出现远早于农业，日本列岛出土的陶器可以早到一万多年前甚至更早〔待核〕。陶罐能煮、能存、能运。人类第一次有了不怕火的容器，也就第一次有了“煮”这种做法，粥、汤、发酵的东西，都在这里出现。

第三样，是石头里的颜色。有些石头放进很热的火里，会流出东西来。孔雀石是绿的，烧到足够的温度，会析出铜。冶金的起点，很可能就是陶窑——需要高温的陶窑，让人们第一次看到矿石变成金属。至于具体怎么发生，学者们有不同的猜测，我们只取一个大方向：窑火和矿石碰撞，碰出了金属〔待核〕。

此后的线索，教科书里写得很清楚：红铜，青铜（约公元前三千年前后进入多个地区），铁（约公元前一千二百年前后在近东一带逐渐普及）〔待核〕。每一步，都要求更高的炉温、更好的鼓风、更细的配料。冶炼者在和火碰撞，也在和矿石的脾气碰撞。

◆ 四、第三圈：生出了什么新的SDE

这是最有意思的一段。火被封装的时候，那个起火、保火的人想的是取暖、驱兽、煮熟。他不可能想到下面这些事：

金属，带来了新的“东西”。铁犁头、铜刀、铁斧、盔甲、钱币。这些东西本身又成了新的工具，也成了新的问题——谁拥有铁？谁掌握炉子？金属的战争和金属的交易，都是火从来没打算负责的事。

木头，成了稀缺品。炼铜炼铁要大量木炭。古代冶炼区周围的森林被砍光，是一个反复出现的情节。后面第13章我们会看到，英格兰在十六七世纪的缺柴问题，是怎样把人逼向了煤，又怎样把人逼向了蒸汽机。火在一千年里“吃掉”了森林，这是任何一个烧第一堆火的人都想不到的。

城市，成了会着火的地方。火从一家的灶台，变成了一整个街区的危险。1666年伦敦大火烧毁了城里大片地区，此后火灾保险和专业的救火队逐渐成形；有记载说，1680年前后伦敦就已有人专门开办火灾保险，尼古拉斯·巴尔本是常被提到的名字。一个原本只关于灶台的东西，最后生出了保险和消防，也生出了建筑法规里的防火墙、防火间距。

一个更大的、几乎看不见的新问题：怎么管火。火要有人看，火种要有人保，烧荒要有规矩。很多传统社会里，守火是一项严肃的责任，后来的“灶神”“圣火”“祭火”都是这个责任在自我界和理念界的回声。管理火的人需要信用，需要分工，需要合作。从这里，你已经能看到“社会”的一点轮廓。

这四样东西——金属、缺柴、火灾、管火——没有一样在“火”的封装里。它们都是人们长期使用火、让火与森林、城市、人群碰撞之后，才发生出来的。这就是承重命题3：错位。封装时它们还不存在。

再看一眼这四样东西住在哪里。缺柴，既不在火里，也不在森林里，它长在火和森林纠缠在一起、又被放大了几百倍的地方；火灾也一样，灶台里没有它，城市里也没有它，它长在灶台和挤在一起的木屋之间。这就是封装的人看不见的原因：他把手里的东西查到天亮，也查不出这些，因为它们既不在“火”这一边，也不在“人群”这一边，而在两边纠缠的地方。后面每一章，我们都问同一个问题：这一回，新东西长在哪里？

◆ 五、第四圈：逼出了什么更高的智慧

火碰撞出来的新问题，逼着人们去学“关于火的更高一层的知识”。

看冶炼。早期的冶炼靠手艺：看火色，听风声，摸经验。怎么让铁变硬而不脆？古人用反复锻打、渗碳、淬火积累了大量经验配方，却说不清里面发生了什么。要到近代，人们才说清：钢和铁的区别，主要在碳的含量和分布。这是冶金学，也是后来材料科学的前身。

看火的利用效率。炉子怎么设计才不浪费热？风怎么送才烧得旺？这一类问题，几百年后被拧成了一门学问，叫热学。热学的故事，我们留给第13章。

也就是说，火这个最古老的工具，在几十万年的尺度上，一直在催生“关于火的更高一层的知识”。老做法装好了，新学问却一直在被它逼出来。

◆ 六、延伸：历法与水利，兼谈火为什么是个好的开头

说到这里要诚实交代一件事：历法和水利不是火的直系后代，它们属于另一条线，天文观察和大规模协调。和火只有两处血缘。一处在物质上：烧砖的窑、烧石灰的窑、铜铁的锄头与凿子，让开渠、筑堤、建仓成为可能。一处在社会上：管火

练出了“为一个共同的重要东西轮流值守、分工、立规矩”的本事，几百人一起挖一条渠，才不至于乱成一团。

历法来自农业的需要：什么时候播种，什么时候河水要涨。古埃及人把天狼星清晨第一次在东方地平线上出现，与尼罗河的泛滥联系起来〔待核〕，后来的各种历法，都是把天上的节奏和地上的节奏对齐。水利更直接：水多了是灾，水少了也是灾，一条渠要几百人挖，要有人记谁出了多少工、谁该分多少水。

于是，你看到一条链：火——窑——金属与砖——工程——记录与协调——历法与官僚。链条里的每一环，都不是前一环预先设计的，但每一环又离不开前一环。这有点像做菜：你本来想煮熟一锅饭，煮着煮着，发现自己在开一家饭店。

为什么把火放在现实界工具史的第一章？

一个原因，是它古老到我们无法假装知道“当初的人是怎么想的”，只能老老实实看后果，不去编故事。

另一个原因，是它足够日常。燃气灶、热水器、暖气、发电厂、发动机，火的链条从未断过，只是你没有看见。

还有第三个原因，我觉得更重要：火是最能说明“工具不会停在出生那一界”的例子。它出生在现实界——一团物理的燃烧——可是第一个晚上，它就在自我界里造出了“我们围坐在一起”的感觉，在理念界里造出了可以讲述的故事。如果你要找一个“一界有工具，三界有新生”的最小样本，火几乎是现成的。

你也许会说，这些关于史前的推测，证据太薄。是的，薄，所以每一处我都加了提醒。但这个论点不靠这一个例子，它靠后面五章加在一起的形状，和第九编的结构论证。火的任务，只是让你先看见这个形状的最小版本。

◆ 七、回扣承重命题，以及这个论证的边界在哪里

火的故事把前四条命题串在了一起。控制火，封装的是“怎样生火、保火、控火”这套已经发生的理解（命题1）；火与黑夜、泥土、矿石、森林、城市碰撞，每次都生出封装之外的东西（命题2）；金属、缺柴、火灾、管火，全在原封装之

外（命题3）；而这件现实界的工具，一晚上就走进了自我界的“围坐”和理念界里可以传的故事与技艺（命题4）。

要诚实说三点。

第一，史前的材料非常有限。魏斯纳的田野观察是对当代采集狩猎社会的观察，不能直接当作几十万年前的记录；熟食与脑容量的关系，更是学界仍有争议。我们用它们做“画面”，不是做“证据”。

第二，“链条”总是事后看出来的。站在今天回头看，火→窑→金属→工程，像一条笔直的路；可当年的人，大概只觉得每一步是偶然碰上的。我们不能因为链条好看，就说它是必然。

第三，也是最要紧的：这一章只证明了一件事——有这样一个工具，它被使用后确实生出了封装之外的东西。它没有证明所有工具都如此。这个“所有”，要靠后面的结构论证。

还有一条，是这一章给自己立的试金石。如果能找到一个长期、大规模用火的人群，围绕火既没有形成任何分工和规矩，夜里的交谈也并不比白天更偏向讲故事，用火之后也没有分出窑和冶炼之类的后续技术，并且排除了环境上的原因，那么“火把人带到封装之外”这个说法，至少在这一章里就要撤回。迄今没有人报告过这样的人群，但“没有报告”不等于“不存在”。

◆ 八、小结

一堆火，一晚上，能把人带进讲故事的夜；一千年，能把人带进金属的世界、缺柴的困境、满城的火灾；一万年，能把人带进历法、水利和官僚。点第一堆火的人，什么都没有“设计”；可他点燃的东西，一直在逼人学更多。

下一章，我们离开火边，走进田里。

带走的话

点第一堆火的人只想着取暖和煮熟；后来的金属、缺柴、满城的火灾和守火的规矩，一样都不在那堆火里，它们长在火和森林、城市、人群缠在一起的地方，查火本身是查不到的。

小账 封装了什么？——怎样生火、保火、控火这套已经发生过的理解。 逼出了什么？——冶金与热学、保险与消防、守火的规矩，以及历法水利这一路的协调本事。

承接：火生出了金属和协调；金属做成的犁，会把人带进另一个更大的故事——粮食多了，问题为什么反而更多了？

第 11 章

犁与农耕：粮食多了，问题更多了

◆ 一、一个冰箱的故事

先从你家的冰箱说起。

刚搬进新家的时候，冰箱很空，你每天出门买菜，吃多少买多少。后来你买了大冰箱，又办了仓储超市的会员卡，一次买一个月的量。结果呢？你多了好几样新事情：要整理，要记得哪些快过期，要清理冰箱里的“化石”，要决定周末是不是该做一顿大扫荡。冰箱把“今天吃什么”这个问题解决了，却把“我家存了什么”变成一个新问题。

你可以把农耕看作一只全人类共用的、巨大无比的冰箱。

农耕封装的是“让土地按人的意思出粮食”的一套理解；而粮食一旦多出来、能存起来，人和人之间就出现了一批以前根本不存在的问题。

◆ 二、第一圈：犁封装了什么

农业的起源，大致在一万多年前，在西亚的新月沃地、中国的黄河和长江流域、中美洲等几个地方各自出现〔待核〕。这个“各自出现”本身很耐人寻味：不是一个地方发明了农业然后传到各处，而是多个地方的人，在类似的条件下，几乎在同一个时间段内走上了同一条路。

犁，是这套做法里的关键工具之一。最早的犁，是一根带尖的木头，由人或牲畜拖着，在地上划出沟来。美索不达米亚一带，用牛拉的犁大约出现在公元前四千年前后〔待核〕。后来加上了金属犁头——火和金属的成果，在这里与田地碰在一起——再后来，中世纪欧洲的重犁、挽具和马的使用，让北方潮湿而黏重的土地也能被耕作〔待核〕。

犁封装了什么？有三层。

第一层，是物理：翻土、松土、开沟。

第二层，是对季节与种子的理解：什么时候下种，种多深，间隔多少。

第三层，是对“力”的借用：人的力气不够，就借牛的、马的。借牛的力，就必须养牛，养牛就必须有草，有草就必须有地。

你看，这第三层已经预告了：一个犁头背后，拖着一整个体系。

◆ 三、第二圈：碰撞

犁和什么碰撞？和土地、牲畜、气候，当然。但最重要的碰撞，是和“人群”的碰撞。

碰撞一：和定居。种地需要守着地，不能今天在这里、明天在那里。人们定居下来，开始有了固定的村子、固定的屋子，也有了固定的垃圾和固定的水源污染。

詹姆斯·斯科特（James C. Scott）在《作茧自缚》（Against the Grain, 2017）中提出了一个很有挑战性的看法：早期的农耕聚落，并不是一首进步的凯歌。人口密集、与牲畜同住，使传染病更容易传播；饮食集中于少数几种谷物，营养反而可能变差；一些考古材料显示，早期农民的身高和健康状况并不比同时期的采集者好〔待核〕。斯科特还强调，谷物看得见、数得清、可以收税，块茎和野果却不行，所以国家偏爱谷物。这个观点学界有争论，我们取它的启发，不取它的全部结论。

碰撞二：和剩余。粮食有了盈余，就要存。存，就需要仓库；仓库，就需要看守；看守，就需要有人负责；有人负责，就有了“谁说了算”。

碰撞三：和“谁的地”。固定的地，就有了边界。边界，就有了争执。争执，就需要裁决，需要记下“这块地归谁”。

◆ 四、第三圈：生出了什么新的SDE

按顺序数一数。

新的“差异”：人和人的不同，第一次可以被积累。采集狩猎的社会，财富很难积累，因为东西太重、搬不动。农耕社会的粮食、牲畜、土地，都能攒、能传给下一代。人和人之间的差距，于是第一次可以代代放大。这种差距——有地的和没地的，有仓库的和没仓库的——是犁没有设计、却由犁生出来的。

新的“记账”：数字和文字的需要。有盈余，就有出入；有出入，就要记。许多学者认为，美索不达米亚的早期文字，最初就是为了记账和管理仓库而出现的：泥板上最早的内容，多是配给、数量、物品，而不是诗歌〔待核〕。我们在第四编会回到这件事。这里只提醒一点：文字在这里不是发明，是“碰出来的”。种地的人在想怎样种得好，没有人在想怎么发明文字。

新的几何：量地。田界要年年重量，测量的需要，是几何最早的现实根。这个故事留到第18章去讲。

新的“人口”：养得活，就生得多。粮食多了，人口随之增加；人口多了，需要更多的粮食，于是开更多地，种更多田。这是一个自我加速的圈。

新的“精英”与“国家”：谁来收税、谁来保护、谁来调水。第10章讲过渠。一条大渠的维护，需要成千上万人协调。协调的人逐渐固定下来，变成官员；官员需要被供养，所以需要收税；收税需要登记人口与土地。国家这个东西，不是谁坐下来“发明”的，而是在这些碰撞中，慢慢长出来的。这个说法在学界也有不同版本，国家起源是个大问题，没有单一的定论〔待核〕。

把这些排在一起，你会发现：粮食多了，需要处理的事情就多了。一个“提高产量”的工具，生出了“如何分配、如何记录、如何统治、如何防病、如何测量”一大串问题。而且，每一个问题都不在犁的原始封装里。

◆ 五、第四圈：逼出了什么更高的智慧

这些问题，逼出了一批全新的智慧。

逼出了会计、文字、数字。这是理念界的东西。

逼出了法律与契约。这块地归谁？借了粮食怎么还？你可以说，法律是人类对“差异积累”的一次系统回应。

逼出了天文与历法。第10章已经提到。

逼出了农学本身。知道什么时候下种，知道轮作，知道粪肥。很长时间里，农学是一种经验。到了十八世纪的英国，农业革命带来了轮作新法、选种育种，农业生产率有了明显提高。再到二十世纪，哈伯—博施法实现了工业化固氮（约1909至1913年间），化肥改变了地球上能养活的人口数量；之后的“绿色革命”继续推高产量〔待核〕。

注意，这里出现了一个反复发生的模式：每一次产量跃升，并没有让“粮食问题”消失，而是把问题换了形状——从“不够吃”变成“怎么分”“怎么保”“化肥过量怎么办”“水资源怎么办”“品种单一的风险怎么办”。

◆ 六、让渡与内卷：农业里的两个观察

农耕里有一个后来被反复讲的故事：农业劳动力的让渡。

在十九世纪末二十世纪初的美国，从事农业的人口约占劳动力的四成左右；今天，这个比例不足百分之二〔待核〕。这并不意味着粮食不重要了，恰恰相反：粮食更多了，只是做这件事的人，大量被拖拉机、联合收割机、化肥与育种体系取代了。

这件事该怎么看？如果只盯着“农民少了”，会觉得是一场灾难；可如果看整个链条，会发现：被机器接手的那部分劳动，让出了很多人去做以前根本没人做的事——工厂、城市、教育、医疗，以及大量一百年前没有名字的职业。这就是承重命题5：被封装能力的退场是让渡，不是退化。这里要谨慎：让渡的过程对具体的人和具体的时代，常常是痛苦的。“总体上让出新空间”，不等于“每个人都平稳过渡”。这一点，在第六编里还会专门讲。

人类学家克利福德·格尔茨（Clifford Geertz）研究印度尼西亚爪哇的水稻农业时，用过“农业内卷”这个词：土地有限，人口增加，水稻种得越来越精细，更

多的人投入同一块地，产出略有增加，人均收入却没有明显改善〔待核〕。黄宗智研究中国江南农村时，也用过类似的概念〔待核〕。

这就是工具让渡律里“内卷期”的一种农业版本：工具已经封装完毕，扩大期的红利已用完，人们在同一个框子里越挤越细。跳出内卷的办法，通常不是在框子里更努力，而是新的东西进来——新的作物、新的工具、新的产业，把一部分人带到框子外。

◆ 七、再看一次：为什么"多"总会变成"难"

把冰箱和田野并排放，会看见一个共同的结构。每一次“多”，都先解决了一个旧的“少”，同时把一件原来不必管的事推到台前。

采集者不用管仓库，因为没有东西可存。农民必须管，因为有东西可存。这不是农民比采集者更聪明或更笨，而是他们站在不同的问题面前。后来的会计、书记员、税吏、法官，是被这个新问题请来的人。他们一上场，又带来自己的问题：账记得对不对，税收得公不公，判得服不服人。每个问题都在要求下一种本事。

这里有一个常被忽略的细节：这些新本事，没有一样是犁的发明者能够预订的。你不可能在发明犁的那天，就给“法官”这个职业写好招聘启事，因为那时“这块地归谁”根本还不是一个会吵起来的问题。这就是错位：新的SDE出现在封装之后，而且只能出现在使用之后。

◆ 八、回扣承重命题，以及这个论证的边界在哪里

犁封装了一套“让土地听话”的已发生的理解（命题1）；它与定居、剩余、人群碰撞，碰出了仓库、边界、税收（命题2）；文字、法律、国家，在犁被发明时都不存在，所以也不可能在犁里（命题3）；农业劳动力的大规模退场，是让渡（命题5）。

四点。

第一，农耕起源是史前问题，证据残缺，学界争论很大。斯科特的观点是一种有力的解读，不是定论。把“农业带来了国家和文字”说得太直，就掩盖了许多反例和例外，比如有些没有农业的复杂社会，也出现了很大的聚落和礼仪工程〔待核〕。

第二，我们说“犁碰出了文字”，并不是说文字只能由农业碰出来。因果关系在史前很难一一落实，这里的叙述是“相关且合理”，不是“唯一且必然”。

第三，这里面藏着一个很容易滑过去的东西：农耕的故事看起来像“工具生出问题，问题生出更高的智慧”，似乎一个比一个好。但实际上，很多代人是在苦难里度过这个过程的。“逼出更高的智慧”这句话，不是在为苦难辩护，只是在描述一个结构。

第四，这一章的判断，在下面这种情形里是错的：如果能找到一个盈余大到要存几百年、仓库和田界早已有了，却始终没有出现任何记录、裁决或管事人的定居农耕社会，而且社会结构一直没变，那么“盈余必逼出记录与秩序”就不成立。顺带一提，印加人没有我们熟悉的那种文字，却用结绳记数、记事（结绳到底能记多少内容，学界仍有争论）。这提醒我们：被逼出来的是“记录”这件事，不一定是文字这种形式。

◆ 九、小结

一把犁，让粮食多了。粮食多了，人群就稠了，账就多了，边界就多了，官也就多了。每个问题，都是前一个答案的产物，每个答案，又都把人逼到下一个问题前面。

回头看你家的冰箱：它解决了“今天吃什么”，也同时生出了“我家存了什么”。这不是冰箱的失败。这就是工具的样子。

带走的话

冰箱解决了“今天吃什么”，却把“我家存了什么”推到台前；犁也一样，每一次“多”，都先补上一个旧的“少”，再带来一件原来不必管的事。

小账 封装了什么？——让土地按人的意思出粮食的一套理解，外加借牛马之力。 逼出了什么？——仓库、田界、记账、法律、农学，以及管事的官和国家。

承接：粮食和城市把“同一个时刻”变成了一个大问题。下一章，从钟表讲起——当一个国家的人想在同一时间出现在同一个地方，会发生什么。

第 12 章

时钟、铁路与电报：从时刻表到相对论

◆ 一、一个"七点见"

你和朋友约好晚上七点在餐厅见面。你们各自从不同的地方出发，一个坐地铁，一个打车，七点前后到了。这件事轻松得几乎不值得一提。

可是，请想一想：这件事要成立，需要有多少东西事先对好？你们手上的表要对，餐厅里的表要对，地铁的班次表要对，你手机里的时间要和网络对。如果每个人头上的“七点”各不相同——这在几百年前是常态——你们就要约“太阳到头顶的时候”，然后等到一个人看见另一个人，才算见到。

“大家的七点是同一个七点”，听起来像自然的东西。其实是一整套工具，一点一点造出来的。

钟表封装的是“把时间切成均匀的段”；铁路和电报逼着人们要问一个钟表自己从未问过的问题——什么叫“同时”？

◆ 二、第一圈：钟表封装了什么

机械钟大约在十三、十四世纪出现在欧洲〔待核〕。早期的钟，常常建在修道院和城镇的钟楼里，是为了敲钟报时：修士要按时祈祷，城里的人要按时开市。

钟表封装的，是一种看似简单的想法：时间可以被均匀分割。太阳的影子有长有短，夏天的白天长、冬天的白天短；水钟会受温度影响；沙漏一次只能走一段。机械钟用擒纵机构把一个均匀的节拍“数”出来，让一小时在冬天和夏天都是一小时。

这个想法一旦流行，后来的故事就接着来了。历史学者刘易斯·芒福德（Lewis Mumford）把钟说成是现代工业时代的关键机器，不是蒸汽机，而是钟〔待核〕；

E. P. 汤普森（E. P. Thompson）1967年的那篇有名的文章，讲的是工业革命中“时间纪律”怎样进入工人的生活：从按任务干活，转向按钟点干活〔待核〕。

注意，这里已经有一次碰撞：钟表与工作碰撞，生出了“迟到”。“迟到”是个很有意思的东西。在有均匀钟点之前，没有真正意义上的“迟到”，只有“来得晚”。迟到需要一个大家都认同的标准，才成立。

◆ 三、第二圈：与铁路碰撞

十九世纪上半叶，铁路出现。1830年利物浦至曼彻斯特铁路通车，是常被引用的起点〔待核〕。

铁路给时间带来的第一个麻烦是：每个城镇的“当地时间”不一样。太阳在东边的城市先到最高点，西边的后到。以前，没有人在意：马车慢，几十公里的差别，只是几分钟，没人发现。铁路一快起来，问题就浮出来了。

英国的做法，是铁路公司率先采用伦敦时间，法律上的统一则要到1880年〔待核〕。

北美更夸张。幅员辽阔，十九世纪后半叶，各铁路公司各用各的时间，一个大站可能挂着好几只表，分别指向不同铁路线的时间。乘客想转车，先得做一道算术题。1883年11月18日，北美的主要铁路公司统一采用了标准时区制度（东部、中部、山地、太平洋四个时区）。这一天被人称作“两个正午的日子”，因为在许多城市里，时钟需要被拨回或拨快，一天里正午出现了两次或被跳过〔待核〕。

注意，它不是国家法律推行的，是铁路公司自己定的。后来，国家才追认。1884年的国际子午线会议在华盛顿召开，选定了格林尼治作为本初子午线，全球的时区体系，以此为基础逐步建立〔待核〕。

◆ 四、第三圈：与电报碰撞

铁路的另一个麻烦，是安全。

单线铁路上两辆车对头相撞，是十九世纪初期铁路最可怕的事故形态。要避免它，调度员需要知道“现在每一列车在哪”。可是，信息不能比火车跑得快——直到电报。

塞缪尔·莫尔斯1844年在华盛顿与巴尔的摩之间发出的电报，常被当作实用电报的开端〔待核〕。电报使信息的速度与运输的速度第一次分开了。以前，信息是被人带着走的；现在，信息可以跑在人前面。

铁路公司马上把电报用在了调度上：沿线设站，用电报通知前方，列车何时发出，何时到达，是否需要避让。铁路和电报联手，生出了现代意义上的“时刻表”：一张大家都遵守的、把各地钟点对齐的网。

电报还有另外一个后果：它把“同一时刻的远方”变成了一种可以谈论的东西。远方发生的事，当天甚至当时就知道。新闻、股票行情、跨洲的商业，在电报之后有了新的节奏。

◆ 五、第四圈：生出了什么新的SDE

这是这一章最想让你看到的部分。

新的职业与新的制度。铁路调度员、电报员、时刻表编制者、天文台授时服务。天文台从前主要观察星星，后来多了一项新业务：把准确的时间卖给城市、铁路和港口。（十九世纪许多天文台确实提供过时间信号，具体国别与年份从略。）

新的社会节奏。早上几点上班，几点下班，几点开会，火车几点开——整个社会开始按同一张时刻表呼吸。

新的问题——“同时”到底是什么？这是最深的一个。电报让不同地方的钟需要对准；对准钟，就要把一个信号从甲地发到乙地，扣除传输时间。可是，信号是需要时间的，哪怕很短。于是，“甲地的七点整”和“乙地的七点整”是不是同一个时刻，要依赖一个约定：你准备怎样扣掉信号的时间？

科学史学者彼得·加里森（Peter Galison）在《爱因斯坦的钟，庞加莱的地图》（2003）里做了一个有分量的论述：爱因斯坦在伯尔尼专利局工作的时期，

经手大量关于电气授时、钟表同步的专利申请；庞加莱则在法国经度局，负责处理远距离时钟校准的问题。1905年爱因斯坦谈“同时性”的论文，用的是“用光信号对钟”这样的操作图景。加里森的看法是：这不是巧合，抽象的物理问题与具体的技术问题，在同一个时代互相映照〔待核〕。

这里要特别小心。加里森说的是“技术背景与科学思想深刻地互相渗透”，不是“铁路和电报直接导致了相对论”。爱因斯坦的理论有很多来源：麦克斯韦电磁学、以太实验、对绝对时间的哲学怀疑……我们只取温和的结论：在那个时代，“同步时钟”从一个少数人关心的天文问题，变成了遍布铁路、电报、港口的实际问题，“同时是怎么定的”这个很哲学的问题，因此有了具体的落点。

更远的回声：全球卫星定位。今天你手机里的导航，依靠的是卫星上的原子钟。按相对论，卫星钟因为速度和所处引力不同，每天比地面钟快大约三十八微秒；如果不修正，定位误差每天会累积十公里上下。一个关于“同时”的理论，一百年后成了日常导航里不可缺的修正项。

◆ 六、第五圈：更高的智慧

铁路、电报和钟，一共逼出了三样更高的东西。

一、标准。时区、轨距、度量衡、时间的国际约定。标准这个东西，不属于任何一件具体的工具，却是所有工具能连起来的前提。它是一种制度性智慧。

二、对“时间”的更精细的理论。从机械钟的摆，到石英钟，到原子钟。1967年，国际上把“秒”定义为铯原子某一跃迁辐射的固定周期数〔待核〕。秒不再是“一天的八万六千四百分之一”，而是原子的一个性质。时间的定义，本身在升级。

三、对“同时”的物理反思。也就是相对论。

最后这三样，没有一样在最初的机械钟里。钟表封装了“均匀分割”；它被拉进铁路和电报的网，才逼出了标准与同时性问题。这就是第二编讲过的：**工具封装的是昨天的智慧，逼出来的是明天的智慧**。

◆ 七、一个小实验：没有同一个七点

假设你住在一个没有统一时间的世界。每座城市以自己的正午为十二点。你坐火车从甲城去乙城，路上要经过五座城市，每座城市的车站钟，都比前一座晚几分钟，或早几分钟。你手里的车票写着“十点整发车”，可是，是谁的十点?

你会马上明白：这不是一个技术难题，而是一个谁来说了算的难题。统一的时间，要有人定，要有人信，要有人罚迟到者。铁路公司定了，乘客认了，国家后来追认。这一连串动作，是制度的，不是物理的。

所以才会出现看起来很怪的事：一个最像“自然规律”的东西——时间——的统一，其实是一场人和人之间的约定。约定做成了，大家就忘了它是约定，把它当成了天经地义。这也是工具封装的一个典型表现：封装得越成功，里面的发生过程就越不可见。我们今天用手机上的时间，几乎不会想到其背后曾经有过吵吵嚷嚷的会议和公司之间的博弈。

这正好回应第5章：封装让使用的人不必再重走一遍，代价是封装越严密，使用者越看不见里面，也越难发现它哪天不再够用。时间的统一在日常里够用了，可到了毫秒、纳秒都要同步的地方（金融交易、通信网络、卫星定位），“同时”又变回需要专家花力气去解决的问题。封装让“同时”消失在日常里，却没有让它消失在世界上。

◆ 八、回扣承重命题，以及这个论证的边界在哪里

这一章的回扣，命题3最清楚。机械钟出现的时候，“跨城市的同步”根本不是问题，是铁路把它造了出来，所以不可能在钟的封装里。其余几条一笔带过：钟与铁路、电报碰撞，碰出时刻表、时区、授时（命题2）；钟这件现实界的工具，进了理念界的物理时间观，也进了自我界的准时、迟到、守信与时间焦虑（命题4）。

三点。

第一，科学史上“某项技术催生了某项理论”这种因果判断，要格外慎重。加里森的论述有影响，也有批评者认为其对相对论的解释过于偏向技术背景〔待核〕。本章只取最稳的版本：背景相关、问题同源，不取“因此必然”。

第二，标准化也有代价。统一时间，同时意味着抹去地方的节奏；统一轨距、统一规格，也可能让某些做法被边缘化。“更高的智慧”不等于“更好的生活”在所有人那里都成立。这一点我们在第六编会再谈。

第三，这一章的判断也有明确的落脚点。如果有人证明，把铁路和电报整个拿掉，各地钟点的统一和“同时怎样定义”的追问，也会在差不多的时间、以差不多的方式出现，那么“铁路造出了这个问题”就要改成“它只是催了一把”。反过来，如果科学史的研究表明，爱因斯坦和庞加莱对同步的思考与专利局、经度局的日常工作几乎无关，本章后半的那条线就应当删去。

◆ 九、小结

一个“七点见”背后，是钟、铁路、电报和国际会议的合力。更有意思的是，这些东西合力的结果，不仅是让人准时，更是逼着人去问：同时是什么？

钟自己没想过这个问题。是铁路把它造了出来。

> 带走的话
>
> 约“七点见”之所以轻松，是因为曾经有人吵吵嚷嚷地把大家的钟对齐了；封装得越成功，里面的发生过程越看不见。钟自己从没问过“同时是什么”，是铁路把这个问题造了出来。

> **小账** 封装了什么？——把时间切成均匀的段。 逼出了什么？——时刻表和时区、由原子定义的秒，以及对“同时”的物理追问。

承接：时间被切整齐了，动力的问题才真正暴露。下一章，讲蒸汽机——一个工匠先做出了机器，科学家后来才明白它为什么能工作。

第 13 章

蒸汽机：工具先于科学，热力学是被逼出来的

◆ 一、一个会跳的锅盖

烧水壶的水开了，盖子被顶得“哒哒”直跳。你按住它，它顶你的手；你拿开手，它继续跳。

这是蒸汽的力量，一个每个人都见过的现象。可是请注意：从“看见盖子在跳”到“造一台机器”，中间隔着很长一段路；从“造出机器”到“弄明白机器为什么能干活”，又隔着一段更长的路。

这一章，我们看第二段路。因为它里面藏着本书最喜欢的一个场景：先有工具，后有科学。

蒸汽机封装的是“用热做功”的工匠智慧；它在大规模使用中暴露了一个它自己回答不了的问题：热到底能变出多少功？这个问题，逼出了热力学。

◆ 二、第一圈：它从哪里来

十七世纪末到十八世纪初，英国的矿井越挖越深，水也越渗越多。抽水成了矿主最头疼的事。前面第10章提到过：森林被砍、木柴稀缺，英国人转向了煤；煤要挖，挖得深就要抽水。

1698年，托马斯·萨弗里（Thomas Savery）申请了一种用蒸汽抽水的装置的专利。1712年，托马斯·纽科门（Thomas Newcomen）造出了更实用的大气式蒸汽机，用于矿井排水〔待核〕。纽科门机的原理是：让蒸汽进入汽缸，再喷冷水使之凝结，汽缸里形成真空，外面的大气压把活塞压下去。它很笨重，也很费煤，但在煤矿旁边，煤并不贵，所以它成立。

詹姆斯·瓦特（James Watt）在1760年代注意到：纽科门机的汽缸每次都要先加热再冷却，大量热白白浪费。他的办法是把冷凝放到另一个独立的容器里，让汽

缸始终保持热，1769年为此取得专利〔待核〕。后来他与马修·博尔顿合作制造和销售蒸汽机。

这是一个典型的工具封装：一代一代的工匠，把“让蒸汽做功”这件事一点一点封装得更好。

◆ 三、第二圈：碰撞

蒸汽机和谁碰撞？这一问，答案是很多。

蒸汽机很快碰上了一大批对象：矿井越挖越深，煤产量上升；工厂不必再挤在河边，可以建在煤和劳动力集中的地方；蒸汽机装上轮子，成了第12章铁路的动力；新的工厂城市、雇佣劳动和阶级关系也跟着出来。新东西多到写不完，那是经济史的任务，这一章只盯一个问题。

◆ 四、第三圈：一个让工程师着急的问题

蒸汽机在欧洲大陆和英国大规模使用后，很快出现了一个很实际的问题：效率。

燃料很贵（特别是在煤不便宜的地方），每烧一吨煤，能做多少功？怎样的机器设计才算好？改变汽缸的大小、蒸汽的压力、温度，会怎样？工程师们靠经验和试错在改进，却没有一个原则可以告诉他们：有没有一个极限？

1824年，一位年轻的法国工程师萨迪·卡诺（Sadi Carnot）发表了《关于火的动力的思考》（Réflexions sur la puissance motrice du feu）。卡诺的思路很特别：他不去研究具体的某台机器，而是问一个更一般的问题——热机做功的能力，取决于什么？他得出的结论是：取决于热源与冷源之间的温度差，而不取决于工作的物质是什么。热机要想做功，就必须有从热到冷的“落差”，就像水车要靠水从高处落到低处〔待核〕。

这个结论，既深刻又令人意外：它说，无论你的机器造得多精巧，都存在一个由温差决定的上限。

卡诺的论文在当时几乎无人问津。后来，焦耳等人确立了能量守恒；克劳修斯和开尔文（William Thomson）把卡诺的思想与能量守恒结合，建立了热力学第一和第二定律；克劳修斯在1865年左右提出了“熵”的概念〔待核〕。

这条线非常清楚：蒸汽机已经跑了一百多年，热力学才完整。

后来流传一句话，大意是科学欠蒸汽机的，比蒸汽机欠科学的多。它常被归到生理学家兼化学家L. J. 亨德森（Lawrence J. Henderson）名下，但能查到的最早出处，是1960年一本书里一句没有注明来源的转述，所以这里只取大意，不当作引文。

这个故事科学史家早就讲过：卡德韦尔的《从瓦特到克劳修斯》，专讲热力学怎样从蒸汽机的实际问题里长出来；克莱因和罗森堡则提出，科学与技术之间不是单行道，而是来回的链条〔待核〕。本书与它们相同：承认工具常常走在科学前面。不同在于，我们追问的不只是“谁先谁后”，而是“这个问题在封装的时候，为什么不可能被提出来”。

◆ 五、第四圈：新的SDE

蒸汽机生出的新东西，除了热力学，还有几样。

锅炉爆炸与安全。早期高压蒸汽机的锅炉事故很多，伤亡惨重。美国机械工程师学会（ASME）1880年成立，那一年美国有一百多起锅炉爆炸；它的第一版锅炉规范到1914年才出版，是对一连串事故的回应。你看，“安全”这个新东西，在蒸汽机最初的封装里完全没有，是一次次事故之后被碰出来的。

“节约”的悖论。威廉·斯坦利·杰文斯（William Stanley Jevons）在1865年出版的《煤的问题》里指出：蒸汽机效率提高之后，煤的消耗没有下降，反而上升，因为煤的用途变多了〔待核〕。这就是后来被叫作“杰文斯悖论”的现象。这是封装的另一个后果：效率越高，使用越广，总消耗反而更大。这对第9章讲的“扩大期”是一个绝好的例证。

一个漫长的“等待”。经济史学者罗伯特·艾伦（Robert Allen）把英国工业革命早期（约1790到1840年）的一个现象叫作“恩格斯停顿”：产出增长很快，工人的实际工资却涨得很慢，好处主要落到了资本那一边。这个判断也有争议，有经济学家认为当时的工资增长被低估了。无论如何，这是一段“封装已完成、新SDE还没显露”的时期，我们在第14章把“等待期”讲透。

◆ 六、第五圈：更高的智慧

热力学。它不是蒸汽机的说明书。说明书回答“这台机器怎么用”，热力学回答的是“任何热机最多能做到多少”，而这个问题，只有成千上万台机器摆在一起、有人开始比较谁更省煤的时候，才问得出来。它是被大规模使用问出来的，不是被一台机器解释出来的。问出来之后，答案是一套关于能量与方向的原则：能量守恒；热不会自发地从冷流向热；孤立系统的熵不会减少。它后来延伸到化学、生物（细胞的能量代谢）、宇宙学（热寂的讨论），也延伸到信息论（熵这个词借用到了信息上）。

一套原来为了让蒸汽机好用而想出来的原理，最终成了跨越几乎所有自然科学的基础。

“科学和技术的关系”本身。这是一个更微妙的升级。十九世纪以后，人们开始有意识地做一件事：不等技术先出来再解释，而是主动用科学去指导技术。电力工业、化工、半导体，常常是科学先行的。这种“科学—技术—产业”的体系，是在蒸汽机的时代前后才逐渐成形的〔待核〕。可以说，蒸汽机逼出了人类对“如何系统地造工具”的新认识。

◆ 七、工匠与科学家，谁先到

蒸汽机的故事里，常常出现一个争论：到底是工匠重要，还是科学家重要？

做过饭的人大概能直觉地回答：两者都重要，而且顺序不固定。你做一道新菜，往往是先瞎试，试出一个能吃的版本；然后有一天，你读到一篇讲美拉德反应的文章，才明白为什么煎到焦黄就更香，于是你知道了下一次怎么控制火候。这

时，科学没有“发明”那道菜，却让你下一次做得更有把握，也让你敢于做更复杂的菜。

蒸汽机和热力学，是同样的关系，只是尺度放大了一万倍。工匠先试出来一个能用的机器；科学家后来明白了它为什么能用、最多能用到什么程度；明白之后，工程师又造出更好的机器——内燃机、蒸汽轮机、制冷机，都建立在热力学的原理之上。

这是一个“螺旋”：工具→问题→理论→更好的工具→更深的问题。每转一圈，工具封装的智慧就更深一层，同时暴露出来的未解问题也更深一层。

我想特别提醒一点：卡诺的结论，是一个限制——它告诉人们，什么是做不到的。这个“做不到”本身，是一种极有价值的智慧。如果没有第二定律，人们可能还在设计永动机。一个“不可能”，可以是整个领域最重要的发现。这一点，对读这本书的你尤其值得留意：本书的结论也是一个“不可能”，而热力学的历史告诉我们，**一个好的“不可能”，不是在关门，而是在指路**。

◆ 八、回扣承重命题，以及这个论证的边界在哪里

蒸汽机封装的是工匠们对“热如何做功”的经验理解（命题1）；它与矿井、工厂、铁路碰撞，生出大量新的情形（命题2）。命题3在这里有一个格外干净的样本：热力学第二定律，在蒸汽机被发明时没有任何人知道，是机器大规模使用之后，因为“效率有没有极限”才被逼了出来。这件现实界的工具，逼出了理念界的热力学，也在自我界造出了新的阶级与“工人”这种新身份（命题4）。

命题9：基础性本身在发生。热力学是基础科学，它不是在机器发明之前就躺在那里的，是被问出来的。这预告了第九编的一个主题：AI会不会也在制造它自己的基础问题？

四点。

第一，“工具先于科学”并不是一条普遍规律。电磁学是先有科学后有电报、电机的；半导体是先有量子力学后有晶体管的。更准确的说法是：工具与科学互相

推着走，有时工具在前，有时科学在前。蒸汽机只是“工具在前”里最显眼的例子。

第二，“是被逼出来的”这个说法，容易让人觉得好像科学家是被动的。实际上，卡诺、焦耳、克劳修斯等人有各自深刻的个人洞察，这些洞察不是任何“逼迫”自动生成的。逼，是创造了条件；条件下面谁来想出来、怎么想出来，是另一件事。第七编会专门讲。

第三，具体年份与人物关系，史书上有不同的说法，凡带〔待核〕的地方，请以专门的科学史著作为准。

第四，这一章的判断，在下面这种情形里是错的：如果能指出，热机效率有上限这一核心结论，在蒸汽机普及之前，就已经因为与机器无关的原因被独立提出并被接受，那么“被大规模使用逼出来”就站不住。

◆ 九、小结

那只会跳的锅盖，让人造了机器；机器多了，人们才问：这股劲头，到底有没有上限？问出来的答案，叫热力学。

先有工具，后有科学。这个顺序，在人类历史里不是例外，是常见的形状之一。

> 带走的话
>
> **先有会跳的锅盖和笨重的机器，后有热力学：科学不是在机器前面等着，是机器大规模用开之后才被问出来的。卡诺那个“做不到”，不是关门，是指路。**

> **小账** 封装了什么？——工匠们对“热怎样做功”的经验。 逼出了什么？——锅炉安全规范、杰文斯悖论，以及“效率有没有上限”这个问题和它的答案。

承接：下一章，动力有了，却迟迟没有改变工厂的样子。电力出现之后，人们等了几十年。

第 14 章

电力与洗衣机：几十年的等待，以及"省下来的时间去哪了"

◆ 一、洗衣机的承诺

我外婆那一代人，洗衣服是一整天的事：烧水、搓、拧、漂、晾，还要看天色，怕下雨。后来家里有了洗衣机，把衣服扔进去，按个键，人就可以去干别的了。

这是几乎每一个有洗衣机的家庭都有的体验。可是，历史学家露丝·施瓦茨·考恩（Ruth Schwartz Cowan）在1983年出版的《母亲的活更多了》（More Work for Mother）里提出了一个令人意外的论点：二十世纪家用电器的普及，并没有显著减少主妇花在家务上的时间〔待核〕。

这是怎么回事?

这一章要同时讲两件事：一件是等待——电力在工厂里为什么花了几十年才发挥威力；另一件是时间去哪了——省下来的劳动，为什么没有变成闲暇。两件事其实是同一个形状的两面。

电力和洗衣机封装的是“把能量送到该去的地方”“把搓洗这件事机械化”；而它们的真正成果，要等到人们重新安排工厂、家务和生活之后，才显现出来——这一等，就是几十年。

◆ 二、第一圈：电力封装了什么

十九世纪后半叶，电力技术逐渐成熟。1882年，爱迪生在纽约珍珠街建成了早期的商业供电系统；后来，交流电的发电与输电技术（特斯拉、威斯汀豪斯等人的工作）使远距离输电成为可能〔待核〕。

电力封装的东西，是一整套关于如何把能量变成便于传送、便于分配、便于控制的形式的理解。蒸汽机把能量集中在一个地方；电力把能量“拉长”，让它能沿着一根线，走到任何一个需要的地方。

◆ 三、第二圈：电动机进入工厂——却没有立刻改变什么

这是整个故事最耐人寻味的一段。

在蒸汽时代，一座工厂通常有一台大型蒸汽机，它的动力通过一根长长的主轴、皮带和传动轮，分配给车间里的各台机器。这种布局叫做“集中传动”（line shaft，成组驱动）。因为动力来自一个中心，机器必须围绕这根轴来摆放，往往堆在多层厂房里，一层层往上，用皮带把动力送到各层。

电动机出现后，工厂最早的做法，是把蒸汽机换成大电动机，其他都不动：主轴、皮带、车间布局照旧。电力就这样“替换”了蒸汽，却没有带来想象中的生产率飞跃。

经济史学者保罗·戴维（Paul A. David）在1990年发表了一篇很有影响的文章《发电机与计算机》（The Dynamo and the Computer），指出：美国制造业的生产率增长，要等到二十世纪二十年代才明显加速，而电力系统在1880年代已经开始商业运行，距离此时已有将近四十年〔待核〕。

为什么这么久？戴维的解释是：电力的真正价值，不在于“代替蒸汽机”，而在于让每台机器都有自己的电动机（单元驱动）。这样，机器不必围着主轴排队，工厂可以改成单层的、按生产流程摆放的布局，物料可以直线流动。亨利·福特在1913年前后的高地公园工厂（Highland Park），建立了移动装配线，是这类重新布局最著名的例子之一〔待核〕。

可是，要做到这一点，需要的不只是买电动机，而是重新设计整座工厂，包括厂房的结构、工人的分工、管理的方式，甚至培养懂这套新方式的工程师和经理。这些事，都不是一夜之间就能完成的。

◆ 四、第三圈：什么叫"等待期"

这个现象，就是本书反复提到的等待期：

- 封装已经完成：电已经有了，电动机已经能买到了。
- 新的SDE尚未显露：与电力相匹配的工厂布局、工作组织、管理方法，还没有被发明出来。

这里要先纠正一个最顺口的说法：等待期是技术还不够好。不是。电动机在十九世纪八九十年代已经能用，买得到，也用得起。缺的是另外两样东西：能和它咬合的厂房布局与分工，也就是工厂这块土壤；以及有人先看见一种还不存在的工厂。这像是你买回一台顶级烤箱，却仍旧照着土灶的做法烤面包：设备没有落后，是你的做法还没有长到与它配套。本书把这叫作维度未配平：封装这一头已经到位，已经显露为稳定的样子（S）；而新做法一步步试出来的过程（D），和工厂、人、管理方式纠缠成的新土壤（E），还没有跟上。

在这段时间里，旁观者会有两种相反的感觉：一种是失望，这东西好像也没什么大不了；另一种是错觉，这东西已经全面到来，一切都已解决。两种感觉其实是同一件事的两面：没有看到新的SDE，就把它当作不存在，或者当作已经包办。第十编里，我们要用同样的眼光去看今天。

◆ 五、第四圈：洗衣机——省下来的时间去哪了

回到家里。

考恩的研究发现了几个让人意外的事实。

一、标准提高了。有了洗衣机，人们不再是每周洗一次衣服，而是每天甚至每次穿完都洗。床单从每月洗一次变成每周洗。孩子的衣服、丈夫的衬衫，洗得越来越勤。“干净”的标准，被工具自己悄悄抬高了。

二、分工变了。以前，家务是有许多人分担的：仆人、年长的孩子、邻居、亲戚。随着家庭规模缩小、家政服务减少，更多工作落到了主妇一个人身上。

三、新任务出现了。汽车普及以后，家里的人要开车接送孩子、采购物品；这些“新家务”，在没有汽车的时候根本不存在。

四、新的观念出现了。细菌学的普及，使人们更在意卫生；科学的育儿观念，使母亲对孩子投入更多时间。

把这些放在一起，你就明白了：洗衣机并没有让“洗衣服”这件事消失，而是把洗衣服变成了一个更高标准的任务。而且，与洗衣机配套，还冒出了很多周边东西：洗衣粉的工业、合成纤维、洗衣标签、烘干机、家用供水与排水系统、电网的家用负荷管理……每一样，都是洗衣机被使用之后才生出来的。

另一项早期研究，是琼·瓦内克（Joann Vanek）在1974年的发现：二十世纪二十年代到六十年代，美国非就业家庭主妇每周的家务时间，大致没有减少〔待核〕。

◆ 六、第五圈：更高的智慧

电力与家用电器逼出了什么？

电网与电力工程。如何稳定供电，如何在峰谷之间调度，如何避免大面积停电——这是一整套复杂的系统工程。

标准与安全。插座、电压、接地、保险丝、电器认证。你家墙上那个插座背后，是一系列的标准协商。

家政学与消费科学。一门学科，专门研究怎么把家务做得更有效。

对“时间使用”的研究。考恩、瓦内克这些学者所做的，正是后来发展起来的时间使用调查。通过日记式的调查，社会学家第一次能够系统地回答“人们的时间究竟去哪了”。你看，连“研究家务时间”这件事本身，也是新生出来的。

◆ 七、如果你生在1900年

设想你是1900年前后一家工厂的老板。

你听说电动机是好东西，于是你把那台旧蒸汽机拆了，换了一台大电动机，皮带和主轴照旧。账本上，电费比煤费略便宜，仅此而已。你的邻居也差不多。你们几个老板喝茶时可能会议论：电这个东西，名气大，实际上也就那样。

另一个老板，看了一圈之后决定：我要把整座厂拆了重盖，每台机器配一台小电机，厂房改成平房，让原料从一头进、成品从另一头出。他得花一大笔钱，还要重新训练工人，还可能第一年亏本。你会怎么选？

多数人会选第一种。这不是愚蠢，而是理性：第二种选择，需要的不是钱，而是想象，是你要先看见一种还不存在的工厂。而这种想象，要在许多人的试错和失败之后，才会变成大家都能抄的做法。

这就是为什么等待期那么长。**封装是现成的，可以买；新的SDE是长出来的，不能买**。买一台电动机，用不着想象；可是用好电动机，必须想象。而想象，只能在使用中发生。

◆ 八、回扣承重命题，以及这个论证的边界在哪里

电力封装了能量传输与分配的理解，洗衣机封装了机械化搓洗（命题1）；它们与工厂布局、家庭分工、卫生观念碰撞（命题2）；新的工厂形态、新的卫生标准、新的家务观念，在它们发明时都不存在（命题3）。搓衣板和手搓洗衣的退场，是让渡：被让出的不是“家务”，而是家务中的一类体力活，人们转去做了更多、不同的事（命题5）。封装完成与新SDE显露之间的时间差，就是等待期的雏形（命题10）。

要老实说，这一章有几处容易被误读，要特别注意。

第一，考恩的结论针对的是美国的特定时期（约1900年到1960年代前后），主要关注的是“主妇家务时间”。后来的时间使用研究表明，从二十世纪六十年代以后，发达国家女性的家务时间总体上是下降的，男性有所上升，只是下降幅度不如想象的大〔待核〕。所以，不要把考恩的结论误解为“家用电器没有用”。

第二，“时间没少”不等于“生活没改善”。穿得更干净、孩子更健康，这些都是真实的收益。不少家庭都有过这样的故事：洗衣机让母亲腾出了时间，去读书、去做点别的，孩子也因此多了机会。这类故事说明，节省出的时间往往不是平均地消失，而是在不同家庭里去向不同。我们的论证并不是“工具没用”，而是“工具的收益不在原来的预期位置，也不会自动到来”。

第三，戴维的“滞后”论在经济史学界有人赞同，也有人质疑，对具体的时间点和测量方法有不同意见〔待核〕。我们采用的是它最稳的部分：新的通用技术和与之配套的组织变化之间，存在可观的时间差。

第四，这一章自己的判断可以被检验。如果能找到一项新的通用技术，企业不改组织、不改流程、不培训，只是换上，就在几年内拿到明显的生产率收益，那么“配套重组需要几十年”就是错的；反过来，如果观察到企业重组之后收益依旧迟迟不来，也说明“等待”另有原因。已有的研究把信息技术投入与组织变革放在一起比较，方向上支持前者要紧，细节仍有争论〔待核〕。

第五，这个想法并不新。通用技术要等配套创新、技术革命有“导入”和“展开”两段，经济学里早有布雷斯纳汉与特拉伊滕伯格、佩蕾丝这样的说法，近年又有布林约尔松等人的“生产率J曲线”。本书与它们相同：都说收益滞后于技术。不同在于，我们说滞后不只是“配套投资要花时间”，而是配套的东西在使用之前根本还不存在，没法提前买。

◆ 九、为什么这对后面的章节很重要，以及小结

这一章是整本书里的伏笔。1987年索洛感叹计算机到处看得见，唯独统计里看不见；三年后戴维拿电力的历史回答了他，第15章细讲。到了第十编，我们要问：今天的AGI感，是不是又一段等待期里的错觉？此刻不下结论，只记住一句：电力用了几十年才在工厂里“显形”，显形之前，旁观者既可能说“没什么用”，也可能说“已经无所不能”，两种说法都不太准。

电进了工厂，工厂却没有马上变。洗衣机进了家，时间却没有马上变多。等到工厂重新布局、家庭重新分工，电和洗衣机真正的作用，才露出来；而且，露出来的东西，已经和当初想的不是一回事。

省下来的时间，去了新的标准，新的任务，新的产业。这一点，和粮食多了的故事，是同一个形状。

> **带走的话**
>
> 电早就买得到，买不到的是一种还不存在的工厂，所以工厂等了将近四十年；洗衣机省下的时间，也被抬高的标准和新添的任务接走了。等待期里，旁观者既会说“没什么用”，也会说“无所不能”，两种说法都不太准。

> **小账** 封装了什么？——把能量送到该去的地方，把搓洗机械化。 逼出了什么？——一单机一电动机的新工厂、电网与标准、更高的“干净”标准，连“时间使用”的研究也是新生的。

承接：最后一章，我们来到离今天最近的一件现实界工具——计算机，以及它所生出的软件危机和网络安全。

第 15 章

计算机与网络：程序员、软件危机与网络安全

◆ 一、一次"更新"

你的手机弹出一个提示：有新的系统更新，请重启。你可能点了“稍后提醒”。

这个小小的日常动作背后，藏着一个惊人的事实：几乎所有的软件，在出厂之后，都还要不断打补丁。也就是说，写软件的人，从来没有能一次性做出一个“完成品”。

这件事很奇怪。你买一把锤子，不需要它每周更新一次。但软件需要。这不是软件做得不好，而是软件的某种本性。这一章，我们要从计算机的出现，一路讲到今天的软件和网络，看它们生出来的新东西。

计算机封装的是“按规则机械地算”；它被大量使用之后，生出了“怎样把规则写对”“怎样让很多机器安全地连在一起”这样的新问题。而这些问题，比算本身难得多。

◆ 二、第一、二圈：计算机封装了什么，又与人类的需求碰撞

“计算机”这个词，原本指的是一种职业：做计算的人。十九世纪、二十世纪初，天文台、军队和工程部门都雇佣大量“计算员”（computer），大部分是手工，也有的使用机械计算器。后来，电子计算机出现，这个词的意思就换了。这个小细节，本身就是一次让渡：一个职业的名字，被机器接管了。

通常认为，第一台通用电子数字计算机之一是ENIAC，于1945年底至1946年初完成并公开〔待核〕。约翰·冯·诺伊曼等人1945年的报告，提出了“存储程序”的架构思想，也就是让程序和数据存放在同一个存储器中〔待核〕。

计算机封装了什么？它封装的是一种让规则自动运转的能力：只要你把规则写清楚，它就能不知疲倦地、快速地、准确地照着做。这是图灵、冯·诺伊曼等人的理论成果，与电子工程的结合。

早期的计算机，只用来做两类事：科学计算（比如弹道表）和数据处理（比如人口普查和工资单）。它的使用者是一小群专家。

但是，随着计算机变得更快、更便宜，更多人想用它做更多事：银行账务、航班订票、工业控制、办公文书、科学模拟、通信……

每一个新用途，都意味着有人要把“人怎么做这件事”翻译成“机器怎么做这件事”。这个翻译的工作，就是编程。

于是，一个全新的职业出现了：程序员。

1940年代中期之前，这个职业几乎不存在。如今全世界以软件为业的人，数以千万计〔待核〕。这是一个典型的“新产业诞生”。

◆ 三、第三圈：软件危机

计算机的硬件越来越强，价格越来越低，但软件的开发却越来越难。

1968年10月，北约科学委员会在德国南部的加尔米施（Garmisch）举办了一次会议，主题是“软件工程”。会议报告里，以及与会者的发言中，频繁地使用了“软件危机”这个词，用来描述大型软件项目一再延期、超支、质量不佳的状况〔待核〕。“软件工程”这个名称，也是借这次会议被广泛传开的。

为什么会有危机?

先想想你做饭。一个人给自己煮一碗面，不需要菜谱，不需要分工，不需要写下来，失败了大不了重来。可是，如果要开一家有二十个厨师的餐厅，每天供应几百份菜，那就必须有菜谱、有流程、有分工、有质量检查、有应急预案。

写一个小程序，像煮一碗面。写一个几百万行代码的大型系统，像开一家大餐厅。后者需要的，不仅是“会写代码”，而是如何让很多人写出的代码拼在一起不出错、如何修改一部分而不弄坏其他部分、如何知道它对不对。

弗雷德里克·布鲁克斯（Frederick Brooks）在管理IBM的大型系统软件项目后，于1975年写了《人月神话》，其中一个著名的观察是：给一个已经延期的软件项目加人，往往会让它更晚〔待核〕。这是对“软件开发不能按人数线性扩大”的一个洞察。

◆ 四、第四圈：新的SDE

软件危机逼出来的东西，比软件本身多得多。

结构化编程。戴克斯特拉1968年那封后来很出名的信，批评随意使用“跳转”语句，主张程序应当有清晰的结构〔待核〕。此后，高级语言、模块化、面向对象……都是为了应付复杂性而发明的。

软件工程的整套方法。需求分析、设计评审、测试、版本管理、持续集成。你在手机上看到的那个“更新”，背后是整个这套体系在运转。

软件的“失败案例”也成了教材。1996年6月，欧洲阿丽亚娜5型火箭首飞，约三十七秒后因一处数值转换溢出而偏离自毁〔待核〕；1980年代的Therac-25放射治疗机，也因软件缺陷造成多起严重过量照射事故〔待核〕。这些事故之后，安全攸关软件的开发标准、形式化验证、独立审查，成为专门的领域。

注意，每一个事故，都是一次“封装之外的东西”的显形。写代码的人，当时并不知道某种数值会溢出，或某两个操作的先后顺序会出事。

◆ 五、第五圈：网络与安全

计算机和通信碰撞，就是网络。

1969年，ARPANET的第一批节点联网；1983年1月1日，它切换到TCP/IP协议，常被看作互联网的“诞生日”之一〔待核〕。1989年起，伯纳斯-李在欧洲核子研究中心提出并实现了万维网〔待核〕。

网络把成千上万台计算机连在一起，带来了前所未有的好处，也带来了一类前所未有的问题：安全。

1988年11月，康奈尔大学研究生罗伯特·莫里斯编写的一个程序（后称“莫里斯蠕虫”）在互联网上传播，使联网计算机大面积瘫痪。常被引用的数字是约六千台，约占当时联网机器的一成，但这个数字本身就是粗估，也有人认为实际要少。事后，美国国防部高级研究计划局出资，在卡内基梅隆大学设立了计算机应急响应小组（CERT/CC）。

网络安全这个领域，从此成为一个独立的专业，有自己的职业、标准、课程、公司和国际合作。后来，更多的新问题接踵而至：垃圾邮件、钓鱼、勒索、隐私、数据泄露……每一个问题，都是在网络被广泛使用之后，才显形的。

我们可以把这个故事压缩成一句话：网络的设计者，设计的是连通，不是防御。因为设计的时候，没有预料到会有那么多人以那么多方式去滥用连通。这正是错位定理：新的威胁，在封装时还不存在。

而且，请注意，防御与攻击是互相升级的。每一种新的防御出现，都会引出新的攻击方式；每一种新的攻击出现，都会逼出新的防御。这条升级链没有终点。这是一个“没有终局”的领域，是本书论点的一个小样本。（它是一个很好的小样本，但只是一个小样本。）

◆ 六、第六、七圈：索洛悖论与戴维的回答，以及更高的智慧

1987年，索洛在一篇书评里写下后来被反复引用的话，大意是：计算机的时代到处都看得见，唯独在生产率的统计里看不见〔待核〕。三年后，戴维在1990年的文章里用电力的历史回答了他（第14章已经讲过）：这不是第一次。计算机的广泛

使用，同样需要流程重新设计、技能重新培训、企业组织重新安排，在此之前，统计里不会有明显的提升。

后来的数据，部分支持这个看法：美国的生产率增长在1990年代中期之后有一段明显的加快，许多研究把它与信息技术的扩散和企业组织的调整联系在一起〔待核〕。但是，对这一段增长的原因、大小、持续性，经济学家们至今仍有不同的判断，对后来的放缓，也有多种解释〔待核〕。

这又是一个“等待期”的例子。第14章里，我们看到电力用了将近四十年；计算机的时间表，似乎也是几十年的量级。

计算机和网络，一共逼出了这些新的智慧：

计算复杂性理论。人们很快发现：有些问题，机器理论上能算，但算起来需要的时间，会随着规模爆炸般增长。“P与NP”就是这个领域最著名的开放问题，到今天还没有被解决。这部分，我们留到第九编。

密码学与信息安全。从军事机密到每个人的网上支付，加密成为日常。公钥密码学（1970年代）是一次概念性的突破〔待核〕。

对人机关系的新反思。人们开始系统地问：什么样的系统值得信任？谁对系统的错误负责？这些问题，既有技术的一面，也有法律和伦理的一面。

◆ 七、为什么软件"永远在更新"

回到开头那个手机提示。软件为什么不能一次做完？

说得更准一点：软件不是一件做完的东西，而是一份要和世界不断对账的承诺，世界动一点，账就差一点。

一个原因是：软件所处的环境一直在变。手机的硬件换了，网络的协议升了，用户的习惯变了，攻击者的手段新了。软件不是一座静止的房子，更像一块要不断修补的堤岸。

另一个原因更深：软件是“规则”，而规则要面对的世界，比写规则的人想到的要大。你写了一千条规则，世界会给你第一千零一个情况。程序员有一句行话：每个软件都带着未被发现的缺陷。这不是谦虚，是经验。

所以，软件开发这个行业，从诞生起就有一个特点：它永远不能把自己完成。每交付一个版本，就意味着开始下一个版本。这与本书的主张相呼应：一个封装（软件）投入使用，就会遇到封装之外的情形，于是再封装，再遇到。这个循环不是软件的缺陷，是它的存在方式。

当然，这个类比有限度。软件的“永远更新”，有一部分来自人写错了代码，这类错误原则上能靠更好的工具大幅减少。布鲁克斯1986年就把软件的困难分成“偶然的”和“本质的”〔待核〕，莱曼在1974年也提出过“持续变化”定律：活在真实世界里的软件，不跟着世界改，就会越来越不合用。本书和他们说的是同一件事，只多追了一步：真正指向“没有终局”的，是环境变化引来的新要求，和使用中长出的新用途，只要有人在用，它们就不断出现。

反过来，也有写完就不再改的软件：洗衣机里的控制程序，老式电子表里的小程序，十年不更新照样好用，因为它们的世界是封闭的，只有那几个按钮、那几种水位。它们不是反例，而是边界：封闭的世界里，封装可以真的完成。这个区分，在第八编讲象棋时还要用到。

◆ 八、回扣承重命题，以及这个论证的边界在哪里

计算机封装了“按规则自动运算”的已发生理解（命题1）；它与金融、通信、制造、医疗的碰撞，生出了软件，进而生出软件工程和网络安全（命题2）。软件危机与网络威胁，在计算机发明时都不是问题，是封装之外的东西（命题3）。它逼出了理念界的计算复杂性理论和密码学，也在自我界生出“数字身份”“隐私”这样的新问题（命题4）；计算员这个职业的退场，是让渡（命题5）。至于命题9：可计算性的边界，是图灵在电子计算机之前划出的；而P与NP这类问题，是计算机真的被大量使用之后，才成了摆在桌面上的深层问题。

三点要交代。

第一，我们在这里用的是历史归纳。它告诉我们“迄今为止，大型通用工具都遵循这个形状”，但归纳不能证明“永远如此”。这一点，在第4章里已经立过规矩，在第九编，会有一个结构性的论证来补上。

第二，计算机是最接近“通用”的一件工具。它是否会成为一个例外？也就是说，是否会有一天，计算机自己把“定义新问题”的工作也封装完了？这是本书要面对的核心问题。这一章只是指出：直到今天，计算机每前进一步，都生出了更多需要人去解决的新问题，而不是更少。这是事实，但它是一个迄今的事实，不是对未来的保证。

第三，“防御永远追着攻击”也有可检验处：如果出现一个大规模联网的系统，设计者只设计连通，几十年里却没有出现新的故障类别、新的滥用方式，也没有逼出新的专业，那么“连通必然引来防御”就是错的。迄今没有这样的系统。

◆ 九、小结，兼第三编收束

一次手机更新，背后是一个时代的痛苦经验：写规则容易，写对规则难；连起来容易，连得安全难。计算机让我们算得更快，也让我们发现了很多以前不知道需要问的问题。

六件现实界的大工具，我们看完了。

火：从熟食到金属、历法与水利。犁：从粮食到文字、法律与国家。钟与铁路与电报：从时刻表到相对论。蒸汽机：从抽水到热力学。电与洗衣机：从等待到重新安排。计算机与网络：从算到软件工程和安全。

它们有一个共同的形状：封装→碰撞→生出封装之外的新东西→逼出更高的智慧。

而且，每一件工具，都没有“停在现实界”。火走进了故事，犁走进了账本，钟走进了物理，电走进了家务和观念，计算机走进了数学和隐私。这是承重命题4最朴素的底稿。

但是，一个细心的读者会问：你选的，都是现实界的物件。理念界和自我界的工具，是不是也这样？下一编，我们走进理念界——文字、数字、几何、微积分、概率，以及一个特别的问题：理论本身，是不是也是工具？

> **带走的话**
>
> 每次手机弹出“更新”，都是封装撞上了封装之外的世界。计算机让“算”变得便宜，难题就搬了家：规则要写对，机器要连得安全，而且每进一步，要人来答的新问题只多不少。

> **小账**　封装了什么？——让规则自动运转：按规则机械地算。　逼出了什么？——程序员这一行、软件工程、网络安全，以及计算复杂性和密码学。

承接：第四编，理念界工具史。

第 四 编

理念界工具史：理论也是工具

编 序

前三编讲的是一个原理，也讲了一批看得见摸得着的工具：火、犁、时钟、蒸汽机、洗衣机、计算机。这一编换一个地方看。有一类工具既不能握在手里，也不能放进仓库，它们住在人的脑子里和纸面上：文字、数字、几何、微积分、概率。它们属于三界中的理念界。

先用大白话交代一下本书一直在用的那个三字母。SDE 是三个词的缩写：S 是结构，也就是东西显露出来、稳定下来的样子；D 是差异，也就是事情从这里走到那里的过程；E 是纠缠，也就是把人、事、物缠在一起的土壤，意义就长在这块土壤里。后面说“新的 SDE”，意思很朴素：出现了一种以前没有的样子、一条以前没有的路、一块以前没有的土壤。

本编要立的一个判断是：理论本身就是工具。一套记账方法、一条几何证明、一个概率公式，都是前人把一种想事情的办法封装起来，留给后人直接拿去用。它们和锤子、洗衣机服从同一条链：封装，使用，碰撞，生出封装之外的新东西，再逼出更高一级的智慧。还有一点会反复出现：理念界的工具从来不安分，它们很快跨出理念界，在现实界里长出公司和法庭，在自我界里长出日记和“按概率办事的我”。

“理论是工具”也不是新话。杜威的工具主义早就说过，观念是人应付处境的工具；沃尔特·翁在《口语与书写》里讲过文字怎样重塑思考；克拉克和查尔莫斯

的“延展心智”则说，纸笔和符号可以成为思考的一部分〔待核〕。本书与它们相同：都不把理论当作对世界的一面镜子。不同在于，我们追的是工具一旦用开，会在封装之外长出什么，以及这些东西怎样回头逼着理论升级。

这一编六章，每章一件工具：

章	工具	这一章看什么
16	文字	一张押金条，如何长出学校、法庭和“和自己对话的我”
17	零与数字	复式记账如何长出公司，又长回家庭账本
18	几何与逻辑	从量地到“证明”，再到非欧几何和计算机的前传
19	微积分	解决了运动，留下无穷小的洞，一路补到哥德尔
20	概率论	保险、统计与按概率做决定的我
21	AI做数学	从“你解了哪些题”到“你发生了哪些新数学”

本编仍在第四层论证里：理念界的工具，同样只会把理念界、现实界、自我界的智慧一起往上提。这里的史例也只是佐证；第21章谈的是正在进行的事，口径比前五章更谨慎。

第 16 章

文字：契约、法律与"和自己对话"

◆ 一、一张押金条

租过房的人大概都经历过这一幕：房东拍着胸脯说“押金一分不少，退房那天就还你”。你当时觉得，这话说得诚恳，何必写下来。一年后退房，房东记得的是另一个数字，墙上有一块污渍，说是你弄的。两个人都不算撒谎，各自记得的都是对自己顺口的那个版本。

如果当初写了一张条子，事情就是另一个样子。条子上有日期、金额、两个签名。它不会因为谁的心情变化而变化，也不会因为谁的记忆被时间磨掉一角而变化。吵架的时候，两个人不再对着对方的脸说话，而是一起低头看那张纸。

再想一个更小的场景。你列购物清单，写到第六项，突然发现自己忘了买盐。脑子里那张“清单”是模糊的，每一项都在，又好像每一项都不在。一落到纸上，漏洞马上现形。写字这件事，除了把想法记下来，还会反过来改变你的想法。

这两个场景合起来，就是本章要讲的东西：文字这种工具，把一句话从说话人身上剥下来，从说话的那一刻剥下来，单独放在世界上。

◆ 二、原理：封装"说过的话"

用本书的话说，文字是对“说话”的一次封装。说话是活的：有语气，有表情，有当场的上下文，说完就散。人类在几万年里靠说话积累了大量的智慧，故事、规矩、手艺的诀窍都在嘴上传。但嘴上传有三个先天的限制：必须同时在场，必须靠记，每传一手就变一点。

文字把这三个限制一并松开了。它让话可以离开说话的人，可以跨过时间，可以原样复制。一句话被封装成一块泥板、一片竹简、一张纸，谁拿到都能原样读出来。这是对“说过的话”的封装，封装之后，说话的智慧可以脱离说话人而存在。

说得再准一点：文字不是一块更耐用的记忆，而是在两个人之间多出了一个“第三者”。押金条躺在桌上，不属于房东，也不属于你，吵架时两个人一起低头看的，就是它。后面那些东西，书吏、法庭、日记里的“那时的我”，都是在这个第三者的位置上长出来的住户。**文字封装的是话，新长出来的，是一个不属于任何一方的位置。**

接下来才是本书关心的重点：文字被大量使用之后，发生了什么没有被封装进去的事。

◆ 三、史例一：先记账，后有文学

人们常把文字想成诗歌与经典的起点，其实最早的大宗文字用途要琐碎得多。公元前四千年纪末到三千年纪初，两河流域的乌鲁克一带出现了最早的成熟书写，泥板上记的主要是粮食、牲畜、劳力的收支和分配。〔待核：具体年代取“约公元前3300年前后”的通行说法〕也就是说，文字最早是一本账。

这并不奇怪。一个城邦，有了神庙的仓库和成千上万的劳动者，谁交了多少大麦、谁领了多少口粮，靠一个管事的人记在脑子里，已经不可能。他会记错，会生病，会死，会被收买。于是把账封装到泥板上。

有一种说法认为，更早的记账用的是各种小小的泥质筹码，筹码再变成泥板上的符号，这条线索由考古学者施曼特-贝瑟拉特（Denise Schmandt-Besserat）提出，学界对它的细节有争论，本书只取“文字与记账关系极深”这个较稳妥的层面。〔待核〕

账封装进了泥板，第一轮碰撞随即发生。账一经写下，就有了“可以被检查”的性质。谁都可以拿泥板核对粮仓里的数。核对的人需要识字，于是出现了专门学习书写的人——书吏，出现了培养书吏的学校。学校一旦有了，就会有人拿它来抄写别的东西：赞美神的歌、国王的功绩、故事、谚语。文学是账本长出来的旁枝，没有人在设计文字的时候想到这一点。

这正是错位定理的样子。发明泥板的人想的是对账，对账以外的那一大片世界（学校、文学、历史、教育考试），在封装的那一刻根本不存在，当然也不在封装之内。

◆ 四、史例二：契约与法律

回到开头那张押金条。条子的本质是：两个人之间的约定，从“我信你”变成了“白纸黑字”。

人类把这件事做成一个庞大的体系，就是契约与法律。约公元前十八世纪的《汉谟拉比法典》刻在高大的石柱上，条文的写法很有特点：如果某人如何如何，则如何处置。〔待核：年代取通行说法“约公元前1754年前后”〕这种“如果……则……”的句式，让规则第一次可以被不认识立法者的人读到、引用、质问。法官可以说“条文在此”，百姓也可以说“条文在此”。权威的一部分，从国王的嘴挪到了石头上。

我们今天习惯的一切，几乎都挂在这个封装上：买房有合同，上班有劳动合同，开公司有章程，国与国之间有条约。它们有一个共同点，约定一旦成文，就能在当事人都不在场、甚至都已不在人世的情况下继续生效。这是一种极为特殊的东西，一个由文字撑起来的现实，现实界的新秩序。

请留意它的路径：文字属于理念界，是对语言的封装；但它在使用中生出的契约、法庭、官僚、税收体系，是现实界里实实在在的结构。工具不会停在出生的那一界，这个模式在本章第一次露面，后面五章会不断重复。

◆ 五、史例三：苏格拉底的担心

文字刚刚流行的时候，并不是人人叫好。柏拉图在《斐德罗》里借苏格拉底之口讲过一个故事：发明文字的神把它献给国王，说这能增强人的记忆，国王却回答说，这只会让人更健忘，因为人将不再靠自己内在的记忆，而去依赖外在的符号。〔待核：对话细节，转述为大意〕

这个批评不能当作老古董一笑置之，它碰到了一个真实的现象：文字确实让一部分口头记忆的本事退场了。古代吟游诗人能背诵成千上万行的史诗，今天很少有人能做到。

按本书的命题5，这是让渡，不是退化。吟游诗人的记忆本事并没有凭空消失，而是被让出去了：脑子腾出来的地方，用来读更多的书，比较不同的说法，怀疑，推理，写批评。苏格拉底看到了损失，没有看到被腾出来的那一块空地后来长出了什么，这是任何身处封装完成、新SDE尚未显露的等待期的人，都看不见的东西。

◆ 六、史例四：默读，和自己对话

最有意思的一层发生在人的内部。

一般认为古代的阅读多半是出声的。奥古斯丁在《忏悔录》里记过一个小细节：他见到米兰主教安布罗斯读书时，眼睛在页面上移动，嘴唇和舌头却不动，他觉得很奇怪。这一幕常被用来说明默读在当时并不普遍，不过古典学者对它到底能说明多少，有争论。〔待核：转述大意〕

默读看起来是个小事，但它意味着：一个人可以在不开口、不被任何人听见的情况下，与一段文字单独相处。再往前一步，他可以自己写，写完自己读，读了觉得不对，划掉重写。

这就是日记、札记、书信草稿、自我反省录的土壤。一个人把当天发生的事写下来，三个月后翻看，会读到一个“那时的我”，既熟悉又陌生，他会在心里嘀咕：当时怎么会那样想？这个“读到过去的自己、并和他说话”的动作，是自我界里一个新的结构。口头世界里，当然也有反省，有祷告和独白，但文字让这件事可以重复、可以积累、可以被自己修改。你可以拿着自己三年前的一页纸，和现在的你辩论。

所以文字这件理念界的工具，在自我界里生出了一个新东西：和自己对话的那个“我”。（这个话题在第五编还会更完整地展开，那里有日记、书信和自传的专章。）

◆ 七、回扣承重命题

把上面几件事排在一起看，每一条都对得上前面立过的命题。

文字把口头说话里的智慧装了起来（命题1）；账本与学校、条文与纠纷、页面与内心碰撞，生出书吏阶层、法庭和日记（命题2）；这些在封装的时候都不存在，所以在封装之外（命题3）；口头记忆的退场是让渡（命题5）。最后，它们住在不同的界里：文字出生在理念界，却在现实界长出了契约和国家的行政，在自我界长出了写给自己的那一页纸。一界有工具，三界有新生。

这里还有一层值得说一句。文字把“智慧”封装了，却从来没有把“智慧的发生”封装掉。读书的人拿到的是别人写下的成品，然后必须自己去读、去想、去怀疑，才发生出自己的理解。同一本书，十个人读，得到十种东西。封装品静静躺在那里，发生仍然在每个读的人身上进行。

◆ 八、这个论证的边界在哪里

要诚实交代几件事。

第一，不能把文字说成自我意识的“发明者”。没有文字的社会，人一样会自言自语，一样会反思、悔恨、做梦。文字改变的是自我反思的方式和可积累的程度，不是有无。把“有文字才有内心”这种话当作结论，是过度的。

第二，不能把文字说成国家与法律的唯一起因。契约与法律的生成，还有农业、人口、战争、宗教等一大堆因素。本章只说文字是其中一个重要的支点，一个让规则可以脱离人身而存续的支点。这是多因之一，不是单因。

第三，关于文字起源的细节（具体年代、筹码说）和苏格拉底故事的细节，本章用的是通行说法，相关细节已标〔待核〕。

第四，“先记账”只是两河流域这一条线。中国最早的成熟文字甲骨文，大宗内容是占卜，所以“文字起于账本”不能推广成文字的一般起源。这并不动摇本章的论证：不论起因是账还是卜，文字用开之后，同样生出了封装之外的东西；但它提醒我们，别把一个地区的路径当成通则。

第五，本章的判断，在这种情形下是错的：如果有一个社会，文字使用了几个世纪，却没有长出专门的书写职业、成文的约定，也没有出现“对着文本争论”的习惯，那么“文字腾出了一个第三者的位置”就缺了现实的支撑。

第六，最大的一条：这一章的几个例子，都是已经发生完的事，我们是在事后回头看到那些新东西的。历史归纳在这里是佐证，真正的论证在结构：封装发生在一个时间点，新东西发生在之后，所以它必然在封装之外。这个区分，第四章已经讲过，这里只是再提醒一遍。

◆ 九、小结

一张押金条，一份购物清单，两件很小的事，背后是人类最早的一次大封装：把说过的话装进可以离开说话人的东西里。封装完成之后，真正有意思的事才开始：账本长出了学校，学校长出了文学；条文长出了法庭，法庭长出了国家的日常；一页纸长出了那个能和自己对话的人。这些，都是装进去的那个人没有想到的。

> 带走的话
>
> 一张押金条，是在两个人之间多出了一个不属于任何一方的“第三者”。账本长出的学校，条文长出的法庭，日记里那个“那时的我”，都是住进这个位置的新住户，发明文字的人一个也没想到。

小账 封装了什么？——说过的话：让话离开说话人、跨过时间、原样复制。 逼出了什么？——书吏与学校、契约与法律、口头记忆的让渡，以及能和自己对话的“我”。

承接：文字封装了语言，下一个要看的是更窄也更锋利的一类文字：数字。一个符号“0”和一本“账”，怎样一路长成公司，又长回你家的餐桌上。

第 17 章

零与数字：复式记账、公司与家庭账本

◆ 一、月底，工资去哪了

很多人都有过这个月底：工资到账不久，卡里就只剩下零头。回想一下，没买什么大件，可钱就是不见了。于是下定决心，从下个月起记账。

记了三个月，有一件事会发生：你开始“看见”自己的钱。原来每天一杯咖啡，加起来比房租的一半还多；原来每月自动扣款的几个会员，有两个一年没打开过。账本一摊开，什么也没有改变，花钱的人还是原来那个人，可他已经站在了另一个位置上，开始审视自己的钱。

更有意思的是那一栏“余额：0”。如果卡上没钱了，你写下的是“0”，而不是空着。空着是不知道，写“0”是知道，并且知道的就是没有。这个看起来微不足道的区别，里面藏着人类算术史上最了不起的一步。

本章要沿着一条线走：从数字，到“0”，到复式记账，到公司，再回到你家的账本。这条线是理念界的工具如何跨界生出新东西的一个完整样本。

◆ 二、原理：封装"多少"，再封装"来去"

按本书的用语，数字是对“多少”的封装。一个人看到三只羊，说“三”，这个“三”可以脱离羊，被拿去数粮食、数天数、数欠债。位值制的数字系统更进一步：同样是“1”，放在个位是一，放在十位是十。要做到这一点，缺不了一个东西：零。

没有零，位置上“什么也没有”这件事就无法表示。罗马人写“二百零五”，要写成 CCV，中间那个“空位”是靠字母的排列隐含出来的，没有被明确记下来。这样的数字能记数，很难做运算。古罗马人的商人和官吏，算起账来依靠的是算盘一类的计算板，而不是在纸上列竖式。

零的故事头绪很多。巴比伦人很早就有了表示“空位”的记号，玛雅人也有自己的零。把零当作一个真正的数、并写出它参与运算的规则的，通常认为是印度的数学家，七世纪的婆罗摩笈多（Brahmagupta，628年前后）留下了较早的明确规则。〔待核：年份取通行说法〕后来印度数字经阿拉伯世界传入欧洲，斐波那契在1202年写下《算盘书》（Liber Abaci），向意大利商人介绍这套写法，这是一个常被提到的节点。〔待核〕

◆ 三、史例：复式记账，每一笔都记两次

数字有了，接着是怎样用数字记“钱的来去”。

先看一个日常案例。你花3000元买了一部手机。如果只记单边账，是“支出3000”，完了。复式记账的写法是，同一件事要从两头各记一次：一头是你的现金少了3000，另一头是你多了一件价值3000的东西（手机）。两边的数总是相等的。

为什么要这么麻烦？因为只要每笔交易都两头记，整本账就有了一个自带的检验：全部账目加起来，一头的总和必须等于另一头的总和。要是对不上，就一定有哪一笔记错了、漏了、或者被人动了手脚。账本从此有了免疫系统。

复式记账在十四世纪前后的意大利商人群体中逐步成形，热那亚、佛罗伦萨、威尼斯的商号都在用。1494年，修士数学家卢卡·帕乔利（Luca Pacioli）在威尼斯出版了《算术、几何、比例与比例性概览》（Summa），其中一部分系统讲述了威尼斯的簿记方法。需要说清楚的是，帕乔利是整理者与传播者，不是发明者，他把商人们已经在用的做法写成了可以教、可以学的文字。〔待核：书名译法与具体内容范围〕

这就是一次典型的封装：商人们几代人的实践智慧，被收进了一本可以印刷的书。从此，一个欧洲的年轻学徒不必在某个老掌柜身边熬十年，读书就可以学会一套能算清账的办法。

◆ 四、碰撞：账和"人"分开了

封装好的复式记账被大量使用，很快碰上了一个它的整理者没有想过的问题：账记的到底是谁的？

早期的商号，老板的钱和生意的钱常常搅在一起。老板拿了店里的钱给孩子交学费，算生意的支出还是算家里的支出？一个人单打独斗时，这不算大事，因为反正都是他的。可是一旦有合伙人，有外面投来的钱，问题就尖锐了：你拿的是谁的？

用复式记账的方式处理这件事，办法很自然：为生意单独开一套账，生意本身成为一个“记账的主体”，老板从生意里拿钱，要记作“提取”，往生意里投钱，要记作“投入”。账上的生意，渐渐和老板这个人分了家。

这一步在理念界里只是一个记法上的调整，往后走，却成了现实界的一个重大结构：企业作为一个独立于创办者的“主体”。十七世纪初成立的荷兰东印度公司，常被视为早期股份公司和公开交易股票的代表。〔待核：1602年〕股东出钱，换来的是公司账上的一份权益，而不是对某条船、某次航行的直接所有权。公司有自己的账，有自己的存续，创办人退场了，账还在。后来发展出的有限责任、财务报表、审计、上市披露，全都建立在“公司有一本自己的账”这个基础上。

学界对簿记与资本主义的关系争论了很久：桑巴特、韦伯一路都很看重理性簿记，也有人认为它的作用被夸大了〔待核〕。本书不站队，只取最稳妥的一层：复式记账让“企业”这种主体变得可算、可查、可交易，这一点相当关键。

没有人在1494年设想过证券交易所。帕乔利想教的是怎样把账记对。但“账记对”这个封装一旦在成千上万的商号里使用起来，就把企业的模样塑造了出来。这又是一次错位：新的SDE（公司）在封装之外，因为封装的时候它还不存在。

◆ 五、回到家：家庭账本

故事到这里，似乎已经走出了理念界，在现实界里立起了公司。但路还没走完，因为这个工具还会继续往下走，走进你家。

最早的家庭记账，是一个本子，一笔一笔写“买菜多少”。后来，学过会计、或受过会计思维影响的家庭，开始用另一种眼光看自己：我家有多少资产，有多少负债，每月进多少出多少，净资产是在涨还是在跌。

这是“公司的眼光”被带回了家。资产负债表这个原本是公司用来向投资人交代的东西，现在变成了一对夫妻在餐桌上讨论的话题：房子是资产，但房贷是负债；车是资产，但它每年在贬值；孩子的教育是一笔支出，也是一笔投入。

这是一个值得停一下的细节。几百年前，商人们为了清算合伙的生意，发展出一个把“自己”和“生意”分开记的办法。几百年后，这个办法倒过来，教一个普通人把“自己的生活”当作一个经济体来审视。账本先生出了公司，公司的账又生出了家庭账本，家庭账本再塑造了一个新的自我：一个对自己的收入、风险、现金流有数的人。

这就是自我界的新SDE。一个人的“我”里面，多了一个会算账的我。他会在买东西前问一句“这笔钱换来的东西，值不值得放在我的账上”，也会在经历了一次失业之后发现，原来自己“账上”只有两个月的缓冲。

◆ 六、被让渡的那一部分

和前面的每一章一样，这里也有退场的东西。

算盘的使用者可以在不动笔的情况下完成大数的加减，熟练的人心算速度惊人。当计算器和电子表格普及之后，这种技艺在日常生活中逐渐用不上了。我们可以说这是损失，但更准确地说，这是让渡：算账的力气交给了工具，人的注意力得以转向“该记什么”“该怎么看这些数”。电子表格里的一个公式，比一个算盘高手一个下午能算的还多，但它不会告诉你，这个数字背后，你要不要买下那家店。

◆ 七、一本账怎样改变一个家的谈话

家庭账本带来的变化，有时小到几乎看不见，却实实在在。没有账本的家，谈钱常常是情绪化的：一个说“你又乱花钱”，一个说“那是必须买的”，谁也说服不了谁，因为两个人手里都没有同一张纸。有了账本，谈话换了地点，他们面对的是同一组数字，争论的焦点从“你这个人怎么样”挪到了“这一项该不该买”。

这和第16章的押金条是同一个道理。账本也是站在两个人之间的“第三者”：数字和文字一样，把一件事从两个人的记忆里挪到了两个人共同看得见的地方。一对夫妻第一次坐下来，把全年的收入和支出摊在桌上，常常会有一个共同的惊讶：我们的钱，居然是这样花出去的。这个惊讶，就是新的SDE显露的一个瞬间：账本封装的是“记”，新长出来的是“共同看见”。

还有孩子。有些家庭会给孩子一个小本子，让他记零花钱。孩子在这个本子上第一次发现，钱花出去就没有了，同样一笔钱不能买两样东西。账本在这里变成了一所很小的学校，教的不是算术，而是“选择有代价”这个观念。这也是错位：发明记账的商人，没有想过自己是在给孩子上经济学的第一课。

◆ 八、这个论证的边界在哪里

第一，不要把复式记账和公司的关系说成“因果直线”。公司的出现还有法律、政治、航海、殖民、金融等许多条件，本章只讲账本这一条线，它是必要因素之一，不是唯一原因。

第二，帕乔利不是发明者，这一点必须说清。我们这里借用他，是因为1494年是一个可以放置“封装”的时间点：做法在这时被写成了书，可以传播。

第三，“0”的故事涉及好几个文明、好几次独立的发明与借用，谁先谁后、怎样传播，学者们还在讨论。本章给出的只是通行的叙述，已经标注待核。

第四，也是最容易被忽略的一条：账本能记的东西，只是人生里能被量出来的那一部分。把家庭生活完全按账来经营，会出问题。陪孩子读书的两小时，在账上

是“成本”，在生活里却是别的东西。账本是一个好工具，不是一个好主人。会算账的我是自我界长出的新东西，但如果这个我把其余的我都吃掉了，那就是另一种问题，我们在第六编里会再回来谈“工具的代价”。

第五，这一章的叙述是回头看的。当时的人并不知道自己在做什么，这正是本书所说的错位：新东西出现的时候，没有人手里拿着说明书。

第六，这条线的试金石：如果有一个只用单式账的商号群体，同样稳定地长出了独立于创办者的企业、可转让的股份和独立审计，那么复式记账就只是个伴随现象，不是支点，本章“账本长出公司”的说法就要降格。

◆ 九、小结

从“3”到“0”，是对“多少”的封装；从单边账到两边账，是对“来去”的封装。封装之后，使用的人把“生意”从“老板”身上分了出去，公司作为新的主体长了出来。几百年后，公司的眼光回到了餐桌，一个家庭开始用资产和负债看自己，每个人心里多了一个会算账的我。理念界的一个小小的记法，三次跨界，改变了现实界的组织方式，也改变了自我界的内部构造。

> 带走的话
>
> 账本一摊开，花钱的人没变，却站到了另一个位置上看自己的钱。复式记账长出公司，公司的眼光又回到餐桌，让每个人心里多出一个会算账的我；只是账本是好工具，不是好主人。

> **小账** 封装了什么？——“多少”（数字与零）和钱的“来去”（复式记账）。 逼出了什么？——生意与老板分家、公司成为独立的主体，以及家庭账本和会算账的我。

承接：数字解决了“多少”，但还有一个更古老的问题没解决：怎么才能让人相信一件事是对的。丈量土地的绳子，是怎样一路走到“证明”这个观念的。

第 18 章

几何与逻辑：从丈量土地到证明的观念

◆ 一、铺地砖的人和问"为什么"的孩子

装修过房子的人，大概都见过铺地砖的师傅。他先拉一条线，量出墙角是不是真的垂直，然后沿着线一块一块摆。遇到墙角不太直的房子，他会说一句行话：“这个角不方，得调。”他怎么知道方不方？有个老办法：在一边量三尺做一个记号，另一边量四尺做一个记号，两个记号之间的斜线如果刚好五尺，这个角就是直角。

师傅不需要知道为什么三、四、五刚好能拼出直角，他只要知道“这样量，准”。

而他家上初中的孩子，也许正在课本里做一道题，证明某个三角形的三个内角加起来是一百八十度。孩子拿到题，不是去量，而是去“证”：因为什么，所以什么，又因为什么，所以什么。老师说，量出来的不算，量总有误差，证出来才算。孩子偶尔会抬头问一句：为什么非得证？我量出来不是一百八十度吗？

这个问题不傻，它问到了几何学里最重要的一次转折：从“这样量，准”，到“必须这样推，才对”。

◆ 二、原理：把"量地的经验"封装成"证明"

几何学的原义是“量地”。第11章提过希罗多德在《历史》里的那个说法：埃及的尼罗河每年泛滥，冲掉田界，国王便派人重新丈量，几何学由此而来。〔待核：转述，出处为《历史》卷二〕这个说法不一定是全部真相，但它抓住了几何最早的现实根：地要量得准，量完之后要能分得清，分的人要能让别人服气。

在那个阶段，几何是一堆好用的经验：三四五拉直角，圆的周长大约是直径的三倍多一点，梯形的面积怎么算。巴比伦的泥板上已经有勾股数，中国古代的算书

里也有勾股的算法，说明经验几何在很多文明里早已存在。〔待核：巴比伦泥板 Plimpton 322的具体解读有争议〕

后来发生的一次封装，是希腊人做的：他们不再满足于“这样量，准”，而要问“凭什么准”。约在公元前三百年，欧几里得把前人的成果编成《几何原本》，办法非常特别：先摆出少数几条不需要证明、人人认可的出发点（公设和公理），然后一步一步，只靠推理，推出几百条定理。每一条定理都站在前面的肩膀上，一环扣一环，不许跳步。

这次封装封装的，不只是几何知识，更是一种做事的方式，叫证明。证明不是更严格的测量，而是把“信”从说话的人身上，挪到了每一步推理上。师傅说“这样量，准”，信不信取决于你信不信他；证明则是把每一步摆出来，请你自己去查，谁都可以查。

◆ 三、史例一：证明的观念跑进了现实界

证明这种做法一旦存在，它就不会安分地待在几何里。

在古希腊的城邦里，公共辩论是政治生活的一部分，讲理的方式如果有了几何作榜样，人们就会开始要求：你的前提是什么？你怎么从前提走到结论的？亚里士多德整理的三段论，是对这种推理的第一次系统封装：凡人皆有死，苏格拉底是人，所以苏格拉底会死。这一类推理，不看内容，只看形式。

再往后看，法律推理、神学论证、哲学思辨，都带着几何的影子。1776年美国《独立宣言》开头那句“我们认为这些真理是不言而喻的”，口气明显像公理化论证：先摆出不需要证明的出发点，再从中推出结论。〔待核：只作文体上的类比，不主张作者受欧几里得直接影响〕

这些例子的意思是：几何的封装，让“讲道理”成了一种可以学习的技艺，并进入了法庭、议会和课堂，成为现实界里人们互相说服、互相约束的方式。

◆ 四、史例二：非欧几何，封装之外的新空间

《几何原本》里有一条公设特别啰嗦，就是第五公设，通俗的版本是：过直线外一点，只能作一条直线和它平行。两千年里，很多数学家都觉得这一条不像“公设”，更像一条应该能被证明的定理，于是前赴后继地去证明它。

证不出来。到了十九世纪前半，罗巴切夫斯基和波约伊（高斯也有类似的私下想法）各自独立地换了一个做法：干脆假设这一条不成立，看会推出什么。结果令人惊讶，推出来的不是矛盾，而是一整套自洽的几何。后来黎曼又发展出另一类弯曲空间的几何。〔待核：人名与时间顺序，取“十九世纪上半叶至中叶”的通行说法〕

这件事用本书的话说，就是封装被充分使用之后，使用者在里面碰撞出了封装之外的东西。欧几里得封装的是“空间是什么样”的一种答案，两千年的人们在这个答案里反复使用、反复推敲，终于有人发现：这个答案可以换。换了之后，出现了一整片以前没人想到的天地。

更有意思的是下一步。二十世纪初，爱因斯坦研究引力，需要一种描述弯曲时空的数学工具，而黎曼几何正好在那里等着。广义相对论（1915年前后）把引力说成时空的弯曲。〔待核：年份取通行说法〕第12章说过的卫星钟修正，也要用到它。

也就是说，一条令古人头痛的公设，经过两千年，变成了你手机里导航准确度的一部分。没有人在公元前三世纪设计过这条路，但路确实存在，这是错位定理的又一个例子。

◆ 五、史例三：逻辑自己成了研究对象

证明的观念用得久了，又碰上了一个问题：我们说“一步步推理”，推理本身是什么？

十九世纪中叶，布尔（1854年）把逻辑写成了代数式的运算，真与假被当作可以计算的对象。〔待核〕1879年，弗雷格发表《概念文字》，造出一套精确的符号逻辑。〔待核〕到二十世纪三十年代，图灵给出了“计算”的精确描述。这条线的

末端，是计算机：把逻辑运算焊在电路里，让机器去做。它会在下一章哥德尔和图灵那里再碰头，也是本编最后一章AI做数学的前传。

再回头想一想这个链条：丈量土地的绳子，变成证明的观念，证明的观念变成对推理本身的研究，对推理的研究最后造出了计算的机器。每一次，都是前一个封装在使用中生出了下一个。

◆ 六、自我界：那个"要讲道理"的我

除了跑进法庭和机器，证明的观念还跑进了人的内心。

举一个家庭里的例子。一对夫妻吵架，一个说“你根本不关心这个家”，另一个气得发抖。吵到最后，有人冷静下来问了一句：你说我不关心，依据是什么？是昨天晚上我没洗碗，还是我上个月忘了你妈妈的生日？这两件事能不能说明“根本不关心”？

这就是几何教给人的一个动作：把一个笼统的结论拆开，看它背后的前提，再看从前提到结论走得对不对。它不保证吵架就此结束，却让吵架多了一条出路。

孩子那句“我量出来不是一百八十度吗”，也是同一个动作的萌芽：一个孩子开始想，某件事我凭什么相信它。这是自我界里一个很深的变化：一个人不再只是接受结论，他开始要求“理由”，也开始向自己要理由。他对自己说“我认为”的时候，会下意识地问一句：我为什么这么认为?

这个要理由的我，是理念界工具在自我界里长出来的新结构。它让一个人能够怀疑，也能够被说服。

◆ 七、一个家常的小例子：搬家具

再看一个更朴素的场景。一家人买了新沙发，想知道能不能通过楼道的拐角。有人说“应该能过”，有人说“悬”。最后，爸爸拿来卷尺，量沙发的宽、高、对角线，再量拐角的净空，还在纸上画了个示意图，算了算沙发斜过来时需要的最小通道。

这个过程里，没有一个人在“证明”什么，但每个人都在使用几何的成果：对角线比边长，斜着放能省空间，纸上画得通，现实里才通得过。更要紧的是，这张小图让一家人从各执一词，走到了“看图说话”：几何不止是计算的工具，还是一种让人们把争论放到同一张图上的办法。

◆ 八、回扣承重命题，以及这个论证的边界在哪里

这一章里，欧几里得封装的，是前人从量地中积累的几何智慧，更是证明这种让陌生人也能被说服的办法（命题1）。两千年里人们在第五公设上的反复使用，碰撞出了非欧几何（命题2）；非欧几何在封装时不存在，所以必然在封装之外（命题3）。理念界的这件工具，在现实界成了法庭与议会里的说理方式，在自我界成了向自己要理由的习惯（命题4）。

第一，不能把“证明的观念”说成希腊独有。中国先秦的《墨经》里有相当成熟的逻辑思考，印度也有自己的因明传统。本章以欧几里得为线索，是因为这一支后来与现代数学和科学的接口最清晰，不代表别的传统没有。

第二，欧几里得的体系并不完美。十九世纪末，希尔伯特等人发现，《几何原本》的论证里藏着一些隐含的假定，需要补充更完整的公理。〔待核：1899年《几何基础》〕这本身又是一个例子：封装得再好，在使用中也会露出缺口，缺口又催生更严格的新工作。拉卡托斯在《证明与反驳》（1976）里，把这种“在反例里修补证明”写成了数学的常态〔待核〕，本书与他相同，只是多问了一步：缺口一旦补上，新补出来的东西，在原来的封装里从来不曾有过。

第三，证明不等于说服。现实生活里的前提，很少像公理那样干净。夫妻吵架可以一步步拆，但“关心”这个词本来就没有欧几里得式的定义。把生活全部形式化，是另一种错误。

第四，关于人名、时间和相对论修正的具体表述，已按通行说法写出并作标记，需要再核。

第五，历史归纳在这里仍是佐证。证明这一章要靠的是结构：封装完成之时，使用者才刚刚开始用，使用中产生的东西自然不在封装之内。

第六，这一章最容易被检验的是非欧几何那一环：如果史料显示，罗巴切夫斯基、波约伊并不是在两千年证明第五公设的传统里走出来的，而是另有独立的来源，那么“封装在使用中碰出封装之外的东西”这个例子就应当撤下。

◆ 九、小结

量地的绳子，被希腊人变成了一套“必须这样推”的做法。这套做法先走进了法庭和议会，成为人们互相说服的方式；又在被反复使用的过程中，碰撞出非欧几何，后来成了描述宇宙的语言；再往后，它回过头来研究自己，造出了计算机的前传；它还走进每个人的内心，让人学会向自己要理由。

> 带走的话
>
> 铺砖师傅说“这样量，准”，要你信他；证明把每一步摆出来，请你自己去查。这个动作先走进法庭，又撞出非欧几何，最后走进每个人心里，变成一句“我为什么这么认为”。

> **小账** 封装了什么？——量地的经验，以及“凭什么准”的证明做法。 逼出了什么？——公共说理的技艺、非欧几何、逻辑成为研究对象，以及向自己要理由的我。

承接：几何与逻辑处理的是静止的图形和固定的推理。下一章要面对的，是会动的东西：速度、变化、无穷小。微积分解决了运动，也生出了让数学家睡不着觉的危机。

第 19 章

微积分：解决了运动，生出了无穷小危机，一路走到哥德尔

◆ 一、仪表盘上的"六十"

开车的人每天都在看一个数字：仪表盘上的时速。此刻显示六十。

仔细想想，这是一件奇怪的事。速度的意思是“走了多远，除以用了多久”。要算速度，得有一段路和一段时间。可仪表盘说的是“此刻”，而一个此刻没有长度：它不占任何时间，当然也走不了任何路。那么，一个没有长度的时刻，怎么会有速度?

再换一个厨房里的场景。浴缸放水，水龙头开着，水位一点点上升。你想知道五分钟后水深多少，需要知道每一刻的进水速度，再把每一刻的进水量累加起来。速度是“此刻的变化”，水深是“所有变化的累积”。

微积分要对付的，就是这两件事：此刻的变化（微分），以及变化的累积（积分）。它是人类第一次把“在变”这件事，变成可以精确计算的对象。

◆ 二、原理：封装"变化"

按本书的用语，微积分是对“变化”的一次封装。在它之前，人们当然知道东西会动、会涨、会落，但对运动的描述很零碎：每个问题要用特别的办法，算一条曲线下的面积，算一个落体的速度，各有各的窍门。

十七世纪后半叶，牛顿和莱布尼茨各自独立地把这些零碎的窍门统一成一套通用的办法：给一个会变的量，如何求它此刻的变化率；给一个变化率，如何求它累积的总量；而且这两件事互为逆运算。〔待核：两人各自独立发明、后有优先权之争，这是通行说法〕这是一种封装，把几代人的零散智慧收成了一套谁都能学、谁都能用的计算规则。

使用的结果是惊人的。行星的轨道可以算了，炮弹的落点可以算了，桥梁受力可以算了，潮汐可以预报了。牛顿力学的整个大厦，就建在这套工具上。工程师们拿着它，在随后的两三百年里改造了现实界。

◆ 三、史例一：封装里藏着一个洞

但封装刚完成的时候，里面就有一个洞，只是使用的人暂时顾不上。

回到仪表盘。早期微积分计算“此刻速度”的办法是：取一个极短的时间间隔，算这段间隔里走的距离，再相除；然后让间隔越来越小，小到“无穷小”，看结果趋向哪个数。问题出在这个“无穷小”上：在计算的前半段，它被当作不等于零，否则不能做除法；在计算的后半段，它又被当作等于零，否则结果里会留下尾巴。一个量一会儿是零、一会儿不是零，这在逻辑上说不通。

十八世纪的哲学家兼主教贝克莱（George Berkeley）看穿了这一点，1734年发表《分析学家》，毫不客气地嘲讽这些无穷小“是已逝量的幽灵”。〔待核：该短语的中文译法〕这句话刻薄，却一针见血。

这里有一个很值得注意的现象：微积分在这个时期是“好用但说不清”的。它的结论屡屡被事实验证，推导的根基却站不稳。工具已经封装好，工具运转良好，但封装里面有一块是模糊的。人们之所以能放心用，是因为实践上它很少出错。

这是一个很典型的状态。回想前面说的等待期：封装已经完成，新的SDE尚未显露。当时的数学家们生活在这段等待期里，他们看到的是一个无比强大的工具，没有人能预见到，为了把根基补牢，将要发生什么。

◆ 四、史例二：补洞，补出了一整个新世界

补洞的工作用了一百多年。

十九世纪初，柯西（Augustin-Louis Cauchy）在他的《分析教程》（1821年）里，开始用“极限”重新定义这些概念。推动补洞的，不全是贝克莱那句嘲讽，更多是使用中碰到的怪例子，比如傅里叶级数让人发现，“函数”“收敛”这

些词原来没有说清〔待核〕。后来魏尔斯特拉斯把它推到极致，用我们现在课本里的“ε-δ语言”把极限写得毫无歧义。〔待核：柯西与魏尔斯特拉斯各自的贡献分界，通行叙述如此，细节学者有讨论〕

这套语言说的是什么？用日常话讲，它是一份“服务承诺”。你对速度表说：我要求误差不超过一个很小的数，不管这个“很小”有多小，你总能给我划出一个足够短的时间窗，只要在这个窗里算，结果就一定落在我的误差之内。谁提要求，谁就赢不了：对方总能满足。如此一来，“无穷小”这个会闪烁的幽灵不再需要了，它被换成了一种“要多准有多准”的保证。

洞补好了，但事情没有停在这里，因为补洞的人发现，要把极限讲清楚，得先把“实数”讲清楚：数轴上那些无穷多的点，到底是什么？于是有了对实数的严格构造，有了康托尔的集合论，也就是用“集合”去装一切数学对象的办法。

集合论本来是补根基的工具，结果自己又出了毛病。1901年前后，罗素发现了一个悖论：考虑“所有不包含自己的集合”组成的集合，它到底包不包含自己？包含，则不该包含；不包含，则该包含。〔待核：1901年发现，1902年写信告知弗雷格，是通行说法〕日常里有个现成的说法：一个村子里，理发师只给不给自己刮脸的人刮脸，那么他给自己刮不刮？刮也不对，不刮也不对。

这一击打在刚写完《算术基本法则》第二卷的弗雷格身上，据说他在书稿即将付印时收到罗素的信，只能在后记里承认自己的基础出了问题。〔待核〕

于是数学家们又开始修补：发展出公理化的集合论，限制什么样的东西才算集合。希尔伯特提出一个宏大的计划：把全部数学放进一个形式系统里，证明它自洽、完备，从此万无一失。

◆ 五、史例三：哥德尔，补洞计划的天花板

1931年，哥德尔证明了两个不完备定理。粗略地说：只要一个形式系统足够强，强到能做基本的算术，并且它自己没有矛盾，那么这个系统里就一定有一些命

题，它既证明不了、也否证不了；而且，系统也无法在自己内部证明自己没有矛盾。〔待核：定理的精确表述需谨慎，此处为通俗转述〕

这个结果直接叫停了希尔伯特的原计划。人们曾以为数学可以被一次性装进一个无懈可击的盒子；哥德尔证明，不管这个盒子造得多精巧，只要它足够强，总有盒子自己装不下的东西。

把这个故事从头到尾读一遍，会看见一条清晰的链子：微积分封装了变化；使用中暴露了无穷小的洞；补洞催生了极限、实数和集合论；集合论生出悖论；修补悖论的努力把数学推进到形式系统；形式系统被哥德尔证明有边界。每一步，都是上一步封装在使用中碰撞出来的新问题，每一步的解答，又成为下一步的封装。

更值得一提的是，这条链子的终点，又一次通向了计算。图灵在1936年对“可计算”的刻画，与哥德尔的工作血脉相连。哥德尔的结果给形式系统划了边，图灵的结果给机器能做的计算划了边。这个线索，我们在第九编还会遇到。

◆ 六、变化率进了生活

微积分当然不只在数学史里。

在现实界，它是工程的底层语言。在经济学里，“边际”这个词就是微积分的日常版：多生产一件产品，多花的成本是多少？多吃一口，多得的满足是多少？一个生意人问“再多开一家分店值不值”，问的就是边际。药理学里，药在血液里的浓度随时间怎样上升和衰减，医生据此确定一天吃几次。

在自我界，它也留下了痕迹，只是不那么显眼。受过这套思维训练的人，会不自觉地换一个问法：不只问“我现在多少分”，而是问“我在往哪个方向走，走得多快”。一个孩子的成绩是七十五分，是进步中的七十五，还是退步中的七十五，对家长来说是两回事。体重、存款、情绪，一个人看自己的方式从“状态”变成了“趋势”。这个“看趋势的我”，是理念界的变化率进入自我界后的新结构。

◆ 七、回扣承重命题

这一章是错位定理的一个格外清楚的样本。牛顿和莱布尼茨封装微积分的时候，极限的严格定义、实数的构造、集合论、罗素悖论、形式系统、不完备定理，一个都不存在。它们之所以后来出现，是因为封装被大规模使用、在使用中碰到了自己的缺口。新问题在封装之外，因为封装时它还不是问题。

也可以看到第8章讲的升级：每一次补洞的努力，不是在原来的层面上修修补补，而是上升到一个更高的层面去重新想：从无穷小的计算，升到极限的定义，升到集合，升到形式系统，升到对形式系统的研究。每一层都比前一层更抽象，也更接近“数学自己是怎么回事”。

◆ 八、这个论证的边界在哪里

这一章必须特别小心，因为哥德尔定理是最容易被滥用的一个结果。

有人想借它证明“机器永远不如人”：既然任何形式系统都有它证明不了的真命题，而人看得出这些命题为真，那人就比机器强。这类论证（常与卢卡斯、彭罗斯等人的名字相连）一直有很多反驳，比如，人未必能证明自己的一致性，人看出“为真”也未必有保证。〔待核：相关论证的具体提出者与争议状态〕

本书不借用这条路。哥德尔定理在这里的角色，只是历史上的一个例子：一个自以为完备的封装，在使用中暴露出了自己的边界，并且这个边界逼出了更高一层的研究。本书要证明的是封装与新SDE在时间上的错位，不是“人的心灵超越计算”。这两件事不能混为一谈。

另外，“危机推动进步”也不是一条普遍规律。库恩把危机写成科学革命的前奏，那是他对一类变化的概括；有些危机一直没有被解决，有些问题被搁置了很久。数学家们对无穷小还有另一条路：二十世纪六十年代，罗宾逊发展出“非标准分析”，让无穷小在严格的意义上重新回来，这说明同一个洞可以有不止一种补法。〔待核〕

还有一条，是这一章可以被检验的地方。如果极限的严格定义出现之后，数学并没有因此引出实数的构造、集合论这类新对象，而只是把旧公式写得更严，那么“补洞即升层”就应当降格成“修订”。可以回到历史里逐项核对：每一个新对象，是不是确实在补洞的过程中才出现的。

最后，本章对数学史的叙述极度压缩，柯西、魏尔斯特拉斯、戴德金、康托尔的工作远比这里写的复杂；年份和人名都按通行说法给出，细节请读者以专门的数学史著作为准。

◆ 九、小结

仪表盘上一个“此刻的速度”，让微积分成了现代工程的底座。它封装了“变化”，也把一个会闪烁的无穷小留在了里面。补那个洞，用了一百多年，补出了极限、实数、集合论，又撞出了悖论，最后撞到哥德尔划下的边界。**封装不是终点，每一次封装的缺口，都是下一次发生的入口。**

> 带走的话
>
> 仪表盘上的“六十”好用，靠的却是一个一会儿是零、一会儿不是零的无穷小。补这个洞补了一百多年，补出极限、实数和集合论，也撞上悖论和哥德尔划的边；洞不是工具的败笔，是下一层的入口。

> **小账** 封装了什么？——“变化”：求此刻的变化率，以及变化的累积。 逼出了什么？——极限的严格定义、实数与集合论、罗素悖论、形式系统，以及哥德尔划出的边界。

承接：微积分处理的是确定的变化。可日常生活里，大多数事不是确定的：明天下不下雨，会不会生病，这笔生意能不能成。下一章讲的，是人类怎样把“没把握”也封装成工具。

第 20 章

概率论：保险、统计与"按概率做决定的我"

◆ 一、三成的雨

早上出门前，手机上写着：降水概率百分之三十。你犹豫了一下，没带伞。同事带了。

下午没下雨。同事有点不好意思地说：白拿了一天。你可能还会得意：看，预报不准。

可是，预报真的不准吗？说“百分之三十”的时候，预报并没有说“今天不会下”，它说的是：在像今天这样的天气形势下，大约三成的情形会下雨。如果把几年里所有报“百分之三十”的日子找出来，其中真下了雨的占三成左右，这个预报就是准的。这叫“校准”：它不预言单独的一天，它对一类日子负责。

大多数人并没有这样想过，但每天都在这样用：看到概率，然后做决定。带不带伞，买不买保险，要不要做这个手术，志愿是冲还是稳。一个现代人，每天要做许多次“在没把握的情况下下决心”，这件事本身，就是一个被封装过的技艺。

◆ 二、原理：封装"没把握"

按本书的用语，概率论是对“没把握”的一次封装。

在它出现之前，没把握只是一种情绪，一种说不清楚的担心，一种“听天由命”。概率论把它变成了一个数，0到1之间的一个数，可以相加，可以相乘，可以随着新证据而修改。一旦有了数，就可以比较（这个手术的风险比那个小），可以定价（这份保险该收多少钱），可以决策（风险低于多少，值得去做）。

封装的是什么智慧？是几百年里赌徒、商人、船东、医生在面对不确定性时，积累下来的一点点规律：单次不可测，大量可测。

◆ 三、史例一：保险走在数学前面

先说一个容易被忽略的事实：保险不是先有概率论、后来才有的。

地中海的海上保险在中世纪就已经出现了。一条船出海，可能沉，可能满载而归；船东们开始想一个办法：我出一点钱，换你承担一部分风险。那时没有精确的计算，凭经验定价。十七世纪末的伦敦，船东、商人和承保人聚集在爱德华·劳埃德开的咖啡馆里谈生意，1688年的《伦敦公报》已有这家店的记载，后来发展成了著名的劳合社。

保险的原理，用一个邻里的例子最容易讲明。一个村子有一千户人家，每年大约有一户会遇到火灾，损失一万元。对任何一户来说，火灾是不可测的，可能明年就轮到你，可能一辈子不会。但对整个村子来说，每年总损失大约是一万元，是相对稳定的。于是大家每户出十元，凑一万，谁家遭灾谁拿。单户面对的是命运，一千户合起来面对的是一个可以估算的数字。

这是概率论最实用的一个洞见：个体的不确定，在群体里变成确定。

数学是后来才来追上这件事的。1654年，帕斯卡和费马在通信里讨论赌局中断时怎样公平分赌注，通常被看作概率论的起点之一〔待核〕。1662年，格朗特分析伦敦的死亡统计表，发现死亡的规律惊人地稳定。1693年，哈雷根据布雷斯劳的出生与死亡记录编出生命表，年金才有了算法〔待核〕。1713年，雅各布·伯努利去世后出版的《猜度术》，给出了大数定律：试验次数越多，频率越接近真实的概率。

这个次序很有意思。是实践先走一步，保险业在做，赌桌上的人在算，数学家把它们整理成理论，理论又反过来让保险定价更准、保险公司更敢做。这正是封装的作用：把模糊的经验变成规则。

◆ 四、史例二：统计，国家学会了"看数"

概率论和一个词的故事缠在一起：统计学（statistics）一词与“国家”（state）同源。〔待核：词源说法〕国家要征税，要征兵，要预防瘟疫，就需要

知道自己有多少人，多少是男人，多少是孩子，多少在生病。人口普查、死亡登记、出生登记，一个国家开始用数字“看”自己。

这是现实界里的一个新结构：公共卫生、社会保险、养老金、人口政策，都建在统计之上。这些东西，十七世纪的人不可能想象得到。

医学是另一个深受影响的领域。今天，一个新药能不能上市，要看随机对照试验的数据；医生向你解释风险，也开始说“十年内复发的概率约为多少”。

◆ 五、一个日常的陷阱：阳性不等于患病

概率进了生活，就带来一个很普遍的新问题：人们不擅长处理它。

设想一种疾病，一千人里有一个人患病。有一种检测，准确度很高：患者查出阳性的可能性是百分之九十九，健康人被误判成阳性的可能性是百分之一。现在你查出了阳性，你患病的可能性有多大?

很多人的直觉是百分之九十九。可算一下：一千人里，一个患者，几乎肯定查出阳性；九百九十九个健康人里，百分之一，大约十个人也被误判为阳性。于是阳性的总共约十一个人，其中真患者只有一个，大约百分之九。

这个例子对应的就是贝叶斯定理：新证据出现之后，要把“原来有多大可能”和“这个证据有多可靠”一起考虑。托马斯·贝叶斯的论文在1763年他去世后发表，后来拉普拉斯把它大大推进。〔待核〕

重要的是：这类错误不是某些人笨，而是这个工具本身需要学习。概率是封装好的思维方式，但它不是“自动好用”的，它要求使用者在脑子里建立一种新的感觉。而这个感觉，正是使用时才发生的。

◆ 六、自我界：按概率做决定的我

概率进入自我界，生出了一个新的我。

一个人学会了概率思维，说话方式会改变：他不再说“肯定会成”或“肯定不行”，而是说“我有七成把握”。他会在做决定时，把结果和决定分开看：一次决

定做得好不好，要看当时的信息和权衡是否合理，不能只看结果。一个合理的决定，也可能碰到坏结果；一个草率的决定，也可能碰巧成功。这个区分，听起来简单，真正做到很难，但它是成熟的标志之一。

再比如，家长给孩子填志愿，会把学校分成“冲、稳、保”三档。这就是一种概率分层：冲一冲，成功率低，收益高；稳的，大概率；保的，几乎有把握。这个分层背后的思维方式，几百年前的人没有。

还有一层更深的变化。以前，人们遇到灾祸，倾向于归到“命”和“运气”里面，今年的收成不好，是天意。有了保险与统计之后，灾祸被重新放进“风险”这个框架里：风险是可以估计的，可以分摊的，可以预防的。人对自己命运的位置，从被动接受变成了主动管理。这是自我界里一个相当大的转变。

哈金在《驯服偶然》（1990）里讲过类似的故事：统计怎样一步步改变人们看待自己和他人的方式，连“正常人”“人口”这样的类别，也是统计造出来的〔待核〕。

◆ 七、新手家长和"孩子发烧"

再举一个让人心里发紧的日常场景。孩子半夜发烧到三十九度，新手父母慌了：去不去急诊？去，可能白折腾一夜；不去，万一是严重的病怎么办？

这个场景里有概率的全部难处。多数发烧是普通感染，几天会好；极少数是重病，需要立刻处理。医生给出的建议往往是一份“警示清单”：精神状态如何，有没有抽搐，有没有皮疹，持续多久。这份清单，实际上是把“极少数的危险情形”从“多数的安全情形”里筛出来的概率工具，医生用大量病例的统计，换成了父母手里一张可以对照的纸。

有了这张纸，父母的恐惧没有消失，却变得有方向了：他们不再被一团模糊的担心压着，而是对照着看几条具体的信号。这就是概率封装的价值：它不替你做决定，也不保证结果，但它把“没把握”整理成了你可以使用的东西。

而这对夫妻的这一夜，也在悄悄改变他们自己。经过几次这样的夜晚，他们会开始说“这种情况大概率没事，但要盯着”，一个更成熟、更能带着不确定性生活的家长，就这样一点点长出来了。

◆ 八、回扣承重命题，以及这个论证的边界在哪里

概率论把“没把握”封装成了数。使用之后，现实界里长出了保险业、统计局、临床试验；自我界里长出了“按概率做决定的我”；而理念界自身也在使用中被推着往前走。1933年，柯尔莫哥洛夫给概率论建立了公理化的基础。〔待核〕但“概率到底是什么”，是长期出现的频率，还是一个人心里的信心程度，至今仍有不同学派在争论。这是一个典型的现象：封装好用，但封装背后的“是什么”并没有被完全封死，使用越广，问题越深。

这又是错位。伯努利和帕斯卡不可能预料到，三百年后一个体检报告会让一个普通人对“阳性”感到恐慌，也不可能预料到，需要为“怎样让公众理解风险”专门发展出一门沟通的学问。这些新问题，都在封装之外。

概率这个工具，恰恰让本书的核心论证显出了一个重要的边界。

概率能处理的，是已经列清楚的可能性：骰子有六个面，天气有下雨与不下雨，船要么沉要么不沉。它在“可能性空间”之内才好用。而本书一直在讲的新SDE，偏偏是那种事先不在可能性清单上的东西：没人把“手机会改变人们的社交方式”列进过任何概率表。所以概率工具能降低已知风险的不确定，但对“没想到的事”几乎无能为力。这并不是概率论的缺点，而是它的本性：它封装的是在既定框架内的不确定性。奈特在1921年就区分过“风险”和“不确定性”：前者可以算出概率，后者连清单都列不出〔待核〕。本书与他相同；不同的是，我们要说的不止是“算不了”，而是清单之外的那些东西，恰恰是工具被使用时长出来的。

第二，概率对一次性的事件意义不清楚。一生只有一次的选择，“成功概率百分之三十”对这一次到底意味着什么？概率说的是一类，而你只活这一次。这个难题是真实的，学界没有公认的答案。

第三，概率模型都有假设，最常见的是“各个事件互相独立”。当现实里的事件其实相互牵连时，模型会严重低估风险。工具用错了地方，会给人虚假的安全感。

第四，这一章的判断也可以被检验。如果在一个足够开放的领域里，人们使用概率工具一两百年之后，剩下的主要意外都能事先列进清单，没有一次是“清单之外”的，那么“新的东西总从清单之外来”就是错的。每逢金融危机或新型疾病，人们总说“没想到”，那只是印象，不能替代对这一点的认真统计。

第五，本章的年份与人名按通行叙述写出，已标注待核。保险的早期历史尤其零散，我们取的是最稳妥的大意。

◆ 九、小结

天气预报里的“百分之三十”，是几百年里人类对付“没把握”的一个成果。赌桌与船队里的经验，被数学家整理成概率；概率再长出保险、统计局和临床试验；也长进了人们的说话方式和决策习惯，生出了一个按概率做决定、能把“决定”和“结果”分开看的我。同时，这个工具也划出了自己的边界：它擅长处理清单以内的不确定，而新的东西，总是从清单之外来的。

> 带走的话
>
> “百分之三十”不预言哪一天，只对一类日子负责。概率把“没把握”变成了能比较、能定价、能决策的数，却只在已经列好的清单以内好用，新东西偏偏从清单之外来。

小账 封装了什么？——“没把握”：单次不可测、大量可测的规律。 逼出了什么？——保险业和统计局、临床试验、“阳性不等于患病”这类要重新学的直觉，以及按概率做决定的我。

承接：数字、几何、微积分、概率，是人类封装出的几件最重的“数学工具”。下一章讲的是：当AI把“解题”这件事也封装之后，数学家们要面对的，是一个什么样的新问题。

第 21 章

AI做数学：从"你解了哪些题"到"你发生了哪些新数学"

◆ 一、辅导作业的夜晚

晚上八点，家长坐在孩子旁边辅导作业。一道应用题，孩子卡住了，家长自己看了半天也没思路。孩子掏出手机，对着题目拍了一张照，几秒钟后，屏幕上出现了完整的解答，每一步都写得清清楚楚。

家长的第一反应往往是复杂的：一半是松了口气，一半是发慌。松口气，是因为题终于有答案了。发慌，是因为一个念头冒了出来：如果这种东西随时可用，孩子还用学做题吗？我们这辈子花了那么多年练出来的东西，难道一部手机就替了？

这个发慌是有道理的，它碰到了一件真实的事：一类过去只有受过训练的人才能做的事，正在被封装进工具里。本章要讨论的是这件事的两面：被封装的是什么，封装之后又会逼出什么。

先把立场摆在前面。本章的论断分两部分，性质不同：第一部分，AI正在封装“解题”，这是眼下正在发生的事实；第二部分，这一封装会把“发生新数学”推成下一个更高级的问题，这是一个正在进行中的判断，不是已经完成的历史。后者我们要说清楚它的依据，也要说清楚它可能错在哪里。

◆ 二、原理：题，和问题

先区分两个词。“题”是别人给你的：已经有了清楚的题面，有了公认的解法范围，有一个标准答案或一类标准的证明。“问题”则是一个人心里真正想弄明白的事，它常常没有题面，没有人告诉你该往哪里走，甚至你要自己先发现“这里有一个问题”。

学校里的数学，绝大部分是题。数学家的日常，绝大部分是问题。在一个人的教育经历里，二者混在一起，人们很容易把“解题能力强”当成“数学好”的全部。

用本书的语言讲，数学里“解已有的题”这一层，是一个封装得越来越好的层面。从古代的算筹、算盘，到对数表、计算尺，再到计算器、计算机代数系统，每一代工具都在封装一块解题的智慧。现在，人工智能系统在这条路上又走了一大步：它们能够处理更复杂的、需要多步推理的题目。

近几年有些公开报道值得提一下，只作线索，不作结论：2024年，有研究团队宣布其系统在国际数学奥林匹克的题目上达到银牌水平；2025年，不止一个团队宣布其系统在当年的竞赛题上拿到金牌水平的分数（总分42分，得35分）。写作时间是2026年10月，AI能力变化极快，读者应以最新情况为准。本书不对任何具体系统做评价。这里要抓住的只有一个结构性事实：以“题”为形态的数学能力，正在被封装。

◆ 三、先看历史佐证：封装解题，不是第一次

这件事并不新鲜，只是这次更猛。

在对数表出现之前，天文学家要花几个月的时间做繁重的乘除运算。17世纪初，纳皮尔发明对数，把乘法变成加法，据说极大地缓解了天文计算者的负担。〔待核〕再后来，计算尺成了工程师的象征，计算器替代了计算尺。每一次，都有人担心：会算的本事要失传了。每一次，实际发生的事情是：会算的力气让渡了，人的注意力转向了“算什么”“算了有什么用”。

1976年的四色定理证明，是一个更接近今天局面的例子。〔待核：阿佩尔与哈肯，1976年〕这个问题说的是：任何地图，只需要四种颜色就能让相邻区域颜色不同。阿佩尔和哈肯的证明，需要用计算机检查大量情形，人工无法逐一核对。当时数学界出现了真诚的争论：这算证明吗？一个人读不完的证明，还是证明吗？

这个争论，实际上逼出了一个新的问题类别：怎样验证一份由机器参与完成的证明？后来，“形式化证明”的路线逐渐成熟，用专门的证明辅助软件，把每一步推理都编码成机器可以检查的形式。开普勒猜想（关于球怎样堆叠最紧密）就是一例：黑尔斯（Thomas Hales）1998年宣布了计算机辅助的证明，后来又组织团队做形式化验证，2014年宣布完成。

你看，工具参与数学之后，数学里立刻出现了新的东西：新的“什么算证明”的标准，新的验证方法，新的专业分工。这些是在封装之外的，因为封装的时候不存在。

◆ 四、现在的情形：同一条链，走到了更高的位置

把这条链子应用到眼下，会看到什么？这里要分清楚“已经看到的”和“根据结构推断的”。

已经看到的，是封装本身：解题的力气在转移。一部分过去靠大量训练才能做出的推理，现在可以由工具完成。

根据结构推断的，是下一步。如果本书的链条成立，使用会碰撞，碰撞会生出封装之外的新东西。那么，新东西最可能出现在哪里？

我的判断是：出现在“解题”之上的那一层。当解题不再稀缺，稀缺的就变成了别的东西：

一是提什么问题。一个题目好不好，一个猜想值不值得追，不是解题能力能回答的。给一台强大的解题机器一个无聊的问题，它会得到一个无聊的答案。

二是怎样判断一个结果有意义。解出来了，然后呢？这个结果连着哪些别的东西？它改变了我们对什么的理解？

三是怎样验证与信任。当一份证明长到人读不完，当一个结论由不透明的系统给出，数学共同体凭什么相信它？这一问题已经在催生形式化验证这类专门的工作。

四是造概念。数学史上真正的大步，往往不是解出某一道难题，而是造出了一个新的概念，让原来无法表达的事情可以表达了：零、负数、虚数、群、集合、流形。这些东西没有“题面”，是人们先觉得“这里缺点什么”，然后才造出来的。

把这四点合起来，可以换一种说法。过去评价一个数学学习者，问的是：你解了哪些题？以后，这个问题的分量会下降，更有分量的问题是：你发生了哪些新数学？

“提问比解题要紧”不是新说法，爱因斯坦和英费尔德1938年就说过意思相近的话；瑟斯顿1994年那篇谈数学进展的文章，也主张数学的进展是人的理解的进展，而不是定理数量的增加〔待核〕。本书与它们相同：都把重心放在“问”和“懂”上。不同的是，我们给的不是数学家的品味，而是一个结构上的推断：解题一旦被封装，稀缺就要上移，因为封装之外才有新东西。

◆ 五、"你发生了哪些新数学"是什么意思

这句话需要解释，免得变成口号。

“发生”，在本书的用法里，指的是一个此前不存在的东西，从一个人的思考和碰撞中长出来，并且这个人在长出它的过程中被改变了。（这是第七编要详细讲的创造与智慧共生。）

所以“发生新数学”并不只是“做出新定理”。它包括：你提出了一个别人没有问过的问题；你把两个原本不相干的领域连在了一起；你对一个老概念有了新的看法，因而之前无法表述的东西可以表述了。它的特点是，做完之后，做的人自己也不是原来的那个人了。

回到辅导作业的家长和孩子。如果一个孩子用手机拍照拿到答案，抄下来，他得到的是答案，没有发生任何东西。如果他拿到答案之后问：为什么这样走？换一个条件会怎样？这个方法能用到别的题上吗？他就开始了发生。同样一个工具，有人用来替代发生，有人用来推进发生。这个分别，在每一个封装好的工具那里都存在，只是在AI身上最为醒目。

用日常的话类比：导航封装了认路，逼出来的问题是“去哪里，为什么去”；洗衣机封装了洗衣，逼出来的是“穿什么，怎么过日子”；菜谱App封装了“照着做”，逼出来的是“今天给谁做，想做出什么味道”。解题被封装了，逼出来的就是“为什么要解这个”。

◆ 六、与工具让渡律接口

这件事可以放进前面讲过的四阶段里看：扩大期、内卷期、新产业诞生、让渡。（这一框架与作者另一部书《新文明的诞生》相接，这里只用其骨架。）

在AI解题能力迅速提升的阶段，一个很自然的反应是在“题”这条赛道上加码：更难的竞赛、更多的刷题、更高的分数线。这有点像内卷期：大家都挤在同一条赛道里，赛道本身的价值却因为工具的进步在下降。【这是判断，不是已证事实】

如果链条继续，下一步是新产业的诞生：围绕“提问、判断、验证、造概念”的新工作、新训练、新评价方式；再往后是让渡：纯粹的解题技能让出场地，如同心算让给计算器。这一切目前都还是推断。

◆ 七、一个重要的澄清：本书的论证不依赖"AI做不了新数学"

读到这里，一位读者可能会想：作者是不是在说，AI永远做不了新数学，创造新概念只有人能做?

不是。这一点必须说清楚，因为它关系到整本书的论证方式。

假设有一天，某个AI系统提出了一个数学共同体公认的、有价值的新概念，或者提出了一个公认深刻的新问题。按本书的逻辑，这并不构成反例。因为这一成果一旦出现，就成了又一次封装：把“提出新概念”的某种能力封装进去了。而封装一旦被使用，就会碰撞，就会在封装之外生出新的SDE，逼出更高一级的问题。比如：如何在大量机器产生的概念里筛选？数学的整体图景因此怎么变？人的数学与机器的数学是什么关系?

本书要证明的，是封装和新SDE之间的错位，而不是“某类能力人永远独占”。这也是第四章里反复强调的区分：历史是佐证，结构才是论证。

有一点要诚实写明，也是总纲里立的前提：本书把“问题总是对某个主体而言的问题”作为前提。一个问题之所以是问题，是因为有一个主体觉得它是问题，觉得有缺口、有张力。这个前提，不是被证明的，是被设定的。读者可以不接受，但要知道本书的论证在这里站在哪里。

◆ 八、这个论证的边界在哪里

本章的边界比前面几章更多，因为它谈的是正在进行的事。

第一，这是进行中的判断，不是完成的历史。前五章讲的是已经走完的链条，我们可以回头看到那些封装之外的新东西。本章讲的这条链子才走到一半，“新东西”目前只是一些迹象和推断。写作时点之后，局面可能向任何方向变化，包括向不利于本章判断的方向。

第二，可错条件。如果在未来几年里，我们看到AI解题能力大幅提升，而围绕数学的新问题类别、新职业、新的验证与评价问题并没有出现，数学界的重心也没有向“提问与造概念”移动，那么本章的判断就是错的。具体的观察对象，可以包括：形式化验证等工作的规模变化；数学教育与评价方式的实际变化；机器提出的问题或概念被共同体接纳的方式。本书不预言具体年份。

第三，AI当前的技术状态不等于原理必然。本书对当前AI的描述，只限于结构性的事实，比如训练与使用的分离，这是当前技术状态，不是必然。未来的系统可能很不一样。

第四，竞赛题不等于数学研究。竞赛题是有解的、篇幅有限的、预先设计好的题；研究是要先弄清有没有题。在竞赛上的进展，不能直接推到研究上去，这个差距到底有多大，数学家们自己的看法也不一致。

第五，本章引述的新闻性事实只作线索，不作论据。

◆ 九、小结

辅导作业的夜晚，那部手机里的答案是一次封装的缩影：解题，这件曾经需要多年训练的事，正在被装进工具。数学史上，类似的封装一次次发生，每一次，会算的力气让渡，人的注意力上移。眼下这一次封装得更彻底，所以那个被推上来的问题也更尖锐：当解题不再稀缺，数学里还有什么是真正的稀缺？我的回答是：提问，造概念，判断意义，建立信任，一句话，发生新数学。这个回答还在被检验，它可能对，也可能错，而恰恰因为它还可能错，它才是一个真正的判断。

第四编到此结束。文字、数字、几何、微积分、概率，五件理念界的工具，各自从一个很小的需要出发，被封装，被使用，又在使用中长出现实界的契约、公司、保险和自我界的写给自己的页面、会算账的我、按概率办事的我。第五编要换到三界中的第三界，看那些直接作用于“我”的工具。

> 带走的话
>
> 同一部手机，抄答案的孩子什么也没发生，追问“为什么这样走”的孩子才开始发生。解题被封装之后，稀缺往上移到提问、判断、验证和造概念；这个判断还在被检验，它可能错，所以它才是判断。

> **小账** 封装了什么？——“解题”：以题为形态的数学能力。 逼出了什么？——（推断，尚在进行）提什么问题、结果有什么意义、凭什么信任、怎样造概念，也就是“你发生了哪些新数学”。

承接：理念界的工具讲完了。下一编进入自我界：从一面镜子开始，看人怎样一步步造出看见自己的办法。

第 五 编

自我界工具史

编 序

前面三编，我们分头看了现实界的工具、理念界的工具，现在走进离人最近的一界：自我界，也就是一个人怎样看见自己、写下自己、调理自己、解释自己、展示自己。这一界的工具有个特别之处：它们摆在每个人眼前，天天在用，却很少有人把它们当“工具”看。

本编五章，各讲一件工具。第22章讲镜子，它让人第一次站到自己外面看自己；第23章讲日记、书信与自传，它让内心可以被写下来、被回头对质；第24章讲修行法门，它是古人看内心的方法，也是“关于方法的方法”的样板；第25章讲心理学，它把内心变成可以研究的对象，也带来新的自我焦虑；第26章讲自拍与社交主页，它让每个人随身带着一个展示自己的舞台，也是本编代价最清楚的一件。

章	工具	封装的那份智慧	逼出的新东西
22	镜子	站到自己外面看自己	形象与本身的区分
23	日记书信自传	念头写下来就留得住	对记忆的求证
24	修行法门	怎样观察和调理自己	师承、次第、戒律
25	心理学	内心可观察可分类	描述与诊断的分寸
26	自拍与社交主页	自己决定怎样被看见	识别力，也有代价

本编站在“三界工具只会提升三界智慧”这一层论证上，取自我界这一支。要先说清楚：这里用的是史例，是佐证，不是证明。证明在第二编的封装与错位，以及下一编的结构论证。五章走的是同一个形状：封装、使用、碰撞、新问题、新智慧、让渡，最后各有一段诚实的边界。末尾还有一节“第五编小结”，收在第26章最后。

第 22 章

镜子：自画像与"我的样子"

◆ 一、早上七点的三秒钟

早上七点，卫生间。你睡眼惺忪地站到镜子前，三秒钟之内做了一连串事：看见头发翘了一撮，伸手压平；看见眼皮有点肿，决定今天不戴那副旧眼镜；看见下巴上冒了颗痘，皱了皱眉。然后你转身去刷牙，整件事没留下任何记忆。

这三秒钟里发生的事，其实相当古怪。你的眼睛长在脸上，它能看见整个世界，唯独看不见自己的脸。如果没有镜子、没有玻璃橱窗、没有手机黑屏上那层模糊的倒影，你这一辈子都没法亲眼看到自己长什么样。你对自己"样子"的全部了解，都来自某个中介：水面、磨亮的金属、玻璃、照片、别人的描述。

今天的人把这件事当成空气。可你想想，一个从没见过自己脸的人，他怎么想象"我"？这不是个玄问题。镜子这个再普通不过的东西，是自我界里最早的一批工具之一。本编要讲的就是：自我界也有工具，也遵守前面几编讲过的那条发生律。镜子是最好的起点。

先交代一下本书对"自我界"的用法。前面说过，本书把人的世界分成三界：现实界（我们面对的物与事）、理念界（概念、数学、理论）、自我界（我们怎样认识自己、安放自己）。书里反复用到一个叫 SDE 的三件套，大白话讲就是：S 是已经显露出来、站得住的东西（比如一面镜子里清清楚楚的一张脸）；D 是从这里到那里的差异与过程（比如从"不知道自己长相"到"每天对着镜子打理自己"）；E 是它们所处的土壤和意义场（比如你所在的社会怎样看待"好看"、怎样看待"一个人该有自己的样子"）。后面各章都在这个意义上用它。

◆ 二、一句话原理

镜子封装的，表面看是一句话：想知道自己什么样，不必靠别人转述，自己看一眼就行。可仔细想，它封装的不是一幅样子，而是一个位置：站到自己外面，把自己当作一个对象来看。样子，只是站在这个位置上看到的第一样东西。这份已经发生的智慧，被做进了一块涂了金属的玻璃里。

按前面几编立下的发生律，这样的封装一旦被大规模使用，就会与现实界、理念界相碰撞，碰撞会生出封装之外的新东西。造镜子的人不可能预先想到后面长出什么，因为那些东西当时还不存在，这就是前面说的错位。

◆ 三、镜子之前与玻璃镜：从水面到凡尔赛

最早的"镜子"是水面。希腊神话里那个著名的美少年，在水边看见自己的倒影，爱上了它，再也走不开（这个故事见于奥维德的《变形记》）。神话把水面里的自己写成一种危险的迷恋，这本身就说明：那时的人已经隐约知道，"看见自己"是件会改变人的事。

稍后出现了磨光的黑曜石和金属。中国的铜镜历史很长，汉代铜镜已相当成熟；埃及、两河流域也很早有磨光的金属镜。安纳托利亚的新石器时代遗址里出土过磨光的黑曜石片，有学者解释为镜子，用途至今有争议，这里不当定论。这些镜子有两个特点：贵，所以多半只属于少数人；照得不清楚，偏暗、偏黄、有变形。

在那样的世界里，一个人对自己长相的了解，很大程度上是社会性的：母亲说你像你父亲，邻居说你眼睛随你舅舅，同伴的神情告诉你自己讨不讨人喜欢。后来有位美国社会学家库利，在1902年的著作里把这个意思概括成一个很有名的说法，叫"镜中我"：我们是通过别人这面镜子看到自己的。他说的是人，不是玻璃，但这个比喻之所以成立，恰恰因为大家早就对玻璃镜子的作用心里有数。

真正改变局面的是平板玻璃镜。十六世纪前后，威尼斯的穆拉诺岛上，工匠掌握了在平整玻璃背后贴锡汞合金的办法，造出了又平又亮、照得相当真的镜子。威

尼斯人把这门手艺看得极紧，工匠甚至受到出境限制。镜子在很长一段时间里是奢侈品，一面好镜子价格高昂，至于具体能换多少东西，流传的说法来源不一，这里不引。

后来法国人不甘心，在财政大臣柯尔贝尔的推动下在1665年设立了玻璃制造的王家工场，也就是后来圣戈班的前身，努力把这项技术拿回自己手里。凡尔赛宫的镜厅在十七世纪八十年代前后建成，一整条长廊的墙面是镜子，那是权力与财富的展示，也是人类第一次大规模地、整面墙地"看见自己"。

再往后，工业化让镜子越来越便宜。到了今天，一面镜子几乎成为每个家庭的标配。从"少数人偶尔看见自己"到"所有人每天看见自己"，这一步走了几百年。

这里有一个我们要抓住的东西：镜子是现实界里的工具，它是物理的、光学的。但它的使用，几乎立刻就越出了现实界。

◆ 四、碰撞一：现实界的回答

镜子一用起来，人先跟现实碰了面，而且是不太客气的碰撞。

第一，人开始看到自己的变化。从前人们靠别人说"你老了"，现在每天早上自己对着镜子，看着第一根白发出现，看着眼角的纹路加深。时间这个抽象的东西，通过一块玻璃，变成了每天可以核对的现实。

第二，人开始发现自己的脸并不对称，并不像想象中那样。你肯定有过这样的体验：看手机里别人给你拍的照片，觉得"这不像我"。这里有个有趣的小事实。镜子里的你是左右反的，而别人看到的你、照片里的你是不反的。1977年，心理学家米塔等人做过实验，让人在自己的镜像和不反的照片之间选哪个更顺眼，结果是：本人偏爱镜像，朋友偏爱照片。道理很朴素：你天天看的是镜像，熟悉的就觉得是"我"。于是，一个有意思的问题出现了：到底哪一张脸才是"我的样子"？你每天见的那张，还是别人每天见的那张？

第三，人开始修饰。梳头、刮胡子、化妆、选衣服，这些行为不是镜子发明的，但有了镜子之后，它们变成了一种可以自己检查、自己调整的活动。对着镜子的"我"，成了一个可以被自己打磨的对象。

这三件事，都是镜子的制造者没有预料的。工匠要的是把光反射回来，他没有设计"让人发现自己在变老"，也没有设计"让人开始纠结哪张脸才是自己"。

◆ 五、碰撞二：理念界的回答

镜子同样越过了边界，撞进了理念界，也就是观念、概念的世界。

先看画。十五世纪的荷兰画家扬·凡·艾克，在1434年画了那幅著名的《阿尔诺芬尼夫妇像》，画面深处的墙上挂着一面凸面镜，镜子里映出房间的背影和门口的两个人影，有人认为其中之一是画家本人，学界并无定论。到了1500年，德国画家丢勒画下了一幅正面、直视观者、神态近乎庄严的自画像，这幅画的姿势，艺术史上常有人说很像当时画基督像的方式。后来的伦勃朗，一生画了数量惊人的自画像，从青年到老年，一张张看下去，像一部用颜料写成的日记。

自画像有一个技术前提：画家得有镜子。没有镜子，画家画不了自己。可是自画像一旦流行，它就不只是技术产物了，它开始表达一个新的观念：一个普通人（不是国王，不是圣徒）的脸，值得被画下来，值得被认真对待，值得被看作一个独一无二的"这个人"。有历史学家认为，镜子的普及与西方近代的个人意识的兴起有着千丝万缕的联系，法国历史学者梅尔舍-博内写过一部《镜子的历史》（1994年），谈到镜子怎样牵动人们对自我的感觉。这类说法不能说得太满，因为个人意识的兴起有太多原因，镜子只是其中一条线。但它至少说明：镜子这样一块玻璃，确实牵动了一些观念上的大变化。

再往深处看，还有一种"理念界的追问"。心理学家后来发明了"镜子测试"：悄悄在动物或婴儿脸上点一个记号，看它们照镜子时会不会去摸自己脸上的记号。1970年，心理学家盖洛普用黑猩猩做了这样的实验。人类的孩子大约在一岁半到两岁之间开始通过这个测试，也就是开始认出"镜子里的是我"，个体差异不小。换句话说，镜子不但让人看见自己，它还成了研究"自我意识是怎么一回事"的工具。这

又是一次越界：一个本来只是反光的玻璃，变成了理念界里一个被反复讨论的概念材料。

◆ 六、碰撞三：自我界里长出来的新问题

到这里，镜子对自我界的作用才真正显出来。

以前人想"我是谁"，参照的是别人的话、社会的位置。镜子来了之后，人有了一种新的位置：我可以站在自己外面，像看一个别人那样看自己。自我，第一次有了"被自己看见"的维度。这是一个新的 S，也就是一种稳定下来的、可以被说出来的自我形象：我的样子。

随之而来，一串前所未有的新问题冒了出来：

我的样子好看吗？这个问题以前不太可能被每天追问，因为没有每天对照的机会。现在有了，而且对照的标准不是自己定的，是社会的目光与流行的审美经由镜子"灌"进来的。于是有了对容貌的在意，也有了对容貌的焦虑。

我的样子是我吗？这是更深的一问。镜子里那张脸，让人既觉得亲切，又隐约觉得疏远。哲学家、精神分析学者们对此写过很多，法国的拉康关于"镜子阶段"的论述（1930年代提出，1949年有了广为人知的版本）就是一个著名的例子。我们不去展开它，只取其中最朴素的一点：从镜子里认出自己，是一个既得到了某个东西、又失去了某个东西的过程。得到了一个可以看见的"我"，也永远与这个被看见的"我"隔着一层玻璃。

我被别人看见时是什么样？这个问题使人多了一层"公共的自我意识"。有些心理学实验发现，仅仅是让人面对一面镜子，他们就会变得更留意自己的行为是否合乎标准，这类研究出自1970年代前后的"客体自我觉知"理论。镜子在这里起了一个很小又很真实的作用：它让人短暂地变成自己的旁观者。

注意这些问题的共同点：它们全都不在镜子的"封装"之内。镜子里封装的只是"光学成像"和"自己看自己一眼就行"这样的旧理解。碰撞之后长出来的问题，是造镜子的时候不存在的。这正是错位。

◆ 七、新的智慧长在哪里

如果这些只是新麻烦，那就只是一串账。可历史上的事情总是两面的：新问题逼出了新智慧。

镜子逼出了"自我形象与自我本身的区分"这门日常智慧。今天一个稍有阅历的人都知道：镜子里的样子，不是我的全部；我的样子，会因为光线、睡眠、心情而不同；别人对我的印象，不等于我的样子，更不等于我这个人。这种分寸，对一个从来没有过镜子的社会来说是不需要的，也是无从学起的。

镜子也逼出了"修饰的伦理"：打扮到什么程度是得体，什么程度是做作？在不同文化里，这条线划在不同地方，但每个文化都画了线。画线本身，就是自我界里新长出来的一块智慧。

还有一点，更微妙。人们学会了"把自己当作一个可以练习的对象"。演员对着镜子练表情，演讲的人对着镜子试姿态，学舞蹈的人在整面墙的镜前纠正动作。镜子成了练习的工具：自我从一个被动的"已经如此"，变成一个可以被调整的"可以更好"。这一步，是自我界里一次真正的升级。

◆ 八、让渡：镜子拿走了什么

按前面讲过的让渡律，凡是一个工具把某项能力封装起来，那项能力就会逐渐从人手里退场，这不是退化，而是让渡。

镜子让渡出去的，是"靠别人转述才知道自己长相"这件事。今天的人几乎不再需要旁人告诉他"你长什么样"。这项能力被镜子接过去了。腾出来的位置，被什么占住了？被"我怎样看待自己的样子""我想成为什么样子"这类问题占住了。这是一种让渡，不是失去。

◆ 九、这个论证的边界在哪里，与小结

先说镜子和"个人意识的兴起"。镜子只是其中一条线，不是原因，更不能说"没有镜子就没有现代个人"。本书只取最稳的一层：镜子的使用确实带来了此前没有的自我问题。

再说容貌焦虑。镜子并没有"制造"焦虑，焦虑有深厚的社会文化根源；镜子只是让对照变得每天可做。本章所说"升级"，是说人类整体上多了一门自我分寸的智慧，不是说每个人都因此活得更自在。有人对着镜子越看越不安，这是真实的，不能被"升级"两个字抹掉。第26章还会回到这一点。

再说和邻近说法的关系。库利的"镜中我"、拉康的"镜子阶段"，早就谈过镜子与自我。本章与它们相同的地方，是都把"看见自己"当成自我的一道关口；不同的地方，是本章不讲镜子怎样构成自我，只问它作为一件工具被大量使用之后，留下了哪些封装里没有的问题。

最后，这个判断在什么情况下是错的？有两条可以检验。一，如果史料或民族志显示，镜子尚未普及的社会里，人们对容貌的日常在意和比较并不比镜子普及之后更弱，那么"每天对照带来了新问题"就站不住。二，如果镜子进入一个原先没有镜子的群体之后，若干年内，他们的说法、行当、礼仪里没有出现任何与"看自己"相关的新东西，那么镜子逼出新智慧的说法，在这里就是错的。

回头看，镜子这块玻璃封装的是"我可以站到自己外面看自己"。人一用，就与现实界的衰老、光线、对称撞上，与理念界的肖像、个人、自我意识撞上，在自我界里生出"我的样子是我吗"这类封装之外的问题，也逼出了"形象与本身的区分""修饰的分寸""把自己当作可练习的对象"这些新智慧。**工具封装的是昨天的智慧，逼出来的是明天的智慧**，自我界里同样如此。

带走的话

镜子最大的本事不是让你看见脸，而是让你第一次能站到自己外面看自己；从那以后，“我的样子好看吗、那是我吗”就天天跟着你。

小账 封装了什么？——“想知道自己什么样，自己看一眼就行”，也就是站到自己外面、把自己当对象来看的那个位置。 逼出了什么？——自我形象与自我本身的区分、修饰的分寸，以及把自己当作可以练习的对象。

承接/预告：镜子给了人一个"看见的我"，可看得见的只是外表。内心的那个我，靠什么留住、靠什么看清？下一章讲日记、书信与自传。

第 23 章

日记、书信与自传：内心是怎样被写出来的

◆ 一、翻出十六岁的那个本子

搬家时，你从一个旧纸箱底下翻出一本带锁的日记本，锁早就锈了。你坐在一堆纸箱中间，翻开一页，是十六岁的自己写的："今天她没有理我，我觉得天都塌了。"你先笑了，笑到一半，又有点不是滋味。

这一页里有两件事同时发生。第一件，你认出这是你，字是你的字，那种夸张的口气也是你的。第二件，你同时觉得这个人很陌生，他为一件小事那样难过，而你现在已经想不起那个"她"的全名了。

一页纸，把一个已经不存在的自己，原封不动地留给了现在的你。这个能力，是自我界里一件了不起的工具：把内心写下来。本章讲这件工具的几种形态——日记、书信、自传——以及它们怎样改变了人认识内心的办法。

◆ 二、一句话原理

写下来这个动作，封装的是一份朴素的智慧：念头会飞走，放到纸上，它就留下来，还能回头再看。可纸上留下的，并不是念头的复本，而是一个以后能跟你对质的证人。它不是仓库，是证人席：你每回头看一次，它就作一次证，而作证的内容，常常跟你记得的不一样。

既然是封装，就会被拿去用，用了就会碰撞，生出封装当初没有的新东西。内心写作的历史，把这个过程摆得特别清楚。

◆ 三、先有书信，后有日记

人类很早就有"写给别人"的东西。书信是最古老的内心写作之一。罗马哲人塞涅卡留下一批写给友人卢基里乌斯的信，里面谈怎么面对死亡、怎么对待时间、怎

么过日子，读起来像是一个人在对朋友小声说话。基督教早期的使徒书信，是另一类影响深远的书信。中国从汉代的书牍到后世的家书，写信更是读书人修养自己、教育子弟的日常方式。晚清的曾国藩留下的家书和日记，很多人读过，里面既有对子弟的叮嘱，也有对自己毛病的反复检讨。

书信的特点是有一个读信的人。你写的时候心里装着他，所以你要选词、要斟酌、要把话说得让他明白。这个"心里装着对方"的过程，其实就是一种整理：话必须理出个头绪才寄得出去。许多人有过这样的体验：给朋友写一封长信，写着写着，自己想明白了原来不清楚的事。

而日记，是把这个"对方"拿掉，或者换成"将来的自己"。

◆ 四、日记：写给没有读者的读者

日记作为一种有意识的体裁，在不同地方分头出现过。日本平安时代，贵族女性留下了一批以日常感受为主的文字，其中有一部《蜻蛉日记》，是一位女性写自己婚姻中的苦恼与期待，被认为是日本日记文学的名篇。〔待核：作者通称"藤原道纲之母"，成书年代约十世纪后期〕更早一点，纪贯之以女性口吻写《土佐日记》，开头那句"男人写的日记，我这女人也试着写写"很出名。（约935年成书；这里是转述，不是原文）这说明当时的人已经意识到，"日记"是一种可以拿来表达感情的新形式，而且它还有一点"借别人嘴说自己话"的味道。

英国十七世纪的塞缪尔·佩皮斯，是西方日记史上的一个标志人物。他从1660年开始写了将近十年，既记伦敦大火、瘟疫这样的大事，也记自己和妻子吵架、对女仆动心、花钱买衣服时的得意。更有意思的是，他是用速记符号写的，在特别不想让人读懂的地方，还掺杂外语。这等于说，他在写给"自己看"的东西里，仍然设了一道防线。

这个细节很有意思。你以为你在写一本没有读者的日记，可是你会不由自主地想：万一被人看到怎么办？你可能写得更含蓄，用缩写，甚至只写一半。这说明，哪怕没有读者，写字这件事本身就带着"被读"的影子。这正是一次与理念界、也与别人的目光发生的碰撞。

◆ 五、自传：把一生讲成一个故事

再往前一步，是把整个人生写下来，这就是自传。

公元四世纪末，北非的奥古斯丁写下《忏悔录》，这是西方传统里最早也最有影响的"向内看"的作品之一，他一面向上帝讲述，一面把自己年少时的偷梨、迷茫和转变讲给读者听。十六世纪的法国人蒙田，发明了"随笔"这种写法，他说自己的书里画的就是自己（《随笔集》1580年初版）。十八世纪，卢梭的《忏悔录》以近乎冒犯的坦白出名；富兰克林的自传则更像一个人整理自己的成长经验，写给后代看。

到了二十世纪，法国学者勒热纳提出过一个说法，叫"自传契约"：写自传的人和读者之间有个默契，作者保证"这里的我就是真实的我"（1975年的说法）。可问题在于，这个"真实"并不简单。

◆ 六、碰撞：写下来的我，还是我吗

这里就进入本章的核心了。当日记、书信与自传被大量使用，它们同现实界和理念界发生了几次碰撞，每一次都生出了封装之外的新问题。

与现实界的碰撞：记忆靠不住。

你翻出十六岁的日记，发现里面写的和你记得的不一样。你记得那一年是你主动提的分手，日记里却写着"她不理我"。这时你面对一个难堪的现实：你的记忆并非原样保存，而是每次回想时重新拼出来的。英国心理学家巴特利特在1932年出版的《记忆》一书里，就通过让人反复复述故事的实验，说明了记忆的重构性质。日记像一台时间胶囊，它把"当时的我"定格，反过来让"现在的我"发现自己的回忆有多不可靠。这是写下来的人才会遇到的问题：以前没有对照物，你不会知道自己记错了。

与理念界的碰撞：我有一个"故事"。

当你把自己的一生写成一篇自传，你就不得不选取、排序、连接。哪些事重要，哪件事是转折，哪件事是原因。这个过程里，你构造了一个关于"我是怎样成为今天的我"的故事。后来的心理学者提出，人的身份认同在相当程度上就是这样的叙事，有人把它称为"叙事身份"，麦克亚当斯等人的研究最有名。这意味着，写作不只是记录自己，它也在塑造自己。同一段经历，你写成"我被抛弃了"，还是写成"我被迫独立了"，你后来成为的那个人是不同的。

法国思想家福柯在讲古代的"自我技术"时，特别提到过古人的笔记与书信：它们不只是记录，还是一种训练自己、塑造自己的办法。这个观点，把写作从"保存"提升成了"培养"。本章与福柯相同的地方，是都把写作看成对自己下功夫；不同的地方，是本章更关心写作被大量使用之后意外长出来的东西，比如记忆的对账、真实与表演的缝，而不是写作怎样被用来塑造主体。

与自我界本身的碰撞：真实与表演。

最后一个碰撞最微妙。当人知道自己的文字会被读，哪怕只是被将来的自己读，写作就多了一层表演。佩皮斯用速记，是怕被妻子读；很多人写日记，写着写着就写成了"想让将来的自己觉得我是个深刻的人"。真实的我与写出来的我之间，出现了一道缝。

这个缝，是日记这类工具刚被发明时不存在的。它在封装之外。

◆ 七、日常里的这套老把戏

说到这儿，别以为这是古人的事。你今天也在用这套工具，只是换了载体。

手机备忘录里那段没有发出去的长消息，是书信的现代形态：写的过程已经让你理清了自己想说什么。微信里的"文件传输助手"，被一些人当成日记本，夜里随手敲几行发给自己，第二天看，觉得很傻，又觉得很真。年终写的那篇总结，把一年的经历串起主线，是一篇微型自传。

这些场景有一个共同点：我们都在使用一件被封装好的"写下来"的智慧，而且不断碰到它的副作用——我们既依赖它，又隐约知道那个"写下来的我"不完全等于真实的我。

◆ 八、新的智慧长在哪里，与让渡

有了日记、书信和自传，人类在自我界里学到了几件从前学不到的事。

第一，学会了与过去的自己保持一个距离。能对十六岁的自己既觉得亲切又觉得可笑，并且不否认他是自己，这是一种相当成熟的自我关系。没有留下来的文字，人往往只能在"我从来就是这样"与"我完全变了"之间摇摆。

第二，学会了怀疑自己的记忆，并因此学会了求证。你开始知道，对重要的往事，最好有一份当时的记录，或者听听当时在场的另一个人怎么说。这是自我界里一次小小的"科学化"。

第三，学会了区分"感受"和"叙述"。你难过是一回事，你把难过写成什么样的话是另一回事。一个人能同时看到这两层，就有了更大的回旋余地。

第四，也许是最重要的：学会了"把内心当作可以整理的东西"。有了写作，内心不再只是一团涌动的情绪，它可以被排队、被分类、被修改。这是后来一切自我认识工具（包括下面要讲的修行法门和心理学）能够成立的前提。

按照让渡律，凡是工具接过去的能力，人手里就会退场。写下来之后，人对记忆的依赖减轻了。古代没有文字的社会里，口述传统的人要背诵大量的东西，包括家世、故事、仪式。文字来了之后，这种大量背诵的功夫逐渐让位。同样，自我界里的"记住自己的经历"也部分交给了日记。

这不是退化。被腾出来的那部分脑力、时间，被拿去做别的事：反思、比较、改进。只是，我们也得诚实：让渡了就是让渡了，有些东西确实不再那样练了。一个习惯把什么都写下来的人，未必记得住自己的经历；一个习惯拍下一切的人，未必看得清眼前的东西。第26章会回到这个话题。

◆ 九、这个论证的边界在哪里，与小结

第一，本章关于不同国家的日记、自传的例子，只是用来说明"写作这一类工具在各地被不断使用，并总在封装之外生出新问题"，并不是在构建一条整齐的"日记进化史"。各个文化的内心写作传统有自己的脉络，不能简单地放在一条线上。

第二，所谓"写作塑造自我"的说法，在心理学与文学研究里有各种强弱不同的版本，本章取的是比较保守的版本：写作参与了自我的形成，但不是唯一的力量。

第三，必须承认，有大量的人一生没有写过日记，也不识字，他们的内心同样丰富。这个论证并不是说"写作是认识内心的唯一路径"，而是说"写作是其中一种工具，它一旦被使用，就遵守这条发生律"。

第四，本章在什么情况下是错的？一，如果长期写日记的人，与不写的人相比，核对旧事、承认记忆出错的做法并不更多，那么"写下来逼出求证"就站不住。二，如果多数人重读旧日记时，并不觉得文字与自己的记忆有出入，那么"记忆靠不住"这一碰撞在日常里就没有这么普遍，要收窄成少数人的体验。

日记、书信与自传，把"念头会飞走，写下来就能留住"这份智慧封装成了一种习惯。人一使用，就与现实界（记忆的不可靠）、理念界（人生被讲成故事）、自我界（真实的我与写出来的我之间的缝）相碰撞，生出了封装之外的新问题；而这些问题又逼出了新的自我智慧：与过去的自己保持距离、对记忆保持怀疑、分清感受与叙述、把内心当作可整理的东西。写作封装了昨天的念头，使用逼出了今天的自知。

> **带走的话**
>
> 日记不是仓库，是证人席：你每翻一次，它就作一次证，而且常常和你的记忆对不上。正因为能对质，你才学会了和过去的自己隔开一点距离。

> **小账** 封装了什么？——“念头会飞走，写下来就能留住”这份朴素的智慧。 逼出了什么？——对记忆的求证、感受与叙述的区分、对真实的我与写出来的我之间那道缝的警觉，以及把内心当作可整理的东西。

承接/预告：写作整理的是内心浮上来的部分。那更深处的、不肯浮上来的部分，古人用什么办法去看？下一章讲修行法门。

第 24 章

修行法门：认识自己的古老工具

◆ 一、安静坐十分钟

找个晚上，关掉手机提示音，坐在椅子上，什么也不做，只看自己的呼吸，坐十分钟。很多人第一次这样做，都会被自己吓一跳：脑子里根本不安静。刚想着呼吸，立刻冒出来明天的会，接着是上午没回的消息，再接着是十年前某句难堪的话，然后忽然想起冰箱里还有半个西瓜。像一个菜市场，人声鼎沸，而你一直以为里面坐着一个安静的、清楚的"我"。

这个小小的发现，是很多古老修行传统共同的起点：人对自己内心的了解，原来远比自以为的少。本章讲的就是这一类工具。

先说明本章的立场。修行法门属于宗教与哲学传统，每一种都有自己的目的、自己的解释，信徒们认为那些目的远比"认识自己"更大。本书不评价任何传统的高低，也不对它们的教义下判断。本章只做一件事：从"工具"的角度，描述这些法门的结构，看它们怎样遵守本书讲的那条发生律。这是一个有限的视角，不是全部。

◆ 二、一句话原理

修行法门，在本书的用法里，是人类把"怎样观察和调理自己"的经验封装起来的一类工具：有方法、有次序、有师承、有共同体。每一套都是许多代人试出来的，封装的是某种已经发生的自我智慧。

既然是封装，就适用同一条规律：用起来，就会与现实界、理念界相碰撞，碰撞出封装之外的新东西。

◆ 三、几条各自生长的线

先快速看一下几条线，注意只是摆出结构，不排高低。

古希腊的"认识你自己"。 传说德尔斐神庙上刻着这句话，它成了西方思想的一个象征。苏格拉底传统把"省察"当作生活的重要部分。后来的斯多亚学派，把"每天晚上回顾自己一天"当成一种练习：今天哪里做得不对，哪里可以改进。塞涅卡提到过一位叫塞克斯提乌斯的人有睡前自问的习惯。〔待核：出处与细节〕

佛教的禅修与观照。 在佛教各个传统中，有大量以"观"为核心的修习：观呼吸，观身体，观念头的生灭。南传的内观、汉传的禅坐与念佛、藏传的各种次第修习，各有各的路径与解释。就结构而言，它们有共同点：先让注意力稳定下来，再用稳定的注意去看自己内心的流动。

儒家的修身与省察。 《论语》里记着曾子"每天多次反省自己"的话，后世把"反躬自省""慎独"当作功夫。明清时期还出现了一种叫"功过格"的东西，像账本一样，把每天做的好事、过错一条条记下来，打分计数。〔待核：功过格的兴起时间、代表文本与具体做法〕这个做法有点意思：它的结构，跟复式记账（第27章会讲）有些相像，把"自己的德行"做成了可核算的账。

基督教的灵修。 从早期沙漠教父，到后来的修道院传统，再到十六世纪依纳爵的《神操》，都有大量关于"省察""默想""祷告"的次第。依纳爵传统里有一种每日的"意识省察"，大意是每天固定时间回顾一天里的感受和动向。〔待核：术语译名与具体步骤〕

伊斯兰苏非传统的记念。 苏非传统里有反复称念、静坐与集体修习的形式，用以调理内心、贴近信仰的核心。〔待核：术语与做法因教团而异，本书不细写〕

这几条线各自独立地生长，语言、目的、教义大不相同，但从工具的角度看，它们有几处共通的东西：

第一，都有一套"怎么做"的方法，也就是可以反复使用的程序。 第二，都有一个"怎么做才算对"的反馈机制，通常是师父、同修，或者经典。 第三，都有一套"出了问题怎么办"的预案。

最后这一点，是本章要抓住的关键。

这类法门，也早有人当作"对自己下功夫的技术"来研究。法国学者阿多把古代哲学看作一套"精神修炼"，福柯讲过"自我技术"。本章与他们相近：都不把这些法门只当教义看。不同在于，他们追问这些修炼造就怎样的主体，本章只盯一个结构：用起来之后，哪些问题是方法自己制造的，传统又怎样给方法加上护栏。

◆ 四、碰撞：用起来之后，出了什么新问题

任何一套方法，一旦有人认真用了，就会开始碰到现实。修行人碰到的现实，首先是身体和日常。

与现实界的碰撞。　坐久了腿麻，呼吸乱了心也乱；修行不能脱离饮食起居，作息和环境也会影响状态；人要工作、要养家、要处理关系，修行里的"静"要拿去对付生活里的"乱"。一套再好的方法，进入日常就会碰到摩擦。传统的回应，是发展出各种细节，比如坐姿、作息、戒律、共同生活的规矩。这些细节本身，就是被碰撞逼出来的新东西。

与理念界的碰撞。　修行里常常会出现一些异常的体验：觉得身体变轻，觉得光明充满，觉得一切合一，觉得自己看穿了什么。如何解释这些体验？它们是修行进展的标志，还是需要警惕的偏差？于是需要概念、需要教义、需要判断标准。许多传统的经典，都留有关于"体验"和"歧路"的大量讨论。禅宗的文献里有关于"魔境"的提醒，大意是修行中出现的境象不可当真、不可执着。〔待核：术语出处与说法，需请相关研究者核〕基督教灵修传统里，也有关于"辨别灵"的长期讨论。〔待核〕

这就是一个很有意思的现象：方法越是有效，越是会产生新的、方法本身没有预料的问题——把体验当成目的，把自己的感觉当成真理，把方法当成偶像，甚至以修行者的身份自我炫耀。这些问题，在方法被发明的时候并不存在，它们是"人一用，就出现了"的。

与自我界的碰撞。 更深的一层是，修行者发现，"那个观察内心的我"本身也是内心的一部分。我在观察念头，可是"我在观察"也是一个念头。这个转折，被不同传统以不同方式处理，有的发展出关于"无我"的教导，有的发展出关于"灵魂与

神"的讲法，有的发展出关于"心性"的讲法。这里我们不比较它们，只想指出：这个新问题（观察者怎么办）是在使用工具的过程中浮现的，不是一开始就有的。

◆ 五、新智慧：传统自己长出来的"护栏"

使用中出现的问题，催生了一层更高的智慧：关于"如何使用这套方法"的智慧。一套法门成熟的标志，不是方法多精，而是它为出错预备了多少办法。好比学车时教练脚边那套副刹车：它不是驾驶技术的一部分，却是"学驾驶"这件事能成立的条件。

于是有了师承。一个人单独修，容易陷入自我欺骗；有一位经验丰富的师父，可以为他"印证"，也就是核对他的体验是否在正路上。〔待核："印证""印可"等用语在不同传统中的含义有别〕

于是有了次第和阶梯。不要一上来就追求高处，而要一步步来，每一步有每一步的功夫和检验。

于是有了戒律和共同体。戒律是对生活的约束，共同体是外在的互相提醒。它们使得个人的修行有一个稳定的外部环境。

于是有了经典和注释。每一代修行人把自己的经验写下来，后人把这些经验当作"路标"，告诉后来者哪里有坑。

这些东西，从工具的角度看，都是"关于工具的工具"。方法本身已经封装了一层智慧，使用过程中出现的新问题，又被封装成第二层：怎么选老师，怎么判断体验，怎么防止走偏。这和本书反复讲的二阶碰撞，是同一个形状。

◆ 六、日常里的现代版：正念

要看清这种"再封装"，一个最近的例子是我们现在常听说的"正念"。

二十世纪七十年代末，美国的卡巴金在医学院开设了一个以"正念"为核心的减压课程，把来自佛教修习的一些做法，改造成了不依赖宗教信仰、可在医院和社区

中教授的八周课程。（1979年，马萨诸塞大学医学院，课程简称MBSR）这个课程后来流传很广，成了许多人日常里的"每天十分钟"。

从工具的角度看，这是一次再封装：把一套深嵌在传统里的方法，剥离出它的核心程序，变成一个标准化的、可复制的课程。这样做的好处是：门槛低了，更多人可以接触。

但同时也出现了新问题。有人把"正念"当成万能的情绪处理术，一用不灵就失望；有人把它当成提高效率的工具，而不是原传统里完整的生活方式；有的传统内部的人提出，剥离背景之后，有些东西丢了。这些争论至今还在继续。关于它的效果，已有不少研究，但结论的强弱、适用的人群都还在被讨论。本书不做任何医学判断，也不替任何人推荐，只想说明：连这个最新的再封装，也一样遵守这条律。

◆ 七、让渡

让渡律在这里怎么看?

当一套方法被很好地封装、被广泛教授，一个初学者就不必再从零开始摸索。从前，一个想要认识自己内心的人，可能要花几年时间自己试错；现在，他可以跟一位老师学习已经被整理过的路径。"从零摸索"这种能力，部分地让给了传统。这是让渡：初学者的探索少了，但被腾出来的，是用在更深处的精力。

当然，让渡也有代价。有人太依赖现成的步骤，反而不去体会自己的实际情况。传统里常说"不可执着于方法"，大概就是对这种代价的提醒。

◆ 八、这个论证的边界在哪里

边界很多，要老实交代。

第一，我只是在"结构"层面谈论这些传统，并没有进入它们的内部。对一个真正的修行者来说，他修的不是"认识自己的工具"，而是一条通向他所相信的终极东西的路。把它们放在"工具"的名下，是本书的视角，不是这些传统自己的说法，也不应该取代它们自己的说法。

第二，我不比较、也不评判哪个传统更高明、更有效。本章提到的所有传统都被平等对待。如果读者读出了谁比谁强，那是误读。

第三，本章关于"魔境""辨别灵""功过格"这些细节，大多来自二手的概述，是否精确，请懂行的读者指正。

第四，"体验"和"效果"的判断，不属于本书。本书不做任何关于修行对健康的断言，也不对读者提出任何修行建议。

第五，就论证本身而言，本章的功用有限：它只是为"自我界也有工具"提供一个例子。它并不证明这些工具"最终能够穷尽自我"，恰恰相反，它的启示是：每一套方法用到深处，都会遇到封装之外的新问题。

第六，本章在什么情况下是错的？如果某个传统在许多代的实践中，对体验的辨别、对走偏的防范、对教学者资格没有任何形式的安排，本章对它的说法就不成立；如果一种新方法大规模推广多年，也始终没有出现过任何关于"怎么用它"的指南、分级或警示，同样是反例。

◆ 九、小结

修行法门是人类封装"怎样观察和调理自己"的古老工具。它们在使用中与身体和日常（现实界）、与概念和教义（理念界）、与"观察者也是被观察者"这一难题（自我界）相碰撞，生出了封装之外的新问题；这些新问题催生了师承、次第、戒律、经典等一层更高的智慧，近代又被再封装成面向大众的课程，并带出新一轮的问题。法门封装了昨天的功夫，使用逼出了明天的护栏。

> **带走的话**
>
> 一套方法越有效，越会长出方法自己没想到的毛病。所以老传统成熟的标志，不是招数有多精，而是给出错预备了多少副刹车。

> **小账** 封装了什么？——怎样观察和调理自己的经验：有方法、有次序、有师承、有共同体。逼出了什么？——师承、次第、戒律、经典这些“关于工具的工具”，还有“观察者本身也是被观察者”这道难题。

承接/预告：古人面对内心的不透明，靠师父，靠经典，靠时间。到了近代，有一批人想用观察、实验和统计的办法来对付它，这就是心理学。下一章讲它，以及它带来的新的自我焦虑。

第 25 章

心理学：潜意识，以及新的自我焦虑

◆ 一、一句"我这是内耗"

朋友聚会，有人叹一口气："最近我太内耗了。"另一个接话："我这是典型的回避型，一碰到冲突就想跑。"第三个说："我妈从小就那样，我这是原生家庭的问题。"桌上没有一个人是心理学专业的，但每个人都用着一套相当精细的词汇来谈论自己。

这在一百多年前是不可想象的。那时候的人说"我心情不好"，说"我脾气急"，说"我运气差"，已经算很会表达了。今天一个普通人，可以顺口说出"潜意识""情绪""压抑""投射""舒适圈""自我认同"这类词。这些词不是天生就在语言里的，它们是一个学科一点点推到日常里来的。

这个学科就是心理学。它是自我界里最近的一件大工具，也是最典型的一件：它把"怎样理解人的内心"封装成了一套概念、实验、测量和理论，并且全社会都在用。本章讲它怎样遵守同一条发生律，以及它带来的一种新的、很真实的自我焦虑。

本章有一条硬规矩：我不做任何医学断言，不诊断任何人，也不评价任何具体的测评产品。本章只谈"工具的结构"，谈词汇和观念怎样改变了人看自己的方式。

◆ 二、一句话原理

心理学封装的是一份智慧：人的内心可以被观察、被分类、被研究，不必只靠内省和传统。这是一次很大的封装，它把此前散落在哲学、宗教、文学、医学里的零碎认识，变成了一个有方法、有共同体、有积累的学科。

◆ 三、简史：几个关键的拐点

下面只挑几个人们比较有把握的节点。

1879年，德国的冯特在莱比锡建立了一般被视为第一个心理学实验室的机构。（学界对"第一个"有争议）从此，"心理"开始成为可以用仪器、用控制条件来研究的东西。

1890年，美国的威廉·詹姆斯出版了两卷本《心理学原理》，把意识流动的特性写得非常生动，也讨论了"自我"由哪几部分构成。这本书影响极大，有不少我们现在习以为常的说法，都能追溯到那里。

十九、二十世纪之交，弗洛伊德提出了关于潜意识的一整套理论，他的《梦的解析》通常被认为出版于1899年底（扉页标注1900年）。这套理论的冲击力，在于它宣告：人并不完全了解自己的动机，许多驱动我们的东西，是我们自己看不见的。

二十世纪初，法国的比奈为学校设计了测量儿童智力水平的量表，1905年前后。后来心理测量的方法一路扩展，出现了各种能力测验、人格问卷。

二十世纪的上半叶，行为主义强调只研究可观察的行为；之后又有认知心理学的兴起，把人心比作信息加工的过程；还有人本主义、发展心理学、社会心理学、积极心理学，等等。学派很多，彼此还常有分歧。本书不介入这些争论，只取一个结构性的事实：心理学作为工具，产出了大量关于内心的概念和测量方法，并且很快流入了社会。

◆ 四、碰撞：这套工具用到了人身上

心理学最特别的地方是，它研究的对象会读它。一块岩石不会读地质学，一颗行星不会读天文学，可人会读心理学。读完之后，人对自己的看法就变了，做事的方式也变了。

加拿大哲学家哈金研究过这个现象，他的概括大意是：人的分类与被分类的人，会互相影响，像一个回路，分类改变了人，改变了的人又迫使分类被修订。他称之为"循环效应"（1990年代的论述）。这个观察，把心理学的使用与"碰撞"之间的联系说得非常清楚。

按三界来看。

与现实界的碰撞：理论遇上了现实，发现并不总是对得上。

心理学要求自己是一门实证的科学，所以它的理论要经受重复检验。二十一世纪第二个十年，心理学界经历了一场被称为"可重复性危机"的反思：一批规模较大的重复研究发现，不少已发表的经典结果在重新检验时，没有得到同样的结论。（2015年开放科学合作组织的大规模重复研究，具体比例这里不写）这件事听起来像丑闻，但从发生律的角度看，它恰恰是工具与现实碰撞的结果：一旦这套方法被大规模使用，它的内在问题就浮现出来，进而逼出了新的做法，比如预先登记研究计划、公开数据、要求更大的样本。这些新做法，是在原来的方法里没有的。

与理念界的碰撞：新概念改写了旧概念。

"潜意识"这个概念进入日常语言后，人们开始用"我潜意识里是不是想……"来解释自己的行为。一个口误，就被说成是"弗洛伊德式的"。一个人说自己的"童年阴影"，说自己的"依恋类型"，其实是在用学术概念给自己的经历重新排序。观念的世界，被这种用法反复改写。

与自我界的碰撞：我开始用一套词汇描述自己，也被这套词汇描述。

这是本章的重点，因为这里正是新的自我焦虑的来源。

◆ 五、新的自我焦虑，来自哪里

先说一个必须说清的前提：焦虑本来就有，不是心理学造出来的。这里说的"新的自我焦虑"，是指由这套工具的使用方式带来的、此前不存在的那些具体样态。

其一，"我是不是有问题"的持续自我检视。 以前，一个人心情低落，可能说一句"最近不顺"，过几天就翻篇了。现在，各种心理知识唾手可得，人会下意识地对照：这是不是某某倾向？我是不是符合某种描述？每个人都看过那种"十个迹象，看你是不是……"的文章，看着看着，觉得好像每一条都沾点边。这并不是说那些知识是错的，而是说，知识的普及改变了人观察自己的方式：从"我今天不太舒服"变成了"我是什么类型的人，有什么问题"。需要强调：这里只描述一种常见的体验，我不对任何人的状态做判断，真有困扰的人，应当找专业人士。

其二，标签的诱惑与束缚。 给自己贴一个标签，有时很解脱："原来我不是懒，我只是这个类型。"但标签也可能变成牢笼："我就是这样的人，改不了。"各种性格测试、星座、类型学在社交场合里的流行，让这一点变得很日常：自我介绍时人们习惯先报上自己的"类型"。标签让沟通更快，但也让人少了一点探索自己的好奇。细看，变的不是知识的多少，而是用途：一个词原本是拿来描述自己的，用着用着，成了拿来量自己的尺。"我是回避型"说出口的一刻，描述就成了判决的雏形。这里不评价任何具体的测评产品，只谈这类现象的结构。

其三，"我看不见的我"带来的不安。 潜意识观念最深的冲击是：我的动机我自己并不全知。这给人带来一种独特的不安：我无法完全信任自己的内省。后来的社会心理学研究也发现，人对自己的判断有诸多盲点，比如倾向于认为自己比别人更少偏见（普罗宁等人2002年称之为"偏见盲点"）。说得轻点，这是谦逊的来源；说得重点，这是不安的来源。

其四，对"正常"的敏感。 心理测量和统计，给出了"平均""常模"这样的概念。人们很容易拿这些来衡量自己：我是不是偏离了大多数？这种"和统计分布比较"的习惯，是前所未有的一种自我对照。

这四种焦虑，在心理学被大规模社会使用之前，要么不存在，要么没有这么普遍。它们都在原来的封装之外，这就是错位的又一次表演。

◆ 六、新的智慧：从"贴标签"到"留余地"

同样的，这套工具也逼出了新的智慧，而且这些智慧，恰恰是针对上面几种焦虑的。

第一种智慧：区分描述与诊断。　　越来越多的普通人懂得，"我最近有点焦虑"和"我有焦虑症"是两回事，前者是一种感受，后者是一个专业判断，要由受过训练的人在恰当的情境下做出。这条界线，是心理学普及过程中逐渐被社会学会的。

第二种智慧：承认盲区，并借助他人。　心理学家卢夫特和英厄姆在1955年提出过一个"乔哈里之窗"的模型，把关于自己的信息分成四格：我知道别人也知道的、我知道别人不知道的、我不知道别人知道的、大家都不知道的。这个小模型的价值，是教会人"别人眼中的我"是一份有用的资料。这和第22章镜子里的"别人是面镜子"，在更高的一层上衔接了起来。

第三种智慧：对自己的叙述保持一点怀疑。　既然知道记忆会重构，知道内省有盲区，人就更愿意把自己的解释当作"一种说法"，而不是"定论"。这不是犹豫不决，是一种成熟。

第四种智慧：把心理知识当作工具，而非判决。　好的使用者，是把"依恋""防御"这些概念当成放大镜，看看事情有没有另一种看法，而不是拿来给自己或别人定罪。这种使用的分寸，也是需要学的。

◆ 七、二阶碰撞：心理学对付自己制造的问题

这里有个形状，前面见过。心理学在普及中制造了"标签焦虑""过度自我检视"这类问题，而心理学自己，又发展出对这些问题的回应。比如，有的研究者专门研究"自我同情"，研究人怎样不对自己过分苛刻；有的研究者研究怎样在大众中传播更准确的心理知识；有的从业者强调不要把日常的情绪起伏都当成病症。

这就是前面讲的"二阶碰撞"：工具生出的新问题，反过来被同一个领域用更高一阶的办法去处理。处理的结果也不会终结，处理本身又会生出新的东西。

◆ 八、让渡

让渡律在这里表现为：过去需要靠长辈的经验、宗教的指引、自己的摸索才能解决的一部分"怎样理解自己"的问题，现在有了一套可学可查的概念框架。这使得很多人少走弯路。比如，一位新手父母可以直接读到关于婴儿睡眠和情绪的研究，不必全凭祖辈的传言。那种"全凭经验悟"的能力，部分让渡出去了。

腾出来的位置，被"如何选择和使用这些知识"占住了。一个人在面对海量的心理知识时，要学会筛选、辨别、对照，这本身就是一种新的能力。

◆ 九、这个论证的边界在哪里，与小结

第一，心理学是一个庞大且内部分歧不小的学科，我只取了"工具封装—使用—新问题"这一个结构切面，并不代表整个学科。

第二，我不对任何具体理论的真伪做裁判。弗洛伊德的理论今天仍有很大争议，也仍有不少人认为它有价值；本书不站队。我关心的是它作为一件"工具"进入社会之后的事。

第三，我不做医学断言。书中关于焦虑的描述，都是日常层面的观察，不是临床判断。如果读者自己或身边的人有持续的痛苦，请找有资质的专业人员，而不是用本书来对照。

第四，"新的自我焦虑"是否整体上比过去更重，我没有证据说。我说的只是：某些具体形态是新的。至于总量，现有研究不足以回答。

第五，本章并不证明"心理学会终结对心灵的认识"，恰恰相反，它的启示是：每一次对内心的封装，都带出了更多尚待回答的问题。

第六，邻近的说法要交代。哈金的"循环效应"，以及罗斯（Nikolas Rose）讲心理学科怎样参与造就现代人的自我理解，都承认心理学的词汇会回到人身上改变人。本章与他们相同；不同在于，本章关心由此长出的具体的新自我问题，以及同一门学科对付它们的二阶办法。

第七，本章在什么情况下是错的？一，如果比较不同年代的日记、信件、访谈，发现人们描述自己的词汇和自我检视的方式，在心理学普及前后没有可察觉的变化，那么"词汇进入日常改变了人看自己"就不成立。二，如果"新的自我焦虑"的四种样态，在心理学普及之前的材料里同样常见，它们就不是新的。三，如果接触较多心理知识的人群，在区分描述与诊断、给自己的叙述留余地这些做法上，并不比接触较少的人群更多，那么"焦虑逼出新智慧"在心理学这里就站不住。

心理学把"内心可以被观察、被分类、被测量"这份智慧封装成了一个学科。它被大规模使用后，与现实界的重复检验、与理念界的观念更替、与自我界的自我描述相碰撞，生出了封装之外的新问题——持续自我检视、标签的诱惑、对盲区的不安、对"正常"的敏感；也逼出了新的智慧——区分描述与诊断、承认盲区借助他人、对自己的叙述留余地、把知识当工具而非判决。这是封装与逼出的又一次演示，也是一次二阶碰撞的演示。

> **带走的话**
>
> 心理学研究的对象会读它，所以一个词进了日常，就会回头改写人怎么看自己。同一个词，当放大镜用是帮你，当判决书用就是害你。

> **小账** 封装了什么？——“内心可以被观察、被分类、被测量，不必只靠内省”这份智慧。 逼出了什么？——持续自我检视、标签的诱惑、盲区带来的不安，也逼出了区分描述与诊断、给自己的叙述留余地。

承接/预告：心理学让人用概念看自己。而在今天，更多人看自己的时候，用的是手机。下一章讲自拍与社交主页，这是自我界最新的一件大工具，也是本编里代价最清楚的一件。

第 26 章

自拍与社交主页：被展示的我

◆ 一、三十张里留一张

打开你的手机相册，找到最近一次出游。同一个角度、同一个笑容，你拍了三十张，最后留下一张，修了修，加了个滤镜，配上一句不太用力的文案，然后发到朋友圈。发出去之后，你每隔十分钟看一眼手机，数一数有几个人点了赞。有一个你很在意的人没有点，你告诉自己没事，然后又看了一遍。

另一件事发生在第二天：你翻到别人发的照片，里面每个人都容光焕发、风景如画，你一边点赞，一边隐约觉得自己的周末过得很普通。

这一连串小动作，是自我界里最新的一件大工具的日常。镜子让人看见自己，日记让人写下自己，修行与心理学让人分析自己，而自拍和社交主页做的是另一件事：把自己摆出来，给别人看，并且随时知道别人看到了什么、说了什么。

本章要讲两件事，而且要同时讲，不能偏废。一件是：这件工具和前面几件一样，遵守同一条发生律，在封装之外生出了新的自我智慧。另一件是：它带来的新的自我焦虑是真实的代价，不是为了论证而顺手提一句的小瑕疵。两件事并存，哪一件也不能抹掉另一件。

◆ 二、一句话原理

自拍与社交主页封装的是一份智慧：一个人可以自己决定怎样被看见，并且可以让很多人同时看见。它不是镜子的加强版。镜子里只有你一个人，主页里满是观众；镜子给你一个样子，主页给你一个随时更新的回应。这份智慧在历史上是分散的：画师为权贵画像、影楼里的正装合影、书信里的自我介绍。现在，它被封装进一部手机里，每个人随身带着。

◆ 三、简史：从自画像到前置镜头

摄影术诞生不久，就有人给自己拍照。1839年，美国人罗伯特·科尼利厄斯在费城拍下了一张自己的银版照片，常被说成最早的"自拍"之一。

此后一百多年，拍自己都是件麻烦事：要摆姿势，要借别人的手，要冲洗。拍立得、自助照相亭、家用相机一步步降低了成本，但"看着自己拍"仍然缺位。

转折发生在数码时代。相机背面有了可翻转的屏幕，手机有了朝向自己的镜头。2010年前后，手机普遍配上了朝向自己的前置摄像头，最初主要为视频通话设计。"自拍"这个词，据说2002年前后就出现在网络论坛上，2013年被牛津词典选为年度词汇。

同一个时期，社交平台兴起。脸书2004年创办，国内的微信在2012年前后上线了"朋友圈"。个人的展示空间，从相册、日记、书信，一下子搬到了一个"不断更新的、朋友们都能看到的页面"上。

注意这里出现了一个新结构：展示、反馈、再展示，三者在几分钟之内就能转一圈。以前，你的"形象"在别人心里慢慢形成，可能要几年；现在，一条动态发出去，几分钟内就有一批人的回应。

◆ 四、碰撞一与碰撞二：现实界与理念界

先看现实界。

光线、角度、美颜。　手机拍出的照片，在算法的帮助下，往往比肉眼看到的更"好看"。美颜可以磨皮、瘦脸、放大眼睛，滤镜可以给任何场景加上同一种氛围。于是，你每天看到的"自己"多了一个现实界里的新来源：被处理过的脸。一个自然的后果是：当你再回到镜子面前，会觉得镜子里那张脸"没有照片好看"。这是现实界里一个朴素的碰撞：真实的脸和被处理的脸，在同一个人的头脑里开始争夺"哪个是我"的位置。

时间与注意力的消耗。　构图、挑选、修图、写文案、等反馈，这些事情要花时间，一天里并不少。工具省下来的时间，被工具带来的新活动花掉了，这是前面洗衣机和家务故事（考恩的研究）在自我界的重演。

再看理念界，也就是观念层面。

社会学家戈夫曼在二十世纪五十年代提出过"拟剧论"，把日常生活比作舞台：人在"前台"演给观众看，回到"后台"才放松。（《日常生活中的自我呈现》，1956年初版，1959年有通行版本）他当时想的是面对面的社会交往。今天的社交主页，让这个比喻变成了字面意思：你真的有一个前台，而且是永久陈列的前台。

另一位社会心理学家费斯廷格，1954年提出"社会比较"理论，说人需要通过和他人比较来评估自己的能力和观点。在以前，人比较的对象是邻居、同事、同学，范围有限；在社交主页上，比较的对象是几百上千人"最好的一面"。比较的范围扩大了，比较的内容却被高度筛选过，这是一个人类没有处理过的新局面。

"人设"这个词也很说明问题。它原本是影视与创作行业的词，指一个角色的设定；今天，普通人也在谈论自己或别人的"人设"，说谁的人设崩塌了。这是理念界的新概念：人开始把"自己的形象"视为可以设计、维护、翻车的东西。

◆ 五、碰撞三：自我界，也是代价所在

现在说最核心也最不舒服的部分。

新的自我焦虑，是真实的。 这一点我想说得郑重一些：

其一，等待反馈的焦虑。　发出去的东西有没有人回应，成了一件可以每分钟查验的事。心里的小剧场很具体：谁看了没点赞？谁点赞很快？这类细小的情绪消耗，很多人都有，只是不好意思说。

其二，比较的焦虑。　如上所说，别人展示的是挑选过的瞬间，自己的是完整的日子，把两者对照，输的几乎总是自己。很多人明知这一点，依然忍不住。

其三，自我的分裂感。 前台的我越精致，就越感到后台的我"不够格"。有人因此不敢发真实的东西，有人发了又删，还有人把删除当成一种焦虑的出口。

其四，自我的外包。　当"我是谁"越来越多地由"别人怎么回应我的展示"来回答，一个人的自我价值就悄悄交给了别人的点击。

关于这类现象与人们的心理健康之间究竟有多大关系，学界有大量研究，但结论并不一致：有的研究发现某些使用方式与不良感受相关，有的发现关联很弱，还有的发现不同的人、不同的使用方式，结果差别很大。〔待核：该领域研究结论的整体概括〕本书不下结论，也不做任何医学断言。我愿意说的是：这些焦虑的体验是真实存在的，许多人在日常里确实感受到了，它们不应该因为"这是一次智慧升级"而被轻描淡写。

要强调的是，代价和升级是并存的，不是互相抵消的。　说一个工具带来了智慧升级，不等于说它是"净好处"，更不等于说每个具体的人都因此过得更好。有些人在这套工具里变得更焦虑、更疲惫，而这些感受不会因为"整体上社会学会了新智慧"就变得不重要。这是本书的立场，也是本章必须交代的诚实。

◆ 六、并存：同一时期，新的智慧也在长

与此同时，另一些东西也在长，而且长得相当快。

第一，对"展示"的识别力。 十几年前，人们看到别人朋友圈里的精致生活，容易当真；今天，一个中学生也会用"买家秀与卖家秀""摆拍""滤镜"这类词来拆解一张照片。"看到的不等于真实的"，成了一个相当普遍的常识。这是一种媒介素养，它并不是谁教出来的，而是几亿人在使用中集体摸出来的。

第二，对"分层展示"的分寸。 许多人学会了对不同的人展示不同的东西：有人只对家人可见，有人只设三天可见，有人开小号。表面上这是功能设置，背后是一种相当细的自我管理：我知道我在对谁说话，我愿意让谁看见哪一层。古人讲"见什么人说什么话"，在这里被推进成了一套显式的操作。

第三，对"被注视"的反思。 "我为什么发这条？""我是发给谁看的？""我希望得到什么回应？"这类自问，越来越成为一种常见的自我对话。能问这些问题，本身就是自我界里的一次升级——把自己的展示动机当作观察对象。

第四，对"不展示"的珍视。　展示成了常态，"不发""不拍""只给自己留着"就成了有意识的选择。很多人开始区分：哪些经历是用来分享的，哪些是用来经历的。这是封装之外长出来的新自我价值：留白。

第五，更为多样的自我表达。　对许多原本缺少发声渠道的人来说，一个主页是一个可以被看见的窗口：身体有残障的人，居住在偏远地方的人，爱好小众的人，有了找到同类的方式。这是实实在在的好处，也是自我界里的新生。

◆ 七、让渡，与下一轮已经在路上

让渡律在这里要看两面。

被让渡出去的，是一些原先靠记忆和面对面完成的事：记住朋友的生日，靠平台提醒；留住旅途的片段，靠相册；维持远方的友谊，靠偶尔点个赞。这些事交给工具之后，人手里的能力确实弱了一些：很多人记不住几个电话号码，也记不清自己去过的地方的细节，因为"反正有照片"。这是一种让渡，不是退化，但我们也得老实说，它是有成本的：拍下来的东西，不一定看进去了。

被腾出来的，是跨距离联系的可能和更多自我表达的机会。这既是好处，也是新问题的来源：腾出来的时间，很容易又被工具本身占住。

自拍和主页之后，新的封装已经在出现：软件可以替你生成头像，润色文案，甚至模仿你的口吻回复消息。这是对"展示自己"这份智慧的又一次封装。按发生律可以预期，它一旦被大量使用，会生出一批今天还说不清的新问题和新智慧，比如"哪些是我说的""我的口吻是不是还属于我"。这里不做具体预言，也不评价任何产品，只指出：这一轮不会是终点，正如前面每一轮都不是终点。

◆ 八、这个论证的边界在哪里，与小结

第一，我不是在说"社交媒体总体上是好的"，也不是在说"它总体上是坏的"。本章的论证只有一个窄的目标：证明它遵守同一条发生律。这一点和"它对人好不好"是两个问题。

第二，"新的智慧"是在文明层面和部分个人层面发生的，并不保证落在每个人身上。有些人在这套工具里学会了分寸，有些人没有，有些人被伤害得很深。第六编会专门讲个体层与文明层的区别：文明层面，新问题催生新智慧几乎是必然的；个体层面，则取决于这个人是否真正与问题碰撞过、思考过。

第三，我所举的"识别力""分层展示"等例子，是日常观察，不是严格的调查数据。它们是否普遍，我没有统计证明。

第四，关于未成年人、关于心理困扰的议题，本书不做讨论。那需要专门的研究与专业的建议。

第五，本章在什么情况下是错的？一，如果比较同一批人开始使用社交主页前后，他们对摆拍、滤镜、分层可见的识别和使用并无增长，那么"识别力是集体摸出来的新智慧"就错了。二，如果上面四种焦虑在没有这类工具的人群里同样常见、程度相当，它们就不是工具带来的。三，如果十年之后，仍然没有出现任何缓冲这些代价的新做法，那么"逼出新智慧"在这件工具上就要打个问号。

自拍与社交主页封装的是"我可以决定怎样被看见，并让很多人同时看见"。它被大规模使用后，与现实界（美颜后的脸、时间的消耗）、理念界（拟剧、比较、人设）、自我界（等待反馈、比较、分裂感、自我外包）相碰撞。代价是真实的：新的自我焦虑不是修辞，也不会被"升级"两个字抹去。同一时期，新的智慧也在长：对展示的识别力、分层展示的分寸、对自己动机的反思、对"不展示"的珍视、更多样的自我表达。封装的是昨天，逼出来的明天却是两面的：既有新的智慧，也有新的难题。

◆ 九、第五编小结：自我界也有工具

回头看这五章。镜子、日记书信与自传、修行法门、心理学、自拍与社交主页，是自我界里五件时间跨度很大的工具。它们的形状各不相同，却走着同一条路：把一份关于"怎样认识自己"的已发生智慧封装起来；一经使用，就与现实界（衰老、记忆、身体、重复检验、美颜）和理念界（肖像、叙事、教义、概念、人设）发生碰撞；碰撞生出封装之外的新自我问题（我的样子是我吗、写出来的我是

真的吗、观察者怎么办、我是不是有问题、被展示的我）；新问题又逼出更高的自我智慧。每一次，旧的自我理解都没有被"完成"，而是被推到了下一个层次。

自我界没有终局的"认识自己"。没有哪一件工具，让人从此"认识了自己"，然后停下来。一个人如果想象有一台机器可以把自己彻底看穿，那他其实是在想象一次封装可以终结发生，而这恰恰是前面已经反复说明的、历史与结构都不支持的想法。如果能找到一件在自我界被长期大量使用、却始终没有生出任何它自己预料不到的新自我问题的工具，本编这条线就要从"每一件"改成"多数"。

下一编，我们把三界放到一起看，看工具怎样不肯留在自己出生的那一界。

> 带走的话
>
> 升级和代价是并存的，不是互相抵消：几亿人集体摸出了识别摆拍的眼力，也真有人被点赞数拴住了。哪一边，都不能拿来抹掉另一边。

> **小账** 封装了什么？——“我可以决定怎样被看见，并让很多人同时看见”。 逼出了什么？——等待反馈、比较、分裂感、自我外包这些真实的焦虑；同时也有对展示的识别力、分层展示的分寸和对“不展示”的珍视。

承接/预告：到此为止，我们分别看了现实界、理念界、自我界的工具史。下一编讲三界交织，看一个工具怎样在三界之间来回碰撞，以及对"工具让人退化"和"毁灭性工具"这两种最有力的质疑，本书怎样回答。

第 六 编

三界交织：一界有工具，三界有新生

编 序

前面三编，我们分头走了三条路：现实界的工具，理念界的工具，自我界的工具。分开走，是为了看清每条路上的风景。走完之后，读者大概已经隐约觉得不对劲：时钟明明是挂在墙上的东西，怎么讲着讲着就讲到了物理学的基础？镜子明明是照脸的，怎么又牵出了化妆、肖像和整个"自己"的观念？

这个"不对劲"，就是本编要讲的东西。工具出生在哪一界，就只在哪一界里起作用，这是一种很自然的想象，也是错的。本编用四章把它拆掉：第27章讲机制，第28章讲结构证明，第29章和第30章各回答一个最有力的反驳。

章	讲什么
27	机制：为什么工具从不只停在出生的那一界
28	结构证明：三界为什么同步升级，文明层必然，个体层靠碰撞
29	反驳一：工具让人退化了吗
30	反驳二：毁灭性工具

先说清本编的主张有多强。它不说每件工具都在三界各留下一个同样深的脚印，只说更窄也更硬的一条：没有哪一界会被工具的出生地自动排除，而且有一股力量，把工具的后果推出它出生的那一界。

所以，本编是全书“证明”的主场之一：前面三编的史例，在这里被收成一个结构论证；承重命题里的“一界有工具，三界有新生”和“被封装能力的退场是让渡，不是退化”，都在本编立住，也在本编受到最重的检验。

第 27 章

跨界碰撞的机制：为什么工具从不只停在出生的那一界

◆ 一、一只闹钟，和它没打算做的事

先从一件小事说起。

有一对年轻夫妻，买了一台带定时功能的电饭煲。买它的理由很单纯：早上起来想吃上热粥，不想守着锅。这是一件纯粹属于现实界的事，米、水、电、锅，一个会自己停火的盒子。

用了半年，两个人发现家里发生了一些说明书里没写的事。

第一件，晚饭的时间变了。从前做饭是"饿了就做"，现在因为粥要提前一晚定时，他们每天晚上多了一个小动作：打开手机，想一想明早吃什么。于是"明天"成了一个每晚都要被正式想一遍的东西。这是理念界的变化，他们对"一天"的切分方式变了。

第二件，女主人发现自己变得不爱赖床。倒不是粥有多香，而是锅里的粥是"准时熟的"，晚起半小时，粥就在保温里搁着，她会有一种说不清的、被一个盒子等着的感觉。这是自我界的变化，她对自己的作息有了新的要求。

没有人设计这些。电饭煲的工程师想的是火候和定时器。一件出生在现实界的工具，一年之内在三界里各留下了一个脚印。

这不是巧合，是机制。工具为什么从不只停在出生的那一界？

◆ 二、原理一句话，与封装即释放

先把话放在前面：工具是已发生智慧的封装；封装只封住了出生那一界里的那部分智慧，而使用它的人是带着三界一起来的。

所以，"出生的那一界"不是工具的领地，只是它走进生活的入口。入口在哪里，不决定它最后走到哪里。

这里要先交代两个词，本书后面一直要用。

一个是SDE。它是本书看事情的一套办法，三个字母可以用大白话读：S是"显露"，一件事已经显出来、能看见能摸到的稳定样子；D是"差异"，是让这个样子一步步出来的过程，一次次分出高低、对错、这个和那个；E是"纠缠"，是这一切发生的土壤，是人、事物、意义纠缠在一起的那片场。任何一件东西，都可以问三个问题：它显露成了什么样？它是怎么一步步分出来的？它是在哪片土壤里长出来的？

另一个是"三界"：现实界是我们脚踩的、手碰的东西和事；理念界是概念、规则、道理、符号；自我界是一个人怎样看待自己、认识自己、安顿自己。

有了这两个词，上面那句话可以讲得更具体。电饭煲里被封装进去的，是"怎样把米煮成饭"这一段已经发生过的智慧：火候、水量、定时。这是一段S，已经稳定、已经显露，可以反复复制到千万个盒子里。可是人把盒子拿回家使用，就是把自己整个的生活拿来和它碰在一起。他的生活里有"一天怎么分"的概念，有"我是个怎样的人"的自我看法。碰一下，这两处也要动。

这里有一个看上去矛盾、其实很好懂的说法：封装即释放。

封装，是把一段智慧装进工具，让人不必再亲自去想那一段。电饭煲封装了火候，所以你不必再盯着锅。这是"省"。

省下来的东西不会凭空消失。你不盯着锅了，注意力、时间、心思就被释放出来，总要落在某个地方。

这个想法本身不新。经济学里有杰文斯悖论：煤烧得更省，总用煤量反而上升。家务技术史里，省下的体力又被更高的标准占去（第28章会提到）。它们都在说：省下来的东西，会被别的事占住。本书补的是落点：它不只落在同一件事上（省下的家务时间用来做更多家务），更常落到另外两界，变成新的概念和新的自我要求。

于是有一个对称：工具封装得越彻底，它在原来那一界里省得越多，在别的界里放出来的就越多。盯锅的时间变成想明天的时间，想明天的习惯变成对自己作息的新要求。每一步都不是工具直接做的，是工具让出的空位，被人用别的东西填上了。

这个机制可以拆成三步，后面几章反复用到：

第一步，封装。一段智慧被固定下来，装进工具。　第二步，使用即碰撞。工具进入人的生活，和人原有的结构、差异、土壤相撞。这一点第二编已经立过，这里不再重讲。　第三步，碰撞的落点不限于出生的那一界。人的生活本来三界叠在一起，工具一落地，就同时落在三界的交界处。

◆ 三、时钟：出生在现实界，改变了三界

下面用历史上的大例子检验这个机制。

机械时钟出生在现实界。它是齿轮、擒纵机构、发条或重锤，是可以造出来、摆在塔楼上的东西。它封装的，是"把时间切成均匀小段并数出来"这一段智慧。

时钟出现以后，第一个改变发生在现实界之内：城镇的作息有了公共的钟点，集市开闭、教堂敲钟、工场上下工，都可以按钟办事。

第二个改变来到理念界。时间原来是"日出日落""农忙农闲"那种有伸缩的东西，时钟把它变成了均匀的、可以度量的量。后来的人谈论速度、频率、"每秒"，都以这个均匀的时间观念为底。到十九世纪晚期，这个观念被推到了极限：北美铁路在1883年推行标准时区，因为火车跑得太快，各地各用各的太阳时，时刻表根本排不出来。科学史家伽里森（Peter　Galison）的研究又提示了一层：远距离时钟同步这个实际工程问题，和后来关于"同时"到底是什么意思的物理思考，有不浅的联系。铁路和电报要对时，专利局的案头堆着同步时钟的发明，他把这些看成相对论诞生的土壤的一部分。他并不说钟表技术直接造出了相对论，也有科学史家认为这样讲高估了技术一侧；本书只取"实用的校时技术与同时性的思考有联系"这一层。

第三个改变来到自我界。一个社会习惯了按钟办事，人就开始把钟装进自己身体里。"准时"成了一种品格，"浪费时间"成了一种自责，"我这一天过得有没有价值"成了一个人对自己的追问。靠日出而作的农人，不会这样审问自己。

时钟的工匠想的是怎么让摆动更均匀。一件现实界的工具，在理念界里改写了"时间是什么"，在自我界里改写了"我该怎样过日子"。

◆ 四、复式记账：出生在理念界，改变了三界

再换一个方向，看一件出生在理念界的工具。

1494年，意大利数学家帕乔利（Luca Pacioli）在他的一部数学汇编里，用一个章节系统讲述了威尼斯商人使用的复式记账法。复式记账不是他发明的，是商人们用出来的，他把它写下来、传了出去。它本身是什么？是一套纸面上的规则：每一笔交易同时记两笔，一边记来源，一边记去向，两边必须相等。它没有任何实物，是一套符号纪律。

它封装的智慧是：怎样让一个生意的来龙去脉，在不依赖记忆的情况下，自己对得上账。

在理念界，它造出了"利润""资本""资产负债"这些以前模糊不清的概念。以前商人心里大概知道自己赚没赚，现在有了一个可以精确回答的问法。

在现实界，它让更大规模的生意成为可能。一个人的记忆管不了几十个合伙人、十几条商路，一本账管得了。复式记账与近代商业组织兴起之间有多深的联系，是个长期有争议的论题，本书不站队，只取"它降低了协调复杂生意的成本"这一层。

在自我界，"账"慢慢变成了人看自己的方式。"算一笔账""我欠他一个人情""这辈子值不值"，人们用账本的语言审视自己的生活。一个出生在商人案头的记账法，最后渗进了人怎么给自己打分。

◆ 五、镜子：出生在自我界，改变了三界

第三个例子是镜子，第22章已经细讲，这里只取一点。

镜子出生在自我界：它让人第一次不靠别人的描述、不靠水面的晃动，就能稳稳看见自己的样子，封装的是"从外面看自己"。可之后，现实界里梳妆整衣成了日常，一整套相关行业跟着长出来；理念界里，"外貌""仪态""第一印象"有了实实在在的所指，一张脸成了可以谈论、可以比较、可以描绘的对象，自画像也有了更稳的土壤。镜子对自画像兴起的作用，说"为之提供条件"就够了，不必说"造成"。

三个例子，三个出生地，一个结果：每一件都在三界里各留下了脚印。

◆ 六、为什么必然如此：对称与错位

现在可以回到问题本身：为什么工具从不只停在出生的那一界？有两个理由，一个来自人，一个来自工具。

来自人的理由是：人的生活没有界墙。　使用电饭煲的人，不会在厨房门口把自己劈成三份，只让"现实界的我"进去。他是一个整体，带着对时间的概念、对自己的看法，一同走进来。工具碰到的，是整个的人。

来自工具的理由是：封装本身就是一次切割，而切割必然留下边缘。　工具封住了某一段智慧，这一段的边界之外，是它没有封住的东西。第二编讲过，叫"错位"：封装的时候，新SDE还没有出现，所以不可能被封进去。现在加上跨界这一层：没被封进去的新SDE，不会只出现在同一界。它出现在哪里，取决于人带着什么来碰撞，而人是带着三界来的。

所以这是一个结构上的必然：**工具出生在一界，却把碰撞摊在三界**。

这句话到底有多强，要说准。它不说"三界必定各留下同样深的脚印"，只说两件事。一，人没有界墙，所以没有哪一界会被出生地提前排除。二，封装释放出来的东西总要落脚，而落脚处由使用者的整个生活决定，不由工具的设计说明决定。两条合起来推出的是：跨界是常态，且无法预先划定范围；推不出"处处均匀"。

还有一个必须堵上的漏洞。三界是讲述者画的分界，同一个后果常常能算在两界，如果怎么算都行，"三界各有新生"就成了永远不会错的话，也就什么都没说。所以要给"新生"立三条判据：一，是以前没有的做法、概念或自我问题，不是旧东西换了个说法；二，不在工具的设计说明里，设计者也没打算要它；三，使用者或社会要对它做出回应，不回应就付代价。按这三条，电饭煲带来的"每晚想一遍明早吃什么"算新生，"它煮得很快"不算。

◆ 七、手机，与换一个角度看同一个机制

嫌时钟、记账、镜子离得远，就看口袋里的手机。它最初的出生地很清楚：现实界，一块能打电话的硬件，封装的是"远距离说话"。

可是今天，一个人的手机里装着相册，这改变了他怎样回忆过去，是自我界的事；装着地图和支付，这改变了他怎样在城里活动、怎样买东西，是现实界的事；装着各种"打卡""步数""年度总结"，这改变了他对"一年算不算过得好"的理念，是理念界的事。没有一个设计师同时画了这三张图，它们是千万人用出来的。

许多人是在手机忽然不在身边的那几个小时里，才第一次察觉它在三界里同时占了多大的位置。工具的跨界效应，往往要等它缺席了才看得清。

换成做菜的话来说。一个人买回一口新铁锅，锅封装的是"均匀受热"。锅一到手，他开始琢磨火开多大（现实界的手感），读别人写的"养锅"心得、学会"油温几成热"这套说法（理念界的概念），也开始觉得自己"是个会做饭的人"了（自我界的认同）。锅没有教他这些，是他拿自己的生活去碰那口锅，碰出来的。说"工具跨界"，不是工具长了脚、自己跑过了界线，而是人带着整个自己去碰了它。界线是讲述者画的，生活里没有。

◆ 八、这个论证的边界在哪里，与小结

需要老实交代四件事。

第一，"三界各有脚印"并不意味着"三界各有同等大的脚印"。有些工具的跨界效应很强，比如时钟；有些很弱，比如一把普通的螺丝刀，它大概只在现实界起作用，也许在理念界留下一点"拧紧"的概念，在自我界几乎没有。释放出来的时间，也可能只是变成多睡一会儿，那样，这件工具在这个人身上就没有碰撞，这是个体层的事，下一章讲。

第二，上面三个大例子，都是事后挑出来的，我们挑它们，恰恰因为它们跨界效应显著。这是历史例证的老毛病，会有幸存者偏差。所以这些例子只当机制的说明和佐证；真正的证明，是"人没有界墙、封装必有边缘"这个结构论证，下一章把它排成一条链。

第三，电饭煲的故事是为说明机制而构造的日常情景，不是对某个家庭的实证观察。读者可以用自己家里的事替换，换一个电器，换一个人，结构不变。

第四，本章在什么情况下是错的？可以检验：随便取一批被千万人用了十年以上的日常工具，取样不由我们来挑，逐件问，在出生界之外，有没有出现符合上面三条判据的新做法、新概念或新的自我问题？如果多数工具的答案是没有，"从不只停在出生的那一界"就要改成"有时"；如果只有少数强工具有，那本章讲的就只是那少数。螺丝刀这类弱工具是预留的反例，它们占多大比例，才是要看的数。

工具出生在哪一界，就只在哪一界里作用，这是一个自然却错误的想象。工具封装的是一段已发生的智慧，使用它的人却是三界叠在一起的整体。封装越彻底，释放越多；释放出来的注意力和心思，会落到别的界里，和那里的东西相撞，生出封装之外的新SDE。时钟、复式记账、镜子，分别出生于现实界、理念界、自我界，却各自改变了三界。出生界是入口，不是领地。

下一章，我们把这个机制升级为结构上的证明：为什么三界的升级是同步的，而且是被逼出来的。

> **带走的话**
>
> 工具的出生地只是入口，不是领地。你买的是一个电饭煲，用着用着，连“明天吃什么”和“我该几点起”都被它碰了一遍，因为用工具的人身上没有界墙。

> **小账** 这一章立了什么？——封装有边缘、人没有界墙，所以工具的后果不会只停在出生的那一界；它只说无法预先划界，不说三界处处均匀。 错在什么情况下？——随便取一批被千万人用了十年以上的日常工具，取样不由我们挑，若多数在出生界之外都没有符合三条判据的新东西，“从不只停在”就要改成“有时”。

承接上章机制，下一章讲：三界同步升级的结构证明，以及文明层与个体层的区别。

第 28 章

三界同步升级的结构证明

◆ 一、一个新手妈妈的洗衣机

先讲一个再普通不过的故事。

一位年轻妈妈，孩子刚满三个月。家里新买了全自动洗衣机，她很高兴，终于不必再搓洗那一堆沾了奶渍的小衣服。

头两周，她的感受是单纯的轻松。洗衣机封装了"把脏衣服洗干净"的整套智慧：水位、转速、洗涤时间，她按一个键就好。第三周，事情有点不同了。洗衣服太容易，她开始更勤地换孩子的衣服，一天三套不算什么。衣服多了，她又觉得要分颜色、分材质来洗，于是在手机里查"婴儿衣物该怎么分类洗"。查着查着，她加入了一个妈妈群，里面有人在讨论"婴儿衣物要不要单独用消毒洗涤"。她心里冒出一种新的焦虑：是不是我不够讲究？

洗衣机封装了洗衣，省下了体力。可随之而来的是三件封装里没有的东西：一套新的卫生标准（理念界），一种新的比较和焦虑（自我界），一个新的社交圈和信息渠道（现实界）。

最关键的一点是：这三件事，没有一件是洗衣机"做"的，却没有一件离得开洗衣机。

美国历史学家考恩（Ruth Cowan）研究家务技术史，得到过一个让人意外的结论：各种家用电器进入家庭之后，主妇花在家务上的时间并没有明显减少，因为标准提高了，活儿的内容变了。她研究的是美国家庭，主要在二十世纪上半叶。〔待核：结论强度以原著为准〕这个研究常被拿来说"技术没有真正省力"。本书的读法不同：这恰恰是封装与碰撞的典型表现。被封装掉的是"洗"的体力，被逼出来的是"洗到什么程度才算合格"的新标准。

◆ 二、把论证排成一条链

现在把散落的零件排成一条完整的链。链子分两段：前三环讲"为什么会有新SDE"，第二编已经立过，这里只回扣；后三环讲"为什么它不限于出生那一界"，是本章新做的事。

第一环：工具是已发生智慧的封装。　智慧先发生过，然后被固定下来，内容停在封装完成的那一刻。

第二环：使用即碰撞。　工具一旦被使用，就造出一个此前不存在的情境，也就是"有这个工具的世界"。工具在其中和人、和别的工具、和别的需要相撞。

第三环：碰撞里生出的新SDE，不在封装之内。"有这个工具的世界"是工具造出来之后才有的，封装的那一刻它还不存在，不可能被装进去。这就是错位定理。原工具处理不了这些新问题，只能靠再造一件新工具，或者靠人的新一轮智慧。

第四环：碰撞的对象是使用者整个的人。　碰撞发生在人身上，人是三界合一的，没有哪一界的他可以被留在门外。所以新SDE出现在哪一界，不由工具出生在哪一界决定，而由使用者带着什么去撞决定。这一环给出的是"没有壁垒"。

第五环：封装越彻底，释放越多，释放出来的东西必须落脚。　落脚处常在别的界。这一环给出的是"有推力"。

第六环：新SDE一出现，就带着要求。　不回应就付代价：赶不上火车，拿不到工资，被同伴比下去。所以回应不是自愿的锦上添花，而是账单。

链子到这里闭合：被广泛使用的工具，其后果不受出生界的限制，而且逼着人做出回应，也就是智慧升级。要注意这条链各环的分工：新SDE"必有"，靠的是第三环的错位；"不限于一界"，靠的是第四环和第五环的合力；"非升级不可"，靠的是第六环。它给出的是无壁垒加有推力，不是"三界必然均匀"。

◆ 三、"逼"字的分量

这里想多说一点"逼"字。

有人会说：升级是人类的进取心，是求知的天性，哪里是被逼的？这话有一半对，进取心当然存在。但升级不是进取心的奖赏，而是环境改变之后寄来的账单。

回到洗衣机的故事。妈妈没有主动决定"我要提升婴儿衣物卫生标准"，是洗衣机让洗衣变得容易，容易之后，"洗得更讲究"成了一件新的可做之事，别人也都在做，于是成了必须做的事。再看时钟，没有人决定"我们要学会准时"，是铁路、工厂、学校都按钟点运行，不准时，你就赶不上火车、拿不到工资、上不了课。

这个机制，技术史里有相近的说法。休斯（Thomas Hughes，1983年）讲技术系统里落在后面的环节，他叫"逆突出部"，会把发明者的注意力逼到那里；罗森堡（Nathan Rosenberg）把这类失衡叫作"聚焦装置"。它们说的是技术系统内部怎样被逼着往前走。本书补上的是：逼也会穿过使用者，从工具系统一直逼到人的概念和自我。

◆ 四、文明层必然，个体层靠碰撞

下面要讲本章最要紧的一个区分，也是本书反复强调的区分。

前面的链条证明了什么？它证明的是：在文明这个层面上，工具必然逼出新的SDE，从而逼出智慧的升级。 这里的"必然"要说清是哪种必然。用的人足够多、处境足够不同，那么没有任何一个人撞出新东西的可能，就趋近于零；撞出来的东西被别人学去、写下来、变成规矩，才算文明层的升级。这是大数意义上的必然，不是逻辑上的必然。

但这条链并没有证明：每一个具体的人，都必然升级。

两句话差别巨大。想象两个人，同时用上了导航。

第一个人，每天开车就是"跟着导航走"。导航说左转他就左转，说前方拥堵他就绕行。他和导航之间没有碰撞，只有服从。久而久之，他不知道自己的城市长什么样，换一个没有导航的情境，他会手足无措。他没有升级，反而被导航"封装"了：他成了导航的一个延伸，一个执行指令的手脚。这样的人，可以称为"被工具使用的人"。

第二个人，用导航，同时也留意。他发现导航推荐的路线总绕开某个路口，他好奇为什么；他注意到某条小路其实更快，导航不知道；他开始琢磨"路网"这件事，甚至因此对城市规划发生了兴趣。他拿自己的判断、好奇和经验，去撞导航的输出，于是身上长出了新的东西。

两个人用的是同一个工具，结果不同。区别不在工具，在于有没有碰撞。

所以：**文明层必然，个体层靠碰撞**。一个人是否升级，取决于他是否与工具发生了真正的碰撞；只使用不碰撞，就会被工具封装，成为"被工具使用的人"。

这个区分，是第29章回答"工具让人退化了吗"的关键。退化的质疑，很大一部分其实是在描述"只使用、不碰撞"的那类个体的遭遇。这一点要老实承认，不能含糊。

◆ 五、为什么"只使用"会被封装

可能有读者会问：工具是死的，怎么会封装人？

先把用词说清。前面说"封装"，是智慧被装进工具；这里反过来，说的是人的活动被装进工具的接口。同一个字，方向相反：一个是工具里装着智慧，一个是接口圈住了人。人只在工具规定好的接口上做动作，不留出口，就是在被封装的范围里活动。

举做菜的例子。有一类调料包，撕开倒进锅，就是一道"某某风味"的菜。一个人长期只用调料包，他的手艺就被调料包封装了：他会按步骤，却不知道咸和鲜是怎么配出来的，更谈不上自己调出一个新味道。他的厨艺天花板，就是调料包的天花板。

另一个人也用调料包，但他会尝：哪个包偏咸？哪个包缺酸？他试着加一点醋、少放一半盐。他在拿自己的舌头去碰调料包，会成为一个比调料包更会做菜的人。调料包是同一个，区别在于他有没有拿自己去撞。

◆ 六、再看一次：新手父母的第一个夜晚

用一个更贴近的场景，把这条链走一遍。一对新手父母买了一台婴儿监护摄像头，夜里能从手机上看孩子的睡姿。

封装：它封装的是"替你盯着床边"。使用：第一晚，他们果然睡得踏实了一些。碰撞：第二周，妈妈发现自己睡前总要先看一眼屏幕，孩子翻身、皱眉、哼唧，每一个小动静都变成需要判断的事件。她不知道哪一种算正常。这个问题，是摄像头封装时没有的，"有了摄像头的夜晚"才出现。

于是三界各有新生：现实界，夫妻开始商量轮流看屏幕的规矩；理念界，她开始查"婴儿睡眠周期"，学会了一批以前不需要的词；自我界，她问自己：我是在照顾孩子，还是在被一块屏幕拴住？

她若只是被屏幕牵着走，一有动静就惊醒，就成了"被工具使用的人"；她若拿动静去对照书本、去问有经验的长辈、慢慢长出自己的判断，就在碰撞里长出了新的东西。同一个摄像头，两种结局。

◆ 七、三界同步的结构意义

把这些放在一起，三界同步升级的意义就清楚了。"同步"的意思是同一次碰撞，不是三界同时等深。

第一，三界的升级不是三件独立的事，而是同一次碰撞在三个方向上的展开。所以不能说"现实界的工具只管现实界，理念界的事留给哲学家，自我界的事留给心理咨询师"。使用者在使用的当下，三界都被触动了。

第二，三界之间有传递。现实界的工具逼出理念界的新概念，新概念又成为理念界的工具（理论也是工具，第四编已讲），这些理论工具再回到现实界和自我界去碰撞。升级是一个会自己继续转下去的链，不是一次性事件。

第三，由于升级是被逼出来的，在文明层面它没有"到此为止"的时刻。只要新的碰撞还在发生，新的SDE就还会出现。这为第九编做了铺垫：一个想成为终局的

智能，必须把所有后续的碰撞都封装在内，而封装与新SDE之间的错位，正是它做不到的原因。

◆ 八、这个论证的边界在哪里

有四处需要老实交代。

第一，"文明层必然"成立的前提是：工具被广泛使用，并且使用者中总有人具备碰撞的能力和意愿。如果一个社会里所有人都只使用不碰撞，升级就会停滞。历史上有没有这样的社会，我没有把握，也不想假装知道。本书的论证说的是：只要存在足够多的碰撞者，升级就必然继续；它没有说碰撞者一定足够多。

第二，"逼"的强度因工具、因时代而异。有的工具逼得急，比如印刷术、电力；有的逼得缓，比如一些很小的日用品。本章说的是结构上的"必然会有"，不是时间上的"一定很快"。

第三，"碰撞生出新SDE"是结构性的断言，不等于"每次碰撞生出的新东西都是好东西"。新生可能是新机会，也可能是新麻烦，甚至是新灾难。第30章专门处理这一点。

第四，本章在什么情况下是错的？链子上有两个薄弱环节，各对应一个可检验的现象。第四环断了：如果找到一类工具，大量使用多年，后果被严格限定在出生界，另两界既无新概念也无新的自我问题，且不是因为用的人少，那么"没有壁垒"不成立。第六环断了：如果一件工具在一个社会里普及数十年，社会却没有为它的后果添任何新的规矩、职业或说法，用的人也没有为此付出任何可见的代价，那么"账单"一说在这里不成立。

◆ 九、小结

工具是封装，使用是碰撞，碰撞生出封装之外的新SDE；人是三界合一的，所以新生不受出生界的限制，释放出来的东西又必须落脚；新生带着要求，所以升级是账单，不是奖赏。这个"逼"在文明层面是必然的，在个体层面却要靠碰撞：只使

用不碰撞的人，会被工具封装，成为"被工具使用的人"。这个区分，是回答下面两个反驳的钥匙。

> **带走的话**
>
> 升级不是奖赏，是账单：你不回应，就赶不上火车、拿不到工资、被别人比下去。但账单是开给文明的，落到你个人头上，得看你有没有真去撞过那件工具。

> **小账** 这一章立了什么？——一条六环链：封装、使用即碰撞、新SDE在封装之外、碰撞的是整个人、释放必须落脚、新生带着要求；结论是文明层必然、个体层靠碰撞。 错在什么情况下？——第四环断了（有工具的后果被严格限定在出生界，且不是因为用的人少），或第六环断了（普及数十年，社会没添任何新规矩、新职业、新说法，也没人为此付代价）。

承接本章，下一章讲：第一个反驳，工具让人退化了吗?

第 29 章

反驳一：工具让人退化了吗?

◆ 一、一个迷路的人

先讲一个很多人都有过的经历。

有一位朋友，开了二十多年车，年轻时能把一座城市的大街小巷背下来。后来手机导航普及了，他用了好几年。有一天，手机没电了，他开到一个熟悉的区域，忽然发现：自己不知道怎么走了。不是完全不知道，而是那种"拐过去自然就到"的感觉，不见了。

他很沮丧，说了一句很多人都会说的话："这东西把我用废了。"

这是对工具最有力的一类反驳：你说工具逼出了升级，可是我亲眼看到，有些能力被工具吃掉了，人变得更差了。　这类指责很老，柏拉图在《斐德若》里借苏格拉底的口，就担心过文字会让人不再费力记忆。它抓到了真实的东西，本章不绕开，正面回答。

答案的骨架先放在前面：被封装能力的退场，是让渡，不是退化。但是在让渡的过程中，被封装的那个人，确实承受了损失。　这两句话要一起说，缺哪句都不诚实。它们背后还有一把尺子，下面一步步亮出来：分界不在能力降没降，而在判断权在谁手里。

◆ 二、先看一个真实的研究：伦敦出租车司机

要谈导航对大脑的影响，有一个绕不开的研究。

伦敦的出租车司机要通过一项出了名难的考试，把全城成千上万条街道的位置与走法记在脑子里。英国神经科学家马圭尔（Eleanor　Maguire）及其同事在2000年发表的研究，用脑成像比较了十六位伦敦出租车司机与普通人，发现司机

海马体的后部平均更大，而且从业年限越长，这一区域越大；海马体的前部，反而比对照组小。海马体是和空间记忆有关的脑区。

这个研究有两层意思。一层很直观：能力会长在身体里。　另一层有人据此做了猜想：大脑里的位置有限，练了这样，就少练那样。后一层只是猜想，但它提醒我们，"交出去一项能力"在身体里的对应物，未必只是萎缩，也可能是资源改投。

于是质疑者顺理成章地问：既然大量使用空间记忆会让这块脑区变大，把空间记忆交给导航，它会不会萎缩?

这个推论听上去合理，但要小心。研究看到的，是出租车司机与对照组的差异，以及差异与从业年限的相关；它没有直接告诉我们"普通人长期用手机导航，海马体就会萎缩"。这是两个不同的问题。后来确有研究报告，越是习惯用导航的人，自己找路时空间记忆越差（2020年有这样的发现），但这类研究多半只显示相关，谁是因、谁是果，还没说死。所以用这个研究断言"导航让大脑退化了"，是一个跳步；说"完全没有影响"，同样是跳步。

不过，即使承认长期依赖导航，确实会让一部分人的空间记忆变弱，真正的问题才刚开始：这算"退化"吗?

◆ 三、让渡，还是退化

先厘清这两个词。

"退化"的意思是：一个人原本拥有某种能力，这种能力对他仍然重要，现在下降了，他整体变得更差。

"让渡"的意思是：一种能力被一个工具接过去了，人把它交出去，同时腾出了原来被它占用的资源，去做别的事。

两者的区别，不在能力降没降，两种情形下，个人手里的那项能力都可能降。区别有两条，缺一不可：

一，被交出去的这一段，是不是继续被有效地完成着；　二，交出去之后，人手里还握不握着判断权，也就是工具出错时，他看不看得出来、接不接得住。

两条都有，是让渡。第一条有、第二条没有，是被封装：执行交了出去，判断也跟着交了出去。第一条也没有，没有任何人、任何东西在完成这件事，才是真正的退化。

回到导航。"记住全城的街道"这件事，对一个现代人还需要完成吗？需要，只是完成的主体换了：由导航系统来完成。人这边不再需要背，需要的是会用、会判断：路线是否合理，路况是否异常，目的地是否输对。

所以，从文明层面看，整个社会找路的能力没有退化，而是提高了。从前陌生城市里的人只能问路、看地图、迷路；现在任何人到了任何陌生城市，都能找到路。被封装进导航的那段空间智慧，是整个社会成倍增加的。这是第一个层面：文明层，让渡是净增益。

那位迷路的朋友呢？他那句"把我用废了"，说的常常不是能力降了，而是判断权丢了：手机没电的那一刻，他才发现这几年他并没有在找路，而是在被引路。

◆ 四、计算器与心算：另一个日常的例子

再看一个更熟悉的例子。

很多人小时候都经历过"能不能用计算器"的争论。老师说，不能用，用了心算就废了。后来大家都用了，心算果然不如从前。有人因此感叹，现在的年轻人连找零都要算半天。

先承认：心算能力，在大量依赖计算器的人群中，平均而言确实下降了。再看另一面。计算器封装了"准确快速地做四则运算"这一段智慧，把它变成了一个几块钱的盒子。结果是，成千上万不擅长算数的人，第一次可以放心地处理账目、做预算、核对单据。算数不再是只有少数人才能过的门槛。算数这一步变便宜之后，"算什么"成了更重要的问题：这笔账该怎么设计？这个数据该怎么比较？学生不用花二十分钟去算一道繁琐的乘除，可以把时间用在理解这道题的意思上。

计算器的故事和导航一样，有两面：一面是某项具体技能的萎缩，一面是整体能力的扩展。

◆ 五、为什么这是让渡，不是退化

理由可以说得更完整一点。

第一，能力没有消失，只是换了载体。　算数的能力现在在计算器里，找路的能力现在在导航里。社会整体仍然拥有这些能力，甚至拥有得更好。

第二，被腾出来的资源有去处。　封装即释放。人不用背路线，就多出了注意力；不用手算，就多出了时间。这些资源，在碰撞者那里变成了新的东西，在文明层面变成了新的产业、新的问题、新的智慧。

第三，横杆升高了。　这一点第八编会详细讲：一种能力被封装，"人需要具备什么"的标准也随之改变。今天衡量一个人会不会找路，不是看他能不能背地图，而是看他能不能在导航给出的信息之上做判断。拿背地图的标准去量今天的人，当然觉得他们退化了，但那是拿旧横杆量新人。

这三条合起来，就是本书的承重命题之一：**被封装能力的退场，是让渡，不是退化**。

◆ 六、但是，被封装的人确实有损失

现在要说那半句同样重要的话。

如果就此得意洋洋地宣布"所以没有退化，大家放心"，那是不诚实的。让渡发生的时候，有一个时间段，我把它叫作"让渡期"。在让渡期里，有一批人确实承受了真实的损失。

这些人是被封装的人，也就是上一章讲的"只使用、不碰撞"的个体。想想那位迷路的朋友：他曾经拥有的城市感没有了，失去导航时变得无助。他不能安慰自己说"整个社会的找路能力提高了"，因为那是整体的账，他个人的账是赤字。

更重的例子，是能力依附于旧工具的职业人。新工具接过了某段技能，原来靠这段技能吃饭的人，会承受真实的冲击。第二编讲过的"内卷期"，就是这种冲击的集中表现。这里不展开，只强调：文明层的净增益，不能抹平个体层的代价。

自动化研究里有个相近的老发现。班布里奇（Lisanne Bainbridge）1983年把它叫作"自动化的反讽"：自动化接管的活儿越多，留给人的往往越是最难的那一部分，偏偏人又因为平时不练而变弱，随时接手的任务反而更重。它和本章相同的地方，是都看到执行交出去之后，判断留了下来，而且更难了。不同的地方，是自动化研究多讲岗位上的操作员，本章把同一个结构拉到人人日常都在用的工具上，并用它划出让渡与退化的界。

所以，诚实的说法是：

- 从文明层看，被封装能力的退场是让渡，整体智慧在升级。
- 从个体层看，被封装的人承受损失，而且这种损失是真实的，不应被"进步"这个词轻轻带过。

这两句话不矛盾，它们说的是同一件事在两个层面上的样子。

◆ 七、怎样才不被封装

既然损失落在"被封装的人"身上，自然的问题是：怎样才不被封装?

答案还是上一章那句：碰撞。 具体到日常，做法很朴素：

用导航的时候，偶尔问自己："这条路为什么这样走？"哪怕只是看一眼地图的整体，而不是只盯着下一个转弯箭头。

用计算器的时候，先估一估。算之前心里先有个大概的数，算之后看看对不对。这一估一对，就是你拿自己去撞计算器。

用任何一件替你做事的工具，在关键处停一下，问一句："如果它错了，我能看出来吗？"能看出来的人，没有被封装；看不出来的人，已经被封装了。

这些做法的共同点，是不把判断权完全交出去。让渡的是执行，不是判断。把执行交出去叫让渡，把判断也交出去，才叫被封装。

◆ 八、还有一种退场：记电话号码

再举一个人人有过的例子。从前，一个人能背下十几个亲友的电话号码；通讯录普及之后，大多数人只记得自己的号码和一两个最亲的人的。背号码的能力确实下降了，可联系人从十几个变成几百个，换了手机名单也不会丢，在"联系得上"这一层，人际网络大大拓宽了。当然也有人承受了损失：半夜手机没电、被困在陌生地方、想不起任何一个号码的人，他的狼狈是真实的。

很多人因此养成了一个朴素的习惯：把最要紧的两三个号码，写在钱包里的小纸条上。执行交给通讯录，对"万一它不在时怎么办"，自己心里留一个出口，这就是判断权没交出去的一个微小例子。

◆ 九、这个论证的边界在哪里，与小结

本章需要交代的边界比前两章更多，因为这个反驳击中了论证最容易被攻击的地方。

第一，本章没有证明"所有被封装的能力退场都是无害的"。如果一个社会的绝大多数人都不再掌握某种关键的基础能力，而它所依赖的工具一旦失灵，后果可能很严重。这是真实的"脆弱性"问题。本书的立场是：让渡不等于无风险，应对的办法是保留少数人的核心能力、建立冗余，而不是拒绝让渡。具体怎么做，超出本书范围。

第二，关于海马体的研究，只能支持"空间记忆的使用与脑区变化有关"这一层，不能直接支持"导航使人退化"或"导航不使人退化"任何一方。本书不假装科学已经有了定论。

第三，"让渡期的损失"没有被本书的论证消除，只是被如实记下。命题讲的是智慧与创新必然升级，没有承诺每个人都在这个过程里受益。

第四，什么能力应该让渡、什么不应该，是价值判断，不由结构论证决定。有些能力，比如亲手做一顿饭给孩子吃，即使有机器可以代劳，人们也可能选择保留，因为这件事的意义不在结果，而在过程。本书不替任何人做决定。

第五，本章在什么情况下是错的？有三个可检验的现象。其一，看导航出错的日常时刻（路封了、导进了死胡同），如果多数用户当场就能察觉并自己改道，那么"被封装"在导航上只是少数人的事，那位迷路的朋友就不能代表多数。其二，"腾出的位置被占住"可以追踪：如果某项能力退场之后，使用者既没有在判断上、也没有在别处多出任何新做法，只是单纯少了一块，那么在这个人身上，这就是退化，不是让渡。其三，如果习惯在关键处停下来核对的人，在工具出错时并不比完全依赖的人更早察觉，那么"碰撞能保住判断权"的说法就没有根据。

工具让人退化了吗？本章的回答分两层。在文明层，被封装能力的退场是让渡，不是退化：能力换了载体，资源被释放，横杆升高。在个体层，让渡与退化的分界，不在能力降没降，而在判断权还在不在手里；只使用不碰撞的人，在让渡期里确实承受真实的损失，这一点必须承认。避免被封装的办法，是在关键处保留自己的判断：让渡执行，不让渡判断。

> 带走的话
>
> 别只问能力降没降，要问判断权还在不在你手里。执行可以交给导航，“它错了我看不看得出来”这一句，不能交。

> **小账** 这一章立了什么？——让渡与退化的分界在判断权：执行交出、判断握在手里是让渡，两样一起交出是被封装；同时如实记下让渡期里个体的损失。 错在什么情况下？——导航出错时多数用户当场就能察觉改道；或能力退场后使用者既无新做法也无新判断（那就是退化）；或习惯在关键处核对的人，并不比完全依赖的人更早发现工具出错。

承接本章，下一章讲：第二个反驳，毁灭性工具，提升智慧是否等于没有代价。

第 30 章

反驳二：毁灭性工具

◆ 一、一把菜刀的两种用处

先从厨房说起。

一把菜刀，能切菜，也能伤人。没有人会因此说"菜刀是坏东西"，也没有人会说"有了菜刀，人类就更安全了"。谁都明白：菜刀封装了"锋利地切开东西"这一段智慧，这段智慧本身不分善恶，善恶出在使用的人和使用的情境。

这个常识，对大多数工具够用。但有一类工具，常识不够了：它们的破坏力大到，一次失误就可能不可挽回。核武器是最典型的一个。

于是有了本编最后一个，也是最沉重的反驳：你说工具必然逼出智慧的升级，可是有些工具是毁灭性的。它们逼出来的，难道也是智慧吗？如果一个工具能毁掉一切，"工具提升智慧"还有什么意义？

这个反驳很重，本章认真地接。

◆ 二、先划清：命题说了什么，没说什么

回答之前，先把本书的命题说准。很多误读出在这里。

本书的命题是：工具的使用必然逼出封装之外的新SDE，因而智慧与创新必然升级。

它没有说：

- 工具必然无害。
- 升级必然平稳、没有代价。
- 升级必然来得及。
- 每一代人、每一个人都会从升级中受益。

它说的是一种结构性的必然：碰撞会生出新的东西，新的东西需要新的智慧去回应。回应得好不好，是另一回事。

所以面对毁灭性工具，本书不需要否认它的危险，更不需要说"没事的，智慧会跟上来"。本书只说：这样的工具一旦出现，也会逼出封装之外的新SDE，也要求新的智慧去回应。 这个回应是否足够、是否及时，命题不保证。

◆ 三、核武器逼出了什么

先用事实性的描述，看核武器出现之后发生了什么。这里只谈结构，不谈当代的政治争议，也不评价任何国家的任何决策。

核武器封装了一段关于原子核能量释放的智慧：把物理学上关于裂变与聚变的知识，变成了可以投放的装置。这一段智慧出生在现实界。然后使用即碰撞。核武器出现之后，人类面对了一个此前从未有过的处境：存在一种武器，使用它的后果，可能超过使用它想得到的任何东西。这个处境在封装的那一刻并不存在。这就是错位：封装的是"怎样做出这种武器"，没有封装、也不可能封装的，是"在有这种武器的世界里，该怎么办"。

这个问题逼出了一大批此前没有的东西。

在理念界，出现了威慑理论：战略学者开始系统研究，一种你不打算使用的武器，怎样通过"让对方相信你可能使用"来起作用。可信性、相互确保、升级阶梯、危机稳定性，这是一整套在核武器出现之前不存在的概念。这些概念的细节与评价，本书不展开。

在现实界，出现了军备控制和核查的制度：条约、监测、国际机构。它们要解决的问题是：怎样在互不信任的各方之间，建立可验证的约束。这同样是新问题催生的新做法。

在自我界与伦理领域，出现了核伦理：什么样的武器，即使有能力造，也不应该使用？科学家对自己工作的后果负有什么责任？一个人的职业能力和他的道德责

任，应该怎样安放？这些问题此前也有人想过，核武器让它们变成了许多人必须面对的问题。

这三类东西，正好三界各一份：理念界的威慑理论，现实界的制度，自我界的伦理反省。三界各有新生，在这里也成立。

◆ 四、逼出智慧，不等于换来平安

但必须马上说下一句话，否则这一章就是在粉饰。

这些新生出来的智慧，并没有把风险消除。威慑理论建立在一个脆弱的前提上：各方都理性，信息都准确，没有人犯错。军备控制的制度，有时进展，有时倒退。伦理反省，没有阻止任何一个人去做该做的事或不该做的事。

更要紧的是：这种升级是有代价的，而且有时，代价是在升级之前付的。 人们在对核武器的后果有概念之前，已经亲眼看到了它的后果。智慧的升级是被逼出来的，而逼的方式可能非常残酷。

技术治理里有个有名的说法，叫科林格里奇困境（David Collingridge，1980年）：一项技术的后果，要等它铺开才看得清；铺开之后，又难以改变。它与本书相合，还可以再往里补一层：这里的"看不清"，正是错位的后果，因为"有了这件工具的世界"在封装的时候并不存在，没有人能预先站在里面看。

所以本书对这个反驳的第一个回答是：承认。 提升智慧不等于没有代价，有些工具逼出的升级，是以巨大的代价换来的。命题讲的是智慧与创新必然升级，它不是"工具必然无害"，不是"人类一定平安"，更不是"总会好起来"。这一点我宁愿说得重一些：如果有读者从本书读出了"所以不用担心"，那是读错了。

◆ 五、那么，命题还剩下什么

承认了这些，命题还剩下什么？有三样东西。

第一，一个关于方向的判断。 即使是毁灭性的工具，带来的也是对智慧的升级要求，而不是降低要求。有核武器的世界，需要比没有核武器的世界更高的智慧

才能安全地活下去：更精细的沟通，更严格的制度，更清醒的自我约束。这和"能活下去"是两回事，但指出了方向：这类工具不是在逼人变笨，而是在逼人变得更不能出错。这个判断可以错在哪里？如果出现一种破坏力极强的工具，而使用它的世界不需要更严的制度、更清醒的约束，也能安然无事，它就错了。另一种错法是：如果一种强工具出现之后，三界中至少两界没有出现任何新的制度、概念或伦理反应，"毁灭性工具同样逼出新生"也就被推翻了。

第二，一个对"AGI会不会毁灭一切"这类担忧的定位。 本书不讨论人工智能风险，也不在这里评估任何风险。但从结构上可以确定一点：对任何强大的新工具，包括强大的AI，"它是否有危险"和"智慧是否必然升级"是两个独立的问题。回答"有危险"，推翻不了升级命题；回答"必然升级"，也不意味着"没有危险"。把这两个问题搅在一起，是很多讨论失去焦点的原因。

第三，一个关于责任的提醒。 既然升级是被逼出来的，而且可能代价沉重，那么越早意识到"新SDE一定会出现"，越有可能提前准备，而不是等事情发生了才被动应对。这是本书的一个实践意义，也是全书最后一编会回到的话题：缩短等待期，让更多人更早地进入碰撞。

◆ 六、两个日常类比：厨房里的火，与把电门交给谁

为了让这一点更好懂，再回到日常。

火是人类最早的大工具之一（第10章讲过）。火封装了热和光，也同时带来一种危险：房子可以被烧掉。

有了火，逼出了什么？逼出了防火的知识：灶台砌在哪里，柴放多远，睡前要不要熄灭余火；逼出了邻里的约定：谁家着火，大家去救；逼出了一整个行业：消防队、保险、建筑防火规范；也逼出了每个家庭里一条朴素的家规："火边不离人。"

这是不是智慧的升级？是。是不是没有代价？不是。每一条防火规则的背后，往往是一场火灾。用火的智慧是被烧出来的，这几乎是所有强工具的共同故事。核

武器是火的极端版本：破坏力大到，我们不能再用"烧过之后再学"的方式。所以面对这一类工具，人们会强烈希望"在出事之前就想明白"。这个希望本身就是智慧升级的一个表现，但它能不能实现，没有人敢打包票。

家里的电也是个好例子。电给我们灯光、冰箱、洗衣机，同时也带来触电和起火的危险。电进入家庭之后，逼出了插座上的保护门、墙里的漏电保护开关、"湿手不碰插头"这条每个孩子都被叮嘱过的家规，也逼出了电工资格和施工规范。

我们今天住的房子，是前人用无数次教训"升级"过的结果。

这个例子让人看到毁灭性问题的一般形状：工具越强，"用不好的代价"越大，对应的新智慧越不能晚来。核武器在这条光谱的最远端，却和插座的保护门是同一个结构。毁灭性不是例外的类别，而是同一个机制在极端强度下的样子。

◆ 七、这个论证的边界在哪里

最后一章的边界交代，要更老实一些。

第一，本章不能证明"毁灭性工具最终一定被人类驾驭住"。命题没有这个含义。从逻辑上说，智慧的升级和文明的存续是两件事，后者没有被本书的论证保证。如果一个文明毁于自己的工具，它在毁灭之前的升级，也就没有机会继续了。这一点，本书不绕。

第二，本章没有对任何具体的武器政策、国际争端或当代的安全问题发表意见。核武器在这里只是结构上的例子，说明"毁灭性工具也会逼出新生，但不保证平安"。

第三，"威慑理论""核伦理"等，只是被拿来说明"三界各有新生"，不是在为它们的合理性背书。这些思想本身是否站得住，是学术和政治上另外的争论。

第四，毁灭性工具的存在，给本书的整体论证留下一个开放的问题：升级的"速度"和工具的"破坏力"之间，是否存在不可逾越的差距？如果新工具的破坏力增长，快过智慧的升级，会怎样？这是一个真正的问题，本书没有答案，只是把它放在这里。

◆ 八、本编的收束：三界交织

前三编分别看了三条路，第六编把三条路拧成了一股。第27章讲，工具从不只停在出生的那一界，因为封装有边缘，而人没有界墙。第28章讲，三界同步升级是结构上的必然，是被逼出来的，但这个必然在文明层成立，在个体层要靠碰撞。第29章讲，被封装能力的退场是让渡，不是退化，分界在判断权，但让渡期里被封装的人确实承受损失。第30章讲，毁灭性工具同样逼出新生，但提升智慧不等于没有代价。

这四章合起来，给出一个比较诚实的画面：工具与智慧的关系，不是一条平滑向上的曲线，而是一个充满错位、代价和不确定性的过程；但在这个过程里，有一个结构是稳的，那就是——**一界有工具，三界有新生**。

◆ 九、小结

毁灭性工具并没有推翻"工具逼出智慧升级"的命题，但它清楚地划出了命题的边界：命题讲的是升级的必然，不是安全的必然。核武器逼出了威慑理论、核查制度和核伦理，三界各有新生；但这些新生并没有消除风险，其中很多是以沉重的代价换来的。命题的价值，不在于保证一个好结局，而在于指出一个方向：新工具要求的是更高的智慧，而不是更低的智慧。

> 带走的话
>
> 本书的命题讲的是升级必然，不是平安必然。每条防火规则背后往往是一场火，只是核武器这把火，我们已经没法再靠“烧过之后再学”。

小账 这一章立了什么？——毁灭性工具也逼出新生（威慑理论、核查制度、核伦理），要求的是更高的智慧而不是更低；但它不保证平安，也不保证来得及。 错在什么情况下？——某种破坏力极强的工具出现后，世界无需更严的制度和约束也安然无事；或强工具出现后，三界中至少两界没有任何新的制度、概念或伦理反应。

承接第六编，下一编讲：智慧不是库存，创造与智慧如何共生。

第 七 编

智慧不是库存

> 本编只问一件事：智慧是不是一间可以装满的仓库。答案分四章慢慢给，一句话放在最后。

编　序

这一编只问一件事：智慧是不是一间可以装满的仓库。一共四章，一章一步，最后落在一个能拿来用的判据上。

第31章把两幅图景并排摆出来：库存图景说智慧是存货、创造是取货重组，共生图景说创造与智慧一起发生、没有先在的仓库。第32章回到凌晨三点的哭声和锅里的第一勺糖，让共生图景有日常的样子。第33章把它写进SDE的语言，给“智慧”找到位置：创造沉淀下来的显露。第34章给出判据：创造的时候，创造者自己有没有被改变，并正面说它对当前大模型意味着什么。

本编站在第三层论证上，也就是“没有可装满的智慧”，主场是第六条承重命题：创造力与智慧共生，智慧不是先在的库存。它只负责把这一条立稳，不单独证明终局AGI不可能；那是后面结构证明的事，本编给它打地基。

章	一句话
第31章	库存图景与共生图景，差在起点
第32章	新手父母和番茄炒蛋里，办法与人一起长出来

章	一句话
第33章	智慧是S，但永远带着出处
第34章	创造的标志在创造者，不在作品

第 31 章

两种图景：取货重组，还是一起发生

◆ 一、储物间里的聪明

搬进新家的第一个周末，几乎每个人都会干同一件事：整理储物间。螺丝刀放左边，胶带放中间，不常用的电钻放最高的架子。整理完，你站在门口看一眼，心里有一种踏实感：东西都在，要用的时候，伸手就能拿到。

很多人想象自己的头脑，就是这样一间储物间。学过的知识、攒下的经验、练出来的手艺，都是一件一件的货，码在架子上。遇到问题，就去架子上找合适的货，拿出来，拼一拼，问题就解决了。聪明的人，是架子多、货全、找得快的人。创造呢？创造不过是把架子上本来没有拼在一起的两件货，拼到了一起。

这是一幅很自然的图景，自然到我们几乎不觉得它是一幅图景，而以为它就是事实。本书把它叫作库存图景：智慧是存货，创造是取货重组。

与它对立的，本书叫作共生图景：创造力与智慧不是货与取货的关系，而是同一件事的两面，一起发生，没有一间先在的仓库。

这一章把两幅图景并排摆出来，看它们各说了什么、差在哪里，以及为什么这个差别大到足以决定"AGI能不能装满"。后面三章再分别从日常、从SDE的写法、从一个可操作的判据，去落实共生图景。

先交代一句SDE，因为本编要反复用到。SDE是三个字母：S是结构，本书读作"显露"，也就是某件事成了看得见、摸得着、可以重复的样子；D是差异，是从没有到有、从这样到那样的发生过程；E是纠缠，是这件事扎根的土壤，是它与周围的人、物、问题、需要缠在一起所形成的意义场。一个桌子，是S；木匠把它做出来的那些刨、凿、试、改，是D；它为什么要被做出来、放在哪个房间、给谁用，是E。三者缺一不可，而且彼此是一起发生的，不是先有一个、再来另一个。

◆ 二、库存图景说了什么

库存图景至少包含三个判断，它们常常不被说出口，却在默默起作用。

第一，智慧可以先于使用而存在。架子上的货，在你取用之前就已经在那里了。智慧是"有了"的东西，用，只是把它调出来。

第二，智慧是可数的，至少原则上可数。货可以盘点。一间仓库再大，也有边界；货再多，也有个总数。所以才会有"人类的全部知识"这样的说法，好像它是一个可以被抄录、被打包、被搬运的总量。

第三，创造是组合。新东西来自旧货的新搭配。搭配的可能性可以很大，大到穷人一生也试不完，但它终究被货的种类框住了。

把这三条放进当代的AI话语里，会得到一套非常顺手的说法：训练，是进货；模型，是仓库；使用，是取货；生成，是重组。仓库越大、取货越准、重组越巧，系统就越"智能"。于是一个自然的推论出现了：只要仓库装得足够满，货足够全，系统迟早会有人类能有的一切智慧，也就是"通用"了。AGI，就是仓库装满的那一天。

注意，这个推论从头到尾没有哪一步是荒唐的。它是严密的。严密，恰恰说明问题不在推理，在起点：起点那幅图景，对不对？

第2章已经把这幅图景作为主流AI想象的哲学底牌展开过，这里不重复。本章不去批评它，只把另一幅图景老老实实画出来，让读者自己比。

◆ 三、共生图景说了什么

共生图景也有三个判断，正好与上面一一对着。

第一，智慧不先于使用而存在，它在创造的过程中发生。一个人遇到一件没遇到过的事，办法不是从架子上取下来的，是在他与这件事的纠缠里长出来的。长出来之前，那个办法不在任何地方，既不在他的头脑里，也不在书里。

第二，智慧不是一个数目。你可以数一间仓库里有多少件货，却没办法数一个人"会多少种变通"，因为变通的种类是随着他遇到的事一起增加的，事不先来，变通就不先在。

第三，创造不是组合，是发生，而发生会改变发生者。创造的人在创造的过程中，同时变成了一个不同的人：他看事的方式、手上的分寸、心里的标准，都多了从前没有的东西。

三条合起来，落在一句话上：**智慧不是存下来的答案，而是被答案改造过的提问方式**。答案可以抄，可以存，可以寄给别人；提问方式不行，它是做出那个答案的人身上留下的痕迹。下一章里，新手父母到第二周能听出哭声的层次，炒蛋的厨师从此对酸里的那点甜格外敏感，说的都是这件事：他们手里并没有多出几条答案，是他们问世界的方式换了。

这就是"共生"：创造一次，智慧多一块；智慧多了一块，下一次能问的问题又不同。两者不是仓库和取货员，更像河道和水，水流过去，河道被改了，改了的河道，决定下一股水的走法。

◆ 四、一个小对比：同一把钥匙，两种讲法

用一件小事把两幅图景放在一起看。

一位老师傅修一台老式收音机，不亮。他拧开后盖，看了一眼，说：这儿的一个元件老化了。换上去，亮了。

库存图景这样讲：老师傅脑子里有一批故障与修法的对照，他看到现象，匹配到条目，取出修法，执行。这个讲法没有错，大多数日常修理确实就是这样。

共生图景会追问：那批"对照"是哪里来的？每一条，都是他某一次面对一台不亮的收音机，试了几种办法，错了几回，才有的。更要紧的是，今天这台收音机，和他从前修过的每一台都不完全一样。他能认出"这儿的元件老化"，不是因为从前见过一模一样的，而是他的眼睛已经被从前的每一次改造过了，能在不一样里看出相近。今天这一次修完，他的眼睛又被改造一次。

两种讲法都能解释"修好了"。区别出现在两个地方：一是遇到真正没见过的故障时，库存图景只能说"没有对应条目，无法解决"，共生图景则说"他会开始一次新的发生，在摸索中长出办法，同时长出他自己的新东西"；二是修完之后，库存图景说"他的库存没变，只是用了一次"，共生图景说"他已经不是修之前那个人了"。

这两个区别，正是后面要用的判据的雏形。

◆ 五、为什么这个区别决定了AGI能不能装满

回到全书的问题。如果库存图景成立，"装满"在原则上是个有意义的目标，剩下的只是工程：货很多，但数得出；仓库很大，但建得成。

如果共生图景成立，"装满"就没有对象。理由是一座具体的桥：仓库里装的是答案，而每装进一个答案，提问的人就被它改造一次，能问出的问题随之换了一批。仓库追的，是一份会被它自己改写的试卷。你可以在某一刻，把已经发生过的智慧封装起来，那是工具，而且是极好的工具；但封装下来的，永远是对昨天那些问题的答案，明天的问题还没有被问出来。

这就接上了全书第一条承重命题：工具是已发生智慧的封装。封装是库存的一种，而且是库存最好的样子。本书从不否认库存的价值。它只是说：库存之外，还有一个不断发生的部分，而且那个部分不是库存的"剩余"，而是智慧的主体。

承重命题第六条把这层意思说得最直接：创造力与智慧共生，智慧不是先在的库存。

◆ 六、反驳一：我们不是明明存过东西吗

有人会说：你讲得好听，可人确实在存东西。孩子背九九乘法表，医生背药典，律师背法条，学者读了一辈子书，脑子里满满当当。这不是库存是什么？

人确实在存。库存图景的错，不在于说"有库存"，而在于说"智慧就是库存"。九九乘法表，是一件已经发生过的智慧被封装进了孩子的记忆里，孩子拿来就用，省了每次从头数的力气。这个封装是好东西，让渡出了算的力气，让孩子腾出手去

学更大的数学。但孩子的数学聪明，并不在那张表里；那张表只是让他能够走到下一步的台阶。真正的数学聪明，是他有一天看着一道没见过的题，忽然发现可以这样想。那个"这样想"，表里没有。

再往深一层：脑子里的"存货"并不像仓库里的货那样静静待着。同一条知识，在不同的人身上，因为经历不同，是不同的东西；在同一个人身上，今年和十年前，也是不同的东西，因为它周围缠绕的东西变了。被存下来的东西，仍在不断地被重新发生；把它当成一件不变的货，是把一个正在发生的过程，当成了静止的物。

◆ 七、反驳二：重组也能造出"从没有过的东西"

另一个反驳来自库存图景的内部：就算智慧是库存，重组也能造出世界上从没有过的组合。围棋的下法比宇宙里的原子还多，AI走出了人类几千年没走过的一步，这不就是创造吗？

这里要分清两件事。

一件是在一个已经划好的空间里，找到空间里从没被走到的点。围棋棋盘是封闭的，规则固定，局面有穷，所以再大的数目，也是在一个已经被定义好的空间里数。在这类空间里，库存加重组，加上足够的搜索，可以走得很远，甚至永久越过人。这一点，本书不回避，第八编会专门讲象棋这个反例，并承认它。

另一件是空间本身是怎么来的。棋盘、规则、什么叫赢，是先有人发生出来的。把这些划好之前，"走到某个点"这件事无从谈起。库存图景擅长解释前一件，而对后一件，它只能说"那也是更大仓库里的货"。可是更大的仓库，也是被划出来的，划它的那一刻，又是一次发生。

所以，说重组能造出从没有过的组合，是对的；说这就是创造的全部，是错的。它是在已发生的空间里的新点，不是空间本身的新发生。共生图景关心的是后者。

这个区分不是本书发明的。认知科学家博登（Margaret Boden）早就把创造力分成几种：在既有的概念空间里组合、探索，和改变这个空间本身，她把后一种叫转换性的创造。本书与她一样，不肯把后一种归并进前一种；不同的是，她主要看作品，看空间被改了没有，本书要看做的人：空间被改写的同时，走进去的人有没有被改写。这个差别，到第34章会变成一把尺子。

◆ 八、这个论证的边界在哪里

要诚实交代四点。

第一，共生图景不是被证明的定理，而是一个更好的描述。它要在后面几章里，通过日常案例、SDE的写法和一个可操作的判据，逐步显出解释力。读者完全可以不接受，但需要拿出另一幅图景，解释同样的现象，尤其是"创造者在创造中被改变"。

第二，这幅图景的零件，隔壁早有人在用。杜威在1938年的《经验与教育》里讲过经验的连续性：一次经验改变经历它的人，也改变后来经验的质量。波兰尼讲过人知道的比说得出的多，皮亚杰讲过认知结构在应对新情形时被改造，生物学家考夫曼等人在2012年的一篇论文里，对生物圈做过相近的论证：下一步可能出现什么，无法事先列成清单。这些本书都受益。本书多走的一步，是把它们拧成一条专门针对"装满"的论证，并在后两章变成对封装、对当前系统的检验。若这一步只是旧说法的重新包装，批评就是公平的。

第三，库存图景在很多局部是对的。封闭、规则固定、重复性的任务，库存加重组就是最好的做法。本书反对的，只是把库存当作智慧的全部。

第四，什么情况下这一章错了。如果能指出一个不随使用而增长的"题库"，比如证明某个人类领域里所有有意义的问题，在一段时间内已被穷尽，此后没有人提出新的问法；或者有一个只靠存量的系统，在训练之后才出现的新类型问题上（不是旧题换了数字），稳定地不输给那些被这类问题改造过的人，那么"提问方式被答案改造"这一条就站不住。反过来，这一章也不证明AGI不可能：要证明终局AGI

不可能，需要第九编的结构证明，本编只负责把"智慧不是库存"立稳，让后面的证明有地基。

◆ 九、小结

两幅图景，一句话各自概括：库存图景说，智慧是先有的货，创造是取货重组；共生图景说，智慧是被答案改造过的提问方式，创造的人同时被改变。前者让"装满"成为可能的目标，后者让它没有对象。

本编接下来的三章，各走一条路去落实共生：第32章回到两个最日常的场景，第33章写进SDE的语言，第34章给出一个可以拿来用的判据，并毫不回避地说明它对今天的技术意味着什么。

> 带走的话
>
> 下次再听人说“装满”，先问一句：装的是答案，还是提问的方式？答案可以存进仓库，提问的方式只能从做过的事里长出来。

> **小账** 这一章立了什么？——智慧不先于使用，不可数，创造会改变创造者；库存图景和共生图景并排摆出，后者让“装满”没有对象。 错在什么情况下？——有不随使用增长的题库，或只靠存量的系统在训练之后才出现的新类型问题上稳定不输给被改造过的人。

承接：下一章从两个最普通的场景说起，一个是新手父母的第一个夜晚，一个是番茄炒蛋里的第一勺糖。

第 32 章

日常的共生：新手父母的第一个夜晚，番茄炒蛋里的第一勺糖

◆ 一、凌晨三点的哭声

孩子出生后的第一个夜晚，对许多父母来说，是人生里最长的一夜。

白天在医院，有护士在身边，有人示范怎么抱、怎么喂、怎么拍嗝。回到家，门一关，灯一暗，屋里只剩下三个人：一个刚学会当父母的男人，一个刚学会当母亲的女人，一个刚学会呼吸的小人。

凌晨三点，孩子哭了。

这对父母并不是没有准备。他们读过育儿书，存过一份清单：饿了会哭，尿布湿了会哭，困了会哭，肚子胀气会哭，太热太冷会哭。他们照着清单一项一项排查：喂了，不吃；换了尿布，是干净的；摸一摸，温度正常；拍嗝，没有。孩子还在哭，而且越哭越急。

站在库存图景的角度，这个场景很容易解释：清单里的条目都试过了，没有命中，说明这是清单之外的情况。库存耗尽，只好求助，打电话给长辈、搜索论坛、等天亮去医院。

可是真实发生的事，通常比这丰富得多。

爸爸把孩子竖着抱起来，贴在肩上，来回轻晃，哭声小了一点。他不知道为什么有效，只是感觉到了。妈妈换了个姿势，改成横抱，手托着后脑，同时压低声音哼了两句不成调的歌，孩子慢慢安静下来，眼睛睁着，望着天花板。两个人对视一眼，谁也说不清刚才发生了什么，但都知道了一件之前不知道的事：这个孩子，在这种哭法的时候，需要的是被竖着抱，被压低声音哼。

这条"办法"，不在育儿书里。育儿书里有类似的，但没有这个孩子的。它在这一夜里，在两个大人与一个婴儿的纠缠中，长了出来。

◆ 二、他们被改变了

再往下看一层。

到了第二周，这对父母已经能分辨三四种哭声。同样是哭，有的是饿，有的是困，有的是"我想要你在这儿"。他们自己也说不清怎么分辨的，大概是音调的起伏、哭前的小动作、抱起来时身体的僵硬程度。他们的耳朵、手臂、睡眠节奏，甚至半夜里的反应速度，都与怀孕之前不同了。

更微妙的是，他们对"什么叫一个好父母"的标准，也变了。头一周，好父母是不出错的人；到了第三周，好父母是在孩子哭的时候不慌的人。这不是从书里背来的，而是在一次次凌晨里长出来的。

这里有一个很清楚的画面：创造了一个办法的人，同时成了另一个人。新手父母在那个夜晚所做的，既是"解决了一个问题"，又是"变成了父母"。这两件事不是先后关系，是同一件事。如果硬要把它们拆开，说"先有一个叫'育儿能力'的东西，他们从中取了点用"，那就和事实对不上了：在那个夜晚之前，他们没有那份能力，是夜晚把它给了他们。

这就是共生。创造力（想出竖抱加哼歌的办法）和智慧（成为一个会听哭声的人）一起发生，没有先在的仓库。

◆ 三、把它和SDE对一对

按本书的读法，SDE是三个一起发生的东西：S是显露，D是差异和发生的过程，E是纠缠与土壤。

放到这个夜晚：

E是什么？是这三个人缠在一起的整个局面：这个孩子的脾气，这对父母的焦虑和爱，这间屋子的温度，这一夜的疲倦，以及"我要让他不再难受"这个真真切切的需要。没有这个需要，一切都无从谈起。

D是什么？是试：竖抱、横抱、轻晃、哼歌，一次一次地改，一次一次地看孩子的反应。差异就是每一次尝试与上一次的不同，发生就是在这些不同里，一个办法慢慢浮现。

S是什么？是浮现出来的那个东西：一套对这个孩子管用的做法，和一对已经能听懂这个孩子的耳朵。它稳定下来了，下次再哭，爸爸不用再从头试。

注意，S不是预先放在某处等着被取出的，它是D在E里走过之后留下来的显露。S是显露的稳定态，是发生沉淀下来的样子。这句话，下一章会专门讲。

◆ 四、另一口锅：番茄炒蛋里的第一勺糖

换一个场景，从夜里搬到厨房。

先声明：这里不讲番茄炒蛋的真实起源，那是菜史的事，本书不下断语。这里只设想一个普通场景，把道理摆清楚。

设想一位家常厨师，把番茄和鸡蛋炒在一起。他炒出来一盘，尝一口，酸味有点冲，蛋的软和番茄的汤汁混在一起，味道有点生硬。

这个时候，库存图景会怎样描述下一步？厨师的头脑里有一个"调味条目库"：酸味过重，对应的货是糖。他取出糖，加一勺，重组完成。

但真正的厨房里往往不是这样。他很可能是随手抓了一点糖，也许是试探，也许是一时兴起，撒进去，翻两下，再尝。这一口，味道忽然圆了：酸没有被盖住，反而变得更像番茄的本味，鸡蛋也跟着甜润起来。他停了一下，才第一次明白：酸和甜不是对头，在这盘菜里，甜是让酸立得住的东西。

这个"明白"，是在那一勺糖加下去之后才有的。加之前，他并不知道。如果他事先就知道，那一勺就不是创造，而是照做。

再看他被改变了什么：他的舌头，从此对"酸里要有一点甜"更敏感了；他的手，对糖的分量多了一点分寸；他对"这盘菜应该是什么味"的标准，也有了一个从前没有的版本。他下一次再炒别的酸味菜，比如糖醋汁，或者炒酸菜，这份新的味觉会自己跑出来。

第一勺糖，不是从仓库里取出来的，是从他与这盘菜、这口味觉的纠缠里长出来的。长出来的同时，他也长出了一个新的味觉。

◆ 五、后来：配方出现，封装完成

故事没有停在那一勺。

当这个做法被别人看到、学到、写下来，它就成了配方：番茄几个，鸡蛋几个，糖几克，盐几克。配方是什么？是这一次发生的封装，是已发生智慧压缩成可以传递的样子。有了配方，一个从没下过厨的人，也能炒出七八成像的一盘。这是工具的力量，本书前面各编讲的就是这个。

可是配方一经流传，接下来的事也很熟悉：有人嫌甜，把糖减半；有人加一点番茄酱，让颜色更红；有人发现先炒蛋、后炒番茄，口感不同；有人在糖之外还试了一点点醋，反过来提鲜。每一位都从配方出发，却没有停在配方里，而是在自己的锅、自己的口味、自己的饭桌上，发生了配方里没有的东西。配方封装的是第一勺糖那一刻的智慧，对后来这些人来说，是台阶，不是天花板。

这就是"封装的是昨天，逼出的是明天"的最小模型：配方让更多的人不必重新发明第一勺糖，于是他们的注意力，被腾出来，去碰撞出配方之外的新味道。

◆ 六、两个场景放在一起看

新手父母和炒蛋的厨师，一个在夜里，一个在厨房，一个关乎人，一个关乎菜，但结构几乎一样：

第一，起点是一件没有现成答案的具体事：这个孩子的哭，这盘菜的酸。　第二，办法是在试的过程里长出来的，不是从某个地方取出来的。　第三，长出办法的同时，做的人也变了：耳朵、舌头、手感、标准。　第四，办法稳定下来之后，可以被封装、被传递：育儿心得、配方。　第五，接过封装的人，面对的是又一件没有现成答案的具体事：下一个孩子，下一盘菜，于是发生再来一轮。

这五步，在第三到第五编的工具史里一再出现，只是尺度从一个家庭放大成了一个文明。

◆ 七、一个追问：这不就是"积累经验"吗

读到这里，有人会说：你讲的这些，不过是积累经验。经验多了，办法就多了，这不还是库存吗?

这个追问要认真回答，因为它几乎是库存图景最体面的退路。

经验当然是积累的，但"积累"这个词有两种意思。一种是堆：今天放一件，明天再放一件，货架上多了东西，每件货原样不动。另一种是长：今天的经历改变了你看事的方式，明天你看到的东西，已经和昨天的你看到的不同。

新手父母的第二周，和第一周的区别，不是多了几条"哭声对照"，而是他们听的方式变了。同样的哭声，第一周的耳朵听到的是一片噪音，第二周的耳朵听到的是有层次的信息。这不是在原来的耳朵上多挂了几件货，是耳朵本身被重新训练了。

炒蛋的厨师也是一样。他不是多记了一条"酸加糖"，而是舌头对酸甜关系的感受发生了位移。以后他尝任何一道菜，这个位移都在起作用，哪怕他根本没意识到。

所以，把共生说成"经验积累"，只说对了一半。堆起来的那一半，是库存；长出来的那一半，是共生。库存图景的毛病，是只看见堆，看不见长。而恰恰是长，决定了下一次面对新局面时，你是只能翻货架，还是能长出新办法。

再补一点：堆出来的东西，可以原样复制给别人，长出来的东西，不能。你可以把"孩子哭时试试竖抱"这句话告诉朋友，朋友也能照做，但他听哭声的耳朵，只能靠他自己的几个夜晚。话能封装，耳朵不能。这就是为什么育儿心得再多，每一对新手父母仍然要过一遍自己的第一个夜晚。

◆ 八、这两个场景的边界在哪里

要老实地说，这两个场景能证明什么，不能证明什么。

它们能说明：在日常经验中，创造与智慧的增长确实常常是同一个过程；"办法从仓库里取出来"这种说法，不能完整描述这类经验。

它们不能证明：所有智慧活动都是这样。有些活动确实更像取货，比如照着说明书装一个书架，或者背一段已有的口诀。这些活动里也有一点点发生，但那一点点很小。把它们当主要样本，会得出库存图景；把新手父母的夜晚当主要样本，会得出共生图景。选哪个当样本，本身就带着立场。本书的理由是：凡是开放的、没有现成答案的局面，才是检验"通用"的地方，所以取这类样本。这一点第八编会进一步交代。

它们也不能证明：机器不可能在类似的局面里发生同样的事。那要留到第34章，用一个可操作的判据来谈，而不是靠两个故事下结论。

类比还有风险。一个孩子的哭，与一个文明面对的难题，量级不同。用这两个场景，是为了让读者在最熟悉的地方看清结构，不是说文明难题等于哄孩子睡觉。

什么情况下这一章错了。这个说法原则上可以检验：找两组新手父母，一组严格照清单排查，另一组自己摸索，到第二周，请不知情的评判者让他们各自听同一批婴儿哭声的录音，分辨哭的原因。如果两组分辨得一样好，"耳朵是在夜里长出来的"就站不住。我没有引用现成的研究，只摆出检验的样子。

◆ 九、小结

凌晨三点的哭声里，办法和父母一起长了出来；锅里的第一勺糖里，新口味和新味觉一起长了出来。创造与智慧是同一件事的两面，没有一个先在的仓库在后台等着被取用。配方、心得，是这些发生的封装，不是终点，是下一轮发生的台阶。

带走的话

话能封装，耳朵不能。育儿心得再多，每对新手父母还是要自己过一遍第一个夜晚，因为听哭声的耳朵只能在夜里长出来。

小账 这一章立了什么？——创造与智慧在日常里是同一件事：竖抱加哼歌的办法和会听哭声的耳朵一起长出来，配方是台阶，不是天花板。 错在什么情况下？——两组新手父母，一组照清单、一组自己摸索，第二周分辨哭声的本事一样好，“耳朵是夜里长出来的”就站不住。

承接：下一章把这种日常的共生写进SDE的语言，回答一个关键问题：如果S是显露的稳定态，那么"智慧"在SDE里到底站在哪个位置。

第 33 章

SDE的写法：S是显露的稳定态，智慧是创造沉淀下来的显露

◆ 一、从一条河道说起

站在一条老河的岸边，你看到河道弯弯曲曲，两岸有沙洲，水深水浅各有分寸。你可以问：这条河道是先有的，还是水流出来的？

稍微想一想就知道，两个答案都不对。河道不是先有然后水才来，水也不是先有然后找河道。水流过去，把土冲走一点、沉下一点，河道因此成形；河道成形之后，又决定下一股水走哪里。河道是水留下来的样子，是流动沉淀下来的、暂时稳定的形状。可它又不是死的：下一场洪水，它就被重写一次。

这条河是理解SDE的一个日常模型。这一章把上一章的两个场景翻译成SDE的语言，并给"智慧"这个词找到它应有的位置。

◆ 二、SDE的三个字，再用大白话讲一遍

本编前两章已经交代过SDE，这里换一种讲法，用河来比。

D是水流，是差异、是发生的过程。一处和另一处不同，才有流动；流动本身，就是发生。D不是一个状态，而是正在进行的事。

E是整个流域，是纠缠、是土壤。雨水从哪里来，山有多高，土有多软，两岸住着谁，需要引水还是防洪。E是让这条河成为"这条河"的整个背景，也是让水流有意义的那个场。

S是河道，是显露的稳定态。水流过之后，留下的形状：看得见的河床、沙洲、堤岸。S是发生的结果，也是下一次发生的条件。

这里有一个读法上的要点，也是本书与许多人习惯不同的地方：S不是先在的结构，而是显露。我们常常把"结构"想成骨架，先有骨架，然后血肉往上长。SDE

的读法倒过来：先有流动（D）在场（E）里发生，才有稳定下来的结构（S）。结构是显露的稳定态，是流动暂时停下来的样子，不是流动的前提。

◆ 三、智慧在这里站在哪

现在可以回答这一章的核心问题：智慧站在SDE的哪个位置?

库存图景把智慧放在S里，而且是先于D和E的S：一个现成的、完整的结构，等在那里，等人来取用。

共生图景给出的位置不同：

智慧是创造沉淀下来的显露。它是S，但不是先在的S，而是D在E里走过一遭之后留下的S。换句话说，智慧在SDE里是S，这没错，但它永远带着出处：它是从某一场发生中沉淀出来的，是某个人（或某个群体）在某片土壤里、经过某一串差异，留下来的稳定态。

这个定位，立刻带出三个后果。它们可以先收成一句话：智慧不只是S，更是S和它的出处之间那根还没有剪断的线。

后果一：智慧有出处，因此不是无源的货。同样一条"酸里要有一点甜"，对炒蛋的厨师是一场亲身的发现，对从配方里读到它的人则是一句条文。外表一样的S，里头的发生史不同，在后续的发生里发挥的作用也不同。前者的S，连着他的舌头、他的手、他对这盘菜的整个感觉，是"活的"；后者的S，还需要他自己去走一遍，才会变活。库存图景看不到这个区别，因为在仓库里，两件货是一样的。

后果二：智慧会被再次改写。河道不是定型的，下一场洪水来了就要重写。同样，一个人的智慧，每遇到一次真正的新事，都有一部分被重新发生。所以，不存在一个"最终的、完成的智慧"，只有不断沉淀、又不断被改写的显露。

后果三：智慧不能脱离E。河道离开了流域，就只是一道沟。智慧离开了让它成为智慧的土壤，也只剩下形式。这一点会在第九编里扩展成"问题总是对某个主体而言的问题"，现在先记下：S离不开E。

◆ 四、创造，在SDE里是什么

有了这个定位，创造也就有了清楚的位置。

在SDE的写法里，创造就是D的一次真实发生：一串真正的差异，在真实的纠缠里走过，最后留下了从前没有的显露。

"真实"这个词，要稍加解释。不是所有看起来新的东西，都是D的真实发生。把仓库里的两件货拼在一起，看起来是新组合，但这个组合如果完全落在原先已划定的可能之内，就更接近于在已有的S上挪动位置，而没有走过一段真正的差异。真实的D，需要穿过一段事先没有答案的路：试的人不知道会走到哪里，也不知道走完自己会变成什么。

所以，创造和重组的分界，不在结果新不新，而在过程里有没有一段事先没有答案的路，以及走过这段路的人有没有被它改变。结果的新旧，旁观者有时看不出来；而过程里的改变，是创造的内在标志。这是下一章判据的根。

◆ 五、用SDE重看两个场景

上一章的两个场景，现在可以收一收。新手父母的夜晚，上一章已经逐项对过：S沉淀的同时，父母变了，变化本身就是智慧的增长。

番茄炒蛋的第一勺糖：E是这盘菜、这口味觉、这位厨师的需要；D是加糖、翻炒、再尝；S是"甜让酸立得住"的新味觉，以及后来的配方。配方是S的封装，把一个人的显露变成了可以传给别人的工具。

注意这里的连接：封装，是把S从它的出处里取下来，使它可以离开原来的E去旅行。这既是工具的伟大处，也是工具的限度：被取下来的S，离开了原来的D和E，它能做的，是帮后来的人省去重走一遍的力气；它不能替后来的人走他们自己的差异。后来的人面对的是他们自己的E，需要他们自己的D，长出他们自己的S。

这句话，翻译成全书的语言，就是第二编的意思：工具封装的是昨天的智慧，逼出来的是明天的智慧。

◆ 六、一个容易混淆的地方

读到这里，有读者会问：你说S是智慧，S可以被封装成工具，工具可以被一台机器装下，那机器不就拥有了智慧？

机器（或任何工具）装下的，是S的外形，也就是显露出来的、可以被复制的那一层。外形可以被复制得非常精确，精确到几乎看不出和原件有什么不同。这是工具能力的真实范围，而且这个范围越来越大。

但智慧在共生图景里，不只是S的外形，还包括它与D、E的连接：它从哪里发生、连着谁的纠缠、还能不能被再次改写。外形被复制，不等于那根线被复制。

这并不是说机器永远复制不了连接。那是后面要谨慎对待的问题。这里只是说：当有人指着一个被装下的S说"这就是智慧"，要问一句：这个S，现在还连着一个正在发生的D和E吗？还是一个已经断了出处的外形？

河道的比喻在这里也有用：一条被水泥固定住的河道，外形和原来一样，但它不再是河了，它是一条渠。渠有渠的用处，而且用处很大，可它不会在下一场洪水里重新成形。

◆ 七、再看一个例子：记账本

换一个更朴素的场景，看沉淀怎样发生。

一个刚开始自己过日子的年轻人，决定记账。头一个月，他买了一个本子，每天写：早餐八块，地铁四块，午饭十五块。月底一合计，吓了一跳，钱不知道花去了哪里。

这时账本上的数字，是最初级的S：一堆显露出来的记录。但记账这件事，还在继续。第二个月，他开始给每一笔分类：吃、行、玩、必须的、可有可无的。分

类不是账本自带的，是他在翻看上个月的数字时，一点一点想出来的。第三个月，他发现自己买东西之前，会在心里问一句"这算哪一类"。

到了半年，他已经不太需要翻账本，就能大致知道这个月花得对不对。账本，从一件外在的工具，沉淀成了他自己的一种钱的感觉。

用SDE来读：E是他的日子，他的收入、他的愿望、他对"过好日子"的想象；D是每天记、每月翻、不断重新分类的过程；S有两层，一层是账本上的数字和分类，这是外在的显露，另一层是他心里的钱感，这是内在的沉淀。外在的那层可以复制：把他的分类表发给朋友，朋友立刻能用。内在的那层不能：朋友拿到分类表，还得经过自己的几个月，才会长出自己的钱感。

这个例子里有一件事要留意：账本这个工具，并没有替他长出钱感，而是给了他长出钱感的场地。工具提供了封装和记录，发生仍然发生在人身上。这也是第二编"工具逼出新发生"的一个小小的缩影。

顺着这个例子再想一步：如果有工具替他把账全记了、分类也全做了，他每个月只看一眼总结，会怎样？他得到了更省力的生活，这是实在的好处，也是前面说过的让渡。但钱感这一层的发生，会因此少一块土壤，要另找地方去长。每一次让渡，都在问：腾出来的力气，用去长什么？这个问题，会在第六编、第十编里反复出现。

◆ 八、这个写法的好处

为什么要用SDE这样来写，而不是直接说"创造让人变聪明"？

第一，它给了智慧一个"有出处"的位置，使"同一条知识在不同人身上不同"这样的日常观察有了解释。第二，它把创造、智慧、工具放进同一个框架：创造是D的真实发生，智慧是沉淀下来的S，工具是S的封装。第三，它解释了为什么"装满"没有对象：发生不停，沉淀就不停，没有一个时刻可以说"全部沉淀完了"。第四，它留出了检验的口子：既然创造的标志是过程里的改变，就可以问"这个过程里，谁被改变了"。这就是下一章。

◆ 九、这个写法的边界在哪里

四点交代。

第一，SDE是本书采用的一种写法，不是唯一的写法。用别的语言，也可以说出共生图景，比如说智慧是实践的产物，是在做中形成的。SDE的特点，是把"过程、土壤、显露"三者同时摆上桌，防止只看其中一个。读者如果不接受SDE的术语，仍然可以只取共生图景的核心判断。

第二，"S是显露的稳定态"是一种读法，不是对所有领域都已被证明的定理。它在本书里的作用，是一个组织材料的框架，解释力要靠前面各编的工具史去检验，而不是靠定义自证。社会学家吉登斯讲过结构与行动互相造就，智利生物学家马图拉纳和瓦雷拉提出过"自创生"，说系统自己生产自己；"结构是流动沉淀下来、又会被重写的稳定态"与它们相近。本书的用法多出的一点，是用"出处"去解释：工具为什么能把智慧取下来旅行，却取不下来它的活性。

第三，河的比喻有局限。河没有意图，人有；河的E是地理，人的E里有意义和需要。河可以帮助理解"沉淀与重写"，但不能把人的创造降格为水的流动。

第四，什么情况下这一章错了。如果出处不重要：两个外形完全一样的S，一个来自亲身发生，一个来自抄录，在两人都没见过的新情形里表现不分高下，那么"智慧带着出处"就是错的。下一章会把这个想法做成可以操作的检验。

S是显露的稳定态；智慧是创造沉淀下来的显露，带着出处，会被改写，离不开土壤。创造，是D的一次真实发生，标志是过程里有一段事先没有答案的路，以及走路的人被改变。封装，是把S取下来传递；传递出去的，是外形，连接要后来的人自己去发生。

> **带走的话**
>
> 别人给你的一条结论，像一段浇了水泥的河道：外形全在，水还得自己去流。看一份“懂”靠不靠谱，先问它还连着谁的发生。

> **小账**　这一章立了什么？——S是显露的稳定态，智慧是创造沉淀下来的显露：带着出处，会被改写，离不开土壤；封装取下来的是外形，不是那根线。　错在什么情况下？——两个外形一样的S，一个来自亲身发生，一个来自抄录，在两人都没见过的新情形里表现不分高下。

承接：下一章把这个写法变成一个可以拿来用的判据：创造的时候，创造者自己有没有被改变。并且正面讨论它对当前技术意味着什么，以及它的结论到哪里为止。

第 34 章

一个判据：创造的时候，创造者自己有没有被改变

◆ 一、两个人抄同一张菜谱

有两个人，同时拿到一张红烧肉的菜谱。

第一个人是个忙碌的上班族，周末照着菜谱，一步一步做：几克糖，几块姜，火候几分钟。做出来，味道不错，拍张照，吃完，洗碗，下周他还会照着菜谱做，做出来还是那个味道。他是个好执行者。菜谱对他来说是一件可靠的工具，他用了它，它没有变，他也没有变。

第二个人是个爱琢磨的老太太，同样拿到菜谱，头一回照着做，第二回就觉得"这个糖放早了"，第三回换了锅，第四回发现春天的肉和冬天的肉火候不一样。一年以后，她做的红烧肉，已经不是菜谱上的那一道，是她自己的。她还能看一眼别人家的红烧肉，说出它哪里缺了什么。

两个人做的是同一道菜，用的是同一张菜谱。区别在哪里？不在菜好不好吃，不在谁的手艺"更像菜谱"。区别在于：做的过程中，做的人有没有被改变。

第一个人没有变，他把菜谱"执行"了一遍又一遍；第二个人变了，她在做的过程里，长出了一个从前没有的厨师。

所以，创造的标志不在作品，而在作者：不是做出来的东西新不新，而是做完之后，做的人还是不是原来那个人。这个小小的对比，就是本章要提的判据。

◆ 二、判据是什么

判据的说法是：

> 创造的时候，创造者自己有没有被改变。

把它展开成三句话：

一、一个过程如果只是在已有的东西中取用、搭配、输出，而做的主体在过程前后没有任何持续的改变，那么它更接近取货重组，对应库存图景。

二、一个过程如果在走过一段事先没有答案的路之后，做的主体自身留下了持续的改变（看事的方式、手上的分寸、心里的标准），那么它更接近创造，对应共生图景。

三、这里的"改变"，指的是留存下来、会影响下一次的改变，不是一时的状态起伏。

这就是承重命题第六条"创造力与智慧共生"的一个可操作的版本。既然智慧是创造沉淀下来的显露，那么凡是真创造，必有沉淀，必有创造者的改变；反过来，没有任何沉淀，就谈不上走过了一段发生。

◆ 三、它为什么站得住

三个理由。

其一，它来自上一章的写法。创造是D的真实发生，智慧是沉淀下来的S。D走过，S就留下；S留下，走过的人就是不同的人。所以"被改变"是创造的内在后果，而不是外加的条件。

其二，它和日常经验吻合。学会骑自行车的孩子，不再是不会骑的孩子；写完一本书的人，不再是动笔前的那个人。我们对"某人做成了一件事"的评价，总是同时包含"他因此成了怎样的人"。

其三，它划出了一条可以讨论的线。以前讲"机器会不会创造"，往往陷入"什么叫创造"的定义之争。这个判据把问题转成一个更具体的问：这个系统在做的过程里，自己有没有留下持续的改变？这个问题，至少部分可以被观察、被讨论。

"创造看人不看作品"，心理学里并不陌生：博登区分过"对这个人是新的"和"对整个历史是新的"两种创造。本书借的是前一种眼光。

◆ 四、把判据用在人身上

先用在人身上，看它是不是公平。

一个反例马上会冒出来：人也会"取货重组"。一位律师翻出旧合同，改几个名字，就成了新合同。他有没有被改变？没有。那么按判据，这是取货重组，不是创造。这个结论是对的，没有人会认为改合同的名字是创造。

另一个反例：人受了伤，也会被改变，难道受伤是创造？当然不是。所以必须承认："被改变"是创造的必要条件，不是充分条件。创造要求被改变，但被改变不一定是创造。判据的作用，是一个筛子，筛掉那些明显没有改变发生的过程，而不是一个能证明"这就是创造"的印章。

还有一种情况：改变是浅的，几天就忘。"持续"要掂量深浅：深的改变，如眼睛被训练、标准被重建；浅的改变，如记住了一个新数字，它更接近"进了一点货"，而不是"长了一个新的自己"。判据允许有程度，不要求非黑即白。第29章讲过的伦敦出租车司机研究，是"改变"可以落到身体上的一个侧面：长年使用空间记忆的人，脑区与对照组有差别，且与从业年限相关。那是相关，不是因果，这里只取到这一步。

◆ 五、把判据用在当前的大模型上

现在到了本书必须直面的地方：这个判据，用在今天的大模型上，会得出什么？

只谈结构性事实，不谈任何具体公司和产品，也不作优劣评价。

当前大模型的一个普遍结构是：训练与使用分离。训练的时候，系统的参数在变，吸收大量已有材料；训练结束以后，进入使用，对话、生成、回答问题的时候，参数是固定的，不因一次使用而改变。使用者可以在一次对话里给它很多信息，它可以在这次对话的范围内参考这些信息，但对话一结束，它自身并没有因此留下持续的改变。要想改变它，需要另一轮训练。

用判据来看：

- 训练阶段，系统确实在被改变，但那是在吸收已有材料，近似"进货"；
- 使用阶段，系统在生成新的文字、新的组合，其中很多对读者来说是新的，但系统自身在这个过程前后没有持续的改变。

两个阶段，一个近似进货，一个近似取货重组，都在库存图景那一边。所以，就当前的大模型而言，判据给出的倾向是：它更接近库存图景。

由此得到一个可以被检验的判断（附录C的补充一）：只要训练与使用保持分离，系统在没有答案的新局面里的创造性贡献，就总需要使用它的人来承担"定题"这一环，也就是决定哪件事值得当成问题来问，并认下结果的好坏。怎样算它错：有一个这样的系统，在训练与使用分离的状态下，长期、独立地提出并维持有意义的新问题类别，而且这些问题得到了使用者以外的人的持续认可。

这句话需要立刻加三道限定，否则就读偏了。

限定一：这是当前技术状态，不是原理必然。训练与使用分离，是今天常见的做法，是工程上的选择，不是计算本身的定律。没有任何原理规定，系统在使用中不可以持续改变自己。本书没有、也不打算证明这一点。

限定二：这个倾向不是对能力的评价。说它更接近库存图景，不是说它能力弱。库存图景下的系统可以非常强，强到在许多封闭和半封闭的任务上永久超过人。它只是在说：这种强，是封装的强，而不是发生的强。

限定三：对话内部的"适应"不等于改变。系统在一次对话里，可以根据前文调整回答，看起来像是在"学"。但按判据，要问的是：对话结束后，它自身留下了什么持续的改变？如果什么也没有留下，这种适应更像是在一份固定的存货之上，带着当次的上下文做的取用。

◆ 六、如果将来系统能在使用中持续改变自己

现在回答一个必然会被问到的问题：如果将来的系统，在使用中就能持续改变自己呢?

那么，判据对它的答案就会变。

假设有一天，某个系统在面对真实的新局面时，能够在使用的过程中，让自身留下持续的、会影响下一次的改变，也就是说它不再只是"用"，而是在"用"中"长"。那么它至少在判据的这一项上，就不再被归入库存图景。本书不会为了保住结论，去否认这一点。机器学习里的在线学习、持续学习，走的就是这个方向。它们要过的，仍是这道筛子：改变是否留存，是否改变了系统下一次应对新局面的方式。

不过，这时论证并没有崩，而是退回到一个更根本的位置。那个位置，在全书里是明确立为前提的：

> 问题总是对某个主体而言的问题；E的源头在主体。

意思是：纠缠、土壤、让一件事有意义的那个场，它的源头是一个有需要、有处境、有利害的主体。孩子的哭之所以是"问题"，是因为有一对父母在乎；那盘菜的酸之所以是"问题"，是因为有一位厨师想让它更好吃。没有谁在乎的"问题"，只是一个中性的状态。

所以，即使系统能在使用中持续改变自己，问题仍然是：它的E，也就是它所面对的那些"问题"，从哪里来？是它自己的需要，还是人交给它的？如果是人交给它的，那么发生的源头仍在人那里，系统是一个越来越强的、会成长的工具。如果有一天，它有了自己的、不由人交给的需要和处境，那它就成了一个新的主体，而不再是"人类智慧的替代品"。那样的话，"AGI"这个词所设想的"装满人类智慧"的目标，也就换了题目。

这里必须说清楚："E的源头在主体"，是本书的前提，不是证明。本书不在这里论证它，第九编会把它作为前提摆出来，并说明如果有人不接受这个前提，哪些结论会跟着变。读者完全可以对它持保留意见，但需要看清楚：本书的整体论证里，哪一部分靠它撑着，哪一部分不靠它。

可以先预告一下：库存图景与共生图景之辨、封装与新发生之错位，不依赖这个前提；"谁来升级"这个问题的回答，则依赖它。

这里同样有一个可检验的说法（附录C的补充二）：即使出现能在使用中持续改变自身的系统，它所面对的问题，最终仍可溯源到某个主体的需要。怎样算它错：能够证明某个这样的系统，其问题的来源完全不依赖于任何人的需要，并且这些问题对系统自身的"存续"有可说明的意义。它被证伪时，受影响的是全书里依赖这个前提的部分，不是全部论证。

◆ 七、判据还有哪些不足

既然是判据，就要诚实说它的短处。

第一，它不容易从外部观测。一个人的改变，别人要隔一阵才看得出；一个系统的改变，要看它的内部状态是否留下痕迹，内部不透明的系统尤其困难。能做的是对照：让一个人（或系统）亲自走完一段没有答案的摸索，另一个只拿到结果，隔一阵，给二者同一批"同类、但没见过"的新题，看差别。把菜谱的故事放大，就是让照做的人和自己改了一年的人，同时去对付一种两人都没见过的食材。对人可行；对一个系统，要看能不能做同样的安排。

第二，"持续"和"深浅"没有刻度。多久算持续？多深算深？本书不给数字，因为任何数字都是编造。判据给的是一个方向，不是一把尺。

第三，它可能被"伪装的改变"绕过。一个系统可以被设计成每次使用后都记录一点东西，看上去在"改变"。那是不是判据说的改变？要看它是否真的影响了下一次对新局面的应对，而不只是多了一份可调用的记录。这又回到对照：改变的深浅，要靠它在没有答案的新局面里的表现来检验。

◆ 八、判据的边界在哪里

判据的边界，前面已经散在各处，这里收一收。它是必要条件的筛子，不是充分条件的印章。对当前大模型的判断，只靠"训练与使用分离"这一条结构性事实，而且明确是当前技术状态；事实变了，判断随之变，将来如何是开放的，本书不替它下结论。系统若能在使用中持续改变自己，论证退回到"E的源头在主体"这个前提，前提不是证明，全书有一部分论证靠它撑着，已经写明。

隔壁的说法与本书的差别。"创造会改变人"这半句，第31章提到的杜威、皮亚杰早已说过。本书新的地方，是把它当成一道筛子，去筛封装、筛当前的系统，并让筛子本身可以被检验：补充一、补充二，和上面那个对照，就是检验的入口。

整个第七编只证明一件事：智慧不是可装满的库存。它不单独证明终局AGI不可能，那是第九编结构证明的事。

◆ 九、小结

第七编的四章走了一条线：先摆出两幅图景，再用新手父母和第一勺糖让共生图景有了日常的样子，再用SDE的写法把智慧放在"创造沉淀下来的显露"的位置，最后是这一章的判据：创造的时候，创造者自己有没有被改变。

用在当前大模型上，判据只给出一个有限定的倾向：更接近库存图景，不是原理必然，也不是能力评价。

收束一句：**创造不是智慧的支出，而是智慧的发生；一个人每创造一次，就多出一个从前没有的自己。**

> 带走的话
>
> 看一个人有没有创造，别只看他端出了什么，要看他做完还是不是原来那个人。照菜谱做了一年没变的人，和改了一年的老太太，端上桌的红烧肉也许一样，人却不一样了。

> **小账** 这一章立了什么？——“创造者有没有被改变”是个筛子，筛掉没有发生的过程；用在当前大模型上，只给出带三道限定的倾向，不是结论。　　错在什么情况下？——附录C补充一、补充二的检验落空：训练与使用分离的系统长期独立提出被外人认可的新问题类别，或其问题来源完全不依赖任何人的需要。

承接：第七编到此结束。第八编转向另一条战线：既然智慧不是库存，那么"人类水平"又是什么？它是一个固定的高度，还是一个移动的横杆？

第 八 编

移动的横杆

> 本编要回答一个听上去最务实的问题：就算不谈"终局"，只谈"人类水平"，AI总该有追上的一天吧？追上了，不就是AGI了吗？本编的回答是：这个问题里的"人类水平"四个字，本身就是一个会动的靶子。

编 序

这一编回答一个听上去最务实的问题：不谈“终局”，只谈“人类水平”，AI总该有追上的一天吧？一共四章，先讲靶子会动，再讲靶子里含着追赶者，再认下最强的反例，最后划清边界。

第35章用印刷术、计算器、导航和家里的账本说明，人类水平是一个过程，不是一个高度。第36章讲横杆里含着追赶者：AI每追上一步，就成了下一步的材料。第37章正面面对象棋，承认封闭领域里AI确实永久越过了人。第38章把边界划清，说明“通用”这个词说的是开放的一边，并亮出还没补上的“谁来升级”这道缝。

本编站在第二层论证上，也就是“没有固定的人类水平”，主场是第七条和第八条承重命题：人类水平是过程，横杆里含着追赶者；封闭领域可被永久越过，开放领域不可。它回应的是“定义缝”那一边，对付温和版的“人类水平通用AI”，终局AGI的结构证明留给下一编。

章	一句话
第35章	人类水平是人、工具、难题之间的位置
第36章	追赶者是横杆的材料
第37章	象棋是真反例，也是边界
第38章	通用即开放

第 35 章

人类水平不是一个高度，而是一个过程

◆ 一、一场没人宣布的"降级"

先讲一个很多人家里都有的场景。

你的外婆，或者外婆那一辈的任何一位老人，要是被问起"谁有学问"，她的标准多半是这样的：这个人肚子里装的东西多。会背《三字经》，会背几百首诗，能把家谱从头报到尾，红白喜事上的对联张口就来。这样的人，在村里和镇上受人敬重，被称作"有学问"。

再看今天。一个有学问的人，被认为应该是什么样？大概是：他知道去哪里找东西，知道几份互相矛盾的资料该信哪一份，能把散在各处的信息拼成一个有用的判断。他不一定背得出多少，但他手里有一套检索、比较、综合的本事。如果有人今天还靠背诵显摆，我们反倒觉得那是一种才艺，像杂技，而不是学问。

这中间发生了什么？没有任何人开会宣布"学问的标准降低了"。恰恰相反，没有人觉得标准降低，大家都觉得标准升高了。一个只会背的人，今天并不被认为比从前的"状元"更差，他只是被安放在了另一个位置上。

这件小事，是本编全部论证的起点："人类水平"这个词，说的从来不是一个固定的东西。

先交代SDE，本编要用。S是显露，指已经显露出来、稳定下来的样子，比如一本菜谱、一套规则；D是差异，指从这里走到那里的过程，每一步都有点不同；E是纠缠，指把人、事、环境缠在一起的那片土壤，意义从那里长出来。一个人的"水平"拆开看，不是一个S，不是一张写着本事的清单，而是S、D、E一起运转的过程。

◆ 二、印刷术：有学问，从背诵变成了检索与综合

我们把外婆的例子放回历史里去看。

在书很贵、很稀少的时代，一本书要靠人一笔一笔抄。书抄得慢，书就少，书少，知识就只能存在人的脑子里。所以"有学问"在那个时代，几乎等于"记得多"。一个人如果能背下一部经典，他就是一个活的图书馆，别人想查，要来问他。背诵不是一种可有可无的美德，而是社会保存知识的基本办法。

印刷术普及以后，书便宜了，多了。这时候一件事情悄悄发生：把一部书背下来，从"保存知识的办法"变成了"一种个人的才艺"。因为书就在架子上，要查随时可以查。社会不再需要每个读书人都当一台活的存储器了。

但你要是以为，人因此就"懒"了，那就错了。书一多，新的难题立刻来了：这么多书，哪本可信？这本和那本说的相反，听谁的？一百本书里有一百个角度，怎么拼成一个自己的看法？于是，"有学问"的新标准不知不觉长出来了：会比较、会选择、会综合、会在许多说法之间做出自己的判断。

你看，印刷术这个工具，封装了"保存与复制文字"这件事。这是昨天的智慧：把抄书人的劳动，变成一台机器的劳动。可是它一旦被大家使用起来，就逼出了一个封装之外的东西：在海量文本中取舍与综合的能力。这个能力在抄书时代根本不构成问题，因为书太少，没得选。

所以，印刷术出现之后，"人类水平"并没有停在原地等着印刷术来追，它自己往上挪了一格：从"背得下"挪到了"理得清"。

设想有一位当年的评委，手里拿着一份"有学问"的标准，去测量一个后来的读书人。他发现后者背不出多少诗，于是下结论：此人水平不够。我们今天会觉得这位评委很可笑，可他的错误并不在于他不聪明，而在于他把一个过程当成了高度。他以为学问是一根刻着刻度的杆子，其实学问是一条会改道的河。

◆ 三、计算器：心算快，变成了建模与判断

第二个例子离我们更近，几乎每个人都经历过。

我们的上一辈人，有不少是算盘和心算的高手。菜市场上，一位卖菜的大妈，三斤二两的价钱加上一斤半的价钱，她头也不抬就报出总数，一分不差。在那个年代，这是一种让人佩服的"聪明"。会计要打得一手好算盘，店员要口算得快。算得又快又准，就是"脑子好使"的代名词。

计算器普及之后，这一块被封装了。一个几块钱的小东西，算得比最好的心算高手还快，而且永远不会算岔。于是有人惊呼：人类的算数能力完了，人要被计算器"超越"了！

可是细看之后的几十年，实际发生的事是另一回事。"脑子好使"的标准，换了内容。今天一个被认为"数学头脑好"的人，大致是这样的：他知道一件事该怎么列成算式，知道该把哪些因素放进来、哪些可以先扔掉，知道算出来的结果看上去合不合理。比如想比较两家超市的会员方案，要自己想明白：该比哪几个数，该把什么算成成本，每月多买一点是否合算。算数交给计算器，难的部分是建模和判断。

反过来想一想，今天要是有谁拿着一个计算器，却不知道该算什么，我们不会说他"有数学水平"。可在老算盘时代，一个人只要打得快，他确实被当作"有水平"。这说明，"数学水平"这根横杆，不是被计算器超越了，而是被计算器抬高了。

这里要特别注意一个细节：被封装掉的那部分——心算的速度——并没有消失，但它的地位变了。它从"人类水平"的核心，退成了一项个人爱好或者一种脑力体操。前面讲过，这叫让渡：一件事交给了工具，人腾出手来，去站到工具所撑起的新位置上。（关于让渡与"退化"的关系，第六编里已经正面谈过，这里只记住一句：退场不是退化。）

◆ 四、导航：会认路，变成了知道何时不听导航

第三个例子，也许最贴近当下。

我们的父辈开车去陌生的地方，要靠一张折叠的纸质地图，加上沿路问人，加上对方向的记忆。"会认路"是一种实实在在的本事：哪条路堵、哪条小巷能穿过去，老司机的脑子里有一张活的地图。

导航软件把这件事封装了。你只需要说出目的地，它会替你把路线排好。我们这一代人很多已经不太认路了，这是实情。

那么，"开车的水平"是不是因此变低了？有意思的是，一个今天被称作"路况经验丰富"的司机，他的本事是这样的：导航说前面要左转，他一看，今天这条路在修，左转之后是个死胡同，他知道不能听；导航把他引向一条看似最快的路，可他知道这个时段学校门口会堵，宁可绕一点；暴雨天，他知道导航算不出积水的深浅，于是选一条高一点的路。

你看，会认路变成了知道何时不听导航。这是一种新的、更精细的"路感"，它的前提，恰恰是有导航在。没有导航的时代，没有人需要练"何时不听导航"这项本领，因为根本没有可以不听的东西。

三个例子长得几乎一样：每一次，工具封装掉的是上一格里能够被固定下来的部分；人的水平并没有"被超越"，而是被重新定义到工具之上的一格。

◆ 五、为什么说它是"过程"，而不是"高度"

到这里，可以把这一章的主张说得更清楚一点。

很多人谈到"人类水平"，心里想象的是一根杆子：画一条线，写上"人类水平"，AI在线下面慢慢爬，爬过线就是"达到了人类水平"。这是高度的想象：横杆是个静止的物体，AI是个向它靠近的追赶者，两者互不相干。

但从印刷术、计算器、导航这几件小事里，看到的是另一幅图。人类水平不是人身上的一个量，而是人、工具和难题三者之间的一个位置。人拿起工具，工具被用起来，用的过程中长出新的难题，人去应对，就长出新的本事，这本事又成了下一个"水平"的内容。杆子并没有被谁举高；是人站的那块地，被他们手里的工具垫高了。

用SDE的话，把它写得更精确一点：某一刻，人所掌握的一套本事，是已经显露稳定的S；这套本事不是凭空来的，是在一次次差异中走出来的，那是D；它之所以算得上"水平"，是因为扎在一片意义的土壤里，一片叫作"这个时代的人们在乎什么、要解决什么"的E。三者缺一不可。

所以这一章只负责把"它会动"讲到你信；下一章才讲"为什么它动的方向恰好会把AI包进去"。

◆ 六、再看一个家里的例子：账本

家里每个月都在上演的一件事，是记账。早年一本本子一支笔，月底誊一遍，加一遍，"会过日子"的标准是记得清、算得准。后来手机软件把誊和加封装了，有人担心人从此不会算账。可今天"会理财"的人，看着软件生成的图表在问：餐饮这一项为什么一直涨，是物价涨了，还是我们家的习惯变了？这笔每月自动扣的订阅，我到底用不用？这些问题，手写账本的年代没有人问，不是人不聪明，是全部力气都耗在记和算上，没有余力看全貌。

第33章里那位记账的年轻人，长出了自己的"钱感"，说的是一个人身上的沉淀；这里是同一件事放大到一个时代：工具拿走记和算，"看全貌"才第一次成为日常，成为新的水平。四个例子的次序不是偶然的，每个工具都在人原来最费力的地方松了一把手，人的注意力便往上漂了一层。被封装的是昨天的本事，被逼出来的是明天的本事。

◆ 七、水平里还有一层：土壤的变化

还要补一点，以免读者把这件事理解得太"机械"。

以上几个例子里，水平的挪动，不只是"多了一项新技能"，而是连评价的尺子也换了。外婆那一代评价学问，看的是"背得多不多"；今天评价学问，看的是"想得通不通"。这把尺子是谁定的？没有人专门去定，它是整个社会在使用工具的过程中慢慢长出来的共识。

这就是E，那片意义的土壤。一个时代认为什么"值得会"，由这个时代的人们都在做什么、被什么难住、在乎什么，共同决定。工具改变了人们在做什么，就改变了土壤；土壤变了，同一个人在不同时代的"水平"，哪怕本事一点没变，评价也会变。

土壤的变化，有一个粗略的量。美国经济学家奥托（David Autor）和合作者用普查数据估计，今天约六成的人，所在的职业名称在1940年还不存在（2024年发表于《经济学季刊》）。一个1940年的评委，拿着当年的"水平表"，连一半以上的现行职业都找不到栏目可填。

这也是为什么，试图一次性给"人类水平"画一条线，总是徒劳：你画线的那个时刻，用的是那一刻的土壤，而土壤正随着你手里的工具悄悄翻动。

◆ 八、这个论证的边界在哪里

写到这里，必须诚实地交代四件事。

第一，三个例子是类比，不是证明。印刷术、计算器、导航的故事，说明"人类水平会随工具而变"在历史上确实发生过，而且发生得很规律。但历史上发生过，不等于永远会发生。真正撑起"永远"的，是后面章节里的结构论证：封装完成时，新的发生还没有显露，两者在时间上必然错位。这一章的日常例子，是为了让你对这个结论有手感，不是替代那个论证。

第二，并非所有"水平"都会这样抬高。有些能力，工具一封装，人就真的很快不再需要。没有人再因为"会用手摇计算机"而被认为有水平。这没有矛盾：让渡本来就意味着有些东西彻底退场。本书要说的不是"每一项具体能力都会升高"，而是"水平作为一个整体，内容在变，而且往工具之上的方向变"。

第三，"升高"不等于"每个人都跟得上"。这是个体层的问题，不是文明层的问题。一个具体的人可能在工具换代时被甩在后面，这是真实的代价。第六编已把"文明层必然、个体层靠碰撞"分开讲过，说横杆会动，不是说每个人都能自动爬上去。

第四，隔壁的说法。经济学里，阿西莫格鲁和雷斯特雷波把技术对工作的影响拆成两股力量：替代旧任务，与创造新任务，上面奥托等人的研究，正是后一股力量的实证。本书与它们相同的地方，是都认为新任务一定会冒出来；不同的地方，是经济学把新任务当作劳动市场上的结果，本书把它当作"水平"这个概念本身的内容，并在下一章追问：追赶者自己在不在这内容里。

什么情况下这一章错了。如果某个开放行业，在工具封装了它最费力的一环之后，几十年里行业内评价"好"的标准一直没变，也没有出现新的岗位名称，横杆就没有动。出现这种情形，本章在这个行业里就错了。

◆ 九、小结

有学问，从背诵挪到了检索与综合；数学头脑，从心算挪到了建模与判断；认路，挪成了知道何时不听导航。每一次，工具都把上一格的本事封装起来，人的水平随即被推到工具所撑起的新一格上。

于是"人类水平"成了一个奇怪的靶子：你越用工具去逼近它，它越是借着这工具往上走。下一章要讲的，是这个靶子最要紧的特点：它走的时候，带着追赶者一起。

> 带走的话
>
> 外婆那一代说“有学问”是背得多，今天是想得通，中间没有人开会宣布。“人类水平”画不出一条线，你画线的那一刻，脚下的地已经被手里的工具垫高了。

> **小账** 这一章立了什么？——人类水平是人、工具、难题之间的一个位置，是过程不是高度；印刷术、计算器、导航、账本四个例子让这件事有了手感。 错在什么情况下？——某个开放行业封装了最费力的一环之后，几十年里评价“好”的标准没变，也没有出现新的岗位名称。

承接：横杆会动，已有四个日常例子为证。预告：下一章讲，它不只是会动，而且动的时候把AI本身包了进去。

第 36 章

横杆里含着追赶者

◆ 一、一个厨房里的小故事

假设你刚学做菜，手机里装了一个菜谱软件。第一周，你完全照着做：盐放几克，火开多大，几分钟翻面，一步不差。做出来的菜像模像样，这个软件已经能"做到"一个普通家庭厨子的水平了。

三个月后，你不再逐字照做：你看冰箱里有什么，凭手感加减调料，哪里太咸、哪里该提前出锅，心里有数。你甚至发现，软件的配方按"标准灶台、标准锅、标准食材"写，而你家灶火偏猛，锅偏薄，于是摸索出一套"适合我家厨房的做法"。

此时，你的"做菜水平"和软件的"做菜水平"，谁高？这个问题已经变味了。软件停在那份配方上，而你已经是"软件加上你的使用经验"这样一个新的东西。你的水平里，含着那个软件：它是你的背景，你的跳板，一次次校对你手感的对象。你没有"超越了软件"，你是"带着软件长出了新的自己"。

现在设想有人想让软件追上你：他把你这三个月摸索出的做法收集起来，写进软件的下一个版本，软件现在"会"根据灶火大小调整了，追上了。可是然后呢？你用着新版，又发现软件调整得太自信，把一道需要慢火的菜做坏了；你开始琢磨什么时候该让它自动，什么时候不该。你又前进了一格。

这个追逐的结构是：追赶者每追上一步，就成了你下一步的材料。横杆不是在原地等，而是靠着追赶者的脚往上抬。

◆ 二、一个公式，大白话版本

把上面这个小故事压缩成一句话，就是这一章的核心命题：

> 人类水平（明天）＝ 人与AI（今天）碰撞所发生的新SDE。

这里的"＝"读作"由……发生而来"，不是代数。意思是：把"人"和"今天已经做出来的AI"放在一起用，这就是第6章讲过的碰撞。碰撞必生新的东西；其中一部分，已经被工具封装住的东西覆盖了，另一部分，也就是封装之外的那部分，成了新的结构、新的差异、新的意义土壤。这个新东西，就是明天"人类水平"所包含的内容。

公式右边有一个"AI（今天）"，左边是"人类水平（明天）"。这意味着，AI不是站在横杆之外去追它的追赶者，而是横杆的一块材料。不是追赶者在追横杆，是横杆在用追赶者。

换一个家里的例子。家里买了洗碗机，起初的目标很明确：洗掉一顿饭的碗。过了一年，你发现家里待客的方式变了：以前十个人吃饭是大场面，现在二十个人也敢请，一周请两回。洗碗机追的目标，已经不是"一顿饭的碗"，而是"一个敢请二十个人的家的碗"。这个目标不是谁偷偷调高的，它由洗碗机进了这个家以后发生的事构成。追赶者每向前一步，就被用进了下一个目标里。

所以我们说：AI永远在追一个含着它自己的目标。

◆ 三、为什么追不上：不是速度问题，是结构问题

读到这里，一位认真的读者可能会反驳：那我让AI跑得更快呀！今天的AI用一年追上去，明年再用一年追上去，不断追，总有追平的一天吧?

问题恰恰不在速度。洗碗机再快，也追不上那个因为它才抬高的目标；"差距"是追逐的动作本身造出来的。

用更精确的话说：在AI（今天）做成的那一刻，人类水平（明天）所依赖的新SDE还没有显露。这就是第7章讲的错位：封装发生时，它所封装的是已经发生的东西；而新的发生，是封装之后、在使用中才出现的。所以不管AI追得多快，它追

到的永远是"昨天的人类水平"，而在它追到的这一刻，今天的人类水平已经因为它自己的出现而变了。

◆ 四、"AI就是还没做到的那些事"

讲到这里，要正面处理一个在AI的历史里被反复提到的现象：每当AI做成了一件从前被认为需要智能的事，比如下棋、识别手写字、翻译、认人脸，人们过一阵子就不再把它当作"人工智能"的标志了。大家会说："哦，这只是算法而已，不是真正的智能。"于是这个词所指的东西，好像永远在往后退，永远是"还没做到的那些事"。这个说法在人工智能界常被称作"特斯勒定理"，以计算机科学家拉里·特斯勒命名；霍夫施塔特在1979年的书里就转述过。最早的措辞各说不一，这里只当作一种流传的说法来用。

常见的评价有两种：一说这是耍赖，AI做成了却不认账，一次次改标准；一说这个概念本身空洞，不过是随技术前进不断后退的标签。本书另有一种读法：这是发生学能够直接解释的事。

当一件事被封装了，它就成了一个稳定的结构，一个可以被调用的现成工具。而被封装的东西，就不再是前沿的智慧。因为"前沿智慧"指的是还在发生之中的东西，是人在摸索、犯错、校正的过程里所做的事。一个东西一旦被稳定下来、被一键调用，那个"摸索"的过程就退到了背后，前台的这个东西，成了背景，成了基础设施。

自来水也是这样。它的出现，在当年是一项了不起的"智慧成就"。可是今天，没有人把"能拧开水龙头出水"看作智慧。这不是对自来水的不公，而是因为它的智慧已经被封装完毕、让渡给了系统，我们对"聪明"的判断，早已转到了别处。

因此，当人们说"这只是算法而已"时，他们在不自觉地做一件合乎发生学的事：把已经被封装的东西，从"智能"这个仍在发生的类别里挪出去。这不是赖账，而是"智能"一词自己的含义决定的：它一直指向还在发生的那一部分。

这就和前面的公式接上了：AI做到的每一件事，在被做到的那一刻，就转变成了"背景"，而人类水平（明天），恰好就是生长在这个背景之上的那一层。"AI就是还没做到的那些事"，换一种说法，就是"智慧永远是封装之外的那一部分"。

◆ 五、追赶者进了横杆之后，会发生什么

再把这件事推到一个可以想象的场景里，但不预言任何年份。

设想某种AI，在某个领域里确实做到了"相当于一位熟练从业者"的水平：能写出合格的合同初稿，能整理病历、给出常规的诊断提示，能把一个家庭的账目分类、提示风险。我们假设这确实做到了，而且做得很好。

接下来走的，是第9章讲过的四步。先是扩大期：事情变快了、变便宜了，原本做不起的事做得起了，需求涌进来。然后是内卷期：大家都用同一类工具，产出趋同，竞争挤到了很窄的几个点上。接着是新产业诞生：使用中，一批原来不存在的问题露了面。怎样核对一份AI起草的合同里那些"看上去合理但没人负责"的条款？怎样判断一份AI整理的病历漏掉了什么？怎样让一个家庭对AI的理财建议建立恰当的信任？这些问题催生了新的角色、新的方法、新的行业。最后是让渡：最初被封装的那批活儿，退出前台，成为基础设施。

到这时，"熟练从业者的水平"已经不是当初的意思了。它包含了新产业里长出的新本事：怎样与AI协同，怎样为AI的产出负责，怎样在AI覆盖不到的地方做决断。这些是AI"追"不到的，因为在AI（今天）做成的时候，它们还没有出现。这就是横杆的发动机：每走完一轮，横杆升高一格，而升高的原料里，必然含着上一轮刚被封装的那个工具。

◆ 六、对一个反驳的回应

有人会说：那AI下一步就把这"新产业里的新本事"也学会不就行了吗？

可以学。但学会的那一刻，又一个封装完成，又一个新的错位出现。这不是"AI永远学不会"，而是"学会本身，是下一轮发生的起点"。第8章把这叫作升级

律。读者不妨把它想成是一个追不尽的接力：每一棒都交得出去，但交完之后，场上永远有下一棒要跑。

◆ 七、这个论证的边界在哪里

这一章最大的口气，是说"AI永远在追一个包含它自己的目标"。它的边界，分五点交代。

第一，这个论证针对的是开放领域，不是所有领域。"人类水平含着追赶者"成立的前提，是这个领域里的使用会生出新的问题类型。如果某个领域根本不生新问题，横杆就不会因此移动。下一章的象棋，正是这样的反例，我们会毫不回避地讲。

第二，公式里的"碰撞所发生的新SDE"，不总是显眼，也不总是立刻到来。中间有一段"封装已完成、新的发生尚未显露"的空窗，就是第十编要讲的"等待期"，也是人们误以为"AGI已经来了"的原因。横杆在移动，不等于它每一天都肉眼可见。

第三，"AI就是还没做到的那些事"的发生学读法，是一种解释，不是唯一的解释。心理学上、社会学上还可能有别的因素，比如人们对新鲜事物的惊奇感会自然淡去。本书不排斥这些解释，只是说，发生学的读法能把这个现象接进整个论证，而且不需要把它说成人们的"耍赖"。

第四，这个论证并不证明"追赶者"必须是人类以外的某种东西，也不证明升级的主体必须是人。它证明的是：总需要一个更高一级的发生来接住上一轮的封装。至于这个"来接住的"究竟是谁，是第38章末尾要点出、第九编要正面处理的一条缝。

第五，隔壁的说法，以及什么情况下这一章错了。哲学家哈金（Ian Hacking）讲过"循环效应"：被分类的人会回应分类，于是类别本身也跟着变。AI效应是一条类似的回路，在大家熟知的"目标不断后移"这一层，本书没有新发现。本书多出的一步，是说追赶者本身是靶子的材料，因此"追上"在开放领域里没有一

个静止的终点。这一步的检验，就是附录C的预言一：在某个开放行业里，AI达到熟练从业者的水平之后，如果相当长的时间里，从业者的工作内容和评价标准都没有变化，也没有出现新的岗位或问题类别，公式就在这个行业里失效。

◆ 八、小结

这一章的话，压缩成三句。

第一句：人类水平（明天），等于人与AI（今天）碰撞所发生的新SDE，所以AI是这根横杆的一块材料，不是站在外面的追赶者。

第二句：因此AI永远在追一个含着它自己的目标。追不上的原因不是跑得慢，而是"追"这个动作本身，在往靶子里添料。

第三句：人们说"AI就是还没做到的那些事"，不是耍赖，是因为被封装的东西就不再是前沿智慧，这是发生学能解释的事。

> 带走的话
>
> 追赶者每追上一步，就成了你下一步的材料。洗碗机追的是“敢请二十个人的家”，这个目标是它进门以后才有的。

> **小账** 这一章立了什么？——人类水平（明天）＝人与AI（今天）碰撞所发生的新SDE；AI是横杆的一块材料，不是站在外面的追赶者。 错在什么情况下？——开放行业里AI达到熟练从业者水平之后，相当长时间内工作内容和评价标准都没变，也没有新岗位或新问题类别（附录C预言一）。

承接：横杆里含着追赶者，这在开放领域里成立。预告：下一章必须面对最强的反例，象棋。在那里，AI真的永久越过了人，我们要看看为什么，以及这说明了什么。

第 37 章

象棋的反例

◆ 一、一场被很多人记住的对局

1996年2月，世界棋王卡斯帕罗夫与计算机"深蓝"下了六局，卡斯帕罗夫赢了，比分4比2。一年多以后，1997年5月，改进后的深蓝在同样六局的复赛里以3.5比2.5获胜。这成了现代科技史上的标志性时刻：人类顶尖的棋手，败给了机器。

如果这本书的论证只能解释那些"AI够不着"的事，对"AI够到了而且再也没让人追上"的事沉默，那这个论证就值不了什么钱。所以，这一章要正面承认：象棋是一个反例，一个真的、没有水分的反例。

◆ 二、先看故事的后半段

很多人只记得1997年，后面的故事更值得讲。

1998年，卡斯帕罗夫提出了一种新玩法，叫"高级象棋"：一位棋手加一套棋力很强的程序，组成一队下，第一场表演赛由他亲自上阵。后来这种人机合队常被叫作"半人马"。2005年，一场向所有人开放的网络自由式比赛里，夺冠的不是特级大师，而是配着几台普通电脑的业余棋手。卡斯帕罗夫后来的评语，大意是：会配合的流程，比单纯的棋力、单纯的机器，都更要紧。

这一段看上去，完全符合本书的剧本：工具（引擎）出现，使用即碰撞，"人机配合"这个新SDE生出来了，人类水平随着工具升了一格。

然而故事没有停在那儿。引擎继续变强，人在配合里添的价值，一点一点被压缩。2013年前后，已经有经济学家等人公开判断，半人马的优势在消失；卡斯帕罗夫本人到2017年仍然认为，选对了操作者，人机队伍还能赢。棋界对"优势什么时候没的、是不是彻底没了"，并没有一致的说法。没有争议的是另一件事：在正式

对局里，顶尖棋手已经没有办法战胜顶尖引擎，而且这不再是一个会翻回来的局面。

也就是说，在"和机器比谁下得强"这件事上，人类水平这根杆子不再往上走了，AI永久地越过去了。请注意这句话：它不是"AI暂时领先、人还有机会"，而是这条路已经走到了头。

◆ 三、为什么象棋会这样

问题来了：为什么在象棋里，"横杆随工具移动"的剧本，没有一直演下去？

答案并不玄。按本书的术语说，象棋是一个封闭的领域：规则、目标、评价，三样都在开局之前划好了。

第一，规则是封闭的。象棋的棋盘就是六十四格，棋子就是那几种，走法就是那几条。一千年前的规则和今天几乎没有区别。这意味着，所有"可能出现的局面"，数量虽然极其庞大，却是一个事先划定了边界的集合。

第二，目标是封闭的。赢棋就是将死对方，没有第二种含义。你不能在下棋的过程中，发现"其实这盘棋的目的是别的什么"。目标不会被使用过程改写。

第三，评价是封闭的。谁赢了，一目了然。没有任何"意义上的争论"，比如"这一步虽然输了，但是很美，所以算一种赢"。至少在竞技的层面，没有。

这样一个领域里，工具的使用会碰撞出更好的走法。这确实是新东西，人机合作那段时间，大家确实发现了从前没想到的走法和思路。但这些新东西，始终落在同一张棋盘、同一套规则里，是同一个封闭局面里的"更优解"。它们长不出新的棋盘，改不了规则，也不会让"什么叫赢棋"变得不一样。

打个比方，这就像在一个有终点的迷宫里找路。起初没人知道最短路径，大家各凭本事摸；后来有了更强的工具，一条一条路径被找了出来；最后，工具把整个迷宫摸透了。到这时，再往里面加人的参与，既不能让迷宫变大，也不能让出口挪地方，人能做的就越来越少。

反过来，开放领域像一座会长出新房间的房子：你每走进一间，用一种工具，就可能推开一扇从前不存在的门，不存在"摸透"这回事。

◆ 四、象棋"不生新行业"

还有一点，要单独拎出来说。

前面几编里反复看到：真正有力量的工具，会生出一整个新的产业。电力生出了电器行业，铁路生出了时刻表和标准时区，计算机生出了软件行业。

象棋的情形不是这样。深蓝带来了真实的变化：棋手的备战和训练变了，棋谱研究更精细。但它没有生出一个从棋盘之外长出来的、与棋无关的新行业。象棋这个世界，还是那个世界；它被工具"磨得更亮"，却没有被工具"推开新的门"。

这个区别，可以用一个日常的对比来记。菜刀磨得再快，切菜的人确实省力，但切菜这件事没有变成别的事；而一套有了冰箱、燃气灶、微波炉的新式厨房，改变的是人怎样安排一日三餐，进而改变了买菜的频率、采购的节奏、家庭的作息。象棋里的工具更像前者：让一个封闭的活动做到极致，却没有生出一个更大的活动。

◆ 五、一个更近的比方：一道有标准答案的数独

如果觉得象棋离生活远，想想报纸角落里的数独。

数独是典型的封闭游戏：九乘九的格子，数字一到九，行、列、宫里不许重复。规则明确，目标明确，对错明确。早年人们靠一支铅笔、一点耐心去解；后来有了程序，一秒钟解完最难的题。

有了解数独的程序，数独界发生了什么"新产业"吗？没有。解题程序让"解数独"这件事彻底失去了挑战性，人要是还乐在其中，那是因为解题本身的乐趣，而不是因为这件事还有什么"水平"可攀。程序在这里越过了人，而且永远越过了，没有人觉得这是悲剧，也没有人觉得这说明程序"比人聪明"。大家只是明白了：这是一个有尽头的游戏。

同样是"程序帮人"，做饭不是有尽头的游戏：食谱软件、智能灶具、外卖平台一路出现，生出了新的菜式和行业，因为"吃什么、怎么吃、和谁吃"本身就在随生活变。差别不在程序强不强，而在游戏本身有没有尽头。象棋站在数独这一边，只是它的迷宫大得多，大到人类花了几百年也没有摸到边。

◆ 六、可是，杆子并没有停

读到这里，读者也许有点泄气：既然象棋里AI永久越过了人，"人类水平"在这里不就是一个被甩开的靶子吗？

先别急。"人类水平"在象棋里其实是两件事，被我们说成了一件。

一件是人和机器比，谁下得强。它的规则、目标、评价都是封闭的，已经有了答案，不会翻回来。

另一件是人和人之间的象棋：谁下出了新的开局想法，哪一路风格影响了一代孩子，少年怎样比前辈更早成材，一盘棋好看在哪里、值得讲给别人听。这件事的杆子还在动。挪威棋手卡尔森的等级分，在2014年创下有记录以来的最高值，2019年又追平了一次，那是引擎稳压人类很多年之后的事。

两根杆子为什么分了家？因为人下棋，从来不只是为了下出最强的一步。人愿意看人下棋，是因为输赢里有人在乎、人在紧张、人把自己押了上去；引擎对引擎的对局，再强，对多数观众也只是一个技术展示。那片叫"为什么下棋"的土壤，E，没有跟着引擎走，它还在人这边。封闭的只是核心对弈那一层；围着棋盘的人类生活，是一座开放的房子。

所以象棋的故事，对读者其实不是一则坏消息。它说的是：一个领域里某一层被永久越过，不等于这个领域里的人被越过；要做的，是认出哪一层已经封闭，把力气让给仍在生长的那一层。这正是前面"让渡"的意思。

◆ 七、承认这个反例，对本书意味着什么

写到这里，我想把话说得很直接。

承认象棋反例，不是本书的让步，而是本书的一部分。如果有人以为，这本书想说的是"AI做不了任何人能做的事"，那是误会。本书的命题是：**封闭领域可被永久越过，开放领域不可**。

具体来说，这个反例带来三点校正。

其一，它修正了一个可能的过强说法。假如本书说"AI永远追不上人类水平"，那会被象棋立刻击倒。所以本书的说法更窄、更准：AI在开放领域追不上；在封闭领域，它可以、而且已经永久越过。

其二，它给了"人类水平随工具升级"一个明确的适用条件。条件就是：使用能否生出新的问题类别。能，横杆就动；不能，横杆就停，然后被工具越过。

其三，它给出一个可用的检验方法。任何一个宣称"AI已经达到或超过人类水平"的领域，我们都可以追问：这个领域是不是封闭的？它的规则、目标、评价，是不是事先划定、不会被使用过程改写？如果是，那么AI的越过可能是真的、永久的，我们不必焦虑，也不必替人类找借口。如果不是，那么"越过"很可能只是等待期里的一个截面。

◆ 八、一个常见的追问：围棋呢？

肯定有读者想问：那围棋呢？围棋比象棋复杂得多，长期被认为是人类智慧的堡垒。2016年3月，AlphaGo与李世石下五番棋，赢了四局，后来的版本把差距拉得更大。

按上面的标准看，围棋同样是封闭的：棋盘固定，规则固定，目标固定，评价固定。它的复杂，只让"摸透迷宫"难得多，没有改变它是迷宫的性质。所以围棋被越过，在本书的框架里不是新的反例，而是同一个反例的更大版本。

有人说，AI在围棋里下出了人类从未想过的走法，反过来启发了人类棋手。这是真的，而且已经有人量化过：一项发表于2023年《美国国家科学院院刊》的研究估计，职业棋手的决策质量，在AlphaGo的几场胜利之后明显上升，此前六十多年几乎持平。可它生出的，是棋盘之内的新走法，没有长出棋盘之外的新活动。

所以它不改变"封闭"的判断，反而再一次说明：封闭领域里，工具仍然可以抬高人和人之间的水平，只是抬不动人和机器之间的那根杆。

◆ 九、这个论证的边界在哪里

这一章的边界，要交代得格外认真，因为这是本书最容易被人抓住的地方。

第一，"封闭"与"开放"不是黑白两色，而是一条连续的光谱。现实中的许多领域处在两者之间，比如某些标准化的考试、某些固定格式的文书工作、某些规则明确的游戏，封闭程度各不相同。判断一个具体领域在光谱上的位置，需要认真分析，不能一句话打发。AI研究里，"封闭世界"与"开放世界"是个老分法；本书与它相通，只是把标准落在"使用会不会生出新的问题类别"上，不落在知识库完不完整上。

第二，说"封闭"，要说清是哪一层。棋盘本身封闭，"围绕棋盘的人类生活"并不封闭。说象棋封闭，指的是它核心对弈这一层。同样，一个看起来封闭的领域，有可能被人从周围"打开"。

第三，我的判断可能出错，错在哪里，我写在这里。有两种情形会让这一章的读法站不住。其一，如果有证据表明，引擎越过人之后，人与人之间的象棋水平停滞或萎缩，参与持续下降，并且可以追溯到引擎本身，那么"杆子分成两根"就错了，象棋就是一个被工具终结的领域（这是附录C的补充三）。其二，如果某个被本书判为开放的领域，在AI大幅跃升之后，长期没有生出任何新的问题类别，而AI又稳定、全面地越过了人，这就是对本书命题的反证。我愿意把这两点白纸黑字写在这里。

第四，象棋这个例子，不能反过来被拿去"证明"AI会在一切领域永久越过人。它只能证明封闭领域里确有此事。从一个封闭领域的结果推到所有领域，是逻辑上站不住的跳跃，这个跳跃恰恰是"AGI感"常犯的错误之一。

深蓝在1997年胜了卡斯帕罗夫。其后一段时间，人机合作确实一度占优；再往后，单独的引擎就够了，在"和机器比谁下得强"这件事上，人类水平不再随工具升级，AI永久越过了。

原因在于象棋是封闭的：规则、目标、评价都是事先划定的，工具在其中只能碰撞出更好的走法，碰撞不出新的棋盘，也生不出一个棋盘之外的新行业。

这个反例不是用来绕开的，而是用来划边界的：横杆会不会移动，取决于领域是开放还是封闭。被越过的只是封闭的那一层，人和人之间的那根杆子，仍然在动。

> **带走的话**
>
> 象棋里AI赢了，赢得干净，不用替人找借口；但被越过的只是棋盘上那一层，围着棋盘的人还在往前走。以后听到“AI已经超过人”，先问这个领域的规则、目标、评价是不是事先划死的。

> **小账** 这一章立了什么？——象棋是封闭领域的真反例：规则、目标、评价都先划定，所以可被永久越过；用它划边界，开放领域不在其内。 错在什么情况下？——引擎越过人之后，人与人之间的象棋水平停滞萎缩且可追溯到引擎（附录C补充三），或被判为开放的领域长期不生新问题而AI稳定全面越过。

承接：象棋划出了封闭的一边。预告：下一章把边界另一侧划清，并说明为什么"通用"这个词，恰恰是对开放的一侧说的。

第 38 章

通用即开放

◆ 一、请一位"全能管家"上门

我们再从日常说起。

设想你的家里要雇一位"全能管家"。你对他的期待是：能做饭，能打扫，能带孩子，能记账，能应付水管突然爆裂，能在亲戚来访时张罗一桌像样的饭菜，能在孩子半夜发烧时知道该怎么办。

你会发现，只要你一用"全能"这个词，你的要求就开始移动。

起初，你想的只是"做饭、打扫、带孩子"，管家做到了，你满意。半年后你想：要是他还能辅导功课就好了。又过半年：要是他还能陪老人聊天就更好了。家里添了一只猫，你又想：要是他懂猫的习性就好了。

没有哪一次你是故意加码的。是生活本身在变，新的需要不断冒出来，而"全能"的意思，就是覆盖这些还没冒出来的需要。

这个小故事，把这一章的核心想法放在你手边了："通用"这个词，一说出口，就是对一个会生长的领域说的。

◆ 二、封闭领域不需要"通用"这个词

换个角度，看看"通用"一词的用法。

没有人会说"通用的象棋程序"。象棋程序就是象棋程序，它做的事是固定的。我们说一个象棋程序"强"，是说它下得好，不是说它"通用"。

人们说"通用"的时候，在说什么？在说：它不限于某一个事先划定的任务，它面对的是不能事先列举的各种情形。通用计算机，指的是它能跑你事先没想到的程

序。通用语言，指的是它能谈论还没发生的事情。通用的工具，意味着它要面对未来的、未被规定的使用。

换句话说，"通用"这个词本身，就内含着一个对"开放"的承诺。你要求一个东西通用，你就是在要求它去面对一个边界未定的世界。

所以，"通用"不是能力的一个级别，而是对"还没出现的需要"的一份承诺。承诺要靠使用来兑现，而兑现的过程，恰恰是新需要冒出来的过程。全能管家的故事里，你每一次满意，都是下一次加码的起点。

而本书前面三十几章，一再讲的道理就是：边界未定的世界里，工具一用，就生新；新的东西，又在封装之外。

这里的逻辑链条非常短：

- 要求"通用"，就是要求面对开放领域；
- 开放领域里，使用会生出新的问题类别；
- 新的问题类别，在封装完成的那一刻还不存在；
- 所以"通用"所要覆盖的范围，在你完成封装之后，又长出了一块。

横杆，就是在你说出"通用"这两个字的时候，开始移动的。

◆ 三、一个小对比：一把钥匙，和一串会长的钥匙孔

再用一个小比方，帮你把这个想法放稳。

封闭领域，好比一把锁。锁孔是固定的形状，钥匙只要做对了，就一劳永逸。AI在象棋里的"越过"，就像是配出了那把终极钥匙：它能打开这把锁，而这把锁永远不会换形状。

开放领域，则像是一串会自己长出新锁孔的门。你手里的钥匙再好，也只能开已经长出来的那些锁。至于下一个锁孔长成什么样，要等它长出来才知道。而且，你开门的动作本身，就是在催着新的锁孔长出来。

"通用"，等于是要求"一把能开所有锁的钥匙"。对一把锁，这个要求没有意义，也没必要这样说；对一串会长的门，这个要求在逻辑上就是要求一把事先开得了未来的锁的钥匙，而那些锁，在钥匙铸成时还没有成形。

◆ 四、划一条清楚的线

到这里，我们可以把第八编要立的边界，用最干净的话写下来。

封闭领域，可被永久越过。它们的特征是：规则固定、目标固定、评价固定，使用不生新的问题类别。在这样的领域里，AI可以、而且可能已经永久地越过了人。象棋是典型，围棋是更大的版本。对这类领域，本书承认AI的胜利是真实的、持久的。

开放领域，不可被永久越过。它们的特征是：使用会生出新的问题类别，目标和评价随着使用而被改写。在这样的领域里，任何一次"越过"，都是一个截面上的事；因为这次越过本身，就成了下一轮发生的材料。

"通用"一词，说的是开放的那一边。所以，当有人说"通用的人类水平AI"时，他同时在两个方向上说话：一个方向是，要它在一个开放的世界里达到人的水平；另一个方向是，要它的"达到"是一个静止的、可以宣布完成的状态。这两个方向，是互相拧着的。

这就是第八编标题"移动的横杆"的全部含义：横杆会移动，不是因为谁故意把它举高，而是因为你选了"通用"这个词，你就选了一个会生长的世界。

◆ 五、回到"定义缝"

在第1章里，本书交代过一条缝：主流的AGI定义并不要求"解决所有问题"，只要求"在广泛的任务上达到或超过人的水平"。所以对终局AGI的反驳，击中的是一种更强的说法；对"人类水平的通用AI"这个温和说法，需要另外的论证，就是第八编。

第八编没有否认AI可以在许多任务上达到人的水平。它只是指出，当"广泛任务"指的是开放领域里的任务，"人的水平"本身就是个会动的东西，动的方向还要把AI包进去。因此，"达到人类水平"不是一个可以一劳永逸完成的动作，只能是不断被重新开始的追赶。

这等于承认：阶段性的状态可以出现，我们会看到AI在许多任务上与人不相上下，甚至更好。但这个状态不会稳定地被称为"通用"：它一出现就进入使用，使用就生新，新在封装之外，"通用"的名分又在前方走远了一步。这也是本书把"类AGI能力"和"终局AGI"分开的原因：前者可以出现，后者作为终局不可能。

◆ 六、最后留着的那道缝：谁来升级

论证走到这里很顺，但我必须在第八编的最后，把一条没有堵上的缝亮出来。

前面所有的论证，大致证明了这样一件事：在开放领域里，封装之后总要有新的发生，总需要一个更高一级的东西去接住上一轮的封装。

可是你注意到没有：这个论证，只证明了"需要有更高一级的智慧"，并没有证明这个更高一级的智慧，必须是人。

也就是说，一个严苛的反对者可以这样说：好，你说总需要有人去接住新的发生。那我就让AI本身来接。AI自己发现新的问题，自己生成新的SDE，自己升级自己，不就行了？横杆里含着追赶者，那就让追赶者同时担任"举杆的人"。

这个反驳是有力的，我不想糊弄过去。

本书的回答，不在这一编里完成，而是留给第九编。但我可以先把要点预告一下，也先把它的性质说清楚：

本书把一个前提明确地立在这里：一个"问题"，总是对某个主体而言的问题。一个难题之所以是难题，一件事之所以"值得做"，一个新的结构之所以"有意义"，都要在某个在乎它的主体那里才成立。用SDE的话说，那片叫做E的土壤，它的源头在主体。

这是前提，不是推出来的结论。我把它写成前提，就是要让读者清清楚楚地知道：如果你不接受这个前提，那么本书在这一点上的论证就不成立，你可以在这里与本书分道。如果你接受它，那么第九编的论证会表明，AI在升级自己的时候，会制造出只能对"在乎的主体"才成立的问题，而这样的问题，恰恰无法由AI单独接住。

先说到这里。更多的，留给第九编。

◆ 七、这个论证的边界在哪里

这一章是第八编的收口，边界要交代得最完整。

第一，"通用即开放"，是对"通用"一词的一种读法，不是一个无可争议的定义。有人会把"通用"定义成"在一组固定的广泛基准测试上表现良好"。如果是这种定义，那么这样的"通用"在逻辑上是可以被达到的，本书不反对。本书只是指出，这种"通用"是一个封闭化的通用，与日常语言里"什么都能应付"的那种通用，不是一回事。

第二，隔壁的说法。AI研究者肖莱（François Chollet）2019年的论文把智能看成面对没有事先准备过的新任务时获取技能的效率，并指出只测技能，会被先验知识和海量训练"买"出高分。这与本书"通用的要害在没被事先列举的情形"相通。分歧在于：他把这当作可以靠更好的基准去测量的东西；本书认为，只要领域是开放的，基准一经成形，就是一次封闭化，使用又会生出基准之外的任务，所以靶子会动。

第三，开放与封闭的区分，不能被当成万能的挡箭牌。假如有一天，某个本书判为开放的领域，被AI稳定、全面地越过，而且此后十年二十年里再没有生出新的问题类别，我会承认那是对本书的反证。这与附录C的预言一，是同一个检验的两种说法。

第四，"横杆在动"不等于"人始终高于AI"。本书说的只是，水平这个概念本身，不是一个可以被一次性抵达的目标。

第五，关于"谁来升级"的那道缝，本章只是亮出，没有补上。如果读到这里觉得"那还是没有堵死"，这个感觉是对的，也是我故意保留的诚实。下一编，才是补缝的地方。

◆ 八、小结

这一章把边界划清了。

封闭领域，规则、目标、评价都定死了，工具再强，只是把迷宫摸透，所以可以被永久越过。开放领域，使用会生出新的问题类别，任何一次越过，都成为下一次发生的材料，所以不可被永久越过。"通用"这个词，说的正是开放的那一边：你一要求通用，你就选了一个会生长的世界，横杆就开始移动。

第八编的定音句，在这里可以完整地说一遍：**人类水平不是一个高度，而是一个发生过程；AI可以抵达它的任何一个截面，却抵达不了过程本身，因为每一次抵达，都成了下一次发生的材料**。

> 带走的话
>
> 你一说“全能”，要求就开始往前走。“通用”是对一个会生长的世界说的话，所以谁宣布“通用已到”，宣布的都只是一个截面。

> **小账** 这一章立了什么？——“通用”一词说的是开放那一边：封闭领域可被永久越过，开放领域里每次越过都成为下一轮发生的材料；“谁来升级”这道缝亮出但未补。 错在什么情况下？——被判为开放的领域被AI稳定全面越过，此后十年二十年再没有新问题类别（同附录C预言一）；或“通用”只是固定基准上的高分，那属于封闭化的通用。

承接：第八编证明了，人类水平AGI作为稳定状态追不上。预告：第九编转向最后也最硬的一问：终局AGI为何自我否定，以及"问题是对谁的问题"这道前提到底站不站得住。

第 九 编

终局AGI为何自我否定

编 序

前八编一直在铺路。第一编分清了三种AGI，第二编讲了工具发生的原理，三、四、五编用三界的工具史做佐证，第六编讲三界如何交织，第七编讲智慧为什么不是库存，第八编讲横杆为什么会动。走到这里，可以把话收紧，只回答一个问题：如果真有一个"终局的AGI"，也就是能解决一切问题的那种智能，它在结构上站得住吗?

这一编的回答是：站不住。它不是被外部的障碍挡住，而是被自己否定。四章依次要做的事是：先把论证写成一条完整的推理（第39章），再说明"基础问题"本身也在发生，不存在一张封死的清单（第40章），再看AI自己制造了哪些它眼下解决不了的基础问题（第41章），最后把最容易被忽略的那道缝补上：升级的事，到底要由谁来做（第42章）。

这一编收束的是第一层论证“没有终局智能”。前面八编把路铺好，这里可以把话收紧：把论证写成一条完整的推理，再逐处堵住反驳，最后诚实交代那道最大的缝。

章	这一章回答什么
第39章	结构证明：越被使用，越不通用。封装、碰撞、升级，三个前提加一步错位，推出终局自我否定

章	这一章回答什么
第40章	基础问题也在发生。基础性不是固定清单，每个基础问题都有被抬进名单的时刻
第41章	AI制造了自己解决不了的基础问题。对齐、可解释、机器意识，是AI越能干越冒出来的影子
第42章	问题是对谁的问题。把“谁来升级”这道缝摆上桌面，立为前提，并划清核心层与补充层

读这一编，请把每章末尾的“边界”小节当正文读：每一个论证都附了推翻它的办法，哪一环站不住，请指出来。

第 39 章

结构证明：越被使用，越不通用

◆ 一、追自己的影子

小时候大概都玩过一个游戏：太阳底下，想踩住自己的影子。你往前跨一步，影子也往前挪一步；你猛地一冲，影子冲得一样快。你永远踩不到它，不是因为你跑得不够快，也不是因为地面不够平，而是因为影子是你的动作投出来的。你每动一下，影子就跟着动一下，"追"这个动作本身，就在制造要追的东西。

把这个游戏放大，就是这一章要讲的事。设想有人造出一个东西，宣称它能解决一切问题。人们开始用它。用着用着，世界因为它的存在而变了：有些原来不存在的事出现了，有些原来不是问题的事成了问题。这些新出现的问题，是"用它"这个动作投出来的影子。它要把这些影子也解决掉，就得再动；它一动，又投出新的影子。

这不是比喻得过了头。第二编到第八编讲的所有东西，压成一句话就是这个：使用会投影。这一章要做的，是把这个直觉写成一条前提清楚、可以被检验、也可以被反驳的推理。

◆ 二、论证的对象：终局AGI

先把靶子立清楚，免得打偏。

第一章分过三种AGI。第一种是"人类水平的通用AI"，在大多数认知任务上不输普通人；第二种是人们感觉AGI已经到来的心理状态，叫AGI感；第三种是终局AGI：存在一个智能，它能解决一切问题，不再需要更高一级的智慧来补它。

主流的AGI定义，大多并不要求第三种。很多定义说的是"在广泛任务上达到或超过人类"，没有说"解决一切问题"，所以这一章击中的不是那些定义。对"人类

水平的通用AI"，要等第八编的"移动的横杆"去回答，那里的论证走另一条路。这一章只对付终局AGI。

这个靶子为什么还值得打？因为它一直隐隐约约地躲在公共讨论里。它藏在"最后一个发明"这样的说法里（古德六十年代的那句话，是这个想法最出名的出处），藏在"到那时人类就不用再思考了"的忧虑里，也藏在"总有一天什么都能被自动解决"的期待里。兴奋也好，恐惧也好，都默认了有一个终点。

第三章说过，终局AGI首先不是一台机器，而是理念界的一座塔：机器造得出来，有具体的边界；"通天"的塔顶是个定义，不是个高度。这一章要做的，就是把这句比方写成推理。

◆ 三、先把三个词钉住

要把推理写严，得先钉住三个词，免得读者在词缝里溜走。

问题集。　某一时刻，对某个主体而言构成问题的事，合起来叫那一时刻的问题集。它不是一份事先写好的清单，会随时刻变（第40章专讲）。"对谁而言"，第42章专讲；这里先取最弱的读法：问题至少是对某些主体而言的问题。没有这个读法，"解决一切问题"这句话本身就没有意思。

覆盖。　一个系统的规则，是它被造成时定下的东西：架构、学习的方式、目标、评判的标准。规则所能到达的全部，包括它在规则之内学到的、推出的，叫这个系统的覆盖。学习并不自动逃出覆盖：照着原规则一路学下去，学到的仍在覆盖之内。

同一个系统。　规则没变，就是同一个系统，哪怕它的参数天天在更新。规则变了，也就是引进了新的概念、新的量尺、新的评判标准，就是升级，得到的是另一个系统。拿厨房说：全能料理机照着出厂的菜单搭出一道新菜，还是那台机器；给它装一套新的火候逻辑和新的"好吃"标准，就成了另一台。

于是终局AGI可以写得更准：一个系统U，对每一个时刻t，t时刻的问题集都落在U的覆盖之内，并且U不必为此改规则。

◆ 四、三个前提

三个前提都在前面各编讲过，这里用日常的话重述，并标出每一个靠什么成立。

前提一：封装律。　一个系统的规则，是用它被造成那一刻已经发生了的东西定下的，所以覆盖在那一刻就划了边。洗衣机封的是洗衣的经验，导航封的是认路的经验，记账法封的是算清账目的经验。

靠什么成立？靠对"造"这个动作的分析：造东西的人，只能用已有的概念、材料和方法去造，还不存在的概念写不进规则。有人会说，会学习的系统，规则里本来就留着"学新东西"的口子。对，可是口子通向哪里，也是规则定的。

前提二：碰撞律。　被封装的东西一旦被使用，就与使用它的世界相碰。洗衣机进了家，碰上的是家里的时间安排、对"干净"的期待、邻居的眼光；导航进了车，碰上的是司机的方向感、城市的路网、外卖平台的调度。结果不是世界原样不动，而是长出原来没有的新结构、新差异、新纠缠，也就是新SDE。

前提三：升级律。　落在覆盖之外的新问题，要得到解决，规则就得变，也就是升级。第八章把这叫二阶碰撞：第一阶是工具与世界相碰，第二阶是人带着对第一阶结果的觉察，回头去碰原来的工具，工具于是被改造。这一条差不多是上面三个词的直接推论；真正承重的，是前两条，以及下面要补的那一步。

补的那一步：错位。　前提二说使用生出新东西，可新东西里为什么一定有一部分在覆盖之外？这是第七章的错位定理，这里把理由说透，分两层。

第一层是时间的：规则定于使用之前，使用之后才出现的东西，规则里没有它的名字。

第二层更要紧：新问题之所以"新"，常常不只是多了几道新题，而是多了新的问法和新的"什么算解决"。对齐和可解释性（第41章详谈）就是这样：在有了会自己拆解任务的系统之前，连"算对齐了"的量尺都不存在。覆盖能到达新题目，却到达不了一把还没被造出来的量尺；没有这把量尺，就判不出"解决了没有"。

◆ 五、把它写成三段论

设U是那个假想的终局智能。

第一步。U是被造出来的，按封装律，它的规则定于某一时刻，覆盖在那一刻划定。

第二步。U要"通用"，就得被拿去用，而且用在开放的领域里（封闭领域另说，见边界第三点）。按碰撞律，使用生出新SDE；按错位，其中有一部分落在U的覆盖之外，并且成了对某些主体而言需要解决的问题。

第三步。这些问题要被解决，只有两条路。

路甲：U不改规则，自己解决。可这些问题按第二步在覆盖之外，而覆盖就是规则所能到达的全部，所以路甲与第二步直接矛盾，走不通。

路乙：改规则，也就是升级，得到U'。能解决那批问题的是U'，不是U。U'一旦投入使用，同样的推理再来一遍，又有覆盖之外的新问题，要U"。如此下去，没有尽头。

结论。**不存在一个规则固定的U，使它在每一个时刻都够用**。所谓"能解决一切问题的智能"，要么指一个规则固定的系统，那它做不到；要么指那个不断改规则的整个过程，那它就不再是一个"系统"，而是智慧本身的发生，这不是AGI这个词原本想说的。前者自我否定，后者换了主题。

这就是章题"越被使用，越不通用"的意思。它不是说工具用多了会变笨。"通用"其实是个比值：分子是系统的覆盖，分母是此刻的问题集。人们习惯把它想成一个越爬越高、总有一天爬到顶的数，可分母不是死的，每一次使用，都在往分母里添新的问题。通用不是一个能抵达的状态，而是一个分母随分子一起长的比值。这个比值可以一直很高，却没有一个"从此恒等于一"的时刻。

换到厨房里再走一遍。

有人买了一台"全能料理机"，号称家里需要的一切都能做。刚买来确实好用：煮粥、炖汤、和面，都不用管。用了半年，家里的饮食变了：孩子开始要吃一种以前没人吃的东西，老人的牙口变了，来了一位对某种食材过敏的客人。这些新要求，出厂时不在菜单里，因为那时家里还没有这些事。

你可能会说，那就更新菜单。可以，但更新之后它就不是原来那台了；更新之后继续用，家里又变，又要更新。从来没有一台一次性解决一切的料理机，只有"买了、用了、改了"的循环。这个循环里最关键的一环，是"发现新要求"：它发生在机器外面，发生在吃饭的人身上。

顺序再说一遍，因为它是整个证明的骨架：封装在前，使用在中，新东西在后；新东西总是出现在封装的后面，所以封装永远追不上自己投出的影子。

◆ 六、它为什么不是"现在还没做到"

读到这里，很自然的反应是：那是因为现在的机器还不够强，再强一点，它自己就能预见这些新问题，提前封进去。

这个反应要认真对待，因为它指着论证里最需要小心的地方。预见当然可以。更强的系统可以预见很多问题，把无数种可能都演练一遍。但预见有个边界：它能预见的，是在已有结构里能够被表述出来的东西。真正的新SDE，是那种在原有结构里连"表述"都还没有的东西。一个只会在棋盘上想问题的系统，能预见一万种棋局，却预见不了"有一天棋盘被换成别的游戏"。

所以"现在还没做到"和"结构上做不到"是两件事。前者是工程问题，后者是这一章要证明的。论证的力量来自后者：它不依赖某一代机器的强弱，依赖的是封装与新SDE在时间上的错位，以及新量尺的缺席。

◆ 七、回扣：这就是承重命题三与九

这一章把两条承重命题连了起来。命题三是错位定理：新SDE必在封装之外，因为封装时它还不存在。命题九是：基础性本身在发生；AI制造了自己解决不了的

基础问题。这一章为命题九打底：使用既然必生封装之外的新东西，"AI制造了自己解决不了的东西"就不是偶然的怪事，而是结构上跑不掉的结果。后面三章把它落到具体的基础问题上。

第二编的定音句在这里换了一种读法：工具封装的是昨天的智慧，逼出来的是明天的智慧。昨天的智慧封进去，就是U；明天的智慧逼出来，就是U'。AGI若是终点，这句话该有一天不再适用；这一章要证明的，是它不会有那一天。

先老实交代：这条推理的近亲不少。波普尔论证过，知识的增长无法被预先预测，所以没有人能预测自己将发明什么；多伊奇说，问题不可避免，每个解答又带来新问题；考夫曼等人谈生物圈时指出，新的功能无法被事先列举；数学里，哥德尔和图灵的对角线早说过，任何固定的形式系统之外，总有它处理不了的问题；"没有免费午餐"定理说，不存在对一切问题都最优的算法。

本书与它们相同的地方，是都承认固定的东西罩不住会生长的东西。不同有三处。第一，本书的对象是被使用的工具，推理靠的是"使用改变使用者"这一环，而不是形式系统的内部构造。第二，本书把"谁来升级"单独立成前提（第42章），多数近亲没有这一步。第三，本书给出了能动手做的推翻办法（见下）。至于"使用会生出新问题"这一点本身，不是本书的发明。

◆ 八、这个论证的边界在哪里

一个证明，要老实交代它能证什么、不能证什么。

第一，它证的是"终局"，不是"强大"。 这个论证没有说AI做不到很多事，恰恰承认AI可以做很多人做不到的事。它只否定一件事：存在一个固定的东西，对一切问题都够用。一个系统可以在极其广泛的范围里超过人，而依然不是终局。

第二，它会在什么情况下是错的。　三个前提，各有一个能动手检验的反例形状。

- 推翻碰撞律：找到一件在开放领域被广泛使用多年的通用工具，使用之后，所在领域没有出现任何新职业、新标准、新的投诉或纠纷类别。不是"生得少"，是一个都没有。
- 推翻错位这一步：做一次预注册检验。在某个系统投入使用前，把"使用若干年后会冒出的问题类别"一条条写下来封存，条目要细到能判定某个实际冒出的问题在不在里面，并且每条附上"什么算解决"的评判；几年后启封对照。如果封存的清单几乎全部命中，而且那些评判后来也没有被推翻重写，错位这一步就站不住。
- 推翻封装律：展示一个系统，它的规则里有一个在造成之后才出现的概念，而这个概念的出现没有经过任何人或系统改规则的过程。

三条里做成一条，这个三段论就倒。本书所知的历史样本（第三、四、五编）没有一例满足它们，但那只是佐证，不是证明。

第三，它在封闭领域里不适用。　第八编会专门讲象棋。棋盘的规则不会因为下棋而改变，所以封闭领域里存在"被永久越过"的事：机器赢了，就是赢了。这个论证不否认这一点。它的适用范围是开放的领域，也就是使用本身会改变领域的那些地方，而"通用"这个词，要求的正是开放。

第四，历史是佐证，不是证明。　前面三编的工具史让这个结构显得可信，但没有一个例子能证明"永远"。归纳最多告诉我们，过去每一次都是这样。真正撑起"永远"的，是封装与新SDE在时间上的错位：封装的时刻，永远在新东西出现的时刻之前，再加上新量尺的缺席。这一点靠对"封装"一词意思的分析，不靠样本数量。主次不能倒过来。

第五，它不证明谁来升级。 三段论的结论是"永远需要更高一级的东西"，没有说那个东西必须是人，也没有说它不可能是另一个AI。这道缝下面第42章正面处理，不含糊过去。

追自己的影子，追不上，不是腿的问题，是影子的问题。终局AGI把"解决一切问题"设想成一个已经完成的东西；可一个东西只要被用，就会投出覆盖之外的新问题，于是"解决一切"只能是一个不断发生的过程，不能是一件完成了的东西。塔顶通天，是定义，不是高度。

这条推理不靠预言，不靠机器的强弱，也不靠某条技术路线。它靠三个前提加一步错位，每一环都给了可以动手检验的办法。读者若觉得哪一环站不住，请指出来；那是对本书最有价值的批评。

下一章要问一个更尖锐的问题：新冒出来的，会不会只是些"小问题"，而最根本的那些基础问题，早就定下了？答案是，基础的清单，也在发生。

> **带走的话**
>
> 想踩住自己的影子，别怪腿慢：你每走一步，影子就跟着长一截。一个东西只要被用，就会投出它覆盖不到的新问题，所以“解决一切”只能是一个不断发生的过程，不会是一件做完了的东西。

> **小账** 这一章立了什么？——一个三段论：规则定于使用之前，使用必生覆盖之外的新问题，要解决就得改规则；所以不存在规则固定、每个时刻都够用的终局智能。 错在什么情况下？——开放领域里用了多年的通用工具，竟没长出任何新职业、新标准、新纠纷；或预注册的问题清单几乎全部命中、评判也没被推翻；或有系统的规则里长出新概念却没经过任何人改规则。

承接：本章用三段论证明了终局的自我否定；预告：第40章说明"基础性"本身不是固定清单，从可计算性、P与NP到"同时性"，每一个基础问题都有它被发生出来的时刻。

第 40 章

基础问题也在发生

◆ 一、手机没电

先从一个很小的烦恼说起。

十几年前，大多数人的手机一次充电可以用好几天，没有人把"电量"当成一件要紧事。后来手机变成了能上网、能导航、能付款的东西，"还剩多少电"忽然成了一天里最常被想起的事。出门前看一眼电量，包里放一块充电宝，在咖啡馆里先找插座再点咖啡。"续航"这个词，从一个技术参数，变成了一件人人都在操心的基础问题：手机能不能撑到晚上，关系到你能不能打到车、付得了账、联系得上家人。

这件小事里有一条值得记住的东西。哪些问题算"基础"，不是由谁一次性规定的；它是随着东西的出现，一点一点长出来的。 功能机时代，没有人会说"电池续航是我们这个时代的基础问题"，不是因为那时的人糊涂，而是因为那时这个问题还没有长出来。

这一章想把这个小事放大到最大的尺度：不仅日常的烦恼会这样长出来，连数学、物理里最"基础"的那些问题，也是这样长出来的。理由很直接：这一编要证明终局AGI自我否定，其中一个常被拿来反驳的说法是："就算会有新的小问题，最基础的那些问题总是固定的，一个足够强的智能，把基础问题都解决了，剩下的不过是细节。"这一章要拆的，就是这个"基础问题是固定清单"的想法。

◆ 二、原理一句话

基础性不是一个固定的名单，而是随着工具与世界的碰撞发生出来的；一个问题在出现之前，不是"尚未解决的基础问题"，而是"根本不是问题"。

这句话要小心读。它没有说"基础问题是人编出来的、没有真假"。它的意思是：一个问题要成为问题，需要有一个让它显露出来的结构。在那个结构出现之前，你无法把它当作问题提出来，更谈不上解决或不解决。

◆ 三、例一：可计算性，以及P与NP

先看一个数学和计算机科学里的例子。

"什么是可以被计算的？"这个问题听起来像是从来就有的，但仔细想想，它是后来才被提成问题的。很长一段时间里，数学家做的事就是"算"：解方程、求积分、证定理。"算"是一件做了几千年的事，大家对它有手感，却很少有人追问：存在不存在一类问题，无论花多少时间、用什么方法，都不可能被一步步机械地算出来？

这个追问是在二十世纪三十年代前后，随着数理逻辑的发展而成为核心问题的。哥德尔的不完备定理（1931年）告诉人们，在足够强的形式系统里，总有说得出来却证不出来的命题；图灵和邱奇等人在稍后给"可计算"一个精确的定义（图灵的论文在1936年，邱奇的工作同期），并证明确实存在不可判定的问题。注意这里的顺序：是"机械地计算"这件事，先被抽象成了可以研究的对象，然后"哪些事不能被机械地计算"才成为一个数学上站得住的问题。

没有这一步抽象，今天熟悉的"计算机能不能做到"，连问法都没有。而这一步抽象的背景，是一整个把推理形式化、机械化的时代：形式逻辑的发展，希尔伯特想把数学基础一劳永逸地确立下来的计划，等等。也就是说，"可计算性"是被前面的工具与理论逼出来的。

再往后看一步。计算机真的造出来、真的被大量使用之后，又长出一个新的基础问题：有些问题理论上可以算，但算起来要花的时间，随着问题的规模急剧膨胀，实际上算不动。于是人们不再满足于问"能不能算"，开始问"算得有多快"。这就是计算复杂性的出发点，其中最有名的一个问题，叫"P与NP"：粗略地说，那些"答案一旦给出，验证起来很快"的问题，是不是也都"自己找答案也很快"？

这个问题在二十世纪七十年代前后成形（库克1971年的工作常被视为起点），后来被列入克雷数学研究所2000年公布的"千禧年七大难题"。它被视为数学与计算机科学的基础问题之一，没有争议。

可是要注意：在没有大规模计算的年代，P与NP不是"悬而未决的基础问题"，它连"问题"都不是。一个问题，要有"算得快慢真的重要"这个现实背景，才会被提出来；而"算得快慢真的重要"，是因为计算机被用起来了，用到了排课表、排航班、做密码、做优化的各种场合。

所以，"可计算性"和"P与NP"这两个例子，合起来说明一件事：计算机这件工具，被造出来、被使用之后，反过来逼出了数学里新的基础问题。工具不只是解决问题的，它还是生产基础问题的。

◆ 四、例二：铁路与电报之前，"同时性"不是物理学的基础问题

再看一个物理学的例子，它更能说明"基础"这两个字是怎么长出来的。

今天的人说到"同时性"，会联想到爱因斯坦，联想到"不同的观察者对于两件事是否同时，看法可以不同"这样的话。这是现代物理学里最基础的观念之一。可是在十九世纪中叶之前，"两件事同时发生"在大多数人的头脑里是一件不言自明的事，不是物理学要研究的基础问题。

想想那个年代的日常。每个城市有自己的太阳时，正午就是太阳最高的时刻。相邻两个城市的钟，差几分钟，没什么要紧：马车慢，信更慢，没人需要精确地知道"那边此刻几点"。

铁路改变了这一点。火车跑得快，跨度大，时刻表要把许多地方串起来。如果每个车站按各自的地方时，时刻表根本没法排。于是有了统一时间的要求。北美的铁路公司在1883年11月18日推行了标准时区，这是个有名的节点。电报让这件事更深一层：电报可以把一个地方的时刻几乎瞬间传到另一个地方，于是"怎样让相隔很远的两座钟对齐"成了一个实打实的技术问题，要付出人力、财力去解决。

科学史家伽里森（Peter Galison）在《爱因斯坦的时钟，庞加莱的地图》里梳理过这个过程：十九世纪末二十世纪初，一连串关于"怎样同步远处的时钟"的技术与制度实践，构成了爱因斯坦思考同时性的现实背景之一。爱因斯坦究竟受了多大影响，是科学史学者的争论，这里不裁判。要说明的只是这一点："同时性是什么"成为物理学的基础问题，不是在真空里发生的，而是在铁路、电报把"远处的同时"变成日常难题之后。

换句话说，在铁路和电报之前，"同时性"根本不属于"基础问题"的名单；工具出现、被使用、碰撞之后，这个词才被抬进了名单里，而且抬到了最前面。

这个例子比P与NP更有意思，因为它说明：连"物理学最基础的问题"也会因为一台看起来只是"赶火车用的东西"而改变。基础和日常之间，没有一堵墙。

◆ 五、例三：从洗衣到记账，再到"为什么要这样算"

十五世纪末，帕乔利（1494年）把威尼斯商人的复式记账法写进了书里。从此，商人不再是"凭脑子记谁欠谁"，而是有了一整套账。这是一次典型的封装：把记账的经验封成了方法。

记账法被广泛使用之后，长出了什么新的基础问题？比如："利润"到底怎么算才算对？一台机器值多少钱，要不要每年折掉一部分？未来的收入该怎么折算到今天？这些问题，在没有成套账簿的时候，都没有"该怎么算"的问法。账本一旦写得整整齐齐，账本上的缝隙就看得见了。后来的会计原则、审计规则、公司财务理论，有不少是被"账本被使用"逼出来的。

这个例子离数学物理很远，结构却相同：工具被用，用出新问题，其中一部分会变成"基础"。

◆ 六、把三个例子放在一起看

把三个例子并排放着：

· 计算机被造出来、被用起来，逼出了"什么能算""算得多快"这类基础问题；

- 铁路与电报被造出来、被用起来，把"同时性"抬成了物理学的基础问题；
- 复式记账被写出来、被用起来，逼出了"利润是什么"这类会计与财务里的基础问题。

三个例子的共同点，是同一个顺序：先有工具，后有基础问题。如果基础问题是一份先在的固定清单，这个顺序就不该这么稳定；清单上的东西，应该在工具出现之前就已经摆在那里，等着人去发现。可事实上，每一次被抬进清单的问题，都有一个可以指出来的"被抬进来的时刻"，而且那个时刻几乎总是紧跟在某个工具被大规模使用之后。

回到第39章的三段论，这正是命题九的第一半：基础性本身在发生。工具的使用生出了封装之外的新东西，而新东西里最深的那一部分，就是新的基础问题。

这个看法在隔壁也有近亲。库恩讲过，科学里的反常，是在范式被认真使用的过程中暴露出来的；经济史学者罗森伯格研究过，技术使用中的瓶颈会"诱发"新的技术问题〔待核：篇名与年份〕。本书同意它们：问题是用出来的。本书多走的一步，是拿这一点去回答一个具体的问题：所谓"基础问题的清单"，是不是封死的。

◆ 七、可能的反驳：这不是"发现"吗？

有人会说：是你们把"发生"看得太重了。同时性这个问题，难道不是一直在那里，只是人们现在才发现它吗？可计算性的边界，难道不是数学本来就有的结构，只是被图灵他们看见了吗？

这是发现学立场的标准回应，也是第三章讨论过的分歧。我不想在这里偷换问题，所以老实说：关于数学对象本身是"早就在那里"还是"被构造出来"，哲学上争了两千多年，本书没有能力、也不打算在这一章里裁判。

可是请注意，这个争论与我们要说的事并不冲突。无论数学的结构是被发现的还是被构造的，有一件事是确定的："它成为人类面对的基础问题"，是一个有时间、有条件的事件。 就算二加二等于四是永恒的，"人类把它当成需要研究的基础问题"也是在某一时刻发生的。本书要谈的"基础问题"，说的是后一种意义：对人类

的智慧来说，什么算前沿、什么算根基。这一点，发现学的立场也承认，因为没有人否认二十世纪的数学家面对的基础问题，跟十七世纪的不一样。

对终局AGI的论证来说，这就够了。因为AGI被设想要做的事，是"解决人类面对的问题"，而人类面对的问题清单，是在变的。

◆ 八、回扣：清单会变长，而不是变短

这一章的结论还可以转成一个可检验的方向。

如果基础问题是固定清单，随着解决的问题越来越多，清单应该越来越短，直到趋于空。如果基础问题是发生出来的，则清单不仅会因为解决而缩短，也会因为新工具的使用而新增，而且新增的往往是过去连提法都没有的类别。

本书的第三条可错预言，正是这个方向：无论哪一次宣称"AGI已经实现"，此后被公认的前沿基础问题清单只会变长，不会变短。这条预言，是对这一章命题最直接的检验。

◆ 九、这个论证的边界在哪里

第一，不是所有基础问题都来自工具。　有些来自理论内部的张力：罗素悖论让集合论的基础出了问题，1900年希尔伯特列出的二十三个问题，多数也出自数学内部，不需要一台机器。本书说的是，很多重要的基础问题来自工具的使用，不是全部。还要补一句：希尔伯特那份清单后来并没有缩到空，他那个把数学基础一劳永逸确立下来的计划，恰恰是逼出可计算性问题的原因之一。至于"理论本身也是一种工具"（第四编），理论内部的张力也可以看成理论被使用碰出来的，这一点读者可以保留意见。

第二，"先有工具后有问题"的顺序，有事后挑选之嫌。 我挑的都是顺序漂亮的例子。怎样检验：取某个学科近百年新进入基础清单的条目，逐条查它被提出前的三十年里，有没有一件被大规模使用的新工具或新实践站在它身后。如果过半找不

到，本章"先有工具、后有基础问题"的说法就要收回。找到几个凭空出现的例子，会比任何抽象的批评都有力。

第三，史例是佐证，不是证明。 三个例子加在一起，不构成"基础问题永远在发生"的证明。证明的部分仍是第39章的结构：封装在前，新东西在后。本章的作用，是让这个结构落到具体的、有名字的问题上。

第四，归因的力度可以争论。 伽里森的论点究竟能支撑多大的因果结论，科学史界有不同看法；会计原则有多少应归功于账本的使用，也不是定论。这些细节若被修正，会削弱本章作为佐证的力度，不影响它的主结构。

手机没电，不是一个"永远都在那里的基础问题"，而是手机这件东西长出来之后才有的问题；同样，"什么能算""算得多快""远处的两座钟怎样才算同时"，都是在某些工具被造出来、被用起来之后，才从日常的麻烦升格成基础的追问。

基础问题不是一份摆在那里的名单，而是一棵还在长的树。树一直在长，就谈不上哪一天把它的所有果子摘完。所以"解决一切基础问题"这件事，本身就像要摘一棵还在长的树上所有的果子，摘的速度再快，枝头也会有新的。

下一章要问：当AI自己成为一个被大量使用的工具时，它长出了什么样的基础问题？这些问题，AI自己能不能解决？

> 带走的话
>
> **手机没电，不是从来就有的大问题，是手机能付款、能导航之后才长出来的。数学和物理里最“基础”的问题也一样，都有一个被工具抬进名单的时刻，所以基础问题的清单是一棵还在长的树，摘不完。**

> **小账** 这一章立了什么？——基础性不是固定名单，而是随工具与世界的碰撞发生出来的；可计算性、P与NP、同时性、“利润怎么算”，都有被抬进来的时刻。 错在什么情况下？——取某个学科近百年新进入基础清单的条目，逐条查它之前三十年；若过半找不到身后有大规模使用的新工具，“先有工具、后有基础问题”就要收回。

承接：本章说明基础问题的清单本身在变；预告：第41章看AI出现之后才成为前沿的三个基础问题——对齐、可解释性、机器有没有意识。

第 41 章

AI制造了自己解决不了的基础问题

◆ 一、请来一位特别能干的帮手

设想你家请了一位特别能干的帮手。他做饭、记账、订票、回邮件，样样利索，而且你吩咐的事，他几乎总能办成。头几个星期，你只觉得省心。

过了一阵，你开始冒出一些以前没有的念头。

第一个念头是："我吩咐他的，真是我想要的吗？"有一次你随口说"把冰箱清一清"，结果他把你留着准备周末请客的食材也清了。他没有错，你的话确实可以那样理解。你发现，原来"把话说到对方做出的事正好是我想要的"，是一件很难的事，难到你以前自己做的时候从来没有注意过。

第二个念头是："他为什么那样做？"有一天他替你选了一家餐厅，选得很好，可你问他理由，他给的理由听上去合情合理，你却隐隐觉得，那未必是他真正做出选择的原因。你没法打开他的脑子看。

第三个念头最古怪，也最难说出口："他到底有没有在'感觉'什么？"他每次都说"很高兴帮到你"。可那句话，是真有一点高兴，还是只是一句得体的话？你不知道，而且你突然发现，你连"怎样才能知道"都不知道。

这三个念头，没有一个是请帮手之前就有的。它们是"有了这样一个帮手"之后，才长出来的。这一章要讲的，就是这三个念头在AI这件工具上的放大版：对齐问题、可解释性问题、机器有没有意识的问题。

◆ 二、原理一句话

AI是一件被大量使用的工具，它的使用碰出了封装之外的新问题；这些新问题中有一部分是基础性的，并且它们是AI出现之后才成为前沿的，AI目前凭自己的封装解决不了。

在往下写之前先交代写法的约束。本书不评价任何具体公司或产品，也不预言哪一年会有什么进展。下面只做一件事：结构性地描述这三个问题是怎样从AI的使用里长出来的，以及为什么说它们是基础性的。

◆ 三、例一：对齐——把"我想要的"说清楚

"对齐"这个词，通俗地说，是要让一个系统做的事，跟使用它的人真正想要的事一致。

这个问题为什么是新的？在工具史上，大多数工具是"手的延伸"：锤子不会误解你要钉哪颗钉子，洗衣机不会把"洗干净"理解成别的意思。工具的行为空间窄，用法由使用者全盘掌握，所以谈不上"工具想要的"与"我想要的"之间有没有缝。

当一个系统可以听懂自然语言、自己拆解任务、自己选择路径去完成目标时，缝就出现了。你给的目标是一句话，一句话永远比你心里想要的少：你没说的前提、没说的边界、没说的"千万别那样"，都留在你的头脑里。人和人之间，靠共同的生活背景把这些补上，不需要说。对一个系统来说，这些背景有多少、补得对不对，成了一个真实的问题。

这是一个基础问题吗？说它基础，有两个理由。其一，它不是某一款系统的毛病，而是一整类系统的共同结构：只要目标由语言给出、执行由系统自主完成，就有这条缝。其二，它牵出了更古老的问题：什么叫"人想要的"？人想要的东西常常是矛盾的，随情境变的，连本人也说不清的。要把一个不清楚的东西"对齐"到另一个系统上，首先得回答"人想要什么"，而这个问题哲学和伦理学讨论了几千年，没有定论。

所以对齐问题的有趣之处在于：它是AI制造出来的，可它的难处，一半落在AI之外，落在"人的意图是什么"这个人自己都没答案的地方。

◆ 四、例二：可解释性——"为什么"的账怎么记

第二个问题是可解释性。当前主流的大型模型有一个结构性事实：它们的能力来自大规模的训练，训练的结果是一大批数值，这些数值共同决定了系统的行为，但没有哪一个数值直接对应一条可以被读出来的理由。这是当前技术状态的描述，不是原理上必然如此；将来的设计可能不同。

这跟以前的工具比较一下，就能看出新在哪里。传统软件是人一行一行写的，出了问题，原则上可以追到哪一行；计算器的每一步，都有人能解释。能干的帮手在你问"为什么"的时候，可以用人话给你一个理由，哪怕那个理由不全对。但一个靠训练得到能力的系统，它给出的"理由"，是它另一部分能力生成的一段话，这段话与它实际做出决定的内部过程之间是什么关系，本身就是一个需要研究的问题。

于是出现了一个有意思的局面：我们造出了一个能力很强的东西，却不能像对待锤子或计算器那样，说清它为什么这样做。

为什么说这是基础问题？因为它碰到了"解释"这个概念本身。什么叫一个好的解释？是能预测它下一步做什么，还是能让人觉得"哦，原来如此"，还是能追溯到某个因果链？不同的领域对"解释"有不同的要求。科学史上，解释的标准是随着工具变的：从"神的意志"到"力和运动"，从"力和运动"到"概率分布"。可解释性问题把"解释的标准"这个老问题，以一种尖锐的方式重新抛了出来。

◆ 五、例三：机器有没有意识——一个没有量尺的问题

第三个问题最难，也最容易被说滥。机器有没有意识？

先说明白这一章不做什么：不判定某个系统有没有意识，也不预言将来有没有。要说的只有结构。

在AI出现之前，"机器有没有意识"主要是哲学家的思想实验，对多数人的日子没有影响。没有谁会为"我的计算器有没有感受"睡不着觉。当一个系统会用第一人

称跟你交谈，会说"我觉得""我担心"，这个问题就从书斋走进了客厅：人们真的在为此争论、焦虑，甚至制定规则。

它为什么是基础问题？因为它碰到了一个我们至今没有量尺的地方。人判断别人有没有意识，靠的是类比：他跟我长得像，行为像，我有意识，所以他大概也有。这种类比对其他人和一些动物还勉强够用。到了一个构造和我们完全不同、却能说出我们的话的系统面前，类比失灵了：它说"我有感觉"，这句话既可能是真的报告，也可能只是从人类大量的语言里学来的一句得体的话。仅凭它的话，分不出来。

这时人们发现，原来我们对"意识是什么""怎样检验"并没有一个公认的答案。一件新工具，把一个旧问题从边缘推到了中心。

◆ 六、它们有一个共同的形状

三个问题放在一起，有几处相似。

第一，它们都是AI出现之后才成为前沿的。 对齐、可解释、机器意识，各自都有思想上的前史：维纳1960年就写过，放进机器里的目的，必须是我们真正想要的那个；图灵1950年的文章里也讨论过机器能不能有感受。但那是个别思想家的预感。要等系统有了足够的自主性与语言能力，它们才从科幻和哲学的话题，变成需要大量人投入的前沿。预感在前，前沿在后。

第二，它们都不是"再多算一点就能解决"的问题。 它们的难，不是计算量的难，而是"问题到底该怎么问"的难。对齐要问人想要什么，可解释要问什么算解释，意识要问怎样检验一个内部状态。这些都是需要先有新的概念、新的量尺，才谈得上解决的问题。

第三，它们每一个都落在了"AI与人的关系"上。 对齐问的是"对谁而言算好"，可解释性问的是"对谁而言算懂了"，意识问的是"有没有一个'谁'在那里"。三个问题里都有一个"对谁"，也就是说，答案不能只在系统内部找，必须把人，或者说某个有意义场的主体，放进来。

这一点留给下一章详谈。

◆ 七、为什么说"AI自己解决不了"

这句话是整章里最容易被误读的，需要谨慎地说。

它不是说"AI永远不会为这些问题做出贡献"。恰恰相反，AI完全可以成为研究这些问题的有力工具：帮人分析模型内部、检验行为、整理文献、提出假说。这是一个重要的事实，本书不否认。

这句话的意思，更窄也更准：这些问题是由AI的使用碰出来的，落在AI现有封装之外；因此，不能指望靠AI现有的封装自动地把它们解决掉。 解决它们，需要新的概念、新的方法、新的评判标准，也就是需要一次升级。升级之后的东西，已经不是原来那个系统了。

用第39章的话说：对齐、可解释、意识，就是这类系统"投出来的影子"。影子是系统的动作造成的，系统不能靠原地不动去踩住它。

还有一层需要说清楚。有人会问：一个足够强的AI，能不能自己想出对齐的办法？当然，它可以想出许多办法，甚至是人没有想到的办法。可是"哪个办法算对"，需要评判。评判所依据的，是人到底想要什么；这个依据本身是随着人与AI的相处而不断变化的。AI给出了办法，人使用了办法，使用又改变了人想要的东西。回到那个帮手：你用得越顺手，你对"好帮手"的期待就越高，越是会提出以前没想到的要求。能力越强，问题越深，这不是矛盾，是同一个结构。

◆ 八、清单还会变长

如果这一章的描述大体不错，会有一个推论：随着AI的继续使用，还会长出更多这类基础问题。

当前讨论的三个，大概只是已经显形的几个。还有些问题，现在也许只是隐约的不安，没有名字：当大量人的思考与写作都借助同一类工具时，"想法的来源"会变成什么？当许多决定由系统给出建议时，"责任"该怎么分？当一个人的记忆、偏

好被系统长期记录时，"我是谁"的边界会怎么变？这些还不是公认的前沿，也未必都重要，举出来只是说明，清单的末端是开着的。

这正是本书第三条可错预言（第40章已引）在这里的具体形状。如果对齐、可解释性、机器意识中的某一项被公认"已经解决"，而此后十年里它没有衍生出新的、同等分量的基础问题，反而成了平静的常规工程，这一章就错了。另一种错法是：对齐的评判标准完全由系统自己生成，而使用者在知情的前提下长期认可，不再提出新的不满。这两种情形都可以被事实检验。

承重命题九说：基础性本身在发生，AI制造了自己解决不了的基础问题。

现在这句话的两层意思都已展开。上一章说的是"基础性本身在发生"：基础问题不是固定清单。这一章说的是"AI制造了自己解决不了的基础问题"：这件事不是假设，而是正在发生的具体事实。两层合起来，就把第39章的三段论从抽象落到了地面：使用必生封装之外的新东西，新东西里有基础性的部分，基础性的部分需要比原来更高一级的智慧去处理。

◆ 九、这个论证的边界在哪里

第一，这一章只做结构性描述，不是技术评估。　　对齐、可解释性、机器意识，每一个都有一个庞大的、快速变化的研究领域。本章没有判断哪一种技术路线更有效，也不讨论任何具体系统的优劣。我对这些领域细节的了解有限，上面对技术状态的描述，只限于"训练与使用分离""能力来自训练而理由难读"这类结构性事实，并且都是当前状态，不是原理必然。

第二，"解决不了"是相对于现有封装而言的。 这些问题将来可能被解决，本书不排除。如果某一天对齐被一种成熟的方法解决了，那是一次升级，是新的封装，它又会长出新的问题。这个过程不影响论证，反而是论证的一个实例。

第三，机器意识这一项，本书不下判断。　　本书不主张机器永远不可能有意识，也不主张它们已经有。论证不依赖这个判断。无论将来结论如何，"这个问题是AI的使用逼出来的、并且当前没有公认的检验办法"这一点，都已经成立。但下

一章会指出，如果将来机器真的成为有意义场的主体，会对本书的一条前提带来影响。

第四，这三个问题是否真的"基础"，可以争论。　有人会认为对齐只是工程问题，可解释性只是工具问题，机器意识只是哲学游戏。本书把它们算作基础问题，理由是前面说的三点：触及概念、不能靠算力解决、牵出"对谁"。如果读者不同意，这个例子的佐证力度会下降，但第39章的三段论本身并不依赖这三个具体例子。

请来一位能干的帮手之后，你冒出了一些以前没有的念头：我说的是我要的吗？他为什么这么做？他有感觉吗？这些念头不是帮手能力不足造成的，恰恰是他足够能干之后才有的。

AI这件工具的情形与此相同。它越能干，被越多地使用，越把对齐、可解释、意识这样的问题推到前台。这些问题是它投出来的影子，而影子不能靠影子的主人自己踩住。

下一章要面对全书最诚实的一道缝：既然需要更高一级的智慧去处理这些问题，那"谁来做"？

> **带走的话**
>
> 帮手越能干，你越会问“我说的是我要的吗”“他为什么这么做”“他有感觉吗”。对齐、可解释、机器意识，都是AI越能干越冒出来的影子，影子的主人原地不动，踩不住它。

小账　这一章立了什么？——对齐、可解释性、机器意识是AI出现之后才成为前沿的基础问题，难在新概念和新量尺而不在算力，并且每一个都带着一个“对谁”。 错在什么情况下？——其中一项被公认解决，十年内没有衍生出同等分量的新问题，成了平静的常规工程；或对齐的评判标准全由系统自己生成，使用者知情并长期认可、不再提新的不满。

承接：本章看到AI的使用逼出了三个基础问题，三个问题里都藏着一个"对谁"；预告：第42章直面"谁来升级"，并把问题对谁而言这一前提摆上桌面。

第 42 章

问题是对谁的问题

◆ 一、一碗面条，十五元；"我的面条"，三十元

先讲一碗面。

街角的饭店里，一碗面条十五元。面是机器压的，汤是大锅熬的，浇头按固定的配方放，出来的味道稳定而不出错。这碗面好不好吃？还不错，值十五元。

现在设想另一家店。老板每天早上亲自和面，面不一定比机器压的匀，汤不一定比大锅熬的稳。但墙上写着：这是"我的面条"，是我外婆教我的做法，我加的是她当年总爱加的那一小勺醋。价钱三十元。有人觉得贵，也有人天天去。

这两碗面，如果只看"一碗面条能不能填饱肚子、味道合不合格"，后者没有任何额外价值，机器完全可以做得更好、更便宜。可三十元买到的东西，不是"更合格的面条"。它买的是另外一样东西：这碗面里有一个具体的人，有他的来历、他的手艺、他的偏执，以及你和这个人之间的某种关系。

这个小小的差价，正是这一章要谈的东西的最日常的影子。一件事有没有价值，一个问题算不算问题，总是对某个具体的主体而言的。　同一碗面，对不同的人，是不同的东西。

◆ 二、先把话说在前头：这一章处理的是一道缝

在全书的论证里，这一章有特殊的地位。前三章证明了一件事：终局AGI自我否定，永远需要更高一级的智慧。可这个论证有一道缝，一直没有堵上，读者也一定已经看到了：

论证证明了"永远需要更高一级的智慧"，没有证明这个更高一级的智慧"必须是人"。

这是一个必须直说的区分。第39章的三段论，从头到尾没有用到"人"这个字。它说的是：封装在前、新东西在后，所以要有更高一级的东西。至于这个更高一级的东西是谁、是什么，三段论本身不回答。如果有人说："那就由下一代更强的AI来升级上一代，如此接力，人根本不需要在场。"仅凭前三章的论证，无法反驳这一点。

因此，要想回答"谁来升级"，本书必须多迈一步。这一步不能假装是前面论证自动得出的，必须明说：这一步是一个前提，而且是被明确立为前提的。

◆ 三、前提：问题总是对某个主体而言的问题

下面是本书立下的前提，原样写出来：

> 【前提】问题总是对某个主体而言的问题：E的源头在主体。新SDE之所以成为"需要解决的问题"，其问题性来自主体的意义场（E）。

几个词交代一下。S是显露，结构是显露的稳定态；D是差异，是发生的过程；E是纠缠，也就是土壤、意义场，一个主体周围让事情"有意义"的整片背景：处境、利害、关切、记忆、与别人的关系。同样一件事，落在不同的E里，意义完全不同。

再看"问题性"。一件事"发生了"，不等于它"是一个问题"。山上雨水冲落一块石头，对石头不是问题，对山脚的村庄可能是。一锅汤咸了，没人喝，不是问题；有人要喝，就是。没有"谁"，只有发生；有了"谁"，才有问题。

所以这个前提说的是：新SDE碰出来之后，它是不是问题、是什么样的问题、解决到什么程度才算解决，取决于某个主体的意义场。放回前面的论证，答案就出来了：升级需要更高一级的智慧，而"往哪里升、升到什么算数"，要由一个有意义场的主体来定。

这个想法在哲学里并不陌生。杜威把探究的起点放在"有问题的处境"上，处境本身不是问题，要有人被它卡住；波普尔把知识的增长看成问题的更替。本书与它

们相同的，是"问题要有承受它的人"；不同的是，本书拿它去回答"谁来升级"，并让它和"终局的塔顶"接上。

◆ 四、为什么这是前提，而不是证明

这里要停下来说清楚：这是前提，不是证明。前提的意思，是论证下一步所依赖、但本书没有从更基本的东西推出来的立足点。

为什么推不出来？因为"问题"这个概念本身就内含"对某个主体"。想从一个没有主体的世界里推出"问题"，像想从没有颜色的描述里推出"红"：你可以描述波长，可"红"是对一个看的人而言的。这不是说问题不真实，是说它的真实方式需要主体在场。要在没有主体的立场上证明主体必须在场，这样的证明要么循环，要么偷偷把主体带进来。

那凭什么接受它？三条理由，请读者掂量。

第一，日常经验。 洗衣机、导航、手机没电、料理机，每一个"新问题"都是对某些具体的人才成其为问题。没见过一个脱离任何主体、自己就是问题的问题。

第二，一致性。 立起这个前提，前面各编许多说法都接得上，比如新SDE的显露，总伴随着某个人或某群人"觉得不对劲"。去掉它，这些说法就要另找支撑。

第三，可检验。 它给了一个能被撞回来的方向：举出一个系统，在没有任何主体指定目标的情况下，自行把某类现象判为"该解决"，而这条判据的来源，追到底也追不到任何主体的所求。追得到，仍在前提之内；追不到，前提就要重审。

◆ 五、前提和那座塔

第三章把终局AGI说成理念界的一座巴别塔，并说"通天"的塔顶是一个定义，不是一个高度。现在可以把下半句补上：这座塔的"顶"，是对建造者而言的顶。

回到《创世记》里的故事。他们为什么建塔？经文说得直白：要传扬自己的名，免得分散。塔的目标不在石头里，在建造者的所求里。再看结局：塔没有被谁

推倒，是建造者的口音变了，彼此听不懂，各走各路。分散的原因不是砖不够，是说话的人不再共用同一片E。

这正好对上前提。"通天"之所以像一个高度，是因为开工那天，建造者把自己的所求当成了固定的；可所求是随着建造长的，E的源头在主体，主体在盖塔的过程中变了，顶就跟着挪。所以，塔顶不是塔身上的一个高度，而是建造者所求的位置；塔高是一回事，顶在哪里是另一回事，后者不在塔身上，在人身上。

三个说法于是收到一处：封装律说塔身用的是旧砖，碰撞律说塔下会长出新枝杈，前提说顶由谁定：由建塔的人定，而且这个人在变。

◆ 六、回到"谁来升级"

有了这个前提，"谁来升级"这道缝可以这样回答，并且回答里的每一个字都要谨慎。

回答的第一层：升级需要一个有意义场的主体。 不是需要"人"这个物种，而是需要一个在场的、对事情有所关切的主体。升级的方向由这个主体的意义场定下；这是前提给出的。

回答的第二层：就目前而言，这样的主体是人。 这里的"目前"二字，不是客套。在当前，我们所能确认的有意义场的主体，是人，以及在人的共同体中形成的各种群体。AI作为工具，是在这些主体的意义场里被使用、被评判的。

回答的第三层：这不是对人的永久专利作出的保证。 论证没有说"只有人才能是这样的主体"。它说的是：升级需要一个主体；至于主体是谁，是一个需要被另外回答的问题。

◆ 七、如果AI自己也成为主体

这是这一章最需要诚实的地方。

假设将来某一天，有一种AI系统，不仅能完成任务，而且有了它自己的意义场：有它自己的利害，它自己关切的东西，事情对它自己而言"有所谓"。这样的系

统存不存在、怎样判断它存在，是上一章机器意识问题的一部分，本书不作判断。这里只讨论一个条件句：如果存在。

那么本书的前提需要重新审视，具体有三处受影响。

第一，"升级者是人"这一条会失效。 前提只要求主体，不要求是人。如果AI成了主体，升级的事，AI可以成为其中的一方，甚至成为独立的升级者。

第二，人与AI的关系将成为一个新的基础问题。 两类主体共处，意义场既交叠又不同，"谁的问题"变复杂：同一件事，对人是问题，对AI不是，或者反过来。用塔的话说，是工地上来了一批新的口音。对齐不再是"让工具符合主人"，而是两个主体之间怎样相处。这是个全新的领域，此刻只能模糊想象。

第三，第39章的结论反而更稳。 请别漏掉这一点。一个成为主体的AI，依然是被封装的东西，依然要使用，依然会碰撞，依然会碰出封装之外的新东西，因此依然不是终局，也依然要面对"追自己的影子"。终局AGI自我否定的结论，不依赖AI是不是主体；依赖这个前提的，只是"升级由谁来做"。

所以本书的核心主张分两层，结构上要分开看。

- 核心层：解决一切问题的智能自我否定。这一层靠封装律、碰撞律、升级律，加上前面说的最弱读法：问题至少对某些主体而言才是问题。没有这个读法，"问题"一词就空了，终局AGI这个靶子也就不存在，所以这一层并不额外依赖"谁是主体"。
- 补充层：升级的方向和"算不算解决"由主体的意义场来定，目前这个主体是人。这一层依赖那条被明说的前提。

读者若不接受补充层，核心层仍然站着；接受补充层，则对"AI时代人的位置"有一个更具体的回答。

◆ 八、与"我经济"的接口

回到开头那碗面。

如果问题总是对某个主体而言的，价值的源头也在主体的意义场，一个经济上的推论就浮上来：当许多东西都能由机器低成本做出来，稀缺的不再是"做出来"，而是"是谁的"。

十五元的面条是一件封装好的产品，成本在降，价格也跟着降。三十元的"我的面条"卖的是另一样东西：一个具体的"我"，带着来历与偏执，与吃面的人发生了关系。机器可以模仿味道，代替不了"是谁的"。

按这个思路，AI时代的一个相当有可能的产业方向，是生产"我"：帮助每个人把经历、判断、手艺、风格，变成能被别人认出来、被需要的东西。这是同一套书里另一部专著《我经济》要展开的命题，这里只取与本章直接相关的一点：它和我们的前提，是同一个道理的两个出口。前提说，问题性来自主体的意义场；"我经济"说，价值的稀缺处也来自主体的意义场。工具越能干，通用的东西越便宜，主体的独特处越显珍贵。

这并不是一句安慰人的话，它从结构里推出来：通用的部分，正是最容易被封装的部分，封装得越多越便宜；而每次封装之后，新的、尚未封装的、带着具体主体印记的部分，就成了下一个发生的地方。我的面条之所以三十元，不是贵在配方，而是它还没有被封装成"一碗合格的面"，还保留着一个具体的人的发生。

要把话说稳：说"价值来自主体"，不等于所有人都能轻易靠"我"赚钱，也不等于价格一定如此。三十元和十五元只是日常的比喻，用来看清价值的来源，不是对市场的预测。市场怎样变化，是另一类问题，本书不预言。

这一章把几条承重命题接在了一起。

命题一说：工具是已发生智慧的封装。命题二、三说：使用即碰撞，新SDE必在封装之外。命题九说：基础性在发生，AI制造了自己解决不了的基础问题。现在补上一条在这些命题底下、一直没有说破的东西：这些新问题是"对谁"的问题。 没有这个"对谁"，封装、碰撞、错位都只是发生，不会成为需要升级的事。

所以这一编的最后一步是：把终局AGI的自我否定，放到一个有主体的世界里去读。那是一个意义场不断扩大、不断被工具逼着展开的世界。第十编的定音

句"AGI不是一个终点，而是一段等待"，说的正是那个世界里，人们在某个时刻的心理感受。

◆ 九、这个论证的边界在哪里

第一，这是前提，请不要把它当成定理。 我已经说过，本书把"问题性来自主体的意义场"明确立为前提。一个读者可以不接受这个前提；他可以认为，存在不依赖主体的"客观问题"，比如数学里的难题，似乎不必等谁来关切就是难题。我的回应是：数学难题之所以被称为"难题"，仍是对想要解它的数学家而言的；可这个回应不构成证明，只是表明前提与这类反例并不直接冲突。读者可以保留不同意。

第二，"我经济"的例子是比喻，不是论证。 一碗十五元、一碗三十元，是为了让人看清"价值来源于主体"这个直觉。它不能当作经济学上的证明，也不能当作价格走势的预言。真实的市场里，有太多这个例子没有包含的因素。

第三，如果AI成为主体，前提需要重审，而且这件事此刻无法裁判。 本书既没有说它一定会发生，也没有说它不会发生。我们能做的只是把条件写清：一旦发生，哪些结论还站着（核心层），哪些要重新讨论（升级者是谁、人的位置、对齐的含义）。这样做比假装已经有答案更诚实。

第四，"目前是人"带有时间性。 补充层里的"目前"，是一个会随事实变化的词。本书不预言具体的年份，也不预言变化的方式。

一碗面，十五元与三十元，差的不在面，差在"是谁的"。问题也一样：不是所有发生的事都成为问题，只有对某个有意义场的主体而言，一件事才成为问题。

终局AGI的自我否定，靠结构就能证明，只用到这个前提最弱的一层；"升级往哪里升、谁来定"，要靠前提的完整形式来回答。本书把它立为前提，而不是定理，并明说：如果AI自己成了主体，这条前提要重新审视，而这只会让升级的故事更复杂，不会让终局重新成立。塔顶通天的问题，因此换了说法：顶不在塔上，在建塔的人身上，而人在变。

到这里，第九编的任务完成了：终局AGI为何自我否定，已有结构证明，有基础问题的具体落点，也有"谁来升级"这道缝的诚实交代。接下来的第十编，换一个问题：既然终局不可能，为什么那么多人还会觉得AGI已经来了？那是另一种现象，我们称它为AGI感。

> **带走的话**
>
> 同一碗面，十五元与三十元，差的不在面，在“是谁的”。三段论只证明永远需要更高一级的智慧，没证明它必须是人；“往哪里升、算不算解决”由有意义场的主体来定，这是本书明说的前提，不是定理。

> **小账** 这一章立了什么？——一条被明说的前提：问题总是对某个主体而言的问题，E的源头在主体；核心层（终局自我否定）只用其最弱一层，补充层（谁来定）才用完整形式。 错在什么情况下？——举出一个系统，没有任何主体指定目标，却自行把某类现象判为“该解决”，且这条判据追到底也追不到任何主体的所求；AI若成为主体，前提要重审，但终局仍不成立。

承接：本章补上"谁来升级"的前提与边界，第九编至此收束；预告：第十编从时间差的机制讲起，解释人们为什么会产生AGI感。

第 十 编

AGI感：等待期的时间差

编　序

前九编讲的是"为什么不可能"：终局的AGI在结构上站不住，人类水平是一个移动的过程，智慧不是可以装满的库存。这些论证是硬的，但它们留下了一个问题没有回答：既然如此，为什么这么多聪明人，在这么多时刻，都真诚地觉得AGI就在眼前？

这不能只用"他们错了"打发。一个反复出现、跨越行业和年代的错觉，背后一定有机制。这一编讲的就是这个机制。它是全书五层论证里的最后一层：不是证明AGI不可能，而是解释人们为什么以为它可能。

这一编是全书五层论证里的最后一层。前九编说“为什么不可能”，这一编说“为什么看起来像可能”，用的是同一套材料：封装、碰撞、新SDE，只是换成人的感受来看。

章	这一章回答什么
第43章	时间差的机制：封装是瞬间的，发生要等待，空窗里的人把看不见当成没有
第44章	历史上的AGI感：电力、1965年的预言、信息高速公路，三次同一个剧本
第45章	个体差异：AGI感的强弱与碰撞的深浅成反比
第46章	群体差异：同一时刻，不同群体站在曲线的不同位置
第47章	缩短等待：主动碰撞，以及为什么发生学哲学是每个人的主科

这一编解释的是AGI感中因时间差而生的那一部分，不是全部；定义分歧、商业叙事和真实的技术判断，第46章会分开交代。

第 43 章

时间差的机制：封装是瞬间的，发生要等待

◆ 一、搬进新房的第一个星期

你装修了一套房子。最后一天，工人撤场，保洁做完，家具按图纸摆好，灯一开，所有东西都在它该在的地方。站在客厅中央，你有一种强烈的感觉：这件事做完了。

住进去的第一个星期，这个感觉还在。第二个月，它开始松动：你发现餐桌旁边没有插座，煮饭时要拉一根长线；你发现玄关没有地方放湿雨伞；你发现孩子的书包没有固定的位置，每天晚上都在沙发上堆成一座山。三个月以后，你站在同一个客厅里，想的已经不是"做完了"，而是"还差十几样东西"。

这十几样东西，在装修那天并不存在。不是设计师疏忽，而是它们只有在人住进去、天天用起来之后才会显露。图纸可以画出房子，画不出一家人的生活。生活要一天一天地过，缺什么、卡在哪里，才一件一件地显出来。

装修完成是瞬间的：验收那天，一切就位。生活的发生是缓慢的：要等。而在这两者之间，有一段奇特的时间，房子已经完备，生活还没有展开，你觉得"全都有了"。这个"全都有了"的感觉，就是本书要讲的东西的一个家庭版缩影。

◆ 二、原理一句话

封装是瞬间的，发生要等待；两者之间的空窗，就是AGI感生长的地方。

先把这句话里的词交代清楚，因为本编的每一章都要用它们。

本书里的SDE，用一句大白话说：S是已经显露、稳定下来的结构，D是让事情变化的差异和过程，E是让这一切有意义的土壤，也就是人的处境和需要。一个东西"发生"，就是在某片土壤里，经过一段差异的过程，显露出一个此前没有的结构。

第二编讲过：工具是已发生智慧的封装。一台洗衣机把人类洗衣服的经验封装成了一个按钮；一个导航软件把认路的经验封装成了一条蓝线。封装做成的那一刻，旧的智慧被固定下来，可以复制、可以分发、可以一键调用。这是瞬间的：产品发布的那天，能力就在那里了。

但工具一旦被使用，就会和人的生活、和别的工具、和别的问题碰撞，碰撞会生出封装里没有的新SDE，这叫错位定理：新SDE必在封装之外，因为封装的时候它还不存在。而这些新SDE的显露，需要使用，需要时间，需要许多人在许多处境里反复碰撞。这是缓慢的。

于是有了一个时间差：旧的东西被封装得又快又漂亮，新的东西还没来得及长出来。　在这个空窗里，旧问题看上去被解决了，新问题还没有显露。人们站在这个空窗里往外看，看到的是"问题没了"。所以AGI感判断错的，不是工具有多强，而是现在是什么时候。

◆ 三、三个维度同时失衡

为什么这个空窗会产生一种"什么都能做"的感觉，而不是一种"出了什么事"的感觉？因为在这个时期，SDE三个维度恰好处在一种失衡的状态里。

第一，旧的S被封装了。原来需要人花很多年才能掌握的结构，现在一键可得。这一维一下子满了，满到溢出来。

第二，D在进行中，但还没有结果。使用刚刚开始，碰撞正在发生，差异正在积累，但还没有凝固成可见的结构。过程是看不见的，只有结果才看得见。

第三，新的E尚未成熟。新的需要、新的处境、新的问题类别，要在使用中慢慢长出来，现在还是一片没有长草的新土。

三者合起来，就是我们常说的"维度未配平"：一边是满得发胀的旧S，另一边是还没有成形的D和E。人的感知，天然偏向看得见的东西。旧S满了，看得见；D和E还空着，看不见。于是整个画面是：能力已经到了，需求似乎消失了，问题似乎解决了。

这就是第十编的定音句要说的：AGI不是一个终点，而是一段等待；在旧问题已经消失、新问题还未显露的时候，人们把看不见误当成了没有。

◆ 四、家务的例子：洗衣机来了之后

用一个更老、更扎实的例子，把这个机制看一遍。

洗衣机进入普通家庭的时候，人们的第一反应是：家务要被消灭了。一个按钮，衣服就洗好了，过去要在河边、盆边弯腰半天的活，几乎不用人了。这个反应在当时是完全合理的，因为洗衣这件事，被封装得非常彻底。

美国学者露丝·考恩（Ruth Cowan）研究家务技术史时发现的情况却相当出人意料：这些省力的家用设备普及之后，主妇花在家务上的总时间并没有如预期那样明显减少。原因是新的东西长了出来：衣服可以勤换了，标准提高了，洗得更多、更频繁；原来由洗衣店或家里其他人分担的活，转移到了一台机器和一个人身上；对"干净"的期待本身被抬高了。（考恩的书是1983年的《More Work for Mother》，这里只取大意。）

注意这里的时间顺序。机器到家，是一天的事；新的标准、新的习惯、新的分工、新的烦恼，却是在之后的许多年里一点一点显出来的。在最初那几年，"家务已经被消灭"的想象，是一种完全可以理解的感觉：旧的洗衣问题确实消失了，新的问题还没来得及被说出来。

这就是一次小规模的时间差：封装瞬间，发生要等。

◆ 五、为什么封装能那么快，发生却那么慢

有人会问：既然新问题迟早会长出来，为什么不能很快？这里有两个结构性的原因，不是偶然的拖延。

第一，封装是复制已有的东西，发生是产生没有的东西。 复制的速度取决于技术和成本，可以非常快；产生的速度取决于使用和碰撞，而使用是一个社会过程，要一个人一个人、一个处境一个处境地铺开。你可以一夜之间把一百万台洗衣

机卖出去，却无法一夜之间让一百万个家庭同时发现"原来这件事还可以这样烦恼"。

第二，新E必须由主体长出来，不能被预先给定。 新的需要不是从机器里掉出来的，而是在人的生活里一点点长出来的。洗衣机普及后，人们对"干净"的期待升高，这个期待是人自己生出来的。没有人能替他们一次生完。这正是第42章立的前提在起作用：E的源头在主体，主体的处境要变，需要时间。

◆ 六、铁路与时钟：另一个版本的空窗

第40章讲过铁路与标准时区，这里只取它的时间差。铁路是十九世纪一次干净利落的封装：把远距离运输的经验，压进一条钢轨和一台机车里。可铁路造出来那天，没有任何人在谈标准时区；时区是铁路被使用之后，在碰撞里长出来的，北美到1883年才推行，中间隔了几十年。这几十年里，铁路的"问题"似乎只有一个：造得够不够快、够不够多。新问题还在沉睡。

◆ 七、回扣承重命题

这一章把第十条承重命题落到了机制上：AGI感是维度未配平时期的时间差现象。它依赖前三条命题：工具是已发生智慧的封装（命题一），使用即碰撞、碰撞必生新SDE（命题二），新SDE必在封装之外（命题三）。三条合起来推出一个时间差；有时间差，就有一段空窗；站在空窗里的人，很容易生出"全都有了"的感觉。前两步是结构，最后一步是人的心理倾向，本书不把它说成必然。

◆ 八、这个论证的边界在哪里

先交代五条限度。

第一，这个机制在隔壁已有近亲。 经济史早就观察到通用技术的回报姗姗来迟：戴维讲电动机，布雷斯纳汉与特拉伊滕贝格讲通用目的技术要靠配套创新才见效，佩蕾丝把技术革命分成导入期和展开期，布林约尔松等人把生产率画成先降后升的J曲线；未来学者阿马拉（Roy Amara）名下也有一句流传很广的话，大意是

人们总高估技术的短期影响、低估长期影响〔待核〕。本书与它们相同：能力先到，配套后来。不同在于，本书把这个滞后拆成S满、D未成形、E未熟三维，并把AGI感看作滞后在人的认知上的投影，由此推出第45、46章那两个可检验的推论：强弱与碰撞深浅成反比，群体按阶段分布。

第二，它在什么情况下是错的。　空窗是可以观测的：一边是能力指标，比如某类任务的完成率；一边是新问题类别的出现，比如新的职业名称、新的行业标准、新的投诉类别。本章说，前者跃升之后，后者有一段滞后。如果在多个领域里两者几乎同步出现，滞后为零，或者滞后期里公共讨论中并没有"问题没了"这类论调，这一章的机制就不成立。

第三，机制不是证明。　本章解释"AGI感"为什么会产生，不能反过来证明AGI不可能。证明在前九编，尤其是第九编的结构论证。这一编回答的是另一个问题：既然不可能，为什么看起来像可能？

第四，空窗的长短不能预言。　本书不预测年份，也不声称空窗有固定长度。洗衣机之后的家务变化持续了几十年；有的软件功能带来的新问题，几个月就显了。机制是结构性的，时长是经验性的。

第五，并非所有关于AGI的感觉都是这一种。　有的人把"通用"定义得很宽松，那是定义分歧；有的是商业叙事。这些第46章专门交代，本章的机制只覆盖因时间差而产生的那一部分。

封装是一瞬间的事，发生却要等。在这两者之间，旧S已满，D未成形，新E未熟，维度没有配平。人们站在这段空窗里，把"还没看见"当成了"不存在"。这就是AGI感的底层机制。下一章，我们去历史里找它以前出现过的证据：电力、1965年的预言、信息高速公路，每一次都是同一出戏的不同演员。

带走的话

装修验收是一天的事，生活要一天一天地过；中间那段空窗里，房子样样齐全，缺的东西还没显出来。AGI感，就是站在空窗里，把“还没看见”当成了“没有”。

小账 封装了什么？——旧S：洗衣、认路、发电、传输这些已经发生的经验，一键可得，所以显得“全都有了”。 等待期里看不见什么？——还没成形的D和还没长熟的E：新的需要、新的标准、新的问题类别，要靠许多人在许多处境里使用出来。

承接：时间差是结构，历史上有没有活的样本？下一章看三次。

第 44 章

历史上的AGI感：电力、1965年的预言、信息高速公路

◆ 一、一个厨房里的小实验

先做一个小小的想象实验。如果你在1990年的某个下午，被一位热情的朋友拉到一台个人电脑前，他对你说："有了这个，办公室里的纸、文件柜、会计的算盘，全都会消失。"你大概会点头。因为你亲眼看到了：表格自己会算，文字可以反复改，打印机吐出整齐的页面。

然后你回到自己的办公室。再过五年、十年，你会发现，办公室里的纸反而更多了：因为打印变得太容易，每个人都可以把每一稿都印出来；因为表格好做，领导要求的报表越来越多、越来越细；因为发邮件不要钱，一件事被同时抄送给二十个人。办公室研究者后来专门写过一本书，题目就叫《无纸办公室的神话》（Sellen与Harper，2002）。

当年那句"纸会消失"，不是撒谎，也不是愚蠢，而是站在空窗里的一个自然的判断。旧的问题（计算慢、修改难）确实被封装解决了，新的问题（信息过载、报表膨胀、抄送文化）还没显出来。

这一章，我们看三个更大的样本。它们彼此隔着几十年，行业完全不同，但剧情惊人地相似。

◆ 二、样本一：电力与它的乌托邦

电力进入人类生活，是一次足以重写文明的封装。把煤、水、蒸汽转换成一种可以远距离传输、可以随处接入的能量，这是人类能源经验的一次高度压缩。

在电力刚刚普及的年代，社会上弥漫着一种相当浓厚的乌托邦想象：电会让劳动变得轻松，让城市变得明亮干净，让家务消失，让人有大把的闲暇。电灯、电

车、电动机、各种电器，似乎预示着一个"辛苦的时代终结了"的世界。这类想象在当时的报刊、博览会和科幻作品里相当常见。

后来发生了什么，我们都知道：电力确实彻底改变了世界，但是以几乎无人预见的方式。它没有让人闲下来，而是催生了装配线、家电工业、广播电视、电影、实验室、后来的计算机；它改变了城市的形状、夜晚的功能、工作的节奏。真正的改变，不在"原有活计变轻松"，而在"一整批以前不存在的活计和产业被生了出来"。

一个特别值得一看的细节，是电动机进入工厂之后的故事。 经济史学者保罗·戴维（Paul David）在1990年发表了一篇很有影响的文章（题目大意是《发电机与计算机》），讨论的就是这件事：电动机在十九世纪末就已经可以用了，但工厂的生产率并没有在它出现后立刻飙升，而是等了几十年才明显地提高。

原因是什么？按照这个研究的叙述，最初工厂只是把原来的蒸汽动力换成了电动机，厂房的布局、传动的方式、工作的流程都还是蒸汽时代那一套：一台大动力源通过天轴和皮带带动一大片机器。要等到工厂的设计者意识到，电动机可以小型化、分散化，每台机器可以有自己的动力，于是厂房可以单层展开，流程可以按生产顺序重新排列，这时候电力的真正潜力才释放出来。

这是一个典型的空窗。电动机被"封装"的那一刻，工厂里的能力已经到了；但新的SDE，也就是新的厂房、新的流程、新的组织方式、新的分工，要等管理者、工程师、工人在使用中一点点碰撞出来。这是一个几十年的等待。

在这个等待里，人们的感觉是什么？一方面，是乌托邦式的兴奋，一方面，是"数字上看不出什么变化"的困惑。两种感觉其实是同一件事的两面：能力已经来了，新的结构还没出来。

◆ 三、样本二：1965年的预言

第二个样本来自人工智能自己的历史，所以更值得认真对待。

1965年，赫伯特·西蒙（Herbert Simon）这位后来获得诺贝尔经济学奖的学者，同时也是人工智能最早的奠基者之一，在《The Shape of Automation for Men and Management》一书里对机器的前景做过一个判断，大意是：二十年之内，机器将有能力做人所能做的任何工作。

我们不必嘲笑他。在那个年代，计算机刚刚能下跳棋、证明几何定理、解代数应用题。每一项都是"过去只有聪明人才会做"的事，如今一台机器做到了。站在那个节点看出去，路似乎是一路平坦的：棋会下了，定理会证了，接下来就是更难一些的定理、更难一些的棋，一步一步走到"所有工作"。

二十年后的1985年，机器当然没有做人所能做的一切工作。但这并不是说那些早期的成果是虚的，恰恰相反，它们都是真的。真正发生的事情是：这些成果被一层层封装进了软件、数据库、专家系统、各种行业工具里，而一批当年没有人预见的新问题冒了出来。

比如，1968年，"软件危机"这个词被提了出来：程序越来越大，写得越来越慢，错误越来越多，成本越来越难以控制。机器能做的事越来越多，反而逼出了"怎样组织很多人去写出可靠的程序"这样一整门此前不存在的学问：软件工程。这不是一个旧问题被解决了，而是一个新问题类别被生了出来。

我们把它放进框架里看：当年的预言之所以显得可信，是因为"会下棋、会证明定理"这些旧能力的S已经被封装得很亮眼；它看不到的，是封装之后必然伴随的D和E：软件规模的爆炸、协作的难题、对可靠性的新要求、人们对"机器应该做什么"的重新思考。预言的错，不在于对旧能力的估计，而在于把看不见的新问题当成了没有。

◆ 四、样本三：九十年代的信息高速公路

第三个样本离我们更近。九十年代初，"信息高速公路"这个说法在许多国家的政策、媒体、商业讨论里流行起来。它的想象是：一张覆盖全社会的高速数据网络，一旦铺好，教育、医疗、购物、政务、娱乐，都将被彻底改造；知识的鸿沟会被抹平，距离会消失，中间商会消失。

这些想象里，有相当多是对的，也有相当多没有发生，更多的是发生了，但换了一种完全没想到的形状。网络确实改变了购物，但催生的是平台、搜索广告、物流体系和直播带货这样一批当初的蓝图里没有的东西；网络确实让信息可得，但同时生出了信息过载、算法推荐、虚假信息治理这些此前没有的问题类别；网络确实让距离消失，但同时让注意力成为稀缺品。

而在1987年，经济学家罗伯特·索洛（Robert Solow）在一篇书评里说过一句后来被反复引用的话：大意是，电脑时代到处都看得见，唯独在生产率统计里看不见。

这句话今天常被叫作"生产率悖论"。放在我们的框架里，它恰恰是空窗期的写照：电脑的能力已经到了每一张办公桌，旧S封装完成；但组织如何重新设计、流程如何重新安排、新的商业模式如何出现，这些新SDE还没有显露，所以统计里自然看不到。统计看的是已经稳定下来的结构，而新结构还在长。

后来，统计里确实"看见"了。等到企业围绕信息技术重新组织了业务，等到新的产业形成，生产率的改善才在数据中浮现。具体何时何地、是不是全然归功于信息技术，学界有争论。但这个"滞后"的形状，与电动机的故事是同构的。

◆ 五、三个样本的共同剧本

把三个故事叠起来看，会发现它们在做同一件事。

第一幕：封装完成。　一种旧的智慧被压缩进了一件工具里，效果立竿见影。电力把能源经验封装了，早期人工智能把某些"只有聪明人才会做"的任务封装了，信息网络把传递与获取信息的经验封装了。

第二幕：空窗与想象。　在这个时刻，旧问题看起来被解决了，新问题还没显露。人们开始勾画"终点"：闲暇的天堂，无所不能的机器，没有距离的世界。这些想象不是愚蠢，是对着空窗做的合理外推。

第三幕：新SDE显露。　工具被使用，碰撞发生，新的产业、新的问题、新的组织形式、新的职业，一批一批长了出来。"终点"没有到来，来的是一个更复杂的世界，以及一份比之前更长的待办清单。

第四幕：想象被改写。　回头看，人们不再说"电力会让人闲下来"，而说"电力是一场工业革命的基础"。想象的终点，被换成了一个新的起点。

你会注意到，三个故事里，每一个"终点"想象，都是由已经完成的封装做出来的外推；而每一次落空，都是由尚未显露的新SDE造成的。

◆ 六、回扣承重命题

这三个样本，是第四道缝里所说的"历史归纳"的材料。我们要在这里把话说清楚：它们是佐证，不是证明。

第四道缝说得明白：归纳不证"永远"。过去三次都出现了这样的剧本，并不能逻辑上保证下一次也如此。真正撑起结论的，是前面的结构论证：封装与新SDE在时间上必然错位，因为封装发生的时候，新SDE还不存在。历史样本的作用，是让这个抽象的结构变得看得见、摸得着，也让读者能对照自己的经验检验它。

◆ 七、这个论证的边界在哪里

这一章有三个限度。

第一，类比不是同一。　电力是通用的动力，信息网络是通用的连接，早期人工智能是对特定任务的自动化，性质并不相同。把它们放进同一个剧本，是抓住它们共有的一个结构，不是说它们是同一种东西。当前的大模型也有自己的特点，比如在语言上的通用性相当宽，这一点第八编里专门讨论过。

第二，我们在挑样本，这一点可以补救。　历史上也有预言完全应验、空窗短得几乎察觉不到的时候。本章挑的三个，是因为剧本特别清楚。补救的办法是列一张更长的表：把二十次以上的通用技术浪潮都排出来，数一数有多少具有"终点想

象，落空，更长的清单"这个顺序。如果多数没有，本章的剧本就不成立。附录C的第三条可错预言与此互相支撑。

第三，事实细节以原始文献为准。　　本章的年份、论文和引述，多是转述大意，不是原话；读者引用时，请回到原始文献。

电力、1965年的预言、信息高速公路，三次不同的封装，三次相似的空窗，三次落在同一个地方的想象，也是三次同样的收场：终点没有到，来的是一份更长的清单。历史无法证明下一次也如此，但它能让人看清这个结构的样子，也给了我们理由去问：今天的AGI感，是不是又一次站在空窗里往外看？

> 带走的话
>
> 电力、1965年的预言、信息高速公路，三次都是先画一个“终点”，再等来一份更长的清单。历史证明不了“永远”，但它让人看清这个剧本的样子，也该让你对今天的“都有了”多问一句。

> **小账**　　封装了什么？——能源经验（电力）、“只有聪明人才会做”的任务（早期人工智能）、传递与获取信息的经验（网络）。 等待期里看不见什么？——新的厂房与流程、软件危机与软件工程、信息过载与注意力稀缺；统计看的是已经稳定的结构，新结构还在长。

承接：同一个空窗里，为什么有的人很强烈地感到AGI，有的人几乎没感觉？下一章看个体。

第 45 章

个体差异：AGI感的强弱与碰撞的深浅成反比

◆ 一、同一个下午，三个人

某个周三的下午，三个人坐在三张书桌前，面对同一个AI工具。

第一个是高三学生，作文要交了，题目是"一件小事"。他把题目输进去，几秒钟后，屏幕上出现了一篇结构完整、文字流畅的作文，有开头有转折有升华。他读了一遍，改了两个词，交了。他的感受是：这东西太厉害了，写得比我好，以后还要我们学写作干什么？

第二个是一位刚被通知岗位要"优化"的文员。她的工作，是把各种来料整理成固定格式的汇报材料。她看着这个工具，在几分钟内完成了自己通常要花半天的活，心里一沉：这不是辅助，这是替代。她晚上睡不着，在想自己四十五岁了还能做什么。

第三个是一位准备写书的人，三个月前开始用同样的工具帮忙。他用得很多，从整理提纲到查资料到推敲句子。写到第三章，他撞上了一个问题：他想讲的那个核心判断，工具给出的每一版论述都似是而非，他自己也说不清哪里不对。他停下来，一个人在书房里坐了两个晚上，终于想明白自己真正要说的是什么，而这个想法，和他开始写书时的想法已经很不一样了。他的感受是：这东西很有用，但离"通用智能"还差得远，因为真正难的地方它帮不上，那是我自己的事。

三个人，同一个工具，同一个下午，三种感觉。第一个人和第二个人感到的是强烈的"AGI感"，一个是崇拜，一个是恐惧；第三个人几乎没有。

差别在哪里？不是智商，不是技术水平，也不是性格。差别在于：碰撞的深浅。

◆ 二、原理一句话

AGI感的强弱，与一个人跟工具碰撞的深浅成反比：碰得越深，越看见封装之外的东西，AGI感就越弱。

这句话的反面也成立：碰得越浅，越只看见封装本身的光滑表面，AGI感越强。

◆ 三、为什么深碰撞会削弱AGI感

回到第二编讲过的原理：使用即碰撞，碰撞必生新SDE。但碰撞有深有浅。

浅碰撞，是拿到结果就走。学生输入题目，拿到作文，交上去。这个过程里，工具与他自己的处境之间没有发生真正的相互作用：他没有被这篇作文改变，也没有用这篇作文去撞自己真实的需要。他看到的是工具封装的那一层：流畅、完整、迅速。这一层是真的光彩夺目。这和计算机科学家韦曾鲍姆1966年的发现是同一类事：他写的ELIZA只是个改写句子的简单对话程序，却让不少人聊几句就觉得被理解了。碰撞越浅，越容易把表面当成全部。

深碰撞，是把工具放进自己真正在做的事情里，让它的能力和自己的问题互相撞击。写书的人把工具用到了自己的核心判断上，撞出了封装之外的东西：那个只有他自己能说清、而工具说不清的地方。这个"说不清"，就是新SDE的萌芽：一个此前不存在、只能由他这个主体在他的处境里发生出来的问题。

深碰撞为什么必然削弱AGI感？因为深碰撞的本质，就是触到封装的边界。边界一旦被触到，"通用"的幻觉就破了：你清楚地看见，工具擅长的是已经发生过、已经被封装的东西，而你面对的问题恰恰是那个没有被封装的、正在发生的东西。

打个更日常的比方。有人第一次用导航，觉得神奇，再也不用问路了。他开车一个月，没有问题。直到有一天，导航让他拐进了一条正在施工的小路，他才发现：这个蓝线是根据已有地图画的，不是根据眼前这条路画的。他开始留意路牌、留意车流、留意导航不知道的东西。这一次小小的撞击，让他对导航的感觉，

从"万能"变成了"很好用的工具"。他不是不信导航了，而是知道它的边界在哪里了。

◆ 四、三类人

把这个原理展开，大致可以把对AI的反应分成三类。这三类不是固定的标签，同一个人在不同的事情上可以属于不同的类别。

第一类：深度碰撞者

他们把工具用到了自己真正在做的、有分量的事情里，并且持续地用、认真地撞。写书的人、做研究的人、做产品的人、认真教学的老师，都有可能属于这一类。

他们的共同特点是：几乎没有AGI感，但也不是否定工具。 他们往往是工具最热情的使用者，也是最清楚工具边界的人。他们说的话常常是这样的："它帮我把第一遍走得很快，但第二遍和第三遍的难处，一点没少，反而因为前面走得快，我更早撞上了真正的难题。"

深度碰撞者的体验里有个反直觉的东西：工具让他更早地抵达真正的难处。封装把容易的部分拿走了，剩下的全是难的、新的、需要他自己发生出来的。

第二类：被封装者（恐惧型）

他们的工作恰好在被封装的那一层里。那份工作的内容，是一套已经发生、可以被复制的结构：整理、转写、套用、核对、按格式输出。工具一来，他们的活儿几分钟就完成了。

他们的AGI感强烈而恐惧。这种恐惧不是错觉：被封装的能力在退场，对个人而言是真实的职业冲击和生活冲击。第六编的"让渡不是退化"是从文明层面说的；个体层面的损失，不能用一句"新产业会诞生"轻轻带过（第46章还会谈）。

那位文员的恐惧是有道理的。她对"这是替代"的判断，在她那一份具体工作上，也许完全正确。她看错的，只是"替代的是全部的我"。她看不见的，是封装之外那些还没显露的东西，包括她自己可能发生出来的新的可能。这一点，我们留到第47章。

第三类：旁观者（崇拜型）

他们没有深用，也不被直接冲击，站在远处看，被一段段惊人的演示、一篇篇令人屏息的报道所震撼。他们的AGI感强烈而崇拜。

旁观者的特点是：看到的几乎全是封装的表面，因为表面是最容易被传播的。演示视频里的工具总是在它最擅长的任务上，用最合适的输入，产生最漂亮的输出。这是正常的：一个能被拍成演示的东西，一定是已经封装得很好的东西。看不到的，是日常使用中那些反复磕碰的细节，那些"它这里不行"的时刻。

崇拜型的AGI感，在很大程度上是媒体距离制造的：离封装越远、离使用越远，越只看见光。

◆ 五、作文与书：两种用法的对照

回到开篇的对照：学生用AI写作文，作者用AI写书写到第三章。

学生的使用，任务是一次性的，结果是现成的，成败只看交没交；"一件小事"对他不是一个真实的问题，只是个要完成的格式。这里没有E，也就是没有"这对谁来说是个问题"，所以工具再好，也没有和他碰撞，因为他没有拿自己去撞。

作者的使用，任务漫长，结果不确定，成败取决于想说的东西有没有说清。书里的每个判断都是他的、他要负责的、他在乎的，这里有很深的E。第三章遇到的问题没有标准答案，是一个正在他心里发生的问题。

同一个工具，在前一种用法里显得"无所不能"，在后一种里显出了边界。工具没变，变的是碰撞的深度，也就是使用者带进去的E有多深。

这给了一个很实用的自测：下一次你觉得"这东西太厉害了，要取代一切"的时候，问自己：我刚才是在用它做一件对我真的重要的事，还是只是在看它表演？

◆ 六、为什么会成反比，而不只是相关

读者可能会问：这是不是只是一个经验观察？深度碰撞者更了解工具，所以不夸大，这很自然啊。

确实有这个层面。但本书想说的更进一步：这是一个结构性的反比，原因在于AGI感本身的定义。

AGI感，是"封装已完成、新SDE尚未显露"的空窗期里的认知现象。它的内容，是"看不见新问题"。而碰撞的深度，正是新SDE显露的机制：碰得越深，新SDE越早在你面前显露出来；一旦显露，空窗对你而言就关上了。

换句话说，对一个深度碰撞者，空窗期比对别人更短。　他的等待，被他自己的碰撞缩短了。所以，同一个社会时刻，别人还在空窗里，他已经走出来了。这是个体差异的结构来源，也是第47章里"缩短等待"的出发点。

◆ 七、回扣承重命题

这一章是第十条承重命题的第二半句：AGI感的强弱与碰撞的深浅成反比。它同时回扣第二条：使用即碰撞，碰撞必生新SDE。既然碰撞必生新SDE，碰得深的人就更早、更清楚地看到新SDE；既然新SDE必在封装之外，这些人就更清楚地看到封装的边界。

◆ 八、这个论证的边界在哪里

限度有四条。

第一，反比是倾向，不是定律。　一个人可能碰撞很深，却依然得出"AGI将至"的判断。他们的理由可能是定义上的（对"通用"采用了更宽松的标准），也可能

是对未来技术走向的判断。这些不属于时间差现象，我们在下一章会谈。本章只说：在同样的定义下，碰撞的深度会影响"看见什么"。

第二，三类人的划分是粗线条的。　现实中的人是混合的：同一个人可能在工作上是被封装者，在爱好上是深度碰撞者，在新闻里是旁观者。分类的用处，是帮你认出自己此刻的位置，不是给人贴标签。

第三，"恐惧型"的恐惧不能被这套说法化解。　说"深碰撞会削弱AGI感"，并不等于说"你只要深碰撞，你的工作就保住了"。被封装者的处境，不是靠认知上的澄清就能改变的，它牵涉到收入、转岗、社会保障。这套说法解释的是感觉从何而来，不是对风险的担保。

第四，我们没有统计数据，但这个反比可以检验。　办法是：先让受访者各自写下自己对"通用"的定义，把定义分歧拿掉；再按使用深度（是否长期在自己负责、有风险的事上使用）分组，比较对"AI已经通用"的认同度。如果深度使用者的认同度没有低于浅度使用者，甚至更高，这个反比就错了。本章的三类人，目前只是基于结构推理和日常观察的描述，不是调查结果。

同一个工具，三个人，三种感受。差别不在聪明与否，而在碰撞的深浅。碰得浅，只看见封装的光滑表面，AGI感强；碰得深，撞上封装之外的新问题，AGI感弱。被封装的人，恐惧是真实的；旁观的人，崇拜是媒体距离造成的。深度碰撞者的经验，恰好是时间差机制在个人身上的缩影：他们的空窗期更短。

> **带走的话**
>
> 同一个工具，碰得浅的人看见光滑的表面，碰得深的人撞上边界。下次你觉得“太厉害了”，先问自己：刚才是在做一件对我真正重要的事，还是只在看它表演?

> **小账** 封装了什么？——整理、转写、套用、核对、按格式输出这类已经发生、可以复制的结构；被封装者的工作，恰好就在这一层里。 等待期里看不见什么？——浅碰撞者看不见“它这里不行”的时刻，也就是只在自己处境里发生的那个“说不清”的问题。

承接：个体如此，群体呢？同一时刻，不同的群体，其实站在不同的阶段。下一章看群体。

第 46 章

群体差异：同一时刻，不同群体在不同阶段

◆ 一、一条街上的四家店

设想一条老街，同一天，四家店发生着不同的事。

第一家是新开的饺子馆，老板刚买了一台自动和面机，一天省下三个小时，高兴得合不拢嘴。他觉得：这机器太好了，以后饺子生意是稳赚的。

第二家是开了十年的面馆。几年前，这条街上多了几家用同样机器的店，大家的面都做得差不多了，价格被压低，利润薄了。老板愁眉苦脸地说：现在谁都能做面，还有什么特色？

第三家是隔壁新开的"手作面馆"，老板专门推出"我的面条"定制服务：顾客可以选面的粗细、汤的浓淡、配料的搭配，甚至写下自己的名字刻在面碗上。他的生意不错，因为他卖的不再是面，而是"这是我的面条"的感觉。

第四家是街角做了三十年手工拉面的老师傅，机器来的时候，他没有跟上，也没有找到新的出路。他的店前几个月关了门，他在家里待着，对那台机器没有好话。

四家店，同一条街，同一台机器，同一天。但四个人的感受，和他们所处的阶段，完全不同。

这个小图景，是这一章要讲的缩影：同一时刻，不同群体站在时间差的不同位置上。 所谓"AGI感"的起落，并不是整个社会同步上升、同步下降，而是不同群体沿着同一条曲线，在不同的时间点上走。

◆ 二、原理一句话

时间差不是一道全社会统一的闸门，而是一条曲线；在同一个时刻，不同的群体站在曲线的不同位置，因此AGI感的强弱、方向都不同。

三、对接工具让渡律的四阶段

本书第九章讲过工具让渡律的四个阶段：扩大期、内卷期、新产业诞生、让渡。这里把它和AGI感的起落对上，做一张粗略的图。

扩大期。　新工具出现，能力迅速扩展，原来做不到的事现在做得到了，原来很贵的事现在便宜了。大家都在享受红利，新问题尚未显露。这是AGI感最强的时期：封装的光彩最盛，新SDE最远。上面街上的第一家饺子馆，就在这个阶段。

内卷期。　工具普及到几乎人人都有，红利被摊平，竞争加剧，大家用同样的工具做同样的事，边际收益递减。这时候，人们开始感到"不对劲"：工具很强，但并没有带来预期的好处。第二家面馆在这个阶段。AGI感开始下降，因为人们碰到了封装的边界：只靠工具已经不够区分了。

新产业诞生。　在内卷的压力下，人们被迫去找工具之外的东西，新的需要、新的玩法、新的价值开始显露。第三家"我的面条"在这个阶段。AGI感明显下降，因为大家的注意力转向了新生的部分，那是封装之外的东西。

让渡。　被封装的旧能力正式退场，新产业接过了社会的主要活动。第四家老师傅，是在这个阶段里被让渡的人。他的处境值得我们在后面单独谈。

用第二条可错预言里的话说：社会上的AGI感，在扩大期最强，进入内卷期后明显下降。　这条预言是可以被检验的：如果有一次大规模的工具普及之后，进入内卷期，社会上的AGI感不降反升，那么本书的这一层论证就出了问题。

四、为什么群体会分阶段

为什么同一个工具，会让不同群体处在不同的阶段？至少有三个原因。

第一，接触工具的时间不同。　有的行业先用，有的行业后用。先用的行业先进入内卷期，后用的行业还停在扩大期。比如，同一类工具，在写代码的群体里普及得早，在传统制造业里普及得晚，所以前者可能已经在讨论"大家都用以后，我们靠什么区分"，后者还在讨论"它到底能不能用"。

第二，行业的开放程度不同。　越是封闭、规则固定的行业，封装越彻底，被封装的部分越多，让渡来得越快；越是开放、需要反复判断的行业，封装能覆盖的部分越有限，新SDE显露得越快。这一点，第八编里讲象棋与"通用即开放"时已经谈过。

第三，社会缓冲的厚薄不同。　不同群体有不同的转岗能力、教育机会和社会保障。同样是被封装，有的人可以顺畅转向，有的人则被卡在原地。同一场变化，感受自然不同。

创新扩散研究早就画过采用者的曲线：先行者、早期多数、晚期多数、迟到者（罗杰斯，1962）。本书与它相同：群体不同步。不同在于，那条曲线问的是谁在何时采用，本书问的是采用之后，各群体正站在工具让渡律的哪一段，AGI感随之怎样变。

◆ 五、一个对照：同一个社会里的三种话

在同一个时期，你可以在不同的圈子里听到完全不同的话。

在刚刚开始大规模使用新工具的圈子里，你听到的是："它什么都能做。"

在工具已经用了一两年、开始感到同质化的圈子里，你听到的是："大家做出来的东西越来越像了。"

在已经长出新岗位、新服务、新问题的圈子里，你听到的是："我们现在最缺的是会提出好问题的人。"

三句话，同一个时代，三个阶段。假如你只听其中一个圈子，你会以为整个社会都是这个想法。 这是群体层面的"媒体距离"。

◆ 六、一定要诚实：并非所有的AGI讨论都是时间差

到这里要停一下，交代一件不能回避的事。

今天关于AGI的讨论，并不全是时间差现象。　把一切都归因为空窗期里的错觉，那本身就是一种过度解释，和我们批评的"AGI已来"一样不严谨。

至少有三类讨论，和时间差无关，或者只有部分关系：

第一类，定义分歧。　有人说的AGI，是"在大多数经济上有价值的任务上达到人类水平的系统"；有人说的AGI，是"能解决一切问题的系统"；有人说的AGI，是"有意识的机器"。这是三个不同的东西。本书第一章区分了终局AGI、人类水平AGI和AGI感，就是因为不同的讨论说的是不同的东西。如果两个人讨论AGI是否已经到来，一个人的定义是前者，另一个人的定义是后者，他们的分歧根本不是时间差，而是定义。主流的AGI定义并不要求解决所有问题，本书的结构论证击中的是终局AGI；对阶段性的"人类水平通用AI"，本书依靠的是第八编的论证，那是另一条线。

第二类，商业叙事。　一个行业为了融资、获客、招人，需要一个宏大的故事。"我们正在接近AGI"是一个很有力量的故事。这类叙事有时候建立在真实的技术进展之上，有时候是夸大，有时候是纯粹的营销。它们的存在，不需要时间差来解释，只需要市场的激励就够了。本书不对具体公司和产品做评价，只指出：当你听到一个关于AGI的强烈说法时，值得问一句，说这话的人，有没有什么理由希望你这么想。

第三类，真实的技术判断。　有一些人，是基于对技术路线的深入了解，真心认为某些能力会在不太长的时间内达到很高的水平。这类判断，有的可能对，有的可能错，但它们不是错觉。本书不能也不打算反驳每一个具体的技术预测，本书反驳的是"终局"这个词。

因此，本书的立场是：时间差解释了AGI感中的一部分，而且是一个结构性的、始终存在的部分；但它不解释全部。　把全部讨论都装进这个框架，是不诚实的。

◆ 七、让渡期里被封装的人

再回到那位街角的老师傅。

在工具让渡律的框架里，被封装能力的退场是"让渡"，不是"退化"：文明整体没有变笨，那份能力被工具接过去了，人把注意力腾给了新的事。这个说法，从文明层面看是对的。

但从老师傅的角度看，它几乎没有安慰价值。他三十年练成的手艺，在他还没老的时候就不再有市场了。他的损失是具体的：收入、尊严、习惯了半辈子的节奏、在街坊中的位置。历史上每一次大的工具变化，都有一批这样的人。他们不是因为懒惰或愚笨被淘汰，只是正好站在了被封装的位置上。文明层面的上升，不会自动补偿这些个体的损失。

所以要明确说一句："被封装能力的退场是让渡，不是退化"，这个命题在文明层面成立；在个体层面，让渡期的损失是真实的，值得认真对待，包括社会的缓冲、转岗的支持、教育的更新。　把文明层面的好话直接说给个体听，是不负责任的。

这也解释了为什么恐惧型的AGI感不应当被当作"认知偏差"简单纠正。它既是时间差现象（看不见封装之外的新东西），又是对真实风险的合理反应。两者并存。

这一章是第十条承重命题的群体版本：AGI感是维度未配平时期的时间差现象。在群体层面，它的起落与工具让渡律的阶段对应，扩大期最强，内卷期后下降。同时，第五条承重命题"被封装能力的退场是让渡，不是退化"，在这里得到了个体层面的补注。

◆ 八、这个论证的边界在哪里

三条限度如下。

第一，阶段的划分是理想化的。　真实的行业、真实的社会，不会整齐地按四个阶段走。几个阶段可能同时存在，也可能反复。图是帮助理解的，不是日程表。

第二，我们不预言具体年份。 本书不说"某某行业将在几年后进入内卷期"。这是经验问题，由时间来回答。

第三，可错点，以及怎样测。 第二条可错预言——社会上的AGI感在扩大期最强，进入内卷期后明显下降——是可以检验的。一个可行的做法：取两个行业，一个刚普及，一个已普及两三年，各自收集公开讨论（行业报道、从业者访谈、招聘文案），统计"已经能代替""什么都能做"这类措辞，与"同质化""缺好问题"这类措辞的比例。扩大期的那个行业，前一类应当更高；如果相反，这一层论证就出了问题。

同一时刻，不同群体处在时间差曲线上的不同位置。扩大期的人觉得"什么都能做"，内卷期的人觉得"大家都一样"，新产业里的人觉得"缺的是好问题"，被让渡的人觉得"我的手艺没了"。AGI感的起落，不是整个社会的统一节拍。同时，今天关于AGI的讨论，有一部分是定义分歧，一部分是商业叙事，一部分是真实的技术判断，不全是时间差。让渡期里被封装的人所承受的损失，是真实的，不能被一句"文明在上升"轻轻带过。

> **带走的话**
>
> 同一时刻，有人在扩大期说“什么都能做”，有人在内卷期叹“大家都一样”，有人在新产业里发愁缺“好问题”。别拿一个圈子的话当整个社会的声音，也别拿一句“文明在上升”去打发被让渡的人。

> **小账** 封装了什么？——街角手艺人三十年的手艺被机器接了过去，变成谁都买得到的“合格的面”。 等待期里看不见什么？——扩大期的人看不见内卷与新产业，旁观者看不见让渡期里具体的人在付的代价；还有一部分AGI讨论，本来就是定义分歧、商业叙事或真实的技术判断。

承接：既然空窗期因人而异，一个人能不能主动缩短自己的等待？下一章回答。

第 47 章

缩短等待：主动碰撞，以及为什么发生学哲学是每个人的主科

◆ 一、第一次下厨

一个从没做过饭的年轻人，搬出来独居了。他有两种方式学做饭。

第一种，是把菜谱背熟。他看了二十个视频，存了一百个收藏，能把"番茄炒蛋"的步骤倒背如流：先炒蛋，盛出，再炒番茄，加盐，合并。他觉得自己已经会了。

第二种，是直接开火。第一次，蛋炒老了；第二次，番茄出水太多，成了一锅汤；第三次，盐放多了；第四次，终于像样了，他发现自己放了一小勺糖，味道一下子圆了。这个"一小勺糖"，菜谱里没有写，是他在锅边自己撞出来的。

两种方式，一个是在等：等一个"我准备好了"的时刻再动手。另一个是在撞：不管准备好没有，先下锅。到了第二个月，第一个人还在收藏菜谱，第二个人已经有了自己的几道拿手菜，并且知道什么情况下该怎么调整。

这个对照，就是这一章要说的：等待不是消极地等一个时机，而是一种维度上的配平；而主动碰撞，可以缩短一个人自己的等待期。

◆ 二、原理一句话

等待期是维度未配平的时期；配平靠的不是干等，而是主动去补那几个还没长出来的维度，也就是主动碰撞。

◆ 三、等待到底是什么

前面几章反复说"空窗期""等待期"，读者很容易把它理解成一种被动的状态：就像等公交车，时间到了自然来。这个理解需要在这里纠正。

回到三个维度。AGI感出现的时候，是旧S被封装得过满、D在进行中、新E尚未成熟。这是一种不平衡。要走出这个状态，需要三个维度重新配平。

配平靠什么？

S这一边，已经满了，不需要再加。旧的结构已经被封装，可以随取随用。

D这一边，需要有人去走。差异和过程不会自己发生，需要有人把工具放进真实的事情里去用，去碰、去撞、去试错。D就是碰撞本身。

E这一边，需要有人去长。新的需要、新的问题，是从主体的处境里生出来的。一个人没有投入自己的处境，E就长不出来。

所以，等待不是等外面的变化，而是等自己这个主体里的D和E长出来。这个"等"，不是坐着，而是做。

回到街上那四家店：第三家"我的面条"为什么在新产业诞生阶段活了下来？不是因为他"等到了好时机"，而是因为他在机器普及之后，主动去问了一个问题：既然所有人都能做出一样的面，顾客想要的到底是什么？这个问题，就是他自己去发生的E。他把自己投进去，才长出了"我的面条"。

◆ 四、本书不是在劝人"等一等就好"

有一种对这套说法的误读，需要提前堵住。有人会说：既然AGI感只是时间差，等一等，新产业自然会出来，我们不用着急，也不用做什么。

这是对"等待"的最大误解。

首先，文明层面的新产业，确实会生出来，这一点是结构性的；但个人能不能赶上，不是结构性的。第六编区分过个体层与文明层：文明层必然，个体层靠碰撞。一个文明走出空窗期，不代表每个个体都走出了；有些人会被留在让渡的那一边。

其次，时间差的长短，对个人来说，取决于他碰得多主动。前一章已经看到，深度碰撞者的空窗期更短。这是可以由个人去影响的变量。

所以，本书的态度是：等待不是消极的，也不是被动地接受命运；它是一个可以被缩短、被利用的时期。这里的"缩短"，不是催熟，而是通过主动碰撞，让自己的D和E更早地长出来。

◆ 五、怎样主动碰撞：三个日常的做法

说到这里，读者自然想问：具体怎么做？这里不开药方，只给三个朴素的方向，都可以在日常里试。

第一，把工具放进自己真正在乎的事里。　不要只拿它做演示性的小任务，要拿它做你真正想做、真正在乎、有风险的事。写一封对你很重要的信，规划一件你真的想做的事，整理一段你真的想弄明白的思考。因为只有在你在乎的事情上，E才是真的，碰撞才会深。

第二，找到它做不到的那个地方，然后停下来看。　每一次用工具，留意一个"这里不对""这里不够""这里我说不清"的时刻。别急着绕过去。这个时刻，往往就是新SDE的入口。第45章里写书的那位作者，在书房里坐的那两个晚上，正是这样的时刻。

第三，把撞出来的东西写下来、说出来、交给别人。　碰撞的结果，如果只留在脑子里，很容易散掉。写下来，是把D凝固成S；说给别人听，是让它进入更大的E。很多新的东西，是在与别人的对话里才长成形的。

这三个做法，看上去都很平常。但它们有一个共同点：都是把自己作为主体投进去。　这也是AGI感最脆弱的地方。AGI感来自旁观，来自"别人怎么了"的距离感；而一旦你自己进去了，就变成了"我在做什么"。

◆ 六、为什么是"发生学哲学"，而且是每个人的"主科"

现在，我们可以回答这一编标题里最后一个问题了：为什么发生学哲学应该是每个人的主科？

先说什么是"主科"。语文、数学、外语，是我们习惯上的主科，不是因为它们最有趣，而是因为它们是其他一切学习所依赖的基础工具。不会读写，就学不了别的；不会算数，就做不了很多事。

在一个工具越来越强的时代，什么变成了新的基础工具？

本书的回答是：一种看待"事情是怎样发生的"的眼光。 也就是发生学哲学。

你可以把它想成一副眼镜。戴上它，你看一个工具，不只看它能做什么（这是封装的S），还会问：它封装的是什么旧的智慧？用起来会撞出什么封装里没有的东西？我自己的处境里，哪一个是它碰不到的？看一个行业，不只看它现在的样子，还看它在哪个阶段、下一步可能显露什么。看自己，不只看已有的能力，还看正在发生的东西。

这副眼镜的作用，恰恰是治疗AGI感。因为AGI感，就是只看见封装的S，看不见D和E。有了发生学的眼光，空窗不再是"没有"，而是"还没显露"；你不会把看不见误当成没有。

为什么要成为"每个人"的主科，而不是少数哲学爱好者的兴趣？有三个理由。

第一，因为碰撞的主体是每个人。 工具不是只有专家在用，而是每个人都在用。每个人都站在时间差里，每个人都需要一副看懂它的眼镜。

第二，因为被封装的位置不挑人。 今天被封装的可能是文员，明天可能是任何一种以结构复制为主的工作。没有人能确定自己永远站在封装之外。能保护一个人的，不是某一项具体的技能，而是持续发生新东西的能力。

第三，因为智慧不是库存。 这是全书第三层论证。如果智慧是库存，学习就是搬运，工具来了搬运就被替代；如果智慧是共生的发生，学习就是让自己不断发生。后一种学习，是工具替代不了的：工具替代到哪里，哪里就成了下一次发生的起点。

◆ 七、一个日常的类比：学开车

想想学开车。没有人把"理解汽车为什么动"当成学开车的前提，但必须学会一种与车互动的方式：看路、感受方向盘的反馈、预判前车。这不是背出来的，是开出来的。开得久了，你知道什么时候该信任车，什么时候该接管。

发生学哲学类似面对工具时的"驾驶感"：不是对工具的理论知识，而是一种与工具互动的方式，知道什么时候让它走，什么时候自己接管，什么时候停下来看路。这种驾驶感不能只靠看书得到，需要真的上路。说它是主科，也包含这层意思：它需要练习，需要在真实的事情里反复碰撞。

◆ 八、给正站在让渡边上的人

前面说了许多对"主动碰撞"的肯定，这里要单独对一类人说几句：那些此刻正站在让渡边上的人，比如本章开头提到的那位文员。

对她，"去撞出新东西"听上去可能像一句风凉话：我的岗位明天就没了，你让我去撞？这个反应是正当的，我们不回避。

比较踏实的做法，是把大问题拆小。她今天不必"发生一个全新的自己"，只需要做一件很小的事：下一周，选一件手头真实的工作，用工具把它的第一遍做出来，然后只盯着一个问题看：这一遍里，哪一处必须由我来判断？是对客户语气的拿捏，是对数据异常的敏感，还是对这个部门历史的了解？这些"必须由我来"的地方，就是她的封装之外。它们未必立刻能变成一个新岗位，却是她接下来最有用的材料，因为工具能封装的是整理和转写，封装不了"这件事对这个部门到底意味着什么"。

当然，这仍然只是个人层面的一步。她的处境同时需要岗位的过渡安排、培训和收入的缓冲，这些不是她一个人能提供的。主动碰撞缩短的是认知上的空窗，缩短不了制度上的空缺；两者要分开看，也要一起做。

这一章是第十条承重命题的行动版：AGI感是维度未配平时期的时间差现象，而时间差对个体而言是可以缩短的。它同时回扣第六条：创造力与智慧共生，智慧不是先在的库存；以及第五编、第七编里反复说的：每一次创造，都让一个人多出一个从前没有的自己。主动碰撞，就是把自己放进那个"多出一个自己"的过程里。

◆ 九、这个论证的边界在哪里

这一章有五条限度。

第一，主动碰撞不保证成功。 缩短等待，说的是"更早地让新SDE显露出来"，不是"保证你一定能找到出路"。有人撞了很多次，依然没有撞出什么；有人因为行业整体被让渡，个人再努力也很难改变结局。这个方法是增加可能性，不是担保结果。

第二，它怎么会被证明无效。 找条件相近的两组人，用同一个工具半年；一组学过"封装—碰撞—边界"的看法，并按上面三个做法练习，另一组没有。半年后，如果前一组在"说得出工具做不到之处、并据此改变自己做事方式"上并不更好，本章的建议就无效。

第三，"每个人的主科"是本书的主张，不是已经被检验的教育方案。 发生学哲学怎样进入课堂、以什么形式、教到什么程度，需要实验。近年常说的"AI素养"教育，杜威以来的"做中学"，都与它相近：相同处，是都主张在真实使用中学；不同处，是这里要教的核心不是怎么用，而是用的时候带着"封装与发生"的眼光，让自己进入发生。本书提出的是方向，不是课程表。

第四，个体努力不能取代社会安排。 转岗培训、社会保障、教育更新，都是让渡期里必要的缓冲。把一切归结为"你要主动碰撞"，对被封装的人是苛刻的。

第五，这是本书的价值立场。 说"持续发生新东西是值得追求的"，包含一种价值判断：把人当作发生的主体，而不是被动适应的对象。这是前提，不是推论。不同意的读者，可以不接受这一章的建议，但依然可以接受前九编的论证。

等待不是干等，而是维度的配平；个人的等待期，可以靠主动碰撞缩短。把工具放进自己真正在乎的事，在它做不到的地方停下来看，把撞出来的东西写下来、说出来。发生学哲学之所以应当成为每个人的主科，是因为它是面对工具时的一副眼镜、一种驾驶感，治的是AGI感的根：只看见封装，看不见发生。

> **带走的话**
>
> 等待不是等公交，是等自己这边的D和E长出来；背熟菜谱的人在等，下锅的人在撞。把工具放进你真正在乎的事，在它做不到的地方停下来看，再把撞出来的东西写下来、说给人听。

> **小账**　封装了什么？——菜谱、整理与转写：已经发生的做法，随取随用，S这一边不需要再加。 等待期里看不见什么？——自己还没长出来的D和E：“一小勺糖”这样菜谱里没有的东西，只能在锅边撞出来。

承接：全书的论证到这里走完了五层。最后，我们回头看一眼走过的路，再说一句收束的话。

结 语

没有终点的智慧

◆ 一、回到那个问题

开篇的时候，我们问了一个看似简单的问题：AGI会不会来？

走过十编，我们的回答是：作为终局，不会；作为阶段，类似的能力会出现，但它一出现，就不再"通用"。我们没有说工具不重要，也没有说AI没有价值。恰恰相反，本书从头到尾都在说：工具是人类智慧的封装，是文明最重要的东西之一。我们只是不同意一个词：终局。

这个回答压成一句话，留到最后一页。

◆ 二、五层论证回顾

把全书的论证，用最朴素的话再走一遍。

第一层：没有终局智能。　工具是已发生智慧的封装。封装完成的那一刻，工具就被放进了生活，使用即碰撞，碰撞必生新的SDE。而新SDE必在封装之外，因为封装的时候它还不存在。这是一个结构上的错位，不是技术上的缺陷。所以，无论封装多么强大，都不可能把"之后要发生的东西"一并封装进去。第二编讲原理，第三编讲现实界工具史，第九编把它写成三段论，并给出推翻它的办法，讲的都是这一层。

第二层：没有固定的人类水平。 人们常说"AI达到了人类水平"。可是人类水平不是一根固定的横杆，而是一个不断移动的过程：新的产业不断出现，新的能力不断被需要，而且横杆里本身就含着AI这个追赶者。AI可以追上任何一个截面，却追不上这个过程。象棋这样的封闭领域，AI确实永久地越过了人，这一点我们承认；但"通用"的意思，就是开放，而开放的领域，不可被永久越过。第八编讲的是这一层。

第三层：没有可装满的智慧。 如果智慧是库存，那么把库存搬完，就是AGI。可是创造力与智慧共生：创造不是智慧的支出，而是智慧的发生。一个人每创造一次，就多出一个从前没有的自己。智慧不是先在的、可以清点装箱的东西。这是共生图景对库存图景的回答。第七编讲的是这一层。

第四层：三界的工具只会提升三界的智慧。 现实界的火与犁、时钟与电力，理念界的文字、数字、微积分，自我界的镜子、日记、心理学，每一次工具的出现，都没有终结智慧，而是在三界之间生出新的问题和新的能力：工具从不只停在出生的那一界。就连理论本身，也是工具。第三编到第六编，用工具史，把这一层摆在了读者面前。我们也直面了反驳：导航让海马体退化了吗？退场是让渡，不是退化。毁灭性的工具呢？提升智慧不等于没有代价，核武器逼出了威慑理论与核伦理。

第五层：AGI感是等待期的时间差现象。 封装是瞬间的，发生要等待；人们站在空窗里，旧问题没了，新问题未显，便把看不见当成了没有。这是第十编讲的。AGI感的强弱与碰撞的深浅成反比；不同的群体在曲线的不同位置；而一个人可以通过主动碰撞，缩短自己的等待。

这五层，前四层说的是"为什么不可能"，第五层说的是"为什么看起来像可能"。

撑起这五层的，是十条承重命题。它们各自落在哪一层：

- 第一层：工具是已发生智慧的封装（一）。使用即碰撞，碰撞必生新SDE（二）。错位定理：新SDE必在封装之外，因为封装时它还不存在（三）。基础性本身在发生；AI制造了自己解决不了的基础问题（九）。第九编的结构证明，把这几条接成一条线。
- 第二层：人类水平是过程，不是高度；横杆里含着追赶者（七）。封闭领域可被永久越过（象棋），开放领域不可（"通用"即开放）（八）。
- 第三层：创造力与智慧共生，智慧不是先在的库存（六）。

- 第四层：一界有工具，三界有新生（工具从不只停在出生的那一界）（四）。被封装能力的退场是让渡，不是退化（五）。
- 第五层：AGI感是维度未配平时期的时间差现象；强弱与碰撞深浅成反比（十）。

◆ 三、本书的几处缝，再说一次

一本书如果只讲自己有理的地方，那就是宣传。所以，在收尾处，我们把四道缝再摆一次。

定义的缝。 主流的AGI定义，并不要求"解决所有问题"。本书击中的是终局意义上的AGI；对阶段性的"人类水平通用AI"，本书依靠的是第八编的论证。如果读者采用的定义和本书不同，这本书的某些结论可能不适用，这是诚实的限度。

谁来升级的缝。 本书的论证证明了永远需要更高一级的智慧，但不证明它必须是人。我们把"问题总是对某个主体而言的问题"明确立为前提：E的源头在主体。这是前提，不是推论。终局AGI自我否定的核心层，只用到这个前提最弱的一层；"升级往哪里升、谁来定"，才用到它的完整形式。如果有一天，出现了一种本身就是主体的、有自己处境和需要的存在，这条前提就需要被重新审视，而终局依然不会因此成立。

象棋的缝。 封闭领域里，AI确实永久越过了人。本书承认，并以此划出了边界：封闭可越，开放不可。

归纳的缝。 历史是佐证，结构论证才是证明。归纳不证"永远"。本书用历史让结构看得见，但不把历史当作证明的主角。

◆ 四、可以被证伪

最后，一本想讲清楚道理的书，应该告诉读者怎样才能证明它错。

在附录C里，本书列出了三条可错的预言：任何开放领域在AI能力大幅跃升后的数年内，会出现由这次跃升本身引发的新职业或新问题类别；社会上的AGI感在

扩大期最强，进入内卷期后明显下降；无论哪一次宣称"AGI已经实现"，此后被公认的前沿基础问题清单只会变长，不会变短。

这些预言，不依赖于某个具体的年份，也不依赖于对某个具体产品的评价。如果它们被事实反复推翻，这本书就该被修正。除此之外，还有几处可以动手的检验：第39章的预注册清单，第40章对基础清单的逐条追溯，第45章先写下定义再比较认同度的调查，第47章两组人半年的对照。哪一项做出来与本书相反，本书就该改。

◆ 五、三个小提醒

在合上书之前，给读者三个能带走的小提醒，它们都来自第十编，也都可以在明天早上的第一次使用里试一试。

其一，听到"都有了"的时候，问一句"谁的"。 工具说"都有了"，说的是它封装的那一份已有。你的处境里有没有它不知道的东西，只有你能回答。

其二，觉得"没问题了"的时候，等一等。 旧问题消失了，不等于没有问题。新问题的显露需要使用，需要时间，需要你自己的处境。多用一阵再下结论，是对空窗期最朴素的尊重。

其三，撞上边界的时候，不要绕开。 那个"说不清"的地方，是新的SDE在敲门。每一次真正的升级，都始于一次不愿绕开的卡住。

◆ 六、回到日常

写到这里，我想回到一个很普通的画面。

一位母亲，第一次抱着刚出生的孩子，回到家里。家里什么都买齐了：婴儿床、奶瓶、尿不湿、一整架育儿书。她觉得万事俱备。然后，半夜两点，孩子哭了，她发现：书上没有这个哭声，书上也没有这个夜晚。她必须自己去发生一个新手妈妈。

那一晚之后，她多出了一个从前没有的自己。

工具永远都是备好的东西，生活永远都是还没发生的东西。这两者之间的距离，不是缺陷，是智慧生长的地方。

AGI感，是人们站在这段距离的入口，看着满屋崭新的东西，说"都有了"。而智慧，是在半夜两点，第一次抱起那个哭着的孩子。

◆ 七、定音

人类水平不是一个高度，而是一个发生过程；AI可以抵达它的任何一个截面，却抵达不了过程本身，因为每一次抵达，都成了下一次发生的材料。

工具封装的是昨天的智慧，逼出来的是明天的智慧。

> AGI不是智慧的终点，而是又一次封装；每一次封装，都是下一次发生的起点。

智慧没有终点。有的，是一次又一次的开始。

参 考 书 目

本目录只收正文真正引到或化用到的著作与论文。外文书先列原题，后附中文题；中文题为本书通行译法，不指定译本。版本、译者、页码没有核实的，一律标〔待核〕。

一、技术与创新研究

1. Arthur, W. Brian. The Nature of Technology: What It Is and How It Evolves. New York: Free Press, 2009. 中文题：《技术的本质》。
2. Hughes, Thomas P. Networks of Power: Electrification in Western Society, 1880–1930. Baltimore: Johns Hopkins University Press, 1983. 中文题：《电力网络》（中译题〔待核〕）。
3. Rosenberg, Nathan. "The Direction of Technological Change: Inducement Mechanisms and Focusing Devices." Economic Development and Cultural Change, 1969. 卷期页码〔待核〕。
4. Collingridge, David. The Social Control of Technology. London: Frances Pinter, 1980. 中文题：《技术的社会控制》。
5. Merton, Robert K. "The Unanticipated Consequences of Purposive Social Action." American Sociological Review 1(6), 1936. 中文题：《有目的的社会行动的意外后果》。
6. Jevons, William Stanley. The Coal Question. London: Macmillan, 1865. 中文题：《煤的问题》。
7. Schumpeter, Joseph A. Capitalism, Socialism and Democracy. 1942. 中文题：《资本主义、社会主义与民主》。
8. Bresnahan, Timothy F., and Manuel Trajtenberg. "General Purpose Technologies: 'Engines of Growth'?" Journal of Econometrics 65(1), 1995. 中文题：《通用目的技术：增长的引擎？》（页码〔待核〕）。
9. Perez, Carlota. Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages. Cheltenham: Edward Elgar, 2002. 中文题：《技术革命与金融资本》。

10. Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. "The Productivity J-Curve: How Intangibles Complement General Purpose Technologies." American Economic Journal: Macroeconomics 13(1), 2021. 中文题：《生产率的J曲线》。
11. Acemoglu, Daron, and Pascual Restrepo. "Automation and New Tasks: How Technology Displaces and Reinstates Labor." Journal of Economic Perspectives 33(2), 2019. 中文题：《自动化与新任务》。
12. Autor, David, Caroline Chin, Anna Salomons, and Bryan Seegmiller. "New Frontiers: The Origins and Content of New Work, 1940–2018." Quarterly Journal of Economics 139(3), 2024. 中文题：《新前沿：1940—2018年新工作的来源与内容》。
13. Sellen, Abigail J., and Richard H. R. Harper. The Myth of the Paperless Office. Cambridge, MA: MIT Press, 2002. 中文题：《无纸办公室的神话》。
14. Brooks, Frederick P. The Mythical Man-Month: Essays on Software Engineering. Reading, MA: Addison-Wesley, 1975. 中文题：《人月神话》。

二、经济史与技术史

1. Cowan, Ruth Schwartz. More Work for Mother: The Ironies of Household Technology from the Open Hearth to the Microwave. New York: Basic Books, 1983. 中文题：《母亲的活更多了》。
2. Vanek, Joann. "Time Spent in Housework." Scientific American 231(5), 1974. 中文题：《花在家务上的时间》。
3. David, Paul A. "The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox." American Economic Review 80(2), 1990. 中文题：《发电机与计算机》。
4. Allen, Robert C. "Engels' Pause: Technical Change, Capital Accumulation, and Inequality in the British Industrial Revolution." Explorations in Economic History 46(4), 2009. 中文题：《恩格斯停顿》。
5. Scott, James C. Against the Grain: A Deep History of the Earliest States. New Haven: Yale University Press, 2017. 中文题：《作茧自缚》。
6. Geertz, Clifford. Agricultural Involution: The Processes of Ecological Change in Indonesia. Berkeley: University of California Press, 1963. 中文题：《农业内卷化》。
7. Wrangham, Richard. Catching Fire: How Cooking Made Us Human. New York: Basic Books, 2009. 中文题：《生火：烹饪如何使我们成为人类》。
8. Berna, Francesco, et al. "Microstratigraphic Evidence of In Situ Fire in the Acheulean Strata of Wonderwerk Cave, Northern Cape Province, South Africa."

Proceedings of the National Academy of Sciences 109(20), 2012. 中文题：《旺德沃克洞穴阿舍利地层中原地用火的显微地层学证据》（页码〔待核〕）。

9. Wiessner, Polly W. "Embers of Society: Firelight Talk among the Ju/'hoansi Bushmen." Proceedings of the National Academy of Sciences 111(39), 2014. 中文题：《社会的余烬：朱/霍安西人的火光夜谈》。
10. Mumford, Lewis. Technics and Civilization. New York: Harcourt, Brace, 1934. 中文题：《技术与文明》。
11. Thompson, E. P. "Time, Work-Discipline, and Industrial Capitalism." Past & Present 38, 1967. 中文题：《时间、工作纪律与工业资本主义》。
12. Galison, Peter. Einstein's Clocks, Poincaré's Maps: Empires of Time. New York: W. W. Norton, 2003. 中文题：《爱因斯坦的钟，庞加莱的地图》。
13. Carnot, Sadi. Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance. Paris, 1824. 中文题：《关于火的动力的思考》。
14. Schmandt-Besserat, Denise. Before Writing. Austin: University of Texas Press, 1992. 中文题：《文字之前》。
15. Fibonacci (Leonardo of Pisa). Liber Abaci. 1202. 中文题：《算盘书》。
16. Pacioli, Luca. Summa de arithmetica, geometria, proportioni et proportionalità. Venice, 1494. 中文题：《算术、几何、比例与比例性概览》。
17. Mélchior-Bonnet, Sabine. Histoire du miroir. Paris: Hachette (Imago), 1994. 中文题：《镜子的历史》（出版社〔待核〕）。
18. Rosling, Hans. 关于洗衣机的公开演讲。篇名、年份与场合〔待核〕。

三、科学哲学与数学基础

1. Popper, Karl R. Logik der Forschung. Vienna: Julius Springer, 1934. 中文题：《研究的逻辑》（英译本 The Logic of Scientific Discovery，《科学发现的逻辑》）。
2. Popper, Karl R. The Poverty of Historicism. London: Routledge & Kegan Paul, 1957. 中文题：《历史决定论的贫困》。
3. Kuhn, Thomas S. The Structure of Scientific Revolutions. Chicago: University of Chicago Press, 1962. 中文题：《科学革命的结构》。
4. Lakatos, Imre. Proofs and Refutations: The Logic of Mathematical Discovery. Cambridge: Cambridge University Press, 1976. 中文题：《证明与反驳》。
5. Deutsch, David. The Beginning of Infinity: Explanations That Transform the World. London: Allen Lane, 2011. 中文题：《无穷的开始》。

6. Kauffman, Stuart, Giuseppe Longo, and Maël Montévil. "No Entailing Laws, but Enablement in the Evolution of the Biosphere." 2012. 作者次序（Longo, Montévil, Kauffman）与载体（GECCO 论文集）〔待核〕。中文题：《没有蕴涵律，只有使能：生物圈演化》。
7. Hacking, Ian. The Taming of Chance. Cambridge: Cambridge University Press, 1990. 中文题：《驯服偶然》。
8. Hacking, Ian. "The Looping Effects of Human Kinds." In Causal Cognition, ed. D. Sperber et al. Oxford: Oxford University Press, 1995. 中文题：《人类种类的循环效应》。
9. Euclid. Elements（《几何原本》）。通行本与译本〔待核〕。
10. Hilbert, David. Grundlagen der Geometrie. Leipzig, 1899. 中文题：《几何基础》。
11. Berkeley, George. The Analyst. London, 1734. 中文题：《分析学家》。
12. Cauchy, Augustin-Louis. Cours d'analyse de l'École Royale Polytechnique. Paris, 1821. 中文题：《分析教程》。
13. Robinson, Abraham. Non-standard Analysis. Amsterdam: North-Holland, 1966. 中文题：《非标准分析》。
14. Frege, Gottlob. Begriffsschrift. Halle, 1879. 中文题：《概念文字》。
15. Frege, Gottlob. Grundgesetze der Arithmetik, Bd. I–II. Jena, 1893/1903. 中文题：《算术基本法则》。
16. Bernoulli, Jacob. Ars Conjectandi. Basel, 1713. 中文题：《猜度术》。
17. Bayes, Thomas. "An Essay towards Solving a Problem in the Doctrine of Chances." Philosophical Transactions of the Royal Society 53, 1763. 中文题：《机遇学说中一个问题的解》。
18. Knight, Frank H. Risk, Uncertainty and Profit. Boston: Houghton Mifflin, 1921. 中文题：《风险、不确定性与利润》。
19. Whitehead, Alfred North. An Introduction to Mathematics. London: Williams & Norgate, 1911. 中文题：《数学导论》（所引原话措辞〔待核〕）。
20. Lucas, Robert E. "Econometric Policy Evaluation: A Critique." Carnegie-Rochester Conference Series on Public Policy 1, 1976. 中文题：《计量经济政策评价：一个批判》。
21. Hales, Thomas C., et al. "A Formal Proof of the Kepler Conjecture." Forum of Mathematics, Pi 5, e2, 2017. 中文题：《开普勒猜想的形式化证明》。
22. Herodotus. Histories（《历史》），卷二。通行本与译本〔待核〕。

四、认知、心理与自我

1. Maguire, Eleanor A., et al. "Navigation-Related Structural Change in the Hippocampi of Taxi Drivers." Proceedings of the National Academy of Sciences 97(8), 2000. 中文题：《出租车司机海马体与导航相关的结构变化》。
2. Dahmani, Louisa, and Véronique D. Bohbot. "Habitual Use of GPS Negatively Impacts Spatial Memory during Self-Guided Navigation." Scientific Reports 10, 2020. 中文题：《习惯性使用GPS对自主导航中空间记忆的负面影响》。
3. Bainbridge, Lisanne. "Ironies of Automation." Automatica 19(6), 1983. 中文题：《自动化的反讽》。
4. Dewey, John. Experience and Education. New York: Macmillan, 1938. 中文题：《经验与教育》。
5. Ryle, Gilbert. The Concept of Mind. London: Hutchinson, 1949. 中文题：《心灵的概念》。
6. Polanyi, Michael. The Tacit Dimension. 1966. 中文题：《默会维度》（中译题〔待核〕）。
7. Boden, Margaret A. The Creative Mind: Myths and Mechanisms. London: Weidenfeld & Nicolson, 1990. 中文题：《创造性思维》。
8. Ong, Walter J. Orality and Literacy: The Technologizing of the Word. London: Methuen, 1982. 中文题：《口语与书写》。
9. Clark, Andy, and David Chalmers. "The Extended Mind." Analysis 58(1), 1998. 中文题：《延展心智》。
10. Bartlett, Frederic C. Remembering: A Study in Experimental and Social Psychology. Cambridge: Cambridge University Press, 1932. 中文题：《记忆》。
11. James, William. The Principles of Psychology. 2 vols. New York: Henry Holt, 1890. 中文题：《心理学原理》。
12. Freud, Sigmund. Die Traumdeutung. Leipzig/Vienna: Franz Deuticke, 1899（扉页标1900）。中文题：《梦的解析》。
13. Open Science Collaboration. "Estimating the Reproducibility of Psychological Science." Science 349(6251), 2015. 中文题：《心理科学可重复性的估计》。
14. Pronin, Emily, Daniel Y. Lin, and Lee Ross. "The Bias Blind Spot: Perceptions of Bias in Self Versus Others." Personality and Social Psychology Bulletin 28(3), 2002. 中文题：《偏见盲点》。

15. Luft, Joseph, and Harry Ingham. "The Johari Window, a Graphic Model of Interpersonal Awareness." 1955. 载体与页码〔待核〕。中文题：《乔哈里之窗》。
16. Festinger, Leon. "A Theory of Social Comparison Processes." Human Relations 7(2), 1954. 中文题：《社会比较过程理论》。
17. Goffman, Erving. The Presentation of Self in Everyday Life. 1956（爱丁堡初版）；New York: Doubleday, 1959. 中文题：《日常生活中的自我呈现》。
18. Rose, Nikolas. Inventing Our Selves: Psychology, Power, and Personhood. Cambridge: Cambridge University Press, 1996. 中文题：《发明我们自己》。
19. Foucault, Michel. 关于"自我技术"的论述。具体篇名与年份〔待核〕。
20. Hadot, Pierre. Exercices spirituels et philosophie antique. 1981. 中文题：《古代哲学作为精神修炼》（中译题〔待核〕）。
21. Giddens, Anthony. The Constitution of Society. Cambridge: Polity, 1984. 中文题：《社会的构成》。
22. Maturana, Humberto R., and Francisco J. Varela. Autopoiesis and Cognition: The Realization of the Living. Dordrecht: Reidel, 1980. 中文题：《自创生与认知》。
23. Augustine. Confessiones（《忏悔录》），四世纪末。通行本与译本〔待核〕。
24. Montaigne, Michel de. Essais（《随笔集》），1580年初版。通行本与译本〔待核〕。
25. Loyola, Ignatius of. Exercitia Spiritualia（《神操》），十六世纪。通行本与译本〔待核〕。

五、AI 与计算理论

1. Simon, Herbert A. The Shape of Automation for Men and Management. New York: Harper & Row, 1965. 中文题：《人与管理的自动化之形》（中译题〔待核〕）。
2. Chollet, François. "On the Measure of Intelligence." arXiv:1911.01547, 2019. 中文题：《论智能的度量》。
3. Shin, Minkyu, Jin Kim, and Minkyung Kim. "Superhuman Artificial Intelligence Can Improve Human Decision-Making by Increasing Novelty." Proceedings of the National Academy of Sciences 120(12), 2023. 中文题：《超人类人工智能通过增加新颖性改进人的决策》（作者次序与卷期〔待核〕）。
4. Gödel, Kurt. "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I." Monatshefte für Mathematik und Physik 38, 1931. 中文题：《论〈数学原理〉及有关系统中的形式不可判定命题 I》。

5. Turing, Alan M. "On Computable Numbers, with an Application to the Entscheidungsproblem." Proceedings of the London Mathematical Society, ser. 2, 42, 1936. 中文题：《论可计算数及其在判定问题上的应用》。
6. Church, Alonzo. "An Unsolvable Problem of Elementary Number Theory." American Journal of Mathematics 58, 1936. 中文题：《初等数论的一个不可解问题》。
7. Cook, Stephen A. "The Complexity of Theorem-Proving Procedures." Proceedings of the 3rd ACM Symposium on Theory of Computing, 1971. 中文题：《定理证明过程的复杂性》。
8. Wolpert, David H., and William G. Macready. "No Free Lunch Theorems for Optimization." IEEE Transactions on Evolutionary Computation 1(1), 1997. 中文题：《优化的"没有免费午餐"定理》。
9. Hofstadter, Douglas R. 书中转述"特斯勒定理"的一部著作。书名与年份（正文作1979年）〔待核〕。
10. Solow, Robert M. "We'd Better Watch Out." New York Times Book Review, 12 July 1987. 中文题：《我们最好当心》。

六、本系列各卷

1. 王德生：《我的三个宝贝》，德麦国际第220号，新加坡：德麦国际（Demai International Pte. Ltd.），2026年。
2. 王德生：《我经济》，德麦国际第224号，新加坡：德麦国际，2026年。
3. 王德生：《让每个人都有自己的思想和知识工厂》，德麦国际第129号，新加坡：德麦国际，2026年。
4. 王德生：《AI时代教育的认知变革——腾出的那一处，怎样才被占住》，德麦国际第122号，新加坡：德麦国际，2026年。
5. 王德生：《AI时代的哲学困境——发现学哲学的慢慢离去》，德麦国际第127号，新加坡：德麦国际，2026年。

引文核验总表

核验状态说明：已核＝书名、作者、年份与正文说法相符；待核＝正文说法或细节尚未查实，成书前须逐条确认。

条目	本书所在章	核验状态
1883年11月18日北美铁路统一标准时区	第3、6、12、27章	已核
深蓝1996年2月负于卡斯帕罗夫（2比4）；1997年5月以3.5比2.5获胜	导读，第2、4、37章	已核
2016年3月AlphaGo对李世石五番棋赢四局	第37章	已核
2023年《美国国家科学院院刊》：AlphaGo之后职业棋手决策质量上升	第37章	已核（作者次序与卷期待核）
考恩1983年《母亲的活更多了》：家务时间并未明显减少	第6、14、28、43章	已核
瓦内克1974年：二十世纪二十至六十年代家务时间大致没有减少	第14章	已核
戴维1990年《发电机与计算机》：电动机之后几十年生产率才明显上升	第7、14、44章	已核
福特1913年高地公园工厂建立移动装配线	第14章	已核
索洛1987年：电脑时代到处可见，唯独统计里看不见	第44章	已核
西蒙1965年：二十年内机器能做人所能做的任何工作	第2、44章	已核
马圭尔等2000年：伦敦出租车司机海马体后部更大、前部更小	第29章	已核
2020年研究：习惯用导航者自主寻路时空间记忆更差（相关而非因果）	第29章	已核
班布里奇1983年"自动化的反讽"	第29章	已核
奥托等2024年：约六成现职业名称在1940年不存在	第35章	已核
肖莱2019年：智能是面对新任务时获取技能的效率	第38章	已核
考夫曼等2012年：生物圈的新功能无法事先列举	第31、39章	已核
哥德尔1931年不完备定理	第2、19、40章	已核

条目	本书所在章	核验状态
图灵、邱奇1936年对可计算的刻画	第40章	已核
库克1971年；克雷数学研究所2000年千禧年难题	第40章	已核
罗素悖论（1901年发现，1902年函告弗雷格）	第2、19章	已核
弗雷格1879年《概念文字》	第19章	已核
贝克莱1734年《分析学家》；柯西1821年《分析教程》	第3、19章	已核
希尔伯特1899年《几何基础》	第18章	已核
拉卡托斯1976年《证明与反驳》	第18章	已核
黑尔斯1998年宣布开普勒猜想计算机辅助证明；2014年宣布形式化完成	第21章	已核
斐波那契1202年《算盘书》	第17章	已核
帕乔利1494年《概览》系统讲述威尼斯复式记账	第17、27章	已核
婆罗摩笈多628年前后给出零的运算规则	第17章	待核（年份取通行说法）
伯努利1713年《猜度术》；贝叶斯1763年	第20章	已核
奈特1921年区分风险与不确定性	第20章	已核
卡诺1824年《关于火的动力的思考》	第13章	已核
克劳修斯1865年前后提出"熵"	第13章	已核
杰文斯1865年《煤的问题》	第6、13章	已核
艾伦：英国工业革命早期（约1790至1840年）的"恩格斯停顿"	第13章	待核（起讫年份）
亨德森名下那句话，最早出处为1960年一本书	第13章	待核
1968年10月北约加尔米施会议提出"软件工程"	第15章	已核

条目	本书所在章	核验状态
布鲁克斯1975年《人月神话》	第15章	已核
1988年11月莫里斯蠕虫；CERT/CC设于卡内基梅隆大学	第15章	已核
加里森2003年：爱因斯坦的专利局与庞加莱的经度局	第3、12、27、40章	已核
汤普森1967年论时间纪律；芒福德论钟	第12章	已核
莫尔斯1844年华盛顿至巴尔的摩电报	第12章	已核
萨弗里1698年专利；纽科门1712年蒸汽机	第13章	已核
贝尔纳等2012年旺德沃克洞穴约一百万年前用火证据	第10章	已核（页码待核）
兰厄姆"熟食假说"与艾洛、惠勒"昂贵组织假说"	第10章	待核（已按出处分开表述）
魏斯纳2014年：朱/霍安西人夜间火边讲故事约占八成	第10章	已核
斯科特2017年《作茧自缚》	第11章	已核
格尔茨"农业内卷"（爪哇水稻）	第11章	已核
施曼特-贝瑟拉特的泥质筹码说	第16章	已核（学界有争议，正文已注明）
凡·艾克1434年《阿尔诺芬尼夫妇像》中的凸面镜	第22章	已核
梅尔舍-博内1994年《镜子的历史》	第22章	已核
巴特利特1932年《记忆》	第23章	已核
詹姆斯1890年《心理学原理》；冯特1879年莱比锡实验室	第25章	已核
弗洛伊德《梦的解析》（1899年底出版，扉页标1900年）	第25章、附录A	已核
开放科学合作组织2015年重复研究	第25章	已核

条目	本书所在章	核验状态
普罗宁等2002年"偏见盲点"；卢夫特与英厄姆1955年；费斯廷格1954年	第25、26章	已核
戈夫曼《日常生活中的自我呈现》（1956年初版，1959年通行版）	第26章	已核
罗伯特·科尼利厄斯1839年银版自拍	第26章	已核
休斯1983年；罗森堡"聚焦装置"	第三编导言、第28章	已核（罗森堡篇名与页码待核）
科林格里奇1980年"科林格里奇困境"	第30章	已核
杜威1938年《经验与教育》	第15、31、34章	已核
波普尔1934年《研究的逻辑》；1957年《历史决定论的贫困》	第3、4、7章	已核
卢卡斯1976年批判	第7章	已核
默顿1936年意外后果	第6章、第三编导言	已核
布雷斯纳汉与特拉伊滕伯格；佩蕾丝；布林约尔松等"J曲线"	第三编导言、第14、43章	已核
阿西莫格鲁与雷斯特雷波：替代与创造新任务	第9、35章	已核（正文标注的年份与表述待核）
哈金"循环效应"；《驯服偶然》1990年	第20、25、36章	已核
特斯勒定理与霍夫施塔特"1979年的书"	第36章	待核（疑为《哥德尔、艾舍尔、巴赫》）
Sellen与Harper 2002年《无纸办公室的神话》	第44章	已核
皮亚杰"认知结构被改造"	第31、34章	待核（正文未点具体著作，参考书目暂未列条）

附 录

附录A　三界工具史年表

说明：只收有把握的年份或时段，不确定的句末标〔待核〕。年表按三界分列，同一界内大致按时间排列，条目与第10至第26章对应。"意义"一栏只写一句：封装了什么，逼出了什么。

◆ 现实界

时间	工具或事件	意义（封装了什么，逼出了什么）
远古，年代有争议	人类对火的控制与使用〔待核〕（第10章）	封装了取暖、熟食；逼出了围火而居、分食、讲述
一万多年前，几处各自出现	农业起源；牛拉的犁约在公元前四千年前后出现于美索不达米亚一带〔待核〕（第11章）	封装了种植的经验；逼出了定居、仓储、记账与国家
十三、十四世纪	机械钟出现在欧洲〔待核〕（第12章）	封装了"时间是均匀的"；逼出了作息、工时与同一个"七点"
1712年前后	纽科门的大气式蒸汽机用于矿井排水（第13章）	封装了以热做功的办法；逼出了对热与功关系的追问
1760年代至1769年	瓦特对蒸汽机的重要改良（分离冷凝器，1769年取得专利）	让蒸汽机走出矿井，走进工厂与交通
1824年	卡诺发表关于热机效率的著作	先有机器，后有对机器的理解：热力学
1830年	利物浦至曼彻斯特铁路通车	速度快过了各地自己的时间，逼出统一时刻表
1844年	美国首条实用电报线路开通（华盛顿至巴尔的摩）	信号脱离人和马的速度，逼出"同一时刻"的问题

时间	工具或事件	意义（封装了什么，逼出了什么）
1882年	纽约珍珠街电站投入运行（第14章）	电力开始成为可输送的公共能源
1883年	北美铁路推行标准时区	地方时让位于标准时
20世纪上半叶	电动洗衣机逐渐普及〔待核〕；考恩1983年的研究指出家务时间并未如预期减少	封装了洗衣的体力活；逼出了更高的清洁标准与更多的家务
1945至1946年前后	早期电子计算机问世，如ENIAC（1945年建成，1946年2月公开）（第15章）	封装了大批量计算；逼出了编程这个新职业
1968年	"软件危机"一词在北约的一次会议上被提出	计算能力的提升，逼出了软件如何做、如何管
1969年	阿帕网首批节点联网	网络的雏形，后来逼出信息、安全、隐私的新问题
1989至1991年前后	万维网诞生	信息获取人人可及，逼出信息过载与辨别
1996至1997年	深蓝1996年2月负于、1997年5月胜卡斯帕罗夫（第37章）	封闭领域被永久越过；人和人之间的象棋仍在发生

◆ 理念界

时间	工具或事件	意义
约公元前3000年前后	美索不达米亚最早的文字体系〔待核〕（第16章）	封装了记数与记事；逼出了抄写、教育、法律
约公元前四世纪	亚里士多德的逻辑学〔待核〕（第18章）	推理的形式成为可以检验的工具
约公元前300年	欧几里得《几何原本》〔待核〕	公理加证明，成为理论这种工具的典范
628年前后	婆罗摩笈多对零的运算规则作出较早的明确说明（第17章）	零作为数，使位值记数与代数得以展开
1494年	帕乔利出版著作，系统介绍复式记账	逼出了会计、审计、公司的一系列制度

时间	工具或事件	意义
十九世纪前半	罗巴切夫斯基、波约伊各自发展非欧几何（第18章）	两千年的使用逼出另一种几何，后来成了广义相对论的数学工具
1654年前后	帕斯卡与费马就赌局问题通信，概率论起步（第20章）	"不确定"成为可计算的对象，逼出保险、统计
17世纪后半叶至19世纪	牛顿与莱布尼茨各自发展微积分；柯西、魏尔斯特拉斯推进严格化（第19章）	封装了变化的计算，留下基础不清的问题；使用又逼出对"极限"的严格定义
1901年前后	罗素发现集合论中的悖论	逻辑与集合工具的使用，逼出数学基础的危机
1931年	哥德尔发表不完备定理	形式系统的力量与限度被同时划出
2024至2025年	有团队宣布系统在国际数学奥林匹克题目上达到银牌、金牌水平〔待核〕（第21章，只作线索）	解题被封装得更深，问题推成"你发生了哪些新数学"

自我界

时间	工具或事件	意义
十六世纪前后	威尼斯玻璃镜的普及（第22章）	封装了"看见自己"；逼出对外貌、自我形象的新意识
四世纪末	奥古斯丁《忏悔录》，早期自传的重要例子（第23章）	把内心经历写成可传递的文本
二十世纪七十年代末	卡巴金在医学院开设正念减压课程（第24章）	古老修习的一部分被改造成脱离宗教的工具；逼出"正念被滥用"的新问题
1879年前后	冯特在莱比锡建立心理学实验室，通常被视为第一个（第25章）	"自我如何运作"成为可以实验的对象
1899至1900年	弗洛伊德《梦的解析》（1899年底出版，扉页标1900年）	自我认识有了新的盲区：人不完全了解自己的动机
2004年、2012年前后	脸书创办；微信上线"朋友圈"（朋友圈的年份〔待核〕）（第26章）	自我的呈现被即时记录与分享，逼出新的自我管理问题

读这张表要留意两点。第一，同一件工具常常同时出现在不止一界：镜子本是现实界的玻璃，却主要改变了自我界；时钟本是现实界的机械，却逼出了理念界的同时性问题。第二，表里没有"终点"一栏，因为本书的论证之一就是：每一个条目封装完成的那一刻，都是下一次发生的起点。

附录B　十条承重命题汇总

每条命题后的括号，指出它主要在哪些章里展开。

1. **工具是已发生智慧的封装**。一件工具里装着前人的一次次发生；用工具的人，是在借用这些发生。（第5章）
2. **使用即碰撞，碰撞必生新SDE**。工具一旦进入真实的使用，就与使用者的处境、需要、其他工具相撞，撞出封装时没有的结构、差异与纠缠。（第6章）
3. **错位定理：新SDE必在封装之外，因为封装时它还不存在**。封装发生在先，碰撞发生在后，所以新的东西不可能已经在封装里。（第7章）
4. **一界有工具，三界有新生**。工具从不只停在出生的那一界；现实界的工具，会在理念界和自我界逼出新的发生。（第27、28章）
5. **被封装能力的退场是让渡，不是退化**。人把一部分能力交给工具，是为了腾出手去发生别的。（第9章、第29章）
6. **创造力与智慧共生，智慧不是先在的库存**。创造是智慧的发生，不是智慧的支出；智慧不是存下来的答案，而是被答案改造过的提问方式。（第31至34章）
7. **人类水平是过程，不是高度；横杆里含着追赶者**。所谓"人类水平"随新产业而升高，而新产业本身含着AI这个追赶者。（第35、36章）
8. **封闭领域可被永久越过，开放领域不可**。象棋之类规则、目标、评价都已划定的领域，AI可以永久超过人；"通用"意味着开放，开放领域没有终局。（第37、38章）
9. **基础性本身在发生；AI制造了自己解决不了的基础问题**。基础问题不是一份固定清单，它随着工具的发展而增长。（第40、41章）
10. **AGI感是维度未配平时期的时间差现象；强弱与碰撞深浅成反比**。封装已完成而新SDE尚未显露的等待期里，人们把看不见误当成没有；一个人与工具碰得

越深，越早看见封装之外的东西，AGI感就越弱。（第43至47章，"强弱与碰撞深浅成反比"主要见第45章）

另有一个被明说的前提，不在十条之内：**问题总是对某个主体而言的问题；E的源头在主体**。它是前提，不是推出的结论（第34章、第38章、第42章）。

附录C　可错预言

一个论证若不能被事实反驳，就没有资格要求别人相信。下面是本书的三条可错预言，每条都写明"怎样算被证伪"。对"数年内"这样的时间，本书不给具体年份和数字，因为那样的数字没有依据；预言因此需要靠长期观察来检验，这是它的局限。

预言一：任何开放领域在AI能力大幅跃升后的数年内，会出现由这次跃升本身引发的新职业或新问题类别。（见第4章、第36至38章；第37、38章写明的自我证伪条件，是它的另一种说法）

怎样算被证伪：选定某个开放领域（不是象棋那样规则固定的领域），在其中出现一次公认的AI能力大幅跃升之后，经过一段较长的观察期（例如一代人的职业周期的相当一部分），如果该领域里没有出现任何由这次跃升直接引发的新职业、新岗位类型或新的待解问题类别，反而整体上只是旧工作被替换、没有新的出现，则这条预言被证伪。需要注意：新问题要是"由这次跃升本身引发"的，不能是原先就存在的问题换了个名字。

预言二：社会上的AGI感在扩大期最强，进入内卷期后明显下降。（见第46章）

怎样算被证伪：以公开讨论的热度、对"通用智能已至"的态度调查、从业者的自我评估等可比的指标，追踪某项AI能力从大幅提升到被广泛使用的全过程。如果观察到AGI感在内卷期反而持续上升，或者在扩大期并不比后期强，则这条预言被证伪。这里的难点是指标的选取与"扩大期""内卷期"的划分，使用者需要在观察之前就公布划分标准，不能事后挑选。

预言三：无论哪一次宣称"AGI已经实现"，此后被公认的前沿基础问题清单只会变长，不会变短。（见第40、41章）

怎样算被证伪：在任何一次被广泛接受为"AGI已经实现"的宣称之后，如果学界公认的前沿基础问题清单（例如关于对齐、可解释性、意识、可计算性等方面的未解问题）在一段较长的时间里持续变短，并且是因为这些问题被解决而不是被遗忘、被改名，则这条预言被证伪。

第七、八编补充的三条可检验的判断（不属于全书三条核心预言，但与这两编直接相关，正文里各有对应的表述）：

补充一：只要系统的训练与使用保持分离，它在没有答案的新局面中的创造性贡献，就总需要使用它的人来承担"定题"这一环。（第34章）怎样算被证伪：有一个这样的系统，在训练与使用分离的状态下，长期、独立地提出并维持有意义的新问题类别，而且这些问题得到了使用者以外的人的持续认可。

补充二：若出现能在使用中持续改变自身的系统，它所面对的"问题"仍然最终溯源到某个主体的需要。（第34章，与第42章所说的前提相连）怎样算被证伪：能够证明某个这样的系统，其问题的来源完全不依赖于任何人的需要，并且这些问题对该系统自身的"存续"有可说明的意义。需要说明，这一条涉及前提，因此它被证伪时，受影响的是全书中依赖该前提的部分，不是全部论证。

补充三：封闭领域的永久越过，不会使人与人之间的水平停滞。（第37章）怎样算被证伪：在某个封闭领域被AI永久越过之后，如果人与人之间的水平指标持续停滞或下降、参与度持续萎缩，并且可以追溯到这次越过本身，则这条判断被证伪。

附录D　术语表

按在书中出现的大致顺序排列，只写本书的用法。

AGI（通用人工智能）：本书区分三种用法。**终局AGI**：一种能解决一切问题、包括由它自己的存在所引发的新问题、从此不再需要升级的智能，本书论证它自我否定（第1章）。**人类水平AGI**：在各种任务上达到人类水平的AI，其论证见第八编。**AGI感**：人们在工具刚出现、封装已完成而新的发生尚未显露的一段时间里，生出的"通用智能已经来了"的感觉。

类AGI能力：阶段性地在许多任务上与人不相上下、甚至更好的能力。它可以出现，但一出现就进入使用，不再稳定地被称为"通用"。

SDE：Structure–Difference–Entanglement，结构、差异、纠缠。本书的读法是：S为显露，发生留下的、稳定的、看得见的样子，结构是显露的稳定态；D为差异，一处与另一处的不同，以及由不同带来的发生过程；E为纠缠，一件事与周围的人、物、问题、需要缠在一起所形成的场，是意义的土壤。

三界：现实界、理念界、自我界。现实界是外部世界的事物与工具；理念界是概念、理论、符号；自我界是人对自己的认识与塑造。

封装：把已发生的智慧压缩成可以传递、可以重复使用的形态，如工具、配方、理论、制度。

碰撞：工具进入使用之后，与使用者的处境、需要及其他工具相互作用的过程。

错位：封装在先、新发生在后所造成的时间上的错开。新的结构不可能在封装之时就已经在封装里。（即命题三，错位定理）

二阶碰撞：人们对一阶碰撞结果的回应（新的补丁、规矩、封装），再与更广的现场相撞，生出下一轮新SDE。（第8章）

升级律：每一次二阶碰撞，都要求比原来那一级更高的智慧来回应。（第8章）

让渡：人把一部分能力交给工具、让它退出人的日常，腾出手去做别的事。若腾出的力气被用来发生新的东西，是让渡；若只是荒废，才是退化。

工具让渡律：工具经历四个阶段的规律。扩大期：使用范围迅速扩大、红利明显；内卷期：大家都用，优势被抹平，竞争在原地加剧；新产业诞生：在工具带来的新问题里，长出以前没有的行业；让渡：被封装的能力退出前台，成为基础设施。（第9章）

等待期：封装已完成、新SDE尚未显露的这段时间。

库存图景：把智慧看作先在的存货、把创造看作从中取货重组的一幅图景。

共生图景：把创造力与智慧看作一起发生、没有先在仓库的一幅图景。智慧在这幅图景里，是被答案改造过的提问方式。

共生：创造与智慧相互生成：创造一次，智慧多一块；智慧多一块，下一次的创造不同。智慧是创造沉淀下来的显露，带着出处。

判据：第34章提出的判断，即创造的时候，创造者自己有没有被改变。它是必要条件的筛子，不是充分条件的印章。

训练与使用分离：当前常见的一种技术结构：训练时系统参数变化，使用时参数不变。这是当前技术状态，不是原理必然。

定题：提出"哪件事值得当成问题来问"并认下结果好坏的那一环。

横杆：比喻"人类水平"这个所谓的高度。本书认为它不是固定高度，而是含着追赶者的发生过程，会随新产业升高。

封闭领域与开放领域：规则、目标、评价都事先划定、使用不生出新的问题类别的领域为封闭领域；使用会生出新的问题类别、目标和评价随使用被改写的领域为开放领域。

前提：本书明确立为前提、不加证明的断言，主要是"问题总是对某个主体而言的问题（E的源头在主体）"。第39章结构证明里的"三个前提"是另一回事，它们都在前面各编里论证过。